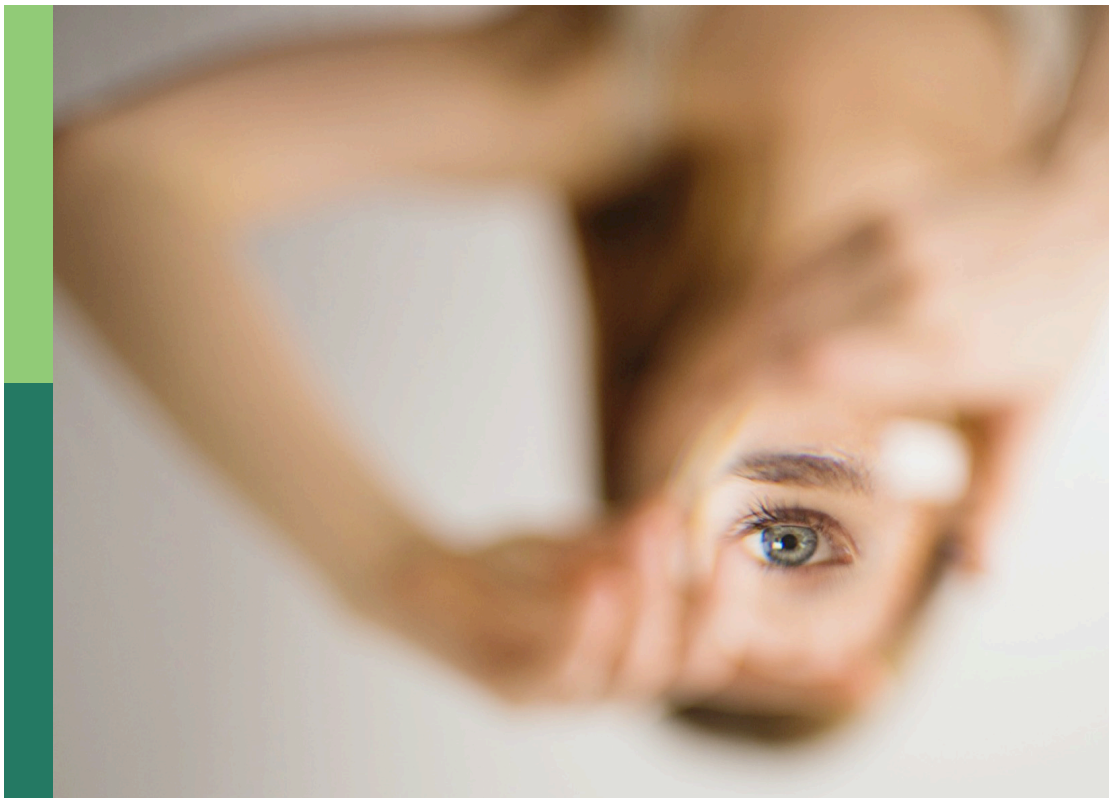
frontiers | Research Topics

Emotion recognition using brain-computer interfaces and advanced artificial intelligence

Edited by
Yizhang Jiang, Maozhen Li, Pengjiang William Qian and Yaoru Sun

Published in
Frontiers in Psychology



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-83250-517-5
DOI 10.3389/978-2-83250-517-5

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Emotion recognition using brain-computer interfaces and advanced artificial intelligence

Topic editors

Yizhang Jiang — Jiangnan University, China
Maozhen Li — Brunel University London, United Kingdom
Pengjiang William Qian — Jiangnan University, China
Yaoru Sun — Tongji University, China

Citation

Jiang, Y., Li, M., Qian, P. W., Sun, Y., eds. (2023). *Emotion recognition using brain-computer interfaces and advanced artificial intelligence*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-517-5

Table of contents

06 Home Textile Pattern Emotion Labeling Using Deep Multi-View Feature Learning
Juan Yang and Yuanpeng Zhang

13 A Novel User Emotional Interaction Design Model Using Long and Short-Term Memory Networks and Deep Learning
Xiang Chen, Rubing Huang, Xin Li, Lei Xiao, Ming Zhou and Linghao Zhang

26 Affective Voice Interaction and Artificial Intelligence: A Research Study on the Acoustic Features of Gender and the Emotional States of the PAD Model
Kuo-Liang Huang, Sheng-Feng Duan and Xi Lyu

42 Optimized Projection and Fisher Discriminative Dictionary Learning for EEG Emotion Recognition
Xiaoqing Gu, Yiqing Fan, Jie Zhou and Jiaqun Zhu

53 A Novel Paradigm to Design Personalized Derived Images of Art Paintings Using an Intelligent Emotional Analysis Model
Yingjing Duan, Jie Zhang and Xiaoqing Gu

67 Research on Emotion Analysis of Chinese Literati Painting Images Based on Deep Learning
Jie Zhang, Yingjing Duan and Xiaoqing Gu

80 A Domain Adaptation Sparse Representation Classifier for Cross-Domain Electroencephalogram-Based Emotion Classification
Tongguang Ni, Yuyao Ni, Jing Xue and Suhong Wang

91 Emotion Analysis of Personalized Color of Beauty Package Based on NCD Image Positioning
Feng Wang, Jifeng Xu, Hanning Zhang and Jun Yin

103 Motivation, Social Emotion, and the Acceptance of Artificial Intelligence Virtual Assistants—Trust-Based Mediating Effects
Shiying Zhang, Zixuan Meng, Beibei Chen, Xiu Yang and Xinran Zhao

113 Erratum: Motivation, Social Emotion, and the Acceptance of Artificial Intelligence Virtual Assistants—Trust-Based Mediating Effects
Frontiers Production Office

114 Visual Expression of Emotion in Dynamic 3D Painting System Based on Emotion Synthesis Model
Shenghe Cheng

124 New Technologies' Commercialization: The Roles of the Leader's Emotion and Incubation Support
Biaoan Shan, Yi Pu, Biao Chen and Shan Lu

133 Research on Migrant Works' Concern Recognition and Emotion Analysis Based on Web Text Data
Zhijie Dou, Zixuan Cheng and Dongmei Huang

143 Relationships Among CEO Narcissism, Debt Financing and Firm Innovation Performance: Emotion Recognition Using Advanced Artificial Intelligence
Lan Zhang, Biming Liang, Datian Bi, Yuan Zhou and Xiaohan Yu

155 Dynamic Evolution Mechanism of Digital Entrepreneurship Ecosystem Based on Text Sentiment Computing Analysis
Jiahui Li and Meifang Yao

169 Facial Expression Emotion Recognition Model Integrating Philosophy and Machine Learning Theory
Zhenjie Song

178 Sentiment Classification of News Text Data Using Intelligent Model
Shitao Zhang

187 A Novel Music Emotion Recognition Model Using Neural Network Technology
Jing Yang

196 Research on Sentiment Analysis and Satisfaction Evaluation of Online Teaching in Universities During Epidemic Prevention
Jing Zhang

203 Emotion Recognition of Chinese Paintings at the Thirteenth National Exhibition of Fines Arts in China Based on Advanced Affective Computing
Jing Li, Dongliang Chen, Ning Yu, Ziping Zhao and Zhihan Lv

217 MIFAD-Net: Multi-Layer Interactive Feature Fusion Network With Angular Distance Loss for Face Emotion Recognition
Weiwei Cai, Ming Gao, Runmin Liu and Jie Mao

228 Research on the Emotions Based on Brain-Computer Technology: A Bibliometric Analysis and Research Agenda
Wei Yan, Xiaoju Liu, Biaoan Shan, Xiangxian Zhang and Yi Pu

240 Foreign Language Teachers' Emotion Recognition in College Oral English Classroom Teaching
Yanyun Dai

247 Multimodal Art Pose Recognition and Interaction With Human Intelligence Enhancement
Chengming Ma, Qian Liu and Yaqi Dang

260 Emotion Recognition of Foreign Language Teachers in College English Classroom Teaching
Lei Li

267 A Study on the Construction of Emotion Recognition Based on Multimodal Information Fusion in English Learning Cooperative and Competitive Mode
Haihua Tu

283 Evaluation of the Virtual Economic Effect of Tourism Product Emotional Marketing Based on Virtual Reality
Manati Talafubieke, Sheng Mai and Nuermila Xialifuhan

294 An EEG Neurofeedback Interactive Model for Emotional Classification of Electronic Music Compositions Considering Multi-Brain Synergistic Brain-Computer Interfaces
Mingxing Liu

305 Position-Enhanced Multi-Head Self-Attention Based Bidirectional Gated Recurrent Unit for Aspect-Level Sentiment Classification
Xianyong Li, Li Ding, Yajun Du, Yongquan Fan and Fashan Shen

316 Exploring the Domain of Emotional Intelligence in Organizations: Bibliometrics, Content Analyses, Framework Development, and Research Agenda
Baobao Dong, Xing Peng and Na Jiang

335 A Study on the Cognition and Emotion Identification of Participative Budgeting Based on Artificial Intelligence
Yuan Zhou, Tianjiao Zhang, Lan Zhang, Zhaoxin Xue, Mingxu Bao and Lingbing Liu

347 Research on Digital Business Model Innovation Based on Emotion Regulation Lens
Shan Lu and Haijing Yu

357 Cross-Modal Transfer Learning From EEG to Functional Near-Infrared Spectroscopy for Classification Task in Brain-Computer Interface System
Yuqing Wang, Zhiqiang Yang, Hongfei Ji, Jie Li, Lingyu Liu and Jie Zhuang

370 Managing Students' Creativity in Music Education – The Mediating Role of Frustration Tolerance and Moderating Role of Emotion Regulation
Lei Wang and Na Jiang

381 Using Facial Micro-Expressions in Combination With EEG and Physiological Signals for Emotion Recognition
Nastaran Saffaryazdi, Syed Talal Wasim, Kuldeep Dileep, Alireza Farrokhi Nia, Suranga Nanayakkara, Elizabeth Broadbent and Mark Billinghurst

404 An Emotional Analysis Method for the Analysis of Cognitive and Psychological Factors in the Change of Second Language Learning Model of Chinese Mainland Students in the Post-epidemic Era
Gang Xie and Xiaona Wang

frontiers in Psychology

ORIGINAL RESEARCH
published: 19 April 2021
doi: 10.3389/fpsyg.2021.666074

Home Textile Pattern Emotion Labeling Using Deep Multi-View Feature Learning

Juan Yang[1] *and Yuanpeng Zhang*[2*]

[1] School of Textile and Clothing, Nantong University, Nantong, China, [2] Department of Medical Informatics, Nantong University, Nantong, China

Different home textile patterns have different emotional expressions. Emotion evaluation of home textile patterns can effectively improve the retrieval performance of home textile patterns based on semantics. It can not only help designers make full use of existing designs and stimulate creative inspiration but also help users select designs and products that are more in line with their needs. In this study, we develop a three-stage framework for home textile pattern emotion labeling based on artificial intelligence. To be specific, first of all, three kinds of aesthetic features, i.e., shape, texture, and salient region, are extracted from the original home textile patterns. Then, a CNN (convolutional neural network)-based deep feature extractor is constructed to extract deep features from the aesthetic features acquired in the previous stage. Finally, a novel multi-view classifier is designed to label home textile patterns that can automatically learn the weight of each view. The three-stage framework is evaluated by our data and the experimental results show its promising performance in home textile patterns labeling.

Keywords: home textile pattern, emotion labeling, deep learning, multi-view learning, feature selection

OPEN ACCESS

Edited by:
Yaoru Sun,
Tongji University, China

Reviewed by:
Chenxi Huang,
Xiamen University, China
Yuexin Li,
Hubei University, China

****Correspondence:***
Yuanpeng Zhang
maxbirdzhang@ntu.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *09 February 2021*
Accepted: *01 March 2021*
Published: *19 April 2021*

Citation:
Yang J and Zhang Y (2021) Home Textile Pattern Emotion Labeling Using Deep Multi-View Feature Learning. Front. Psychol. 12:666074. doi: 10.3389/fpsyg.2021.666074

INTRODUCTION

Emotion is the spiritual essence of home textile design. Fabric pattern is an important component of home textiles, which contains rich emotional information, including aesthetics and values. Therefore, fabric patterns rich in connotation and emotion are more and more respected by designers, which can meet the multiple needs of consumers. However, the pattern materials in home textile design and production are increasing day by day, and there are tens of thousands of patterns in the sample database. It is difficult for designers to make full use of the existing rich fabric patterns for home textile design. Therefore, how to integrate the objective characteristics (color, texture, pattern, etc.) and perceptual experience of fabric patterns into a mathematical model for aesthetic evaluation, emotional classification and retrieval, and emotion labeling of fabric patterns is one of the important topics for computer vision and textile design researchers.

With the continuous development of computer science, AI (artificial intelligence) and CV (computer vision) provide ideas and methods to solve this problem. Gan et al. (2014) made use of deep self-taught learning to obtain hierarchical representations, learn the concept of facial beauty, and produce human-like predictor. Datta and Wang (2010) established the first image aesthetics scoring website ACQUINE in 2010. Although the accuracy of the evaluation results is not high, it has shown that calculable aesthetics is feasible. Li and Chen (2009) adopted the features of color and composition in artistic aesthetic features to realize the classification of high and low aesthetic sense of paintings, and achieved a classification accuracy of more than 70%.

Lo et al. (2012) studied image aesthetic classification from the aesthetic perspective of image color, layout, edge, and other features, and the results showed that image aesthetic features could be used for image sentiment analysis. With the development of deep learning, many deep learning based methods are also used for image aesthetic classification. Lu et al. (2015) considered both the local and global perspectives of images, designed the CNN (convolutional neural network) model for feature learning and classifier training, and evaluated the aesthetic quality of 250,000 images from AVA database. Compared with the traditional manual feature methods, this method has a great improvement in classification performance. Dong et al. (2015) used the CNN model trained by ImageNet large physical classification database to extract image features and classify high and low aesthetic values. The effectiveness of this method was proved in two image quality evaluation data sets. The above CNN-based aesthetic classification methods all use the pixel values of sample images in the large database as the input of CNN, without integrating the existing mature manual features. It is often very difficult to obtain a large number of home textile design patterns. Using relatively small and limited image aesthetic data sets for training will easily to lead to problems of overfitting and unstable convergence.

In this study, to achieve home textile emotion labeling, we propose a multi-view learning framework that contains three main components. The first component is used to extract initial multi-view features from the shape, texture, and salient region perspectives. The second component is used to extract deep features from the initial multi-view features by CNN. The last component is used to collaboratively learn from multi-view deep features.

DATA AND METHODS

Data

We employed 20 students from the School of Textile and Clothing of Nantong University to collect 5000 home textile patterns from the Internet. All images were re-sized into 256*256 for further use. We also invited another 10 students to conduct subjective aesthetic evaluation on these collected patterns. Students calibrated all the images from three aspects of high aesthetic feeling, low aesthetic feeling, and uncertainty. When the emotion evaluation labels with eight or more participants were the same, this label was determined as the final label of the pattern. After removing the uncertain labels, we finally obtained labeled 4,480 patterns. Sample images and the number of sample images of high and low aesthetic are shown in **Figure 1**.

We used image clipping and horizontal flipping to enhance the patterns, and intercepted 4 corners and the middle image for mirror processing. After data enhancement, there are 25,000 patterns in each class (high aesthetic patterns and low aesthetic patterns).

Ethics

The studies involving human participants were reviewed and approved by ethics committee of Nantong University.

Written informed consent to participate in this study was provided by the participants.

Methods

In this study, we construct a three-stage method for home textile emotion labeling. The first stage is to extract initial shape features, texture features, and salient region features from home textile patterns. The second stage is a CNN-based feature extractor that is used to extract deep features from different aesthetic views from original home textile patterns. With deep features, in the second stage, we design a multi-view classifier to realize emotion labeling. The three-stage framework of emotion labeling is shown in **Figure 2**.

Acquisition of Initial Multi-View Features

As shown in **Figure 2**, the initial aesthetic features of home textile contain shape features, texture features, and salient region features. Shape is one of the main characters of design pattern, which can be described by the edge. The edge of an image is a collection of points where the gray value is discontinuous or the gray value changes dramatically. In this study, we use the Sobel operator (Gao et al., 2010) to detect the edge of an image. Texture is a very important content in fabric pattern, which contains many aesthetic features that affect the sense of beauty. The frequency and direction of Gabor filter (Mehrotra et al., 1992) are similar to that of human visual system, which is suitable for image texture feature description. Therefore, in this study, Gabor features are used to represent texture features. The saliency region of an image is the region that attracts the most visual attention and has a more significant influence on the aesthetic feeling of the image. The saliency value of a pixel is defined by the contrast between the pixel and other pixel in the image. Pixels of the same color have the same salience. In this study, we use the LC (Luminance Contrast) (Zhai and Shah, 2006) algorithm to extract salience region features. **Figure 3** illustrates an example of initial multi-view features.

Deep Multi-View Feature Learning

In this study, CNN (Sainath et al., 2013; Kalchbrenner et al., 2014; O'Shea and Nash, 2015; Albawi et al., 2017) is employed to extract deep features from home textile patterns from different views. The initial features in shape, texture, and salience region are first constructed following the methods discussed in section 'Acquisition of Initial Multi-View Features'. The network structure of our deep feature extractor is shown in **Figure 4**.

(1) Input layer: in this layer, we reduce each original three-channel image of 256*256*3 size to 10 images of 224*224*3 size by random cropping.
(2) The first convolution layer (Con-1): in this layer, the input images are reduced into 55*55 feature maps by 48 convolutional kernels (the kernel size is 11*11 and the step size is 4). Because the response result of ReLU (Rectified Linear Units) is unbounded (it can be very large), normalization is required. Here, LRN (Local Response Normalization) (Robinson et al., 2007) is used to perform local response normalization.

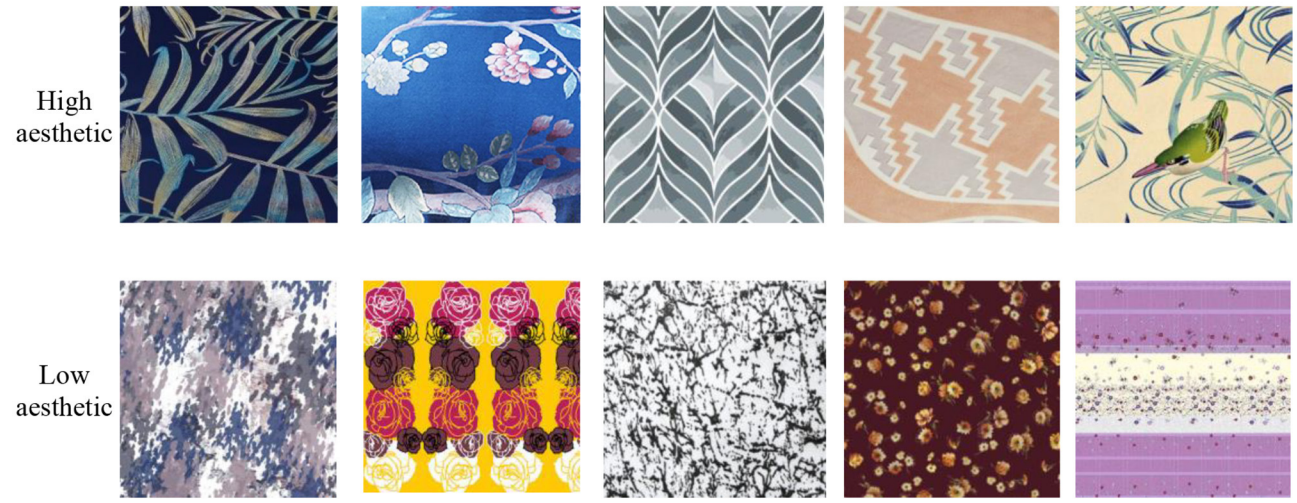


FIGURE 1 | An example of high and low aesthetic home textile patterns.

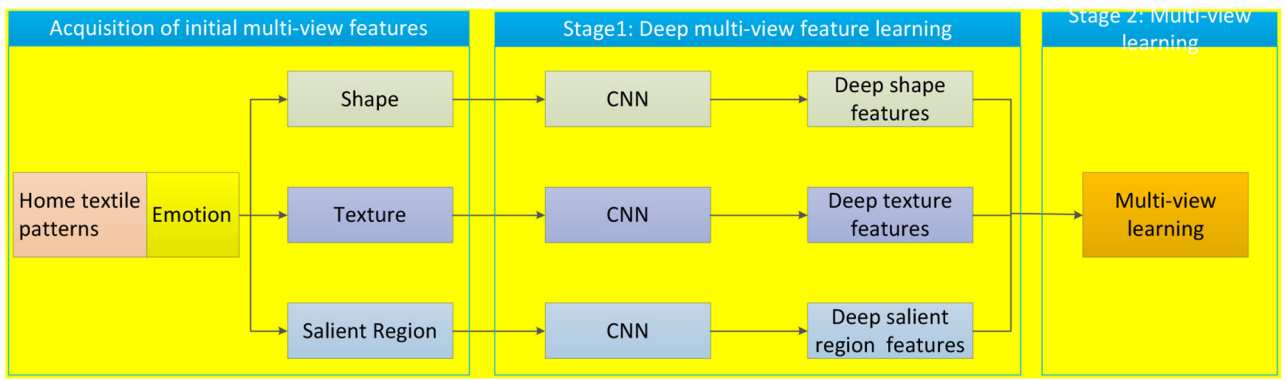


FIGURE 2 | Three-stage framework of emotion labeling.

(3) The second convolution layer (Con-2): in this layer, we use 128 convolutional kernels (the kernel size is 5*5 and the step size is 2) to further extract features from the 48 feature maps (the kernel size is 27*27) generated by the last layer.
(4) The third convolution layer (Con-3): in this layer, we use 192 convolutional kernels (the kernel size is 3*3 and the step size is 1) to generate 192 feature maps (13*13).

Our feature extractor uses the final output to calculate the approximation error and performs back propagation to update the network parameters during training. When the training process is done, the output of the penultimate layer is selected as the deep features learned by our feature extractor. Therefore, the deep features we extracted not only have lower dimensionality than the original features but also possess better discrimination ability to enhance the generalizability of the subsequent classification model.

Multi-View Learning

In this section, we will develop a multi-view classifier (Jiang et al., 2016; Qian et al., 2018) for emotion labeling based on the deep multi-view features that are extracted by the CNN-based deep feature extractor shown in **Figure 2**. The basic idea is that the Shannon Entropy is introduced to the Ridge Regression (SERR) model to automatically identify the weight of each view. With view weights, the deep shape features, deep texture features, and deep salient region features are combined to achieve collaborate learning. Suppose X_k represents the training feature set of the kth view, Y represents the emotion label of the training set, then the objective function of the new classifier SERR is formulated as follows:

$$J = \sum_{k=1}^{K} \omega_k \|X_k W_k - Y\|^2 + \gamma \sum_{k=1}^{K} \|W_k\|^2 + \delta \sum_{k=1}^{K} \omega_k \ln \omega_k$$
$$s.t. \sum_{k=1}^{K} \omega_k = 1 \quad (1)$$

where ω_k represents the weight of the kth view, W_k represents the corresponding transformation matrix, and γ and δ are two parameters to control the penalization terms. The objective function in Equation (1) can be solved by introducing Lagrangian multipliers. To be specific, by introducing the Lagrangian multipliers τ, the corresponding Lagrangian function can be formulated as

$$L = \sum_{k=1}^{K} \omega_k \|X_k W_k - Y\|^2 + \gamma \sum_{k=1}^{K} \|W_k\|^2 + \delta \sum_{k=1}^{K} \omega_k \ln \omega_k + \tau \left(\sum_{k=1}^{K} \omega_k - 1 \right) \quad (2)$$

By setting $\frac{\partial L}{\partial W_k} = 0$ and $\frac{\partial L}{\partial \omega_k} = 0$, we have

$$W_k = \left(\omega_k (X_k)^T X_k + \gamma I \right)^{-1} (X_k)^T Y \quad (3)$$

$$\omega_k = \frac{\exp\left(- \|X_k W_k - Y\|^2 / \delta \right)}{\sum_{t=1}^{K} \exp\left(- \|X_t W_t - Y\|^2 / \delta \right)} \quad (4)$$

With W_k and ω_k, we can use the alternate iteration to search for the optimal solution. The detailed steps of home textile emotion labeling using deep multi-view feature learning are shown as in Algorithm 1.

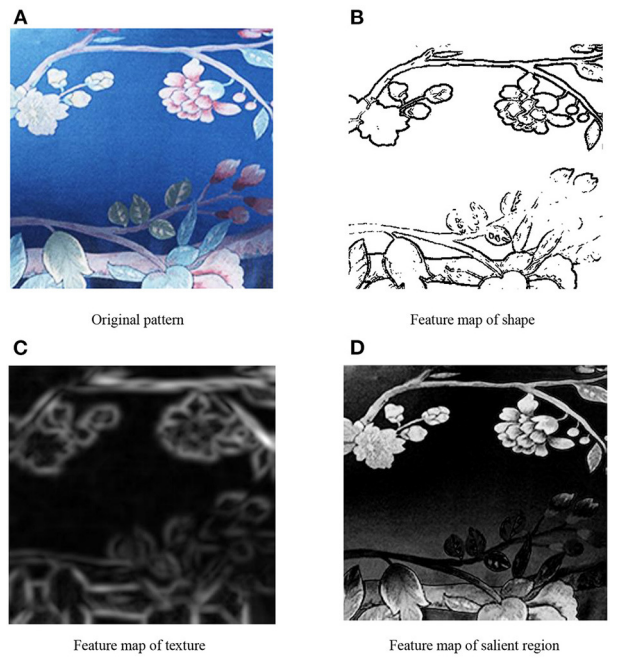


FIGURE 3 | An example of original multi-view features. **(A)** Original pattern, **(B)** Feature map of shape, **(C)** Feature map of texture, and **(D)** Feature map of salient region.

Algorithm 1: Model training

Input: original image set D, emotion label Y, parameters γ and δ
Output: transformation matrix of each view W_k, weight of each view ω_k

Procedure:
1. Use Sobel operator, Gabor filter, and LC to extract initial multi-view features.
2. Use deep feature extractor to extract deep multi-view features, X_k.
3. Randomly assign ω_k under $\sum_{k=1}^{K} \omega_k = 1$.
4. Set $t = 0$.
5. Repeat.
6. Use Equation (2) to update W_k based on current ω_k.
7. Use Equation (3) to update ω_k based on current W_k.
8. $t = t + 1$.
9. Until $(\|W_k(t+1) - W_k(t)\| < \varepsilon)$.

After the training procedure is done, for an unseen textile pattern, we use Algorithm 2 to perform emotion labeling.

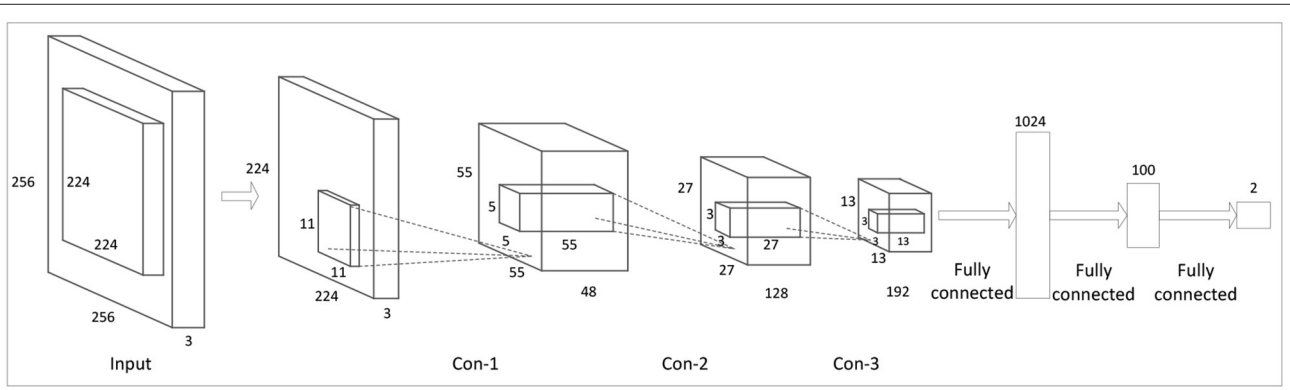


FIGURE 4 | CNN-based deep feature extractor.

Algorithm 2: Model testing

Input: unseen textile patterns, transformation matrix of each view W_k, weight of each view ω_k
Output: emotion labels of unseen textile patterns

Procedure:
1. Use the feature index obtained from training to select features from the unseen textile patterns.
2. Emotion of unseen textile patterns can be determined by

$$Y' = sign(\sum_{k=1}^{K} \omega_k X'_k W_k)$$

TABLE 1 | Settings of feature selection methods.

Feature selection methods	Settings
MRMR (Togaçar et al., 2020)	Use default setting recommended by Togaçar et al. (2020)
l_{21}-norm (Nie et al., 2010)	The regularized parameter is searched from [0.001, 10]
PCA (Karamizadeh et al., 2013)	Use default setting recommended by Karamizadeh et al. (2013)
Relief (Urbanowicz et al., 2018)	Use default setting recommended by Urbanowicz et al. (2018)

RESULTS

In this section, we will evaluate our new emotion labeling method from two perspectives, i.e., the effectiveness of the deep feature extractor and the effectiveness of the multi-view classifier.

Settings

To evaluate the effectiveness of the deep feature extractor, we first reshape the original shape features, texture features, and the saliency region features from two-dimensional matrices to one-dimensional vectors and then introduce several traditional feature selection methods for discriminant feature selection. The settings of the introduced feature selection methods are shown in **Table 1**.

Additionally, to evaluate the effectiveness of the multi-view classifier SERR we proposed, we directly concatenate all features from different views and use classic classifiers SVM (Support Vector Machine), KNN (K-Nearest Neighbor), NB (Naive Bayes), and DT (Decision Tree) for classification. The settings of the introduced feature selection methods are shown in **Table 2**.

All experiments are conducted on a PC with Intel® Core™ i7-9700 @3.00 GHz Dual, 32G memory, and RTX 2080 Ti. The coding platform is Matlab 2012b.

Experimental Results

In this section, we will report our experimental results from two aspects. First of all, our deep feature extractor is used to extract deep features from each initial view. For comparison studies, four commonly used feature selection methods MRMR, l_{21}-norm, PCA, and Relief are also introduced to select discriminant features from each initial view. The Ridge Regression model is taken as the classifier for the classification tasks in each view. **Tables 3–5** show the classification results in terms of Accuracy, Sensitivity, and Specificity on each view.

TABLE 2 | Settings of classifiers.

Classifiers	Setting
SVM (Joachims, 1998)	The Gaussian kernel is adopted. The kernel width is searched from [10^{-5}, 10^5], the center is searched from [10^{-5}, 10^5], and C is set to 100
KNN (Zhang et al., 2017)	K is set to 5
NB (Rish, 2001)	Use default setting recommended by Rish (2001)
DT (Myles et al., 2004)	Use default setting recommended by Myles et al. (2004)
SERR	γ and δ are searched from [0.001, 10]

Additionally, to highlight our proposed multi-view learning method SERR, with the deep features from different views, we introduce SVM, KNN, DT, and NB as classifiers for comparison studies. **Table 6** shows the classification results in terms of Accuracy, Sensitivity, and Specificity on each view.

DISCUSSION AND CONCLUSION

With the improvement of automation of home textile production and design and the increasing number of stored home textile pattern images in enterprises, the traditional retrieval methods can no longer meet the needs of home textile manufacturers. It is necessary to conduct aesthetic evaluation and emotional analysis of home textile pattern, so as to provide better services to enterprises and consumers.

Currently, there are two main ways for home textile enterprises to search home textile design patterns. One is to manually classify and number home textile patterns, which is mainly used for enterprise management. However, a large amount of management storage will cause a waste of resources in all aspects. The other is to establish an image retrieval system for the pre-classified home textile patterns, but it needs to manually classify each pattern. However, it requires human classification of each pattern, which is time-consuming and labor-intensive, and not all home textile patterns can be expressed with keywords or symbols, so it is difficult to meet the different retrieval needs of different searchers.

With the development of AI, especially deep learning, in this study, to achieve home textile emotion labeling, we propose a multi-view learning framework that contains three main components. The first component is used to extract initial multi-view features from the shape, texture, and salient region perspectives. The second component is used to extract deep features from the initial multi-view features by CNN. The last component is used to collaboratively learn from multi-view deep features. We demonstrate our method from two perspectives. From the results shown in **Tables 3–5**, we see that the features extracted by our deep

TABLE 3 | Classification performance in terms of accuracy.

Feature selection methods	Shape features	Texture features	Saliency region features
MRMR	0.9458 ± 0.0030	0.7781 ± 0.0028	0.7517 ± 0.0057
l_{21}-norm	0.9784 ± 0.0028	0.7558 ± 0.0047	0.7510 ± 0.0047
PCA	0.9555 ± 0.0036	0.7697 ± 0.0027	0.7478 ± 0.0092
Relief	0.9420 ± 0.0014	0.7784 ± 0.0015	0.7149 ± 0.0102
Deep feature extractor	0.9816 ± 0.0021	0.7870 ± 0.0099	0.7579 ± 0.0111

TABLE 4 | Classification performance in terms of sensitivity.

Feature selection methods	Shape features	Texture features	Saliency region features
MRMR	0.9478 ± 0.0023	0.5541 ± 0.0103	0.5147 ± 0.0101
l_{21}-norm	0.9578 ± 0.0030	0.5458 ± 0.0088	0.5368 ± 0.0117
PCA	0.9412 ± 0.0026	0.5269 ± 0.0201	0.5429 ± 0.0098
Relief	0.9541 ± 0.0026	0.5578 ± 0.0152	0.5025 ± 0.0100
Deep feature extractor	0.9612 ± 0.0037	0.5670 ± 0.0189	0.5435 ± 0.0125

TABLE 5 | Classification performance in terms of specificity.

Feature selection methods	Shape features	Texture features	Saliency region features
MRMR	0.9236 ± 0.0036	0.8436 ± 0.0056	0.8178 ± 0.0142
l_{21}-norm	0.9547 ± 0.0021	0.8268 ± 0.0017	0.8362 ± 0.0073
PCA	0.9632 ± 0.0054	0.8817 ± 0.0026	0.8555 ± 0.0089
Relief	0.9785 ± 0.0023	0.8557 ± 0.0053	0.8521 ± 0.0107
Deep feature extractor	0.9818 ± 0.0025	0.8929 ± 0.0081	0.8617 ± 0.0133

TABLE 6 | Classification performance of SERR, SVM, KNN, NB and DT.

Classifiers	Accuracy	Sensitivity	Specificity
SVM	0.9369 ± 0.0024	0.9257 ± 0.0016	0.9478 ± 0.0042
KNN	0.9478 ± 0.0025	0.9327 ± 0.0019	0.9087 ± 0.0015
NB	0.9368 ± 0.0074	0.9264 ± 0.0015	0.9457 ± 0.0014
DT	0.9698 ± 0.0025	0.9524 ± 0.0036	0.9644 ± 0.0025
SERR	0.9865 ± 0.0014	0.9782 ± 0.0045	0.9654 ± 0.0024

feature extractor drives the best classifier in each kind of features in terms of Accuracy, Sensitivity, and Specificity, respectively. This superiority indicates that deep features are more discriminant than the initial features obtained in the first stage. Additionally, from **Table 6**, we see that our proposed multi-view classifier SERR performs better than the traditional classifiers, SVM, KNN, DT, and NB, which means that collaborative learning in multiple feature space is more reliable than direct feature concatenation.

The experimental results show that the emotional labeling method proposed in this study realizes the emotional labeling of home textile patterns, provides an auxiliary retrieval method for consumers who want to buy home textile with certain emotional semantics, provides convenience for enterprises in the production and design of home textile patterns, and can meet the multiple needs of consumers. In our future work, we will consider more kinds of deep features and develop more deep feature extractors for emotion labeling.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because we will further enrich them. Requests to access the datasets should be directed to Yuanpeng Zhang, maxbirdzhang@ntu.edu.cn.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Nantong University. Written informed consent to participate in this study was provided by the participants.

AUTHOR CONTRIBUTIONS

JY and YZ designed the whole algorithm and experiments. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the Philosophy and Social Science Foundation of Jiangsu Province under No. 18YSC009.

ACKNOWLEDGMENTS

We thank the reviewers whose comments and suggestions helped improve this manuscript.

REFERENCES

Albawi, S., Mohammed, T. A., and Al-Zawi, S. (2017). "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)* (IEEE), 1–6. doi: 10.1109/ICEngTechnol.2017.8308186

Datta, R., and Wang, J. Z. (2010). *Mir 2010 - Proceedings of the 2010 ACM SIGMM International Conference on Multimedia Information*. Philadelphia, PA: Association for Computing Machinery (ACM), 421–424.

Dong, Z., Shen, X., and Li, H. (2015). *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Sydney, NSW: Springer Verlag, 524–535.

Gan, J., Li, L., Zhai, Y., and Liu, Y. (2014). Deep self-taught learning for facial beauty prediction. *Neurocomputing* 144, 295–303. doi: 10.1016/j.neucom.2014.05.028

Gao, W., Zhang, X., Yang, L., and Liu, H. (2010). "An improved Sobel edge detection," in *2010 3rd International Conference on Computer Science and Information Technology*. IEEE, 67–71.

Jiang, Y., Deng, Z., Chung, F. L., Wang, G., Qian, P., Choi, K. S., et al. (2016). Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system. *IEEE Trans. Fuzzy Syst.* 25, 3–20. doi: 10.1109/TFUZZ.2016.2637405

Joachims, T. (1998). *Making Large-Scale SVM Learning Practical*. Technical report, Technische Universität, Dortmund, Germany.

Kalchbrenner, N., Grefenstette, E., and Blunsom, P. (2014). A convolutional neural network for modelling sentences. *arXiv[Preprint].arXiv:1511.08458*. doi: 10.3115/v1/P14-1062

Karamizadeh, S., Abdullah, S. M., Manaf, A. A., Zamani, M., and Hooman, A. (2013). An overview of principal component analysis. *J. Sign. Inform. Process.* 4:173. doi: 10.4236/jsip.2013.43B031

Li, C., and Chen, T. (2009). Aesthetic visual quality assessment of paintings. *IEEE J. Select. Topics Sign. Process.* 3, 236–252. doi: 10.1109/JSTSP.2009.2015077

Lo, K. Y., Liu, K., Chen, C. S. (2012). *Proceedings - International Conference on Pattern Recognition*. Tsukuba: Institute of Electrical and Electronics Engineers, Inc, 2186–2189.

Lu, X., Lin, Z. L., Jin, H., Yang, J., and Wang, J. Z. (2015). RAPID: rating pictorial aesthetics using deep learning. *IEEE Trans. Multimedia* 17, 2021–2034. doi: 10.1109/TMM.2015.2477040

Mehrotra, R., Namuduri, K. R., and Ranganathan, N. (1992). Gabor filter-based edge detection. *Patt. Recogn.* 25, 1479–1494. doi: 10.1016/0031-3203(92)90121-X

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., and Brown, S. D. (2004). An introduction to decision tree modeling. *J. Chemometr.* 18, 275–285. doi: 10.1002/cem.873

Nie, F., Huang, H., Cai, X., and Ding, C. (2010). Efficient and robust feature selection via joint $\ell 2$, 1-norms minimization. *Adv. Neural Inform. Process. Syst.* 23, 1813–1821. doi: 10.1109/iccv.2013.146

O'Shea, K., and Nash, R. (2015). An introduction to convolutional neural networks. *arXiv[Preprint].arXiv:1404.2188*.

Qian, P., Zhou, J., Jiang, Y., Liang, F., Zhao, K., Wang, S., et al. (2018). Multi-view maximum entropy clustering by jointly leveraging inter-view collaborations and intra-view-weighted attributes. *IEEE Access* 6, 28594–28610. doi: 10.1109/ACCESS.2018.2825352

Rish, I. (2001). "An empirical study of the naive Bayes classifier," in *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence* (Santa Barbara, CA). Vol. 3, 41–46.

Robinson, A. E., Hammon, P. S., and de Sa, V. R. (2007). Explaining brightness illusions using spatial filtering and local response normalization. *Vis. Res.* 47, 1631–1644. doi: 10.1016/j.visres.2007.02.017

Sainath, T. N., Mohamed, A., and Kingsbury, B. (2013). "Deep convolutional neural networks for LVCSR.," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE), 8614–8618. doi: 10.1109/ICASSP.2013.6639347

Togaçar, M., Ergen, B., and Cömert, Z. (2020). A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models. *Irbm* 41, 212–222. doi: 10.1016/j.irbm.2019.10.006

Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., and Moore, J. H. (2018). Relief-based feature selection: introduction and review. *J. Biomed. Inform.* 85, 189–203. doi: 10.1016/j.jbi.2018.07.014

Zhai, Y., and Shah, M. (2006). "Visual attention detection in video sequences using spatiotemporal cues," in *Proceedings of the 14th ACM International Conference on Multimedia* (Santa Barbara, CA), 815–824. doi: 10.1145/1180639.1180824

Zhang, S., Li, X., Zong, M., Zhu, X., and Cheng, D. (2017). Learning k for knn classification. *ACM Trans. Intell. Syst. Technol.* 8, 1–19. doi: 10.1145/2990508

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Yang and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 20 April 2021
doi: 10.3389/fpsyg.2021.674853

A Novel User Emotional Interaction Design Model Using Long and Short-Term Memory Networks and Deep Learning

Xiang Chen[1], *Rubing Huang*[2], *Xin Li*[1], *Lei Xiao*[1], *Ming Zhou*[1] *and Linghao Zhang*[1]*

[1] School of Design, Jiangnan University, Wuxi, China, [2] Faculty of Information Technology, Macau University of Science and Technology, Macau, China

Emotional design is an important development trend of interaction design. Emotional design in products plays a key role in enhancing user experience and inducing user emotional resonance. In recent years, based on the user's emotional experience, the design concept of strengthening product emotional design has become a new direction for most designers to improve their design thinking. In the emotional interaction design, the machine needs to capture the user's key information in real time, recognize the user's emotional state, and use a variety of clues to finally determine the appropriate user model. Based on this background, this research uses a deep learning mechanism for more accurate and effective emotion recognition, thereby optimizing the design of the interactive system and improving the user experience. First of all, this research discusses how to use user characteristics such as speech, facial expression, video, heartbeat, etc., to make machines more accurately recognize human emotions. Through the analysis of various characteristics, the speech is selected as the experimental material. Second, a speech-based emotion recognition method is proposed. The mel-Frequency cepstral coefficient (MFCC) of the speech signal is used as the input of the improved long and short-term memory network (ILSTM). To ensure the integrity of the information and the accuracy of the output at the next moment, ILSTM makes peephole connections in the forget gate and input gate of LSTM, and adds the unit state as input data to the threshold layer. The emotional features obtained by ILSTM are input into the attention layer, and the self-attention mechanism is used to calculate the weight of each frame of speech signal. The speech features with higher weights are used to distinguish different emotions and complete the emotion recognition of the speech signal. Experiments on the EMO-DB and CASIA datasets verify the effectiveness of the model for emotion recognition. Finally, the feasibility of emotional interaction system design is discussed.

Keywords: interaction design, emotion recognition, LSTM, speech, self-attention mechanism

OPEN ACCESS

Edited by:
Yaoru Sun,
Tongji University, China

Reviewed by:
Yuexin Li,
Hubei University, China
Yufeng Yao,
Changshu Institute of Technology, China

***Correspondence:**
Linghao Zhang
linghaozhang20@126.com

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *02 March 2021*
Accepted: *18 March 2021*
Published: *20 April 2021*

Citation:
Chen X, Huang R, Li X, Xiao L, Zhou M and Zhang L (2021) A Novel User Emotional Interaction Design Model Using Long and Short-Term Memory Networks and Deep Learning. Front. Psychol. 12:674853. doi: 10.3389/fpsyg.2021.674853

INTRODUCTION

In the 19th century, some designers in Europe proposed that the fundamental purpose of design is to serve people. Design should be people-oriented and emphasize the importance of people in the composition of design. If the design only has a good appearance, but does not have practicality and functionality, then the design is not a qualified work. Especially in today's society, people have

begun to pursue spiritual pursuits. Designers must actively focus on user experience and emphasize user experience functions in design. Only in this way, design is what society needs and recognizes. Therefore, user experience has become an extremely important element, and it is the key that any designer should pay attention to. The relationship between user experience and interaction design is shown in **Figure 1**.

Good interaction design focuses on user experience. A good user experience is able to adapt to different types of users and the different emotions of each user. Therefore, emotion recognition is the key to good interaction design. The accuracy of emotion recognition is the top priority of interaction design. Emotion recognition is the process of using computers to automatically recognize human emotional states (Shu et al., 2018; Marco-Garcia et al., 2019; Ahmed et al., 2020; Assed et al., 2020; Nyquist and Luebbe, 2020). Due to the richness of human emotion expression, emotion recognition can use different combinations of multiple emotion expression materials to make judgments. In the historical process of emotion recognition research, scholars have researched from single-modal modeling (Edgar et al., 2020; Gosztolya, 2020; Panda et al., 2020), hybrid multi-modal modeling (Ayata et al., 2020; Zhang et al., 2020), to complex deep neural networks (Abdulsalam et al., 2019; Aouani and Ayed, 2020; Sharma et al., 2020). The relationship between human emotion and its multiple expressions is gradually being unearthed. There are many ways for humans to express emotions, and the carriers of emotions are also very rich, but the most intuitive among the many emotional carriers are language and expressions. If a person tends to use language to express their emotions, then the person's speech data may contain more emotion discrimination information. If a person tends to use facial expressions to express their emotions, then the person's video data may carry more emotion discrimination information.

The current emotion recognition is mainly developed from the following aspects. One is the different types of experimental data. Various forms of user physiological and behavioral data will be generated during the interaction. A variety of sensors are usually used in detecting emotional features, and the physiological state or behavior data of the user is captured without interpreting the input data. For example, a camera can capture facial expressions, body postures or gestures, a microphone can capture speech information, and a physiological sensor can directly measure physiological values such as skin temperature and heartbeat. Current popular researches mainly focus on physiological signals (Shu et al., 2018; Ayata et al., 2020), facial expressions (Martinez-Sanchez et al., 2017; Dey et al., 2019; Sivasangari et al., 2019), body postures (Bijlstra et al., 2019; Lenzoni et al., 2020) and speech information (Issa et al., 2020; Parthasarathy and Busso, 2020; Wang et al., 2020; Zhang et al., 2021). For different types of data, different types of features are extracted. Visual features are commonly used in sentiment analysis. Many studies have shown that the generation of human emotions is inseparable from what people see. For example, people often feel fear in a dark environment, and they will feel comfortable to see the symmetrical rules. Visual features include color features, texture features, SIFT features (Jayasimha and Reddy, 2019), HOG features (Bharate et al., 2019) and so on. Audio features include MFCC features (Sheikhan et al., 2012), Audio-Six features (Krishna et al., 2019). Attribute characteristics include Classemes characteristics (Torresani et al., 2010), Senti Bank characteristics (Borth et al., 2013). Second, the method of feature extraction is different. As emotion recognition uses different data types, the feature extraction methods for different types of data will also be different. Commonly used feature extraction methods are CNN (Brousseau et al., 2020; Cabada et al., 2020), LSTM (Wu et al., 2019; Zou et al., 2020), PCA (Smallman et al., 2020), etc. The third is the difference in classification models. Commonly used classification models are Support Vector Machine (SVM; Vapnik, 1999), Random Forest (RF; Svetnik et al., 2003), Hidden Markov Model (HMM; Lay et al., 2003), Gaussian Mixture Model (GMM; Navyasri et al., 2018), Artifical Neural Network (ANN; Powroznik, 2014) and the Soft-max regression classifier widely used in deep learning (Liu and Zheng, 2007).

Among the many researches on emotion recognition, the research on emotion recognition based on speech is the most popular. In speech-based emotion recognition, deep learning can be used not only for classification, but also for feature extraction. Deep learning has excellent performance in extracting discriminative speech emotional features, especially in the extraction of high-level emotional feature representations of large samples. Literature (Jiang et al., 2019) proposed a speech emotion recognition system based on recurrent neural network (RNN). The system uses a powerful learning method with bidirectional long and short-term memory (BLSTM) to extract high-level representations of emotional states from the time dynamics of emotional states, which improves recognition accuracy. Literature (Trigeorgis et al., 2016) proposed an end-to-end method that directly uses speech signals as input. Through the combination of convolutional neural network (CNN) and LSTM network, "context awareness" is realized. The performance of this method on the RECOLA database is better than traditional methods based on signal processing technology to extract features. Literature (Sak et al., 2014) uses LSTM for speech emotion recognition, LSTM converges fast and the model is small. Multiple units in the LSTM network form a tandem structure. The LSTM is created to avoid the long-term dependence of RNN affected by the state of a long time ago. In the three gating units, the calculation of each unit will lose the information processed in the previous sequence, and it is difficult to ensure the integrity of the information and the accuracy of the output at the next moment. Due to the low accuracy of the above methods in speech emotion recognition and the lack of information in the extracted speech signal features, this study uses a speech emotion recognition method based on LSTM and self-attention mechanism. This method uses MFCC as the emotional feature, uses LSTM to learn the temporal relevance of the speech sequence, and calculates the weight of each frame of speech signal in the emotional feature through the attention mechanism. Since this method combines the advantages of self-attention mechanism and LSTM, the accuracy and performance of emotion recognition will be improved. The main work of this article is listed as follows:

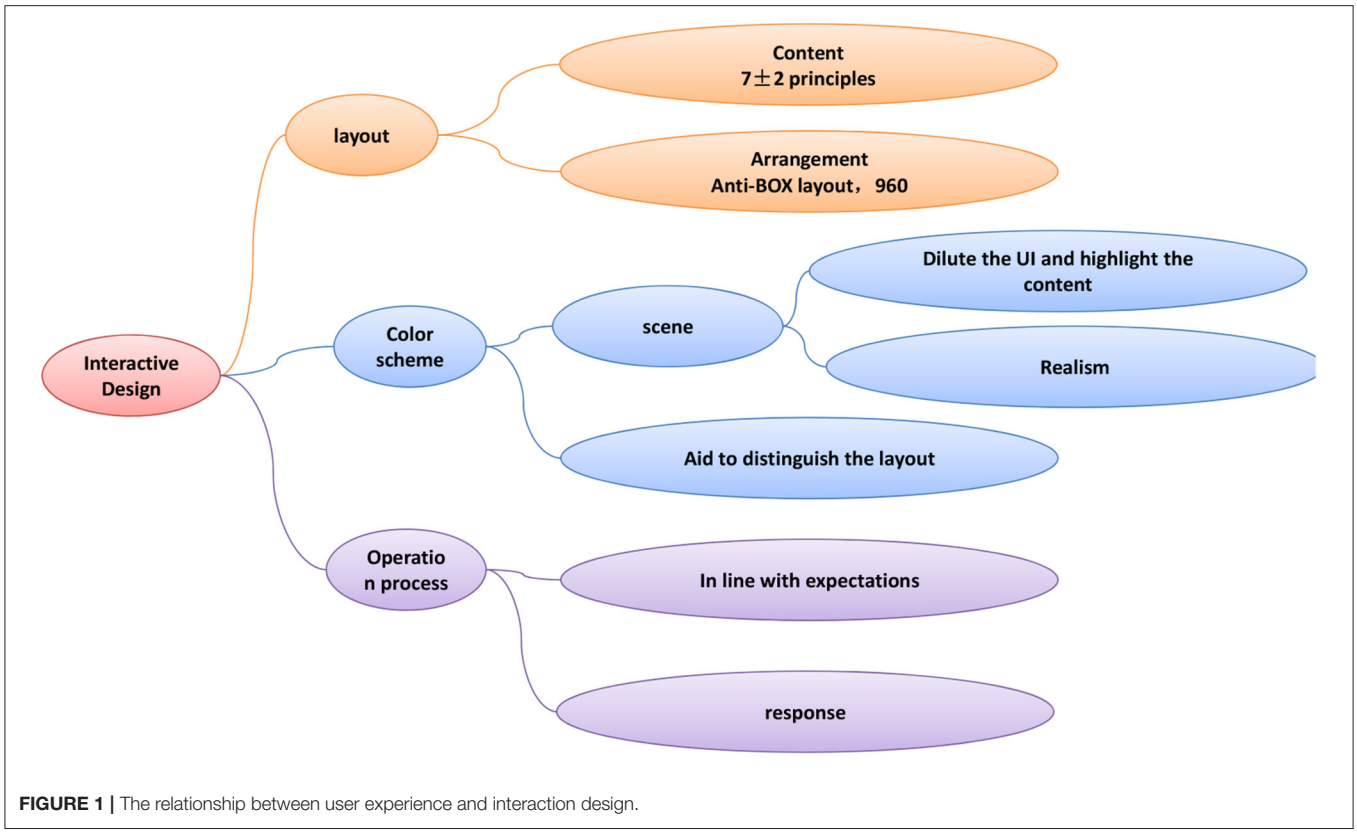


FIGURE 1 | The relationship between user experience and interaction design.

(1) An improved LSTM model is used for speech feature extraction. The improved LSTM adds a peephole connection. In the original LSTM structure, the previous unit state c<t-1> is connected to the forget gate and input gate, and the state of the previous unit is added to the forget gate and input gate. Calculating. The current state will not lose the information obtained by the upper state, and the output at the next moment will be more specific and complete.
(2) A self-attention mechanism is introduced. This mechanism is a strategy calculated according to the importance of different parts of things. The goal is to allocate more attention to the key parts of things. Through the calculation of the attention probability distribution, a larger weight is assigned to a certain key part. After the speech features are output by LSTM, the weight value of each frame is calculated. The introduction of this mechanism allows information with a large contribution rate to be assigned a high weight.
(3) Based on the self-attention mechanism and ILSTM, a speech-based emotion recognition method is proposed. Compared with the traditional LSTM, the method used in this paper has improved the accuracy of emotion recognition on multiple data sets.
(4) Based on the emotion recognition function, an emotional interaction design plan is proposed. The advantage of emotional interaction design proposed in this research lies in more accurate emotion recognition and different interaction designs based on the recognition results. The ultimate goal of this research is to enhance the user's product experience and truly put people first.

RELATED INFORMATION

Emotional Design Concept

People's happiness, anger, worry, thought, fear, and shock are all emotions. Emotions refer to the feelings produced by the interaction between life phenomena and the human heart. Emotional design is to give and satisfy people's emotional needs for a product. By fully considering the needs of human nature, the product that had no vitality was added to the emotional element to realize the spiritual function of the product. In 2002, the "emotional" concept of product design was proposed by American cognitive psychologist Donald Norman. In 2004, the book "Emotional Design" was published, which pushed the research of emotionalization from behind the scenes to the front, and attracted widespread attention. Nowadays, the Internet environment is becoming more mature, and market competition is increasingly encouraging. Products that simply solve functional requirements can no longer meet the needs of users. People's needs are developing toward higher levels such as emotion and interaction. This requires the combination of emotional design theory and interaction design, and through emotional interaction to meet the increasingly individual and diverse needs of users.

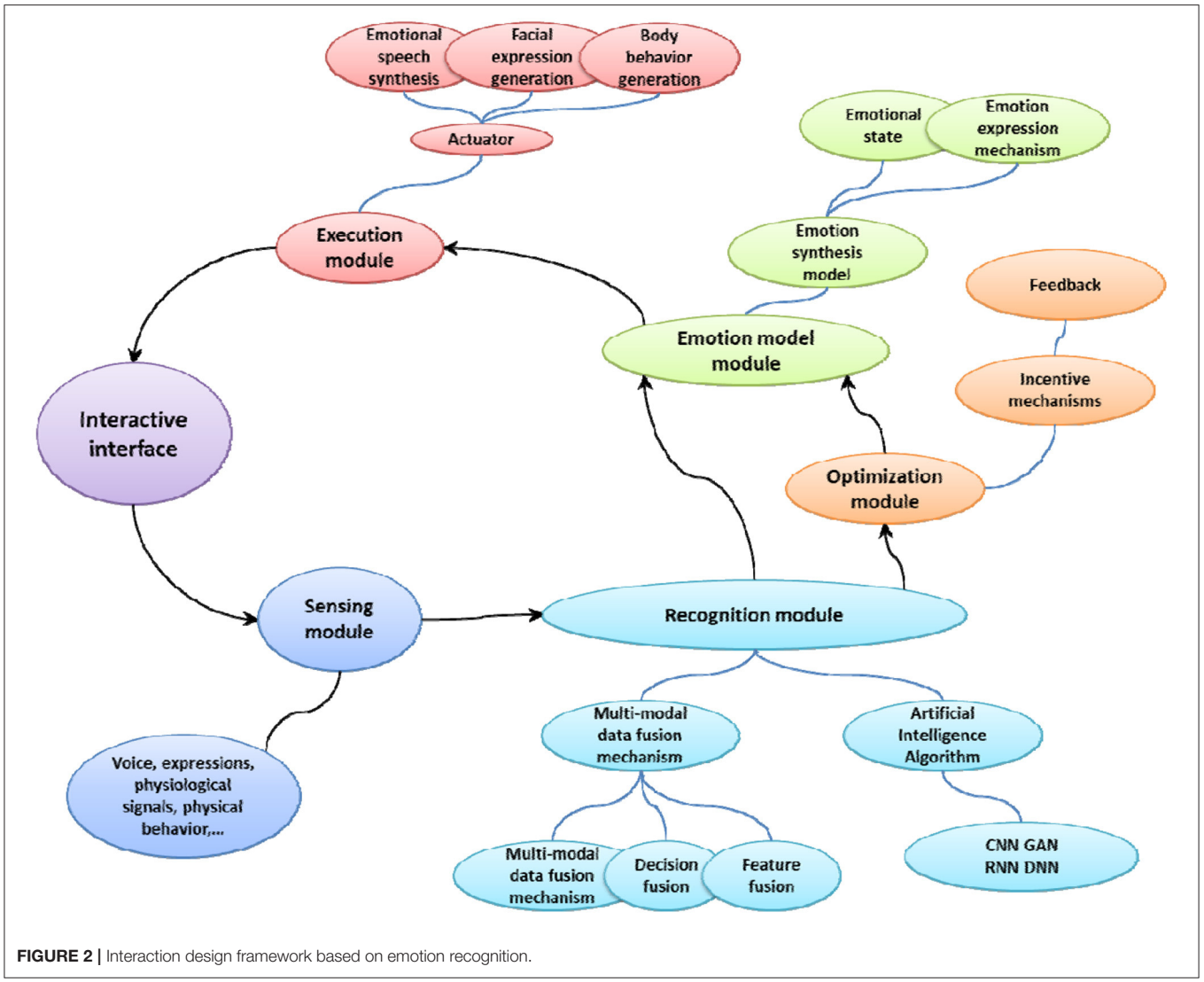


FIGURE 2 | Interaction design framework based on emotion recognition.

Principles of Emotional Design

Principles of Emotional Expression

The interactive design of the product needs to satisfy the user's emotional expression, as much as possible to satisfy the user's desire for expression, and cannot ignore the user's emotional appeal. For example, when using various social software, users can set an avatar, cover or change the theme. This is to satisfy the user's desire for expression as much as possible while transmitting information.

The Principle of Fun

The expression of the design should be humanized, as far as possible to draw the distance between man and machine. For example, in the process feedback, the two modes of "loading" and "loading hard" are displayed, the latter is more likely to give users a good impression. For the 404 page prompt, the non-emotional design is to only display "the requested page does not exist or the connection is wrong." The emotional design is to display "the page you were looking for ran away from home," coupled with a anthropomorphic picture full of sorry. As a user, the latter is more likely to forgive or understand emotions. Emotional design links the functions of the product with the emotions between users, which greatly increases the user's acceptance.

Interaction Design Framework Based on Emotion Recognition

The goal of emotional interaction is to make the computer make reasonable adjustments under the premise of understanding the human emotional state, and adapt to the transfer of user emotions. The focus of emotional interaction design is to use a variety of perception methods to recognize, interpret and respond to human emotions. **Figure 2** shows the framework of interaction design based on emotion recognition. The interactive design system mainly includes a sensing module, an execution module, a recognition module, an emotion calculation module, and an optimization module.

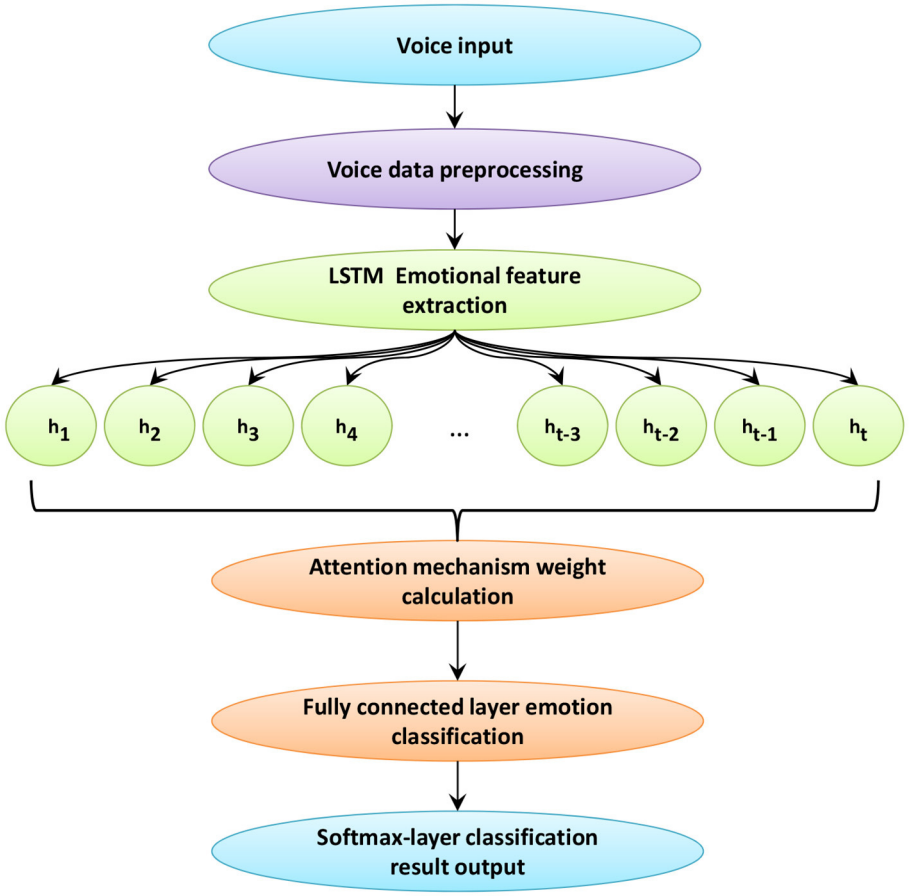


FIGURE 3 | Speech recognition framework based on the method in this paper.

The sensor module is the basis of the emotional interaction system. It uses a variety of sensor devices such as microphones, cameras, wearable devices and eye trackers to collect biometric data, including electrical signals in the brain, ECG signals, electrothermal signals, speech signals, facial expressions and physical behaviors, etc. The sensor module is equivalent to the neuron on the human body, feeling the stimulation of external substances all the time.

The recognition module is the basis of emotional interaction. First, it is necessary to preprocess the collected data. Then, the features of the data sample are extracted. Finally, the feature data is classified to obtain the recognition result. Relevant studies have shown that the use of deep learning methods, such as recurrent neural networks (RNN), convolutional neural networks (CNN), and adversarial neural networks (GAN), have higher speculation potential. They can not only automatically generate the best method for detecting the subjective experience of the user's emotion, but also model the user's long-term behavior through the storage device. In addition, emotion recognition based on multi-modal user information can choose different modal combinations, and there are multiple choices for the data fusion method between modalities according to actual conditions.

The emotion model module is a key part of emotion recognition and emotion expression. By establishing a mathematical model of emotional state, a more reasonable emotional understanding and feedback can be achieved. The emotional model module needs to study various psychological phenomena of the user, and make a reasonable response to the user's emotions by judging the composition of the emotional information. The expression of emotion, as a branch of experience and classic theoretical research, needs to build a human emotion model with inherent utility, summarize cognition from reality, and verify it through experiments.

The execution module is responsible for the information output task of the emotional interaction system. Perform corresponding feedback according to the decision made by the emotion model module, and finally complete the emotional interaction through emotional expression such as facial expressions, body movements, and synthesized speech.

In the emotional interaction system, the optimization module is the key link to provide personalized services. Since it is impossible to find an emotional model solution suitable for all systems, emotional modeling is still highly dependent on user feedback information, so it is particularly important to have an

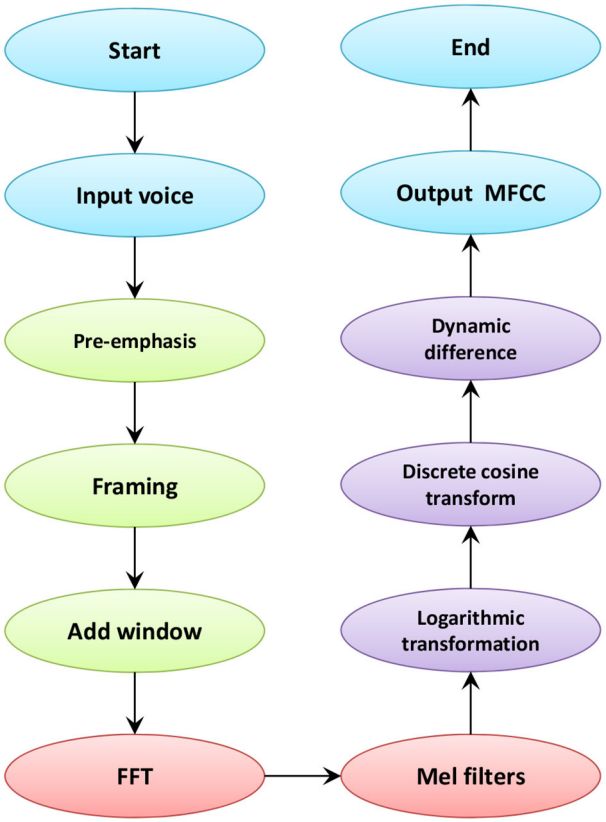


FIGURE 4 | The process of MFCC extraction.

adaptive optimization module. The optimization plan is usually completed in a controlled interactive environment, and the user's response to a specific product needs to be monitored. Monitoring information can provide valuable feedback information for the system, so as to establish a good and personalized reward and punishment optimization mechanism to achieve a better user experience.

EMOTION RECOGNITION BASED ON LSTM

This article uses a speech-based emotion recognition framework, the framework diagram is shown in **Figure 3**. $h_i, i = 1, 2, \ldots, t$ is the speech sequence of each frame obtained after passing through LSTM. t is the number of frames of the speech signal. In this framework, the input speech signal is preprocessed first. Then, perform MFCC extraction on the speech signal through the opensmile tool. The features extracted by MFCC are input into the ILSTM model. By adding ILSTM with peephole connection, a complete speech sequence is obtained. Then, the matrix output in ILSTM is used as the input of the attention mechanism layer. In the attention mechanism layer, through similarity calculation, the attention weight of each frame of speech signal relative to the recognition target is learned. Multiply the learned weight value with the input matrix to get the final weight value. Finally, the obtained information is classified by the fully connected layer, and the output of the speech emotion recognition result is realized by the Softmax layer.

MFCC Extraction

Traditional speech features are mainly divided into acoustic features, prosodic features and sound quality features. Due to the complexity of speech emotions, many emotions are difficult to effectively recognize based on prosody and sound quality alone, which results in a single original feature that is not well distinguished. This requires different speech features to be combined for speech recognition. Moreover, the non-stationary random process of speech features has a strong time-varying nature. Therefore, in order to increase the practicability of feature parameters and reduce the complexity of feature extraction, MFCC is selected as the speech emotion feature in this research. The calculation relationship between MFCC and human ear frequency can be expressed by Eq. (1):

$$M(f) = 2595 \times \log_{10}\left(1 + \frac{f}{700}\right) \quad (1)$$

Where M represents the Mel frequency function and f represents the linear frequency. Eq. (1) shows the relationship between Mel frequency and linear frequency. Within the Mel frequency, the human perception of audio is linear. MFCC coefficient is to construct characteristic parameters by simulating human ear characteristics and human hearing characteristics. The process of MFCC extraction is shown in **Figure 4**.

As shown in **Figure 4**, the speech signal first passes through the first-order high-pass filter of the transfer function for pre-emphasis. This kind of preprocessing can enhance the high-frequency components of the signal and compensate for the lost speech signal. Then, in order to ensure the short-term stability of the speech signal, it is necessary to perform frame processing. The speech signal is divided into several segments, and each short segment is called a frame, which improves the continuity between the left and right ends of the frame. After that, windowing is performed. The traditional MFCC coefficient windowing operation is to add a Hamming window. In this paper, the ordinary Hamming window is changed to a fourth-order new Hamming window. Eq. (2) is the calculation formula that adds the second harmonic component and fourth harmonic component of cosine.

$$w(n) = \left(a + b\cos\frac{2\pi n}{N-1} + c\cos\frac{4\pi n}{N-1} + d\cos\frac{8\pi n}{N-1}\right) R_N(n) \quad (2)$$

Where a, b, c, d are the orientation coefficients of the window function, N is the window length, and R_N is the time domain sequence of the rectangular window of the current window length. The data in the windowed speech undergoes Fast Fourier transform frame by frame. After taking the square of the absolute value, the frequency spectrum of each frame of

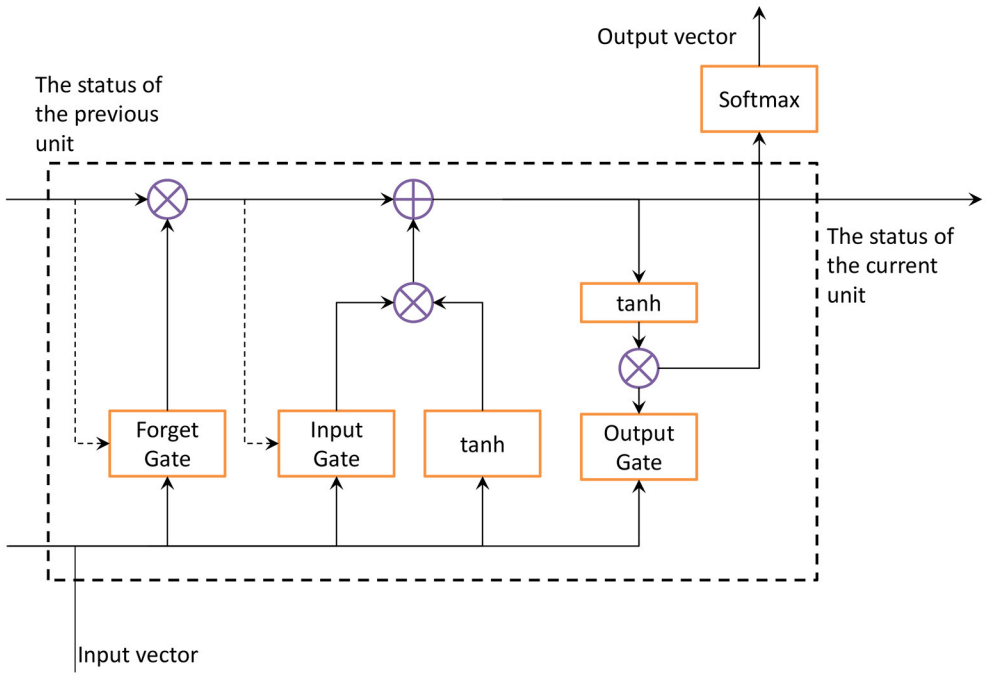


FIGURE 5 | ILSTM structure diagram.

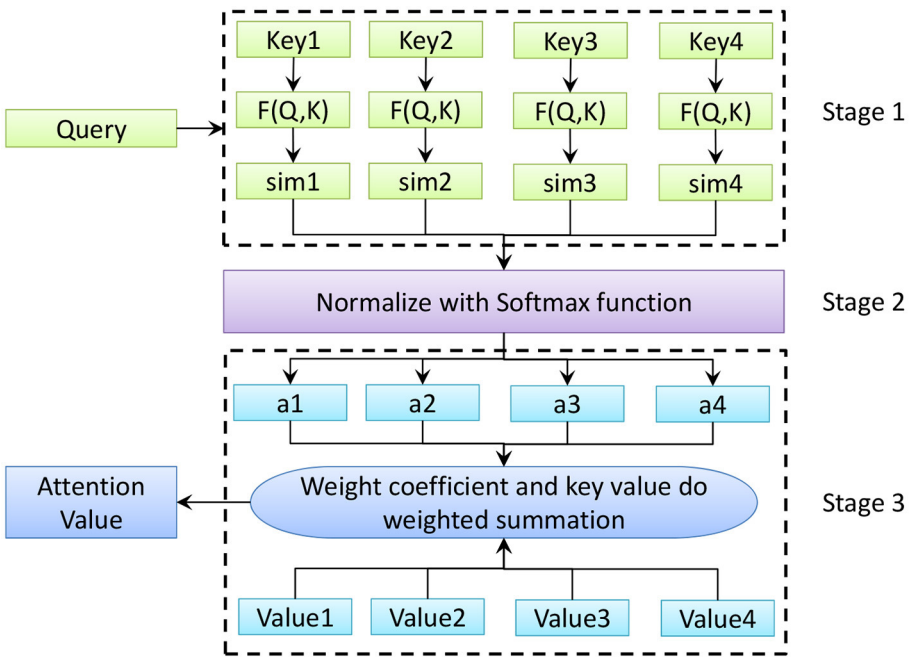


FIGURE 6 | Self-attention mechanism calculation process.

speech signal can be obtained. Perform modulo operation and square operation on the acquired spectrum to generate the power spectrum of the speech signal. The energy spectrum is passed through a set of Mel-scale triangular filters to eliminate the effect of harmonics. Usually a filter bank contains 22–26 filters. The number of Mel filters used in this article is 128. After calculating the logarithmic energy output by each filter bank, the calculation result is brought into the discrete cosine

transform. Most of the signal data is concentrated in the low frequency area. To eliminate redundant data, the first 25 data are taken. Then, the logarithmic energy of one frame is added to extract the dynamic difference parameters to calculate the signal difference of different frames. Finally, the global speech signal cepstrum parameters are used as static features, and the difference spectrum of static features is used as dynamic features. The calculation method of differential parameters combines static characteristics and dynamic characteristics to improve the recognition performance of the system.

Feature Extraction Based on ILSTM

The feed forward neural network of the LSTM network accepts the input of a specific structure, and the LSTM cyclically transmits the state in its own network. Therefore, the input range of time series structure type data is wider, and it has the function of describing dynamic time behavior. The key to LSTM is the unit state. The transfer effect of the unit state exists in the entire chain structure. There is a small amount of linear interaction in the process of transmission, so that information is easy to get. LSTM removes or adds information to the cell state through three gates. A gate is a structure that allows information to pass through selectively. An LSTM unit is composed of forget gate, input gate and output gate. Multiple units in the LSTM network form a series structure. LSTM was created to avoid the dependence of RNN from being affected by the long-term previous state. In the three gating units, the calculation of each unit will lose the information processed in the previous sequence. To ensure the integrity of the information and the accuracy of the output at the next moment, this study uses the ILSTM algorithm for speech emotion recognition. ILSTM adds peephole connections to the traditional LSTM network, and the model structure is shown in **Figure 5**. The dotted line in **Figure 5** is the added peephole connection. The function of this connection is to connect the last unit state c<t-1> in the traditional LSTM structure with the forget gate and the input gate. By adding the state of the previous unit to the calculation of the forget gate and input gate, the current state will not lose the information already obtained by the upper state, making the output at the next moment more complete.

The function of the forget gate is to select the information to be discarded from the cell state. The gate reads the output vector and the input of the memory unit, and outputs a value in the interval (0, 1) through a sigmod function. The values at both ends of the interval represent completely discarded or completely reserved. ILSTM adds the state of the previous unit to the sigmod function. The state of the previous unit is taken into account in the calculation of the forgetting gate, which makes the selection of the forgetting layer more accurate and prevents part of the effective information from being forgotten. The calculation expression of the forget gate at time t is:

$$F_t = \sigma\left(\mathbf{W}_F\left[c^{\langle t-1\rangle}, \alpha^{\langle t-1\rangle}, \mathbf{x}^{\langle t\rangle}\right] + \mathbf{b}_F\right) \tag{3}$$

where F_t represents the forgetting gate at time t, c, a, and x represent the state of the memory unit, the output vector and input vector of the memory unit, respectively.$\mathbf{W}_F$ and $\mathbf{b}_F$ represent the weight matrix and bias vector in the forgetting gate unit. The input gate is calculated according to the last output and current input value, and the addition of new state information is controlled by the degree of update. The unit state of the upper layer is also added to the input gate. The information added to the unit state must be important so that the update effect can be more accurate. The update formula of the input gate and current state at time t is:

$$I_t = \sigma\left(W_I\left[c^{\langle t-1\rangle}, \alpha^{\langle t-1\rangle}, \mathbf{x}^{\langle t\rangle}\right] + \mathbf{b}_I\right) \tag{4}$$

$$c^{\langle t\rangle} = I_t \times \bar{c}^{\langle t\rangle} + F_t \times \bar{c}^{\langle t-1\rangle} \tag{5}$$

$$\bar{c}^{\langle t\rangle} = \tanh\left(\mathbf{W}_c\left[\alpha^{\langle t-1\rangle}, \mathbf{x}^{\langle t\rangle}\right] + \mathbf{b}_c\right) \tag{6}$$

where I_t represents the input gate at time t, and W_I and $\mathbf{b}_I$ represent the weight matrix and bias vector in the input gate unit. After adding the previous unit state to the forget gate and input gate, the current unit state will be more accurate as the two gates are improved. The forget gate and the input gate and the update unit simultaneously control the current unit status, thereby improving the accuracy of the current unit status. Finally, the output result obtained by the unit state dot multiplication output gate is more comprehensive. To reduce complexity, no peephole connection is added to the output gate. Then the formula of the output gate at time t is:

$$\mathrm{O}_t = \sigma\left(\mathbf{W}_o\left[\alpha^{\langle t-1\rangle}, \mathbf{x}^{\langle t\rangle}\right] + \mathbf{b}_o\right) \tag{7}$$

where O_t represents the output gate at time t, and $\mathbf{W}_o$ and $\mathbf{b}_o$ represent the weight matrix and bias vector in the output gate unit. The output gate controls the effect of long-term memory on the current output. The final output of the network is determined by the input gate, forget gate, output gate, the state of the previous memory unit and the state of the current memory unit.

Self-Attention Mechanism

The essence of the self-attention mechanism is a mapping relationship between keys and values. Key-value query contains three basic elements: query, key, and value. The weight coefficient of each key value is obtained by calculating the correlation between each query item and each key. Then the weight and the corresponding key value are weighted and summed. After the speech signal is input to ILSTM, the weight value of each frame is calculated. The calculation process of the self-attention mechanism is shown in **Figure 6**. In **Figure 6**, Key, Query, and Value, respectively represent the key word of the input feature, the query value and the weight value of the current key word. F represents the function for calculating the weight coefficient. *Sim* represents the similarity obtained by the weight coefficient. a represents the weight coefficient of the corresponding key value.

As shown in **Figure 6**, the self-attention mechanism can be divided into three stages. In the first stage, the weight coefficient of each keyword's corresponding value is obtained by calculating the correlation between each query value and each keyword.

TABLE 1 | Data set introduction.

Database	Number of categories	Category breakdown	Number of samples	Sampling rate
EMO-DB	7	Neutral, angry, scared, happy, sad, disgusted, bored	535	48 kHz
CASIA	6	Happy, sad, scared, angry, surprised, neutral	9,600	16 kHz

There are generally three calculation methods, namely the vector dot product method, the cosine function method and the introduction of additional neural network evaluation methods. The three calculation formulas are as follows

$$Similarity(Query, Keyi) = Query \square Keyi \tag{8}$$

$$Similarity(Query, Keyi) = \frac{Query \square Keyi}{\|Query\| \square \|Keyi\|} \tag{9}$$

$$Similarity(Query, Keyi) = MLP\left(Query, Keyi\right) \tag{10}$$

In the second stage, the Softmax function is used to normalize the weights. Eq. (11) is the normalized processing formula. L_x represents the length of the corresponding data source.

$$\boldsymbol{\alpha}_i = soft\max\left(Sim_i\right) = \frac{e^{Sim_i}}{\sum_{j=1}^{L_x} e^{Sim_j}} \tag{11}$$

In the third stage, the weight coefficient and the corresponding key value are weighted and summed to obtain the final attention value, as shown in Eq. (12).

$$Attention\left(Query, Source\right) = \sum_{i=1}^{L_x} \alpha_i \cdot Value_i \tag{12}$$

Suppose there are T frames after the speech data is divided, and the latitude of each frame is the number of neurons in the ILSTM network. So the size of the matrix obtained after passing through the ILSTM network is n*T, where n is the dimension of each frame of speech. The encoding output of the attention layer is obtained through the calculation of Eq. (13).

$$A = soft\max\left(g\left(\mathbf{H}^T\mathbf{W}_1\right)\mathbf{W}_2\right) \tag{13}$$

$\mathbf{W}_1$ and $\mathbf{W}_2$ represent the most suitable parameter matrix given manually. **H** represents the input matrix extracted by ILSTM. Finally, the weight value calculated by the self-attention mechanism is multiplied by the input matrix **H**, and the result of the multiplication is input to the fully connected layer for classification.

EXPERIMENTAL PROCESS AND RESULT ANALYSIS

Experimental Data Set

Two experimental data sets are used in this experiment. The basic situation of each data set is shown in **Table 1**. During the experiment, 30% of the entire data set was selected as the test data set and 70% as the training data set.

TABLE 2 | Experimental parameter settings.

Parameter	Setting instructions
Number of Mel filters	128
Number of hidden layer units of LSTM/ILSTM	128
Learning rate	0.01
Speech signal	16 bit

Experimental Setup

The comparison methods used in the experiment are LSTM, ILSTM, and the method used in this article. The evaluation index is accuracy. The main hardware configuration of the experimental platform used in this article is: the processor is Intel Xeon processor E5-2600 @ 3. 40 GHz, the GPU is NVIDIA GeForce GTX1080Ti, and the memory is 64G. Other experimental parameter settings are shown in **Table 2**.

Discussion of Experimental Results

Emo-DB

To compare the actual recognition effect of each method on the Emo-DB database, this section applies each method to the Emo-DB data set. The experimental results are shown in **Tables 3–5**. Comparing the data in the three tables, it can be seen that the recognition rates of the four emotions all exceed 60%. These four emotions are anger, boredom, disgust, and sadness. The recognition rate of the remaining three emotions is <50%. Natural emotions are easily confused with boring emotions, and angry emotions are easily confused with fear emotions. The recognition rate based on the traditional LSTM model is 62.35%, the recognition rate based on the ILSTM model is 63.57%, and the recognition rate based on this method is 65.29%. Compared with traditional LSTM, the recognition rate of ILSTM has increased by nearly 3%. The recognition rate of this method has increased by nearly five percentage points. This fully demonstrates the effectiveness and superiority of this method.

Comparing the recognition of various emotions by traditional LSTM and ILSTM, the recognition rates for the five emotions of boredom, disgust, fear, happiness, and nature have all increased by more than 2.5%. The recognition rate of anger and fear emotions has decreased slightly. But on the whole, ILSTM's recognition rate of emotion has improved. This shows that because ILSTM adds peephole connections in the input gate and the forget gate, the information reception is more complete

TABLE 3 | Recognition rate of traditional LSTM in Emo-DB.

Emotion category	Angry	Bored	Disgust	Fear	Happy	Natural	Sad
Angry	**75**	2	3	3	15	1	1
Bored	0.5	**88**	0	0	2	5	4.5
Disgust	15	11	**67**	1.5	0.5	5	0.0
Fear	21	14	5	**32**	19.5	3.5	5
Happy	54.5	0	3	1	**42**	0	0.5
Natural	1	57	0	0.5	0.5	**39**	2.0
Sad	1.5	5	0	0	0	0	**93.5**

The bold values represent the recognition rate of the adopted method.

TABLE 4 | Recognition rate of ILSTM in Emo-DB.

Emotion category	Angry	Bored	Disgust	Fear	Happy	Natural	Sad
Angry	**74**	3	3	2.5	17	0	0.5
Bored	0.5	**91**	0	0	1	4	3.5
Disgust	13	11	**70.5**	1	0	4.5	0.0
Fear	20	16	5	**29.5**	19	4	5.5
Happy	50	0	3	1	**45**	0	1
Natural	1	49	0	0.5	0.5	**44**	5.0
Sad	1	7	1	0	0	0	**91**

The bold values represent the recognition rate of the adopted method.

TABLE 5 | Recognition rate of this method in Emo-DB.

Emotion category	Angry	Bored	Disgust	Fear	Happy	Natural	Sad
Angry	**77**	2	3	2	16	0	0
Bored	0	**93**	0	0	1	4	2
Disgust	13	10	**71**	1	0	5	0.0
Fear	20	16	4	**31**	19	5	5
Happy	49	1	3	1	**46**	0	0
Natural	1	48	0	0	0	**46**	5.0
Sad	0.5	5.5	1	0	0	0	**93**

The bold values represent the recognition rate of the adopted method.

and the state extraction is more sufficient. In the calculation of each unit layer, the unit state of the upper layer is input into the input gate and the forget gate, which makes the speech signal reception more complete, thereby improving the overall recognition performance.

Comparing the recognition of various emotions by traditional LSTM and the method in this paper, the recognition rates of the five emotions of boredom, disgust, fear, happiness, and nature have all increased by more than 4%. The recognition rate of angry emotions increased by 2%. The recognition rate of sad emotions decreased slightly. It can be seen from the results that the method used in this paper has the best recognition rate. This is because ILSTM is sufficient for information extraction. The speech data is divided into frames and then input into the attention mechanism. The importance of each frame of signal is learned in the attention layer, and the output is weighted. Because the output information of ILSTM is more specific, the weight of each frame of speech is more accurate.

CASIA

The experimental results of each method on the CASIA data set are shown in **Tables 6–8**. Comparing the data in the three tables, it can be seen that the recognition rates of the four emotions all exceed 60%. These four emotions are anger, happiness, calm, and surprise. The recognition rate of the remaining two emotions is close to 50%. The data in the table shows that it is easy to confuse fear and sadness. The recognition rate based on traditional LSTM is 68.75%, the recognition rate based on ILSTM is 70.75%, and the recognition rate based on this method is 74.17%. Compared with traditional LSTM, the recognition rate of ILSTM is increased by 2%. The recognition rate of this method has increased by five percentage points. This fully demonstrates the effectiveness and superiority of this method.

Comparing the recognition of various emotions by traditional LSTM and ILSTM, the recognition rate of the five emotions of anger, happiness, fear, surprise, and sadness has increased, with an average increase of 2.8%. The recognition rate of calm

TABLE 6 | Recognition rate of traditional LSTM in CASIA.

Emotion category	Angry	Happy	Fear	Calm	Surprised	Sad
Angry	**77**	11	2	1	7.5	1.5
Happy	8	**69.5**	5	3.5	13	1
Fear	2	6	**53**	2	4	33
Calm	1.5	7.5	10	**77**	2	2
Surprised	8	7	2	3	**79**	1
Sad	3	1.5	33	3.5	2	**57**

The bold values represent the recognition rate of the adopted method.

TABLE 7 | Recognition rate of ILSTM in CASIA.

Emotion category	Angry	Happy	Fear	Calm	Surprised	Sad
Angry	**77.5**	10.5	1	1	8	2
Happy	5	**75**	2	3	14	1
Fear	3	5	**57**	1	2.5	31.5
Calm	2	8	9.5	**75**	1.5	4
Surprised	7	6	3	2	**81.5**	0.5
Sad	2	3	34	2.5	0	**58.5**

The bold values represent the recognition rate of the adopted method.

TABLE 8 | Recognition rate of this method in CASIA.

Emotion category	Angry	Happy	Fear	Calm	Surprised	Sad
Angry	**78**	9	3	0	9	1
Happy	4	**78**	5	2	11	0
Fear	1.5	2.5	**59**	1	5	31
Calm	1	4	10	**77**	6.5	1.5
Surprised	6	5	4	1	**83**	1
Sad	2	2	30.5	1.5	3	**60**

The bold values represent the recognition rate of the adopted method.

emotion dropped by 2%. But on the whole, ILSTM's recognition rate of emotion has improved. Comparing the recognition of various emotions by traditional LSTM and the method in this paper, the recognition rate of the five emotions of angry, happy, scared, surprised and sad has increased, with an average increase of 4.5%. The recognition rate of calm emotion has not changed. On the whole, ILSTM's recognition rate of emotion has improved. Comparing the recognition results of the three methods, it can be seen that the recognition rate of this method is the best.

CONCLUSION

The key to emotional interaction design is accurate emotional recognition results. Based on this requirement, this article uses an emotion recognition framework based on LSTM. The framework first extracts the MFCC features of the speech information. Second, ILSTM performs feature extraction on MFCC. The output of the entire ILSTM is used as the input of the attention layer, which is input to the fully connected layer and the Softmax layer after weight calculation to obtain the final recognition result. The recognition results on Emo-DB and CASIA data sets prove that the method in this paper has the best recognition effect. With the development of emotion science, computer science, and electronics, the widespread application of intelligent algorithms and wearable technology has made emotional interaction possible. These new technologies have become part of people's real lives and help to continuously improve the quality of life. In addition, these technologies can obtain large amounts of data, which can be used to develop complex artificial intelligence algorithms to achieve reliable emotional algorithm calculation models. Many methods have been developed for emotional design, and new research questions, challenges and opportunities have emerged, which have made people's understanding and cognition of emotional design a step forward.

In the next step, the method of this paper will be introduced into the fields of machine translation and polygraph detection, to test the continuous emotion corpus and improve the calculation of attention scores to further improve the speech emotion recognition rate.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

XC and LZ conceived and developed the theoretical framework of the manuscript. XC, RH, and XL carried out model design, theoretical verification, and training under the guidance of RH. XC, LX, and MZ conducted the evaluation and analysis of experimental data under the guidance of RH. Everyone participated in the process of writing and revising the manuscript.

FUNDING

The study was supported by Cultural & Creative Products Research Center of Jiangnan University and the Fundamental Research Funds for the Central Universities (2019JDZD02), the Fundamental Research Funds for the Central Universities (JUSRP120109), and the National Key R&D Program of China (Grant No. 2017YFB0202303).

REFERENCES

Abdulsalam, W. H., Alhamdani, R. S., and Abdullah, M. N. (2019). Facial emotion recognition from videos using deep convolutional neural networks. *Int. J. Mach. Learn. Comput.* 9, 14–19. doi: 10.18178/ijmlc.2019.9.1.759

Ahmed, F., Bari, A., and Gavrilova, M. L. (2020). Emotion recognition from body movement. *IEEE Access* 8,11761–11781. doi: 10.1109/ACCESS.2019.2963113

Aouani„H., and Ayed, Y. B. (2020). Speech emotion recognition with deep learning. *Proc. Comput. Sci.* 176, 251–260. doi: 10.1016/j.procs.2020.08.027

Assed, M. M., Khafif, T. C., Belizario, G. O., Fatorelli, R., and Serafim, A. D. P. (2020). Facial emotion recognition in maltreated children: a systematic review. *J. Child Fam. Stud.* 29,1493–1509. doi: 10.1007/s10826-019-01636-w

Ayata, D., Yaslan, Y., and Kamasak, and, M. E. (2020). Emotion recognition from multimodal physiological signals for emotion aware healthcare systems. *J. Med. Biol. Eng.* 40,149–157. doi: 10.1007/s40846-019-00505-7

Bharate, V. D., Chaudhari, D. S., and Chaudhari, M. D. (2019). Emotion recognition from facial expressions using GFE, LBP and Hog feature extraction techniques. *Int. J. Innov. Technol. Explor. Eng.* 8, 2835–2840. doi: 10.35940/ijitee.L3030.1081219

Bijlstra, G., Holland„R. W., Dotsch„R., and Wigboldus, D. H. J. (2019). Stereotypes and prejudice affect the recognition of emotional body postures. *Emotion* 19, 189–199. doi: 10.1037/emo0000438

Borth, D., Tao, C., Ji, R., and Chang, S. F. (2013). "SentiBank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content," in *ACM International Conference on Multimedia ACM* (Amsterdam), 459–460. doi: 10.1145/2502081.2502268

Brousseau, B., Rose, J., and Eizenman, M. (2020). Hybrid eye-tracking on a smartphone with CNN feature extraction and an infrared 3D model. *Sensors* 20, 1–21. doi: 10.3390/s20020543

Cabada,R. Z., Rangel,H. R., Estrada,M. L. B., and Lopez, H. M. C.(2020). Hyperparameter optimization in CNN for learning-centered emotion recognition for intelligent tutoring systems. *Soft Comput. Fusion Foundations Methodol. Appl.* 24, 7593–7602. doi: 10.1007/s00500-019-04387-4

Dey, A., Sujithra, T., and Anand, B. (2019). Emotion recognition system based on facial expressions. *Int. J. Eng. Adv. Technol.* 8, 939–943. Available at: https://www.ijeat.org/download/volume-8-issue-4/

Edgar, P. T. P., Torres„E. A., Hernández-Lvarez, M., and Sang, G. Y. (2020). EEG-based BCI emotion recognition: a survey. *Sensors* 20:E5083. doi: 10.3390/s20185083

Gosztolya, G. (2020). Using the fisher vector representation for audio-based emotion recognition. *Acta Polytech. Hung.* 17, 7–23. doi: 10.12700/APH.17.6.2020.6.1

Issa, D., Demirci, M. F., and Yazici, A. (2020). Speech emotion recognition with deep convolutional neural networks. *Biomed. Signal. Process. Control* 59:101894. doi: 10.1016/j.bspc.2020.101894

Jayasimha, Y., and Reddy, R. (2019). Venkata Siva, a robust face emotion recognition approach through optimized SIFT features and adaptive deep belief neural network. *Intell. Decis. Technol.* 13, 379–390. doi: 10.3233/IDT-190022

Jiang,P., Fu, H., and Tao, H.(2019). Speech emotion recognition using deep convolutional neural network and simple recurrent unit. *Eng. Lett.* 27,1–6. Available at: http://www.engineeringletters.com/issues_v27/issue_4/index.html

Krishna, P. M., Reddy, P. R., Narayanan, V., Lalitha, S., and Gupta, D. (2019). Affective state recognition using audio cues. *J. Intel. Fuzzy Syst.* 36, 2147–2154. doi: 10.3233/JIFS-169926

Lay,N. T., Wei,F. S., Silva,D., and Liyanage C,(2003). Speech emotion recognition using hidden Markov models. *Speech Commun.* 41, 603–623. doi: 10.1016/S0167-6393(03)00099-2

Lenzoni, S., Bozzoni, V., Burgio, F., Gelder, B. D., and Semenza, C. (2020). Recognition of emotions conveyed by facial expression and body postures in myotonic dystrophy (DM). *Cortex J. Dev. Study Nerv. Syst. Behav.* 127, 58–66. doi: 10.1016/j.cortex.2020.02.005

Liu, Y., and Zheng, Y. F. (2007). Soft SVM and its application in video-object extraction. *IEEE Trans. Signal. Process.* 55, 3272–3282. doi: 10.1109/TSP.2007.894403

Marco-Garcia, S., Ferrer-Quintero, M., Usall, J., Ochoa, S., and Huerta-Ramos, E. (2019). Facial emotion recognition in neurological disorders: a narrative review. *Rev. De Neurol.* 69, 207–219. doi: 10.33588/rn.6905.2019047

Martinez-Sanchez, F., Fernández-Abascal, E. G., and Sánchez-Pérez, N. (2017). Recognition of emotional facial expressions in Alexithymia. *Stud. Psychol.* 59, 206–216. doi: 10.21909/sp.2017.03.741

Navyasri, M., RajeswarRao, R., DaveeduRaju, A., and Ramakrishnamurthy, M. (2018). Robust features for emotion recognition from speech by using Gaussian mixture model classification. *Smart Innov. Syst. Technol.* 84, 437–444. doi: 10.1007/978-3-319-63645-0_50

Nyquist, A. C., and Luebbe, A. M. (2020). An emotion recognition-awareness vulnerability hypothesis for depression in adolescence: a systematic review. *Clin. Child Fam. Psychol. Rev.* 23, 27–53. doi: 10.1007/s10567-019-00302-3

Panda, R., Malheiro, R. M., and Paiva, R. P. (2020). Novel audio features for music emotion recognition. *IEEE Trans. Affect. Comput.* 11, 614–626. doi: 10.1109/TAFFC.2018.2820691

Parthasarathy, S., and Busso, C. (2020). Semi-supervised speech emotion recognition with ladder networks. *IEEE/ACM Trans. Audio Speech Lang. Process.* 28, 2697–2709. doi: 10.1109/TASLP.2020.3023632

Powroznik, P. (2014). Polish emotional speech recognition using artifical neural network. *Adv. Sci. Technol. Res. J.* 8, 24–27. doi: 10.12913/22998624/562

Sak, H., Senior, A., and Beaufays, F. (2014). Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. *Comput. Sci.* 2014, 338–342. AQ doi

Sharma, R., Pachori, R. B., and Sircar, P. (2020). Automated emotion recognition based on higher order statistics and deep learning algorithm. *Biomed. Signal Process. Control* 58:101867. doi: 10.1016/j.bspc.2020.101867

Sheikhan, M., Gharavian, D., and Ashoftedel, F. (2012). Using DTW neural-based MFCC warping to improve emotional speech recognition. *Neural Comput. Appl.* 21:176501773. doi: 10.1007/s00521-011-0620-8

Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., et al. (2018). A review of emotion recognition using physiological signals. *Sensors* 18:2074. doi: 10.3390/s18072074

Sivasangari, A., Bhanu Prakash, V., and Rajesh, G. V. V. (2019). Recognition of emotion from facial expression for autism disorder. *Int. J. Recent Technol. Eng.* 8, 530–532. doi: 10.35940/ijrte.B1095.0782S319

Smallman, L., Underwood„W., and Artemiou, A. (2020). Simple Poisson PCA: an algorithm for (sparse) feature extraction with simultaneous dimension determination. *Comput. Stat.* 35, 559–577. doi: 10.1007/s00180-019-00903-0

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., and Feuston, B. P. (2003). Random forest: a classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* 43:1947. doi: 10.1021/ci034160g

Torresani, L., Szummer, M., and Fitzgibbon, A. (2010). "Efficient object category recognition using classemes," in *European Conference on Computer Vision* (New York, NY: Springer-Verlag), 776–789. doi: 10.1007/978-3-642-15549-9_56

Trigeorgis, G., Ringeval, F., Brueckner, R., Marchi, E., and Zafeiriou, S. (2016). "Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network [C]," in *2016 IEEE International Conference On Acoustics, Speech And Signal Processing (ICASSP)*. (Shanghai: IEEE), 5200−5204. doi: 10.1109/ICASSP.2016.7472669

Vapnik, V. N. (1999). *The Nature of Satistical Learning Theory*. New York, NY: Springer-Verlag Inc. 988–999. doi: 10.1109/72.788640

Wang, W., Watters, P. A., Cao, X., Shen, L., and Li, B. (2020). Significance of phonological features in speech emotion recognition. *Int. J. Speech Technol.* 23, 633–642. doi: 10.1007/s10772-020-09734-7

Wu, Y. X., Wu, Q. B., and Zhu, and, J. Q. (2019). Data-driven wind speed forecasting using deep feature extraction and LSTM. *IET Renew. Power Gener.* 13, 2062–2069. doi: 10.1049/iet-rpg.2018.5917

Zhang, G., Luo, T., Pedrycz, W., El-Meligy, M. A., and Li, Z. (2020). Outlier processing in multimodal emotion recognition. *IEEE Access* 8:55688–55701. doi: 10.1109/ACCESS.2020.2981760

Zhang, S., Tao, X., Chuang, Y., and Zhao, X. (2021). Learning deep multimodal affective features for spontaneous speech emotion recognition. *Speech Commun.* 127, 73–81. doi: 10.1016/j.specom.2020.12.009

Zou, J., Balter, J. M., and Cao, Y. (2020). Estimation of pharmacokinetic parameters from DCE-MRI by extracting long and short time-dependent features using an LSTM network. *Med. Phys.* 47, 3447–3457. doi: 10.1002/mp.14222

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Chen, Huang, Li, Xiao, Zhou and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 04 May 2021
doi: 10.3389/fpsyg.2021.664925

Affective Voice Interaction and Artificial Intelligence: A Research Study on the Acoustic Features of Gender and the Emotional States of the PAD Model

Kuo-Liang Huang[1]*, *Sheng-Feng Duan*[1] and *Xi Lyu*[2]

[1] Department of Industrial Design, Design Academy, Sichuan Fine Arts Institute, Chongqing, China, [2] Department of Digital Media Art, Design Academy, Sichuan Fine Arts Institute, Chongqing, China

New types of artificial intelligence products are gradually transferring to voice interaction modes with the demand for intelligent products expanding from communication to recognizing users' emotions and instantaneous feedback. At present, affective acoustic models are constructed through deep learning and abstracted into a mathematical model, making computers learn from data and equipping them with prediction abilities. Although this method can result in accurate predictions, it has a limitation in that it lacks explanatory capability; there is an urgent need for an empirical study of the connection between acoustic features and psychology as the theoretical basis for the adjustment of model parameters. Accordingly, this study focuses on exploring the differences between seven major "acoustic features" and their physical characteristics during voice interaction with the recognition and expression of "gender" and "emotional states of the pleasure-arousal-dominance (PAD) model." In this study, 31 females and 31 males aged between 21 and 60 were invited using the stratified random sampling method for the audio recording of different emotions. Subsequently, parameter values of acoustic features were extracted using Praat voice software. Finally, parameter values were analyzed using a Two-way ANOVA, mixed-design analysis in SPSS software. Results show that gender and emotional states of the PAD model vary among seven major acoustic features. Moreover, their difference values and rankings also vary. The research conclusions lay a theoretical foundation for AI emotional voice interaction and solve deep learning's current dilemma in emotional recognition and parameter optimization of the emotional synthesis model due to the lack of explanatory power.

Keywords: voice-user interface (VUI), affective computing, acoustic features, emotion analysis, PAD model

OPEN ACCESS

Edited by:
Pengjiang William Qian, Jiangnan University, China

Reviewed by:
Xiaoqing Gu, Changzhou University, China
Yuanpeng Zhang, Nantong University, China

***Correspondence:**
Kuo-Liang Huang
shashi@scfai.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *06 February 2021*
Accepted: *18 March 2021*
Published: *04 May 2021*

Citation:
Huang K-L, Duan S-F and Lyu X (2021) Affective Voice Interaction and Artificial Intelligence: A Research Study on the Acoustic Features of Gender and the Emotional States of the PAD Model. Front. Psychol. 12:664925. doi: 10.3389/fpsyg.2021.664925

INTRODUCTION

Nowadays, the core technologies of artificial intelligence (AI) are becoming increasingly mature. People face a new bottleneck in giving the "emotional temperature of humans" to a cold, intelligent device (Yonck, 2017). The conversational voice-user interface (VUI) is the most natural and instinctive interactive mode for humans. Recently, natural language processing (NLP) has

improved significantly due to the development of deep learning (DL) technology. The VUI demands of the new type of intelligent products transform communication to include emotional listening and feedback of users (Hirschberg and Manning, 2015; Dale, 2016; Chkroun and Azaria, 2019; Harper, 2019; Nguyen et al., 2019; Guo et al., 2020; Hildebrand et al., 2020). Giving computers similar emotional mechanisms and emotional intelligence concepts as humans is becoming increasingly critical in the information and cognitive sciences. The goal of "affective computing" is to endow computers with abilities of understanding and generating affective characteristics. Finally, the computer can become intimate with the nature and makeup of vivid interactions, like people. This involves interdisciplinary study in the areas of psychology, sociology, information science, and physiology (Picard, 2003, 2010) and is becoming a hot spot of laboratory research in academic and industrial circles (Bänziger et al., 2015; Özseven, 2018). Although VUI has considerable potential, effective semantic and emotional communication not only requires the subtle understanding of the physics and psychology of voice signals but also needs a method of extracting and analyzing voice features from human voice data (Picard, 2003; Guo et al., 2020; Hildebrand et al., 2020).

Affective computing is crucial to implementing man–machine emotional interactions through intelligent products (Picard, 2010; Dale, 2016). In the past, many studies of emotional voice recognition and synthesis have been reported. Nevertheless, they mainly establish acoustic models and systems based on information science. Abundant voice data have been input into the DL core of AI and several affective factors of acoustic features summarized from the 3-D pleasure-arousal-dominance (PAD) emotional state model on a "continuous dimension." A mathematical model was constructed and abstracted using mathematical knowledge and computer algorithms. Subsequently, the computer was able to learn from the data and make predictions by combining training data and its large-scale operation capability (Ribeiro et al., 2016; Rukavina et al., 2016; Kratzwald et al., 2018; Vempala and Russo, 2018; Badshah et al., 2019; Heracleous and Yoneyama, 2019; Guo et al., 2020). Although these practices can gain accurate prediction results quickly, they do not provide an understanding of where the results come from (e.g., black box) and lack explanatory ability (Kim et al., 2016; Ribeiro et al., 2016; Murdoch et al., 2019; Molnar, 2020). As a result, understanding how to adjust the model parameters is a problem that has yet to be solved, requiring an urgent empirical study of the connection between acoustic features and psychology as the theoretical basis for adjustment of model parameters (Ribeiro et al., 2016; Skerry-Ryan et al., 2018; Evans et al., 2019; Molnar, 2020). Research into voice rhythms from the cognitive psychology perspective has mainly focused on fundamental frequency, sound intensity, voice length, and other features (Juslin and Scherer, 2005). Emotional classifications are described quantitatively, which is different from the "continuous dimension" in existing intelligent systems. None of these studies yields 3-D coordinates through transformation to provide affection matching.

As a result of these shortcomings, an empirical study on the correlation between information enabling the emotional evaluation of acoustic features concerning emotional voice state and psychology is required in AI emotional voice interaction using a PAD model, which is the theoretical basis for adjustment of model parameters (Ribeiro et al., 2016; Skerry-Ryan et al., 2018; Evans et al., 2019; Molnar, 2020). Different average speech characteristics between males and females in human conversations have been reported in most studies (Childers and Wu, 1991; Feldstein et al., 1993). Furthermore, males and females show different emotional expressions. This study connected emotional states and voice features of male and female users through cross informatics and cognitive psychology from the voice interaction application scenes of intelligent products. Hence, this study focuses on the influences of "gender" and "emotions" on the "physical features of voices" in human–computer interactions as well as the quantitative expressions of the "physical features of voices." The research conclusions lay a theoretical foundation for AI emotional voice interaction and solve DL's current dilemma in emotional recognition and parameter optimization of the emotional synthesis model due to lack of explanatory powers.

LITERATURE REVIEW

Studies on Emotions and Classification

According to research within psychology and the neurosciences, there is extensive interaction between the emotions and cognition of humans (Osuna et al., 2020), displaying behavioral and psychological features (Fiebig et al., 2020) that have a profound impact on the expression, tone, and posture behavior of people in daily life (Scherer, 2003; Ivanović et al., 2015; Poria et al., 2017). In the past 20 decades, studies on emotions have increased significantly (Wang et al., 2020). At present, there are two mainstream affective description modes. One is to make a qualitative description of an emotional classification using adjectives from the perspective of "discrete dimensions," such as the six basic emotion categories proposed by Ekman and Oster (1979). The other is to describe the consequence determined by common affective factors of a "continuous dimension." The emotional states can be characterized and divided by quantitative emotional coordinates on different dimensions (Sloman, 1999; Bitouk et al., 2010; Chauhan et al., 2011; Harmon-Jones et al., 2016; Badshah et al., 2019). Specifically, 1-D space focuses on positive or negative emotional classification, and 2-D spatial emotional states are generally expressed by two coordinates, such as peace–excitement and happiness–sadness. The 3-D space is proposed by Schlosberg (1954), Osgood (1966), Izard (1991), Wundt and Wozniak (1998), and Dai et al. (2015), respectively.

Quantitative measurement of emotions is a requirement of affective computing (Dai et al., 2015). Because three-dimensional space is easy to compute, computational models of emotion (CMEs) in the current AI system adopt the continuous dimension; the most used is the PAD model proposed by Mehrabian and Russell in 1994. The PAD model hypothesizes that users have three emotional states according to the situation stimulus, including pleasure, arousal, and dominance. These 3-D axes act as an emotional generation mechanism (Mehrabian and Russell, 1974; Wang et al., 2020). For example, emotions are

TABLE 1 | Mapping of the eight Mehrabian basic emotions in PAD space.

	Trait combination	Emotional state
P (pleasure-displeasure): emotional state's positivity or negativity	+P+A+D	Exuberant
	–P–A–D	Bored
A (arousal-nonarousal): physical activity and mental alertness	+P+A–D	Dependent
	–P–A+D	Disdainful
D (dominance-submissiveness): feeling of control	+P–A+D	Relaxed
	–P+A–D	Anxious
	+P–A-D	Docile
	–P+A+D	Hostile

Datasource: Mehrabian (1996b).

divided into eight states with eight blocks of 3-D negative (–) and positive (+) combinations in the three dimensions as seen in **Table 1** (Mehrabian, 1996b).

As a CME, PAD can distinguish different emotional states effectively (Russell, 1980; Gao et al., 2016) and break from the traditional tag-description method. As one of the relatively mature emotional models (Mehrabian and Russell, 1974; Mehrabian, 1996a; Gunes et al., 2011; Jia et al., 2011; Chen and Long, 2013; Gao et al., 2016; Osuna et al., 2020; Wang et al., 2020), the PAD model measures the mapping relationship between emotional states and typical emotions by "distance" to some extent, thus transforming the analytical studies of discrete emotional voices into quantitative studies of emotional voices (Mehrabian and Russell, 1974; Mehrabian, 1996a; Gunes et al., 2011; Jia et al., 2011; Chen and Long, 2013; Gao et al., 2016; Osuna et al., 2020; Wang et al., 2020). It has been extensively applied in information processing, emotional computing, and man–machine interaction (Dai et al., 2015; Weiguo and Hongman, 2019). PAD is beneficial for establishing an external stimulus emotional calculation model to realize emotional responses during personalized man–machine interaction (Weiguo and Hongman, 2019).

Affective Computing and Emotions in Voice Interaction

Voice signals are the most natural method of communication for people (Weninger et al., 2013). On the one hand, voice signals contain the verbal content to be transmitted. On the other hand, rhythms in the vocalizations contain rich emotional indicators (Murray and Arnott, 1993; Gao et al., 2016; Noroozi et al., 2018; Skerry-Ryan et al., 2018). Each emotional state has unique acoustic features (Scherer et al., 1991; Weninger et al., 2013; Liu et al., 2018). For example, various prosodic features, including different tones, velocity, and volume, can express the speaker's different emotional states (Apple et al., 1979; Trouvain and Barry, 2000; Chen et al., 2012; Yanushevskaya et al., 2013).

Huttar (1968) further demonstrates that prosodic features of voice play an important role in emotions and suggests simulating these features (e.g., tone, velocity, and volume) in the interface by using artificial voices to express the emotional states of the speaker (Sauter et al., 2010). Subsequently, Professor Picard proposed affective computing (Picard, 2000) and attempted to endow computers with a similar affective mechanism to intelligently understand human emotions in man–machine interactions and, thus, realize effective interactions between an artificial voice and users. It is necessary to gain a subtle understanding of voices using an interdisciplinary approach, including physics and psychology, to understand how to extract and analyze phonetic features (Schwark, 2015; Guo et al., 2020). In addition to the automatic speech recognition (ASR) and text-to-speech (TTS) found in artificial speech, the process involves the emotional analysis of users (Tucker and Jones, 1991; Guo et al., 2020; Hildebrand et al., 2020). In **Figure 1**, the relationship between artificial acoustic waves and emotional states and the role of artificial acoustic waves in the voice interaction systems of intelligent products are reviewed. Specifically, a user's current emotional state in the PAD model is identified through affective computing according to emotional acoustic features in voice interactions. The user receives responses in an empathic voice expression of the computer in the AI product.

A Dimensional Framework of the Acoustic Features of Emotions

From a physiological perspective, loosening and contracting the vocal cords leads to rhythm changes in the voice, indicating emotions (Johar, 2016). From the perspective of psychology, relevant studies have proved that prosodic features of voices, such as basic frequency, velocity, and volume, are closely related to any emotional states (Williams and Stevens, 1972; Bachorowski, 1999; Kwon et al., 2003; Audibert et al., 2006; Hammerschmidt and Jürgens, 2007; Sauter et al., 2010; Quinto et al., 2013; Łtowski, 2014; Johar, 2016; Dasgupta, 2017; Hildebrand et al., 2020; Kamiloglu et al., 2020). Murray and Arnott (1993) introduce the concept of utterances and people's emotions, finding three major aspects that influence voice parameters of emotional impacts: utterance timing, utterance pitch contour, and voice quality. Among them, utterance timing and utterance pitch contour are prosodic features. In the past, most studies focused on prosodic features. Although these parameters gave certain differences in emotional distinction, some studies also find disadvantages for intelligent products in judging the emotions of the speaker, including voice quality (spectrum) (Toivanen et al., 2006). Jurafsky and Martin (2014). Experts in both linguistics and computers point out that each acoustic wave can be described completely by the four dimensions of time, frequency, amplitude, and spectrum. Connections between these four dimensions of acoustic waves and emotions in relevant studies are summarized in **Table 2**.

The first dimension is *time*, determined by the duration of a vibration from the sound maker (Sueur, 2018; Wayland, 2018) and measured in seconds or milliseconds of acoustic waves. Previous studies explore the influence of gender on velocity. Some studies demonstrate that the velocity of males is higher than females (Feldstein et al., 1993; Verhoeven et al., 2004; Jacewicz et al., 2010); however, most studies on people who speak English find no differences between males and females (Robb et al., 2004; Sturm and Seery, 2007; Nip and Green, 2013). Velocity

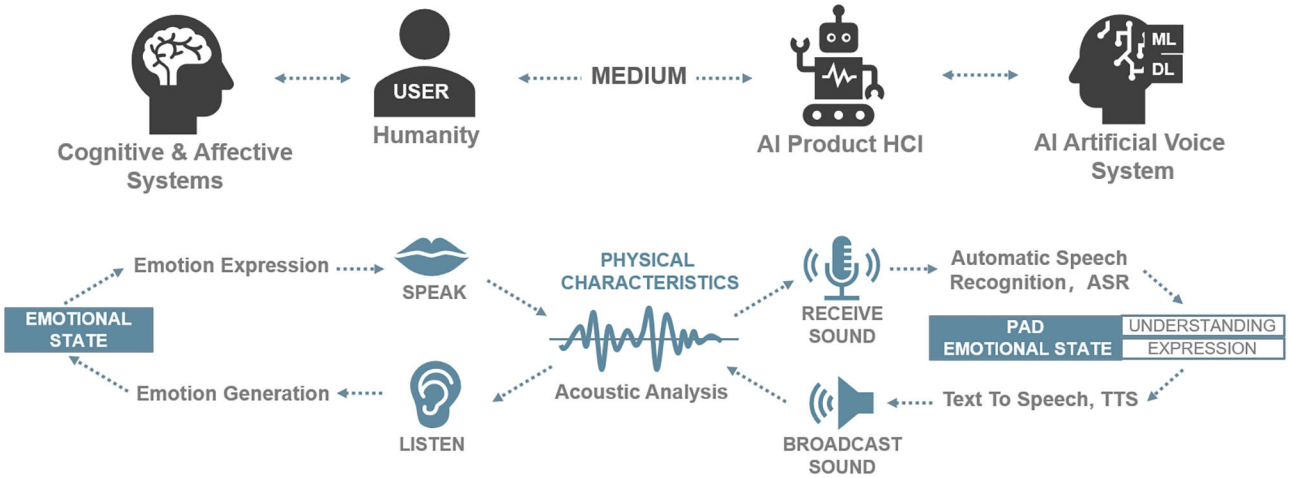


FIGURE 1 | The relationship between artificial acoustic waves and emotional states in the voice interaction systems of intelligent products. Source: Drawn by the authors.

TABLE 2 | Connections between the four dimensions of acoustic features and emotions.

Dimensions	Acoustic features	Emotional state correlations	Selected research
Time	Velocity of speech: average time per word (seconds)	anger (+), competence (+), contemplation (–), dominance (–), enthusiasm (+), extraversion (+), fear (+), happiness (+), persuasiveness (+), sadness (–), stress (+), tenderness (–) competence (–), contemplation (+), extraversion (–)	Williams and Stevens, 1972; Miller et al., 1976; Brenner et al., 1994; Tusing and Dillard, 2000; Mohammadi and Vinciarelli, 2012; Dasgupta, 2017
Frequency	Mean Pitch: Fo (Hz)	anger (+), competence (–), confidence (–), empathy (–), extraversion (+), fear (+), happiness (+), nervousness (+), persuasiveness (–), sadness (–), stress (+), tenderness (–), trustworthiness (–)	Williams and Stevens, 1972; Apple et al., 1979; Scherer and Giles, 1979; Brenner et al., 1994; Kwon et al., 2003; Bänziger and Scherer, 2005; Quinto et al., 2013; Bowman and Yamauchi, 2016; Guyer et al., 2019
	Fo *SD*: Pitch variability	anger (+), extraversion (+), happiness (+), sadness (–), shyness (–), sociability (+), tenderness (–)	Apple et al., 1979; Ray, 1986; Burgoon et al., 1990; Abelin and Allwood, 2000; Juslin and Laukka, 2003
Amplitude	Intensity: mean-sones intensity (dB)	aggression (+), anger (+), annoyance (+), dominance (+), extraversion (+), fear (–), happiness (+), tenderness (–), sadness (–), shyness (–), stress (+)	Mallory and Miller, 1958; Scherer and Giles, 1979; Brenner et al., 1994; Johnstone and Scherer, 1999; Kwon et al., 2003; Scherer, 2003; Asutay and Västfjäll, 2012; Quinto et al., 2013
Spectrum	Jitter%: a ratio of variation of fundamental frequency and mean	anger (+), annoyance (+), happiness (+), sadness (–), stress (+)	Johnstone and Scherer, 1999; Li et al., 2007
	Shimmer%: intensity perturbations	anger (+), confidence (+), joy (–), stress (+),	Juslin and Laukka, 2003; Li et al., 2007; Jacob, 2016; Jiang and Pell, 2017
	HNR: Proportion of periodic part and noises in signals (dB)	confidence (+), happiness (+), interest (+), lust (–), pleasure (+)	Jiang and Pell, 2017; Kamiloglu et al., 2020

Data source: organized in this study.

can indicate the emotional state of the speaker, generally with a high velocity in positive and negative emotional states (e.g., anger, fear, and happiness), but a low velocity in low-wakefulness states (Juslin and Laukka, 2003).

The second dimension is *frequency*, expressed by the number of vibrations of the acoustic wave per second (unit: Hz). The scale of this objective physical quantity corresponds to the fundamental frequency (Fo) of the vocal cord vibrations. Pitch is a subjective psychological quantity of sound, its value determined by the frequency of the acoustic waves (unit: Mel) (Juslin and Laukka, 2003; Colton et al., 2006). Pitch can represent different emotional states. The pitch is increased when a person is feeling anger, happiness, or fear and decreased when a person is sad or bored (Murray and Arnott, 1993; Johar, 2016). With respect to gender, the Fo of a male adult's voice is often lower than a female adult's voice (Mullennix et al., 1995; Pernet and Belin, 2012).

The third dimension is *amplitude*, which determines the intensity of sound (unit: dB). Loudness is the scale of a subjective

psychological index of intensity and results from a subjective judgment of a pure tone (unit: phon) (Sueur, 2018; Wayland, 2018). Generally speaking, the loudness of people is about 70 dB (Awan, 1993; Brown et al., 1993). Higher loudness is generally believed to relate to greater dominant traits or aggressiveness (Scherer and Giles, 1979; Abelin and Allwood, 2000; Asutay and Västfjäll, 2012; Yanushevskaya et al., 2013); relatively low loudness indicates people are fearful, sad, or gentle (Johar, 2016). Additionally, males' intensity of sound is slightly higher than that of females (Awan, 1993; Brockmann et al., 2011).

The fourth dimension is *spectrum*, referring to the energy distribution of signals (e.g., voice) in the frequency domain; it is expressed in graphs by analyzing perturbations of acoustic waves or periodic features (Sueur, 2018). The degree of "sound instability" during the formation of voices has been summarized (Hildebrand et al., 2020), reflecting voice quality (Kamiloglu et al., 2020). Vocal jitter is a measure of the periodic variation in fundamental frequency, indicating uneven tones of the speaker. A nervous speaker has instability in the voice (high perturbations) and a quiet speaker has a steady and stable sound (low perturbation) (Farrús et al., 2007; Kamiloglu et al., 2020). Specifically, jitter percentage expresses each basic frequency period's irregularity, that is, the degree of frequency perturbation. It is the ratio between the fluctuations of the fundamental frequency and mean values. A high numerical value indicates that the tone quality is unstable. Shimmer percentage refers to differences in repeated amplitude changes, that is, the degree of amplitude perturbation. It describes the ratio of the mean amplitude variation and respective mean. A high numerical value of shimmer percentage indicates greater changes in sound volume. HNR reflects the ratio of periodic segments and noises in signals (unit: dB). Lower noise energy in voices reflects fewer components of noises and better sound quality (Baken and Orlikoff, 2000; Ferrand, 2007). Some studies have proved that gender has no significant influences on jitter percentage, shimmer percentage, or HNR (Wang and Huang, 2004; Awan, 2006; Brockmann et al., 2008; Ting et al., 2011).

Research Directions on Connections of Acoustic Features and Emotional States

Studies on the emotional rhythm of voice have pointed out that people's sounds, characterized by pitch, loudness or intensity, and velocity, transfer different emotional information to listeners (Sauter et al., 2010). During a conversation, emotions can be recognized from video clips as short as 60 ms (Pollack et al., 1960; Pell and Kotz, 2011; Schaerlaeken and Grandjean, 2018). The same words and phrases can be expressed differently through fluctuation of different emotional states (Dasgupta, 2017); for example, rumination is related to low velocity and an extended dwell time. Anger is generally related to the loudness of voice (Juslin and Laukka, 2003; Clark, 2005). Fear is related to variations in pitch (Juslin and Laukka, 2003; Clark, 2005). The affective computing team from MIT analyzed variations in acoustic parameters, such as fundamental frequency and duration, during different emotional states; their results show that acoustic features of affective sounds (e.g., happy, surprise, and anger) are similar with the sad acoustic feature being relatively obvious (Sloman, 1999). In brief, the formation of human spoken language involves the interaction of individual traits and emotional states, used as a communication means to understand voices. To recognize and extract information for voice analysis, it is necessary to measure voice quality properties (Johar, 2016; Schaerlaeken and Grandjean, 2018).

To effectively establish an emotional identification and expression system, emotional identification and synthesis based on DL have considerable potential in human–machine interactions (Schuller and Schuller, 2021). Recognizing emotions through the automatic extraction of acoustic features and generating expressions through emotions are the main strategies for relevant research development. It has been proven that a generative adversarial network (GAN) can improve the machine's performance in emotional analysis tasks (Han et al., 2019). Additionally, people begin to think about transfer learning applications in relevant tasks and voice emotional computing modes (Schuller and Schuller, 2021).

Based on the above literature review, research can primarily presently be divided into two types. On the one hand, some studies based on information science strive to gain accurate emotional identification and natural voice expressions through DL. However, these studies lack the explanation for establishing a mathematical model (Ribeiro et al., 2016; Murdoch et al., 2019), thus resulting in the absence of a theoretical foundation for parameter optimization and adjustment. On the other hand, some studies are based on cognitive science and emotional states from the "discrete dimension." Most of these studies use prosodic features only and have shortages in emotional identification and expression (Toivanen et al., 2006). Studies rarely use the PAD model's emotional states in the intelligent product VUI as the framework for incorporating acoustic features of the spectrum and gender impacts. Hence, interdisciplinary studies are needed to solve the black box problems caused by DL.

METHODS

This study aims to connect humans' emotions and acoustic features from across information, acoustics, and psychology disciplines based on acoustic and cognitive psychology concepts.

Research Design

Both the purpose of this study and the literature review results have directed the current research to investigate the correlation of two independent variables, namely "gender" and "emotional state." The emotional state, different from other emotional classification models, considers each emotion has sole coordinates in the PAD space, enabling different emotions to show acoustic features independently. Therefore, the PAD model uses the eight basic emotions for emotional classification and neutral emotions as the benchmark. The dependent variables are seven main features associated with emotional states in the four dimensions of emotional voice sound waves.

Subjects and Materials

A total of 31 male and 31 female respondents were recruited by the stratified random sampling mode. Respondents have clear cognition with the nine basic emotions of PAD and display explicit oral expression. This study focuses on vocalizations from voice signals, and verbalizations are not transmitted; therefore, the recording of voice data used neutral words and verbalizations transmitted by "你好" (Chinese). Because it is easy to induce and simulate emotional recordings that can express real and natural emotions to some extent, PPT was used to provide films as the emotional stimuli to induce and guide recording of the participant (**Figure 2**). The provided film was confirmed by three relevant experts and then predicted and modified to assure effective induction and prompts.

Setting and Program of Experiments

Setup of experiments for data acquisition: An empirical study using laboratory experiments was carried out. All respondents engaged in the experiments, and voices were recorded in the same environment using the same settings. The input sound volume was fixed at 70 dB SP. The recording formula was mono channel; sampling frequency: 44.1 kHz; and resolution: 16 bits and WAV file. The relevant program is shown in **Figures 3**, **4**.

The audio recording process: First, selected respondents, in the closed experimental space without disturbance, were introduced to the experimental process and audition by the same prompts. Second, respondents wore a headset microphone in a closed space, and a provided laptop played the stimulus and prompted the film using Adobe Audition 2019. Respondents provided data of nine emotions: neutral, exuberant, bored, dependent, disdainful, relaxed, anxious, docile, and hostile. The content of the audio recordings from each respondent was then confirmed, and residual contents were preprocessed, including polishing and numbering. Finally, acoustic features were analyzed using the Praat 6.13 voice software (**Figure 5**).

Analysis of the spectrum was done using the calculation formulas of jitter percentage, shimmer percentage, and HNR as outlined below (Boersma, 1993; Fernandes et al., 2018; Sueur, 2018). Nine emotional voices were selected and analyzed by Praat, and characteristic parameter data of seven emotional voices were directly extracted.

FIGURE 2 | Emotional induction and guidance cases during voice recording of different emotions. **(A)** Emotional stimulus is induced. **(B)** Text to remind the emotion, and then record. Subsequently, **(C)** Interval shady, and then enter the next emotional stimulus to induce. The complete contents are shown in the **Appendix: Supplementary Material**.

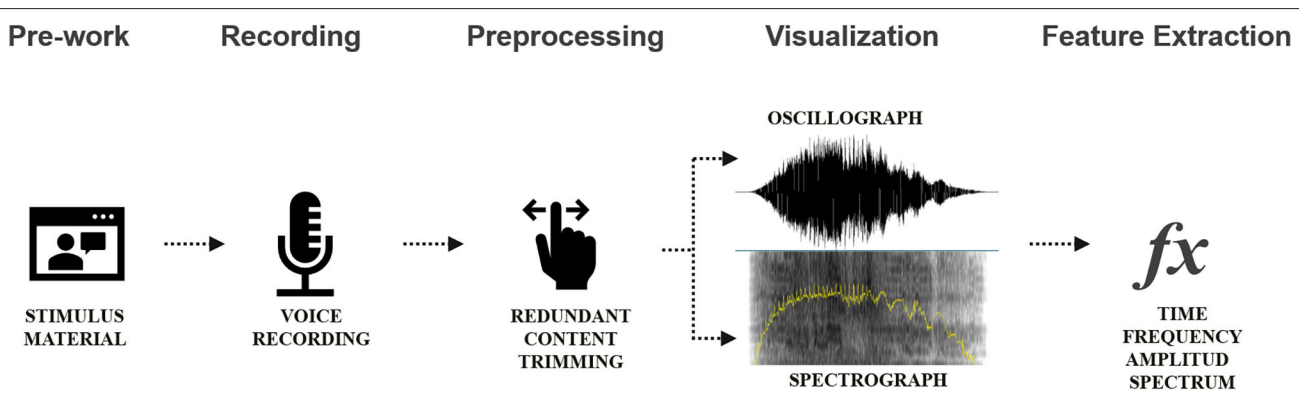


FIGURE 3 | Data collection procedure. Source: drawn by the authors.

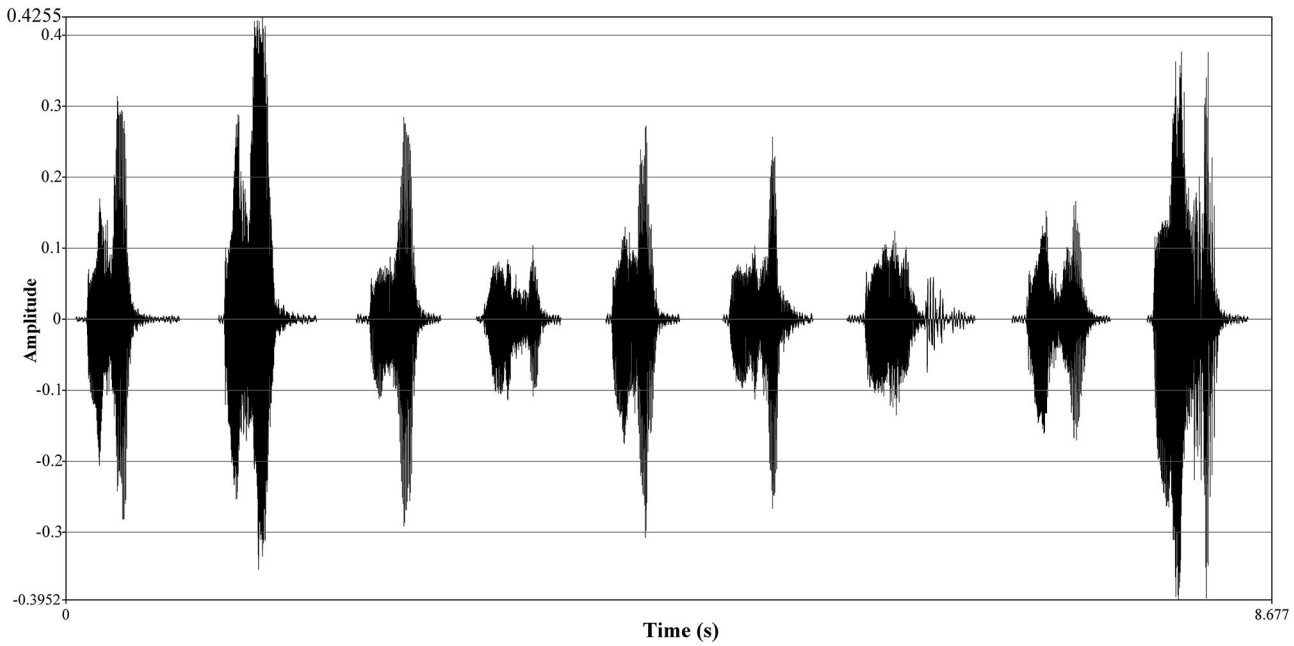


FIGURE 4 | Oscillograph comparison of different emotional voices of respondents. The *Y*-axis of the oscillograph expresses time (unit: s). The *X*-axis, amplitude, has different units of expressions, either decibel (dB) or relative values; it ranges between [−1, 1] and can be expressed by a percentage or frequency value (Sueur, 2018; Wayland, 2018). From the left to the right, a respondent records nine emotions of "你好" from ID.1 to ID.9.

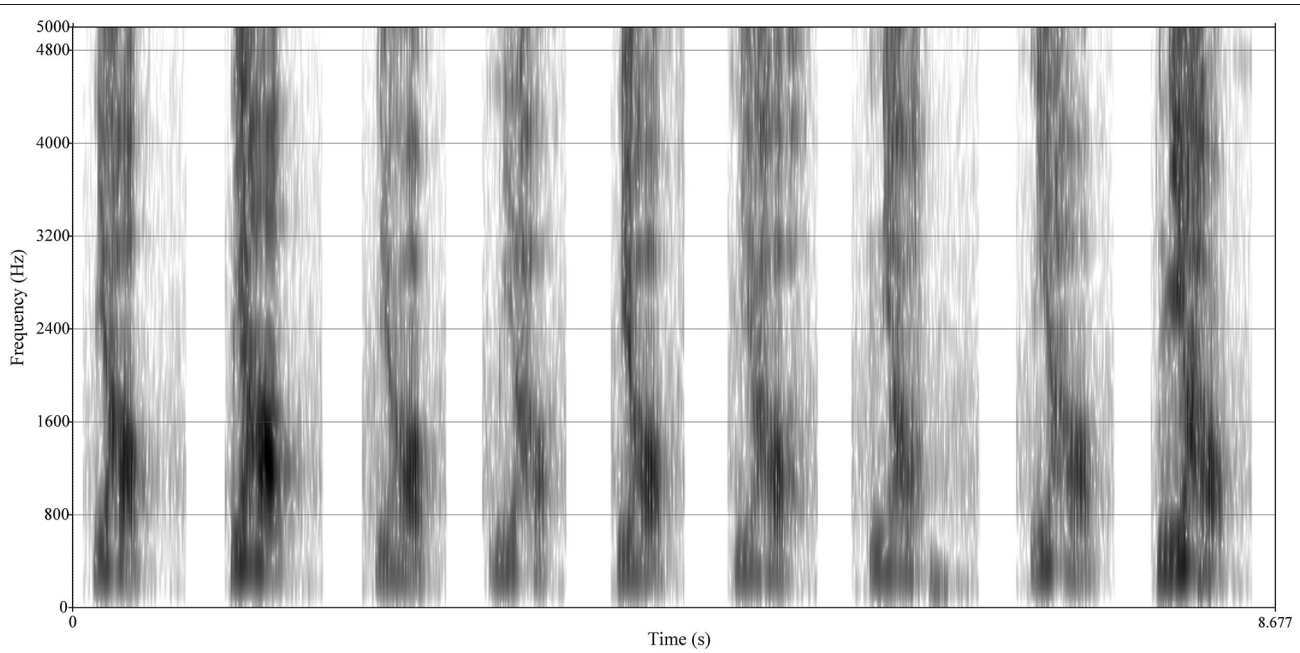


FIGURE 5 | Comparison of spectrographs of respondents among different emotions. The *Y*-axis of the spectrograph is the same as the waveform and expresses the amplitude. The *X*-axis represents frequency (unit: Hz). The frequency spectrum is the variation of voice energy with frequency. In addition, different amplitudes (or loudness) were expressed by the color gradient of data points.

In phonetics, jitter reflects the fast repeated changes of the fundamental frequency, and it primarily describes the variation amplitude of any fundamental frequency. As shown below,

$$jitter_{absolute} = \sum_{i=2}^{N} |T_i - T_{i-1} / (N-1) \quad (1)$$

T_i is the duration of the pitch period i (unit: ms), and N is the quantity of all pitch periods. $\text{Jitter}_{\text{absolute}}$ calculates the absolute mean of differences between any two adjacent pitch periods. The mean period is calculated using

$$meanPeriod = \sum_{i=1}^{N} T_i/N \tag{2}$$

The jitter percentage is calculated using

$$jitter\% = jitter_{absolute}/meanPeriod \tag{3}$$

The $jitter_{absolute}$ is divided by the *meanPeriod*, deriving the ratio between perturbation of fundamental frequency and mean during the pronunciation.

Calculation of Shimmer Percentage

Shimmer percentage reflects changes of amplitude among different periods and is calculated using

$$shimmer_{absolute} = \sum_{i=2}^{N} \left|A_k - A_{k-1}\right| / (N-1) \tag{4}$$

$$meanShimmer = \sum_{i=1}^{N} A_k/N \tag{5}$$

$$Shimmer\% = shimmer_{absolute}/meanShimmer \tag{6}$$

The mean of amplitude changes between two adjacent periods is calculated from $shimmer_{absolute}$. The *Shimmer%* is the ratio between the mean variation of amplitudes and the average value.

Calculation of HNR

HNR refers to the ratio of the periodic and noise parts in speech signals, and it primarily reflects the hoarse degree of voices. The calculation used to determine HNR is explained below.

The autocorrelation function ($r(x)$) of the voice delay signal x is defined as

$$r(x) = \int s(t)\, 2(t+X)dt \tag{7}$$

where $s(t)$ is the stable time signal, and the function achieves the global maximum when $x = 0$. If the function has global maximum points at other moments in addition to $x = 0$, a period of T_0 *is assumed*. For any positive integer (n), then

$$r(nT_0) = r(0) \tag{8}$$

If no other global maximum points in addition *to* $x = 0$ are detected, then other local maximum points may exist, where

$$r'(\tau) = r_x(x)/r(0) \tag{9}$$

$s(t)$ is defined as the periodic signal with a period of T_0, and N(t) is a noise signal. At $x = 0$, the voice signal is $r(0) = T_H(0) + T_N(0)$. As $r(0) = r_H(0) + r_N(0)$, the following equations can be applied:

$$r'(X_{\max}) = r_H(0)/r(0) \tag{10}$$

$$1 - r'(X_{\max}) = r_H(0)/r(0) \tag{11}$$

$r'(X_{\max})$ describes the size of the relative energy of periodic parts in the voice signals and its complementary set $1 - r'(X_{\max})$ describes the size of the relative energy of noises in the voice signal. HNR can be further defined as

$$\text{HNR (in dB)} = 10 * log_{10} \frac{r'_x(\tau_{\max})}{1 - r'_x(\tau_{\max})} \tag{12}$$

The function has a global maximum when $\tau = 0$, where x(t) is a steady time signal and a global maximum when $\tau = 0$.

RESULTS

The extracted seven-feature data of different emotions of different genders were analyzed using SPSS V.26 to conduct a two-way ANOVA, mixed design. Gender was used as the independent variable, and emotional state was used as the dependent variable to understand the variation in seven acoustic features of different genders under different emotions.

General Conditions of Respondents

A total of 62 respondents, including 31 males and 31 females, were recruited. These participants can be grouped according to age: 21–30 years old: nine females and eight males; 31–40 years old: eight females and eight males; 41–50 years old: eight females and eight males; and 51–60 years old: six females and seven males.

Difference Test Analysis of the Acoustic Parameters

To show significant differences in acoustic features under different emotions and gender, the same respondents were repeatedly measured, testing the seven acoustic features of emotions. Results of the correlation analyses are shown below.

Velocity: relevant data of seconds per word are listed in **Tables 3**, **4**.

The interaction tests for gender and emotional state ($SS = 0.01$; $Df = 2.53$; $MS = 0.00$; $F = 0.25$; $P > 0.05$) did not yield any significant results, i.e., participants' velocity in expressing

TABLE 3 | Fine grids and marginal means of emotional states and gender on acoustic features.

		Gender		Marginal means
		Female	Male	
State	Neutral	0.29	0.26	0.28
	Exuberant	0.28	0.24	0.26
	Bored	0.52	0.48	0.50
	Dependent	0.38	0.36	0.37
	Disdainful	0.29	0.25	0.27
	Relaxed	0.34	0.30	0.32
	Anxious	0.26	0.21	0.23
	Docile	0.33	0.28	0.31
	Hostile	0.28	0.24	0.26
Marginal means		0.33	0.29	0.29

the nine different emotions was not significantly correlated to gender.

Gender main effect: The influence of velocity on overall emotional states varies significantly between males and females ($F = 2587.76$, $p < 0.05$). The velocity ($M = 0.33$) of female respondents under different emotional states is significantly lower than that of males ($M = 0.29$).

State main effect: Velocity under different emotional states varies significantly for the overall factor, gender ($F = 76.37$, $p < 0.05$). According to the multiple comparison, the state anxious ($M = 0.23$) shows the highest velocity, followed by exuberant and hostile ($M = 0.26$), disdainful ($M = 0.27$), neutral ($M = 0.28$), docile ($M = 0.31$), relaxed ($M = 0.32$), dependent ($M = 0.37$), and bored ($M = 0.5$), successively.

Fo (Hz): The interaction test showed significant results for both gender and emotional state ($SS = 72887.47$; $Df = 1.85$; $MS = 39437.31$; $F = 15.90$; $p < 0.05$; $\omega^2 = 0.21$), i.e., participants' Fo (Hz) varied across gender and emotional state. Relevant data abstracts of mean pitch are listed in **Table 5**.

Gender simple main effect: Females show significantly different effects of Fo on emotional states ($F = 111.30$, $p < 0.05$), according to the results of *post hoc* comparisons: (1) > (4); (2) > (1)–(6), (9); (3) > (5); (4) > (5); (6) > (1), (3) (5); (7) > (1), (3)–(6), (9); (8)> (1), (3)–(6), (9); (9) > (1), (3)–(5). Males ($F = 103.96$, $p < 0.05$) also show differences, according to results of *post hoc* comparisons: (1) > (4); (2) > (1)–(6), (8), (9); (3) > (4); (5) > (1), (3), (4); (6) > (1), (3)–(5), (8)–(9); (7) > (1), (3)–(9); (8) > (1). (3)–(5), (9); (9) > (1). (3)–(5). These results demonstrate that ranks of emotional states are different between males and females.

State simple main effect: With respect to influences of Fo (Hz) on gender under different emotional states, F-values of neutral, exuberant, bored, dependent, relaxed, disdainful, anxious, docile, and hostile states are 198.83, 113.02, 147.32, 324.47, 49.67, 51.28, 43.98, 66.71, and 207.12, respectively ($p < 0.05$). According to the results of *post hoc* comparisons, females have a significantly higher Fo than males.

Fo SD: The interaction test was significant across gender and emotional state ($SS = 13144.75$; $Df = 3.67$; $MS = 3586.29$; F

TABLE 4 | Two-way ANOVA abstract of emotional states and gender on velocity.

Variable	*SS*	*Df*	*MS*	*F*	*post hoc* comparisons
Gender	0.24	1	0.24	**2587.77*** **	Female > Male
State$_b$	3.37	2.53	1.33	**76.37*** **	(1) > (2); (2) > (7); (3) > (1)-(9); (4) > (1)-(2), (6)-(9); (5) > (7), (9); (6) > (1)-(2), (5), (7), (9); (8) > (1)-(2), (5), (7), (9); (9) > (7);
Gender X state	0.01	2.53	0.00	0.25	
Block	1.24	60	0.02		
Error	2.65	151.75	0.02		

It indicates that b is the interval design factor (dependent factor).
(1) neutral (2) exuberant (3) bored (4) dependent (5) disdainful (6) relaxed (7) anxious (8) docile (9) hostile.
****p < 0.001.*

TABLE 5 | Test of simple main effect in the mixed design of gender and emotional state in Fo.

Variable	*SS*	*Df*	*MS*	*F*	*post hoc* comparisons
State					
Female	621352.46	1.59	391784	**111.30*** **	(1) > (4); (2) > (1)-(6), (9); (3) > (5); (4) > (5); (6) > (1), (3)-(5); (7) > (1), (3)-(6), (9); (8)> (1), (3)-(6), (9); (9) > (1), (3)-(5).
Male	372745.37	1.54	241754	**103.96*** **	(1) > (4); (2) > (1)-(6), (8), (9); (3) > (4); (5) > (1), (3), (4); (6) > (1), (3)-(5), (8)-(9); (7) > (1), (3)-(9); (8) > (1). (3)-(5), (9); (9) > (1). (3)-(5).
Gender					
Neutral	93367.55	1	93367.55	**198.83*** **	Female ($M = 214.73$) > Male ($M = 137.12$)
Exuberant	143410.15	1	143410.15	**113.02*** **	Female ($M = 314.45$) > Male ($M = 218.26$)
Bored	100950.05	1	100950.05	**147.32*** **	Female ($M = 214.73$) > Male ($M = 134.03$)
Dependent	142056.14	1	142056.14	**324.47*** **	Female ($M = 215.45$) > Male ($M = 118.72$)
Disdainful	39185.22	1	39185.22	**49.67*** **	Female ($M = 188.35$) > Male ($M = 138.03$)
Relaxed	89769.08	1	89769.08	**51.28*** **	Female ($M = 283.05$) > Male ($M = 205.95$)
Anxious	129113.74	1	129113.74	**43.98*** **	Female ($M = 309.57$) > Male (M =218.30)
Docile	290924.98	1	290924.98	**66.71*** **	Female ($M = 309.00$) > Male ($M = 171.10$)
Hostile	188635.07	1	188635.07	**207.12*** **	Female ($M = 279.94$) > Male ($M = 169.62$)

(1) neutral (2) exuberant (3) bored (4) dependent (5) disdainful (6) relaxed (7) anxious (8) docile (9) hostile.
****p < 0.001.*

$= 10.80$; $p < 0.05$; $\omega^2 = 0.15$), i.e., participants' Fo *SD* varied across gender and emotional state. Relevant data abstracts of pitch variability are listed in **Table 6**.

Gender simple main effect: Females show significantly different effects of Fo *SD* on emotional states ($F = 2.43$, $p > 0.05$), according to the results of *post hoc* comparisons: (1) > (4)–(6); (2) > (1), (4)–(8); (3) > (1), (4)–(8); (6) > (4); (7) > (1), (5); (8) > (4)–(6); (9) > (1), (4)–(8). Males ($F = 2.43$, $p > 0.05$) show no significant differences.

State simple main effect: Concerning influences of Fo *SD* on gender under different emotional states, *F* values of exuberant, bored, dependent, and hostile states are 47.88, 92.90, and 9.52, respectively ($p < 0.05$). According to the results of *post hoc* comparisons, females give significantly higher values than males; however, males > females with respect to the dependent variable.

Intensity (dB): The interaction test was significant across gender and emotional state ($SS = 7624.57$; $Df = 1.99$; $MS = 314.08$; $F = 9.25$; $p < 0.05$; $\omega^2 = 0.13$), i.e., participants' intensity varied across gender and emotional state. Relevant data abstracts of mean-sones intensity are listed in **Table 7**.

Gender simple main effect: Both males and females show significantly different effects of intensity (dB) on emotional states: Females ($F = 64.11$, $p < 0.05$) and males ($F = 52.60$, $p < 0.05$). According to *post hoc* comparisons, results of females are (1) > (3)–(4), (7)–(8); (2) > (1)–(8); (4) > (3); (5) > (1), (3)–(4), (7)–(8); (6) > (1), (3)–(4), (7)–(8); (7) > (3); (8) > (3); (9) >

TABLE 6 | Simple main effect test of mixed design of gender and emotional states in Fo *SD*.

Variable	*SS*	*Df*	*MS*	*F*	*post hoc* comparisons
State					
Female	19634.95	2.75	7153.90	**24.69*** **	(1) > (4)-(6); (2) > (1), (4)-(8); (3) > (1), (4)-(8); (6) > (4); (7) > (1), (5); (8) > (4)-(6); (9) > (1), (4)-(8).
Male	3935.73	2.45	1605.58	2.43	
Gender					
Neutral	53.16	1	53.16	0.17	
Exuberant	2462.42	1	2462.42	**47.88*** **	Female ($M = 33.86$) > male ($M = 21.25$)
Bored	8920.32	1	8920.32	**92.90*** **	Female ($M = 37.20$) > male (M =13.21)
Dependent	858.95	1	858.95	**9.52** **	Male (M =20.75) > Female ($M = 13.30$)
Disdainful	2.74	1	2.74	0.01	
Relaxed	381.33	1	381.33	3.92	
Anxious	17.25	1	17.25	0.08	
Docile	11.01	1	11.01	0.15	
Hostile	6328.66	1	6328.66	**15.38*** **	Female ($M = 37.57$) > male ($M = 17.36$)

(1) neutral (2) exuberant (3) bored (4) dependent (5) disdainful (6) relaxed (7) anxious (8) docile (9) hostile.
***$p < 0.01$; ***$p < 0.001$.*

TABLE 7 | Simple main effect test using mixed design of gender and emotional states on intensity (dB).

Variable	*SS*	*Df*	*MS*	*F*	*post hoc* comparisons
State					
Female	5071.70	1.41	3596.94	**64.11*** **	(1) > (3)-(4), (7)-(8); (2) > (1)-(8); (4) > (3); (5) > (1), (3)-(4), (7)-(8); (6) > (1), (3)-(4), (7)-(8); (7) > (3); (8) > (3); (9) > (1), (2) -(8).
Male	2939.66	2.26	1300.21	**52.60*** **	(1) > (3)-(4), (7); (2) > (1), (3)-(9); (3) > (7); (4) > (7); (5) > (1), (3), (4), (7); (6) > (1), (4)-(8); (8) > (3), (4), (7); (9) > (3), (4), (5), (7), (8).
Gender					
Neutral	151.32	1	151.32	3.55	
Exuberant	143.85	1	143.85	2.20	
Bored	820.46	1	820.46	**17.46*** **	Male ($M = 69.65$) > Female ($M = 62.38$)
Dependent	363.48	1	363.48	**8.23** **	Male ($M = 69.52$) > Female ($M = 64.68$)
Disdainful	80.62	1	80.62	1.24	
Relaxed	261.91	1	261.91	3.58	
Anxious	45.29	1	45.29	1.42	
Docile	546.12	1	546.12	**9.88** **	Male ($M = 71.71$) > Female ($M = 65.78$)
Hostile	0.15	1	0.15	0.00	

(1) neutral (2) exuberant (3) bored (4) dependent (5) disdainful (6) relaxed (7) anxious (8) docile (9) hostile.
***$p < 0.01$; ***$p < 0.001$.*

(1), (2)–(8). Results of males are (1) > (3)–(4), (7); (2) > (1), (3)–(9); (3) > (7); (4) > (7); (5) > (1), (3), (4), (7); (6) > (1), (4)-(8); (8) > (3), (4), (7); (9) > (3), (4), (5), (7), (8). The results demonstrate that ranks of emotional states are different between males and females.

State simple main effect: Concerning influences of intensity (dB) on gender under different emotional states, F-values of bored, dependent, and docile are 17.46, 8.23 and 9.88, respectively ($p < 0.05$). According to the results of *post hoc* comparisons, males give significantly higher results than females.

Jitter%: The interaction test resulted in significant outcomes considering gender and emotional state ($S = 230.33$; $Df = 2.60$; $MS = 88.67$; $F = 32.05$; $p < 0.05$; $\omega^2 = 0.35$), i.e., participants' Jitter% varied across gender and emotional state. Relevant data abstracts of the ratio between the fundamental frequency changes and the mean are listed in **Table 8**.

Gender simple main effect: With respect to Jitter% of males and females under different emotional states, females ($F = 25.87$, $p < 0.05$) and males ($F = 37.01$, $p < 0.05$) both have significant effects. According to *post hoc* comparisons, females show (1) > (2), (8); (3) > (2), (8); (4) > (2), (8); (5) > (2), (8)-(9); (6) > (1)–(5), (7)–(9); (7) > (1)–(5), (8)–(9). Males show (1) > (2)–(6), (9); (2) > (5)–(6), (9); (4) > (3)–(6), (9); (7) > (1)–(6), (8)–(9); (8) > (2)–(6), (9). These results demonstrate that ranks of emotional states are different between males and females.

State simple main effect: Concerning influences of Jitter% on gender under different emotional states, F-values of neutral, exuberant, bored, dependent, relaxed, anxious, and docile are 82.90, 63.04, 8.11, 14.52, 23.77, 35.51, and 65.22, respectively ($p < 0.05$). According to the results of *post hoc* comparisons, females > males for relaxed and males > females for the remaining six emotional states.

Shimmer%: The interaction test yielded significant results considering gender and emotional state ($S = 1712.65$; $Df = 4.29$; $MS = 399.46$; $F = 49.4$; $p < 0.05$; $\omega^2 = 0.45$), i.e., participants' Shimmer % varied across gender and emotional state. Relevant data abstracts of intensity perturbations are listed in **Table 9**.

Gender simple main effect: With respect to Shimmer% of males and females under different emotional states, females ($F = 240.70$, $p < 0.05$) and males ($F = 241.26$, $p < 0.05$) both have significant effects. According to *post hoc* comparisons, females show (1) > (2), (4), (8)–(9); (2) > (8)-(9); (3) > (2), (4), (8)-(9); (4) > (8)–(9); (5) > (1)–(4), (8)–(9); (6) > (1)–(4), (8)–(9); (7) > (1)–(4), (8)–(9); (8) > (9). Males show (1) > (2)–(9); (2) > (8)–(9); (3) > (2), (8)–(9); (4) > (2)–(3), (6), (8)–(9); (5) > (2), (6), (8)–(9); (6) > (8)–(9); (7) > (2)–(9); (8) > (9). These results demonstrate that ranks of emotional states are different between males and females.

State simple main effect: Concerning influences of Shimmer% on gender under different emotional states, *F*-values of neutral, exuberant, bored, dependent, disdainful, relaxed, anxious, docile, and hostile are 82.90, 63.04, 8.11, 14.52, 19.99, 23.77, 35.51, 65.22, and 7.58, respectively ($p < 0.05$). According to *post hoc* comparison results, females are significantly higher than males concerning disdainful and relaxed, which is the opposite of the remaining emotional states.

HNR: The interaction test yielded significant results considering gender and emotional state ($SS = 1071.63$; $Df = 3.76$; $MS = 284.69$; $F = 37.42$; $p < 0.05$; $\omega^2 = 0.38$), i.e., participants' HNR varied across gender and emotional state. Relative data abstracts of the ratio of periodic part and noise in signals are listed in **Table 10**.

Gender simple main effect: With respect to HNR of males and females under different emotional states, females ($F = 45.87$, $p < 0.05$) and males ($F = 30.90$, $p < 0.05$) both show a significant effect. According to *post hoc* comparisons, females show (1) > (3), (5)–(7); (2) > (1), (3)–(7); (3) > (6)–(7); (4) > (3), (5)–(7);

TABLE 8 | Simple main effect test of mixed design of gender and emotional states in Jitter%.

Variable	*SS*	*Df*	*MS*	*F*	*post hoc* comparisons
State					
Female	78.25	2.66	29.372	**25.87*** **	(1) > (2), (8); (3) > (2), (8); (4) > (2), (8); (5) > (2), (8)-(9); (6) > (1)-(5), (7)-(9); (7) > (1)-(5), (8)-(9).
Male	419.94	1.88	223.091	**37.01*** **	(1) > (2)-(6), (9); (2) > (5)-(6), (9); (4) > (3)-(6), (9); (7) > (1)-(6), (8)-(9); (8) > (2)-(6), (9).
Gender					
Neutral	78.81	1	78.81	**82.90*** **	Male ($M = 4.19$) > Female ($M = 1.93$)
Exuberant	16.00	1	16.00	**63.04*** **	Male ($M = 2.59$) > Female ($M = 1.57$)
Bored	2.01	1	2.01	**8.11*** **	Male (M = 2.33) > Female ($M = 1.97$)
Dependent	11.24	1	11.24	**14.52*** **	Male ($M = 2.89$) > Female ($M = 2.04$)
Disdainful	0.01	1	0.01	0.02	
Relaxed	18.76	1	18.76	**23.77*** **	Female ($M = 3.18$) > Male ($M = 2.08$)
Anxious	124.25	1	124.25	**35.51*** **	Male ($M = 5.70$) > Female ($M = 2.87$)
Docile	130.76	1	130.76	**65.22*** **	Male ($M = 2.10$) > Female ($M = 1.75$)
Hostile	1.86	1	1.86	2.61	

Note. (1) neutral (2) exuberant (3) bored (4) dependent (5) disdainful (6) relaxed (7) anxious (8) docile (9) hostile.
****$p < 0.001$.*

TABLE 9 | Simple main effect test of mixed design of gender and emotional states in Shimmer%.

Variable	*SS*	*Df*	*MS*	*F*	*post hoc* comparisons
State					
Female	7026.62	2.69	2617.24	**240.70*** **	(1) > (2), (4), (8)-(9); (2) > (8)-(9); (3) > (2), (4), (8)-(9); (4) > (8)-(9); (5) > (1)-(4), (8)-(9); (6) > (1)-(4), (8)-(9); (7) > (1)-(4), (8)-(9); (8) > (9).
Male	9662.34	3.42	2825.46	**241.26*** **	(1) > (2)-(9); (2) > (8)-(9); (3) > (2), (8)-(9); (4) > (2)-(3), (6), (8)-(9); (5) > (2), (6), (8)-(9); (6) > (8)-(9); (7) > (2)-(9); (8) > (9).
Gender					
Neutral	1276.10	1	1276.10	**184.14*** **	Male (M = 18.73) > Female (M = 9.65)
Exuberant	22.72	1	22.72	**5.78***	Male (M = 9.24) > Female (M = 8.03)
Bored	320.89	1	320.89	**7.45****	Male (M =11.12) > Female (M = 9.84)
Dependent	320.89	1	320.89	**45.81*** **	Male (M = 12.93) > Female (M = 8.38)
Disdainful	218.49	1	218.49	**19.99*** **	Female (M = 15.55) > Male (M = 11.79)
Relaxed	50.92	1	50.92	**6.48***	Female (M = 12.30) > Male (M =10.48)
Anxious	94.24	1	94.24	**9.02****	Male (M = 15.81) > Female (M = 13.34)
Docile	0.05	1	0.05	**11.33****	Male (M = 0.33) > Female (M = 0.28)
Hostile	0.03	1	0.03	**7.58****	Male (M =.28) > Female (M = 0.24)

(1) neutral (2) exuberant (3) bored (4) dependent (5) disdainful (6) relaxed (7) anxious (8) docile (9) hostile.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.*

TABLE 10 | Simple main effect test of mixed design of gender and emotional states in HNR.

Variable	*SS*	*Df*	*MS*	*F*	*post hoc* comparisons
State					
Female	1500.08	1.99	754.38	**45.87*** **	(1) > (3), (5)-(7); (2) > (1), (3)-(7); (3) > (6)-(7); (4) > (3), (5)-(7); (5) > (6)-(7); (8) > (1)-(7); (9) > (1), (3)-(7).
Male	759.60	2.92	259.89	**30.90*** **	(2) > (1), (7); (3) > (1)-(2), (4)-(5), (7)-(9); (4) > (1); (5) > (1), (7); (6) > (1)-(2), (4), (7)-(9); (8) > (1)-(2), (7); (9) > (1)-(2), (7).
Gender					
Neutral	369.81	1	369.81	**62.35*** **	Female (M =13.08) > Male (M = 8.19)
Exuberant	261.34	1	261.34	**40.59*** **	Female (M =15.14) > Male (M = 11.03)
Bored	38.21	1	38.21	**8.50****	Male (M = 12.99) > Female (M = 11.42)
Dependent	65.90	1	65.90	**18.12*** **	Female (M =13.54) > Male (M = 11.48)
Disdainful	3.56	1	3.56	0.44	
Relaxed	246.12	1	246.12	**49.74*** **	Male (M = 13.09) > Female (M = 9.10)
Anxious	10.16	1	10.16	4.02	
Docile	210.75	1	210.75	**22.60*** **	Female (M =15.97) > Male (M = 12.29)
Hostile	132.10	1	132.10	**15.43*** **	Female (M =14.93) > Male (M = 12.01)

(1) neutral (2) exuberant (3) bored (4) dependent (5) disdainful (6) relaxed (7) anxious (8) docile (9) hostile.
**$p < 0.01$; ***$p < 0.001$.*

(5) > (6)–(7); (8) > (1)–(7); (9) > (1), (3)–(7). Males show (2) > (1), (7); (3) > (1)–(2), (4)–(5), (7)–(9); (4) > (1); (5) > (1), (7); (6) > (1)–(2), (4), (7)–(9); (8) > (1)–(2), (7); (9) > (1)–(2), (7). These results demonstrate that ranks of emotional states are different between males and females.

State simple main effect: With respect to influences of HNR on gender under different emotional states, F-values of neutral, exuberant, bored, dependent, relaxed, docile, and hostile are 62.35, 40.59, 8.50, 18.12, 49.74, 22.60, and 15.43, respectively ($p < 0.05$). According to the results of *post hoc* comparisons, males give significantly higher values than females in terms of bored and relaxed although the opposite phenomenon is observed for the remaining five emotional states.

DISCUSSION AND CONCLUSIONS

This study focuses on physical quantities of acoustic features and their differences according to gender and the emotional states of the PAD model during emotion–voice interactions of AI. The study found significant differences in users' gender and emotional states of the PAD model with respect to seven major acoustic features: (1) With respect to gender and emotional states, Fo (Hz), Fo *SD*, intensity (dB), Jitter%, Shimmer%, and HNR have interactions, and velocity displays no interaction. (2) There are significant gender differences in terms of velocity of eight emotional states in PAD. Moreover, males show significantly higher velocity (M = 0.29) compared to females

($M = 0.33$). (3) Males show no significant differences in six of the acoustic features, except Fo *SD*. Looking at the gender simple main effect, there are significant gender differences in terms of degree and ranking of emotional states. Looking at the state simple main effect, Fo (Hz) shows significant differences among different emotional states. Fo *SD* is significantly different in terms of exuberant, bored, dependent, and hostile states. Intensity (dB) is significantly different with respect to bored, dependent, and docile states. There are significant differences in Jitter% in neutral, exuberant, bored, dependent, relaxed, anxious, and docile states. Shimmer% has significant differences. HNR presents significant differences in neutral, exuberant, bored, dependent, relaxed, docile, and hostile states. The above analyses found physical quantities of relevant parameters and rankings as shown in the results. Specifically, the voice-affective interaction of intelligent products was used as the preset scene. Therefore, the PAD model is different in terms of emotional classification from the emotional classification found in the literature review (Williams and Stevens, 1972; Johnstone and Scherer, 1999; Abelin and Allwood, 2000; Quinto et al., 2013; Bowman and Yamauchi, 2016; Dasgupta, 2017; Hildebrand et al., 2020). Moreover, some acoustic features are different, and it is impossible to compare directly. Directionality of classification is compared with research results, which has not been investigated in past empirical studies; however, there are significant differences in rhythms of different emotions. For gender, previous studies mainly found that men speak more quickly than women (Feldstein et al., 1993; Verhoeven et al., 2004; Jacewicz et al., 2010), but it has also been found that there is no significant difference between men and women (Robb et al., 2004; Sturm and Seery, 2007; Nip and Green, 2013). This study further compared expressions of emotional states and concluded that men speak more quickly than women.

We comprehensively explored the influence of eight emotional states of the PAD model and gender on affective recognition and expression of acoustic features (e.g., velocity, Fo, frequency spectra) in a systematic method. In terms of theoretical implications, the PAD model of intelligent products provides an emotional model that is different from previously used models. In emotional computing, the PAD model is conducive to understanding the influences of gender and emotional states on the connection between acoustic features and psychology in AI affective-voice interaction, including physical variables and their differences. This aids in understanding the acoustic features of affective recognition and expression. In terms of practical applications, in view of the development trends of intelligent products on the market, man–machine interaction will be popularized in intelligent-home life, travel, leisure, entertainment, education, and medicine in the future. This study will help to improve the affective-voice interaction scenes of intelligent products and connections between the emotional states and acoustic features of the speaker. The analysis of acoustic features under different emotions and genders provides an empirical foundation for adjusting the parameters of the affective-voice interaction mathematical models and offsets limitations of current deep learning acoustic models' "explanatory" power. The research results can provide a reference for the adjustment of model parameters during optimization of affective recognition and affective expression.

This study was designed for theoretical and practical application; however, the recorded voices only used Chinese materials. There may be some differences with different languages, which deserves particular attention for generalization of the results. Subsequent studies can further investigate correlations between emotional classification of PAD and voice rhythm of different genders in the PAD model to provide a theoretical basis and supplement shortages of deep learning. This study aims to strengthen emotional integration during man–machine interaction, allowing users and products to generate the empathy effect and, thus, expand the human–computer relationship and highlighting the value of products.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

K-LH: conceptualization and writing. K-LH and S-FD: methodology and formal analysis. K-LH and XL: investigation. S-FD and XL: resources. K-LH: organized the database and analyzed and interpreted the data.

FUNDING

This study was supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJZD-K201901001) and (Grant No. KJZD-M201801001).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.664925/full#supplementary-material

REFERENCES

Abelin, A., and Allwood, J. (2000). "Cross linguistic interpretation of emotional prosody," in *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion* (Newcastle), 110–113.

Apple, W., Streeter, L. A., and Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *J. Personal. Soc. Psychol.* 37, 715–727. doi: 10.1037/0022-3514.37.5.715

Asutay, E., and Västfjäll, D. (2012). Perception of loudness is influenced by emotion. *PLoS ONE* 7:e38660. doi: 10.1371/journal.pone.0038660

Audibert, N., Vincent, D., Aubergé, V., and Rosec, O. (2006). "Expressive speech synthesis: evaluation of a voice quality centered coder on the different acoustic dimensions," in *Proc. Speech Prosody: Citeseer*, 525–528.

Awan, S. N. (1993). Superimposition of speaking voice characteristics and phonetograms in untrained and trained vocal groups. *J. Voice* 7, 30–37. doi: 10.1016/S0892-1997(05)80109-2

Awan, S. N. (2006). The aging female voice: acoustic and respiratory data. *Clin. Linguist. Phone.* 20, 171–180. doi: 10.1080/02699200400026918

Bachorowski, J.-A. (1999). Vocal expression and perception of emotion. *Curr. Dir. Psychol. Sci.* 8, 53–57.

Badshah, A. M., Ahmad, J., Rahim, N., and Baik, S. W. (2019). "Speech emotion recognition from spectrograms with deep convolutional neural network," in *2017 International Conference on Platform Technology and Service* (PlatCon), 1–5. doi: 10.1109/PlatCon.2017.7883728

Baken, R. J., and Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice.* San Diego, CA: Singular Thomson Learning.

Bänziger, T., Hosoya, G., and Scherer, K. R. (2015). Path models of vocal emotion communication. *PLoS ONE* 10:e0136675. doi: 10.1371/journal.pone.0136675

Bänziger, T., and Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Commun.* 46, 252–267. doi: 10.1016/j.specom.2005.02.016

Bitouk, D., Verma, R., and Nenkova, A. (2010). Class-level spectral features for emotion recognition. *Speech Commun.* 52, 613–625. doi: 10.1016/j.specom.2010.02.010

Boersma, P. (1993). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in *Proceedings of the Institute of Phonetic* Sciences (Citeseer), 97–110.

Bowman, C., and Yamauchi, T. (2016). Perceiving categorical emotion in sound: the role of timbre. *Psychomusicol. Music Mind Brain* 26, 15–25. doi: 10.1037/pmu0000105

Brenner, M., Doherty, E. T., and Shipp, T. (1994). Speech measures indicating workload demand. *Aviat. Space Environ. Med.* 65, 21–26.

Brockmann, M., Drinnan, M. J., Storck, C., and Carding, P. N. (2011). Reliable Jitter and Shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task. *J. Voice* 25, 44–53. doi: 10.1016/j.jvoice.2009.07.002

Brockmann, M., Storck, C., Carding, P. N., and Drinnan, M. J. (2008). Voice loudness and gender effects on jitter and shimmer in healthy adults. *J. Speech Lang. Hear. Res.* 51, 1152–1160. doi: 10.1044/1092-4388(2008/06-0208)

Brown, W. S., Morris, R. J., Hicks, D. M., and Howell, E. (1993). Phonational profiles of female professional singers and nonsingers. *J. Voice* 7, 219–226. doi: 10.1016/S0892-1997(05)80330-3

Burgoon, J. K., Birk, T., and Pfau, M. (1990). Nonverbal behaviors, persuasion, and credibility. *Hum. Commun. Res.* 17, 140–169. doi: 10.1111/j.1468-2958.1990.tb00229.x

Chauhan, R., Yadav, J., Koolagudi, S. G., and Rao, K. S. (2011). "Text independent emotion recognition using spectral features," in *International Conference on Contemporary Computing,* ed A.S.et al. (Springer). doi: 10.1007/978-3-642-22606-9_37

Chen, X., Yang, J., Gan, S., and Yang, Y. (2012). The contribution of sound intensity in vocal emotion perception: behavioral and electrophysiological evidence. *PLoS ONE* 7:e30278. doi: 10.1371/journal.pone.0030278

Chen, Y., and Long, R. (2013). Trainable emotional speech synthesis based on PAD. *Pattern Recogn. Artif. Intell.* 26, 1019–1025.

Childers, D. G., and Wu, K. (1991). Gender recognition from speech. Part II: Fine analysis. *J. Acoust. Soc. Am.* 90, 1841–1856. doi: 10.1121/1.401664

Chkroun, M., and Azaria, A. (2019). LIA: a virtual assistant that can be taught new commands by speech. *Int. J. Hum. Comp. Interact.* 35, 1596–1607. doi: 10.1080/10447318.2018.1557972

Clark, A. V. (2005). *Psychology of Moods*. Hauppauge, NY: Nova Publishers.

Colton, R. H., Casper, J. K., and Leonard, R. (2006). *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment.* Philadelphia, PA: Lippincott Williams & Wilkins.

Dai, W., Han, D., Dai, Y., and Xu, D. (2015). Emotion recognition and affective computing on vocal social media. *Inf. Manag.* 52, 777–788. doi: 10.1016/j.im.2015.02.003

Dale, R. (2016). The return of the chatbots. *Nat. Lang. Eng.* 22, 811–817. doi: 10.1017/S1351324916000243

Dasgupta, P. B. (2017). Detection and analysis of human emotions through voice and speech pattern processing. *Int. J. Comput. Trends Technol.* 52, 1–3. doi: 10.14445/22312803/IJCTT-V52P101

Ekman, P., and Oster, H. (1979). Facial expressions of emotion. *Ann. Rev. Psychol.* 30, 527–554. doi: 10.1146/annurev.ps.30.020179.002523

Evans, B. P., Xue, B., and Zhang, M. (2019). "What's inside the black-box? a genetic programming method for interpreting complex machine learning models," in *Proceedings of the Genetic and Evolutionary Computation Conference* (Prague: Association for Computing Machinery).

Farrús, M., Hernando, J., and Ejarque, P. (2007). "Jitter and shimmer measurements for speaker recognition," in *Eighth Annual Conference of the International Speech Communication Association)*, 778–781.

Feldstein, S., Dohm, F.-A., and Crown, C. L. (1993). Gender as a mediator in the perception of speech rate. *Bull. Psychon. Soc.* 31, 521–524.

Fernandes, J., Teixeira, F., Guedes, V., Junior, A., and Teixeira, J. P. (2018). Harmonic to noise ratio measurement - selection of window and length. *Proc. Comput. Sci.* 138, 280–285. doi: 10.1016/j.procs.2018.10.040

Ferrand, C. T. (2007). *Speech Science: An Integrated Approach to Theory and Clinical Practice*. Boston, MA: Pearson/Allyn and Bacon.

Fiebig, A., Jordan, P., and Moshona, C. C. (2020). Assessments of acoustic environments by emotions–the application of emotion theory in soundscape. *Front. Psychol.* 11: 3261. doi: 10.3389/fpsyg.2020.573041

Gao, F., Sun, X., Wang, K., and Ren, F. (2016). "Chinese micro-blog sentiment analysis based on semantic features and PAD model," in *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)* (Okayama), 1–5. doi: 10.1109/ICIS.2016.7550903

Gunes, H., Schuller, B., Pantic, M., and Cowie, R. (2011). "Emotion representation, analysis and synthesis in continuous space: a survey," in *2011 IEEE International Conference on Automatic Face and Gesture Recognition (FG)*(Santa Barbara, CA,), 827–834. doi: 10.1109/FG.2011.5771357

Guo, F., Li, F., Lv, W., Liu, L., and Duffy, V. G. (2020). Bibliometric analysis of affective computing researches during 1999 2018. *Int. J. Hum. Comp. Interact.* 36, 801–814. doi: 10.1080/10447318.2019.1688985

Guyer, J. J., Fabrigar, L. R., and Vaughan-Johnston, T. I. (2019). Speech rate, intonation, and pitch: Investigating the bias and cue effects of vocal confidence on persuasion. *Personal. Soc. Psychol. Bull.* 45, 389–405. doi: 10.1177/0146167218787805

Hammerschmidt, K., and Jürgens, U. (2007). Acoustical correlates of affective prosody. *J. Voice* 21, 531–540. doi: 10.1016/j.jvoice.2006.03.002

Han, J., Zhang, Z., Cummins, N., and Schuller, B. (2019). Adversarial training in affective computing and sentiment analysis: recent advances and perspectives [review article]. *IEEE Comput. Intell. Mag.* 14, 68–81. doi: 10.1109/MCI.2019.2901088

Harmon-Jones, C., Bastian, B., and Harmon-Jones, E. (2016). The discrete emotions questionnaire: a new tool for measuring state self-reported emotions. *PLoS ONE* 11:e0159915. doi: 10.1371/journal.pone.0159915

Harper, R. H. R. (2019). The role of HCI in the age of AI. *Int. J. Hum. Comp. Interact.* 35, 1331–1344. doi: 10.1080/10447318.2019.1631527

Heracleous, P., and Yoneyama, A. (2019). A comprehensive study on bilingual and multilingual speech emotion recognition using a two-pass classification scheme. *PLoS ONE* 14:e0220386. doi: 10.1371/journal.pone.0220386

Hildebrand, C., Efthymiou, F., Busquet, F., Hampton, W. H., Hoffman, D. L., and Novak, T. P. (2020). Voice analytics in business research: Conceptual foundations, acoustic feature extraction, and applications. *J. Bus. Res.* 121, 364–374. doi: 10.1016/j.jbusres.2020.09.020

Hirschberg, J., and Manning, C. D. (2015). Advances in natural language processing. *Science* 349, 261–266. doi: 10.1126/science.aaa8685

Huttar, G. L. (1968). Relations between prosodic variables and emotions in normal American English utterances. *J. Speech Hear. Res.* 11, 481–487. doi: 10.1044/jshr.1103.481

Ivanović, M., Budimac, Z., Radovanović, M., Kurbalija, V., Dai, W., Bădică, C., et al. (2015). Emotional agents-state of the art and applications. *Comput. Sci. Inf. Syst.* 12, 1121–1148. doi: 10.2298/CSIS141026047I

Izard, C. E. (1991). *The Psychology of Emotions*. New York, NY: Plenum Press. doi: 10.1007/978-1-4899-0615-1

Jacewicz, E., Fox, R. A., and Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *J. Acoust. Soc. Am.* 128, 839–850. doi: 10.1121/1.3459842

Jacob, A. (2016). "Speech emotion recognition based on minimal voice quality features," in *2016 International Conference on Communication and Signal Processing (ICCSP)*(Melmaruvathur), 0886–0890. doi: 10.1109/ICCSP.2016.7754275

Jia, J., Zhang, S., Meng, F., Wang, Y., and Cai, L. (2011). Emotional audio-visual speech synthesis based on PAD. *IEEE Trans. Audio Speech Lang. Process.* 19, 570–582. doi: 10.1109/TASL.2010.2052246

Jiang, X., and Pell, M. D. (2017). The sound of confidence and doubt. *Speech Commun.* 88, 106–126. doi: 10.1016/j.specom.2017.01.011

Johar, S. (2016). "Psychology of voice," in *Emotion, Affect and Personality in Speech* (Cham: Springer), 9–15.

Johnstone, T., and Scherer, K. R. (1999). "The effects of emotions on voice quality," in *Proceedings of the XIVth International Congress of Phonetic Sciences* (Citeseer), 2029–2032.

Jurafsky, D., and Martin, J. H. (2014). *Speech and Language Processing*. Pearson Education.

Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814. doi: 10.1037/0033-2909.129.5.770

Juslin, P. N., and Scherer, K. R. (2005). *Vocal Expression of Affect*. Oxford: Oxford University Press.

Kamiloglu, R. G., Fischer, A. H., and Sauter, D. A. (2020). Good vibrations: a review of vocal expressions of positive emotions. *Psychon. Bull. Rev.* 1–29. doi: 10.3758/s13423-019-01701-x

Kim, B., Khanna, R., and Koyejo, O.O. (2016). "Examples are not enough, learn to criticize! criticism for interpretability," in *Advances in Neural Information Processing Systems*, 2280–2288.

Kratzwald, B., Ilić, S., Kraus, M., Feuerriegel, S., and Prendinger, H. (2018). Deep learning for affective computing: text-based emotion recognition in decision support. *Decision Supp. Syst.* 115, 24–35. doi: 10.1016/j.dss.2018.09.002

Kwon, O.-W., Chan, K., Hao, J., and Lee, T.-W. (2003). "Emotion recognition by speech signals," in *Eighth European Conference on Speech Communication and Technology* (Geneva).

Li, X., Tao, J., Johnson, M. T., Soltis, J., Savage, A., Leong, K. M., et al. (2007). "Stress and emotion classification using Jitter and Shimmer Features," in: *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, IV-1081-IV-1084. doi: 10.1109/ICASSP.2007.367261

Liu, X., Xu, Y., Alter, K., and Tuomainen, J. (2018). Emotional connotations of musical instrument timbre in comparison with emotional speech prosody: evidence from acoustics and event-related potentials. *Front. Psychol.* 9:737. doi: 10.3389/fpsyg.2018.00737

Łtowski, T. (2014). Timbre, tone color, and sound quality: concepts and definitions. *Arch. Acoust.* 17, 17–30.

Mallory, E. B., and Miller, V. R. (1958). A possible basis for the association of voice characteristics and personality traits. *Speech Monogr.* 25, 255–260. doi: 10.1080/03637755809375240

Mehrabian, A. (1996a). Analysis of the big-five personality factors in terms of the PAD temperament model. *Aust. J. Psychol.* 48, 86–92. doi: 10.1080/00049539608259510

Mehrabian, A. (1996b). Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament. *Curr. Psychol.* 14, 261–292. doi: 10.1007/BF02686918

Mehrabian, A., and Russell, J. A. (1974). *An Approach to Environmental Psychology*. Cambridge, MA: The MIT Press.

Miller, N., Maruyama, G., Beaber, R. J., and Valone, K. (1976). Speed of speech and persuasion. *J. Personal. Soc. Psychol.* 34, 615–624. doi: 10.1037/0022-3514.34.4.615

Mohammadi, G., and Vinciarelli, A. (2012). Automatic personality perception: prediction of trait attribution based on prosodic features. *IEEE Trans. Affect. Comput.* 3, 273–284. doi: 10.1109/T-AFFC.2012.5

Molnar, C. (2020). *Interpretable Machine Learning*. Lulu.com.

Mullennix, J. W., Johnson, K. A., Topcu-Durgun, M., and Farnsworth, L. M. (1995). The perceptual representation of voice gender. *J. Acoust. Soc. Am.* 98, 3080–3095. doi: 10.1121/1.413832

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. U.S.A.* 116, 22071–22080. doi: 10.1073/pnas.1900654116

Murray, I. R., and Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *J. Acoust. Soc. Am.* 93, 1097–1108. doi: 10.1121/1.405558

Nguyen, Q. N., Ta, A., and Prybutok, V. (2019). An integrated model of voice-user interface continuance intention: the gender effect. *Int. J. Hum. Comp. Interact.* 35, 1362–1377. doi: 10.1080/10447318.2018.1525023

Nip, I. S., and Green, J. R. (2013). Increases in cognitive and linguistic processing primarily account for increases in speaking rate with age. *Child Dev.* 84, 1324–1337. doi: 10.1111/cdev.12052

Noroozi, F., Marjanovic, M., Njegus, A., Escalera, S., and Anbarjafari, G. (2018). A study of language and classifier-independent feature analysis for vocal emotion recognition. *arXiv* [Preprint] arXiv:1811.08935.

Osgood, C. E. (1966). Dimensionality of the semantic space for communication via facial expressions. *Scand. J. Psychol.* 7, 1–30. doi: 10.1111/j.1467-9450.1966.tb01334.x

Osuna, E., Rodríguez, L.-F., Gutierrez-Garcia, J. O., and Castro, L. A. (2020). Development of computational models of emotions: a software engineering perspective. *Cogn. Syst. Res.* 60, 1–19. doi: 10.1016/j.cogsys.2019.11.001

Özseven, T. (2018). Investigation of the effect of spectrogram images and different texture analysis methods on speech emotion recognition. *Appl. Acoustics* 142, 70–77. doi: 10.1016/j.apacoust.2018.08.003

Pell, M. D., and Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE* 6:e27256. doi: 10.1371/journal.pone.0027256

Pernet, C., and Belin, P. (2012). The role of pitch and timbre in voice gender categorization. *Front. Psychol.* 3:23. doi: 10.3389/fpsyg.2012.00023

Picard, R. W. (2000). *Affective Computing*. Cambridge, MA: MIT Press.

Picard, R. W. (2003). Affective computing: challenges. *Int. J. Hum. Comp. Stud.* 59, 55–64. doi: 10.1016/S1071-5819(03)00052-1

Picard, R. W. (2010). Affective computing: from laughter to IEEE. *IEEE Trans. Affect. Comput.* 1, 11–17. doi: 10.1109/T-AFFC.2010.10

Pollack, I., Rubenstein, H., and Horowitz, A. (1960). Communication of verbal modes of expression. *Lang. Speech* 3, 121–130. doi: 10.1177/002383096000300301

Poria, S., Cambria, E., Bajpai, R., and Hussain, A. (2017). A review of affective computing: from unimodal analysis to multimodal fusion. *Inf. Fusion* 37, 98–125. doi: 10.1016/j.inffus.2017.02.003

Quinto, L., Thompson, W., and Keating, F. (2013). Emotional communication in speech and music: the role of melodic and rhythmic contrasts. *Front. Psychol.* 4:184. doi: 10.3389/fpsyg.2013.00184

Ray, G. B. (1986). Vocally cued personality prototypes: an implicit personality theory approach. *Commun. Monogr.* 53, 266–276. doi: 10.1080/03637758609376141

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). Model-agnostic interpretability of machine learning. *arXiv* [Preprint] arXiv:1606.05386.

Robb, M. P., Maclagan, M. A., and Chen, Y. (2004). Speaking rates of American and New Zealand varieties of English. *Clin. Linguist. Phonet.* 18, 1–15. doi: 10.1080/026992031000105336

Rukavina, S., Gruss, S., Hoffmann, H., Tan, J.-W., Walter, S., and Traue, H. C. (2016). Affective computing and the impact of gender and age. *PLoS ONE* 11:e0150584. doi: 10.1371/journal.pone.0150584

Russell, J. A. (1980). A circumplex model of affect. *J. Personal. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714

Sauter, D. A., Eisner, F., Calder, A. J., and Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Q. J. Exp. Psychol.* 63, 2251–2272. doi: 10.1080/17470211003721642

Schaerlaeken, S., and Grandjean, D. (2018). Unfolding and dynamics of affect bursts decoding in humans. *PLoS ONE* 13:e0206216. doi: 10.1371/journal.pone.0206216

Scherer, K. R. (2003). Vocal communication of emotion: a review of research paradigms. *Speech Commun.* 40, 227–256. doi: 10.1016/S0167-6393(02)00084-5

Scherer, K. R., Banse, R., Wallbott, H. G., and Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motiv. Emot.* 15, 123–148. doi: 10.1007/BF00995674

Scherer, K. R., and Giles, H. (1979). *Social Markers in Speech*. Cambridge: Cambridge University Press.

Schlosberg, H. (1954). Three dimensions of emotion. *Psychol. Rev.* 61, 81–88. doi: 10.1037/h0054570

Schuller, D. M., and Schuller, B. W. (2021). A review on five recent and near-future developments in computational processing of emotion in the human voice. *Emot. Rev.* 13, 44–50. doi: 10.1177/1754073919898526

Schwark, J. D. (2015). Toward a taxonomy of affective computing. *Int. J. Hum. Comp. Interact.* 31, 761–768. doi: 10.1080/10447318.2015.1064638

Skerry-Ryan, R., Battenberg, E., Xiao, Y., Wang, Y., Stanton, D., Shor, J., et al. (2018). Towards end-to-end prosody transfer for expressive speech synthesis with tacotron. *arXiv* [Preprint] arXiv:1803.09047.

Sloman, A. (1999). Review of affective computing. *AI Mag.* 20:127.

Sturm, J. A., and Seery, C. H. (2007). Speech and articulatory rates of school-age children in conversation and narrative contexts. *Lang. Speech, Hear. Serv. Schools* 38, 47–59. doi: 10.1044/0161-1461(2007/005)

Sueur, J. (2018). *Sound Analysis and Synthesis with R.* Springer. doi: 10.1007/978-3-319-77647-7

Ting, H. N., Chia, S. Y., Abdul Hamid, B., and Mukari, S. Z.-M. S. (2011). Acoustic characteristics of vowels by normal Malaysian Malay young adults. *J. Voice* 25:e305–e309. doi: 10.1016/j.jvoice.2010.05.007

Toivanen, J., Waaramaa, T., Alku, P., Laukkanen, A.-M., Seppänen, T., Väyrynen, E., et al. (2006). Emotions in [a]: a perceptual and acoustic study. *Logoped. Phoniatr. Vocol.* 31, 43-48. doi: 10.1080/14015430500293926

Trouvain, J., and Barry, W. J. (2000). "The prosody of excitement in horse race commentaries," in *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion* (Newcastle), 86-91.

Tucker, P., and Jones, D. M. (1991). Voice as interface: an overview. *Int. J. Hum.Comp. Interact.* 3, 145–170. doi: 10.1080/10447319109526002

Tusing, K. J., and Dillard, J. P. (2000). The sounds of dominance. Vocal precursors of perceived dominance during interpersonal influence. *Hum. Commun. Res.* 26, 148–171. doi: 10.1111/j.1468-2958.2000.tb00754.x

Vempala, N. N., and Russo, F. A. (2018). Modeling music emotion judgments using machine learning methods. *Front. Psychol.* 8:2239. doi: 10.3389/fpsyg.2017.02239

Verhoeven, J., De Pauw, G., and Kloots, H. (2004). Speech rate in a pluricentric language: a comparison between Dutch in Belgium and the Netherlands. *Lang. Speech* 47, 297–308. doi: 10.1177/00238309040470030401

Wang, C.-C., and Huang, H.-T. (2004). Voice acoustic analysis of normal Taiwanese adults. *J.-Chinese Med. Assoc.* 67, 179–184.

Wang, Z., Ho, S.-B., and Cambria, E. (2020). A review of emotion sensing: categorization models and algorithms. *Multimedia Tools Appl.* doi: 10.1007/s11042-019-08328-z

Wayland, R. (2018). *Phonetics: A Practical Introduction*. Cambridge: Cambridge University Press. doi: 10.1017/9781108289849

Weiguo, W., and Hongman, L. (2019). Artificial emotion modeling in PAD emotional space and human-robot interactive experiment. *J. Harbin Inst. Technol.* 51, 29–37.

Weninger, F., Eyben, F., Schuller, B., Mortillaro, M., and Scherer, K. (2013). On the acoustics of emotion in audio: what speech, music, and sound have in common. *Front. Psychol.* 4:292. doi: 10.3389/fpsyg.2013.00292

Williams, C. E., and Stevens, K. N. (1972). Emotions and speech: some acoustical correlates. *J. Acoust. Soc. Am.* 52, 1238–1250. doi: 10.1121/1.1913238

Wundt, W., and Wozniak, R. H. (1998). *Outlines of Psychology*. Oxford: Thoemmes Press.

Yanushevskaya, I., Gobl, C., and and, N.í, Chasaide, A. (2013). Voice quality in affect cueing: does loudness matter? *Front. Psychol.* 4:335. doi: 10.3389/fpsyg.2013.00335

Yonck, R. (2017). *Heart of the Machine: Our Future in a World of Artificial Emotional Intelligence*. New York, NY: Simon and Schuster

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Huang, Duan and Lyu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 28 June 2021
doi: 10.3389/fpsyg.2021.705528

Optimized Projection and Fisher Discriminative Dictionary Learning for EEG Emotion Recognition

Xiaoqing Gu[1], *Yiqing Fan*[2], *Jie Zhou*[3]* *and Jiaqun Zhu*[1]*

[1] School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, China, [2] Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States, [3] School of Electrical and Mechanical Engineering, Shaoxing University, Shaoxing, China

Electroencephalogram (EEG)-based emotion recognition (ER) has drawn increasing attention in the brain–computer interface (BCI) due to its great potentials in human–machine interaction applications. According to the characteristics of rhythms, EEG signals usually can be divided into several different frequency bands. Most existing methods concatenate multiple frequency band features together and treat them as a single feature vector. However, it is often difficult to utilize band-specific information in this way. In this study, an optimized projection and Fisher discriminative dictionary learning (OPFDDL) model is proposed to efficiently exploit the specific discriminative information of each frequency band. Using subspace projection technology, EEG signals of all frequency bands are projected into a subspace. The shared dictionary is learned in the projection subspace such that the specific discriminative information of each frequency band can be utilized efficiently, and simultaneously, the shared discriminative information among multiple bands can be preserved. In particular, the Fisher discrimination criterion is imposed on the atoms to minimize within-class sparse reconstruction error and maximize between-class sparse reconstruction error. Then, an alternating optimization algorithm is developed to obtain the optimal solution for the projection matrix and the dictionary. Experimental results on two EEG-based ER datasets show that this model can achieve remarkable results and demonstrate its effectiveness.

Keywords: EEG signal, emotion recognition, dictionary learning, fisher discrimination criterion, brain computer interface

OPEN ACCESS

Edited by:
Yaoru Sun,
Tongji University, China

Reviewed by:
Yanhui Zhang,
Hebei University of Chinese Medicine, China
Yufeng Yao,
Changshu Institute of Technology, China

***Correspondence:**
Jie Zhou
sxuj_zhou@163.com
Jiaqun Zhu
angelsoar@cczu.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *05 May 2021*
Accepted: *31 May 2021*
Published: *28 June 2021*

Citation:
Gu X, Fan Y, Zhou J and Zhu J (2021) Optimized Projection and Fisher Discriminative Dictionary Learning for EEG Emotion Recognition. Front. Psychol. 12:705528. doi: 10.3389/fpsyg.2021.705528

INTRODUCTION

Brain–computer interface (BCI) has been one of the research hotspots in recent years in health monitoring and biomedicine (Edgar et al., 2020; Ni et al., 2020b). The BCI does not rely on muscles and the peripheral nervous system. It establishes a direct information transmission channel between the brain and the outside world. The electroencephalography (EEG) signals captured by the BCI system are a powerful tool to analyze neural activities and brain conditions. EEG has the advantages of convenience (i.e. non-invasive, non-destructive and simple) and validity (i.e. sensitivity, validity and compatibility) (Sreeja and Himanshu, 2020). EEG signal is an important tool for revealing the emotional state of human beings. It has been shown that when people are in different thinking and emotional states, the rhythm components of EEG signals are different from their waveform. In BCI, the operation of emotion recognition (ER) starts from

external stimuli to subjects, which induce specific emotions such as happiness, sadness, and anger. These stimuli may be videos, images, music, and so on. During the session, EEG data are recorded by EEG devices. Subsequently, the first step is to extract and preprocess useful features obtained from the recorded EEG. The next step is to train the classifier and optimize the parameters. The final step is to test the training model with new EEG data that are not used in the training process.

Traditional machine learning classifiers have been widely used in EEG-based ER, such as support vector machine (SVM) (Zheng et al., 2019), deep learning (Hwang et al., 2020; Song et al., 2020), nearest neighbor classifier (Li et al., 2019), random forest (Fraiwan et al., 2012), and probabilistic neural networks (Nakisa et al., 2018). In recent years, dictionary learning-based methods have achieved great success in EEG-based recognition tasks for BCI (Ameri et al., 2016; Gu et al., 2020; Ni et al., 2020a). In general, dictionary learning-based classification methods often learn the discriminative and robust dictionaries from training samples. The test sample is sparsely represented as a sparse linear combination of atoms by the learned dictionary, and then, the classification task can be carried out according to the reconstruction error and/or the sparse coefficients. Dictionary learning works well-even with noisy EEG signals. Barthélemy et al. (2013) developed an efficient method to represent EEG signals based on the adapted Gabor dictionary and demonstrated on real data that the learned multivariate model is flexible and the learned representation is informative and interpretable. Abolghasemi and Ferdowsi (2015) developed a dictionary learning framework to remove ballistocardiogram (BCG) artifacts from EEG. Given the advantage of the noise-robust sparse dictionary, a new cost function was proposed, which can model BCG artifacts and then remove them from the original EEG signals. Kashefpoor et al. (2019) developed a correlational label consistent K-SVD dictionary learning method applied to EEG-based screening tool. This method was applied to speckle extraction of EEG signals and extracted spectral features in both time and frequency domains. Aiming at the problem that eye movement and blinking can cause artifacts, Kanoga et al. (2019) proposed a multi-scale dictionary learning method to eliminate eye artifacts from single-channel measurement. Specifically, the time-domain waveforms related to repetitive phase events in EEG signals were learned within the framework of dictionary learning. And the proposed multi-scale dictionary learning method was used to represent the signal components on different timescales. To achieve the highly accurate classification of EEG in BCI, Huang et al. (2020) developed a signal identification model using sparse representation and fast compressed residual convolutional neural networks (CNNs). The authors used the common spatial patterns to extract EEG signal features and build a redundant dictionary using these features. Then, the proposed deep model as a classifier recognized the input EEG signals.

Although machine learning has achieved good classification performance in some application scenarios, the accuracy and applicability of the classification do not go far enough. Since EEG data provide comprehensive information across different frequency bands to characterize emotions, it was expected to design an ER method, which utilizes the specific discriminative information of each frequency band and preserves the common discrimination information shared by multiple band signals. After the success of dictionary learning, in this study, we propose optimized projection and Fisher discriminative dictionary learning (DDL) for EEG-based ER. According to the Fisher discrimination criterion of minimum within-class sparse reconstruction error and maximum between-class sparse reconstruction error, we learn the discriminative projection to map the multiple band signals into a shared subspace and simultaneously build a shared dictionary that establishes the connection between different bands and represents the characteristics of signals well. Therefore, the joint learning of projection and dictionary ensures the common internal structure of multiple frequency bands of signals to be mined in the subspace.

The main contributions of this study are as follows:

(1) A multiple frequency band collaborative learning is introduced in dictionary learning for the EEG-based ER. This learning mechanism can efficiently integrate the band-independent information and inter-band correlation information.
(2) Through the feature projection matrix, the data of multiple frequency bands are projected into a common projection subspace to keep the latent manifold of EEG signals. Meanwhile, the discriminative dictionary is learned by enforcing the classification criterion so that the learned sparse code has a strong representation and discrimination ability.
(3) This joint optimization method has some benefits. Learning independent projection matrices makes this model easily extensible; meanwhile, learning a dictionary in a subspace allows abandoning extraneous information in the original features. In addition, the alternating optimization procedure ensures the dictionary and projection are optimized at the same time.
(4) These extensive experiments on the SEED and DREAMER datasets demonstrate that the multiple band collaborative learning is effective, and this method can improve the discrimination ability of sparse coding in EEG-based ER.

BACKGROUND

Datasets

The experimental data in this study are taken from two public EEG emotion datasets: SEED (Zheng and Lu, 2015) and DREAMER (Katsigiannis and Ramzan, 2018). **Table 1** briefly describes the information of the two datasets. Both SEED and DREAMER datasets are collected when subjects watched emotion-eliciting movies. In the SEED dataset, each subject participated in three experiments, which were separated into three time periods, corresponding to three sessions, and each session corresponds to 15 EEG data trials. Thus, a total of $15 \times 3 = 45$ trials are formed per subject. The SEED provided five frequency bands: δ band (1–3 Hz), θ band (4–7 Hz), α band (8–13 Hz), β band (14–30 Hz), and γ band (31–50 Hz). For the DREAMER dataset, the data recorded by each

subject contain three parts: 18 experimental signal segments, 18 baseline signal segments corresponding to relaxation state, and 18 corresponding labels. The DREAMER data provided EEG features with frequency bands θ , α , and β .

Machine Learning-Based EEG Signal Processing Program

For machine learning-based EEG-based ER, feature extraction and emotion classification are the critical procedures. Considering the SEED dataset as an example, the process of constructing five frequency band sequences is described in **Figure 1**. Firstly, EEG signals collected by BCI are preprocessed by filtering. Then, according to the characteristics of different rhythms of EEG signals, EEG signals usually can be divided into several rhythmic signal components ranging from 0 to 50 Hz. Secondly, EEG features can be extracted by various strategies.

TABLE 1 | The basic information of SEED and DREAMER datasets.

SEED dataset	DREAMER dataset
15 subjects	23 subjects
15 trials using Chinese movie clips (length of film clips about 4 min) to evoke emotions as negative, positive, and neutral	18 trials using movie clips (length of film clips 65–393 s) to evoke 9 emotions
In one session, 5 s hint before each clip, 45 s self-assessment, and 15 s rest after each clip	At least 61 s of pretrial baseline data available
Emotion rating metric: negative, positive, and neutral	Emotion rating scale: valence, arousal, and dominance on a continuous scale from 1 to 5
62-channel electrode cap	14-channel electrode cap
Sampling rate 1,000 Hz	Sampling rate 128 Hz
Frequency band from 0 to 75 Hz	Frequency band from 4 to 30 Hz

Time-domain, frequency-domain, and non-linear analysis methods are the three types of most commonly used EEG feature extraction methods. The time-domain features aim to capture the temporal information of EEG signals, such as higher-order crossings (HOC) (Petrantonakis and Hadjileontiadis, 2010), Hjorth features (Petrantonakis and Hadjileontiadis, 2010), and event-related potential (ERP) (Brouwer et al., 2015). The frequency-domain features aim to capture primarily the EEG emotion information from a frequency perspective. Then, EEG features can be extracted by various methods, such as rhythm (Bhatti et al., 2016), wavelet packet decomposition (WPD) (Wu et al., 2008), and approximate entropy (AE) (Ko et al., 2009). Non-linear features are extracted from the transformed phase space. Non-linear features contain quantitative measures that represent the complex dynamic characteristics of the EEG signals, such as Lyapunov exponent (Lyap) (Kutepov et al., 2020) and correlation dimension (CorrDim) (Geng et al., 2011). Finally, many machine learning methods are established to handle EEG emotion classification on extracted feature sets.

A common approach to deal with multiple bands of EEG data using traditional dictionary learning methods is to directly concatenate features of multiple bands together in the high-dimensional space and treat this single feature vector as the input to the model. However, dictionary learning may not perform well because different band features usually carry different characteristics of EEG emotion.

Dictionary Learning

Let $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_n] \in \mathbf{R}^{m \times n}$ be a set of m-dimensional n training signals. To minimize the reconstruction error and satisfy the sparsity constraints, the sparse representation and dictionary learning of $\mathbf{X}$ can be accomplished by

$$\begin{aligned} &\min_{\mathbf{D},\mathbf{S}} \|\mathbf{X} - \mathbf{DS}\|_F^2, \\ &s.t. \ \|\mathbf{s}_i\|_0 \leq T, \end{aligned} \tag{1}$$

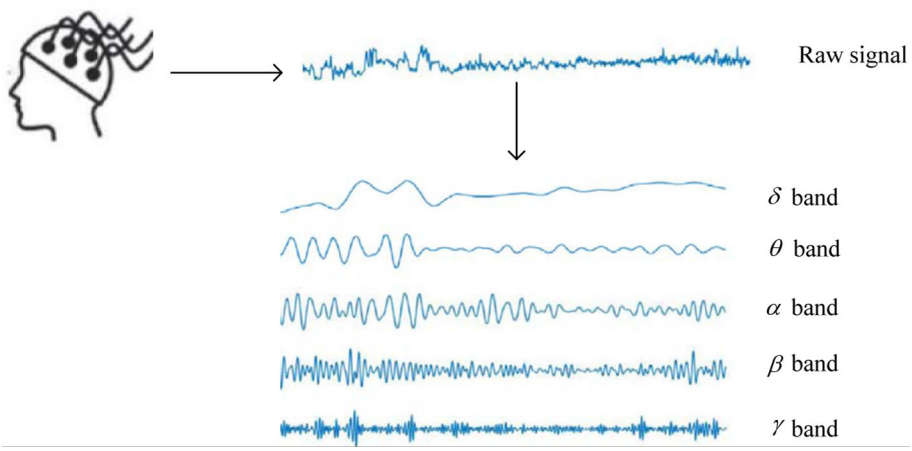


FIGURE 1 | The process of constructing multiple frequency band sequences (Wei et al., 2020).

where $\mathbf{D} = [\mathbf{d}_1, ..., \mathbf{d}_K] \in \mathbf{R}^{m\times K}$ is a dictionary with K atoms $\mathbf{S} = [\mathbf{s}_1, ..., \mathbf{s}_n] \in \mathbf{R}^{m\times K}$ is the sparse coefficient matrix of signals $\mathbf{X}$, and $\mathbf{s}_i$ is the sparse coefficient vector of $\mathbf{x}_i$ over $\mathbf{D}$. T is the sparse constraint factor. The $\|\mathbf{s}_i\|_0 \leq T$ term requires the signal $\mathbf{x}_i$ to have fewer than T non-zero items in its decomposition.

It is not easy to find the optimal sparse solution using ℓ_0-norm regularization term; thus, an alternative formulation of Equation (1) is to replace it with ℓ_1-norm regularization as

$$\begin{array}{l} \min\limits_{\mathbf{D},\mathbf{S}} \|\mathbf{X} - \mathbf{D}\mathbf{S}\|_F^2 + \lambda\|\mathbf{S}\|_1, \\ s.t.\ \|\mathbf{d}_i\|_2^2 = 1, \forall i \end{array}, \tag{2}$$

Equation (2) can be optimized by many efficient ℓ_1 optimization methods, such as the famous K-SVD algorithm (Aharon et al., 2006; Jiang et al., 2013). However, Equation (2) is an unsupervised learning framework. To learn a discriminative dictionary for classification tasks, different kinds of loss functions or Fisher discrimination criterion are considered in the dictionary learning. Fisher discrimination constraints on atoms of the dictionary (Peng et al., 2020) or sparse coefficient $\mathbf{S}$ (Li et al., 2013) or reconstruction error of $\mathbf{X}$ (Zheng and Sun, 2019; Zhang et al., 2021) strive to preserve the class distribution and geometric structure of data.

Suppose data matrix $\mathbf{X}$ consists of samples from C different classes, from $\mathbf{X}$, both the sub-dictionary $\mathbf{D}_i$ and sub-sparse coefficient matrix $\mathbf{S}_i$ are learned for the i-th class data (i = 1, 2,...,C). The whole dictionary $\mathbf{D}$ is represented as $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, ..., \mathbf{D}_C]$. Let $\mathbf{W}_w$ and $\mathbf{W}_b$ denote the within-class scatter and between-class reconstruction error of $\mathbf{X}$, respectively, then

$$\mathbf{W}_w = \sum_j (\mathbf{x}_j\text{-}\mathbf{D}\delta_{l_j}(\mathbf{s}_j)) \times (\mathbf{x}_j\text{-}\mathbf{D}\delta_{l_j}(\mathbf{s}_j))^T, \tag{3}$$

and

$$\mathbf{W}_b = \sum_j (\mathbf{x}_j\text{-}\mathbf{D}\zeta_{l_j}(\mathbf{s}_j)) \times (\mathbf{x}_j\text{-}\mathbf{D}\zeta_{l_j}(\mathbf{s}_j))^T, \tag{4}$$

where $\delta_{l_j}(\,)$ function returns the sparse codes consistent with the class of $\mathbf{x}_j$ (j = 1,2,...,n), and $\zeta_{l_j}(\,)$ function returns the sparse codes not consistent with the class of $\mathbf{x}_j$.

Then, a discriminative dictionary can be learned by reducing the within-class diversity and by increase between-class separation using Equations (3) and (4) (Zheng and Sun, 2019; Zhang et al., 2021).

OPTIMIZED PROJECTION AND FISHER DISCRIMINATIVE DICTIONARY LEARNING

Objective Function

Here, we describe in detail the optimized projection and Fisher DDL model for collaborative learning of multiple frequency band EEG signals. The training framework of the OPFDDL model is illustrated in **Figure 2**. We learn a discriminative projection to map multiple frequency band EEG signals into a common subspace; simultaneously, we learn a common discriminative dictionary to encode the band-invariant information of multiple frequency bands. In particular, to promote the discrimination ability of the model, we utilize the model according to the Fisher discrimination criterion (Gong et al., 2019) under the structure of dictionary learning.

Let $\mathbf{X}^r = \{\mathbf{x}_j^r\}$denote the signal set $\mathbf{X}$ of the frequency band r, where R is the number of frequency bands (r =1,..., R). $\mathbf{x}_j^r$ is the jth sample in $\mathbf{X}^r$. To build the connection between different frequency bands and exploit the specific characteristic of each representation, we project $\mathbf{x}_j^r$ into a feature subspace as $\mathbf{z}_j^r = \mathbf{Q}^r\mathbf{x}_j^r$ by using a transformation matrix $\mathbf{Q}^r \in \mathbf{R}^{m\times d^r}$. Therefore, we obtain $\{\mathbf{z}_j^r\}_{r=1}^R$ by $\{\mathbf{Q}^r\mathbf{x}_j^r\}_{r=1}^R$ as the feature representations for R frequency bands. Then, we denote the within-class reconstruction error J_w^r and between-class reconstruction error J_b^r of the rth frequency band in the projection subspace

$$\begin{aligned} J_w^r &= tr(\sum_{j=1}^{n_r} (\mathbf{Q}^{r\,T}\mathbf{x}_j^r\text{-}\mathbf{Q}^{r\,T}\mathbf{D}\delta_{l_j}(\mathbf{s}_j^r)) \times (\mathbf{Q}^{r\,T}\mathbf{x}_j^r\text{-}\mathbf{Q}^{r\,T}\mathbf{D}\delta_{l_j}(\mathbf{s}_j^r))^T) \\ &= tr(\mathbf{Q}^{r\,T}\sum_{j=1}^{n_r} (\mathbf{x}_j^r\text{-}\mathbf{D}\delta_{l_j}(\mathbf{s}_j^r)) \times (\mathbf{x}_j^r\text{-}\mathbf{D}\delta_{l_j}(\mathbf{s}_j^r))\mathbf{Q}^r) \\ &= tr(\mathbf{Q}^{r\,T}\mathbf{W}_w^r\mathbf{Q}^r) \end{aligned}, \tag{5}$$

and

$$\begin{aligned} J_b^r &= tr(\sum_{j=1}^{n_r} (\mathbf{Q}^{r\,T}\mathbf{x}_j^r\text{-}\mathbf{Q}^{r\,T}\mathbf{D}\zeta_{l_j}(\mathbf{s}_j^r)) \times (\mathbf{Q}^{r\,T}\mathbf{x}_j^r\text{-}\mathbf{Q}^{r\,T}\mathbf{D}\zeta_{l_j}(\mathbf{s}_j^r))^T) \\ &= tr(\mathbf{Q}^{r\,T}\sum_{j=1}^{n_r} (\mathbf{x}_j^r\text{-}\mathbf{D}\zeta_{l_j}(\mathbf{s}_j^r)) \times (\mathbf{x}_j^r\text{-}\mathbf{D}\zeta_{l_j}(\mathbf{s}_j^r))\mathbf{Q}^r) \\ &= tr(\mathbf{Q}^{r\,T}\mathbf{W}_b^r\mathbf{Q}^r) \end{aligned}, \tag{6}$$

where $\mathbf{W}_w^r = \sum\limits_{j=1}^{n_r} (\mathbf{x}_j^r\text{-}\mathbf{D}\delta_{l_j}(\mathbf{s}_j^r)) \times (\mathbf{x}_j^r\text{-}\mathbf{D}\delta_{l_j}(\mathbf{s}_j^r))^T$ is the within-class scatter matrix for sparse coding of the rth frequency band, and $\mathbf{W}_b^r = \sum\limits_{j=1}^{n_r} (\mathbf{x}_j^r\text{-}\mathbf{D}\zeta_{l_j}(\mathbf{s}_j^r)) \times (\mathbf{x}_j^r\text{-}\mathbf{D}\zeta_{l_j}(\mathbf{s}_j^r))^T$ is the between-class scatter matrix for sparse coding of the rth frequency band.

From the classification point of view, minimizing within-class scatter and maximizing between-class scatter in the dictionary learning-based classifier can be represented as

$$\begin{array}{l} \min\limits_{\mathbf{Q}^1,...,\mathbf{Q}^R,\mathbf{D}} \frac{\sum\limits_{r=1}^{R} tr(\mathbf{Q}^{r\,T}\mathbf{W}_w^r\mathbf{Q}^r)}{\sum\limits_{r=1}^{R} tr(\mathbf{Q}^{r\,T}\mathbf{W}_b^r\mathbf{Q}^r)}, \\ s.t.\ (\mathbf{Q}^r)^T(\mathbf{Q}^r) = \mathbf{I},\ \forall r \end{array} \tag{7}$$

With definition as $\tilde{\mathbf{Q}} = [\mathbf{Q}^1, \mathbf{Q}^2, ..., \mathbf{Q}^R]$, $\tilde{\mathbf{W}}_w = \begin{bmatrix} \mathbf{W}_w^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{W}_w^R \end{bmatrix}$, and $\tilde{\mathbf{W}}_b = \begin{bmatrix} \mathbf{W}_b^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{W}_b^R \end{bmatrix}$, Equation (7) can be written as follows:

$$\begin{array}{l} \min\limits_{\tilde{\mathbf{Q}},\mathbf{D}} \frac{tr(\tilde{\mathbf{Q}}^T\tilde{\mathbf{W}}_w\tilde{\mathbf{Q}})}{tr(\tilde{\mathbf{Q}}^T\tilde{\mathbf{W}}_b\tilde{\mathbf{Q}})}, \\ s.t.\ (\tilde{\mathbf{Q}})^T(\tilde{\mathbf{Q}}) = \mathbf{I}, \end{array} \tag{8}$$

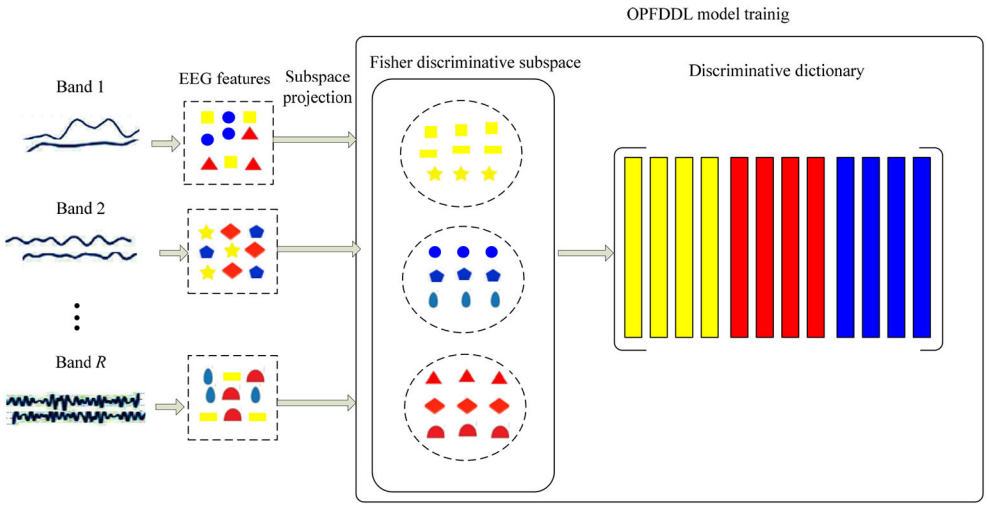


FIGURE 2 | The training framework of optimized projection and Fisher discriminative dictionary learning (OPFDDL) model.

The projection matrix $\tilde{\mathbf{Q}}$ is limited to be orthogonal, which is highly effective in the optimization process. The solution of Equation (8) refers to the complex inverse operation and is computationally intensive. Thus, we translate it into the following quadratic weighted optimization (QWO) problem and obtain the objective function of OPFDDL

$$\begin{aligned} &\min_{\tilde{\mathbf{Q}},\mathbf{D},\mu} \mu^2 tr(\tilde{\mathbf{Q}}^T\tilde{\mathbf{W}}_w\tilde{\mathbf{Q}}) - \mu tr(\tilde{\mathbf{Q}}^T\tilde{\mathbf{W}}_b\tilde{\mathbf{Q}}), \\ &s.t.\ \tilde{\mathbf{Q}}^T\tilde{\mathbf{Q}} = \mathbf{I}, \end{aligned}, \tag{9}$$

It is noted that the parameter μ is an adaptive weight that can be obtained by a closed-form solution but is not a manually adjusted parameter.

Optimization

In the following, the alternating optimization approach is used to update the parameters $\{\tilde{\mathbf{Q}}, \mathbf{D}, \mu\}$ in Equation (9).

(1) Update step for $\tilde{\mathbf{Q}}$. With $\mathbf{D}$ and μ fixed, and with the known $\tilde{\mathbf{W}}_w$ and $\tilde{\mathbf{W}}_b$, the optimization of $\tilde{\mathbf{Q}}$ can be solved by

$$(\mu^2\tilde{\mathbf{W}}_w - \mu\tilde{\mathbf{W}}_b)\tilde{\mathbf{Q}} = \gamma\tilde{\mathbf{Q}}, \tag{10}$$

The projection matrix $\tilde{\mathbf{Q}}$ is constituted by the feature vector corresponding to the first d minimum eigenvalues of Equation (10).

(2) Update step for $\mathbf{D}$. With the definition of $\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{X}^R \end{bmatrix}$, $\boldsymbol{\delta}_{l_j}^r = [\delta_{l_j}(\mathbf{s}_1^r), \delta_{l_j}(\mathbf{s}_2^r), ..., \delta_{l_j}(\mathbf{s}_{n_r}^r)]$, $\boldsymbol{\zeta}_{l_j}^r = [\zeta_{l_j}(\mathbf{s}_1^r), \zeta_{l_j}(\mathbf{s}_2^r), ..., \zeta_{l_j}(\mathbf{s}_{n_r}^r)]$, $\boldsymbol{\Omega} = [\boldsymbol{\delta}_{l_j}^1, \boldsymbol{\delta}_{l_j}^2, ..., \boldsymbol{\delta}_{l_j}^R]$, and $\boldsymbol{\Theta} = [\boldsymbol{\zeta}_{l_j}^1, \boldsymbol{\zeta}_{l_j}^2, ..., \boldsymbol{\zeta}_{l_j}^R]$, and with the known $\tilde{\mathbf{Q}}$ and μ, Equation (9) can be written as

$$\begin{aligned} &\max_{\mathbf{D}} \mu^2 tr[(\tilde{\mathbf{Q}}^T\tilde{\mathbf{X}} - \tilde{\mathbf{Q}}^T\mathbf{D}\boldsymbol{\Omega})(\tilde{\mathbf{Q}}^T\tilde{\mathbf{X}} - \tilde{\mathbf{Q}}^T\mathbf{D}\boldsymbol{\Omega})^T] \\ &-\mu tr[(\tilde{\mathbf{Q}}^T\tilde{\mathbf{X}} - \tilde{\mathbf{Q}}^T\mathbf{D}\boldsymbol{\Theta})(\tilde{\mathbf{Q}}^T\tilde{\mathbf{X}} - \tilde{\mathbf{Q}}^T\mathbf{D}\boldsymbol{\Theta})^T], \end{aligned} \tag{11}$$

For each column of $\tilde{\mathbf{X}}$, i.e., $\tilde{\mathbf{X}}_k$, the optimization of $\mathbf{D}$ can be solved by the following problem:

$$\frac{\partial L(\tilde{\mathbf{X}}_k)}{\partial \mathbf{D}} = 2\tilde{\mathbf{Q}}\tilde{\mathbf{Q}}^T\mathbf{D}(\boldsymbol{\Omega}\boldsymbol{\Omega}^T + \boldsymbol{\Theta}\boldsymbol{\Theta}^T) - 2\tilde{\mathbf{Q}}\tilde{\mathbf{Q}}^T\tilde{\mathbf{X}}_k(\boldsymbol{\Omega}^T + \boldsymbol{\Theta}^T), \tag{12}$$

Then, D can be updated by

$$\mathbf{D} = \mathbf{D} - \frac{\lambda_D}{n}\sum_k \frac{\partial L(\tilde{\mathbf{X}}_k)}{\partial \mathbf{D}}, \tag{13}$$

where λ_D is the step size.

(3) Update step for $\mathbf{S}$. When $\mathbf{D}$, $\tilde{\mathbf{Q}}$, and μ are learned, the sparse code $\mathbf{s}$ for each signal $\mathbf{x}$ in the subspace can be obtained by

$$\min_{\mathbf{s}} \left\|\mathbf{Q}^T\mathbf{x} - \mathbf{D}\mathbf{s}\right\|_F^2 + \lambda \|\mathbf{s}\|, \tag{14}$$

(4) Update step for μ . By recalling that the matrixes, $\tilde{\mathbf{W}}_b$ and $\tilde{\mathbf{W}}_w$ can be built according to the obtained $\tilde{\mathbf{Q}}$ and $\mathbf{S}$. When $\tilde{\mathbf{Q}}$ and $\tilde{\mathbf{Q}}$ are learned, the solution of μ is $\frac{\partial L}{\partial \mu} = 0$, and it can be obtained by a closed-form solution

$$\mu = \frac{tr(\tilde{\mathbf{Q}}^T\tilde{\mathbf{W}}_b\tilde{\mathbf{Q}})}{2tr(\tilde{\mathbf{Q}}^T\tilde{\mathbf{W}}_w\tilde{\mathbf{Q}})}, \tag{15}$$

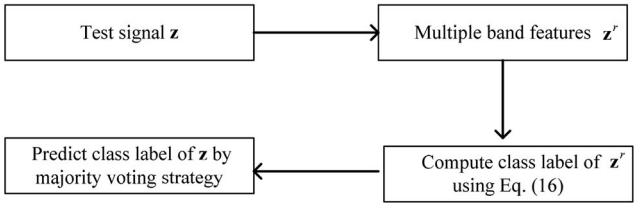


FIGURE 3 | The testing procedure of OPFDDL.

Based on the above analysis, the implementation process of OPFDDL is described in Algorithm 1. We initialize the sub-dictionary for each class by the K-SVD algorithm, and then, we integrate them to form the initialization dictionary **D**.

When the projection matrix **Q** and dictionary **D** are learned, we perform the following procedure to run testing work. The testing procedure of OPFDDL is illustrated in **Figure 3**. For each testing EEG signal **z**, its *r*th frequency band feature is denoted as $\mathbf{z}^r$. We map $\mathbf{z}^r$ into projection subspace using $\mathbf{Q}^r$ and classify its class label according to the smallest reconstruction error on each class as follows:

$$y^r = \arg\min \left\| (\mathbf{Q}^r)^T \mathbf{z}^r - \mathbf{D}_j \tilde{\mathbf{D}}_j \mathbf{z}^r \right\|_2, \tag{16}$$

where $\tilde{\mathbf{D}}_j = (\mathbf{D}_j^T \mathbf{D}_j)^{-1} \mathbf{D}_j^T$ is the pseudo-inverse of dictionary $\mathbf{D}_j$ of class *j*.

Finally, we use the majority voting to identify the class label of signal **z**, i.e.,

$$y = \arg\max_j \Delta_j, \tag{17}$$

where Δ_j is the number of votes for class *j*.

Algorithm 1 The OPFDDL algorithm

Input: Multiple frequency band EEG signals $\mathbf{X}^r (r = 1, 2, ..., R)$ with their class labels.
Output: Projection matrix **Q** and dictionary **D**.
Initialization: Initialize **D** and **S** using the K-SVD algorithm and initialize $\tilde{\mathbf{Q}}$ such that $\tilde{\mathbf{Q}}\tilde{\mathbf{Q}}^T = 1$.
Repeat
Sparse codes update: Compute sparse code **s** for each training sample using Equation (14);
Projection matrix update: Compute $\tilde{\mathbf{Q}}$ using Equation (10);
Dictionary update: Compute **D** using Equations (11–13);
Adaptive weight update: Compute μ using Equation (15);
Until convergence

EXPERIMENT

Experimental Settings

Following the study of Li Y. et al. (2019), we used extracted methods of three features on the SEED dataset, including differential entropy (DE), power spectral density (PSD), and fractal dimension (FD). We investigated the EEG features over all frequency bands per second with no overlap in each channel. We used the random 10 trials in each subject for model training and the rest 5 trials for testing. The classification performance corresponding to each period is recorded for each subject. For DREAMER dataset, to balance the number and length of the segments, we divided the 60-s EEG signals into 59 blocks with an overlap rate of 50%. The DE feature extraction method was carried out and 14-dimensional features for each frequency band were obtained. For each subject, we trained our model using the random 12 trials and the rest 6 trials for testing.

We compared our proposed model with five machine learning methods, including SVM (Cortes and Vapnik, 1995), K-SVD (Aharon et al., 2006), PCB-ICL-TSK (Ni et al., 2020b), DDL (Zhou et al., 2012), and dictionary pair learning (DPL) (Ameri et al., 2016). The Gaussian kernel and Gaussian fuzzy membership were used in SVM and PCB-ICL-TSK, respectively. The parameters in comparison methods were set according to the default settings in corresponding methods. In OPFDDL, the dimension of the projection subspace was set as 90% of the dimension of the EEG signal features. The number of atoms in each class was selected in {10, 15, 20, 25, 30, 35}. The λ parameter in Equation (2) was set as 0.01. We used the 5-fold cross-validation method to select the optimal parameters, and we performed five independent runs to evaluate the classification accuracy of all methods.

Experiment Results on the SEED Dataset

In this subsection, we performed the comparison experiments on the SEED dataset using various combinations of frequency bands and various features. The average accuracy performances of all methods with three feature methods are summarized in **Table 2**. From these results, we have the following observations: (1) Under different frequency band combinations, the results of total frequency bands of all methods are the best. For example, the classification accuracy of OPFDDL using all frequency bands is 6.70, 3.83, and 2.17% higher than that using frequency bands $\beta + \gamma$, $\alpha + \beta + \gamma$, and $\theta + \alpha + \beta + \gamma$. The classification accuracy of SVM using all frequency bands is 3.57, 2.90, and 1.76% higher than that using frequency bands $\beta + \gamma$, $\alpha + \beta + \gamma$, and $\theta + \alpha + \beta + \gamma$. In addition, in most cases, the SDs of all methods are small in all five bands. It demonstrates that multiple bands are helpful for EEG-based ER, due to that the features of each band have discrimination ability and five bands are complementary for distinguishing EEG emotions. (2) The classification performance of the three features is comparable. The performance of the DE feature is slightly better and shows an advantage in most of the cases. The classification accuracy of OPFDDL using the DE feature is 88.87%. It indicates that the DE feature is suitable to deal with EEG emotion signals. (3) OPFDDL outperforms all comparison methods, especially in the case of all five bands. It is because that OPFDDL can effectively integrate band-independent information and inter-band correlation information. The encouraging results indicate that direct concatenation of five frequency bands of EEG data

TABLE 2 | The average accuracies (SDs) of all methods under four combinations of frequency bands and three feature methods.

Features	Methods	$\beta+\gamma$	$\alpha+\beta+\gamma$	$\theta+\alpha+\beta+\gamma$	$\delta+\theta+\alpha+\beta+\gamma$
PSD	SVM	77.41	78.08	79.22	80.98
		(12.65)	(11.21)	(11.01)	(10.38)
	KSVD	77.03	78.00	79.46	81.28
		(12.28)	(11.76)	(11.28)	(10.47)
	PCB-ICL-TSK	77.53	80.10	80.29	81.26
		(11.76)	(11.29)	(10.63)	(10.23)
	DDL	79.79	80.85	81.78	83.16
		(10.49)	(10.56)	(10.01)	(9.45)
	DPL	80.46	81.12	82.28	83.42
		(10.58)	(10.02)	(9.21)	(8.64)
	OPFDDL	**81.00**	**83.27**	**84.93**	**87.10**
		(10.26)	(9.75)	(9.32)	(8.32)
DE	SVM	79.58	80.46	80.89	81.89
		(12.76)	(12.88)	(11.43)	(10.93)
	KSVD	78.32	78.83	80.07	82.06
		(12.77)	(12.05)	(12.11)	(10.85)
	PCB-ICL-TSK	79.66	80.55	81.36	82.49
		(11.63)	(11.09)	(10.35)	(10.04)
	DDL	80.26	81.52	83.75	84.19
		(10.56)	(10.09)	(9.43)	(9.11)
	DPL	81.76	82.12	83.37	84.34
		(10.74)	(10.12)	(9.21)	(8.65)
	OPFDDL	**82.09**	**85.34**	**86.92**	**88.87**
		(10.30)	(9.65)	(9.21)	(8.29)
FD	SVM	78.02	78.94	79.42	81.25
		(12.43)	(12.32)	(11.86)	(11.01)
	KSVD	77.56	78.18	79.79	81.61
		(12.48)	(12.02)	(11.63)	(11.12)
	PCB-ICL-TSK	78.34	80.43	80.98	81.65
		(11.57)	(11.21)	(10.68)	(9.95)
	DDL	80.21	81.23	82.46	83.64
		(10.75)	(10.12)	(9.24)	(9.15)
	DPL	80.74	81.81	82.79	84.09
		(10.35)	(10.07)	(9.31)	(9.17)
	OPFDDL	**81.94**	**83.96**	**85.61**	**87.98**
		(10.14)	(9.64)	(9.32)	(9.01)

The highest average accuracy is shown in bold type.

cannot well-exploit the inherent distinguishing characteristics of data. Considering the common information of multiband and band-specific information shared in each band, it is important to jointly learn multiple band representations. (4) Compared with K-SVD and ODFDL, OPFDDL generates the shared common dictionary on all frequency bands in the projected subspace, which can maintain the data structure of multiple frequency bands. In addition, based on the Fisher discrimination criterion of maximizing within-class compactness and minimizing between-class separation, OPFDDL can well learn the discriminative dictionary from the cooperation of multiple frequency bands.

To further validate the discrimination ability of OPFDDL, in **Figure 4**, we reported the confusion matrix of the OPFDDL model with the DE feature. As shown in **Figure 4**, OPFDDL achieves a better classification performance on positive and neutral emotions than negative emotions in all cases. The performance result of OPFDDL is similar to that in references (Zheng and Lu, 2015; Li Y. et al., 2019). This suggests that subjects may have different EEG signals when experiencing negative

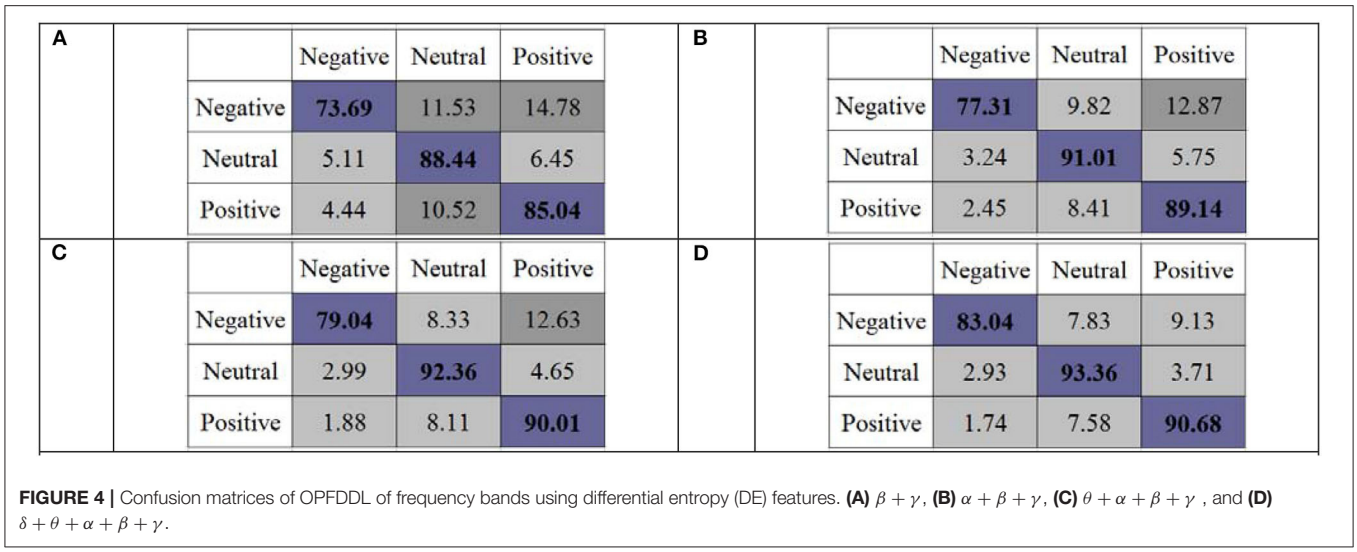


FIGURE 4 | Confusion matrices of OPFDDL of frequency bands using differential entropy (DE) features. **(A)** $\beta + \gamma$, **(B)** $\alpha + \beta + \gamma$, **(C)** $\theta + \alpha + \beta + \gamma$, and **(D)** $\delta + \theta + \alpha + \beta + \gamma$.

TABLE 3 | Average accuracies (SDs) of all methods using DE feature on DREAMER dataset.

Methods	Valence	Arousal	Dominance
SVM	86.35	86.19	86.43
	(5.21)	(5.04)	(4.89)
KSVD	86.77	86.94	87.38
	(5.28)	(5.08)	(4.89)
PCB-ICL-TSK	86.79	87.66	86.90
	(5.26)	(5.02)	(4.87)
DDL	87.54	87.85	87.99
	(5.26)	(5.00)	(4.93)
DPL	87.61	88.33	87.98
	(5.21)	(4.94)	(4.90)
OPFDDL	**89.84**	**90.11**	**89.96**
	(5.19)	(4.90)	(4.74)

The highest average accuracy is shown in bold type.

TABLE 4 | Average accuracies of each subject for OPFDDL model using differential entropy (DE) feature on DREAMER dataset.

	Valence	Arousal	Dominance
Subject 1	93.52	93.13	95.65
Subject 2	90.45	90.08	86.04
Subject 3	86.26	85.53	81.46
Subject 4	96.87	96.32	96.84
Subject 5	93.71	94.56	95.22
Subject 6	81.05	81.04	78.04
Subject 7	84.46	84.02	82.64
Subject 8	85.42	84.97	86.73
Subject 9	81.27	82.43	85.05
Subject 10	88.95	88.04	90.43
Subject 11	82.43	83.10	87.65
Subject 12	92.48	92.15	93.04
Subject 13	89.26	92.07	90.65
Subject 14	95.19	94.66	94.86
Subject 15	98.53	98.75	98.05
Subject 16	87.16	86.78	86.43
Subject 17	87.84	89.01	88.05
Subject 18	89.45	89.02	87.21
Subject 19	94.77	95.43	94.05
Subject 20	91.01	88.03	89.23
Subject 21	93.16	94.20	94.07
Subject 22	94.83	96.04	95.11
Subject 23	88.35	93.21	92.75

emotions and have similar EEG signals when experiencing positive and natural emotions. In addition, it can be seen that OPFDDL achieves the best classification performance (see **Figure 4D**) which uses all five frequency bands.

Experiment Results on the DREAMER Dataset

In this subsection, we performed the comparison experiments on the DREAMER dataset. We performed the comparison experiment using the DE feature. Similar to the SEED dataset, our model is compared with the abovementioned five methods. In the experiment, we verified the performance of OPFDDL according to valence, arousal, and dominance. The accuracy performance of all methods under the frequency band $\theta + \alpha + \beta$ is shown in **Table 3**. From the results, we can see that OPFDDL obtained the highest accuracies of 89.84, 90.11, and 89.96% in terms of arousal, valence, and dominance. Similar to the performance results on the SEED dataset, OPFDDL performed best among all comparison methods. Based on the Fisher discrimination criterion, OPFDDL can well-learn the intrinsic relationships of EEG bands and can obtain the discriminative dictionary from multiple frequency bands cooperation in the projection subspace. In addition, the joint optimization strategy, which

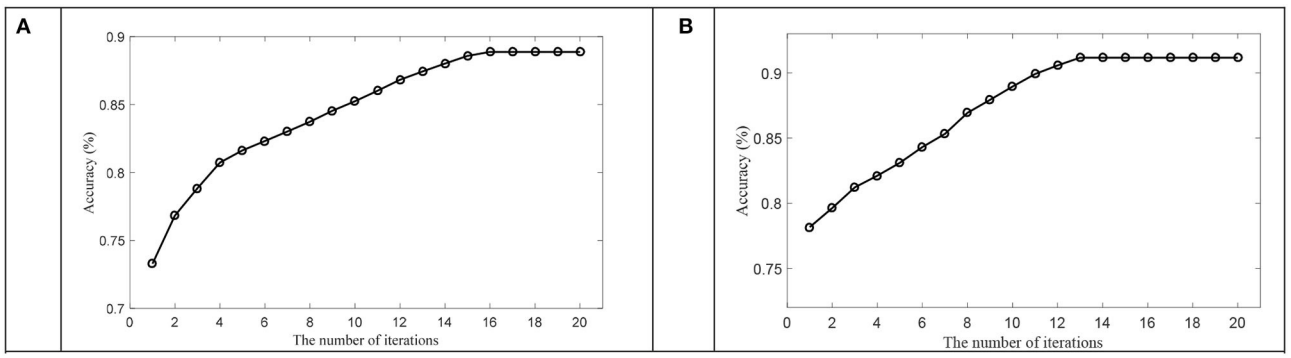


FIGURE 5 | The accuracy with the number of iterations on **(A)** SEED dataset and **(B)** DREAMER dataset.

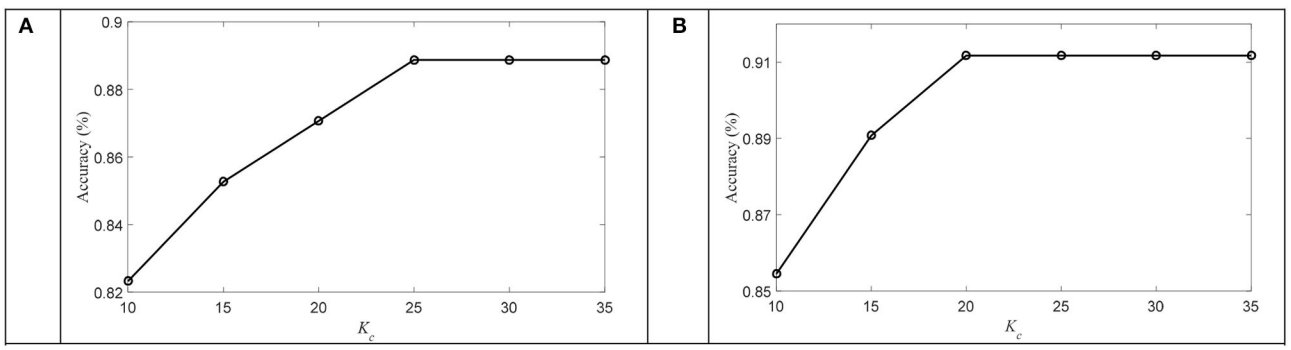


FIGURE 6 | The accuracy with different K_c on **(A)** SEED dataset and **(B)** DREAMER dataset.

addresses the shared projection subspace and dictionary learning, also can incrementally enhance the recognition performance of our proposed model. Thus, our proposed model can utilize more distinctive representations of multiple frequency bands of EEG signals.

Then, we recorded the average accuracies of each subject for the OPFDDL model using the DE feature in terms of arousal, valence, and dominance. The experimental results are shown in **Table 4**. The proposed OPFDDL model had achieved satisfactory recognition performance for all three dimensions of arousal, valence, and dominance. Based on the structure of dictionary learning and the principles of projection and Fisher discrimination criterion, OPFDDL can make better use of discriminative information of different frequency band data and has stronger generalization ability, so it can be effectively used in EEG emotion classification task.

Parameter Variations

In this subsection, we first discussed the convergence of OPFDDL using the DE feature of total frequency bands on SEED and DREAMER datasets. The threshold for iteration stop was set as 10^{-3}. **Figure 5** plots the accuracy that varies with the number of iterations on one subject in two datasets. The results verify the convergence of OPFDDL. It can be seen that the OPFDDL model can achieve convergence within 20 iterations.

Then, we discussed the number of atoms used in OPFDDL. The number of atoms in each class K_c was increased from 10 to 35 in increments of 5. **Figure 6** plots the accuracy that varies with the parameter K_c. The results show that after an initial dramatic increase, the classification accuracy of OPFDDL becomes stable after $K_c = 20$. In addition, the variation trend of accuracy is consistent on two datasets. Thus, the classification performance of OPFDDL is acceptable for small dictionary sizes.

CONCLUSIONS

Most previous machine learning methods focus on extracting feature representations for total frequency bands together without considering specific discriminative information of different frequency bands. In this study, we propose collaborative learning of multiple frequency bands for EEG-based ER. In particular, our model is an integration of projection and dictionary learning based on the Fisher discrimination criterion. For subspace projection optimization, a shared subspace is employed for each frequency band such that the band-specific representations and shared band-invariant information can be simultaneously utilized. For dictionary learning optimization, a shared dictionary is learned from the projected subspace where the Fisher discrimination criterion is used to minimize

within-class sparse reconstruction error and maximize between-class sparse reconstruction error. The joint learning strategy allows the model to extend easily. Consequently, we obtain a discriminative dictionary with a small size. We have performed the experiments and proved the performance of OPFDDL on two real-world EEG emotion datasets, i.e., SEED and DREAMER. For further studies, we will try to utilize and test more discriminative sparse representation criteria in our model. In addition, we only consider subject-dependent classification in EEG emotion identification. Applying this model to subject-independent classification is a challenging work.

DATA AVAILABILITY STATEMENT

The SEED dataset analyzed for this study can be found in this link (http://bcmi.sjtu.edu.cn/~seed/seed.html). The DREAMER dataset analyzed for this study can be found in this link (https://zenodo.org/record/546113).

AUTHOR CONTRIBUTIONS

XG, JZho, and JZhu conceived and developed the theoretical framework of the study. All authors carried out experiment and data process, and drafted the study.

FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61806026 and by the Natural Science Foundation of Jiangsu Province under Grant No. BK20180956.

REFERENCES

Abolghasemi, V., and Ferdowsi, S. (2015). EEG-fMRI: dictionary learning for removal of ballistocardiogram artifact from EEG. *Biomed. Signal Process. Control* 18, 186–194. doi: 10.1016/j.bspc.2015.01.001

Aharon, M., Elad, M., and Bruckstein, A. (2006). K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* 54, 4311–4322. doi: 10.1109/TSP.2006.881199

Ameri, R., Pouyan, A., and Abolghasemi, V. (2016). Projective dictionary pair learning for EEG signal classification in brain computer interface applications. *Neurocomputing* 218, 382–389. doi: 10.1016/j.neucom.2016.08.082

Barthélemy, Q., Gouy-Pailler, C., Isaac, Y., Souloumiac, A., Larue, A., and Mars, J. (2013). Multivariate temporal dictionary learning for EEG. *J. Neurosci. Methods* 215, 19–28. doi: 10.1016/j.jneumeth.2013.02.001

Bhatti, A., Majid, M., Anwar, S., and Khan, B. (2016). Human emotion recognition and analysis in response to audio music using brain signals. *Comput. Hum. Behav*. 65, 267–275. doi: 10.1016/j.chb.2016.08.029

Brouwer, A., Zander, T., van Erp, J., Korteling, J., and Bronkhorst, A. (2015). Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Front. Neurosci*. 9:136. doi: 10.3389/978-2-88919-613-5

Cortes, C., and Vapnik, V. (1995). Support vector machine. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018

Edgar, P., Torres, E. A., Hernández-Lvarez, M., and Sang, G. Y. (2020). EEG-based BCI emotion recognition: a survey. *Sensors* 20:5083. doi: 10.3390/s20185083

Fraiwan, L., Lweesy, K., Khasawneh, N., Wenz, H., and Dickhaus, H. (2012). Automated sleep stage identification system based on time-frequency analysis of a single EEG channel and random forest classifier. *Computer Methods Prog. Biomed.* 108, 10–19. doi: 10.1016/j.cmpb.2011.11.005

Geng, S., Zhou, W., Yuan, Q., Cai, D., and Zeng, Y. (2011). EEG non-linear feature extraction using correlation dimension and hurst exponent. *Neurol. Res.* 33, 908–912. doi: 10.1179/1743132811Y.0000000041

Gong, C., Zhou, P., Han, J., and Xu, D. (2019). Learning rotation-invariant and fisher discriminative convolutional neural networks for object detection. *IEEE Trans. Image Process.* 28, 265–278. doi: 10.1109/TIP.2018.2867198

Gu, X., Zhang, C., and Ni, T. (2020). A hierarchical discriminative sparse representation classifier for EEG signal detection. *IEEE/ACM Trans. Comput. Biol. Bioinformat*. doi: 10.1109/TCBB.2020.3006699. [Epub ahead of print].

Huang, J., Li, Y., Chen, B., Lin, C., and Yao, B. (2020). An intelligent EEG classification methodology based on sparse representation enhanced deep learning networks. *Front. Neurosci.* 14:808. doi: 10.3389/fnins.2020.00808

Hwang, S., Hong, K., Son, G., and Byun, H. (2020). Learning CNN features from DE features for EEG-based emotion recognition. *Pattern Anal. Applicat.* 23, 1323–1335. doi: 10.1007/s10044-019-00860-w

Jiang, Z., Lin, Z., and Davis, L. (2013). Label consistent K-SVD: learning a discriminative dictionary for recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*35, 2651–2664. doi: 10.1109/TPAMI.2013.88

Kanoga, S., Kanemura, A., and Asoh, H. (2019). Multi-scale dictionary learning for ocular artifact reduction from single-channel electroencephalograms. *Neurocomputing* 347, 240–250. doi: 10.1016/j.neucom.2019.02.060

Kashefpoor, M., Rabbani, H., and Barekatain, M. (2019). Supervised dictionary learning of EEG signals for mild cognitive impairment diagnosis. *Biomed. Signal Process. Control* 53:101559. doi: 10.1016/j.bspc.2019.101559

Katsigiannis, S., and Ramzan, N. (2018). Dreamer: a database for emotion recognition through EEG and ECG signals from wireless low cost off-the-shelf devices. *IEEE J. Biomed. Health Informat.* 22, 98–107. doi: 10.1109/JBHI.2017.2688239

Ko, K., Yang, H., and Sim, K. (2009). Emotion recognition using EEG signals with relative power values and bayesian network. *Int. J. Control Automat. Syst*. 7:865. doi: 10.1007/s12555-009-0521-0

Kutepov, I., Dobriyan, V., Zhigalov, M., Stepanov, M., Krysko, A., Yakovleva, T., et al. (2020). EEG analysis in patients with schizophrenia based on Lyapunov exponents. *Informat. Med. Unlock.* 18:100289. doi: 10.1016/j.imu.2020.100289

Li, L., Li, S., and Fu, Y. (2013). "Discriminative dictionary learning with low-rank regularization for face recognition," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition* (Shanghai: IEEE).

Li, X., Fan, H., Wang, H., and Wang, L. (2019). Common spatial patterns combined with phase synchronization information for classification of EEG signals. *Biomed. Signal Process. Control* 52, 248–256. doi: 10.1016/j.bspc.2019.04.034

Li, Y., Zheng, W., Cui, Z., Zong, Y., and Ge, S. (2019). EEG emotion recognition based on graph regularized sparse linear regression. *Neural Process. Lett.* 49, 555–571. doi: 10.1007/s11063-018-9829-1

Nakisa, B., Rastgoo, M., Tjondronegoro, D., and Chandrana, V. (2018). Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors. *Expert Syst. Applicat*. 93, 143–155. doi: 10.1016/j.eswa.2017.09.062

Ni, T., Gu, X., and Jiang, Y. (2020a). Transfer discriminative dictionary learning with label consistency for classification of EEG signals of epilepsy. *J. Ambient Intell. Hum. Comput*. doi: 10.1007/s12652-020-02620-9. [Epub ahead of print].

Ni, T., Gu, X., and Zhang, C. (2020b). An intelligence EEG signal recognition method via noise insensitive TSK fuzzy system based on interclass competitive learning. *Front. Neurosci.* 14:837. doi: 10.3389/fnins.2020.00837

Peng, Y., Liu, S., Wang, X., and Wu, X. (2020). Joint local constraint and fisher discrimination based dictionary learning for image classification. *Neurocomputing* 398, 505–519. doi: 10.1016/j.neucom.2019.05.103

Petrantonakis, P., and Hadjileontiadis, L. (2010). Emotion recognition from EEG using higher order crossings. *IEEE Trans. Informat. Technol. Biomed*. 14, 186–197. doi: 10.1109/TITB.2009.2034649

Song, T., Zheng, W., Song, P., and Cui, Z. (2020). EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Trans. Affect. Comput.* 11, 532–541. doi: 10.1109/TAFFC.2018.2817622

Sreeja, S., and Himanshu, D. (2020). Distance-based weighted sparse representation to classify motor imagery EEG signals for BCI applications. *Multimed. Tools Applicat*. 79, 13775–13793. doi: 10.1007/s11042-019-08602-0

Wei, C., Chen, L., Song, Z., Lou, X., and Li, D. (2020). EEG-based emotion recognition using simple recurrent units network and ensemble learning. *Biomed. Signal Process. Control* 58:101756. doi: 10.1016/j.bspc.2019.101756

Wu, T., Yan, G., Yang, B., and Sun, H. (2008). EEG feature extraction based on wavelet packet decomposition for brain computer interface. *Measurement* 41, 618–625. doi: 10.1016/j.measurement.2007.07.007

Zhang, G., Yang, J., Zheng, Y., Luo, Z., and Zhang, J. (2021). Optimal discriminative feature and dictionary learning for image set classification. *Informat. Sci.* 547, 498–513. doi: 10.1016/j.ins.2020.08.066

Zheng, W., Liu, W., Lu, Y., Lu, B., and Cichocki, A. (2019). Emotion meter: a multimodal framework for recognizing human emotions. *IEEE Trans. Cybernet.* 49, 1110–1122. doi: 10.1109/TCYB.2018.2797176

Zheng, W. L., and Lu, B. L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Autonom. Ment. Dev.* 7, 162–175. doi: 10.1109/TAMD.2015.2431497

Zheng, Z., and Sun, H. (2019). Jointly discriminative projection and dictionary learning for domain adaptive collaborative representation-based classification. *Patt. Recognit.* 90, 325–336. doi: 10.1016/j.patcog.2019.01.004

Zhou, W., Yang, Y., and Yu, Z. (2012). "Discriminative dictionary learning for EEG signal classification in brain-computer interface," in *2012 12th International Conference on Control Automation Robotics & Vision (ICARCV)* (Guangzhou), 1582–1585. doi: 10.1109/ICARCV.2012.6485382

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Gu, Fan, Zhou and Zhu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 09 July 2021
doi: 10.3389/fpsyg.2021.713545

A Novel Paradigm to Design Personalized Derived Images of Art Paintings Using an Intelligent Emotional Analysis Model

Yingjing Duan[1], *Jie Zhang*[1]* *and Xiaoqing Gu*[2]

[1] School of Design, Jiangnan University, Wuxi, China, [2] School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, China

With the development of artificial intelligence (AI), it is imperative to combine design methods with new technologies. From the perspective of the personalized design of derived images of art paintings, this study analyzes the new user demand generated by the current situation and background of personalized design, puts forward a new method of derivative design based on AI emotion analysis, verifies the feasibility of the new method by constructing a personalized design system of derived images of art paintings driven by facial emotion features, and explores the method of combining AI emotion recognition, emotion analysis, and personalized design. This study provides new ideas for the design of art derivatives for the future with massive personalized demand. Thinking and practicing from the perspective of the development of new technology will promote the change of design paradigms in the digital age.

Keywords: artificial intelligence, emotional analysis, emotional features, art derivatives, design paradigm

OPEN ACCESS

Edited by:
Yaoru Sun,
Tongji University, China

Reviewed by:
Lijun Xu,
Nanjing Institute of Technology (NJIT), China
Tingyang Li,
Wuhan Polytechnic University, China

***Correspondence:**
Jie Zhang
Zhangjie@jiangnan.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *23 May 2021*
Accepted: *11 June 2021*
Published: *09 July 2021*

Citation:
Duan Y, Zhang J and Gu X (2021) A Novel Paradigm to Design Personalized Derived Images of Art Paintings Using an Intelligent Emotional Analysis Model. Front. Psychol. 12:713545. doi: 10.3389/fpsyg.2021.713545

INTRODUCTION

In art derivative design, designers use modern design methods to popularize artistic connotations into the modern landscape by thinking, transforming, and designing after fully understanding the connotation of the derived original art. It enables derivatives to meet the spiritual needs of users by carrying a certain artistic connotation. Unlike ordinary products, the willingness of users to buy art derivatives is more out of a desire for the exhibit or works of art in the exhibit. Therefore, it is important for art derivatives to meet the emotional needs of users as much as possible. With the advent of the era of participatory consumption, an increasing number of derivative designs have begun to incorporate personalized factors; only carrying the artistic connotation of art derivatives cannot simply meet the emotional needs of users. The addition of personalized elements can help to realize the prominent and differentiated embodiment of the personal meaning of consumers, accordingly establishing the emotional connection between derivatives and users. Therefore, the combination of derivatives and personalization is an inevitable trend in the design of art derivatives. However, from the current situation of derivative design, the current product form is immutable, the design method is monotonous, and the cost to meet personalized design needs is high. Therefore, future derivative design has to not only meet the demand for the thousands of needs of people in quantity but also realize the meaning connection between individuals and products from the perspective of consumer psychology to be meaningfully accurate to the individual users.

With the rise of GPU parallel computing, especially the explosion of big data, the development of artificial intelligence (AI) technology has become increasingly faster. Among them, the emergence of deep learning is particularly important (Tseng and Ho, 2012; Al-Saffar et al., 2017; Kumar et al., 2019; Mccormack et al., 2020; Somasundaram et al., 2020; Hughes et al., 2021). Initially, deep learning was used in the field of automated speech recognition, which resulted in a significant increase in recognition rates (Krizhevsky et al., 2017). Then, the application of convolutional neural networks (CNNs) has achieved remarkable success in image recognition. Subsequently, generative adversarial nets (GAN) (Goodfellow et al., 2014) that emerged in 2014 was a huge success in areas such as image generation and style migration. In the field of design, deep learning has also been fruitful. For example, an automatic generation system of user portraits was built by Salminen et al. (2020) and An et al. (2018) to isolate behavioral and demographic segments for persona creation *via* aggregated user data. The generation model proposed by Wang et al. (2017) and Bharadhwaj et al. (2018) can be used for the construction of potential user portraits. The creative-stimulating models built by Chen et al. (2019) through semantic associations and the ability to design models for a rapidly built prototype interface introduced by the design software Figma help designers to build quick prototypes. The advancement of deep learning challenges the original design paradigm, and there are feasible ways to meet the needs of "thousands of needs" and "emotional connection."

THE PROCESS OF PERSONALIZED DESIGN OF ARTISTIC IMAGE DERIVATIVES BASED ON EMOTIONAL ANALYSIS

Facing these problems, based on AI technology, this study constructs a new method of generating art derivative images driven by user emotional representation and forms a set of personalized design processes of art image derivatives based on emotion analysis. The process is divided into five main stages: input, extraction, transformation, implantation, and generation. Because emotion is abstract, in extracting features, this study chooses the figurative expression of emotion, i.e., facial expression, as the source of extracting data. Transformation refers to the transformation of facial emotion data into an emotional swatch. Implantation is the replacement of the color on the original art painting with the color of the emotional swatch of the user, and the resulting production of derivative images of the art painting is realized. Among them, extraction, transformation, and implantation are the key steps under this method, as shown in **Figure 1**.

Extraction: Feature Extraction of Facial Expressions

Facial expression extraction can use image recognition technology. Image recognition technology is an important research category of artificial intelligence, and its recognition process is divided into image input, image preprocessing, image feature extraction, recognition result output, and label classification (Anwar et al., 2018). With the introduction of the concept of deep learning, the efficiency of image recognition technology has been greatly improved, and its application has become increasingly extensive. In image recognition technology, the most widely used is CNN (Krizhevsky et al., 2017). The model can convert an image into machine-understood data and process it by an algorithm so that the feature value can be extracted from the image data. Using CNN for facial and emotional recognition has been a major research topic in recent years, and many researchers have made improvements based on classic CNN to propose more accurate facial expression recognition methods for people. For example, Mollahosseini et al. (2016) combined AlexNet with the Google Net model to achieve better facial expression recognition. Lopes et al. (2017) combined some specific feature extraction methods with CNNs to improve the accuracy of expression recognition. Verma et al. (2019) proposed a network with branches of visual and facial identity that achieved a high recognition rate on the CK-plus dataset.

Due to the advantages of image recognition technology, in "extraction," image recognition technology can be used to shoot users in real time through a camera so that the facial emotional characteristics of the user are extracted quickly and accurately. The main purpose of extraction is to convert the facial emotional characteristics of the user into computer data that the computer can understand. The system inputs photos of the user into the CNN model after certain processing and uses the model to obtain results composed of different kinds of facial emotional characteristics data. After normalization processing, the proportion data of various emotional categories are formed, and finally, the system transfers the data to the next process to complete the generation of an emotional palette, as shown in **Figure 2**.

Implantation: The Transformation of Facial Expression Data into an Emotional Palette

The main purpose of the "transformation" session is to form an emotional palette that is accurate to the individual user based on data extracted from facial emotional characteristics to facilitate the replacement of the original color of the art in the next steps. Taking Van Gogh as an example, this study classifies colors as emotional symbols by reference to the research of symbolism and psychology of colors, focusing on Van Gogh (Cela-Conde et al., 2010). In color habits of Van Gogh, yellow or orange colors represent positive and happy; saturated yellows and peaches represent surprise and joy; light colors with lower saturation, such as sky blue, pale green, and goose yellow, represent calm and leisure; red represents anxiety and anger; heavy green or blue green represents despair and sadness; blue represents a conflict between disgust and inner desire; and black and crimson represent depression and fear. According to the above color law, each type of facial emotion will be extracted to correspond to the color system of related emotion expression. Each color system has 10 colors, and finally, a complete type of facial emotion to color mapping will be constructed. After

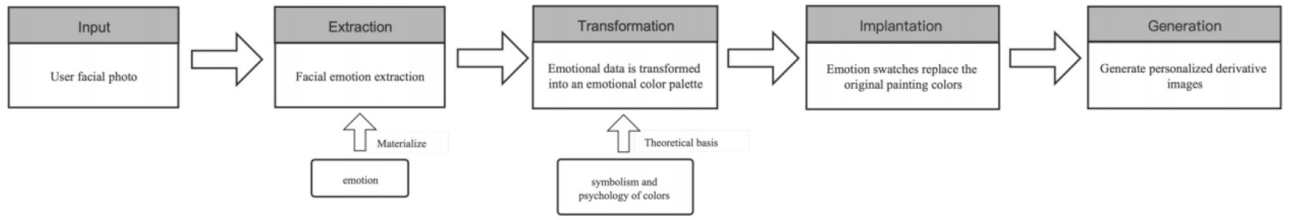


FIGURE 1 | The design process based on emotion analysis.

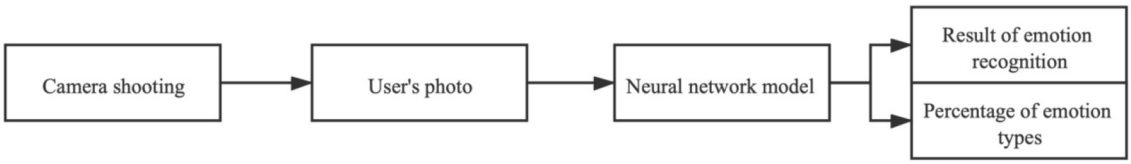


FIGURE 2 | The process of extraction.

FIGURE 3 | Emotional palette.

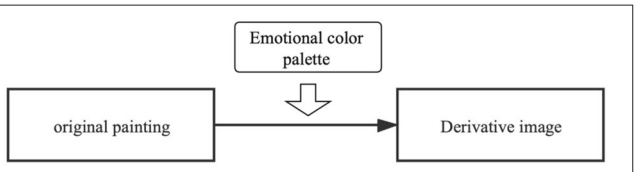


FIGURE 4 | The color replacement process.

constructing the relationship between emotion data and color mapping, the individual emotion data are matched with the specific color by an algorithm, and the emotional palette is formed accurately to the individual to ensure the diversity and richness of the emotional palette. The emotional palette is shown in **Figure 3**.

Implantation: Replacement of the Color of the Original Painting by an Emotional Palette

The main purpose of implantation is to automatically replace the color of the original painting with an emotional palette and to preserve the main features of the original painting, as shown in **Figures 4**, **5**. To date, many scholars have worked out ways to recolor the image, such as Li et al. (2015), who used geodesic line spacing to complete the recoloring of the image while ensuring color coordination throughout the image in 2015. In the same year, Chang et al. (2015) proposed an interactive color replacement system based on palettes that can visualize images, and Cho et al. (2017) trained PaletteNet using a framework of deep neural networks, dramatically reducing the time to recolor in 2017. Replacing the color of the original painting with an

FIGURE 5 | The transformation of original painting to the derivative image.

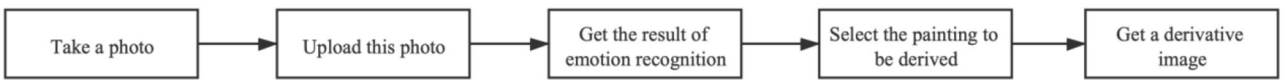


FIGURE 6 | The user operation flow chart.

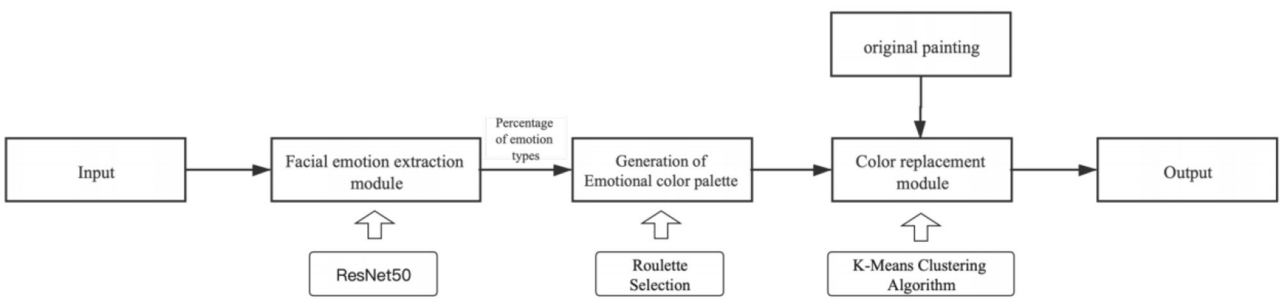


FIGURE 7 | The System flow chart.

emotional palette that is accurate to the individual user and using the "color" in the derived image as a personalized variable manipulated by the emotional characteristics of different users can realize the differentiated design of thousands of people and establish the emotional connection between the derived image and the user.

SIMULATION EXPERIMENT

Based on paintings of Van Gogh, this study combines the traditional method with deep learning to construct a complete personalized derivative image design system, which is driven by emotion analysis for simulation testing, to verify the feasibility of the new method. The system is divided into three main modules, namely, the facial emotional feature extraction module, the emotional palette generation module, and the original color replacement module. The actions of the user are as follows: first, the user needs to make an expression in front of a camera to shoot. Then, from the images provided to the user by the system, the user selects the original painting that will be derived. Finally, the system changes the color of the original image by emotional analysis and image processing and generates the corresponding image derivatives, as shown in **Figure 6**. Correspondingly, the operating process of the system is as follows: first, the system reads the photos taken by the camera and passes them into the facial emotional feature extraction module to finish the transformation of facial emotional characteristics into computer data; then, the extracted data are fed into the emotional palette generation module to create a corresponding emotional palette, and finally, the emotional palette is entered into the color replacement module to replace the original color, as shown in **Figure 7**.

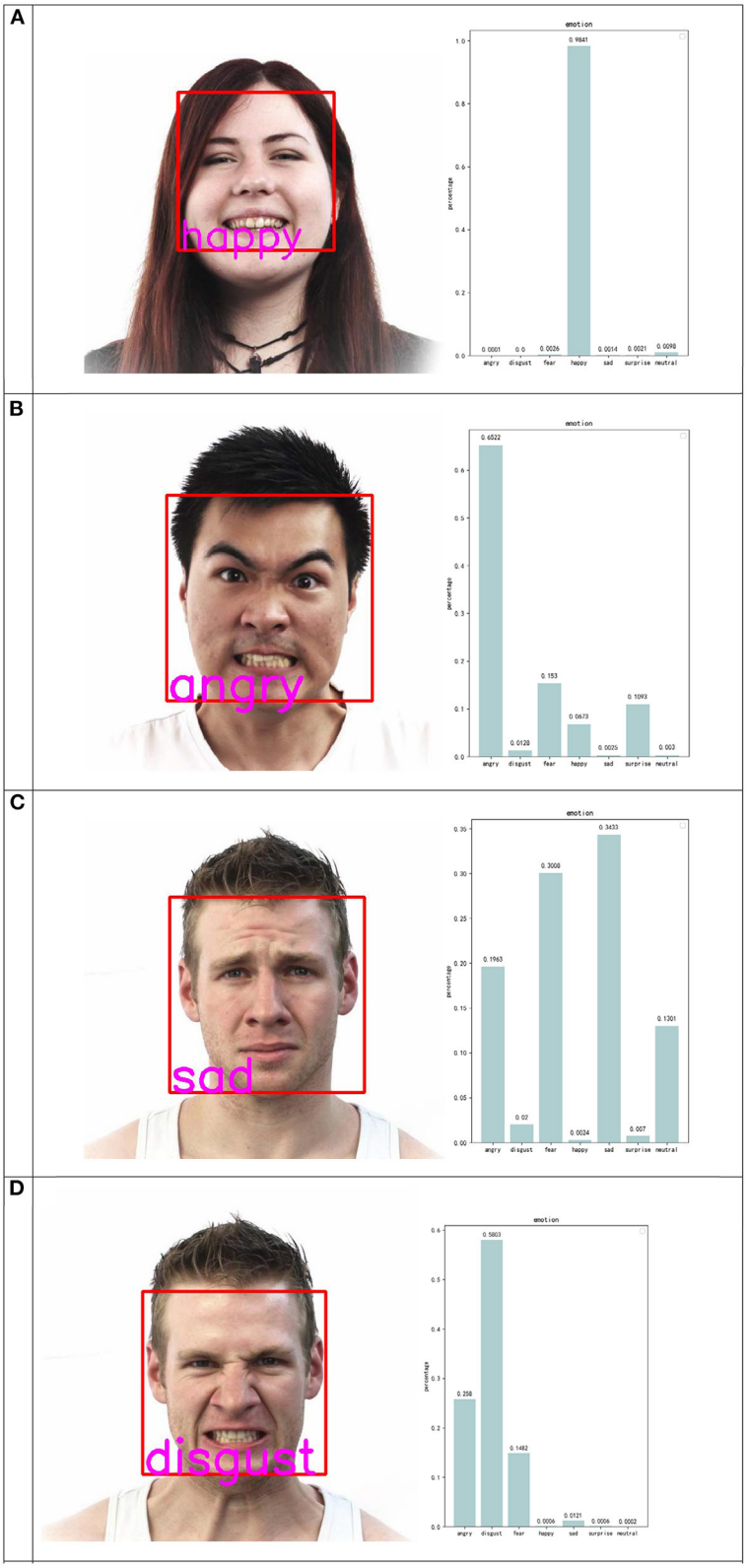


FIGURE 8 | Continued

FIGURE 8 | A visualization chart of facial emotional photos: **(A)** happy, **(B)** angry, **(C)** sad, **(D)** disgust, **(E)** surprise, **(F)** fear, and **(G)** neutral. Facial images are from the public dataset fer2013, which can be downloaded at https://www.kaggle.com/deadskull7/fer2013.

Module 1: The Facial Emotional Feature Extraction Module Based on ResNet50

The main purpose of facial emotion extraction is to transform the emotions of the user into computer-understandable data through recognition and extraction to push the design process and influence the design results. In this study, the CNN model is constructed using Keras in the back end of Tenserflow. In the selection of the data set, both the training set and the test set used Kaggle's fer2013 dataset, a large data set that contains all the facial expressions of people, containing 28,000 training photos and 30,000 test photos, each of which is stored at 48× 48 pixels. To make the extraction of facial emotional characteristic data more accurate, this study uses three neural network models, classical CNN (Krizhevsky et al., 2017), VGG16 (Dubey and Jain, 2020), and ResNet50 (He et al., 2016), and finally chooses ResNet50, whose accuracy is 70.89%, ranking among the top three. In the extraction of facial emotional characteristics, the more critical step is to classify emotions. Therefore, in the process of building neural networks, the more important step is to store facial expressions as values. In this process, Keras generates an output array that contains seven different emoticon scores: happy, angry, sad, disgust, neutral, surprise, and fear. After finishing the training, the system will see the category whose value is largest as the result of emotion recognition. After normalizing these emoticon scores into proportional values in a certain way, the system outputs a visualization chart of facial emotional data and a photo with the recognition result in the form of a bar chart, as shown in **Figure 8**.

Although the module solves the problem of facial emotional feature extraction, there are still some problems. First, in the extraction, the accuracy of the available model is not high enough, especially for the recognition of microexpressions, and is prone to misjudgment. For example, when the corners of the mouth of a person are slightly sagged, the brow is slightly wrinkled, and

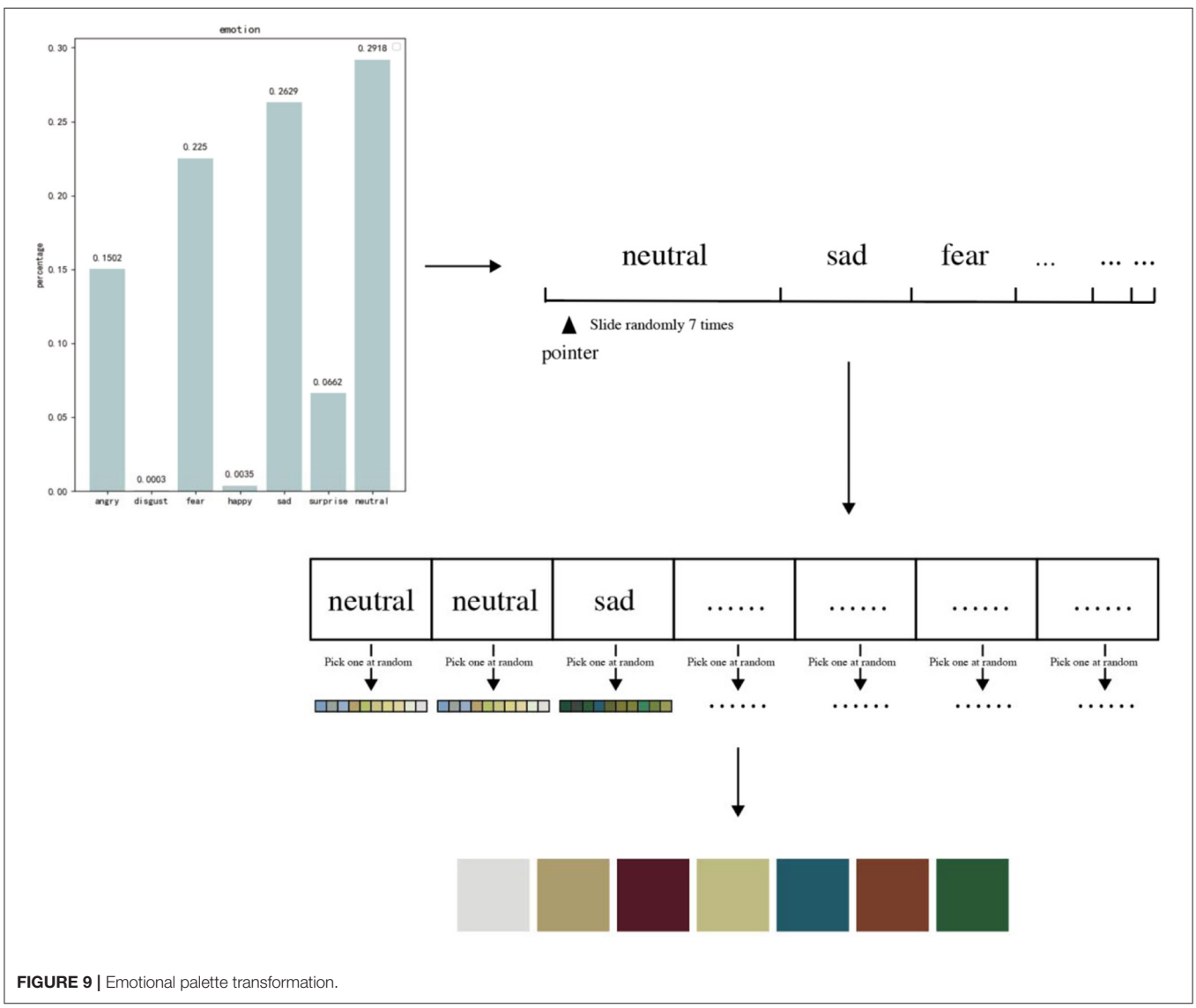


FIGURE 9 | Emotional palette transformation.

the eyes are full of grievances, the result of the system judgment is "calm." Obviously, this is not in line with the judgment of a human, in which the result should be "sad." Therefore, in the follow-up study, we can improve this by expanding the data set or adjusting the neural network model to improve the accuracy of the emotion analysis and realize the precise recognition of the microexpression of the characters. In addition, the existing facial emotion extraction methods can only analyze static pictures and cannot perform emotion recognition in dynamic images. It is precisely because of this limit that users are required to shoot facial expressions in front of a fixed camera. Such restrictive behavior makes the interaction stiff and not natural enough, and timely feedback on emotion from the users cannot be obtained when they are enjoying the exhibition. Therefore, in subsequent studies, how to extract data from a timely camera video and how to use these data in the design process are the key issues to optimize the interaction.

Module 2: The Emotional Palette Generation Module Based on Roulette Wheel Selection

Based on the established facial emotion data and color mapping relationship, this study uses roulette wheel selection (Lipowski and Lipowska, 2012) to complete the matching of emotion data and specific colors to form an individual emotional color palette. Roulette wheel selection is often used in genetic algorithms to increase the diversity of individual generations. The principle is similar to a turntable for a lottery, in which seven regions are divided according to the type of emotion, each corresponding to an emotional color. The proportion of the region is determined by the proportion of emotion types, and a higher proportion of emotion will have a higher probability of being selected. The pointer of the turntable rotates randomly, and where the pointer stops, a color is randomly selected from the 10 colors of the emotional color. To avoid multiple selections of the same color, the selected color will be rejected after the selection. The emotional palette chosen in this study is composed of seven colors. To make the final formation of the palette, seven emotional corresponding color classes are selected. The system drives the pointer to rotate seven times through roulette wheel selection, selecting an emotion class at a time and randomly selecting a specific color with equal probability in this class. The system follows the above process seven times to generate an emotional palette, which can make the color class corresponding to the larger emotion species have more opportunities to be selected, ensure the consistency of the resulting emotional palette tone, and clearly reflect the facial emotional characteristics in the photo of the user.

In the process, the roulette wheel selection integrates the probability graph, which is originally fan-distributed, into a straight line, divides the interval (0, 1) into seven segments according to their respective emotional classes, and then generates a random floating point located within the interval (0, 1). The system selects the emotion type through pointer selection, randomly selects a color from the color class corresponding to the emotion type to add to the emotional palette, and then deletes the color from the color library to avoid repeatedly selecting the same color. After repeating the above process seven times, the system chooses seven colors to form the final emotional palette. This method used in this study can not only ensure that users generate completely different emotional palettes for each shot but also avoid the situation in which users with similar emotions have emotional palettes. This method can increase the randomness of palettes. In addition, in the testing process, we found that the facial emotion data occasionally have extreme situations. For example, when a person is laughing, the emotion extraction system will give "happy" a very high value. When transferred into proportional data, it occupies a large area, and the proportion of other emotions is compressed into a small amount. Since the proportion of emotion is directly related to the emotional color palette, to avoid the problem that the final generated color palette is all from the same color class and the color tone of the new image is too single after color replacement, this study adds the "if" condition to the roulette wheel selection. When the probability of a single type of color class selection does not exceed 50%, the tone of the new image is relatively balanced. Therefore, the conditional discrimination statement detects the probability of each type of color class, and when the probability of a color class is more than 50%, the conditional discrimination statement limits it to 50%, with the remaining 50% divided equally by other classes according to their proportions, as shown in **Figure 9**.

Here, we present the code of our adopted model in **Algorithm 1**.

Algorithm 1 |The adopted model in this study

```
begin
    Number the emotions in turn E∼[1, 7],
    Divide the interval (0, 1) into 7 parts, in order of
    probability, and record it as interval 1-7,
    N←1
    While(N>7) do
        P=random(0,1) //P is the number of random floats
        generated within the interval(0, 1).
        if  P is within the interval i      // i∼[1, 7]
            then E←i
        end if
        Randomly select a color from Mood E to add to
        the palette
        N←N+1
    end
end
```

However, as far as the expected effect is concerned, the method of generating emotional palettes proposed in this study meets the need to match emotion and color, but there are still some shortcomings. The main problem is that, although the method fully satisfies the randomness of palette generation, it neglects the harmony of color matching. In these randomly generated palette, because of the lack of boundary between color and color, it will lead to a great difference between the result of

color replacement and the expected effect. The appearance of color can not meet the aesthetic standard, which will affect the user's aesthetic experience. Therefore, in follow-up research, it is necessary to establish a mathematical model according to the requirements of color harmony and aesthetic standards, and the system needs to limit the original algorithm or find a new and more suitable algorithm.

Module 3: Color Replacement Module Based on the K-means Clustering Algorithm

The purpose of the color replacement module is to replace the original palette with the emotional palette to complete the color transformation of the original image. As the modules are built, this study refers to the project of the team of Kanyaraasi, Palette-based Photo Recolor (Chang et al., 2015), to build the color replacement system. The image recoloring system proposed by the Kanyaraasi team has a complete image recoloring interface that allows users to replace the color of palettes one by one by mouse interaction, depending on their personal preferences. Based on this system, this study proposes an image recoloring system driven by an emotional palette corresponding to facial emotional characteristics, which realizes the full automation of the process. The color transformation system is based on the K-means clustering algorithm, which extracts the palette of the original painting and realizes the color transformation processing of the image. To improve the clustering efficiency of the K-means algorithm, first, the three channels of the image RGB are normalized, and then, based on the statistical results of the image histogram, the channels are divided into 16 sets of histogram columns. Each histogram column contains several pixels, and then $16 \times 16 \times 16$ statistical histogram columns can be obtained. The Lab spatial color mean of the pixels in each statistical hist bar is calculated as the cluster color center of each statistical hist square bar, and then, the $16 \times 16 \times 16$ cluster color centers are clustered using the K-means clustering algorithm. The cluster center is divided into seven categories. The color of each cluster center is a kind

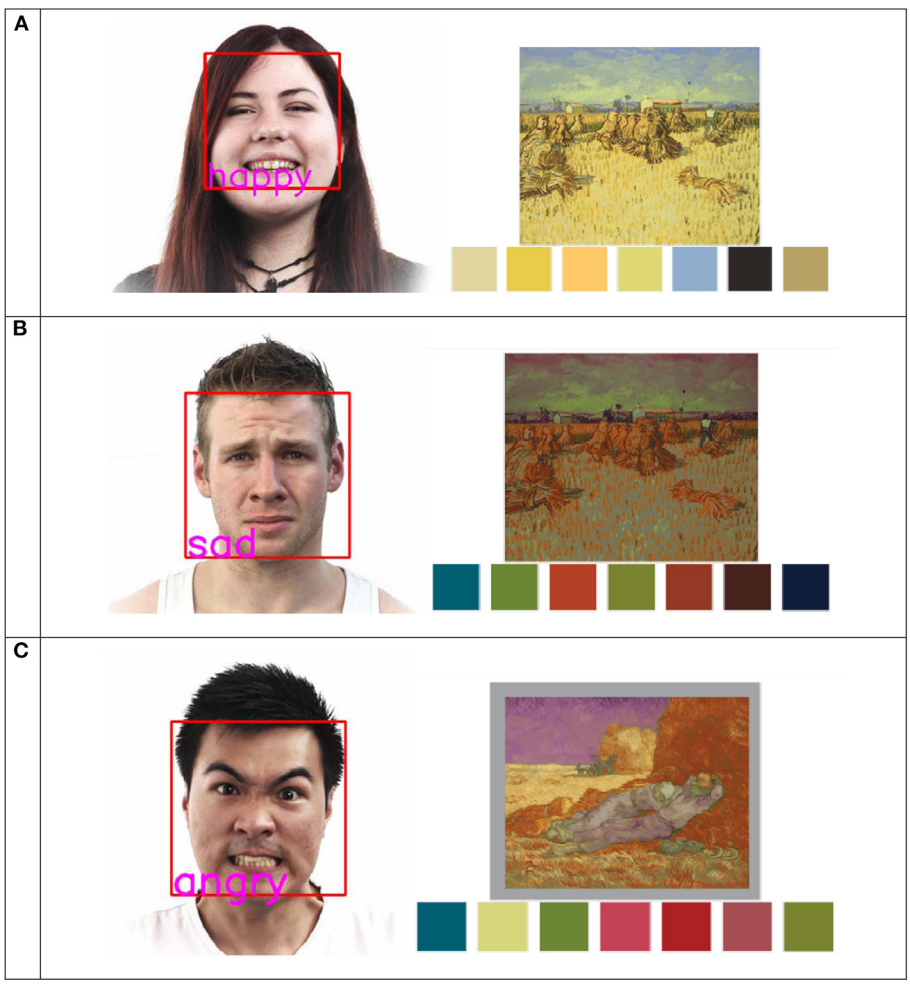


FIGURE 10 | Continued

FIGURE 10 | A derivative picture of **(A)** happy, **(B)** sad, **(C)** angry, **(D)** surprise, **(E)** disgust, **(F)** fear, and **(G)** neutral. Facial images are from the public dataset fer2013, which can be downloaded at https://www.kaggle.com/deadskull7/fer2013.

of palette, and then, the original image of seven palettes is extracted. Then, with the emotional swatch proposed above, according to a specific method of one-to-one corresponding processing, the original palette is replaced to achieve the color conversion of the image. Eventually, the system outputs the image with the emotional characteristics of the user as a derived image to the user, and the result of color substitution is shown in **Figure 10**.

Although the existing color replacement module has completed the color replacement of the original image, there are still some problems. The main problem with the replacement is that the derived image cannot retain the light and dark layers of the original color distribution. When the original art is transformed into a derivative image, the problem will cause the color layer of the image to be greatly weakened or become insufficiently harmonious. Moreover, it does not conform to the

context of the picture, and the meaning expressed by the picture is the opposite. Therefore, in the follow-up research process, researchers need to start from the painting semantics and color light and dark structures in the process of color replacement to limit the brightness of the picture so that the derivative image more closely fits the original rather than only from the picture color ratio of the area to make a simple color replacement.

DESIGN OF THE PERSONALIZED SYSTEM BASED ON EMOTION ANALYSIS

Through the above research, the new design method is implementable in combination with new technology, and the intervention of new technology can effectively solve the existing problems in personalized design. Even so, there are two issues in the existing methodology that can be improved. First, in the existing personalized design process, the solution proposed in this study focuses on the user of the characteristic extraction stage. In other stages, using different technologies to achieve more personalized user needs of the design is still worth exploring; it is also a direction worth studying later.

In addition, in the construction of the personalized design simulation system in this study, the main technology is machine learning, and other computer technologies are secondary technologies. This phenomenon verifies that the development of new technology, especially the wide application of machine learning, can promote the change of design methods and provide the possibility for the emergence of new design paradigms. As far as the entry point of this study, i.e., personalized design, there are two possible changes in the design method: one is the function changes of the designer in the design process, and the other is that the connection between user characteristics and personalized elements is becoming increasingly diverse and complex.

Functional Changes in the Design Process of a Designer

Traditional personalized products are roughly divided into four links from design to production: user personalized demand analysis, personalized element extraction, personalized design, and personalized product production. In this design process, the designer occupies a dominant position, and user data are only involved in the analysis of the individual needs of the user in the early stage. In the subsequent design links, the designer controls the subsequent process as the lead, as shown in **Figure 11**. To better serve users and meet their emotional needs, future intelligent personalized design methods need to allow more user data to participate in the entire design process, and the entire design process should be more dominated by users or user data. The sense of participation of the user should be enhanced to achieve true participatory consumption. The role of the designer in this process should be a designer of the personalized design system. Its main task is to combine new technology to complete the design of the personalized system and the interactive process of user participation in the personalized design, as shown in **Figure 12**. After finishing these tasks, the designer transfers the leading power to the user, allowing the personalized system and the user to cooperate with each other to complete the design process together and allowing the user to complete the customization of personalized products according to their personal preferences and realize the real sense of "thousands of needs" design.

The Connection Between User Characteristics and Personalized Elements

In traditional personalized design, the source of user characteristics mostly comes from user research; designers need to extract user characteristics and preferences from the results and, then, through certain means, transform the characteristics and preferences of the user into personalized elements and integrate them into personalized products, which is also the key step to establish personalized product user emotional connections. However, due to the limited amount of data and the limited ability of designers to process data, it is difficult to meet the needs of all when extracting user characteristics and building user emotional connections. However, with the development of big data and artificial intelligence and with the help of intelligent technology, an increasing number of user features can be mined and extracted. In addition to the advantages of "many quantities," intelligent technology with the advantages of user characteristic extraction also has the advantages of high speed and high precision. With such technical support, the emotional feature extraction of the user can be multidimensional, fast, and precise so that the extraction of user characteristics can be more comprehensive and more diverse to find more dimensions that can serve the personalized design. As far as the project studied in this study is concerned, in the future design, more kinds of emotional characteristics of users can be extracted, and the elements used to extract the emotional characteristics of the user are the facial expressions of the user. Facial expression is the materialized embodiment of the emotion and expression of the user of the direct emotion of the user. In addition, it can also be extracted by extracting the voice, behavior, brain waves, and other elements of the user that directly reflect emotional characteristics to form an emotional connection with personalized products or by extracting color matching, wearing style, and other elements that indirectly reflect emotional elements to establish the meaning of the user and personalized products. As the number of emotional feature elements that can be extracted increases, so does the number of personalized elements in the corresponding product, while each element representing the emotional characteristics of the user can be associated with different personalized elements, such as shape, style, and texture, in addition to color. Therefore, compared with the relationship between the emotional characteristic elements and the personalized elements of the user in the past, the relationship between the two is more complex and diverse in intelligent personalized design in the future, as shown in **Figure 13**.

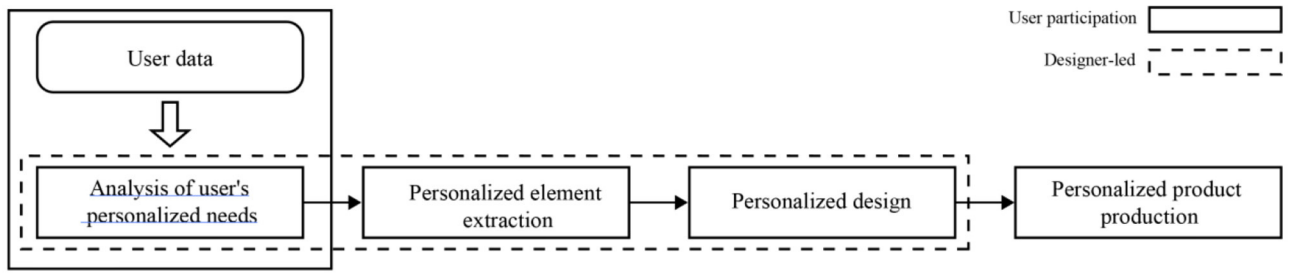


FIGURE 11 | Traditional personalized design approach and role engagement.

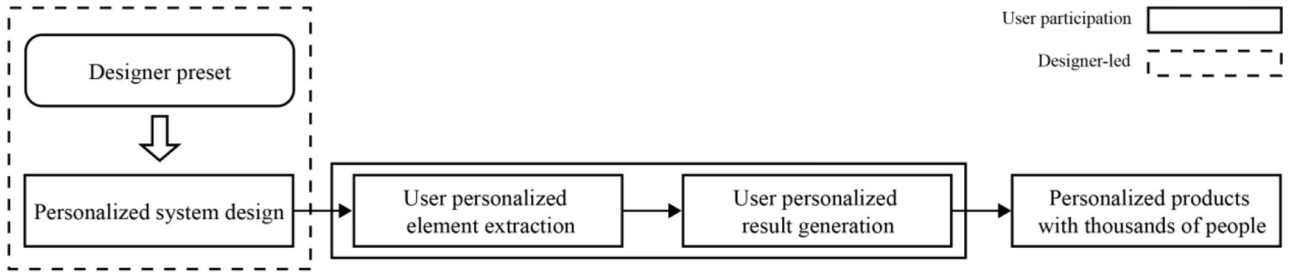


FIGURE 12 | Combining new technology to complete the design of the personalized system.

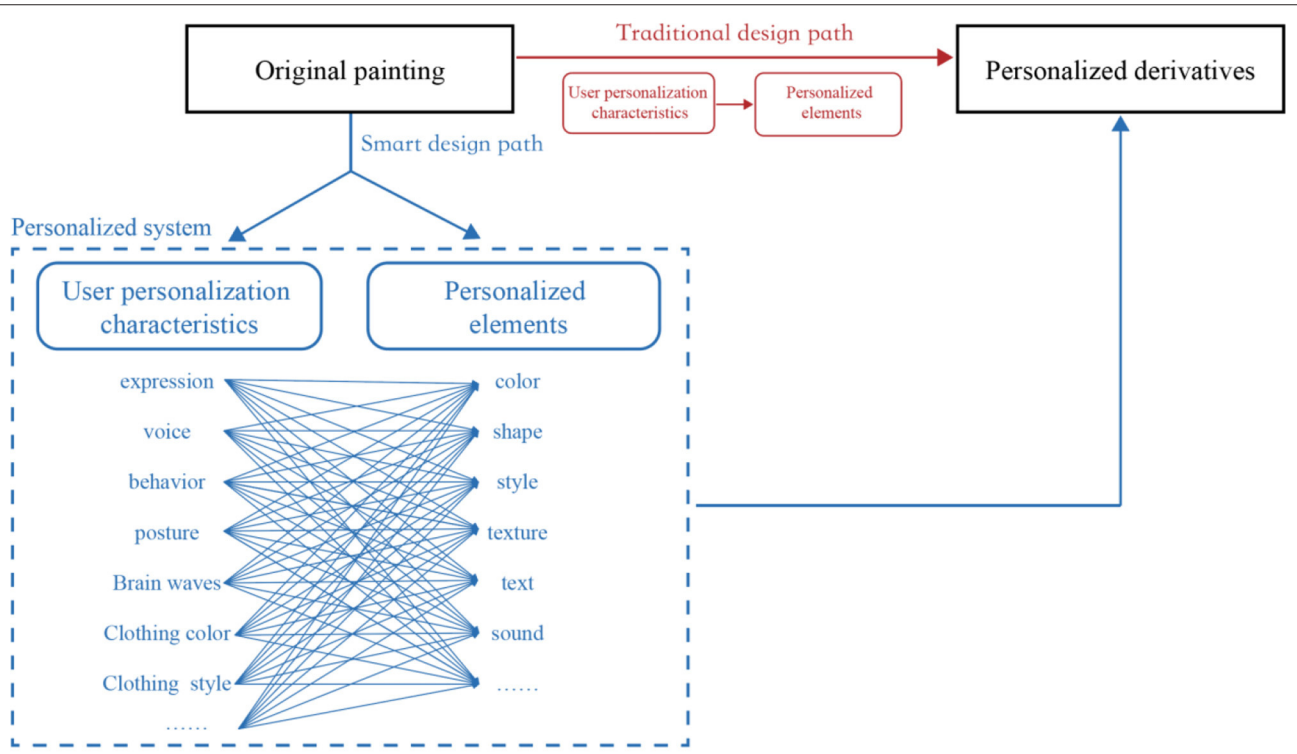


FIGURE 13 | The relationship between user emotional characteristics and individual elements.

CONCLUSION

This study proposes a new method of art image derivative design based on emotion analysis. First, the significance of personalization is analyzed in the design of art derivatives. Then, the problems existing in the personalized design method are pointed out, the new needs of users in the background of intelligent technology are analyzed, new design ideas combined with new technology are put forward, the main technical principles of the technology are analyzed, and practical solutions are proposed to explore the new design paradigm. Finally, by constructing a personalized design system of image derivatives driven by user facial emotion data based on paintings of Van Gogh, this study verified the feasibility of the solution. It was found that innovation of the design paradigm is imperative. It is significant to change the original design paradigm and introduce new ones, which can make the design better combined with current new technology and better serve the needs of the public.

There are also some problems that can be improved. The facial emotion feature extraction module mainly uses ResNet50 to extract the facial emotion data of the user. Its shortcomings are that the accuracy of this module is not high enough and extracting emotion data from video and processing them cannot be achieved; the emotion color palette generation module uses roulette wheel selection to realize the generation of emotional palettes, but its disadvantage is that the color matching of the generated palettes cannot fully meet the aesthetic standards. The original color replacement module uses the K-means clustering algorithm to achieve the color replacement of the original image based on the emotional palette. The disadvantage is that it cannot completely retain the beauty of the light and dark levels of the original color.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.kaggle.com/deadskull7/fer2013.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

YD, JZ, and XG developed the theoretical framework and the model in this study and drafted the manuscript. YD implemented the algorithm and performed experiments and result analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the Jiangsu University Philosophy and Social Science Research Fund under Grant 2018SJA0808.

REFERENCES

Al-Saffar, A. A. M., Tao, H., and Talab, M. A. (2017). "Review of deep convolution neural network in image classification," in *2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET),* 26–31.

An, J., Kwak, H., Jung, S., Salminen, J. O., and Jansen, B. (2018). Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data. *Soc. Network Anal. Mining* 8, 1–19. doi: 10.1007/s13278-018-0531-0

Anwar, S. M., Majid, M., Qayyum, A., Awais, M., Alnowami, M., and Khan, M. K. (2018). Medical image analysis using convolutional neural networks: a review. *J. Med. Syst.* 42, 1–13. doi: 10.1007/s10916-018-1088-1

Bharadhwaj, H., Park, H., and Lim, B. Y. (2018). "RecGAN: recurrent generative adversarial networks for recommendation systems," in *Proceedings of the 12th ACM Conference on Recommender Systems*, 372–376.

Cela-Conde, C. J., Marty, G., Munar, E., Nadal, M., and Burges, L. (2010). The "style scheme" grounds perception of paintings. *Percept. Motor Skills* 60, 19–34. doi: 10.2466/pms.2002.95.1.91

Chang, H., Fried, O., Liu, Y., Diverdi, S., and Finkelstein, A. (2015). Palette-based photo recoloring. *ACM Trans. Graph.* 34, 1–11. doi: 10.1145/2766978

Chen, L., Wang, P., Dong, H., Shi, F., Han, J., Guo, Y., et al. (2019). An artificial intelligence based data-driven approach for design ideation. *J. Visual Commun. Image Represent.* 61, 10–22. doi: 10.1016/j.jvcir.2019.02.009

Cho, J., Yun, S., Lee, K., and Choi, J. Y. (2017). "PaletteNet: image recolorization with given color palette," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1058–1066.

Dubey, A. K., and Jain, V. (2020). Automatic facial recognition using VGG16 based transfer learning model. *J. Inform. Optim. Sci.* 41, 1589–1596. doi: 10.1080/02522667.2020.1809126

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial networks. *Adv. Neural Inform. Proces. Syst.* 2014, 2672–2680. doi: 10.1145/3422622

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Hughes, R. T., Zhu, L., and Bednarz, T. (2021). Generative adversarial networks–enabled human-artificial intelligence collaborative applications for creative and design industries: a systematic review of Current approaches and trends. *Front. Artif. Intel.* 4:604234. doi: 10.3389/frai.2021.604234

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90. doi: 10.1145/3065386

Kumar, V., Rajan, B., Venkatesan, R., and Lecinski, J. (2019). Understanding the role of artificial intelligence in personalized engagement marketing. *Calif. Manage. Rev.* 61, 135–155. doi: 10.1177/0008125619859317

Li, X., Zhao, H., Nie, G., and Huang, H. (2015). Image recoloring using geodesic distance based color harmonization. *Comput. Vis. Media* 1, 143–155. doi: 10.1007/s41095-015-0013-5

Lipowski, A., and Lipowska, D. (2012). Roulette-wheel selection via stochastic acceptance. *Phys. A Stat. Mech. Appl.* 391, 2193–2196. doi: 10.1016/j.physa.2011.12.004

Lopes, A. T., Aguiar, E. D., Souza, A. F. D., and Oliveira-Santos, T. (2017). Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. *Pattern Recog.* 61, 610–628. doi: 10.1016/j.patcog.2016.07.026

Mccormack, J., Hutchings, P., Gifford, T., Yee-King, M., Llano, M. T., and D'Inverno, M. (2020). Design considerations for real-time collaboration with creative artificial intelligence. *Org. Sound* 25, 41–52.

Mollahosseini, A., Chan, D., and Mahoor, M. H. (2016). "Going deeper in facial expression recognition using deep neural networks," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1–10.

Salminen, J., Guan, K., Nielsen, L., and Jung, S. (2020). "A template for data-driven personas: analyzing 31 quantitatively oriented persona profiles," in *International Conference on Human-Computer Interaction* (Cham: Springer), 125–144.

Somasundaram, M., Junaid, K. A. M., and Mangadu, S. (2020). Artificial intelligence (AI) enabled intelligent quality management system (IQMS) for personalized learning path. *Proc. Comp. Sci.* 172, 438–442. doi: 10.1016/j.procs.2020.05.096

Tseng, Y. S., and Ho, M. C. (2012). "Creating sustainable emotional value through personalized design," in *Design for Innovative Value Towards a Sustainable Society,* eds M. Matsumoto, Y. Umeda, K. Masui, and S. Fukushige (Dordrecht: Springer), 257–260.

Verma, M., Kobori, H., Nakashima, Y., Takemura, N., and Nagahara, H. (2019). "Facial expression recognition with skip-connection to leverage low-level features," in *2019 IEEE International Conference on Image Processing (ICIP)*, 51–55.

Wang, J., Yu, L., Zhang, W., Gong, Y., and Zhang, D. (2017). "IRGAN: a minimax game for unifying generative and discriminative information retrieval models," in *SIGIR 2017-40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 515–524.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Duan, Zhang and Gu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 28 July 2021
doi: 10.3389/fpsyg.2021.723325

Research on Emotion Analysis of Chinese Literati Painting Images Based on Deep Learning

Jie Zhang[1,2]*, *Yingjing Duan*[1] and *Xiaoqing Gu*[3]

[1] School of Design, Jiangnan University, Wuxi, China, [2] Graduate School of Management, Management and Science University, Shah Alam, Malaysia, [3] School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, China

Starting from a pure-image perspective, using machine learning in emotion analysis methods to study artwork is a new cross-cutting approach in the field of literati painting and is an effective supplement to research conducted from the perspectives of aesthetics, philosophy, and history. This study constructed a literati painting emotion dataset. Five classic deep learning models were used to test the dataset and select the most suitable model, which was then improved upon for literati painting emotion analysis based on accuracy and model characteristics. The final training accuracy rate of the improved model was 54.17%. This process visualizes the salient feature areas of the picture in machine vision, analyzes the visualization results, and summarizes the connection law between the picture content of the Chinese literati painting and the emotion expressed by the painter. This study validates the possibility of combining deep learning with Chinese cultural research, provides new ideas for the combination of new technology and traditional Chinese literati painting research, and provides a better understanding of the Chinese cultural spirit and advanced factors.

Keywords: emotional analysis, Chinese literati painting, deep learning, computer vision, machine learning

OPEN ACCESS

Edited by:
Yaoru Sun,
Tongji University, China

Reviewed by:
Tao Tao,
Anhui University of Technology, China
Qin Qin,
Henan Institute of Engineering, China

***Correspondence:**
Jie Zhang
zhangjie@jiangnan.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *10 June 2021*
Accepted: *29 June 2021*
Published: *28 July 2021*

Citation:
Zhang J, Duan Y and Gu X (2021) Research on Emotion Analysis of Chinese Literati Painting Images Based on Deep Learning. Front. Psychol. 12:723325. doi: 10.3389/fpsyg.2021.723325

INTRODUCTION

In ancient China, most literate people appreciated painting and calligraphy, and many well-educated people, such as Dongpo Su, were talented at painting. Because he was good at painting, poems of Dongpo Su included images depicting the fusion of poetry and painting, which can be understood as textual expression and visual performance to a high consistency. In contrast, a painter with no cultural accomplishment and whose paintings lack thought and emotion can only be called an artisan. After the Song Dynasty, almost all great Chinese painters were literati painters. The appreciation and research of literati paintings usually require researchers to have rich historical and cultural knowledge, associating the era of the author with personal behavior, emotion, knowledge, self-cultivation, feelings, and other factors (Edward, 1997; Wong, 2000; Vigneron, 2011; Bowie, 2014; Adriana, 2019). However, today, the image information of literati paintings has become the main information transmission channel for cohesive emotion. A traditional Chinese painting has complex forms and various types. In the present research, although there is no shortage of either samplings from many kinds of traditional Chinese painting forms or samplings combined with computer technology with a variety of picture characteristics for the analysis, there is still a lack of machine vision-based Chinese literati painting emotional analysis. Therefore, this study attempts to analyze the emotions of literati paintings by using artificial intelligence. We also constructed a dataset of Chinese literati paintings, visualized the characteristic

area corresponding to the emotional keywords, and explored the relationship between the picture content and the expression of emotion based on machine vision to help people better understand the semantic emotion in literati painting.

Emotional Characteristics of Chinese Literati Painting

Literati painting is the emotional product of specific groups of people in ancient Chinese society. Literati paintings are all made by literati or painters with literati nature. The special identity of the painters makes literati paintings have the nature of literati and contain the tastes of literati. However, this study does not aim at depicting the appearance of the object but pays more attention to the cultivation of extra painting work and the expression of emotions, generally through the reflection of the literatito form an indoctrination of the external society or through the expression of elegant tastes to achieve the purpose of self-entertainment. Outside the picture, the ancient literati pursued "*Lide*," "*Ligong*," and "*Li Yan*." Most of the painters used painting into the Tao to obtain spiritual relief and transcendence. Painters with personal quality and sentiment strived to learn how to create a "simple" and "ethereal" mood as well as other moods to express emotions, express personality, expound philosophy, and still include the unique aesthetic taste of emotional development. Within the picture, the literati used a unique screen perspective: white background, pen and ink changes, metaphors, and other ways to express emotions with a scenery. They also used the rationality of poetry to expand the expressiveness of painting.

Emotion Analysis of Paintings in the Context of Machine Learning

With the explosion of big data and the development of artificial intelligence, the use of advanced artificial intelligence algorithms to analyze painting works has become a major research hot spot at the intersection of art and computer vision. Current research results show that, using image analysis and feature extraction technology, computers can learn much knowledge from paintings, extract effective features of painting images, and realize painting value evaluation, painting protection, painting author identification, painting style comparison, and painting semantic emotion analysis (Prabowo and Thelwall, 2009; Xia et al., 2011; Feldman, 2013; Medhat et al., 2014; Xu et al., 2019; Ortis et al., 2020; Xue et al., 2020; Zhang et al., 2020, 2021; Castellano and Vessio, 2021; Guo et al., 2021). The analysis of traditional Chinese paintings based on computer vision technology is still in the development stage. Liu and Jiang (2014) proposed a convolutional neural network (CNN)-based feature description and a feature-weighted feature-prioritized algorithm to achieve overwhelmingly better classification performances of Chinese painting. Sun et al. (2016) analyzed the length, curvature, and density of traditional Chinese painting strokes and proposed new methods for extracting related features. Fan and Zhang (2020) reported the results of evaluating the influence of extracted features on visual order in Chinese ink paintings using a regression model.

The process of dataset production is divided into four steps: obtaining data, filtering data, processing data, and classifying data based on sentiment analysis. Each step can be subdivided into several steps, as shown in **Figure 1**.

Obtain Data

Literati paintings in the Song Dynasty were greatly developed, and a large number of famous literati painters emerged from this time period, such as Yuan Dong, Ran Ju, Cheng Li, Kuan Fan, Tong Wen, Shi Su, and Fu Mi. The literati also received much development in the Yuan, Ming, and Qing dynasties. This study selected 19 representative literati painters: Yuan Dong, Kuan Fan, Shi Su, Fu Mi, and Youren Mi from Song dynasty; Gongwang Huang, Zan Ni, Meng Wang, Zhen Wu, Mengfu Zhao, and Xuan Qian from Yuan dynasty; Shen Zhou, Zhengming Wen, Wei Xu, Chun Chen, Fu Wang, and Qichang Dong from Ming dynasty; and Tao Shi, and Bada Shanren from Qing dynasty. The image data of the representative works of the above painters were mainly obtained from network engine searching and picture album picture scanning, and the names of these painters were used as keywords for retrieving data. The picture data retrieved from the network were then downloaded and stored, and the picture album of the artist was scanned as a supplement to complete the dataset. Finally, the number of literati painting images was 635.

Filter and Process Data

After obtaining the literati painting data, this study next filtered and processed the data. In the acquired data, there was a large number of literati paintings with inscriptions or stamps on the white space of the paintings. The existence of these redundant artificial traces affected the original white space layout of the picture. Therefore, in the process of data processing, this study used image processing tools to remove redundant traces to restore the original appearance of the paintings, as shown in **Figure 2**. Additionally, to improve the accuracy of image feature extraction by the neural network, based on preserving the original blank features of the painting, this study used image processing tools to split the picture with too large aspect ratios to ensure the consistency of the pictures in the dataset on the frame. This avoided the problem of the aspect ratio of some pictures being too exaggerated affecting the overall training effect of the neural network. After filtering and processing the data, this study finally obtained 795 images to build the literati painting dataset.

Data Classification Based on Emotion Analysis

After filtering and processing the data, this study classified the data based on the emotional characteristics expressed in the pictures. As a typical representative of a literati painter, Shi Su emphasized that art should express his subjective emotions. He regarded painting as an important part of improving self-cultivation and lifestyle. He pursued the expressive interest of strokes instead of similarity. In terms of themes, in addition to the theme of landscape, Shi Su extended to objects such as flowers and birds, figures, plums, orchids, bamboo, and chrysanthemums—objects also endowed with noble character

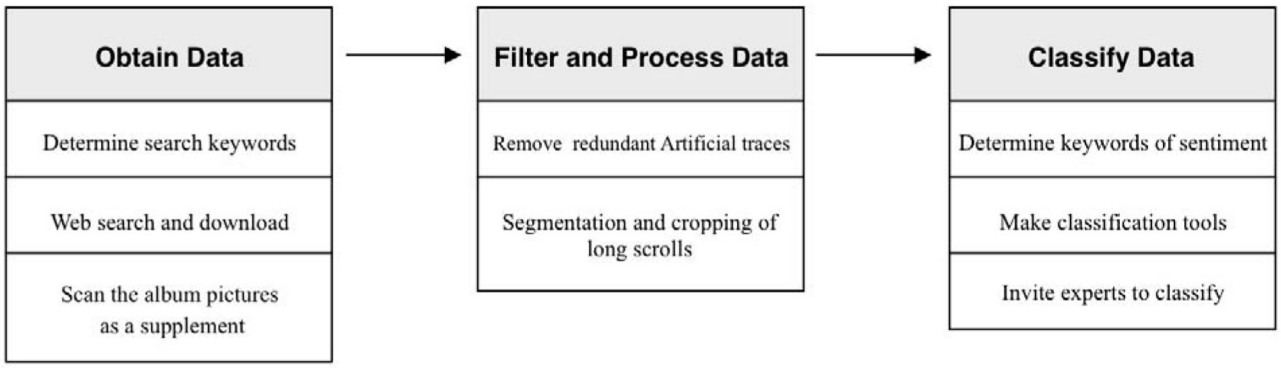


FIGURE 1 | Process of making a dataset.

and emotion. On the basis of strokes, composition, and objects, for Shi Su, painting is lyrical and expresses aspirations. Therefore, the emotional characteristics that literati paintings often present are closely related to the sentiment of the painter and are usually described as open-mindedness, moral loftiness, ethereality, leisure, seclusion, and so on. This study extracts three words that are frequently used, have a wide range, and have a certain distinction between each other, namely, morally lofty, seclusion, and leisure, for classification. Then, we created a literati painting sentiment classification system in the form of a web page, so that researchers can classify and choose the option of the image while browsing the images. An example of the use of the keywords mentioned above is shown in **Figure 3**. After creating the emotion classification system, this study invited 10 experts in the field of literati painting to select the picture emotions in the classification system, and the option selected in the largest proportion by people for each picture is the result of emotion classification. The number of specific categories under emotional qualifiers is as follows: 282 for morally lofty, 241 for seclusion, and 272 for leisure.

USE OF CNNS TO LEARN AND ANALYZE THE EMOTIONAL CHARACTERISTICS OF LITERATI PAINTINGS

In this study, based on the built-up literary painting emotion dataset, a CNN was used as an experimental model to identify the emotion of literati paintings. To explore the correlation between picture features and emotions under machine vision, we visualized the picture features recognized by the machine and explored the relationship between the picture features and emotions that the machine can extract through the study of the feature aggregation area.

Improving Model Generalization Capabilities by Data Augmentation

Considering the limited number of training samples produced to improve the training effect, the method of data enhancement was used to improve the robustness of the model so that the collected data can be fully utilized in the model training and enhance the generalization ability of the model. In general, common data augmentation methods included brightness enhancement, contrast enhancement, image rotation, image flip, image cropping, and Gaussian noise addition. In this study, the data of each classification in the dataset were divided into a training set and a test set according to a ratio of 7:3. The data in the training set were then expanded in several common ways, namely, image brightness enhancement, contrast enhancement, image rotation, and image flip. To ensure the integrity of the picture layout and to avoid random cropping caused by missing pictures, we did not choose image cropping. In terms of brightness enhancement, an increment of 10–50% random was selected to change the dependence of the model on the brightness of a large area. In contrast to enhancement, a random increment of 10–50% was selected to change the dependence of the model on the contrast of a large area. In the aspect of image rotation, a random angle in the range of approximately −180–180° was selected for rotation so that the model can learn the object of interest from many angles and improve the robustness of the model. With regard to image reversal, each image was flipped left or right with a 50% probability to change the position of the object of interest and reduce the dependence of the model on the target position. In terms of increasing noise, increasing Gaussian noise was chosen. The effect is shown in **Figure 4**. As the data were expanded, the amount of data in the training set eventually increased to six times the original amount.

Testing and Improvement of Neural Networks

Classic Network Analysis

After data augmentation, this study focused on the literati painting dataset using the classic networks of VGG16 (Liu et al., 2020), ResNet50 (He et al., 2016), MobileNet (Chechliński et al., 2019), GoogleNet Incepetion V1 (Szegedy et al., 2015), and GoogleNet Incepetion V3 (Szegedy et al., 2015) for sentiment classification. The test accuracies are shown in **Table 1**.

It can be seen from **Table 1** that VGG16 and ResNet50 have better accuracy than other networks in the test of

FIGURE 2 | Six gentlemen figure before data processing (left) and after data processing (right).

emotion classification of literature. To select a suitable model for improving the accuracy of sentiment classification step by step, this study analyzed the two models based on their characteristics. VLG16 mainly improved the detection performance by increasing the network depth. Therefore, deep network of VGG16 is more focused on a specific target than the shallow network and less focused on global information. ResNet50 introduced a residual network, which deepened the network depth and was able to integrate the shallow information into the deep information so that the model could utilize the global information of the picture to some extent. In summary, due to the overall effect of Chinese literati painting on the expression of emotion, this study chose ResNet50 to obtain better test results in the classification and recognition of the emotion of literati painting.

Improvement of the ResNet50 Network

For deep learning networks, the model effects are theoretically better and more expressive as the number of layers in the network increases. However, it has been found that, when the CNN network reaches a certain depth and continues to deepen the network, the classification performance will not improve. Rather, there is slower network convergence and lower accuracy. Thus, even if the problem of overfitting is solved by extending the dataset, the classification performance and accuracy cannot be improved. In 2015, when Kaiming et al. found that a residual network could solve the problem, they proposed ResNet50.

The ResNet50 network is composed of two basic residual modules: the Conv Block and the Identity Block. The structure of the Conv Block is shown in **Figure 5**. From the figure, it can be seen that the Conv Block consists of two parts. The left part

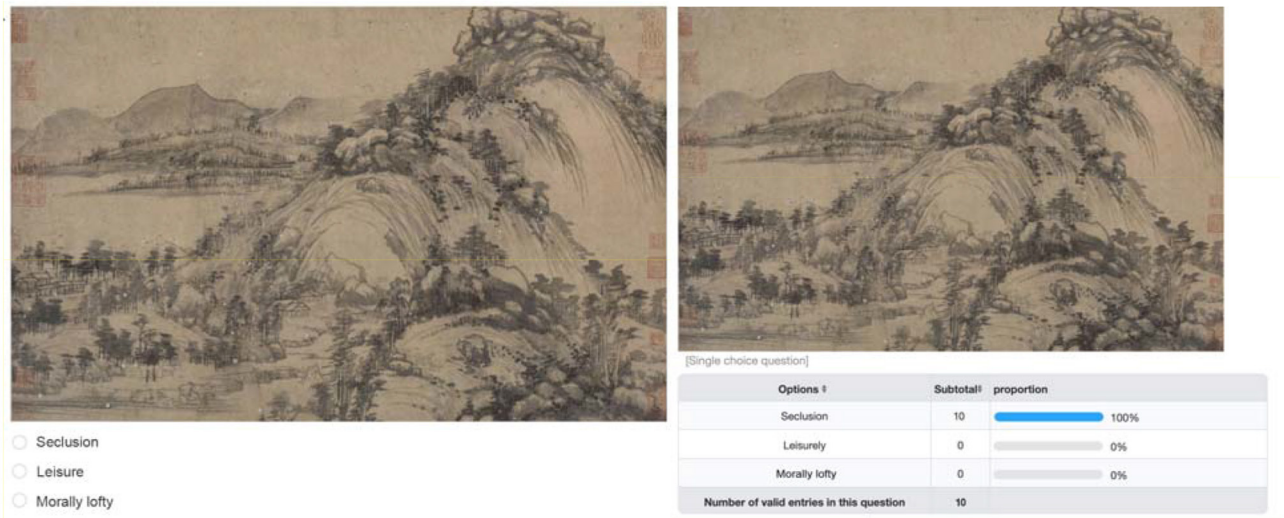


FIGURE 3 | Literati painting sentiment classification system.

FIGURE 4 | Results of data augmentation (left to right: original picture, brightness enhancement, contrast enhancement, image flip, image rotation).

is the main part, which is composed of the two convolution operations, standardization and ReLU activation function blocks, and one convolve and standardization block. The right part is the residual edge part, which is composed of a convolution block and a standard block. Then, the main part is added to the residual side part and the output is obtained by non-linear processing of the ReLU activation function. The residuals of the Conv Block go through a convolution operation that causes the input and output dimensions to be inconsistent, so they cannot be concatenated continuously. Their purpose is to change the dimensions of the network. The structure of the Identity Block is shown in **Figure 6**. As seen from the figure, the Identity Block is composed of two parts: the left part is the main part, which is consistent with the Conv Block. The right part is the residual side part, which adds directly to the trunk output without any processing and then outputs it through the non-linear processing of the ReLU activation function. The entity block input dimension is the same as the output dimension and can be concatenated to deepen the network. Finally, based on the above two types of residual modules, the Conv Block and the Identity Block constitute a complete ResNet50 network. The core modules are composed of 3, 4, 6, and 3 residual modules using the Conv Block module 4 times and Identity Block 12 times.

TABLE 1 | Test accuracies of the different models.

Model	Accuracy (%)
MobileNet	38.89
GoogleNet Incepetion V1	41.82
GoogleNet Incepetion V3	45.19
VGG16	52.40
ResNet50	51.92

To improve the accuracy of the emotional classification of literary painting, this study referred to the method of feature recalibration in deep network aggregation (Wang et al., 2021) to improve the main part of the residual module. This was done by converging the main part of the multiple convolution layer characteristic information, and then the information of the integration of each layer is fed back to the output of the main part for reintegration. The process is shown in **Figure 7**. After global average pooling, the residual module and the main part of

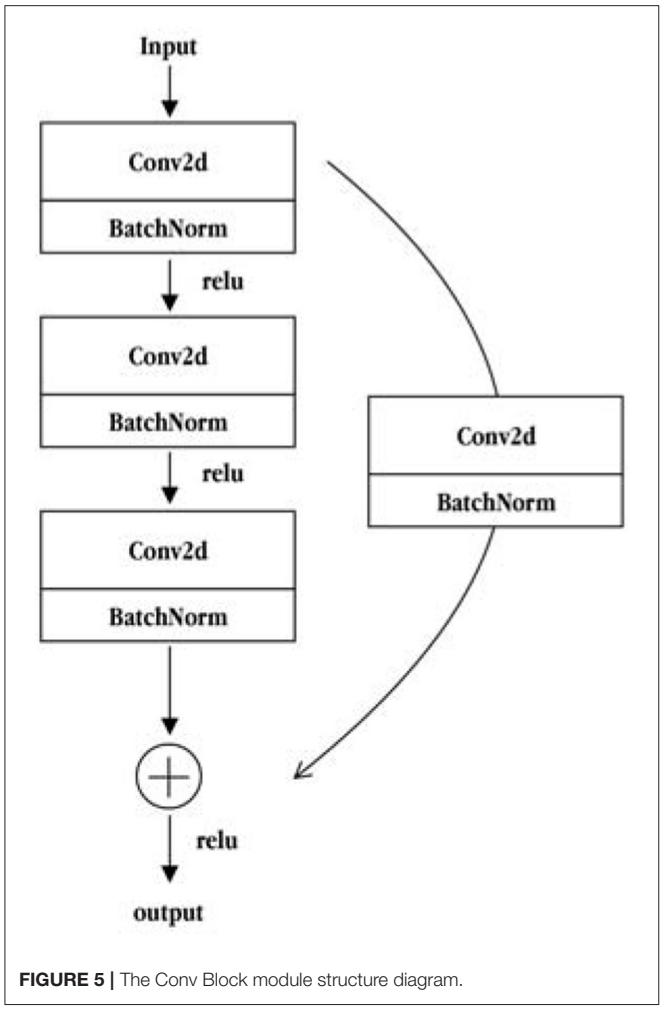


FIGURE 5 | The Conv Block module structure diagram.

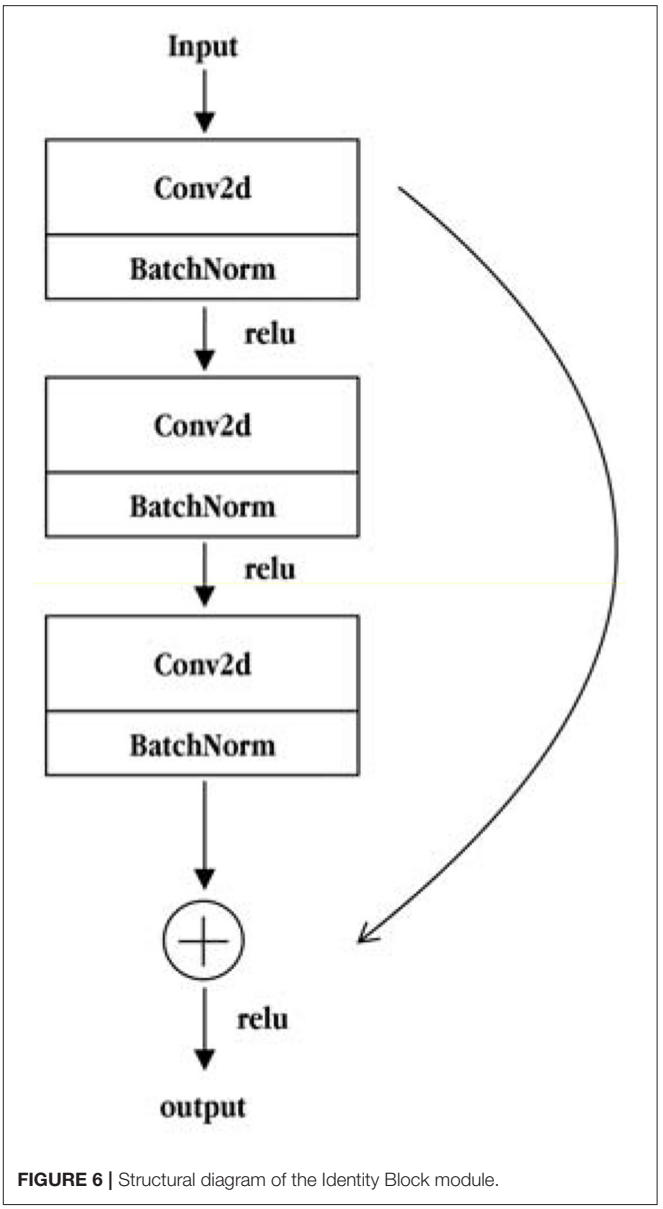


FIGURE 6 | Structural diagram of the Identity Block module.

the first two convolution layer characteristic information merges. The merged vector dimension of the fully connected layer are combined with the globally averaged feature information of the third convolution layer. Then, they form a bottleneck structure through a lower-dimensional full-connection layer and a layer-up full-connection layer to enhance the feature information extraction ability. Finally, the characteristic information of the last three convolutional layers is used as the weight, and the third convolution layer output information of the main part is integrated to form the final main part convolution layer characteristic fusion residual module. As shown in **Figure 7**, on the basis of the residual module, the main convolution layer feature fusion residual module integrates the three-layer convolution layer information of the main dry part of the module to feeds it back to the output of the main edge. This further enhances the use of deep information to integrate shallow information for better analysis of the emotion of the literati painting.

Based on the main coil feature fusion residual module proposed above, three sets of improvement schemes were adopted: feature fusion improvement only for the Conv Block module, feature fusion improvement only for the Identity Block module, and feature fusion improvement for both modules at the same time. The resulting improvement is shown in **Table 2**.

It can be seen from the data in **Table 2** that only the feature fusion of the Conv Block module and only the feature fusion of the Identity Block module exhibited a certain improvement in the optimal model accuracy, which is 2.25% more accurate compared to the classic ResNet50, and 1.77% more accurate compared to the classic VGG16. However, when the two modules are optimized for feature fusion at the same time, the optimal model accuracy is lower than the accuracy of the classic model. This phenomenon may be due to the repeated reuse of shallow information in the network so that deep semantic information is more focused on specific objectives. Thus, better accuracy cannot be achieved.

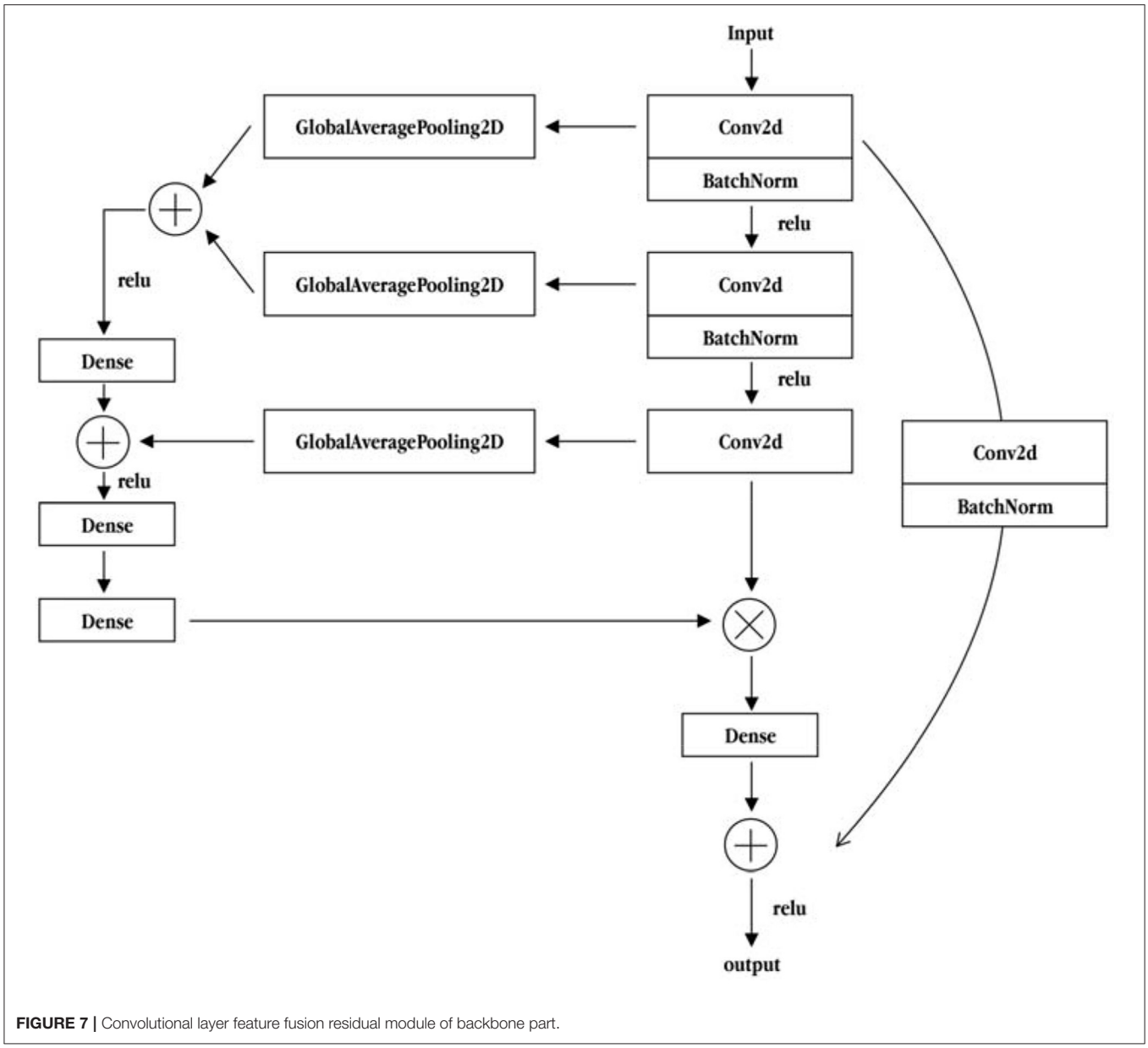


FIGURE 7 | Convolutional layer feature fusion residual module of backbone part.

In addition, to explore why the optimal accuracy rates of the first two improved models were the same, this study analyzed the change curves of the accuracy of the two types of improved models to evaluate the stability of the two improved models. The curves of the accuracy of the two types of improved models with the number of training rounds are shown in **Figures 8**, **9**.

By analyzing the above two kinds of curves, we found that they have some common characteristics. First, the accuracy of the first 50 rounds of training was low because, in order to speed up the efficiency of training in the first 50 rounds that used the frozen training ideas, only part of the data training played a warmup and preparation role. Another common feature was that, after 100 rounds of training, the accuracy of the model tended to stabilize; however, this was not the optimal model, and analysis may suggest increasing the number of model training rounds. On the one hand, the model exhibited the overfitting phenomenon; on the other hand, the model of a specific target understanding further deepened so that the overall information of the model, with the increase in the number of training rounds, was decreased. Based on the above

TABLE 2 | Three improvement schemes and their model accuracies.

Improve Proposals	Accuracy of Model (%)
Feature fusion is only performed on the Conv Block module	54.17
Feature fusion is only performed on the Identity Block module	54.17
Feature fusion of two modules at the same time	50.93

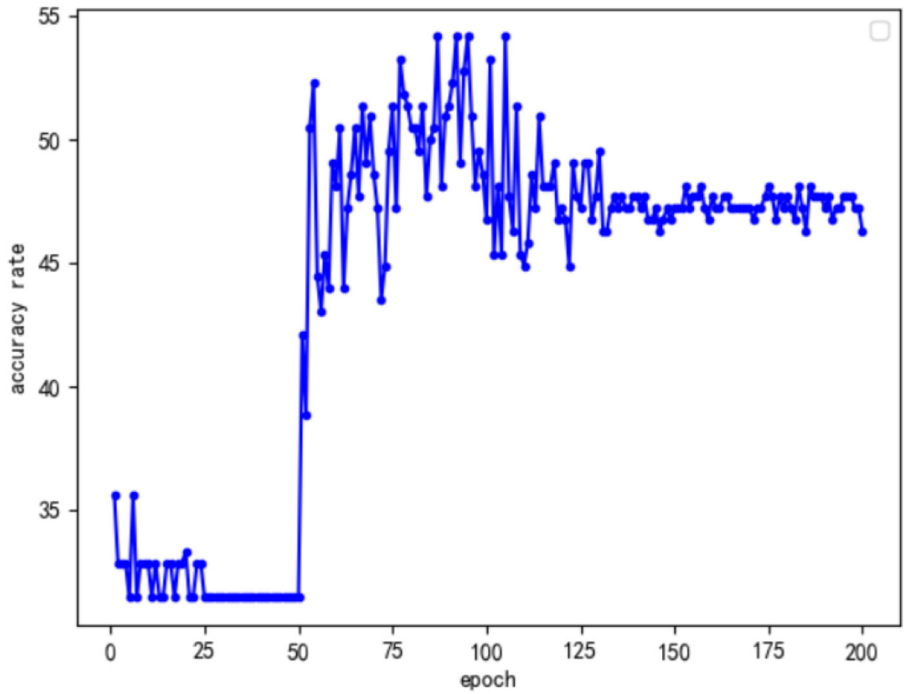


FIGURE 8 | Only perform feature fusion model accuracy curve for the Conv Block module.

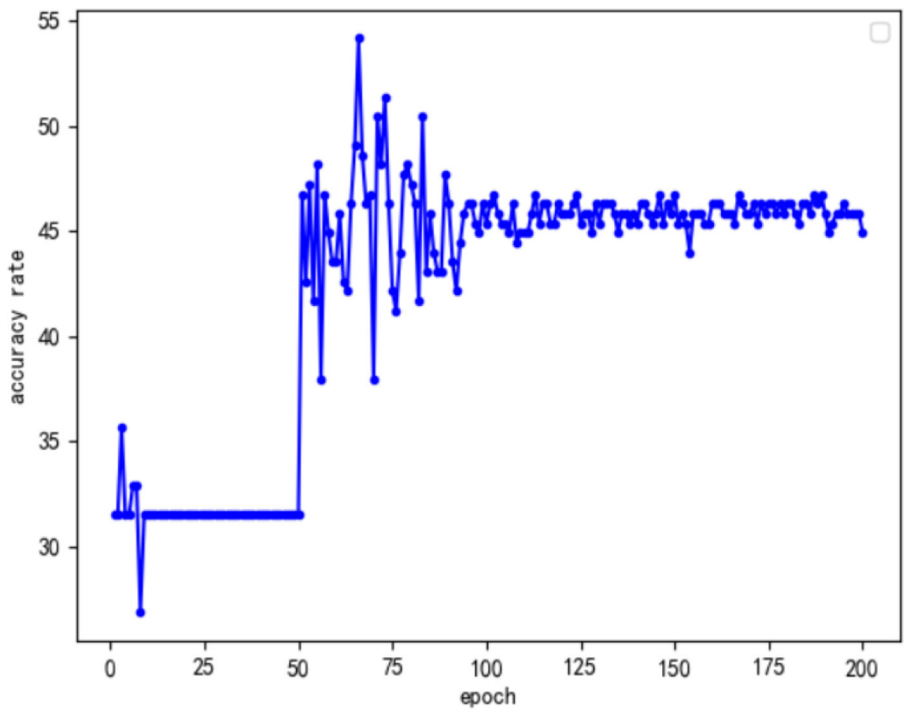


FIGURE 9 | Only perform feature fusion model accuracy curve for the Identity Block module.

analysis, this study focused on the training between round 50 and round 100, and the results are as follows: Comparing the model accuracy of the two types of optimization models during the 50 to 100 rounds, it was found that only the Conv Block module feature fusion model accuracy was stable, as there were many peaks of optimal accuracy. Furthermore, only for the

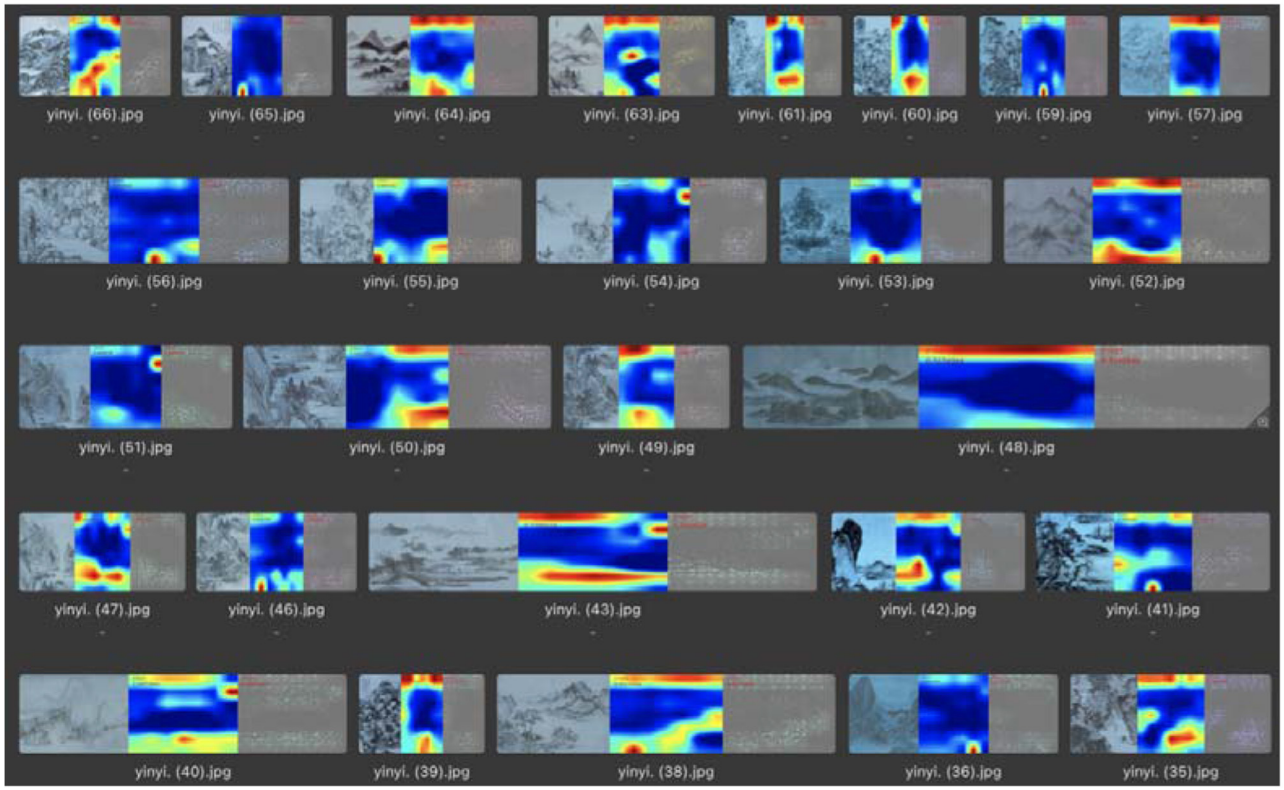


FIGURE 10 | Feature visualization results in the "seclusion" category.

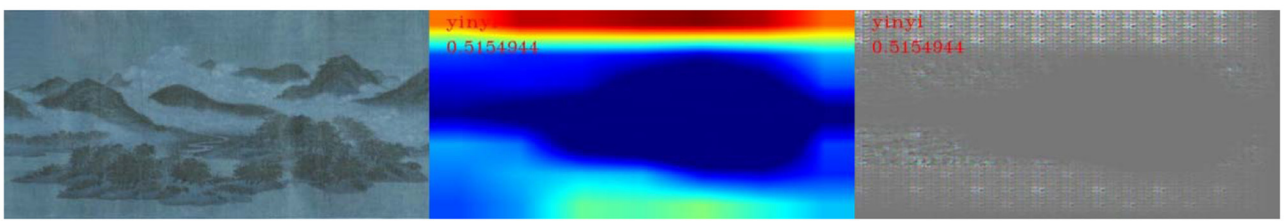


FIGURE 11 | Sky and river corresponding to the "seclusion" feature.

accuracy of the identity block module feature fusion model, after the first peak with an increase in the number of epochs, the overall model accuracy showed a clear downward trend. Combined with the analysis of the structure of ResNet50, it was not difficult to see that there were 16 residual modules in the ResNet50 network, only 4 Conv Block modules and 12 Identity Block modules. Thus, as the number of trainings increased, the Identity Block module was more affected by specific targets than the Conv Block module. Given the effect of the Conv Block module in increasing the network dimensions, the Identity Block module has the effect of increasing the depth of the network. In view of the above analysis, it can be concluded that, in future research, we can improve the neural network model from the perspective of improving the model dimensions in order to achieve a better emotional analysis of the effect of literati painting.

Visualize the Feature Gathering Area

This study proposes a new method of art image derivative design based on emotion analysis. First, the significance of personalization was analyzed in the design of art derivatives. Then, the problems existing in the personalized design method were pointed out, the new needs of users in the background of intelligent technology were analyzed, new design ideas combined with new technology were put forward, the main technical principles of the technology were analyzed, and practical solutions were proposed to explore the new design paradigm. Finally, by constructing a personalized design system of image derivatives driven by user facial emotion data based on the paintings of Van Gogh, this study verified the feasibility of the solution. It was found that the innovation of the design paradigm is imperative. It was significant for us to change the original design paradigm and

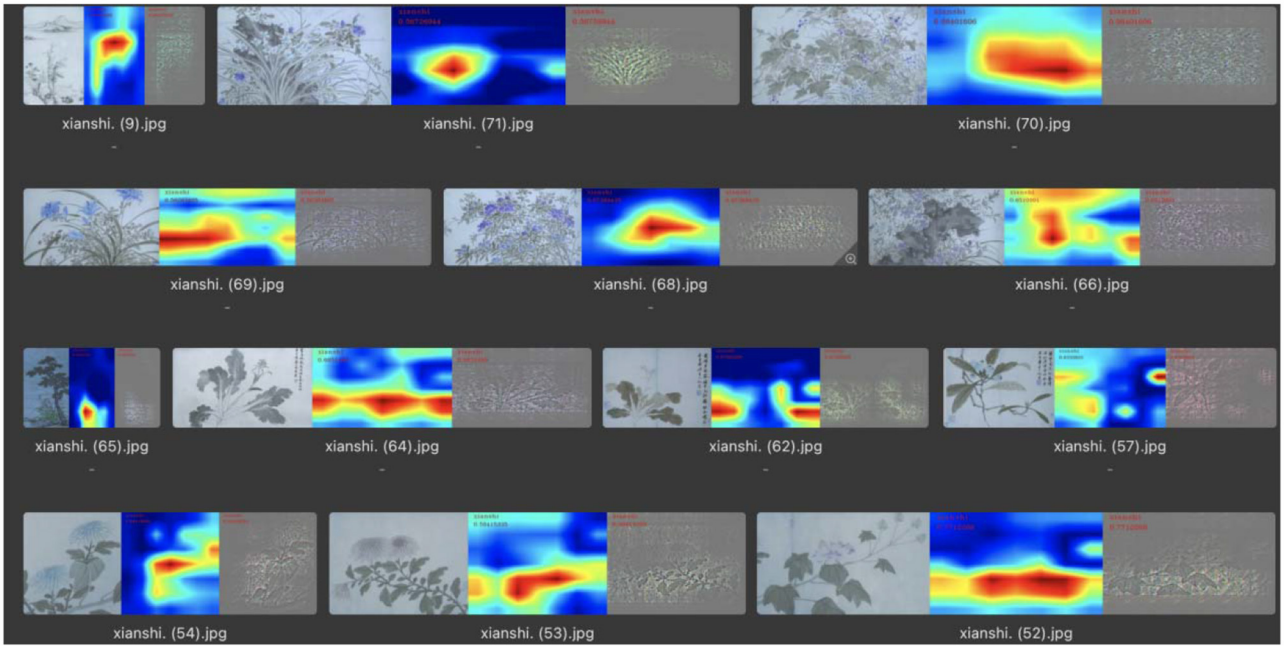


FIGURE 12 | Feature visualization results in the "leisure" category.

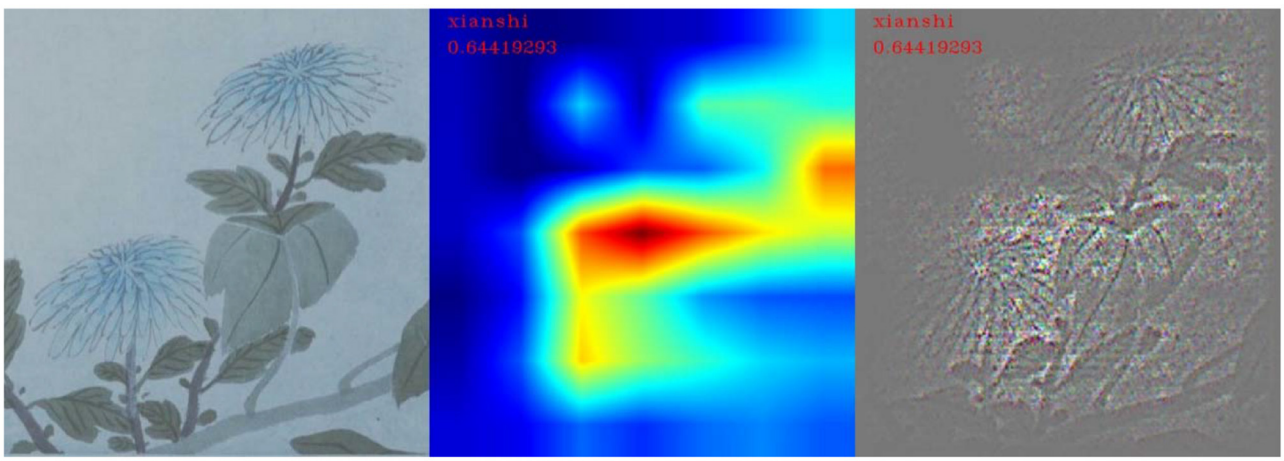


FIGURE 13 | Flowers and leaves corresponding to the "leisure" feature.

introduce new technology, which could improve the design using current new technology and better serve the needs of the public.

To explore the cognitive basis of the emotional analysis of Chinese literati paintings by machines from computer vision, we visualized the characteristic areas related to emotional keywords in the sample of the paintings of literati according to the model visualization method, Guided Grad_CAM (Selvaraju et al., 2020), and then explored the relationship between the content and emotion of the picture from the computer vision. The Grad_CAM process of visualizing the emotional characteristics in the sample of literati painting consisted of the following steps. First, the image was obtained after the last convolution of the feature extraction. In this study, the feature map was the final layer of the ResNet50 convolution layer res5c_branch2c. Then, we selected the node with the highest softmax value, the category with the highest confidence to reverse propagation, found the gradient for the last convolution layer, and took the mean gradient of each feature map as the weight of the feature map. Next, we multiplied each feature map by the weight to obtain the weighted feature map, activated it with the ReLU activation function, normalized it, and obtained the

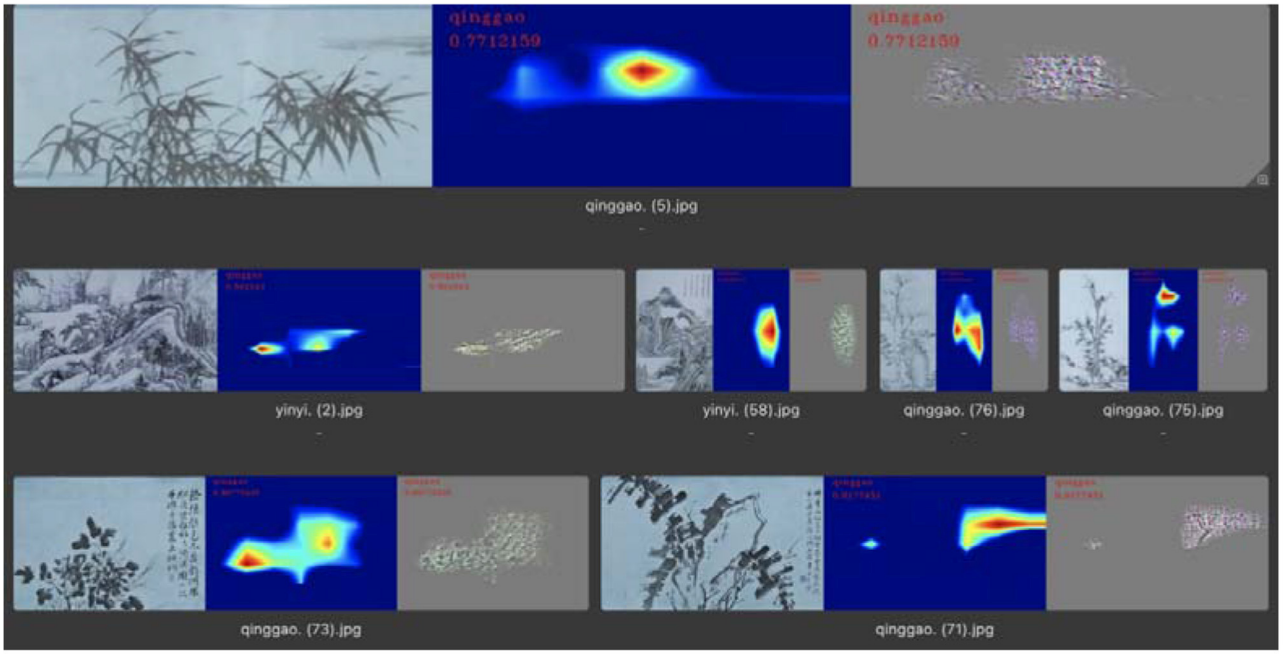


FIGURE 14 | Feature visualization results in the "morally lofty" category.

thermal effort (Grad_CAM heatmap). The ReLU function was chosen to remove feature diagrams that contributed negatively to classification, i.e., useless features, and to preserve only feature maps that have a positive contribution to feature classification, so that the resulting positioning map only reflected the category features with the highest confidence. The softmax value was then guided through backpropagation, that is, when propagation was reversed, only values with inputs and gradients >0 were passed, which resulted in a feature image (G-B image) containing deep semantic information. Finally, the thermal effort (Grad_CAM heatmap) and the feature image (G-B image) containing deep semantic information were merged to form the final Guided Grad_CAM image.

After completing the training of the model, the pictures in the test set were put into the model for emotional classification, the characteristic gathering area of the correct images was visualized, and the connection between the picture content and the expression of emotion was explored by analyzing the visualization results. The visualization results were analyzed as follows:

In the recognition of "seclusion," the number of test pictures was 68, and the number of correct classifications was 55. In this category, the visualized significant feature areas were mainly concentrated at the top and bottom of the picture, each occupying approximately a quarter of the entire area, which also corresponded to the whitening in the picture. Thus, it can be concluded that the recognition of the "seclusion" emotion in the picture by the machine is mainly based on the analysis of the location and area of the white areas left in the picture to obtain the classification results, as shown in **Figure 10**. A comprehensive analysis concluded that, in the machine vision of the literary painting emotion analysis, the expression of the "seclusion" emotion of the painting at the top and bottom of the canvas each have a quarter of its size in white. Combined with the semantic analysis of the literati painting, these areas are mostly sky and river, as shown in **Figure 11**.

In the identification of "leisure" emotions, the number of test pictures was 63, and the number of correct classifications was 32. In this category, the visualized area of remarkable features was mainly concentrated on the objects in the picture, so the extraction of the features of the picture by neural networks was mainly in the recognition of objects in the picture, as shown in **Figure 12**. In the correctly classified test images, the main objects in the picture labeled "leisure" were mostly flowers, birds, rocks, and plants with leaves, as shown in **Figure 13**. Therefore, the machine mainly recognizes the abovementioned objects to determine whether it is a "leisure" emotion.

In the identification of the "morally lofty" emotion, the number of test pictures was 71, and the number of correct classifications was 19. Because of the similarity between the picture features of "seclusion" and "morally lofty," the effect of the machine in the emotional classification was not good, and the paintings expressing "morally lofty" emotion were often misjudged as "seclusion," thus affecting the visualization of features under the category of "morally lofty." However, an analysis of a small number of classified and accurate graphs revealed that the visualized area of distinctive features was the center of the main concentrated picture and that the area corresponded to the blank in the picture, as shown in **Figure 14**.

In summary, through the results of visual analysis, it can be concluded that, in paintings expressing "seclusion" emotions, the feature extraction of the machine is mainly concentrated in the blank part of the screen, and the top and bottom of this type of screen both occupy ∼4% of the screen area. In one part of the blank spaces in the paintings that express "leisure" emotions, the feature extraction of the machine was mainly concentrated on the main objects in the picture. The main objects in this type of picture were generally flowers, birds, rocks, and plants with leaves. In the paintings expressing the emotion of "morally lofty," the extraction of features by the machine was mainly concentrated in the blank part of the picture. In this type of painting, there was a small blank area in the center of the picture.

CONCLUSION

This study combined existing machine learning technology with a self-built Chinese literati painting sentiment dataset to train five classic neural network models, VGG16, ResNet50, MobileNet, GoogleNet Incepetion V1, and GoogleNet Incepetion V3. Through a comparison of the accuracy of models, ResNet50 was selected as the most appropriate model for the emotion classification of literati paintings. Then, based on the multilayered aggregation feature of the recalibration algorithm, the study improved ResNet50 and obtained three models with different improvement schemes. By visualizing the accuracy under different improved methods after each round of training, it was found that the improved Conv Block model was the most stable. Finally, the present study put the images in the test set into the trained model for emotion classification and visualized the salient feature regions of the correctly classified pictures. By summarizing the rules, a series of picture contents and the connection relationship between the sentiments the paintings expressed were obtained.

Although the existing research results have been effective, there are still areas for improvement. First, due to the limited number of pictures in the literati painting sentiment classification dataset, the training effect of the model was greatly reduced, which resulted in the accuracy of the model being unable to meet expectations. In addition, finding a more appropriate data augmentation method can improve the accuracy of the emotion classification of the model. Therefore, in subsequent research, researchers can improve the performance of the dataset by improving the richness and professionalism of the dataset, comparing multiple parameters, and choosing the best data augmentation method to improve the performance of the dataset, thereby training the generalization ability to be a stronger literati painting sentiment of the classification model. At the same time, by improving the model of the neural network from the perspective of increasing the model dimension, the accuracy of the model can be improved.

Second, the effect of the existing feature region visualization was still not sufficiently ideal to explore the differences in the content of pictures with similar characteristics in different classifications and find the reasons for the above problem. Therefore, in future research, researchers can enhance the ability of the model to visualize the extracted features to obtain a clearer calibration map of salient feature regions. With this, we can better summarize and analyze literati painting images under different expressions of sentiments, which helps us to more accurately understand the connection between the content of the image and the sentimental expression from machine vision. In addition, this type of method can be used for the emotional analysis of paintings of other styles, for the exploration of the inner relationship between the aesthetic laws and emotional expression of the creator through the assistance of computer vision and for assisting people in understanding the emotion of paintings in a variety of styles. Furthermore, in addition to being connected with emotions, the characteristics of paintings can be connected with external factors such as the life experience of the creator and social events affecting artistic creation. This can help researchers to explore the relationship between the characteristics of the picture and the reasons for the mutual influence between the characteristics of the picture and the external factors. The sentiments of literati paintings are confusing, but this special group of literati mastered the highest culture and the highest thoughts in ancient China. Therefore, literati paintings, as sentimental products of the literati, must convey the culture, thoughts, and emotions of ancient Chinese society. The study of literati painting helps to understand the spirit of Chinese culture and art at a deeper level and helps to grasp the artistic characteristics and advanced factors of Chinese painting as a whole.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/zzj-dyj/classification-of-literati-painting-sentiment/tree/master/datasets.

AUTHOR CONTRIBUTIONS

JZ, YD, and XG developed the theoretical framework and model in this study and drafted the manuscript. YD implemented the algorithm and performed the experiments and result analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Social Science Foundation of China Key Project of Art Science "Research on Chinese Animation Creation and Theoretical Innovation under the Construction of National Cultural Image" (Grant No. 20AC003) and by Jiangsu University Philosophy and Social Sciences Research Fund under Grant No. 2018SJA0808.

REFERENCES

Adriana, I. (2019). The kwanyin clan: modern literati graffiti writers. *Annali di Ca' Foscari: Serie Orientale.* 55 (1)

Bowie, T. (2014). Oriental art: susan bush, the Chinese literati on painting: Su Shih (1037-1101) to Tung Ch'ich'ang (1555-1636). *Art J.* 32.4, 457–462. doi: 10.2307/775710

Castellano, G., and Vessio, G. (2021). Deep learning approaches to pattern extraction and recognition in paintings and drawings: an overview. *Neural Comput. Appl.* doi: 10.1007/s00521-021-05893-z

Chechliński, U., Siemitkowska, B., and Majewski, M. (2019). A system for weeds and crops identification-reaching over 10 FPS on raspberry Pi with the usage of mobilenets, densenet and custom modifications. *Sensors* 19:3787. doi: 10.3390/s19173787

Edward, H. O. (1997). Aesthetic considerations in understanding Chinese literati musical behaviour. *Ethnomusicol. Forum.* 6, 35–49. doi: 10.1080/09681229708567260

Fan, Z. B., and Zhang, K. (2020). Visual order of Chinese ink paintings. *Visual Comp. Indust. Biomed. Art.* 3, 1–9. doi: 10.1186/s42492-020-00059-5

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Commun. ACM.* 56, 82–89. doi: 10.1145/2436256.2436274

Guo, X., Yu, W., and Wang, X. (2021). An overview on fine-grained text sentiment analysis: Survey and Challenges. *J. Phys. Conf. Series.* 1757, 012038. doi: 10.1088/1742-6596/1757/1/012038

He, K., Zhang, X., and Ren, S. (2016). "Deep residual learning for image recognition". in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV). 770–778.

Liu, C., and Jiang, H. (2014). Classification of traditional Chinese paintings based on supervised learning methods. 2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC). *IEEE.* 641–644. doi: 10.1109/ICSPCC.2014.6986272

Liu, Z., Wu, J., Fu, L., Majeed, Y., and Cui, Y. (2020). Improved kiwifruit detection using pre-trained VGG16 with RGB and NIR information fusion. *IEEE Access.* 8, 2327–2336. doi: 10.1109/ACCESS.2019.2962513

Medhat, W., Hassan, A., and Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Eng. J.* 5, 1093–1113, doi: 10.1016/j.asej.2014.04.011

Ortis, A., Farinella, G. M., and Torrisi, G. (2020). Exploiting objective text description of images for visual sentiment analysis. *Multimedia Tools Appl.* doi: 10.1007/s11042-019-08312-7

Prabowo, R., and Thelwall, M. (2009). Sentiment analysis: a combined approach. *J. Inform.* 3, 143–157, doi: 10.1016/j.joi.2009.01.003

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2020). Grad-cam: visual explanations from deep networks *via* gradient-based localization. *Int. J. Comp. Vision.* 128, 336–359. doi: 10.1007/s11263-019-01228-7

Sun, M., Zhang, D., Wang, Z., Ren, J., and Jin, J. S. (2016). Monte Carlo convex hull model for classification of traditional Chinese paintings. *Neurocomputing* 171, 788–797. doi: 10.1016/j.neucom.2015.08.013

Szegedy, C., Wei, L., Jia, Y., Sermanet, P., and Rabinovich, A. (2015). "Going deeper with convolutions", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Boston, MA). 1–9.

Vigneron, F. (2011). Le Songe Creux-between literati theory, minimalism and maximalism. *World Art.* 1, 101–124. doi: 10.1080/21500894.2011.533473

Wang, Z., Guo, J., Huang, W., and Zhang, S. (2021). High-resolution remote sensing image semantic segmentation based on a deep feature aggregation network. *Meas. Sci. Technol.* 32:095002. doi: 10.1088/1361-6501/abfbfd

Wong, A. Y. (2000). A new life for literati painting in the early twentieth century: Eastern art and modernity, a transcultural narrative? *Artibus Asiae.* 60, 297–326, doi: 10.2307/3249921

Xia, R., Zong, C., and Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. *Inform. Sci.* 181, 1138–1152, doi: 10.1016/j.ins.2010.11.023

Xu, J., Huang, F., Zhang, X., Wang, S., Li, C., Li, Z., et al. (2019). Sentiment analysis of social images *via* hierarchical deep fusion of content and links. *Appl. Soft Comput.* 80, 387–399. doi: 10.1016/j.asoc.2019.04.010

Xue, L., Mao, Q., Huang, X., and Chen, J. (2020). NLWSNet: a weakly supervised network for visual sentiment analysis in mislabeled web images. *Front. Inform. Technol. Electronic Eng.* 21, 1321–1333. doi: 10.1631/FITEE.1900618

Zhang, J., Chen, M., Sun, H., Li, D., and Wang, Z. (2020). Object semantics sentiment correlation analysis enhanced image sentiment classification. *Knowledge-Based Syst.* 191:105245. doi: 10.1016/j.knosys.2019.105245

Zhang, K. E., Zhu, Y., Zhang, W., and Zhu, Y. (2021). Cross-modal image sentiment analysis *via* deep correlation of textual semantic. *Knowledge-Based Syst.* 216:106803. doi: 10.1016/j.knosys.2021.106803

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhang, Duan and Gu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 29 July 2021
doi: 10.3389/fpsyg.2021.721266

A Domain Adaptation Sparse Representation Classifier for Cross-Domain Electroencephalogram-Based Emotion Classification

Tongguang Ni[1], *Yuyao Ni*[2], *Jing Xue*[3]* *and Suhong Wang*[4]*

[1] School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, China, [2] School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China, [3] Department of Nephrology, Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi, China, [4] Department of Clinical Psychology, The Third Affiliated Hospital of Soochow University, Changzhou, China

OPEN ACCESS

Edited by:
Yaoru Sun,
Tongji University, China

Reviewed by:
Qin Qin,
Henan Institute of Engineering, China
Lijun Xu,
Nanjing Institute of Technology (NJIT), China

***Correspondence:**
Jing Xue
xuejing@njmu.edu.cn
Suhong Wang
yarmine@foxmail.com

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: 06 June 2021
Accepted: 28 June 2021
Published: 29 July 2021

Citation:
Ni T, Ni Y, Xue J and Wang S (2021) A Domain Adaptation Sparse Representation Classifier for Cross-Domain Electroencephalogram-Based Emotion Classification. Front. Psychol. 12:721266. doi: 10.3389/fpsyg.2021.721266

The brain-computer interface (BCI) interprets the physiological information of the human brain in the process of consciousness activity. It builds a direct information transmission channel between the brain and the outside world. As the most common non-invasive BCI modality, electroencephalogram (EEG) plays an important role in the emotion recognition of BCI; however, due to the individual variability and non-stationary of EEG signals, the construction of EEG-based emotion classifiers for different subjects, different sessions, and different devices is an important research direction. Domain adaptation utilizes data or knowledge from more than one domain and focuses on transferring knowledge from the source domain (SD) to the target domain (TD), in which the EEG data may be collected from different subjects, sessions, or devices. In this study, a new domain adaptation sparse representation classifier (DASRC) is proposed to address the cross-domain EEG-based emotion classification. To reduce the differences in domain distribution, the local information preserved criterion is exploited to project the samples from SD and TD into a shared subspace. A common domain-invariant dictionary is learned in the projection subspace so that an inherent connection can be built between SD and TD. In addition, both principal component analysis (PCA) and Fisher criteria are exploited to promote the recognition ability of the learned dictionary. Besides, an optimization method is proposed to alternatively update the subspace and dictionary learning. The comparison of CSFDDL shows the feasibility and competitive performance for cross-subject and cross-dataset EEG-based emotion classification problems.

Keywords: electroencephalogram, domain adaptation, emotion classification, cross-subject, cross-dataset

INTRODUCTION

Emotion is the attitude experience and corresponding behavior response of human beings to objective things, which has an important influence on human behavior and mental health. How to accurately identify emotions has an important significance in practical application. For example, in the medical field, emotion recognition is helpful to guide and diagnose patients with mental

diseases or expression disorders, and in the education field, different teaching methods according to the emotion of the listener can improve the teaching efficiency.

Emotion recognition using a variety of modal emotion signals has now gained a lot of attention from researchers. Typically, emotions can be perceived in the form of a variety of signals. One type of visual signal can be directly observed from external behavior and characteristics, such as facial expressions, voice intonation, body movements, etc. The other type is those physiological signals, such as electroencephalography (EEG), electromyography (EMG), electrocardiogram (ECG), skin conductance, pulse, heartbeat, skin temperature, and respiratory signals; however, facial expressions, voice, and other non-physiological signals are easily restricted by environmental or social factors. The emotional information transmitted by physiological signals is more objective and can reflect the psychological emotion more reliably (Zheng et al., 2019; Doma and Pirouz, 2020; Ni et al., 2020a).

Brain-computer interface (BCI) is a human-computer interaction system that provides a communication channel for human brain interaction with the external environment and without depending on the peripheral nervous system and muscles (Zhang and Wu, 2019; Liu et al., 2020; Ni et al., 2020b). EEG plays a dominant role in emotion recognition based on physiological signals. The illustration of EEG-based emotion classification in BCI is shown in **Figure 1**. The operation of emotion classification begins with the presentation of stimuli to the user, which induces specific emotions. The stimuli may be music, videos, and images, etc. During the session, EEG samples are recorded by EEG devices. The next step is usually to extract features from the recorded EEG and train a classifier. The final step is to test new EEG samples to classify emotion labels (Liu et al., 2018).

The existing EEG-based emotional classification in BCI requires a large amount of label data and a lot of time in the training phase. A relatively simple and direct method is to reuse previously collected EEG data and train a new classifier, without considering differences between individuals. These classification methods are based on the assumption that training and test data are independently and identically distributed. This assumption is often difficult to hold for BCI, because EEG signals have non-stationary characteristics, and the performance of the classifier fluctuates significantly between subjects and datasets. When the same classifier is applied to EEG data of other subjects or from other datasets, the performance will be significantly reduced.

Domain adaptation learning is a fast and effective solution for developing a classifier that selectively trains a new classifier in TD using auxiliary data (source domain, SD) and less training data in the new scenario (target domain, TD) (Fahimi et al., 2019; Ni et al., 2020c). Different from multi-task learning that aims to benefit the classifier in both source and target tasks, domain adaptation learning mainly aims to benefit the classifier in TD. For example, Yang et al. (2019) proposed a support vector machine (SVM) combined with the significance test and sequential backward selection strategy for cross-subject EEG-based emotional classification. Instead of utilizing features on raw EEG signals, this method analyzed and selected features based on the significant differences between positive and negative trials. Li et al. (2020a) proposed a two-stage multi-source semi-supervised transfer learning method, in which the work of the first stage was source domain selection and the second stage was to learn style transfer mapping. This method selected the appropriate sources and projected source data to the destination *via* an affine mapping, so that only a few labeled data was used in the calibration sessions. Subsequently, Li et al. (2020b) developed a joint distribution adaptation method for EEG-based emotion classification in cross-subject and cross-session scenarios. The label information in the SD was used to train the model, and it also took an important part in reducing the difference of conditional distribution. This method achieves domain adaptation by combining marginal distributions and conditional distributions in the framework of neural networks. Morioka et al. (2015) also developed a cross-subject and cross-session recognition method, which learned the common spatial bases underlying both SD and TD by using unsupervised dictionary learning. The spatial transforms technology was found to be efficient in extracted common brain activities. Lan et al. (2018) developed a domain adaptation method to reduce discrepancies across datasets and inter-subject variance. This method designed a linear transformation function to adapt subspaces feature to match the marginal distributions of SD and TD.

Although domain adaptation for EEG-based emotion classification has been extensively studied, most of the studies focus on cross-subject and cross-session adaptation within the same dataset, i.e., the samples of SD and TD came from the same dataset. Domain adaptation across datasets is more challenging. Because cross-dataset domain adaptation is restricted to the different datasets, EEG signals are collected from the different devices and different stimuli, etc. (Li et al., 2018; Yang et al., 2019; Cimtay and Ekmekcioglu, 2020).

It is our opinion that although the distribution of common characters in EEG signals shows differences between subjects and datasets, it is expected that there might be some certain common knowledge that is potentially independent of the subjects and datasets. In addition, the shared common knowledge could be preserved in a shared projection subspace. Thus, we propose a domain adaptation sparse representation classifier (DASRC) to address the EEG-based emotion classification in cross-subject and cross-dataset scenarios. We consider learning the common component in both SD and TD by exploring a common dictionary in a shared subspace. Thus, we adopt the local information preserved criterion to reduce the domain distribution discrepancies in the learned subspace. We learn the common domain-invariant dictionary, which builds a connection between SD and TD. In addition, the principal component analysis (PCA) and Fisher criteria are exploited in this model to promote the recognition ability of the learned classifier.

The main contribution of the study is as follows. First, DASRC exploits the common characteristics of EEG data in SD and TD to yield a domain-invariant dictionary in the shared subspace. It takes advantage of the local data information preserved in both SD and TD. This allows enhancing the domain adaptation in

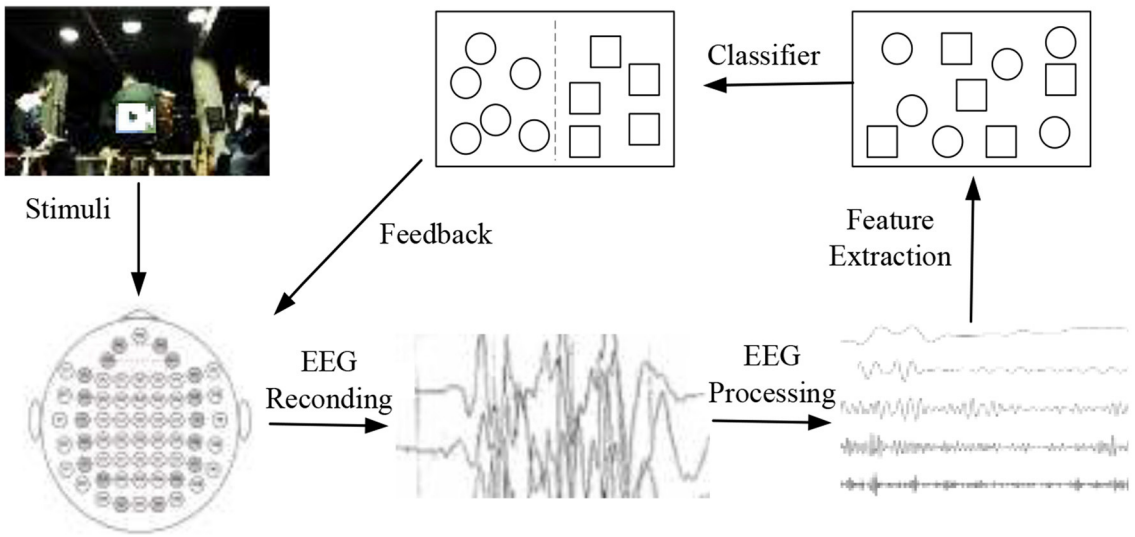


FIGURE 1 | The illustration of EEG-based emotion classification in BCI.

subspace. Second, using PCA and Fisher criteria, the objective function of DASRC is directly related to the classification rule. This strategy can promote the recognition ability in the domain-invariant subspace. Mathematically, an alternating optimization algorithm is proposed to solve the subspace and dictionary learning problem. Third, experiments on SJTU emotion EEG dataset (SEED) (Zheng and Lu, 2015) and dataset for emotion analysis using EEG, physiological and video signals (DEAP) (Koelstra et al., 2011) demonstrate that dictionary learning in subspace is effective and DASRC outperforms the advanced methods in cross-subject and cross-dataset scenarios.

BACKGROUND

Sparse representation is a data analysis method to estimate the sparse representation of measurable signals completely. It originated from neuroscience and has been used in signal processing, such as denoising and compression (Kanoga et al., 2019; Gu et al., 2020). In pattern recognition, sparse representation has also been proved to be suitable for classification. In sparse representation, the data matrix $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n]$ can be decomposed into a linear combination of a few atoms on the dictionary,

$$\mathbf{Y} \approx \mathbf{DA}, \tag{1}$$

where $\mathbf{D}$ is the dictionary matrix, and $\mathbf{A}$ is the coefficients matrix.

An adequate approximation makes $\mathbf{DA}$ the sparse representation as a reasonable estimation of$\mathbf{Y}$. Based on this concept, Equation (1) can be rewritten as follows:

$$\begin{aligned} &\arg\min_{\mathbf{D},\mathbf{A}} \ \|\mathbf{Y} - \mathbf{DA}\|_F^2, \\ &s.t. \ \forall i, \|\mathbf{a}_i\|_0 \leq l_0, \end{aligned} \tag{2}$$

where l_0 represents the sparsity constraint, $\mathbf{a}_i$ is the sparse coefficient vector to represent $\mathbf{y}_i$ over D. K-singular value decomposition (KSVD) algorithm is one of the most representative to solve Equation (2), in which the sparse coding and dictionary are updated alternately (Aharon et al., 2006).

DOMAIN ADAPTATION SPARSE REPRESENTATION CLASSIFIER

In this study, we consider the EEG data from two different domains SD and TD. The SD is with sufficient labeled samples $\mathbf{Y}^s = [\mathbf{y}_1^s, \mathbf{y}_2^s, ..., \mathbf{y}_{n_s}^s] \in R^{d \times n_s}$ and the TD is with limited labeled samples $\mathbf{Y}^t = [\mathbf{y}_1^t, \mathbf{y}_2^t, ..., \mathbf{y}_{n_t}^t] \in R^{d \times n_t}$, such that the data distribution is $P(\mathbf{Y}^s \neq \mathbf{Y}^t)$ and $P(\mathbf{Y}^s|\tilde{\mathbf{L}}^s) \approx P(\mathbf{Y}^t|\tilde{\mathbf{L}}^t)$, where $\ddot{\mathbf{Y}}\tilde{\mathbf{L}}^s$ and $\tilde{\mathbf{L}}^t$ are the class label set of samples in SD and TD, respectively. The main idea of the proposed model is shown in **Figure 2**.

Local Information Preserved in DASRC

Because the EEG signals are collected from different domains, domain discrepancies exist between SD and TD. Usually, directly adapting the existing classifier in SD may perform poorly to new samples in TD. Domain adaptation is adopted in this study to find a latent and domain-invariant subspace and maps $\mathbf{Y}^s$ and $\mathbf{Y}^t$ by projection matrixes $\mathbf{M}^s$ and $\mathbf{M}^t$, respectively. In this domain-invariant subspace, the discrepancy between SD and TD is reduced. Finally, we can train a new classifier for TD in the subspace with the help of discriminative knowledge from labeled samples in SD.

Domain adaptation method should strive to preserve the class distribution and local characteristics of training samples. Therefore, the local geometric structure of the samples is considered, so that the samples in SD and TD can be validly represented in the domain-invariant subspace. We construct the similarity matrixes $\mathbf{G}^s$ and $\mathbf{G}^t$ of SD and TD by k-nearest

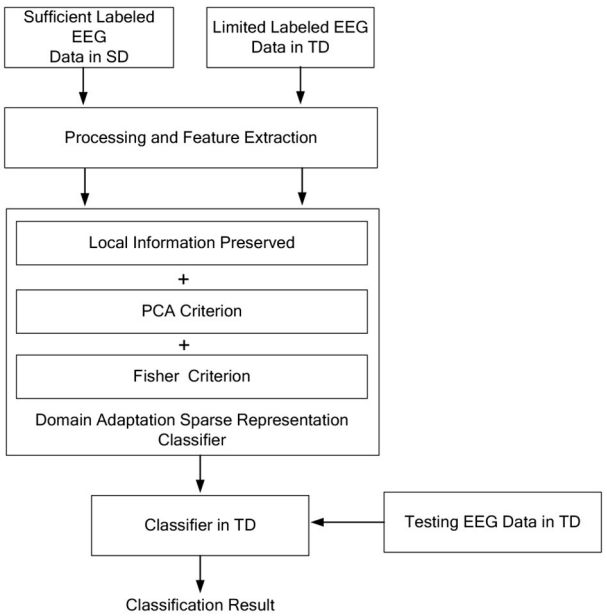


FIGURE 2 | The main idea of the proposed model.

neighbor graphs, respectively. The elements in $\mathbf{G}^s$ and $\mathbf{G}^t$ can be computed by the following equations,

$$G^s_{i,j} = \begin{cases} \exp(-\frac{\left\|\mathbf{y}^s_i - \mathbf{y}^s_j\right\|_2}{\sigma}) & if\ \mathbf{y}^s_i \in KNN(\mathbf{y}^s_j) \\ 0 & else \end{cases} \quad (3)$$

$$G^t_{i,j} = \begin{cases} 1 & if\ \mathbf{y}^t_i\ and\ \mathbf{y}^t_j\ are\ of the\ same\ class \\ 0 & else \end{cases} \quad (4)$$

where $KNN(\mathbf{y}^s_j)$ is a set that contains k nearest neighbor samples of $\mathbf{y}^s_j$. The element $G^s_{i,j}$ presents the similarity between the ith and jth samples in SD. As one of the most commonly used similarity metrics in graph learning, the Gaussian kernel is adopted in Equation (3). σ is the kernel parameter. Since the number of samples in TD is insufficient, Equation (4) ensures that the limited number of samples in TD is assigned with a given one-to-one weight. Therefore, we construct the local preserved constraint to maintain the intra-domain local information.

$$\min_{\mathbf{M}^s,\mathbf{M}^t} J_1 = \sum_{i,j} G^s_{i,j} \left\|(\mathbf{M}^s\mathbf{Y}^s_i - \mathbf{M}^s\mathbf{Y}^s_j)\right\|^2_2 + \sum_{i,j} G^t_{i,j} \left\|(\mathbf{M}^t\mathbf{Y}^t_i - \mathbf{M}^t\mathbf{Y}^t_j)\right\|^2_2. \quad (5)$$

PCA Criterion in DASRC

In addition, with the aim of classification, the discriminative knowledge of SD and TD should be enforced in subspace projection. We consider $\mathbf{M}^s$ and $\mathbf{M}^t$ are the bases of the subspaces based on the PCA criterion for SD and TD. Following Gong et al. (2019) and Ma et al. (2011), the PCA criterion is used to preserve discriminative knowledge in the subspace. To this end, we minimize the following optimization problem:

$$\min_{M^s,M^t} J_2 = -\mathbf{M}^s\mathbf{Y}^s(\mathbf{Y}^s)^T(\mathbf{M}^s)^T - \mathbf{M}^t\mathbf{Y}^t(\mathbf{Y}^t)^T(\mathbf{M}^t)^T,$$
$$s.t. \quad \mathbf{M}^s(\mathbf{M}^s)^T = \mathbf{I},$$
$$\mathbf{M}^t(\mathbf{M}^t)^T = \mathbf{I}. \quad (6)$$

where I is the identity matrix.

The Laplacian matrixes are denoted as $\mathbf{L}^s = \mathbf{G}^s - \mathbf{W}^s$ and $\mathbf{L}^t = \mathbf{G}^t - \mathbf{W}^t$, where $\mathbf{G}^s$ and $\mathbf{G}^t$ are diagonal matrixes whose elements in the principal diagonal are defined as $\mathbf{W}^s_{i,i} = \sum_j \mathbf{G}^s_{i,j}$ and $\mathbf{W}^t_{i,i} = \sum_j \mathbf{G}^t_{i,j}$, the term $J_1 + J_2$ can be written as

$$\min_{\mathbf{M}^s,\mathbf{M}^t} J_1 + J_2 = Tr(\mathbf{M}^s\mathbf{Y}^s(\mathbf{L}^s - \alpha\mathbf{I})(\mathbf{Y}^s)^T(\mathbf{M}^s)^T$$

TABLE 1 | Cross-subject accuracy % (std %) of DASRC using different features on SEED dataset.

	PSD	FD	DE	MAV
Subject 1	73.76	79.43	79.84	74.02
	(4.56)	(6.33)	(5.24)	(5.61)
Subject 2	74.54	80.01	80.36	74.89
	(5.63)	(4.24)	(5.23)	(5.74)
Subject 3	63.99	69.35	69.49	64.37
	(5.34)	(6.17)	(6.02)	(7.26)
Subject 4	83.71	89.31	89.97	84.28
	(6.49)	(7.11)	(6.13)	(5.18)
Subject 5	73.29	78.23	78.62	73.95
	(6.69)	(7.11)	(7.13)	(7.07)
Subject 6	86.47	90.32	90.83	86.99
	(7.15)	(6.98)	(7.31)	(6.86)
Subject 7	67.85	72.06	72.65	67.82
	(6.49)	(6.75)	(6.71)	(6.66)
Subject 8	75.02	70.26	70.98	75.42
	(7.04)	(6.48)	(7.01)	(6.87)
Subject 9	79.66	84.09	84.26	79.95
	(7.42)	(7.16)	(7.05)	(7.04)
Subject 10	70.91	76.35	76.94	71.16
	(6.89)	(7.12)	(7.30)	(6.99)
Subject 11	74.90	80.14	80.47	75.25
	(7.73)	(7.56)	(7.49)	(7.58)
Subject 12	75.73	80.25	80.41	75.97
	(7.29)	(7.33)	(7.48)	(7.80)
Subject 13	81.28	86.45	86.78	81.44
	(7.31)	(7.45)	(7.68)	(7.54)
Subject 14	75.12	80.11	80.51	75.43
	(6.43)	(6.37)	(6.39)	(6.97)
Subject 15	75.32	81.26	81.93	75.75
	(6.32)	(6.27)	(6.66)	(6.74)
Mean	75.44	79.84	80.27	75.78
	(6.57)	(6.73)	(6.73)	(6.80)

TABLE 2 | Cross-dataset accuracy % (std %) of DASRC using different features on SEED→ DEAP.

	PSD	FD	DE	MAV
SVM	33.34	35.75	36.66	33.94
	(8.04)	(9.00)	(8.36)	(9.10)
LC-KSVD	33.77	35.61	37.97	33.80
	(7.34)	(8.08)	(8.85)	(8.78)
TCA	41.69	44.12	45.08	41.90
	(7.27)	(9.35)	(9.01)	(9.70)
MIDA	43.61	45.13	45.86	42.96
	(9.50)	(10.81)	(8.52)	(6.65)
ASFM	44.67	47.14	47.97	44.88
	(8.96)	(9.04)	(10.48)	(8.66)
MMD	46.71	50.66	50.54	47.22
	(9.89)	(10.36)	(8.99)	(9.80)
DASRC	**50.57**	**53.26**	**53.54**	**50.60**
	(9.30)	(8.86)	(8.77)	(9.89)

*The bold values in **Tables 2**, **3** mean the best values in comparison experiments.*

$$
\begin{aligned}
&+Tr(\mathbf{M}^t\mathbf{Y}^t(\mathbf{L}^t-\alpha\mathbf{I})(\mathbf{Y}^t)^T(\mathbf{M}^t)^T,\\
&s.t.\ \mathbf{M}^s(\mathbf{M}^s)^T=\mathbf{I},\\
&\quad\mathbf{M}^t(\mathbf{M}^t)^T=\mathbf{I}.
\end{aligned}
\tag{7}
$$

where α is the regularization parameter.

Fisher Criterion in DASRC

To train the discriminative dictionary D in the projection space, where the learned dictionary can sparsely represent samples from each class in SD and TD, the representation errors of intraclass and interclass in SD and TD are required to be minimized and maximized, respectively. Inspired by the Fisher criterion (Peng et al., 2020), the ratio of intra-class scatter to inter-class scatter are minimized on the coding coefficients in SD and TD as follows:

$$
\begin{aligned}
&\arg\min_{\mathbf{M}^s,\mathbf{M}^t,\mathbf{D}} J_3=\frac{\sum_j\left\|\mathbf{M}^s\mathbf{y}_j^s-\mathbf{D}\varphi_{l_j^s}(\alpha_j^s)\right\|_F^2+\sum_j\left\|\mathbf{M}^t\mathbf{y}_j^t-\mathbf{D}\varphi_{l_j^t}(\alpha_j^t)\right\|_F^2}{\sum_j\left\|\mathbf{M}^s\mathbf{y}_j^s-\mathbf{D}\theta_{l_j^s}(\alpha_j^s)\right\|_F^2+\sum_j\left\|\mathbf{M}^t\mathbf{y}_j^t-\mathbf{D}\theta_{l_j^t}(\alpha_j^t)\right\|_F^2},\\
&s.t.\ \left\|d_k\right\|_2^2\le 1,\ \forall k
\end{aligned}
\tag{8}
$$

where the function $\varphi_{l_j}()$ returns the coefficient vectors of the same class of y_j, and the function $\theta_{l_j}()$ returns the coefficient vectors of the different class of y_j.

To simplify Equation (8), we denote

$$\Theta^s=[\varphi_{l_1^s}(\alpha_1^s),\varphi_{l_2^s}(\alpha_2^s),...,\varphi_{l_{n_s}^s}(\alpha_{n_s}^s)],\tag{9.1}$$

$$\Theta^t=[\varphi_{l_1^t}(\alpha_1^t),\varphi_{l_2^t}(\alpha_2^t),...,\varphi_{l_{n_t}^t}(\alpha_{n_t}^t)],\tag{9.2}$$

$$\Delta^s=[\theta_{l_1^s}(\alpha_1^s),\theta_{l_2^s}(\alpha_2^s),...,\theta_{l_{n_s}^s}(\alpha_{n_s}^s)],\tag{9.3}$$

$$\Delta^t=[\theta_{l_1^t}(\alpha_1^t),\theta_{l_2^t}(\alpha_2^t),...,\theta_{l_{n_t}^t}(\alpha_{n_t}^t)],\tag{9.4}$$

Equation (8) can be re-written as

$$
\begin{aligned}
&\arg\min_{\mathbf{M}^s,\mathbf{M}^t,\mathbf{D}} J_3=\frac{\left\|(\mathbf{M}^s\mathbf{Y}^s-\mathbf{D}\Theta^s)\right\|_F^2+\left\|(\mathbf{M}^t\mathbf{Y}^t-\mathbf{D}\Theta^t)\right\|_F^2}{\left\|(\mathbf{M}^s\mathbf{Y}^s-\mathbf{D}\Delta^s)\right\|_F^2+\left\|(\mathbf{M}^t\mathbf{Y}^t-\mathbf{D}\Delta^t)\right\|_F^2},\\
&s.t.\ \left\|d_k\right\|_2^2\le 1,\ \forall k
\end{aligned}
\tag{10}
$$

when the minimization problem of J_3 is solved, a shared dictionary is learned to establish an intrinsic relationship between different domains so that the discrimination information learned from SD can be transferred to TD in a cross-domain scenario.

The DASRC Model

For cross-domain EEG-based emotion classification, we take into account three objective functions J_1, J_2, and J_3 together, joint constraints on the local information preserved, PCA and Fisher criteria to optimize the shared dictionary and domain-specific projections. Thus, the optimization problem of DASRC can be formulated as

$$
\begin{aligned}
&\arg\min_{\mathbf{M}^s,\mathbf{M}^t,\mathbf{D},\mathbf{A}^s,\mathbf{A}^t}\ J_1+J_2+J_3,\\
&s.t.\ \mathbf{M}^s(\mathbf{M}^s)^T=\mathbf{I},\\
&\quad\mathbf{M}^t(\mathbf{M}^t)^T=\mathbf{I}.
\end{aligned}
\tag{11}
$$

To see all the components clearly, Equation (11) is expanded by

$$
\begin{aligned}
&\arg\min_{\mathbf{M}^s,\mathbf{M}^t,\mathbf{D}}\frac{\left\|(\mathbf{M}^s\mathbf{Y}^s-\mathbf{D}\Theta^s)\right\|_F^2+\left\|(\mathbf{M}^t\mathbf{Y}^t-\mathbf{D}\Theta^t)\right\|_F^2}{\left\|(\mathbf{M}^s\mathbf{Y}^s-\mathbf{D}\Delta^s)\right\|_F^2+\left\|(\mathbf{M}^t\mathbf{Y}^t-\mathbf{D}\Delta^t)\right\|_F^2}\\
&+Tr(\mathbf{M}^s\mathbf{Y}^s(\mathbf{L}^s-\alpha\mathbf{I})(\mathbf{Y}^s)^T(\mathbf{M}^s)^T\\
&+Tr(\mathbf{M}^t\mathbf{Y}^t(\mathbf{L}^t-\alpha\mathbf{I})(\mathbf{Y}^t)^T(\mathbf{M}^t)^T,\\
&s.t.\ \mathbf{M}^s(\mathbf{M}^s)^T=\mathbf{I},\\
&\quad\mathbf{M}^t(\mathbf{M}^t)^T=\mathbf{I}.
\end{aligned}
\tag{12}
$$

Let $\tilde{\mathbf{M}}=[\mathbf{M}^s,\mathbf{M}^t]$, $\tilde{\mathbf{Y}}=\begin{bmatrix}\mathbf{Y}^s & 0\\ 0 & \mathbf{Y}^t\end{bmatrix}$, $\tilde{\mathbf{\Theta}}=[\Theta^s,\Theta^t]$, $\tilde{\mathbf{\Delta}}=[\Delta^s,\Delta^t]$, $\tilde{\mathbf{L}}=\begin{bmatrix}\mathbf{L}^s & 0\\ 0 & \mathbf{L}^t\end{bmatrix}$, the simplified formulation of Equation (12) can be approximately written as

$$
\begin{aligned}
&\min_{\tilde{\mathbf{M}},\mathbf{D}}\frac{\left\|\tilde{\mathbf{M}}\tilde{\mathbf{Y}}-\mathbf{D}\tilde{\mathbf{\Theta}}\right\|_F^2}{\left\|\tilde{\mathbf{M}}\tilde{\mathbf{Y}}-\mathbf{D}\tilde{\mathbf{\Delta}}\right\|_F^2}+Tr(\tilde{\mathbf{M}}\tilde{\mathbf{M}}(\mathbf{L}-\alpha\mathbf{I})\tilde{\mathbf{Y}}^T\tilde{\mathbf{M}}^T,\\
&s.t.\ \tilde{\mathbf{M}}\tilde{\mathbf{M}}^T=\mathbf{I}.
\end{aligned}
\tag{13}
$$

Optimization of the Objective Function

Equation (13) becomes a least square problem with quadratic constraints and could be solved by many methods. The Lagrangian of the optimization problem in Equation (13) is

$$
\begin{aligned}
F(\tilde{\mathbf{M}},\mathbf{D},\beta,\gamma)=&\left\|\tilde{\mathbf{M}}\tilde{\mathbf{Y}}-\mathbf{D}\tilde{\mathbf{\Theta}}\right\|_F^2+Tr(\tilde{\mathbf{M}}\tilde{\mathbf{Y}}(\mathbf{L}-\alpha\mathbf{I})\tilde{\mathbf{Y}}^T\tilde{\mathbf{M}}^T\\
&-\beta\left\|\tilde{\mathbf{M}}\tilde{\mathbf{Y}}-\mathbf{D}\tilde{\mathbf{\Delta}}\right\|_F^2-\gamma(\tilde{\mathbf{M}}\tilde{\mathbf{M}}^T-\mathbf{I}),
\end{aligned}
\tag{14}
$$

where β and γ are Lagrange multipliers.

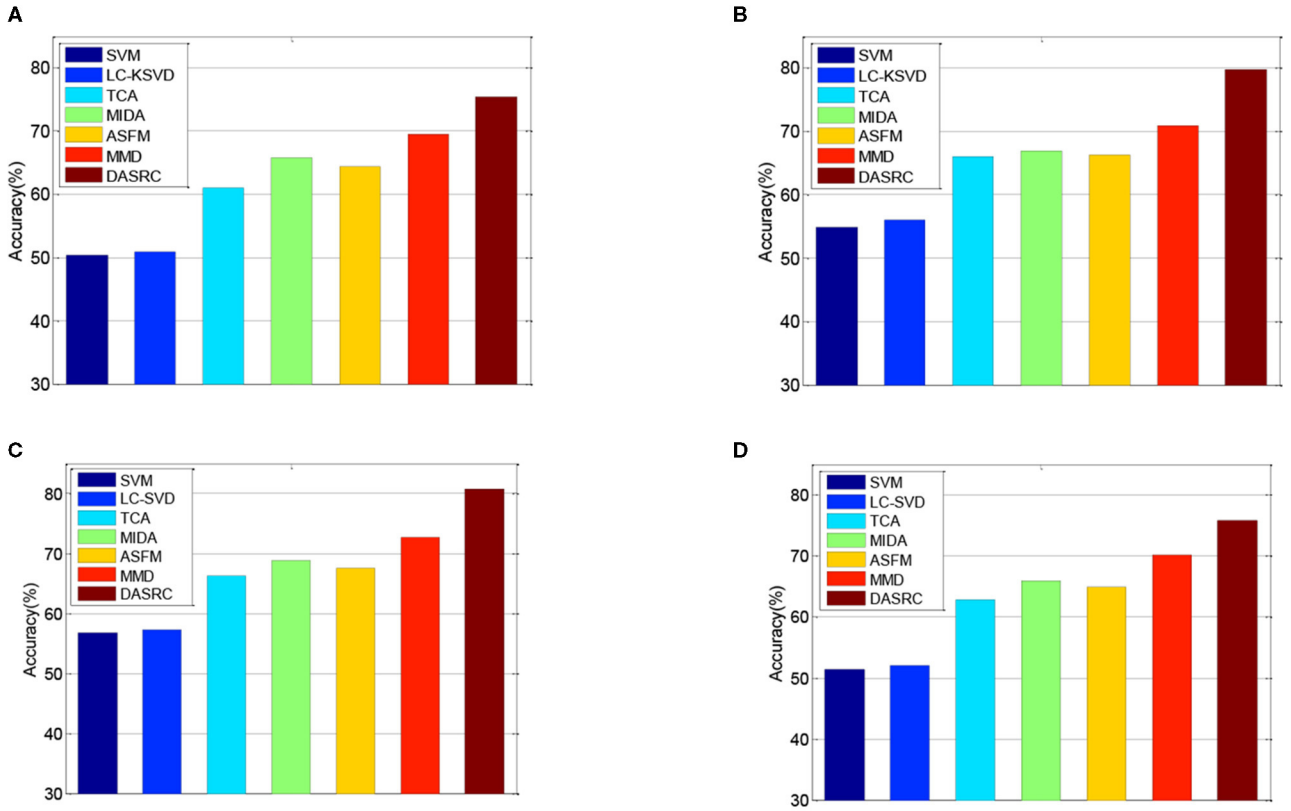


FIGURE 3 | Cross-subject accuracy on SEED dataset with different features, **(A)** PSD, **(B)** FD, **(C)** DE, **(D)** MAV.

We use the alternating optimization method to solve Equation (14). When fixing $\mathbf{D}$, β, and γ, we take the first-order partial derivatives of Equation (14) over $\tilde{\mathbf{M}}$,

$$F'(\tilde{\mathbf{M}}) = \tilde{\mathbf{M}}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T - \mathbf{D}\tilde{\mathbf{\Theta}}\tilde{\mathbf{Y}}^T + \tilde{\mathbf{M}}\tilde{\mathbf{Y}}(\mathbf{L} - \alpha\mathbf{I})\tilde{\mathbf{Y}}^T - \beta(\tilde{\mathbf{M}}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T - \mathbf{D}\tilde{\mathbf{\Delta}}\tilde{\mathbf{Y}}^T) - \gamma\tilde{\mathbf{M}}, \quad (15)$$

and the optimal $\tilde{\mathbf{M}}$ can be computed in the closed-form

$$\tilde{\mathbf{M}} = (\mathbf{D}\tilde{\mathbf{\Theta}}\tilde{\mathbf{Y}}^T - \beta\mathbf{D}\tilde{\mathbf{\Delta}}\tilde{\mathbf{Y}}^T)(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T + \tilde{\mathbf{Y}}(\mathbf{L} - \alpha\mathbf{I})\tilde{\mathbf{Y}}^T - \beta\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T - \gamma\mathbf{I})^{-1}. \quad (16)$$

When fixing $\tilde{\mathbf{M}}$, β, and γ, we take the first-order partial derivative of Equation (14) over $\mathbf{D}$,

$$F'(\mathbf{D}) = -\tilde{\mathbf{M}}\tilde{\mathbf{Y}}\tilde{\mathbf{\Theta}}^T + \mathbf{D}\tilde{\mathbf{\Theta}}\tilde{\mathbf{\Theta}}^T + \beta\tilde{\mathbf{M}}\tilde{\mathbf{Y}}\tilde{\mathbf{\Delta}}^T - \beta\mathbf{D}\tilde{\mathbf{\Delta}}\tilde{\mathbf{\Delta}}^T, \quad (17)$$

and the optimal $\mathbf{D}$ can be computed in the closed-form

$$\mathbf{D} = (\beta\tilde{\mathbf{M}}\tilde{\mathbf{Y}}\tilde{\mathbf{\Delta}}^T - \tilde{\mathbf{M}}\tilde{\mathbf{Y}}\tilde{\mathbf{\Theta}}^T)(\tilde{\mathbf{\Theta}}\tilde{\mathbf{\Theta}}^T - \beta\tilde{\mathbf{\Delta}}\tilde{\mathbf{\Delta}}^T)^{-1}. \quad (18)$$

When fixing $\tilde{\mathbf{M}}$, $\mathbf{D}$, and γ, we take the first-order partial derivative of Equation (14) over β,

$$F'(\beta) = -\left\|\tilde{\mathbf{M}}\tilde{\mathbf{Y}} - \mathbf{D}\tilde{\mathbf{\Delta}}\right\|_F^2, \quad (19)$$

Then β can be optimized by

$$\beta_{new} = \beta + \lambda_\beta \frac{F(\beta)}{F'(\beta)}, \quad (20)$$

where λ_β is the length size.

When fixing $\tilde{\mathbf{M}}$, $\mathbf{D}$, and β, we take the first-order partial derivative of Equation (14) over γ,

$$F'(\gamma) = \mathbf{I} - \tilde{\mathbf{M}}\tilde{\mathbf{M}}^T, \quad (21)$$

Then γ can be optimized by,

$$\gamma_{new} = \gamma + \lambda_\gamma \frac{F(\gamma)}{F'(\gamma)}, \quad (22)$$

where λ_γ is the length size.

The proposed DASRC model is given in **Algorithm 1**.

Testing

With all the optimization steps discussed above, we summarize the optimization procedure of the DASRC model in Algorithm 1. When the projection matrix $\tilde{\mathbf{M}}$ and the shared dictionary D are obtained by Algorithm 1, we use the following step to recognize the new EEG signal y in TD. The sparse coefficient vector α over dictionary D can be solved as,

$$\min_{\alpha} \left\|\tilde{\mathbf{M}}y - \mathbf{D}\alpha\right\|_2^2 + \mu\left\|\alpha\right\|_2^2, \quad (23)$$

The sparse coefficient vector α can be obtained as

$$\alpha = (\mathbf{D}^T\mathbf{D} + \mu\mathbf{I})^{-1}\mathbf{D}^T\tilde{\mathbf{M}}y, \tag{24}$$

The classification label of x can be derived as

$$\text{label}(\mathbf{y}) = \arg\min_i \{\left\|\tilde{\mathbf{M}}y - \mathbf{D}\varphi_i(\alpha)\right\|_2^2 / \left\|\varphi_i(\alpha)\right\|_2^2\}, i = 1, 2, ..., C \tag{25}$$

where C is the number of classes.

Algorithm 1 | DASRC: domain adaptation sparse representation classifier.

Input: Source domain dataset $\mathbf{Y}^s$and target domain $\mathbf{Y}^t$;
Output: The dictionary $\mathbf{D}$ and subspace $\tilde{\mathbf{M}}$;
1: Initialize $\mathbf{M}^s$ and $\mathbf{M}^t$ using PCA method and $\mathbf{D}$ using KSVD algorithm;
2: Construct k-nearest neighbor graphs in SD and TD;
3: Construct matrixes $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{L}}$;
While
  4: Construct matrixes $\tilde{\mathbf{M}}$,$\tilde{\mathbf{\Theta}}$, and $\tilde{\mathbf{\Delta}}$;
  5. Update $\tilde{\mathbf{M}}$ using Equation (16);
  6: Update $\mathbf{D}$ using Equation (18);
  7: Update β using Equations (19)–(20);
  8: Update γ using Equations (21)–(22);
Return $\tilde{\mathbf{M}}$ and D.

EXPERIMENT

Datasets and Experimental Settings

We conduct the experiments to evaluate the efficacy of the proposed classifier on two public EEG emotion datasets, SEED and DEAP. The EEG signals in the SEED dataset are recorded from 15 participants across three different sessions. Their emotions are stimulated by the Chinese file clips using an ESI NeuroScan system with 62-channel electrodes at a sampling rate of 1,000 Hz. Each film clip is related to three emotions as positive, neutral, or negative, and each emotion has five corresponding film clips. The EEG signals in the DEAP dataset are recorded from 32 participants by watching 40 videos with 32-channel electrodes. DEAP dataset labels the valence and arousal rating scores from 1 to 9, which is closely related to emotions. We manually label the valence values above 4.5 as positive and the values smaller than 4.5 as negative. For a comprehensive study, we extract four different features in terms of time analysis, frequency analysis, and non-linear analysis for each EEG channel. Time analysis includes mean absolute value (MAV) (Shim et al., 2016), frequency analysis includes power spectral density (PSD) (Jenke et al., 2014), and non-linear analysis includes fractal dimension (FD) (Li et al., 2019) and differential entropy (DE) (Zheng and Lu, 2015). The size of the extracted feature is 10 dimensions for each channel.

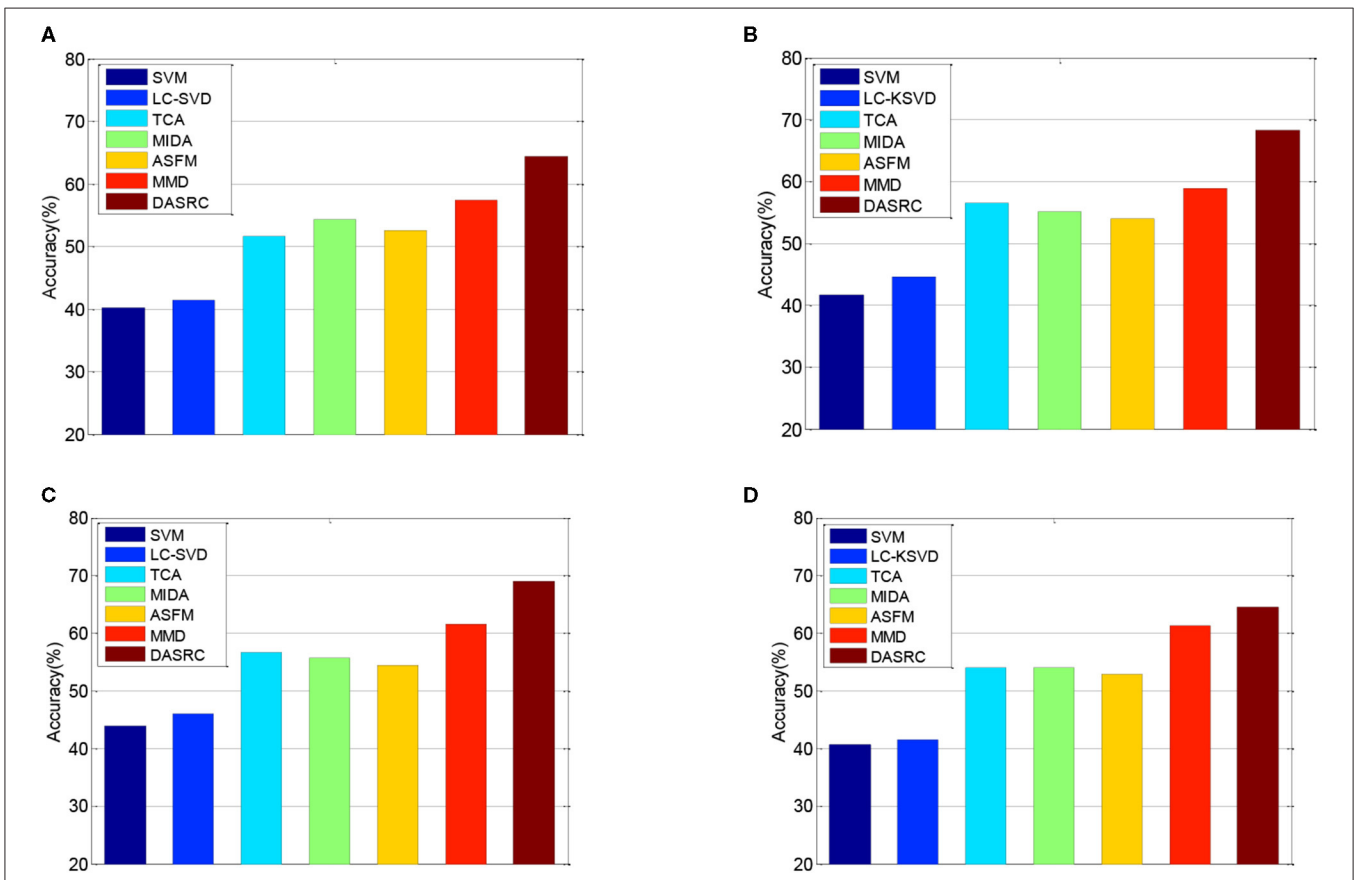


FIGURE 4 | Cross-subject accuracy on DEAP dataset with different features, **(A)** PSD, **(B)** FD, **(C)** DE, **(D)** MAV.

We evaluate the DASRC model on cross-subject and cross-dataset scenarios. We compared DASRC with two baseline methods and four domain adaptation methods. For the baseline methods, we compared the DASRC with label consistent K-SVD (LC-KSVD) (Jiang et al., 2013) and SVM (Cortes and Vapnik, 1995). For these two methods, the training data in SD and TD are combined as the input samples. The Gaussian kernel is used in SVM, and the kernel and penalty parameters are searched in the grid $\{10^{-3}, 10^{-2}, \ldots, 10^{3}\}$. The number of atoms in each class is selected in {50, 60,…, 200}. We compared the DASRC model with four domain adaptation methods, including transfer component analysis (TCA) (Pan et al., 2011), adaptive subspace feature matching (ASFM) (Chai et al., 2017), maximum mean discrepancy (MMD) (Sejdinovic et al., 2013), maximum independence domain adaptation (MIDA) (Yan et al., 2018). The latent dimension in MIDA and TCA is determined by searching the grid {20, 30,…, 100}. The subspace dimension in ASFM is set as 70. The threshold parameter in ASFM was set at 0.45. In DASRC, the subspace dimension is determined by searching the grid {20, 30,…, 100}. The number of atoms in each class is selected in {10, 15, 20, 25, 30, 35}. All the algorithms are implemented in MATLAB.

Cross-Subject EEG-Based Emotion Classification

For cross-subject evaluation, one subject is left out as the test subject, and the remaining different subjects are used as training data to feed the model. In the SEED dataset, one subject contains 925 samples in each session. We randomly select 300 samples from each subject and combine them as training data. We repeat the procedure 10 times.

First, we evaluate how the classification accuracy of the classifier varies with the different features. For each subject, **Table 1** depicts the experimental results of mean accuracies and SD on the SEED dataset. We can see that for DASRC, the classification accuracy is relatively stable, and the value of SD is acceptable. Meanwhile, the classification accuracy of some subjects is relatively high and of some subjects is relatively low. DE and FD features achieve better performance than PSD and MAV features.

Then, DASRC is compared with two baseline methods and four domain adaptation methods. **Table 2** depicts the experimental results of mean accuracies of all models using PSD, MAV, DE, and FD features. From **Table 2**, we can see that the classification accuracies of all methods on DE and FD features are higher than PSD and MAV features. It may suggest that non-linear analysis features may be more suitable when compared to EEG-based emotion classification. The performance of the DE feature is better than that of the FD feature, and the best results in all methods are obtained using the DE feature.

In the DEAP dataset, one subject contains 180 samples. As such, we randomly select 100 samples from each subject and the training set contains 3,100 samples. **Figure 3** shows the classification accuracies of six comparison methods on the SEED dataset when four different kinds of features (PSD, MAV, DE, and FD) are used. **Figure 4** shows the accuracy results of five comparison methods on the DEAP dataset for positive and negative classes. According to the experimental results, we can see that first, single-domain classification methods SVM and LC-KSVD cannot obtain satisfactory classification performance in subject-to-subject scenarios on SEED and DEAP datasets. After all, they are not proposed to address the cross-domain data scenarios. Second, among the domain adaptation methods, the proposed DASRC model based on shared dictionary and subspace learning perform better than the methods using some other shared components. The main factor is that the shared dictionary can learn more discriminative knowledge to encode the EEG signals. Third, the accuracies of all models on the DEAP dataset are lower than those obtained on the SEED dataset. It may be the reason that the labeling quality of EEG signals in DEAP is poorer.

TABLE 3 | Cross-dataset accuracy % (std %) of DASRC using different features on DEAP→ SEED.

	PSD	FD	DE	MAV
SVM	48.87	51.05	51.66	48.94
	(9.00)	(9.43)	(8.57)	(8.08)
LC-KSVD	49.30	51.73	51.97	49.86
	(9.76)	(10.72)	(7.70)	(6.68)
TCA	56.08	58.57	58.32	56.55
	(8.02)	(8.83)	(9.11)	(7.06)
MIDA	57.96	59.21	59.99	58.31
	(8.41)	(8.64)	(7.15)	(9.61)
ASFM	58.77	61.74	62.01	59.76
	(9.68)	(7.45)	(7.40)	(8.05)
MMD	61.04	63.60	63.89	61.25
	(8.82)	(7.00)	(8.64)	(7.24)
DASRC	**62.75**	**64.74**	**64.97**	**62.71**
	(8.20)	(7.33)	(7.67)	(7.05)

Cross-Dataset EEG-Based Emotion Classification

For cross-dataset evaluation, the SD and TD are from different datasets. We perform comparison experiments across SEED and DEAP datasets. According to the given exact threshold, the samples in the DEAP dataset are divided into positive and negative classes, which correspond to the positive and negative classes of the SEED dataset. As the domain adaptation method requires the same feature space in SD and TD, we use the 32 channels shared between DEAP and SEED. In two experiments, we randomly select 3,000 samples in SD for training and 2,000 samples in TD for testing. We, then, repeat the procedure 10 times. **Tables 2**, **3**, respectively, show the accuracy results of five comparison methods on SEED→ DEAP and DEAP→ SEED, when PSD, MAV, DE, and FD features are used. According to the experimental results, we can see that first, the performance of single-domain and domain adaptation methods exhibit

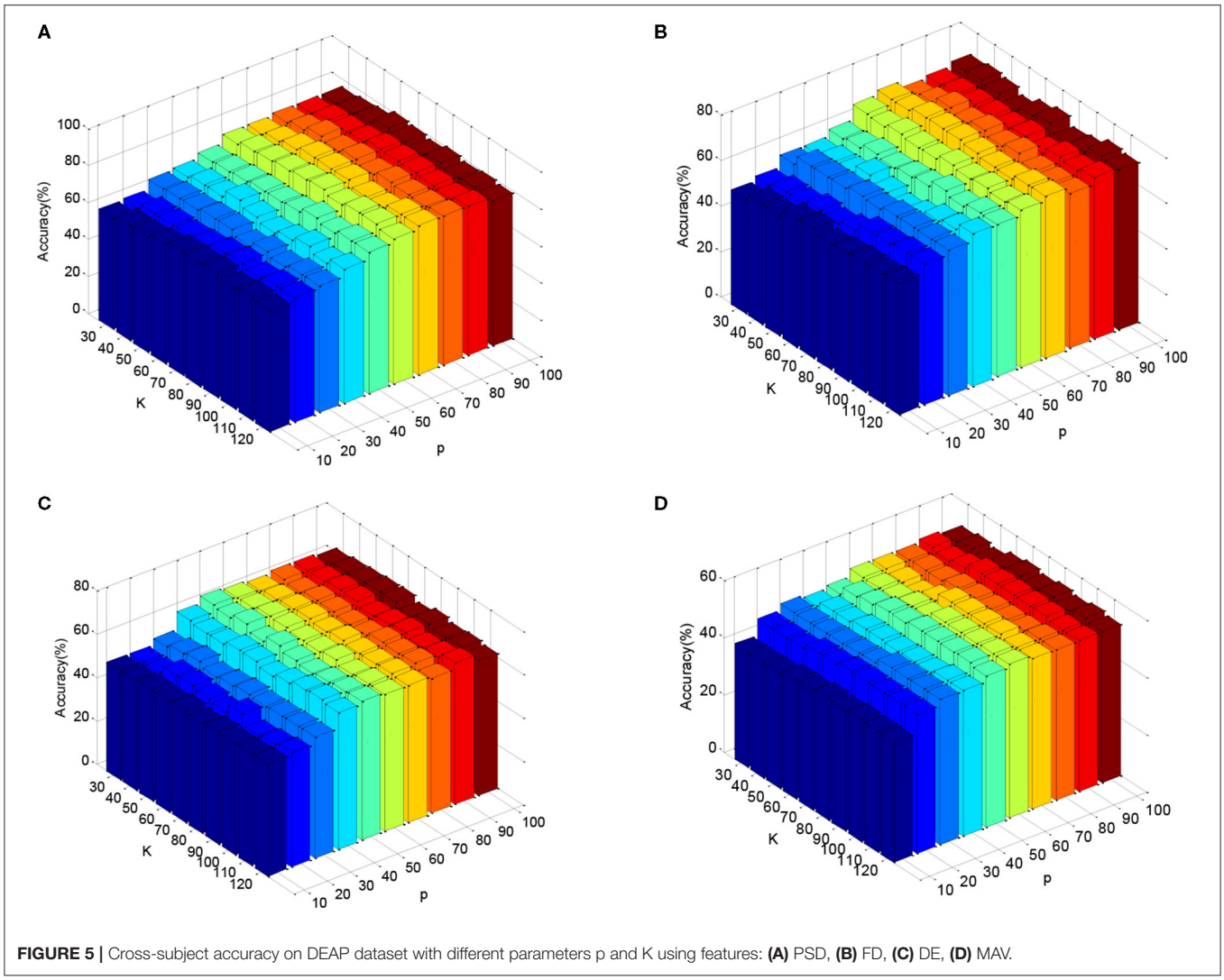


FIGURE 5 | Cross-subject accuracy on DEAP dataset with different parameters p and K using features: **(A)** PSD, **(B)** FD, **(C)** DE, **(D)** MAV.

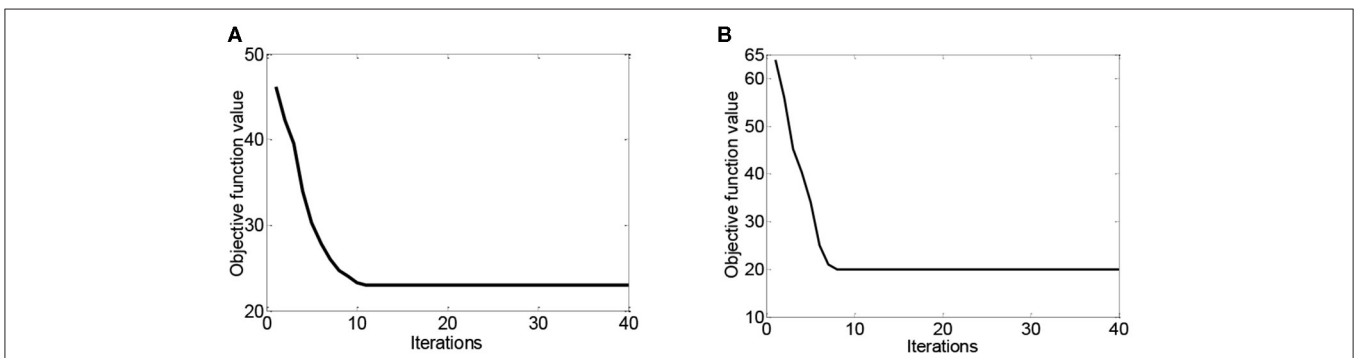


FIGURE 6 | The convergence curves of DASRC on two scenarios, **(A)** cross-subject on SEED, **(B)** cross-dataset on DEAP→ SEED.

evident differences, with three domain adaptation methods show significant improvements in classification accuracy. Second, DASRC is the best-performing classifier in two cross-dataset EEG emotion classifications. DASRC achieves 16.88% accuracy gains over the compared single-domain methods and 8.70% accuracy gains over the compared domain adaptation methods. The reason is that single-domain methods fail to reduce the domain shift. On the contrary, DASRC learns the shared

dictionary to build the connection between SD and TD such that the discriminative knowledge from SD can be transferred to TD. In addition, besides the advantage of the cross-domain Fisher criterion and local information preserved technology, the recognition ability also promotes the classification performance of DASRC.

Parameter Analysis

In this subsection, we validate the DASRC in cross-subject and cross-dataset scenarios. The subspace dimensions, p, and size of dictionary atoms, K, are the key parameters, and they are determined by the cross-validation method. We empirically search p in {10, 20,. . . , 100} and K in {30, 40,. . . , 120}. **Figure 5** plots the mean accuracy of DASRC with varying p and K, while using DE features. From **Figure 5**, we can see that DASRC can achieve stable performance with small p and K. The best accuracies are achieved, in general, when p is great than 60 and K is great than 80. This result indicates that the proposed DASRC can exploit the common knowledge in a relatively low dimensional subspace. Based on the results in **Figure 5**, the subspace dimension and dictionary size are suggested to be set to 60 and 80, respectively.

The domain adaptation methods often produce extra computational overhead. **Figure 6** plots the convergence curves of DASRC in cross-subject and cross-dataset scenarios while using DE features. From this figure, we can see that DASRC can converge within a small number of iterations. Thus, we can set the iteration bound to 40.

CONCLUSION

In this study, the DASRC model is proposed, which solves the EEG-based emotion classification across different subjects and datasets. Three criteria are considered to jointly learn sunspace and shared dictionary in DASRC. The local information preserved criterion is exploited to project samples in SD and TD into the shared subspace, where both PCA and Fisher criteria are exploited to transform discriminative knowledge through the shared dictionary. Experimental testing using SEED and DEAP datasets demonstrates the effectiveness of DASRC for dealing with the domain discrepancy for EEG-based emotion classification. For future work, we will explore more local preserved strategies in domain adaptation dictionary learning, such as local salience information. In addition, we will study the semi-supervised domain adaptation scenario, in which the unlabeled samples in TD rather than limited labeled samples participate in the model training. How to prevent negative transfer will also be considered in the next stage of work.

DATA AVAILABILITY STATEMENT

Publicly available datasets SEED and DEAP were analyzed in this paper. These data can be found in the following links, respectively: https://bcmi.sjtu.edu.cn/home/seed/ and http://www.eecs.qmul.ac.uk/mmv/datasets/deap/.

AUTHOR CONTRIBUTIONS

TN, JX, and SW conceived and designed the proposed model. YN and JX performed the experiment. TN and SW wrote the manuscript. All authors read and approved the manuscript.

FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grants 61806026, the research project of maternal and child health in Jiangsu Province under Grants F202060, and the Natural Science Foundation of Jiangsu Province under Grants SBK2021022110.

REFERENCES

Aharon, M., Elad, M., and Bruckstein, A. (2006). K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* 54, 4311–4322. doi: 10.1109/TSP.2006.881199

Chai, X., Wang, Q., Zhao, Y., Li, Y., Wang, Q., Liu, X., et al. (2017). A fast, efficient domain adaptation technique for cross-domain electroencephalography (EEG)-based emotion recognition. *Sensors* 17:1014. doi: 10.3390/s17051014

Cimtay, Y., and Ekmekcioglu, E. (2020). Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition. *Sensors* 20:2034. doi: 10.3390/s20072034

Cortes, C., and Vapnik, V. (1995). Support vector machine. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018

Doma, V., and Pirouz, M. (2020). A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals. *J. Big Data* 7:18. doi: 10.1186/s40537-020-00289-7

Fahimi, F., Zhang, Z., Goh, W. B., Lee, T. S., Ang, K. K., and Guan, C. (2019). Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI. *J. Neural Eng.* 16:026007. doi: 10.1088/1741-2552/aaf3f6

Gong, C., Zhou, P., Han, J., and Xu, D. (2019). Learning rotation-invariant and Fisher discriminative convolutional neural networks for object detection. *IEEE Trans. Image Process.* 28, 265–278. doi: 10.1109/TIP.2018.2867198

Gu, X., Zhang, C., and Ni, T. (2020). A hierarchical discriminative sparse representation classifier for EEG signal detection. *IEEE/ACM Trans. Comput. Biol. Bioinform.* doi: 10.1109/TCBB.2020.3006699. [Epub ahead of print].

Jenke, R., Peer, A., and Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* 5, 327–339. doi: 10.1109/TAFFC.2014.2339834

Jiang, Z., Lin, Z., and Davis, L. (2013). Label consistent K-SVD: learning a discriminative dictionary for recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 2651–2664. doi: 10.1109/TPAMI.2013.88

Kanoga, S., Kanemura, A., and Asoh, H. (2019). Multi-scale dictionary learning for ocular artifact reduction from single-channel electroencephalograms. *Neurocomputing* 347, 240–250. doi: 10.1016/j.neucom.2019.02.060

Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., et al. (2011). Deap: a database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* 3, 18–31. doi: 10.1109/T-AFFC.2011.15

Lan, Z., Olga, S., Wang, L., Reinhold, S., and Muller-Putz, G. R. (2018). Domain adaptation techniques for EEG-based emotion recognition: a comparative study on two public datasets. *IEEE Trans. Cogn. Dev. Syst.* 11, 85–94. doi: 10.1109/TCDS.2018.2826840

Li, J., Qiu, S., Du, C., Wang, Y., and He, H. (2020b). Domain adaptation for EEG emotion recognition based on latent representation similarity. *IEEE Trans. Cogn. Dev. Syst.* 12, 344–353. doi: 10.1109/TCDS.2019.2949306

Li, J., Qiu, S., Shen,Y. Y., Liu, C. L., and He, H. (2020a). Multisource transfer learning for cross-subject EEG emotion recognition. *IEEE Trans. Cybern.* 50, 3281–3293. doi: 10.1109/TCYB.2019.2904052

Li, X., Song, D., Zhang, P., Zhang, Y., Hou, Y., and Hu, B. (2018). Exploring EEG features in cross-subject emotion recognition. *Front. Neurosci.* 12:162. doi: 10.3389/fnins.2018.00162

Li, Y., Zheng, W., Cui, Z., Zong, Y., and Ge, S. (2019). EEG emotion recognition based on graph regularized sparse linear regression. *Neural Process. Lett.* 49, 555–571. doi: 10.1007/s11063-018-9829-1

Liu, M. J., Zhou, M. M., Zhang, T., and Xiong, N. X. (2020). Semi-supervised learning quantization algorithm with deep features for motor imagery EEG recognition in smart healthcare application. *Appl. Soft Comput.* 89:106071. doi: 10.1016/j.asoc.2020.106071

Liu, Y. J., Yu, M. J., Zhao, G. Z., Song, J., Ge, Y., and Shi, Y. C. (2018). Real-time movie-induced discrete emotion recognition from EEG signals. *IEEE Trans. Affect. Comput.* 9, 550–562. doi: 10.1109/TAFFC.2017.2660485

Ma, J., Bayram, S., Tao, P., and Svetnik, V. (2011). High-throughput ocular artifact reduction in multichannel electroencephalography (EEG) using component subspace projection. *J. Neurosci. Methods* 196, 131–140. doi: 10.1016/j.jneumeth.2011.01.007

Morioka, H., Kanemura, A., Hirayama, J. I., Shikauchi, M., Ogawa, T., Ikeda, S., et al. (2015). Learning a common dictionary for subject-transfer decoding with resting calibration. *NeuroImage* 111, 167–178. doi: 10.1016/j.neuroimage.2015.02.015

Ni, T., Gu, X., and Jiang, Y. (2020c). Transfer discriminative dictionary learning with label consistency for classification of EEG signals of epilepsy. *J. Ambient Intell. Humaniz. Comput.* doi: 10.1007/s12652-020-02620-9. [Epub ahead of print].

Ni, T., Gu, X., and Zhang, C. (2020b). An intelligence EEG signal recognition method via noise insensitive TSK fuzzy system based on interclass competitive learning. *Front. Neurosci.* 14: 837. doi: 10.3389/fnins.2020.00837

Ni, T., Zhang, C., and Gu, X. (2020a). Transfer model collaborating metric learning and dictionary learning for cross-domain facial expression recognition. *IEEE Trans. Comput. Soc. Syst.* doi: 10.1109/TCSS.2020.3013938. [Epub ahead of print].

Pan, S. J., Tsang, I. W., Kwok, J. T., and Yang, Q. (2011). Domain adaptation via transfer component analysis. *IEEE Trans. Neural Netw.* 22, 199–210. doi: 10.1109/TNN.2010.2091281

Peng, Y., Liu, S., Wang, X., and Wu, X. (2020). Joint local constraint and fisher discrimination based dictionary learning for image classification. *Neurocomputing* 398, 505–519. doi: 10.1016/j.neucom.2019.05.103

Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Stat.* 41, 2263–2291. doi: 10.1214/13-AOS1140

Shim, H. M., An, H., Lee, S., Lee, E. H., Min, H. K., and Lee, S. (2016). EMG pattern classification by split and merge deep belief network. *Symmetry* 8:148. doi: 10.3390/sym8120148

Yan, K., Kou, L., and Zhang, D. (2018). Learning domain-invariant subspace using domain features and independence maximization. *IEEE Trans. Cybern.* 48, 288–299. doi: 10.1109/TCYB.2016.2633306

Yang, F., Zhao, X., Jiang, W., Gao, P., and Liu, G. (2019). Multi-method fusion of cross-subject emotion recognition based on high-dimensional EEG features. *Front. Comput. Neurosci.* 13:53. doi: 10.3389/fncom.2019.00053

Zhang, X., and Wu, D. (2019). On the vulnerability of CNN classifiers in EEG-based BCIs. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 814–825. doi: 10.1109/TNSRE.2019.2908955

Zheng, W. L., and Lu, B. L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Ment. Dev.* 7, 162–175. doi: 10.1109/TAMD.2015.2431497

Zheng, W. L., Zhu, J. Y., and Lu, B. L. (2019). Identifying stable patterns over time for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* 10, 417–429. doi: 10.1109/TAFFC.2017.2712143

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Ni, Ni, Xue and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 05 August 2021
doi: 10.3389/fpsyg.2021.727788

Emotion Analysis of Personalized Color of Beauty Package Based on NCD Image Positioning

Feng Wang[1], *Jifeng Xu*[1], *Hanning Zhang*[2] *and Jun Yin*[1]*

[1] School of Design, Jiangnan University, Wuxi, China, [2] School of Art, Southeast University, Nanjing, China

Purpose: Based on the Nippon Color & Design (NCD) system, this study aims to accurately discover the color preferences and image positioning of different female groups in China and, thus, to establish a color reference system, which is suitable for the packaging of beauty products.

Methods: We first selected middle-aged women in the rural area of northern China as the studied group. Then, we conducted an extensive questionnaire survey on their color preferences focusing on their psychological characteristics and perception rules of emotional needs for colors. After that, we extracted colors from samples and classified them using the NCD system research method. Finally, we conducted a systematic analysis and verification to determine the color emotional preference of this group.

Conclusion: We investigate the color preference of targeted customers to establish the image space of beauty product packaging. By applying SPSS for the factor analysis, we study the rural women in northern China, who are less concerned about fashion. We focus on the accurate positioning between the interaction of color, women, and environment throughout the color design process. We have conducted a series of comparative analyses of popular beauty products among the rural women. The results can accurately reflect the brand definition and positioning of color semantics and, thus, provide invaluable information for future beauty packaging designs and marketing promotion strategy.

Keywords: NCD image positioning, color emotion analysis, beauty package, color system, color semantics

OPEN ACCESS

Edited by:
Yaoru Sun,
Tongji University, China

Reviewed by:
Yanhui Zhang,
Hebei University of Chinese Medicine, China
Hongru Zhao,
Soochow University, China

***Correspondence:**
Jun Yin
junyin2021@126.com

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *19 June 2021*
Accepted: *05 July 2021*
Published: *05 August 2021*

Citation:
Wang F, Xu J, Zhang H and Yin J (2021) Emotion Analysis of Personalized Color of Beauty Package Based on NCD Image Positioning. Front. Psychol. 12:727788. doi: 10.3389/fpsyg.2021.727788

INTRODUCTION

Due to the rapid development of the economy of China in the past two decades, the per capita disposable income of Chinese residents has been growing swiftly, which has led to a substantial increase in the beauty product market. According to the report of China's National Bureau of Statistics in 2020, the sales in the cosmetics market of China have increased from 204.9 billion yuan in 2015 to 340 billion yuan in 2020. Currently, China is the second largest cosmetics consumption country in the world after the United States. In order to compete in the fast-growing market, many researchers start to study the purchasing behaviors of customers and their psychological states. How to lure customers to purchase their products becomes their top priority. Hence, emotional designs have become the primary concern in the design industry.

According to the survey shown in *A study about package design and color elements of digital products* by Koh, Seon, the colors of the product packaging design have a great impact on the purchase desires of consumers; 78% of the survey participants said the color would affect their purchase desires. The outcome of the survey shows that the color of the product packaging design is essential to sales. As the primary element in visual impression, the color of packaging design directly affects the physiological response and psychological cognition of people. Relevant theoretical foundations include, for example, *A Study of Color Emotion and Color Preference*. In this study, three color preference models for single colors were developed (Ou et al., 2004), *Atypical Color Preference in Children with Autism Spectrum Disorder* (Marine and Nobuo, 2016), *Color Preference and Food Choice among Children* (Walsh et al., 1990), and *Relations between Their Color Preference and Colors of Their Clothes Color Preference of Little Children* (Haruko and Emiko, 1972).

Color emotion is a psychological reaction that associates with the long-term understanding of people of objective things and their related colors, and it manifests itself as the emotional expression of color and the emotional experience brought by color. According to the survey shown in *Color Preferences in Adolescence and Early Adulthood and Their Relationship with Emotional Intelligence* by Bolshunova and Khromova (2021), the article is to identify the age specificity of color preferences and their connection with some personality traits presented in the parameters of emotional intelligence in adolescence and early adulthood. In *Seasonal Variations in Color Preference* by Schloss et al. (2017), empirical support for this distinction shows that pair preference and harmony both increase as hue similarity increases, but preference relies more strongly on component color preference and lightness contrast (Schloss and Palmer, 2011).

Color itself can express certain moods and emotions. There are many factors affecting color psychology. The psychological association and emotional representation of color are multifaceted and complex. First, color perception is closely related to age, occupation, and gender. In *Comparative Studies on Color Preference in Japan and Other Asian Regions, with Special Emphasis on the Preference for White* (Saito, 1996), factor analysis and cluster analysis indicated some relationship between color preference and the lifestyles of subjects. Besides the factors of age and sex, associative images based on environmental and cultural aspects may be an important factor influencing color preference. Second, the application of color in different situations will bring people different emotional experiences. Finally, the symbolic meanings of colors in different countries, regions, and nations are also different. When designing products, we must fully understand the influencing factors of color psychology in order to correctly grasp the application rules of color emotion in the design field. We have to carry out color collection, summarization, and symbolic expression, establish the expression of color intentions, and incorporate them into the color design process.

Of the 1.4 billion people in China, 40.42% are farmers, and female employees account for 43.7% of total employees (Sin et al., 2001). Northern China is a traditional agricultural province, with Qinling Mountains and Huaihe River as the dividing line between the northern and southern regions. Northern women are deeply influenced by traditional culture and customs, especially rural women who are the typical representative of traditional Chinese women. Judging from the composite consumption index of China in recent years, the female consumer market in China occupies an increasingly important position (Tam and Tai, 1998).

The research objective of this study is aimed at the minority group of Chinese women who are ignored by the present designers and the market. They are traditional conservative women with introverted personality and low educational level in the northern rural areas. In this study, the research on color emotional needs of this group can reflect the problem awareness of the dynamic development of the cosmetic market and, thus, highlight the research and practical value of this topic.

NCD Color Image Coordinate System

The NCD semantic positioning system was developed by the Japan Institute of Color Design, which is based on the theories of perceptual engineering and color psychology, and established through experimental parameters reflecting the mapping relationship between color and semantics. This color matching system provides a valuable reference for modern design practices. The color psychologist Kobayashi Shigshun established the NCD color system through more than 40 years of research studies. He systematically categorized the meaning of color from the image perspective on the basis of the principle of western color science as well as through the basic platform of sociology, iconography, and psychology, and, thus, constructed a system that is suitable for local human environment and more in line with modern design concepts. According to the survey shown in *a study of color emotion and color preference. Part i: color emotions for single colors* by Ou et al. (2010), this article classifies color emotions for single colors and develops science-based color emotion models. The factor analysis identified three color-emotion factors: color activity, color weight, and color heat. The three factors agreed well with those found by Kobayashi and Sato et al. Four color-emotion models were developed, namely, warm–cool, heavy–light, active–passive, and hard–soft.

The NCD semantic system is a positioning system that innovatively establishes emotional semantic words based on a variety of sensory experiences of different users. Through rational design logic thinking, the image space between color and perceptual vocabulary is constructed by three psychological axes of cold and warm axis (W–C), soft and hard axis (S–H), and clear turbidity (K–G), and the image system is established on this basis. **Figure 1** shows the hue and tone system. There are 120 types of colors arranged in the horizontal and vertical orders according to the hue, and 10 types of chromatic colors, totally 130 types of colors. **Figure 2** shows the monochrome image coordinates, which are made according to the data obtained from the psychological survey of language sense, and the colors of the same tone are connected by lines.

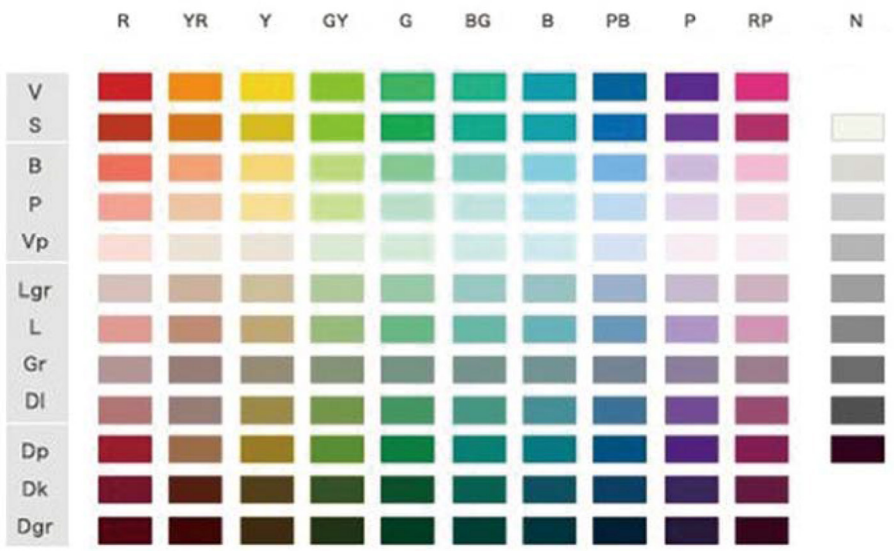


FIGURE 1 | Hue and tone system.

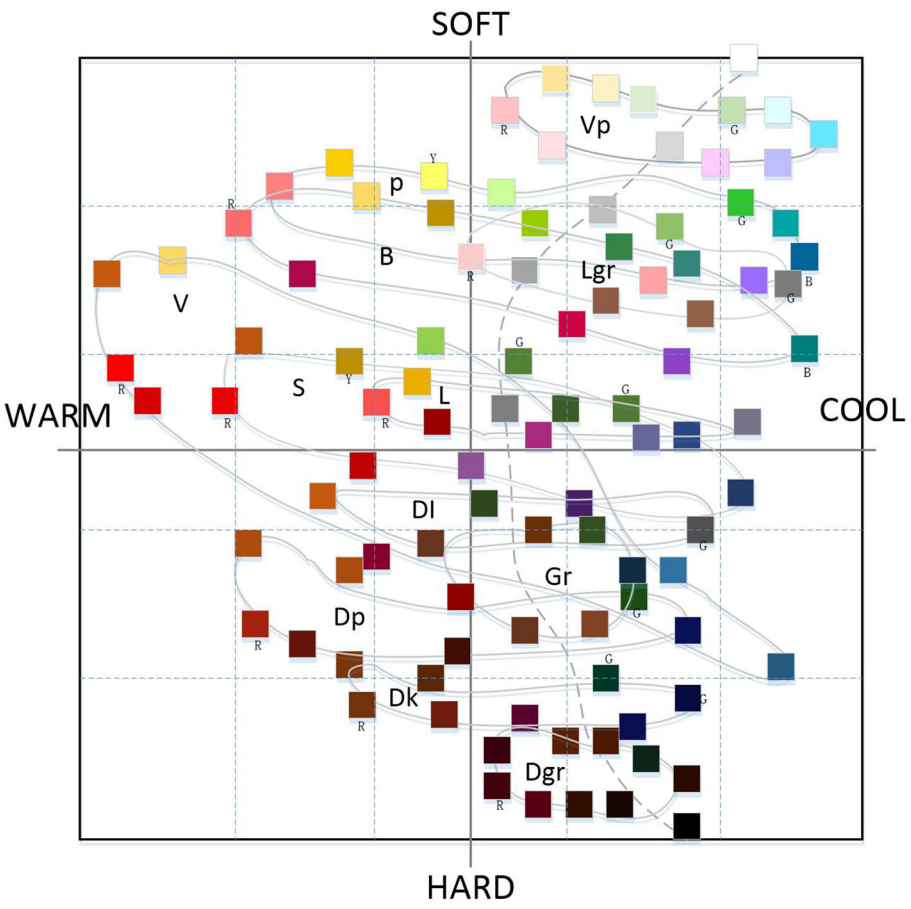


FIGURE 2 | Monochrome image coordinates.

Color Positioning of Beauty Packaging at Home and Abroad

The NCD semantic system is color-matching of a semantic coordinate system, which is formed by analyzing the perspective of color image based on the three basic attributes of color (hue, brightness, purity). The NCD semantic system converts the perceptual psychological needs of users, such as sensory impressions and aesthetic preferences, into semantic symbols, matches the design of cold, warm, soft, and hard colors to meet the perceptual needs of users, and finally proposes solutions to meet the color emotional preferences of different users.

This research investigates, analyzes, and classifies the color status of beauty packaging at home and abroad. We try to explore an accurate and effective color design method to design the color of beauty packaging to enhance the emotional experience of customers. Combining the research theme of color imagery, the practical color coordinate system (PCCS) is used to refine and abstract the color of selected beauty packaging samples, and to conduct statistical and visual analysis of color distribution and visualization of the color style positioning and emotional tendency of beauty brands at home and abroad. Based on the NCD semantic system, its essence is to achieve the consistency of the mapping between color semantics and user perception on the premise that color image positioning is consistent with the perceptual needs of the user. A consistent quantitative basis is established, and the perceptual engineering and quantitative color systems is applied to complete the quantitative research of matching objects at both ends of the mapping, which provides the feasibility for establishing consistency in orientation.

1) The hue and tone distribution of beauty packaging at home and abroad

After preliminary investigations, we selected basic skin care products as the subject of this research. Contemporary women are very demanding on their own appearances, 78% of women are not satisfied with their skin conditions, and their requirements for beauty are getting higher and higher. Among them, anti-aging has become their main demand. CBNData research shows that from the consumption of different effects in the online female skin care industry in 2020, the demand for anti-aging skin care accounted for 34%, exceeding the long-term concern of most women for moisturizing and moisturizing needs (33.4%). The significance of choosing this type of beauty product research lies in three aspects. First, basic skin care products are the most representative products in the current beauty industry. Second, the purchase frequency of skin care products is relatively fixed, and their seasonal characteristics are strong. The main purchases are quarterly basis, with one to two times a quarter being the mainstream purchase frequency. Third, the consumption level of skin care products tends to be even, and there are a considerable number of consumer groups in the middle, high, and low categories.

The survey collected 200 samples of beauty packaging at home and abroad. The main containers and packaging of skin care products are usually plastic bottles, glass bottles, hoses, and airless bottles. Factors such as different painting processes and reflective surfaces of the outer packaging are excluded, and the final color application situation is shown in **Figure 3**. The survey results show that the overall hue of basic skin care products is more blue-green than red-orange. This aspect is related to the use of the product by a consumer, mainly for cleansing, moisturizing, and moisturizing. The color is cooler, giving a refreshing and clean feeling. From the perspective of different origins and brand image positioning, achromatic gray and white are used more frequently in the packaging of European and American skin care products. In addition, cool hues such as blue-green and blue appear more frequently. Japanese and Korean brands tend to have warm red-orange colors, which are softer and warmer than European and American brands. The colors of Southeast Asian brands tend to be more gorgeous, the colors of domestic Chinese brands are more diverse, and product segmentation is closer to the specific needs of consumers.

2) Color distribution of cosmetics packaging in image space

The distribution of the packaging colors of basic skin care products at home and abroad in the image space is shown in **Figure 4**. The distribution is mainly concentrated on the upper right of the color arc, showing a tendency to be colder overall, with relatively more cool colors. The color group below the central axis, the low-brightness dark color group, is obviously less than the high-brightness color group above the central axis. The lightness and purity of the two ends of the axis are relatively more tones. The color is relatively soft in terms of softness and hardness, and the soft colors are more concentrated, indicating that the main color and light color have the same lightness and, thus, the emotional demand for the product is more consistent.

COLOR EMOTION ANALYSIS OF MIDDLE-AGED WOMEN IN RURAL AREAS OF NORTHERN CHINA ON BEAUTY PACKAGING

Extraction of Representative Semantics of Beauty Packaging Color Image

This research is based on the collection and selection of color image adjective pairings of beauty packaging. First, after determination of the representative color test samples of beauty packaging in the previous steps (**Figures 3**, **4**), we will explore the color design and planning process of beauty packaging. In order to gain more comprehensive and accurate image cognition of testing the beauty packaging, all the colors of the beauty packaging were paired together, and then an integral color image of this product series was formed. The color positioning and emotional semantics produced by this overall image are relatively clear, rich, and perceptible. The main references include: Kobayashi Shigshun, Analysis of Color Psychology; the website "NIPPON COLOR & DESIGN RESEARCH INSTITUTE"; Fujisawa Hideaki, Color psychology; ChiQianYanHideaki, The Incredible Psychological Colors; Tadashi Oyama, The Psychology of Color; Heller & Wu Tong, and The Character of Color. Among

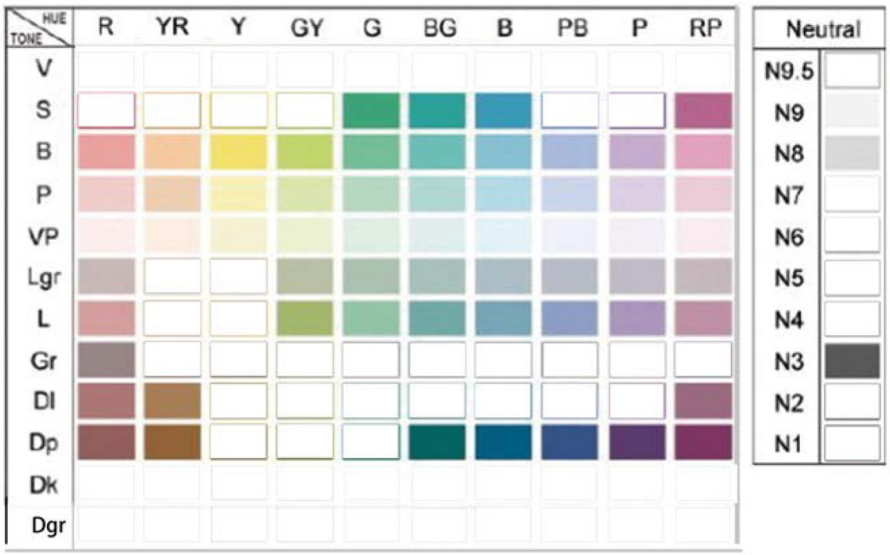


FIGURE 3 | Hue and tone distribution of cosmetics packaging in China and abroad.

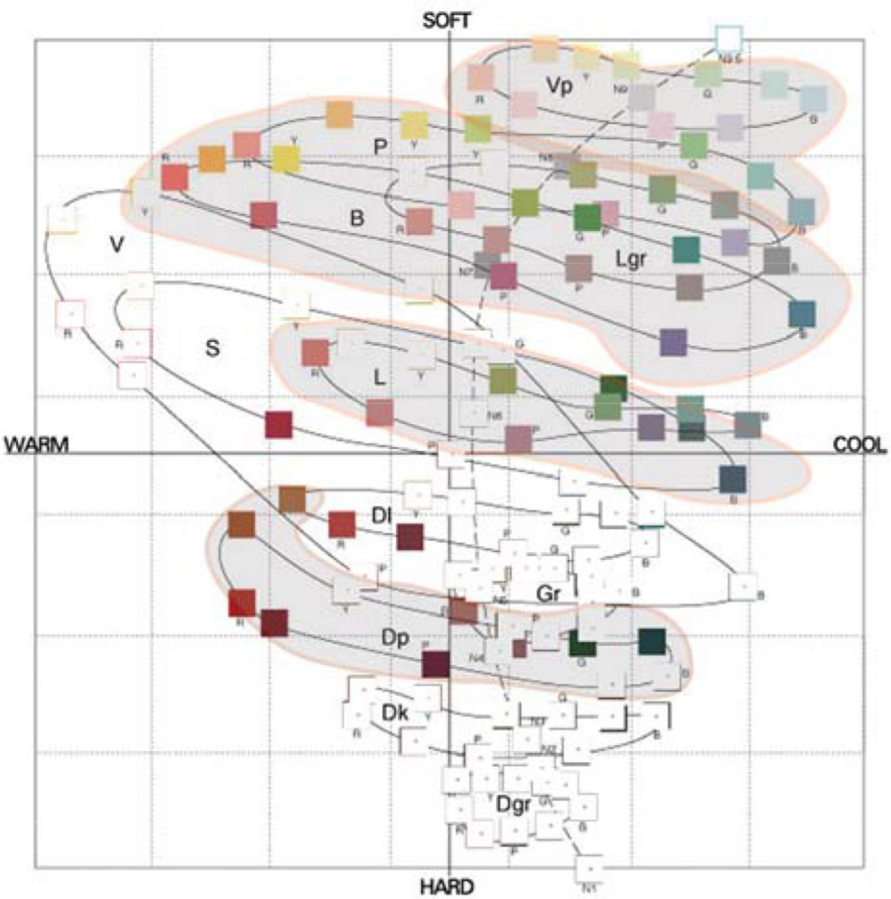


FIGURE 4 | Color distribution of cosmetic packaging in image space.

TABLE 1 | Selected 30 sets of color image adjective pairs.

1. Emotional–intellectual	11. Youthful–mellow	21. Intellectual–progressive
2. Elegant–dandy	12. Modern–traditional	22. Complex–simple
3. Sweet–bitter	13. Stylish–classic	23. Abundant-rich–pure
4. Masculine–feminine	14. Lighthearted–intense	24. Natural–artificial
5. Casual–formal	15. Provocative–sober	25. Alluring–conservative
6. Dynamic–calm	16. Agile–sedate	26. Provincial–urban
7. Gorgeous–plain	17. Speedy–smooth	27. Luxurious–rustic
8. Lighthearted–nostalgic	18. Fascinating–rustic	28. Romantic–practical
9. Fashionable–old-fashioned	19. Passionate–grave	29. Negative–positive
10. Western–eastern	20. Innocent–elaborate	30. Delicate–coarse

them, 180 adjectives commonly used in color image analysis summarized by Japan Color Research Institute are involved.

On this basis, we perform a detailed analysis of the specific meaning of each group of adjectives, and deleted the adjective groups with similar meanings, similarities, and unclear expressions. We also search relevant literature, referring to the cognitive suggestions of cosmetic design practitioners on the color of beauty packaging, and then obtain the semantic description related to the color image of beauty packaging. Finally, we select 30 sets of color image adjective pairs as the final test vocabulary, as shown in **Table 1**.

Determine Middle-Aged Women in Northern Rural Areas of China as the Research Object

The female consumer market in China is huge, and the rapidly growing market has led to changes in the market structure. The growth of female users will stimulate the overall market. Exquisite product packaging and unique emotional experience often tend to arouse the consumer psychology of this group and stimulate consumer impulse. Therefore, studying the color preferences of female consumer groups has great potential and value for the market. White collar workers are the mainstream consumers of skin care products. They will choose first- and second-tier mid- to high-end products. As far as the development of domestic brands is concerned, the market value of domestic brands is that they are better at developing channels in second- and third-tier cities than international brands, and quickly seize the mass market in second- and third-tier cities and rural areas.

In this case, the middle-aged female consumer group of about 30–55 years old is selected as the study subject. There are three main reasons for studying the northern rural women of this age group:

(1) The consumption behaviors of northern rural women under the age of 30 have not yet been fixed, and they are mostly seeking differentiation and have a strong sense of group imitation consumption;
(2) Among the northern rural woman consumers who are over 55 years of age, consumption behaviors and consumption habits have reached a fixed pattern in the middle and young ages, and their personal purchasing intentions have weakened, and new consumer products will be purchased by their children;
(3) Among the northern rural women, those who are 30–55 years of age have a leading role in daily commodity decision-making and purchasing activities. They are not only the decision-makers and purchasers of their own consumption activities but also of their husbands, children, and parents. Due to the influence of different consumer demand, factors such as different personalities, different regions, and different cultural education levels, female consumers in this age group have a certain degree of representativeness. When they purchase cosmetic products, they are more susceptible to the consumer psychology of product appearance and demands for aesthetic emotion.

FACTOR ANALYSIS OF COLOR IMAGE SEMANTICS IN BEAUTY PACKAGING

The method of factor analysis was widely used in semantics analysis research. For example, the study investigates the relationship between color perceptual attributes and color emotions, as well as the influence of different cultural backgrounds. It was concluded that chroma and lightness were the most important factors on color emotion, whereas the influences of hue and cultural background were very limited. In *Analysis of cross-cultural color emotion* by Gao et al. (2010), semantic-differential data of color emotions for color pairs were collected and examined by factor analysis (FA). In *Analyses of Color Emotion for Color Pairs with Independent Component Analysis and Factor Analysis* by Hanada (2013), word associations or verbal synesthesia between concepts of color and emotions were studied in Germany, Mexico, Poland, Russia, and the United States. With emotion words as the between-subjects variable, 661 undergraduates indicated on a 6-point scale to what extent of anger, envy, fear, and jealousy 12 terms of color reminded them in *Correspondence Analysis of Color–Emotion Associations* by Hanada (2018). In *The Color of Music: Correspondence through Emotion* by Barbiere et al. (2007), it was concluded that music-color correspondences occur *via* the underlying emotion common to the two stimuli, and this study investigated the relationship between color perceptual attributes and color emotions, as well as the influence of different cultural backgrounds.

Based on 97 questionnaires, the average value of the semantic evaluation of each color sample is computed. We use SPSS to perform factor analysis, and by the correlation analysis of each variable, the suitability of factor analysis is judged, and the principal component analysis method is used for variable conversion. Using new variables to construct and find several relatively independent new variables can better explain the changes of the original semantic factors, thus constituting a common factor. Then, according to the corresponding variables and samples, the cognitive image space map is drawn to facilitate the formation of color cognition of beauty packaging.

In the first factor analysis, there were a total of 31 factors, and the results showed that there was a high factor load in many

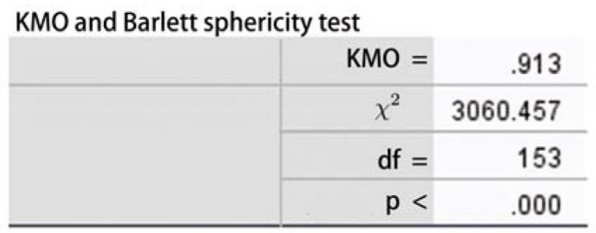


FIGURE 5 | Kaiser–Meyer–Olin (KMO) and Bartlett's test.

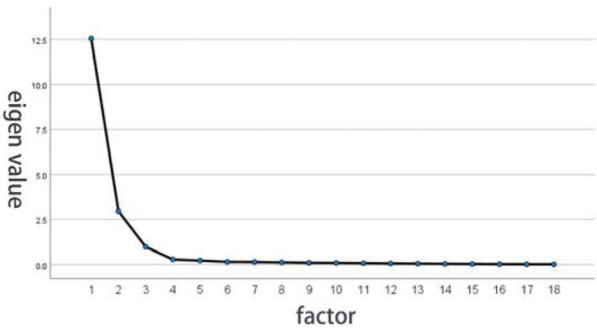


FIGURE 6 | Eigenvalue factor analysis of screen plot.

factors. Therefore, the topic was deleted and then re-analyzed. After the second factor analysis, the results showed that there was still a high factor load in many factors, so the topic was deleted, and we repeated the analysis.

Test on the Stability of Factor Analysis

In order to test whether the original data are suitable for factor analysis, it is necessary to pass the correlation coefficient matrix, Kaiser–Meyer–Olin (KMO) and Bartlett sphericity test, and the test of the common degree of variables. The variants are not irrelevant to each other ($X^2 = 3060.475$, $df = 153$, $p < 0.001$) and the result shows KMO = 0.913, as shown in **Figure 5**. According to the measurement standard of KMO provided by Kaise (0.9< KMO <1, very suitable), the material is suitable for factor analysis. When using the screen plot shown in **Figure 6**, three factors should be chosen, and these variables are suitable to conduct the factor analysis.

In order to determine the degree of correlation between the extracted common factors and the original variables, it is necessary to test the degree of commonality of the variables. For example, **Table 2** shows the common degree of the variable output by the software. The common degree of the first variable "delicate-coarse" in the table is 0.98, indicating that all the extracted factors together explain the 0.98% of the variation information produced by the variable "delicate/coarse." From the results, the common degree of the 17 variables is >0.8, which means that extracting three common factors is an ideal state.

Three factors are extracted using the principal axis method and carrying out the promax. As shown in **Table 3**, the element loading martix shows that the first factor contains eight questions, whose contents are mostly related to elegant sense of quality (such as delicate/coarse, intellectual/progressive, fashionable/old-fashioned, lighthearted/intense, elegant/dandy, natural/artificial, lighthearted/nostalgic, and sweet/bitter). The second factor contains six questions, whose contents are mostly related to dynamic sense of youth (such as youthful/mellow, agile/sedate, speedy/smooth, dynamic/calm, provocative/sober, and emotional/intellectual). The third factor contains four questions, whose contents are mostly related to area sense of luxury (such as luxurious/rustic, gorgeous/plain, western/eastern, and fascinating/rustic).

TABLE 2 | Common degrees of variables.

	Variable	Original	Extract
01	Emotional–intellectual	1.000	0.69
02	Elegant–dandy	1.000	0.95
03	Sweet–bitter	1.000	0.86
04	Masculine–feminine	1.000	0.33
05	Casual–formal	1.000	0.47
06	Dynamic–calm	1.000	0.84
07	Gorgeous–plain	1.000	0.89
08	Lighthearted–nostalgic	1.000	0.90
09	Fashionable–old-fashioned	1.000	0.95
10	Western–eastern	1.000	0.87
11	Youthful–mellow	1.000	0.98
12	Modern–traditional	1.000	0.61
13	Stylish–classic	1.000	0.41
14	Lighthearted–intense	1.000	0.95
15	Provocative–sober	1.000	0.83
16	Agile–sedate	1.000	0.97
17	Speedy–smooth	1.000	0.86
18	Fascinating–rustic	1.000	0.87
19	Passionate–grave	1.000	0.59
20	Innocent–elaborate	1.000	0.47
21	Intellectual–progressive	1.000	0.97
22	Complex–simple	1.000	0.46
23	Abundant-rich–pure	1.000	0.32
24	Natural–artificial	1.000	0.94
25	Alluring–conservative	1.000	0.73
26	Provincial–urban	1.000	0.61
27	Luxurious–rustic	1.000	0.98
28	Romantic–practical	1.000	0.66
29	Negative–positive	1.000	0.53
30	Delicate–coarse	1.000	0.98

Determination of the Number of Extracted Factors

It can be seen from the output data shown in **Table 4** that the eigenvalue of the first factor solution is 12.548, which explains 69.711% of the total variation information of all the 18 variables and is a principal component with the largest variance contribution. The eigenvalue of the fourth factor

solution is 0.289. All the eigenvalues are <1 and, hence, only the first three factors are extracted as common factors. Besides, the first three factors show 91.587% of the total variation information of all variables, reaching a good level. Therefore, it is relatively appropriate to extract three common factors. After orthogonal rotation, the variance contributions of factors are different; the purpose is to increase the load difference of each variable and make the variation in the correlation matrix to be scattered to different factors as far as possible, so the respective contribution rates after rotation are 69.711, 86.149, and 91.587%.

TABLE 3 | Pattern matrix diagram.

		Element		
		1	2	3
01	Delicate–coarse	0.98		
02	Intellectual–progressive	0.97		
03	Fashionable–old-fashioned	0.95		
04	Lighthearted–intense	0.95		
05	Elegant–dandy	0.95		
06	Natural–artificial	0.94		
07	Lighthearted–nostalgic	0.90		
08	Sweet–bitter	0.86		
09	Youthful–mellow		0.98	
10	Agile–sedate		0.97	
11	Speedy–smooth		0.86	
12	Dynamic–calm		0.84	
13	Provocative–sober		0.83	
14	Emotional–intellectual		0.69	
15	Luxurious–rustic			0.98
16	Gorgeous–plain			0.89
17	Western–eastern			0.87
18	Fascinating–rustic			0.87

Factor Extraction and Naming

Tables 5, **6** show the output factor loading matrix without rotation and the factor loading matrix after rotation, respectively, with the corresponding relationship between the original variables and the common factors. The factor loading matrix without rotation shows that the loads of 18 variables are higher on the first factor. The loads of the variables "delicate/coarse," "intellectual/progressive" are the highest, 0.99 and 0.98, respectively. The two variables "youthful/mellow" and "agile/sedate" have relatively high loads on the second factor, which are 0.99 and 0.97, respectively. The variable "luxurious/rustic" has the highest load, which is 0.98 on the third factor. Therefore, the variance maximization method is used to further differentiate the size of orthogonal rotation load, so that the corresponding relationship between variables and factors is clearer, and that the main variables affected by each factor can be identified more easily.

The first factor mainly affects three groups of variables: delicate/coarse, intellectual/progressive, and fashionable/old-fashioned, among which the load of delicate/coarse is the highest, which is 0.98. These three groups mainly describe the rational cognition of color images, which can be called "cognitive factors."

The second factor mainly affects three groups of variables: youthful/mellow, agile/sedate, and speedy/smooth, among which the load of youthful/mellow is the highest, which is 0.98. These three groups are mainly about direct experience and feelings of color images, which can be called "activity factors."

The third factor mainly affects four groups of variables: luxurious/rustic, gorgeous /plain, western/eastern, and

TABLE 4 | Explanation of total variance.

Component	Total	Original value	Cumulative	Total	Quadratic sum	Cumulative	Total after rotation
1	12.548	69.711	69.711	12.548	69.711	69.711	11.065
2	2.959	16.438	86.149	2.959	16.438	86.149	10.973
3	0.979	5.438	91.587	0.979	5.438	91.587	9.045
4	0.289	1.606	93.193				
5	0.228	1.268	94.461				
6	0.157	0.871	95.332				
7	0.148	0.822	96.154				
8	0.125	0.693	96.847				
9	0.105	0.582	96.429				
10	0.096	0.535	97.965				
11	0.082	0.455	98.419				
12	0.069	0.381	98.800				
13	0.059	0.330	99.131				
14	0.046	0.253	99.384				
15	0.038	0.209	99.593				
16	0.029	0.163	99.756				
17	0.024	0.131	99.887				
18	0.020	0.113	100.000				

TABLE 5 | Factor loading matrix without rotation.

		Element		
		1	2	3
01	Delicate–coarse	0.99		
02	Intellectual–progressive	0.98		
03	Fashionable–old-fashioned	0.96		
04	Lighthearted–intense	0.95		
05	Elegant–dandy	0.91		
06	Natural–artificial	0.94		
07	Lighthearted–nostalgic	0.87		
08	Sweet–bitter	0.43		
09	Youthful–mellow		0.99	
10	Agile–sedate		0.97	
11	Speedy–smooth		0.87	
12	Dynamic–calm		0.86	
13	Provocative–sober		0.83	
14	Emotional–intellectual		0.70	
15	Luxurious–rustic			0.98
16	Gorgeous–plain			0.88
17	Western–eastern			0.88
18	Fascinating–rustic			0.86

TABLE 6 | Factor loading matrix after rotation.

		Element		
		1	2	3
01	Delicate–coarse	0.98		
02	Intellectual–progressive	0.97		
03	Fashionable–old-fashioned	0.95		
04	Lighthearted–intense	0.95		
05	Elegant–dandy	0.95		
06	Natural–artificial	0.94		
07	Lighthearted–nostalgic	0.90		
08	Sweet–bitter	0.86		
09	Youthful–mellow		0.98	
10	Agile–sedate		0.97	
11	Speedy–smooth		0.86	
12	Dynamic–calm		0.84	
13	Provocative–sober		0.83	
14	Emotional–intellectual		0.69	
15	Luxurious–rustic			0.98
16	Gorgeous–plain			0.89
17	Western–eastern			0.87
18	Fascinating–rustic			0.87

fascinating/rustic, among which the load of luxurious/rustic is the highest, which is 0.98. These four groups mainly describe the social and physical attribution of color, which can be called "attributive factors."

The Corresponding Relation Between Sample Colors and Factor Axis

After determining the three common factors, the system will transform the factor scores of all the samples into a visual scatter diagram, by which we can more intuitively judge and analyze the spatial distribution of color images in beauty packaging. **Figure 7** shows the distribution of color semantics in the three-dimensional cognitive space (composed of three factors). **Figure 8** shows that the two-dimensional cognitive space matrix is composed of three factors matching with each other. The influence of semantics on public factors can be determined by its position in space, while the distance between them indicates the similarities and differences of semantics. Hence, when evaluating the color image of beauty packaging, the corresponding evaluation semantic system can be selected according to the distribution position and distance of cognitive space.

Through the comparative analysis of the representative semantics of the three common factors (factor axis) extracted from the factor analysis and the color attributes of the corresponding representative samples, the corresponding relationship and correlation characteristics between the sample color attributes and the semantics conveyed by factor axis are discussed. It can be seen from **Figure 9** that the standardization coefficient of each path is <1, and P-values are more than 0.8, indicating that the cognitive attribute, activity attribute, and attribution attribute of cosmetic packaging color will affect the willingness of consumers to decide which beauty packaging to choose, and that all of them will have a significant positive impact. Among them, the cognitive attribute plays the most important role, the attribution attribute of beauty packaging plays a second role in the choices of consumers, and the last is the activity attribute.

NARRATIVE ANALYSIS OF QUESTIONNAIRE ON BEAUTY PACKAGING COLOR

Based on the color sample statistics of 97 beauty packaging and using a Likert scale, the representative semantics of the three common factors are extracted. **Figure 10** shows that the average value of "delicate/coarse" in the "cognitive factor" is 3.38, the average value of "youthful/mellow" in the "activity factor" is 4.03, and the average of "luxurious/rustic" in the "attributive factor" is 4.37. It can be seen that at the level of rational cognition, the group prefers "delicate" to the color tendency of beauty packaging; at the level of experience and feelings, the group prefers "mellow"; and at the level of social geography, the group prefers "rustic."

Color Attributes Characteristics of Semantics Represented by Cognitive Factor

The first common factor (cognitive factor) extracted from the factor analysis shows that its representative vocabularies

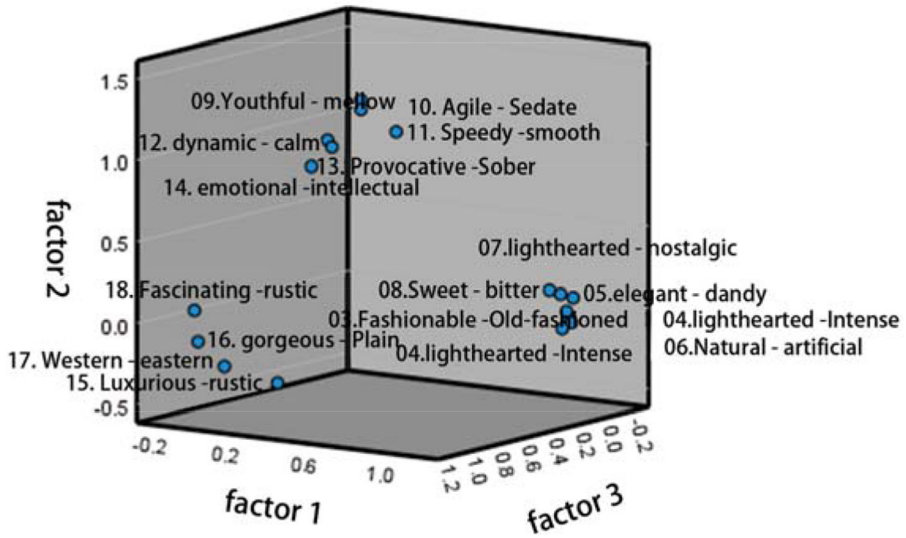


FIGURE 7 | Distribution of color semantics in the three-dimensional cognitive space.

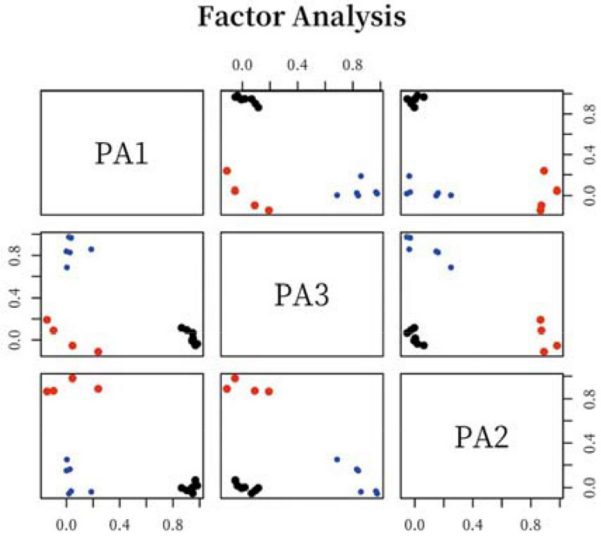


FIGURE 8 | Distribution of color semantics in the two-dimensional cognitive space.

are delicate/coarse, intellectual/progressive, and fashionable/old-fashioned. The most remarkable samples of negative axes (delicate, intellectual, fashionable) on this factor are: YR/B, Y/B, G/L, BG/L, B/L, G/S, BG/S; the most remarkable samples of positive axes (coarse, progressive, and old-fashioned) are R/DGR, YR/DGR, and Y/DGR. By analyzing the color attributes of typical sample colors, we observe that the color features corresponding to approximate sensory images such as delicate, intellectual, and fashionable, are mainly as follow: the color has a visual perception of weak brightness, and the change in chromaticity is also inclined to increase the content of white. In terms of hue, blue-green color is the most remarkable, and there are also some colors of orange and yellow, which show the psychological feelings of visual cognition and prefer soft colors, while the color brightness and purity of main-tune tendency prefer stable color combinations with little fluctuation. It shows that women of this age group with a natural sense of closeness to rural life prefer neutral and cold colors in hue, and turbid colors, which have low color purity and tend to gray system. The women in northern China are reserved, and the colors they prefer have the characteristics of low brightness and low purity, and basically form the conservative and cautious semantics, which is consistent with their personalities.

Color Attributes Characteristics of Semantics Represented by Activity Factor

The second common factor (activity factor) extracted from the factor analysis shows that its representative vocabularies are youthful/mellow, agile/sedate, and speedy/smooth. The most remarkable samples of negative axes (youthful, agile, speedy) on this factor are: RP/V, R/S, B/S; the most remarkable samples of positive axes (mellow, sedate, smooth) are: PB/B, PB/GR, B/GR. By analyzing the color attributes of typical sample colors, we observe that the color features corresponding to approximate sensory images such as mellow, sedate, and smooth, are mainly as follow: the color has a visual perception of weak brightness, and the change in chromaticity is also inclined to increase the content of black. In terms of hue, blue-purple color is the most remarkable; chromaticity also tends to the middle gray scale of achromatic colors. Colors of gray and blue-gray are the main hues, which show the reserved and conservative characteristics of the group. The tone is mainly cloudy and gray, which is closer to the quiet and calm inner feelings, contrary to the semantic characteristics such as being open and enthusiastic.

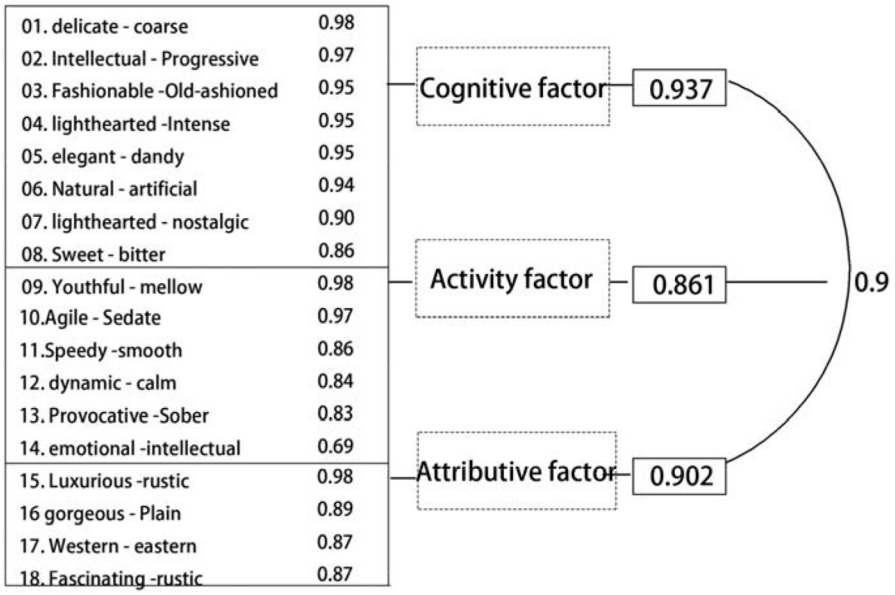


FIGURE 9 | Numerical analysis diagram of three common factors.

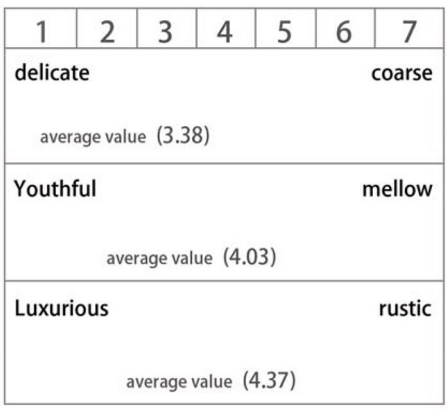


FIGURE 10 | Average value of semantics represented by three common factors.

Color Attributes Characteristics of Semantics Represented by Attributive Factor

The third common factor (attributive factor) extracted from the factor analysis shows that its representative vocabularies are luxurious/rustic, gorgeous/plain, and western/eastern. The most remarkable samples of negative axes (luxurious, gorgeous, western) on this factor are R/DP and YR/DP; the most remarkable samples of positive axes (plain, eastern, rustic) are GY/Lgr, G/Lgr, BG/Lgr, P/P, and N8. By analyzing the color attributes of typical sample colors, we observe that the color features corresponding to approximate sensory images such as plain, eastern, and rustic, are mainly as follow: in terms of brightness, it is better to choose middle brightness, and the change in chromaticity is also inclined to increase the content of white. In terms of hue, the colors of blue, purple, and gray, which can provide the best visual experience by reflecting the oriental life, also reflect the color emotional experience of the group in the specific cultural background.

CONCLUSION

Nowadays, perceptual engineering and color psychology are both emerging fields in the design industry. They explore the relationship between rational thinking and perceptual impression of humans, how aesthetic experience of people is formed, and how to construct the relationship between human emotions, environment, and objects. A good communication bridge between the participants and finding effective emotional design methods are issues worthy of further investigation. NCD color image positioning, as the front-end system of color design, is the scientific color theory basis for fashion product designers. This study constructs a complete color reference system by investigating and positioning the color image of middle-aged women in the rural areas of northern China, focusing on exploring the co-existence of the color design process on color, women, and the environment. The proposed color-oriented design method conforms to the development trend of scientific design, clarifies the color logic relationship of "color impression and emotional semantics," and is oriented to enhance the emotional experience of the user, which is conducive to accurately creating aesthetic

characteristics of the local market. The product color with cultural connotation highlights the application value. At the same time, the construction of the color system meets the aesthetic needs of the Chinese population, and has the commercial practice value of more optimized design and more scientific structure.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: www.wjx.cn, https://www.wjx.cn/mobile/statnew.aspx?activity=111860856&reportid=#1, https://www.wjx.cn/vm/tU4Ai4A.aspx.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This study was supported in part by National Social Science Foundation of China Key Project of Art Science Research on Chinese Animation Creation and Theoretical Innovation under the Construction of National Cultural Image with Grant No: 20AC003, and by Jiangsu University Philosophy and Social Sciences Research Fund with Grant No: 2018SJA0809.

REFERENCES

Barbiere, J., Vidal, A., and Zellner, D. (2007). The color of music: correspondence through emotion. *Empir. Stud. Arts* 25, 193–208. doi: 10.2190/A704-5647-5245-R47P

Bolshunova, N. Y., and Khromova, T. I. (2021). *Color Preferences in Adolescence and Early Adulthood and Their Relationship With Emotional Intelligence*. Krasnoyarsk: Federal State-Funded Educational Institution of Higher Education "Krasnoyarsk State Pedagogical University named after V.P. Astafiev". doi: 10.25146/1995-0861-2021-55-1-260

Gao, X. P, Xin, J. H., Sato, T., Hansuebsai, A., Scalzo, M., and Kajiwara, K., et al. (2010). Analysis of cross-cultural color emotion. *Color Res. Appl.* 32, 223–229. doi: 10.1002/col.20321

Hanada, M. (2013). Analyses of color emotion for color pairs with independent component analysis and factor analysis. *Color Res. Appl.* 38, 297–308. doi: 10.1002/col.20750

Hanada, M. (2018). Correspondence analysis of color–emotion associations. *Color Res. Appl.* 43, 224–237. doi: 10.1002/col.22171

Haruko, T., and Emiko, W. (1972). Relations between their color preference and colors of their clothes color preference of little children. *J. Home Econ. Jpn.* 23, 190–195. doi: 10.11428/jhej1951.23.190

Marine, G., and Nobuo, M. (2016). Atypical color preference in children with autism spectrum disorder. *Front. Psychol.* 7:1976. doi: 10.3389/fpsyg.2016.01976

Ou, L.C., Luo, M.R., Woodcock, A., and Wright, A. (2010). A study of colour emotion and colour preference. Part III: colour preference modeling. *Color Res. Appl.* 29, 381–389. doi: 10.1002/col.20047

Ou, L. C., Luo, M. R., Woodcock, A., and Wright, A. (2004). A study of colour emotion and colour preference. Part I: colour emotions for single colours. *Color Res. Appl.* 29, 232–240. doi: 10.1002/col.20010

Saito, M. (1996). Comparative studies on color preference in japan and other asian regions, with special emphasis on the preference for white. *Color Res. Appl.* 21, 35–49. doi: 10.1002/(SICI)1520-6378(199602)21:1<35::AID-COL4>3.0.CO;2-6

Schloss, K. B., Nelson, R., Parker, L., Heck, I. A., and Palmer, S. E. (2017). Seasonal variations in color preference. *Cogn. Sci.* 41, 1–24. doi: 10.1111/cogs.12429

Schloss, K. B., and Palmer, S. E. (2011). Aesthetic response to color combinations: preference, harmony, and similarity. *Attent. Percept. Psychophys.* 73, 551–571. doi: 10.3758/s13414-010-0027-0

Sin, L., So, S., Yau, O., and Kwong, K. (2001). Chinese women at the crossroads: an empirical study on their role orientations and consumption values in chinese society. *J. Cons. Mark.* 18, 348–367. doi: 10.1108/07363760110393010

Tam, J., and Tai, S. (1998). Research note: the psychographic segmentation of the female market in greater china. *Int. Mark. Rev.* 15, 61–77. doi: 10.1108/02651339810205258

Walsh, L. M., Toma, R. B., Tuveson, R.V., and Sondhi, L. (1990). Color preference and food choice among children. *J. Psychol.* 124, 645–653. doi: 10.1080/00223980.1990.10543258

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Wang, Xu, Zhang and Yin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 13 August 2021
doi: 10.3389/fpsyg.2021.728495

Motivation, Social Emotion, and the Acceptance of Artificial Intelligence Virtual Assistants—Trust-Based Mediating Effects

Shiying Zhang[1], *Zixuan Meng*[2], *Beibei Chen*[2], *Xiu Yang*[3]* *and Xinran Zhao*[2]

[1] Business and Economic Research Institute, Harbin University of Commerce, Harbin, China, [2] School of Business, Dalian University of Technology, Dalian, China, [3] Accounting and Auditing College, Guangxi University of Finance and Economics, Nanning, China

The complexity of the emotional presentation of users to Artificial Intelligence (AI) virtual assistants is mainly manifested in user motivation and social emotion, but the current research lacks an effective conversion path from emotion to acceptance. This paper innovatively cuts from the perspective of trust, establishes an AI virtual assistant acceptance model, conducts an empirical study based on the survey data from 240 questionnaires, and uses multilevel regression analysis and the bootstrap method to analyze the data. The results showed that functionality and social emotions had a significant effect on trust, where perceived humanity showed an inverted U relationship on trust, and trust mediated the relationship between both functionality and social emotions and acceptance. The findings explain the emotional complexity of users toward AI virtual assistants and extend the transformation path of technology acceptance from the trust perspective, which has implications for the development and design of AI applications.

Keywords: AI virtual assistant, motivation, social emotion, trust, acceptance, mediating effects, inverted U relationship

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Fei Hou,
Beijing Normal University, Zhuhai, China
Jing Xue,
Wuxi People's Hospital Affiliated to Nanjing Medical University, China

***Correspondence:**
Xiu Yang
2017110016@gxufe.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *21 June 2021*
Accepted: *20 July 2021*
Published: *13 August 2021*

Citation:
Zhang S, Meng Z, Chen B, Yang X and Zhao X (2021) Motivation, Social Emotion, and the Acceptance of Artificial Intelligence Virtual Assistants—Trust-Based Mediating Effects. Front. Psychol. 12:728495. doi: 10.3389/fpsyg.2021.728495

INTRODUCTION

With the advancement of AI technology, there are increasing numbers of Artificial Intelligence (AI) applications, such as service robots, chatbots, and AI virtual assistants (Gummerus et al., 2019). Regarding AI virtual assistants, which can offer convenience and more efficient services to users (van Doorn et al., 2017; Fernandes and Oliveira, 2021), people's interest and frequency of use is gradually increasing. Since technology acceptance is a key variable reflecting whether AI virtual assistants are accepted by users (Fernandes and Oliveira, 2021), it is important for product developers and corporate investors to explore the drivers of AI virtual assistant acceptance and their mechanisms of action.

However, current research on the acceptance of AI virtual assistants still suffers from the following three deficiencies. First, the current studies focus more on the impact of functionality and social emotions (Wirtz et al., 2018) than on the acceptance of AI virtual assistants (AVA), which helps reveal consumers' intention to use AI virtual assistants. However, the dual satisfaction of technical and social needs does not induce users to trust AI virtual assistants, which leads to low loyalty. In this case, it will be difficult for human users to collaborate with AI virtual assistants, thus limiting their application in society and making it difficult for AI virtual assistants to be truly accepted by human users.

In fact, trust can reduce human users' negative emotions about and affect their tendency to accept new technologies (Sparks and Browning, 2011). Consequently, the applicability of trust in the field of AI still needs further verification. Second, some scholars have tried to explore the potential mechanisms of users' trust in AI virtual assistants from the perspective of trust (Hassanein and Head, 2007; Glikson and Woolley, 2020); however, the relationship between trust and functionality and social emotions and whether the existing drivers have an impact on trust remain to be further explored. Finally, there is still a lack of effective transformation paths between AVA and their drivers, and it is unclear whether trust can carry the transformation between the two. This research scenario is not conducive to expanding the potential paths and intrinsic functions of AVA from a trust perspective.

This paper describes the following research findings. First, trust can reduce users' rejection of new things, thus promoting users' acceptance of AI virtual assistants at the psychological level, which is a direct driver of AI virtual assistant acceptance. Second, trust building depends on the user's motivation and perception of using the AI virtual assistant, that is, it is affected by functionality and social emotion. In this process, a positive experience contributes to user trust; for example, users are more concerned with whether something is useful or convenient. Nevertheless, users' perceptions of AI virtual assistants hardly have a coherent effect on trust, which includes the positive effects of perceived social presence and perceived social interaction as well as the inverted U-shaped effects of perceived humanity. In other words, the satisfaction of users' social needs by AI virtual assistants can effectively increase users' trust in them, but as perceived humanity tends to contribute to the trust transition effect, AI virtual assistants should be designed to maintain a moderate level of perceived humanity so that users can trust their services better. Finally, this paper reveals the transformation path between functionality, social emotion and AVA, namely, the mediation effect of trust is examined. In this procedure, trust presents two different effect mechanisms, which are partially mediated between functionality and acceptance and fully mediated between social emotion and acceptance, with two different degrees of mediation effects also indicating the effectiveness of relying on trust as a transformation path.

RESEARCH FRAMEWORK AND HYPOTHESIS DEVELOPMENT

Referring to the technology acceptance model and service robot acceptance model, this paper contains three levels of research on AVA: first, it investigates the relationship between trust and acceptance; second, it investigates the relationship between trust and functionality and social emotion; third, it investigates the mediating effect of trust between both functionality and social emotion and acceptance. Based on the above theoretical models, this study's theoretical model and hypotheses are shown in **Figure 1**.

Trust is defined as the user's confidence that the AI virtual assistant can reliably deliver a service (Wirtz et al., 2018). The services of AI virtual assistants are based on artificial intelligence algorithms, but due to the inherent black box problem of artificial intelligence technology (Asatiani et al., 2020), users will not fully trust the information or services provided by artificial intelligence virtual assistants (Kaplan and Haenlein, 2019). The existing research shows that only meeting the technical and social needs of users does not truly increase their loyalty to AI virtual assistants (Hassanein and Head, 2007). Trust will prompt users to subjectively reduce their negative emotional perception of AI virtual assistants, which contributes to their reduced complexity and vulnerability and plays a key role in improving the acceptance of AI virtual assistants (Shin, 2021). Therefore, this paper introduces trust variables to explore the mechanism underlying their effect on acceptance and makes the following hypothesis:

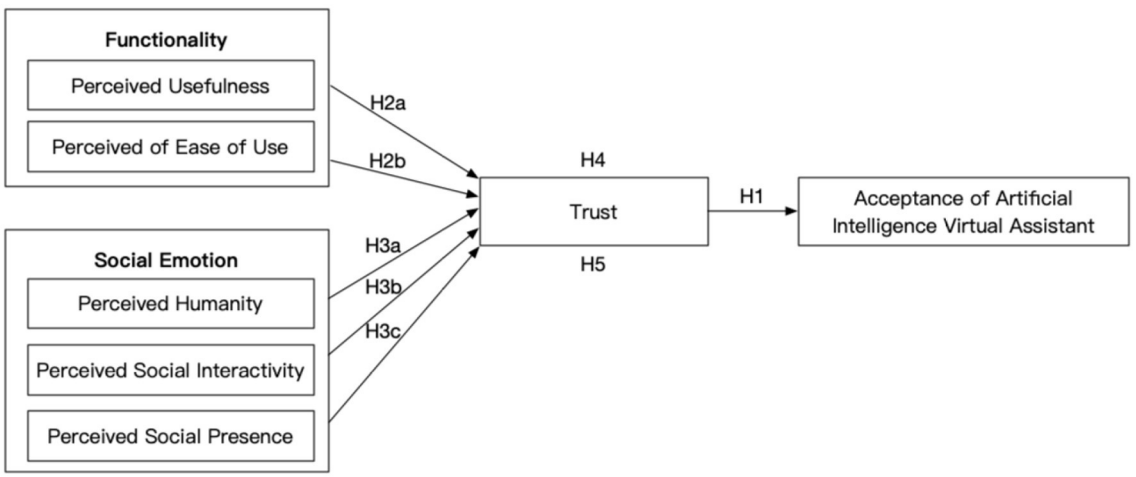


FIGURE 1 | Artificial intelligence virtual assistant acceptance model.

H1: Users' behavior of trusting AI virtual assistants is positively correlated with the AVA.

Perceived usefulness is the degree to which an individual perceives that a technology improves their performance and is an important factor in determining user acceptance, adoption, and usage (Kulviwat et al., 2007; Jan and Contreras, 2011). Perceived ease of use is the degree to which an individual's perception that using a technology requires minimal physical and mental effort, which is an important driver of technology acceptance and adoption (Kulviwat et al., 2007). Wirtz et al. (2018) put perceived usefulness and perceived of ease of use at the core of the service robot acceptance model. McLean and Osei-Frimpong (2019) found that utility benefits (namely, perceived usefulness and perceived of ease of use) have a positive impact on users' use of AI virtual assistants. The results show that perceived usefulness and perceived of ease of use are important antecedents that influence the formation of consumer trust. Venkatesh and Bala (2008), on the other hand, found that perceived usefulness and perceived of ease of use were significant predictors of behavioral intention. Perceived usefulness and perceived of ease of use have a positive effect on individuals' acceptance of the technology, and the higher the indicators of P perceived usefulness and perceived of ease of use, the more positive are users' attitudes toward the AI virtual assistant, which leads to trust in the AI virtual assistant. Glikson and Woolley (2020) argued that trust formation also depends upon machine competence (i.e., the extent to which it does its job properly). Therefore, this paper makes the following hypotheses.

H2a: Perceived usefulness is positively correlated with users' behavior of trusting AI virtual assistants.
H2b: Perceived ease of use is positively correlated with users' behavior of trusting AI virtual assistants.

Perceived humanity, also known as anthropomorphism, refers to whether the user perceives the AI virtual assistant as human during interaction with it. Perceived humanity is an important determinant of customer use of AI virtual assistants (van Doorn et al., 2017). Scholars hold different views on, with research showing that users tend to use anthropomorphized AI assistants (Epley et al., 2008). Drawing on the "Uncanny Valley" effect, a highly anthropomorphic AI virtual assistant will make users more inclined to measure human-computer interaction by the rules of human interaction and form higher expectations. When the AI virtual assistant makes a low-level mistake, the inconsistency between the high form of anthropomorphism and the mistake behavior defies the user's expectation and creates a sense of disgust. Once the AI virtual assistant is anthropomorphized, the user will experience a sense of connection with it (van Pinxteren et al., 2019), but because the AI virtual assistant is not human, it will create a sense of unnaturalness and may even cause the user's interaction with the AI virtual assistant to be completely interrupted (Tinwell et al., 2011; van Doorn et al., 2017). Users who are more sensitive to perceived humanity believe that AI virtual assistants with human-like characteristics threaten human specificity and self-identity (Gursoy et al., 2019). In addition, humans must interact with AI virtual assistants and learn how to interact with AI virtual assistants, thus increasing the burden on consumers to use AI devices (Kim and McGill, 2018). According to existing research, moderate perception of humanity will enhance the trust between human users and AI virtual assistants, while excessive perception of humanity will make human users feel threatened or even fearful, which may even cause human users to interrupt their interactions. Based on this, the paper makes the following assumptions:

H3a: Perceived humanity has an inverted U-shaped relationship with user trust in artificial intelligence virtual assistants.

Interaction means that people connect with information through interaction, in which they communicate and exchange emotions, energy, resources and other contents while generating judgments and reactions to the activities and words of others. From the perspective of information dissemination, interaction is based on the relationship between people and computers, using new technologies to enhance the interaction between users and computers (Shin, 2020). Current research suggests that interaction promotes both emotional and behavioral loyalty to technology (Wirtz et al., 2018; Sundar, 2020) and that enhanced interactions can increase user satisfaction with a website (Song and Zinkhan, 2008; Jiang et al., 2019). Hence, the interactivity of AI virtual assistants can engage users with positive effects. Perceived social interactivity can be defined as the perception that the AI virtual assistant displays appropriate actions and "emotions" according to societal norms (Wirtz et al., 2018). If an AI virtual assistant interacts in a social manner, demonstrates its social capabilities, and provides favorable service to the user, its social appeal increases (McLean and Osei-Frimpong, 2019), thus promoting trust in it (Chattaraman et al., 2019). Therefore, the following assumptions are made in this paper:

H3b: Perceived social interactivity is positively correlated with user trust in artificial intelligence virtual assistants.

Perceived social presence refers to the degree to which the user perceives the AI virtual assistant as a social entity. Drawing on social presence theory, perceived social presence is an inherent quality of AI virtual assistants. Perceived social presence means that the user has a perception of human interaction with the AI virtual assistant that is personal, social, warm and sensitive. If an AI virtual assistant conveys a sense of interpersonal and social connection to the user, the user will have a positive experience with it (Holzwarth et al., 2006; Wirtz et al., 2018) and perceive the AI virtual assistant as a real social entity. In cases where AI virtual assistants demonstrate a higher perceived social presence, users will build stronger trust in them (Wang and Emurian, 2005). Moreover, AI virtual assistants have real-time communication, voice, politeness, and other language-based communication skills, which can meet the social needs of human users and generate positive emotions, creating a harmonious social atmosphere, thus prompting human users to develop trust in AI virtual assistants (Fernandes and Oliveira, 2021). Therefore, the following assumptions are made in this paper.

H3c: Perceived social presence is positively correlated with user trust in AI virtual assistants.

Functionality refers to the degree of usefulness and convenience of the AI virtual assistant. Social emotion refers to the social experience of human users during their interaction with the AI virtual assistant. However, there is no effective transformation path between the two dimensions and acceptance; therefore, this paper introduces the trust variable to explore its mechanism of action between usage motivation and perception and acceptance. According to the brand effect, AI virtual assistants that can provide favorable services and information tend to bring users a sense of cozy experience, forming a positive cycle, which means that a positive product usage experience drives users to trust AI virtual assistants more, thus increasing their loyalty and acceptance. In environments lacking social emotion, users tend to hide information and reduce their trust behavior. Therefore, users will trust AI virtual assistants more in contexts where social emotions are stronger (Glikson and Woolley, 2020); that is, social emotions are necessary for trust development (Hassanein and Head, 2007). Research shows that trust is influenced by both rational (i.e., functionality) and emotional (i.e., social emotion) dimensions, and trust mediates between the rational and emotional dimensions and user acceptance (Palmatier et al., 2006; Glikson and Woolley, 2020). If the AI virtual assistant can bring more trust to the user, it will reduce the user's suspicion of it, and it will improve the AVA. Therefore, this paper explores the mediating role of trust between functionality and social emotion and acceptance and proposes the following hypotheses:

H4: User trust behavior has a mediating role between functionality and acceptance.
H5: User trust behavior has a mediating role between social emotion and acceptance.

DATA AND RESEARCH METHODOLOGY

Scale Design and Data Sources

At present, AI virtual assistants are widely used in daily life, and people of all ages and industries are able to access the use of AI virtual assistants and can grasp the scenario of this study more accurately. Therefore, this paper selects the public as the investigation target. This paper uses Wenjuanxing to generate online questionnaires and sends out questionnaire links and QR codes through WeChat groups, friend circles, QQ groups, online forums, and so on to invite the public to visit the questionnaire. This paper collects sample data with the help of the Questionnaire Star platform and sends questionnaires by QQ, WeChat, friends circles, among other routes. A total of 240 valid questionnaires were obtained. The descriptive statistics of the sample are shown in **Table 1**.

TABLE 1 | Results of descriptive statistics of the sample ($N = 240$).

Content	Category	Sample size	Proportion (%)
Gender	Male	101	42.08
	Female	139	57.92
Age	20<	36	15
	21–25	111	46.25
	26–30	32	13.33
	31–35	27	11.25
	36–40	13	5.42
	>41	21	8.75
Education	Undergraduate	152	63.33
	Master	37	15.42
	Ph.D.	8	3.33
	Other	43	17.92
Marital status	Unmarried	153	63.75
	Married with children	70	29.17
	Married with no children	17	7.08

Variable Measurement

To ensure the reliability and validity of the measurement scales, mature measurement scale items were selected for this study and appropriately adapted to the study scenario. In particular, functionality is based on the technology acceptance model and incorporates the findings of Venkatesh and Davis (2000) and others to classify functionality into two dimensions, perceived usefulness and perceived of ease of use, each of which contains four question items. Social emotion is based on the service robot acceptance model, which is divided into three dimensions: perceived humanity, perceived social interactivity, and perceived social presence. Combined with the study of Fernandes and Oliveira (2021), perceived humanity contains four question items, perceived social interactivity contains two question items, and perceived social presence contains three question items. Trust is based on the service robot acceptance model, containing one dimension of trust and four question items based on Shin (2021). The AVA contains three question items according to Fernandes and Oliveira (2021). The question items in this paper are all scored on a 5-point Likert scale, with 1–5 indicating very poor to fully conforming.

Reliability and Validity Tests

The descriptive statistics and correlation coefficients are shown in **Table 2**. The results show that perceived usefulness is positively related to trust ($r = 0.510$, $p < 0.01$), perceived of ease of use had a positive correlation with Trust ($r = 0.464$, $p < 0.01$), perceived social interactivity displayed a positive correlation with trust ($r = 0.547$, $p < 0.01$), and perceived social presence is found to be positively correlated with trust ($r = 0.537$, $p < 0.01$). These results preliminarily supported hypotheses 2a, 2b, 3b, and 3c.

As shown in **Table 3**, the Cronbach α of the seven variables are >0.7, indicating that factor analysis was appropriate.

Confirmatory factor analysis was conducted using SPSSAU to directly test the validity through confirmatory factor analysis of mature scales (as shown in **Table 4**). The results showed that standard load factors were all in acceptable range (>0.400), which indicates a strong correlation between the latent variables and the analytic term measures. And the results show that the AVE is >0.5 and the CR value is >0.7, which means the aggregation validity is high.

The results of KMO and Bartlett's test are shown in **Table 5**. The KMO value is more than 0.9 and p-value is <0.05, indicating that the validity of the study data was feasible.

TABLE 2 | Descriptive statistics results with correlation coefficients.

Variables	Average	Standard deviation	1	2	3	4	5	6	7
1 Perceived usefulness	3.579	0.835	1						
2 Perceived of ease of use	3.561	0.719	0.722**	1					
3 Perceived humanity	2.79	1.062	0.325**	0.162*	1				
4 Perceived social interactivity	3.333	0.911	0.679**	0.594**	0.463**	1			
5 Perceived social presence	3.189	1.026	0.555**	0.469**	0.551**	0.714**	1		
6 Trust	3.526	0.708	0.510**	0.464**	0.320**	0.547**	0.537**	1	
7 AVA	3.782	0.727	0.634**	0.528**	0.141*	0.506**	0.432**	0.467**	1

*$p < 0.05$ **$p < 0.01$.

TABLE 3 | Reliability test results.

Variables	Items	Cronbach α
Perceived usefulness	4	0.794
Perceived of ease of use	4	0.906
Perceived humanity	4	0.924
Perceived social interactivity	2	0.733
Perceived social presence	3	0.898
Trust	4	0.818
AVA	3	0.871

EMPIRICAL TESTING AND ANALYSIS

Selection of Research Method

The purpose of this study is to investigate the factors influencing acceptance and their mechanisms of action. Based on the existing literature, this paper identifies a study of acceptability consisting of four dimensions: functionality, social emotion, trust, and AVA. This paper uses multilevel regression analysis to analyze the correlations among the variables. The core of multilevel regression analysis is regression analysis; the difference is that hierarchical regression can be divided into multiple layers, where each layer adds more polynomials on top of the previous layer. This approach can solve the problem of whether more polynomials have explanatory power for the model. In addition, this paper used the product coefficient test for mediating effects; specifically, the bootstrap sampling method was used for testing. The basic idea of bootstrap sampling is to construct an estimated confidence interval with the help of multiple sampling with partial sample release when the full sample is unknown. This method has relatively high test efficacy and does not impose restrictions on the mediating sampling distribution.

Correlation Test

In this paper, a multilevel regression analysis was used to test the hypotheses using SPSSAU software.

As shown in **Table 6**, this stratified regression analysis involved a total of three models. There are three models involved in this hierarchical regression analysis. The independent variables in model 1 are control variables (Gender, Age, Education and Marital Status), model 2 adds perceived of ease of use and perceived usefulness to model 1, and model 3 adds perceived humanity, perceived humanity squared, perceived social interactivity, and perceived social presence to model 2.

The explanatory variable in this study is Trust, model 1 examines the effect of control variables, and an F-test of the model reveals that the model does not pass the F-test ($F = 1.516$, $p > 0.05$). This indicates that the four control variables of Gender, Age, Education, and Marital Status do not have a significant effect on the logical path of Trust.

The results of model 2 showed that the variation in the F-value was significant ($p < 0.05$) after adding perceived of ease of use and perceived usefulness to model 1. This means that perceived of ease of use and Perceived usefulness added explanatory meaning to the model. In addition, the *R*-squared value increased from 0.025 to 0.296, implying that perceived of ease of use and perceived usefulness can explain 27.0% of the strength of trust. In particular, the regression coefficient value for perceived of ease of use is 0.219 and shows significance ($t = 2.772$, $p = 0.006 < 0.01$), which implies that perceived of ease of use has a significant positive relationship with trust. The regression coefficient value for perceived usefulness is 0.287 and shows significance ($t = 4.184$, $p = 0.000 < 0.01$), which implies that perceived usefulness has a significant positive relationship with Trust.

For model 3, the addition of perceived humanity, perceived humanity squared, perceived social interactivity, and perceived social presence to model 2 produced a significant change in the F-value ($p < 0.05$), implying that the addition of Perceived humanity, perceived humanity squared, perceived social interactivity, and perceived social presence explained the significance of the model. In addition, the R-squared value increased from 0.296 to 0.525, implying that perceived humanity, perceived humanity squared, perceived social interactivity, and perceived social presence can explain 23.0% of the strength of Trust. Specifically, the regression coefficient value for perceived humanity was 1.208 and demonstrated significance ($t = 8.100$, $p = 0.000 < 0.01$), implying that perceived humanity can have a significant positive influence on Trust. The regression coefficient value of perceived humanity squared was −0.212 and showed significance ($t = -8.178$, $p = 0.000 < 0.01$), implying that perceived humanity squared would have a significant negative influence on Trust, which is the inverted U relationship between perceived humanity and Trust. The regression coefficient value

TABLE 4 | Scale items and validity tests.

Factor (latent variable)	Measurement items (significant variables)	Standard load factor
Perceived usefulness (CR = 0.906, AVE = 0.708)		
	1. I find that using an AI virtual assistant will improve my daily work performance	0.817
	2. I find that using an AI virtual assistant will help me in my daily work	0.866
	3. I find that using an AI virtual assistant will improve my daily work productivity	0.860
	4. I find that using an AI virtual assistant will be useful for my daily work	0.821
Perceived of ease of use (CR = 0.809, AVE = 0.523)		
	1. I think it will be easy to use the AI virtual assistant	0.778
	2. I find that the interaction with the AI virtual assistant is clear and easy to understand	0.841
	3. I find that the AI virtual assistant is difficult to use	0.533
	4. I find that it is easy to get the AI virtual assistant to do what I want it to do	0.692
Perceived humanity (CR = 0.926, AVE = 0.759)		
	1. I think the AI virtual assistant has a mind of its own	0.874
	2. I think the AI virtual assistant has consciousness	0.933
	3. I think the AI virtual assistant has its own free will	0.900
	4. I think the AI virtual assistant can experience emotions	0.773
Perceived social interactivity (CR = 0.743, AVE = 0.595)		
	1. I think the AI virtual assistant is easy to get along with	0.705
	2. I think the AI virtual assistant can understand me	0.824
Perceived social presence (CR = 0.900, AVE = 0.751)		
	1.There is a sense of interacting with a human being when interacting with an AI virtual assistant	0.843
	2. There is a sense of social interaction with the AI virtual assistant	0.892
	3. There is a sense of humanity in interacting with the AI virtual assistant	0.864
Trust (CR = 0.819, AVE = 0.534)		
	1. In my experience, the AI virtual assistants are honest	0.692
	2. In my experience, the AI virtual assistant cares about the user	0.772
	3. In my experience, the AI virtual assistant provides favorable service	0.688
	4. In my experience, the AI virtual assistant is trustworthy	0.752
AVA (CR = 0.872, AVE = 0.694)		
	1. I will try to use an AI virtual assistant in the future	0.807
	2. I plan to use the AI virtual assistant in the future	0.848
	3. I intend to use the AI virtual assistant in the future	0.841

of perceived social interactivity is 0.176 and shows significance ($t = 2.919$, $p = 0.004 < 0.01$), implying that perceived social interactivity will have a significant positive influence on Trust. The regression coefficient value of perceived social presence was 0.206 and showed significance ($t = 4.174$, $p = 0.000 < 0.01$), implying that perceived social presence will have a significant positive influence on Trust.

As shown in **Table 7**, there were 2 models involved in this hierarchical regression analysis. The independent variables in model 1 are control variables (Gender, Age, Education, and Marital Status), and model 2 adds Trust to model 1.

The explanatory variable in this study is AVA, model 1 examines the effect of control variables, and an F-test of the model reveals that the model does not pass the F-test ($F = 0.432$, $p > 0.05$). This indicates that the four control variables of Gender, Age, Education, and Marital Status do not have a significant effect on the logical path of AVA. The results of model 2 showed that the variation in the F-value was significant ($p < 0.05$) after adding Trust to model 1. This means Trust added explanatory meaning to the model. In addition, the R-squared value increased from 0.007 to 0.224, implying that Trust can explain 21.6% of the strength of trust. In particular, the regression coefficient value for Trust is 0.483 and shows significance ($t = 8.074$, $p = 0.000 < 0.01$), which implies that Trust has a significant positive relationship with AVA.

TABLE 5 | KMO and Bartlett's test.

KMO value		0.914
Bartlett Sphericity test	Approximate cardinality	4155.013
	df	276
	p-value	0

TABLE 6 | Results of multilevel regression tests (Explanatory variable: Trust).

Category	Variable	Model 1	Model 2	Model 3
Control variables	Gender	−0.010 (−0.110)	−0.015 (−0.189)	0.049 (0.734)
	Age	0.052 (1.368)	0.051 (1.575)	0.039 (1.433)
	Education	−0.012 (−0.253)	−0.023 (−0.569)	−0.018 (−0.544)
	Marital status	0.084 (0.929)	0.044 (0.561)	−0.019 (−0.297)
Explanatory variables	Perceived usefulness		0.287** (4.184)	0.112 (1.759)
	Perceived of ease of use		0.219** (2.772)	0.172* (2.495)
	Perceived humanity			1.208** (8.100)
	Perceived humanity squared			−0.212** (−8.178)
	Perceived social interactivity			0.176** (2.919)
	Perceived social presence			0.206** (4.174)
Model explanatory degree	R^2	0.025	0.296	0.525
	Adjusted R^2	0.009	0.277	0.505
	F Value	$[F_{(4,235)} = 1.516, p = 0.198]$	$[F_{(6,233)} = 16.292, p = 0.000]$	$[F_{(10,229)} = 25.343, p = 0.000]$
	ΔR^2	0.025	0.27	0.23
	ΔF Value	$[F_{(4,235)} = 1.516, p = 0.198]$	$[F_{(2,233)} = 44.716, p = 0.000]$	$[F_{(4,229)} = 27.712, p = 0.000]$

*$p < 0.05$ **$p < 0.01$.

TABLE 7 | Multilevel regression test results (Explanatory variable: AVA).

Category	Variable	Model 1	Model 2
Control variables	Gender	−0.013 (−0.140)	−0.019 (−0.224)
	Age	0.040 (1.023)	0.015 (0.441)
	Education	0.046 (0.826)	0.048 (0.976)
	Marital status	−0.028 (−0.303)	−0.066 (−0.813)
Explanatory variables	Trust		0.483** (8.074)
Model explanatory degree	Sample size	240	240
	R^2	0.007	0.224
	Adjusted R^2	−0.01	0.207
	F Value	$[F_{(4,235)} = 0.432, p = 0.785]$	$[F_{(5,234)} = 13.477, p = 0.000]$
	ΔR^2	0.007	0.216
	ΔF Value	$[F_{(4,235)} = 0.432, p = 0.785]$	$[F_{(1,234)} = 65.184, p = 0.000]$

*$p < 0.05$ **$p < 0.01$. *The t-values are in parentheses.*

TABLE 8 | Summary of intermediary role test results.

Items	Total effect	Intermediary effect	95% BootCI	Direct effect	Effectiveness ratio	Test conclusion
Functionality-trust-AVA	0.584**	0.058	0.009∼0.107	0.525** (0.400∼0.655)	10.017%	Partial mediation
Social emotion-trust-AVA	0.079	0.058	0.011∼0.127	0.017 (−0.091∼0.136)	100%	complete mediation

*$p < 0.05$, **$p < 0.01$.

Intermediation Effect Test

In this paper, the bias-corrected nonparametric percentile bootstrap method was applied to test the mediating effect (as shown in **Table 8**), and the confidence level was set at 95%.

In the path from functionality to acceptance, the 95% BootCI was (0.012∼0.109), excluding zero; the trust interval for direct effects was (0.398∼0.652), excluding zero, indicating that it was partially mediated. In the path from social emotion to acceptance, the 95% BootCI was (0.013∼0.129), excluding zero; the direct effect trust interval was (−0.096∼0.130), including zero, indicating that trust plays a full mediating effect in the process of moving from social emotion to acceptance, and hypotheses 4 and 5 are confirmed.

The intermediary role test results were shown in **Figure 2**.

DISCUSSION

This paper develops an AI virtual assistant acceptance model based on the technology acceptance model and the service

FIGURE 2 | Intermediary role test results.

robot acceptance model. Overall, the AVA model extends the potential path of AVA and improves the study of the acceptance transformation mechanism from the trust perspective. The model shows high predictive power for AVA and can well explain the differences in user acceptance and the reasons for the formation of differences. The results of the study can be summarized in three aspects as follows.

First, this paper verifies the positive effect of trust on the acceptance of AI virtual assistants. The services of AI virtual assistants are executed based on AI algorithms, but users do not fully trust the information or services provided by AI virtual assistants due to the inherent black-box problem of AI technology. Existing studies suggest that trust is an essential driver of technology acceptance and can positively influence users' acceptance of new technologies (Kaplan and Haenlein, 2019; van Pinxteren et al., 2019; Asatiani et al., 2020). This paper confirms through empirical research that trust is significantly and positively correlated with AI virtual assistant acceptance, and that its ability to reduce users' negative emotions toward AI virtual assistants plays a key role in improving AI virtual assistant acceptance.

Second, this study explored the relationship with trust at the functionality and social emotional levels. Most of the existing studies focus on the relationship between functionality and acceptance (King and He, 2006), but there is a lack of research on the relationship with trust. Based on the technology acceptance model, this paper examines the effect of functionality on trust in terms of two dimensions, perceived usefulness and perceived of ease of use, which confirmed that both usefulness and ease of use dimensions were significantly and positively related to trust. This shows that efficient service experience helps users develop trust, which means that effective services of AI virtual assistants will encourage users to increase their interactions with them and thus develop trust based on familiarity with their functions.

In addition, the degree of user trust in AI virtual assistants depends on their ability to satisfy users' social-emotional and relational needs. Drawing on the service robot acceptance model, this paper divides social emotion into three dimensions: perceived humanity, perceived social interactivity, and perceived social presence. Currently, there are two different views on perceived humanity. Social reaction theory and social presence theory believe that higher anthropomorphism will lead to a positive customer response, which means that perceived humanity will lead to user trust (Qiu and Benbasat, 2009; van Doorn et al., 2017). Conversely, some scholars have argued that the positive effects of highly anthropomorphic AI virtual assistants fail to be proven in many scenarios and can even increase users' negative emotions (Tinwell et al., 2011; van Doorn et al., 2017). Consequently, this paper explores the relationship between perceived humanity and trust through empirical tests and finds a non-linear inverted U-shaped relationship between perceived humanity and trust, meaning that only moderate perceived humanity can promote trust among users. In addition, the maturation of AI technology will enable AI virtual assistants to have certain human-like attributes, such as voice, real-time interaction, verbal communication skills, and social etiquette, that enable human users to perceive their social presence. Scholars believe that these attributes can generate positive emotions and establish favorable social relationships, which can enhance the interaction between users and AI virtual assistants to help increase their trust level (Fernandes and Oliveira, 2021). This study further confirmed that perceived social interactivity and perceived social presence were positively related to trusting behavior, which is consistent with the findings of existing studies.

Finally, this paper explores the mediating role of trust between functionality and social emotion and acceptance. Trust motivates AI virtual assistants to build a favorable image and suppresses users' perceptions of various risks, which in turn positively motivates users' acceptance behavior toward them. According to Schmitt (1999), customers' purchasing behavior is the result of a combination of rational and emotional factors. Drawing on the customer delivered value theory, AI virtual assistants should make every effort to provide customers with quality services, obtain customer satisfaction, and help customers generate willingness to use. AI virtual assistants that provide favorable services give users a comfortable experience, which generates a brand effect, increases user loyalty and trust, and forms a positive cycle. This paper confirms the mediating role of trust between functionality and acceptance, with the service experience being the bridge between users and AI virtual assistants to build a well-trusted relationship. The social emotions of AI virtual assistants enable users to

create sublimation and association with social systems, which leads to satisfaction at the psychological level. Trust is the prerequisite for user identification and the key factor that determines whether users are willing to interact deeply with the information source (Wirtz et al., 2018). Based on the above studies, this paper confirms the mediating role of trust between social emotion and acceptance, which is important for enhancing the contextualized services of AI virtual assistants from a perceptual perspective.

In summary, this study establishes a new acceptance model for AI virtual assistants, verifying the inverted U-shaped effect of perceived humanity on trust and the mediating role of trust in the acceptance transformation mechanism. It complements the gap of existing technology acceptance models at the trust level and expands the transformation path of AVA. To a certain extent, it extends the boundary and application space of existing theories and helps to solve the user acceptance problem from the trust perspective.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

SZ and ZM: writing. BC and XY: providing revised advice. XZ processing data. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Social Science Foundation of Heilongjiang Province of China (18TQC239), Guangxi Science and Technology Plan Project (Guangxi Science and Technology Base and Talent Special Project: AD20159069).

REFERENCES

Asatiani, A., Malo, P., Nagbøl, P. R., Penttinen, E., Rinta-Kahila, T., and Salovaara, A. (2020). Challenges of explaining the behavior of black-box AI systems. *MIS Q. Executive* 19, 259–278. doi: 10.17705/2msqe.00037

Chattaraman, V., Kwon, W., Gilbert, J., and Ross, K. (2019). Should AI-Based, conversational digital assistants employ social-or task-oriented interaction style? a task-competency and reciprocity perspective for older adults. *Comput. Hum. Behav.* 90, 315–330. doi: 10.1016/j.chb.2018.08.048

Epley, N., Akalis, S., Waytz, A., and Cacioppo, J. (2008). Creating social connection through inferential reproduction. *Psychol. Sci.* 19, 114–120. doi: 10.1111/j.1467-9280.2008.02056.x

Fernandes, T., and Oliveira, E. (2021). Understanding consumers' acceptance of automated technologies in service encounters: drivers of digital voice assistants adoption. *J. Bus. Res.* 122, 180–191. doi: 10.1016/j.jbusres.2020.08.058

Glikson, E., and Woolley, A. W. (2020). Human trust in artificial intelligence: review of empirical research. *Acad. Manag. Ann.* 14, 627–660. doi: 10.5465/annals.2018.0057

Gummerus, J., Lipkin, M., Dube, A., and Heinonen, K. (2019). Technology in use-characterizing customer self-service devices. *J. Serv. Mark.* 33, 44–56. doi: 10.1108/JSM-10-2018-0292

Gursoy, D., Chi, O. H., Lu, L., and Nunkoo, R. (2019). Consumers acceptance of artificially intelligent (AI) device use in service delivery. *Int. J. Inf. Manage* 49, 157–169. doi: 10.1016/j.ijinfomgt.2019.03.008

Hassanein, K., and Head, M. (2007). Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. *Int. J. Hum. Comput. Stud.* 65, 689–708. doi: 10.1016/j.ijhcs.2006.11.018

Holzwarth, M., Janiszewski, C., and Neumann, M. M. (2006). The influence of avatars on online consumer shopping behavior. *J. Mark.* 70, 19–36. doi: 10.1509/jmkg.70.4.019

Jan, A. U., and Contreras, V. (2011). Technology acceptance model for the use of information technology in universities. *Comput. Hum. Behav.* 27, 845–851. doi: 10.1016/j.chb.2010.11.009

Jiang, C., Rashid, R. M., and Wang, J. (2019). Investigating the role of social presence dimensions and information support on consumers' trust and shopping intentions. *J. Retail. Consum. Serv.* 51, 263–270. doi: 10.1016/j.jretconser.2019.06.007

Kaplan, A., and Haenlein, M. (2019). Siri, siri in my hand, who is the fairest in the land? on the interpretations, illustrations and implications of artificial intelligence. *Bus. Horiz.* 62, 15–25. doi: 10.1016/j.bushor.2018.08.004

Kim, H. Y., and McGill, A. L. (2018). Minions for the rich? financial status changes how consumers see products with anthropomorphic features. *J. Consum. Res.* 45, 429–450. doi: 10.1093/jcr/ucy006

King, W. R., and He, J. (2006). A meta-analysis of the technology acceptance model. *Inform. Manag.* 43, 740–755. doi: 10.1016/j.im.2006.05.003

Kulviwat, S., Bruner, I. I., G. C., Kumar, A., Nasco, S. A., and Clark, T. (2007). Toward a unified theory of consumer acceptance technology. *Psychol. Mark.* 24, 1059–1084. doi: 10.1002/mar.20196

McLean, G., and Osei-Frimpong, K. (2019). Hey Alexa… examine the variables influencing the use of artificial intelligent in-home voice assistants. *Comput. Hum. Behav.* 99, 28–37. doi: 10.1016/j.chb.2019.05.009

Palmatier, R. W., Dant, R. P., Grewal, D., and Evans, K. R. (2006). Factors influencing the effectiveness of relationship marketing: a meta-analysis. *J. Mark.* 70, 136–153. doi: 10.1509/jmkg.70.4.136

Qiu, L., and Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: a social relationship perspective to designing information systems. *J. Manag. Inform. syst.* 25, 145–182. doi: 10.2753/MIS0742-1222250405

Schmitt, B. (1999). Experiential marketing. *J. Mark. Manag.* 15, 53–67. doi: 10.1362/026725799784870496

Shin, D. (2020). How do users interact with algorithm recommender systems? the interaction of users, algorithms, and performance. *Comput. Hum. Behav.* 109:106344. doi: 10.1016/j.chb.2020.106344

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *Int. J. Hum. Comput. Stud.* 146:102551. doi: 10.1016/j.ijhcs.2020.102551

Song, J. H., and Zinkhan, G. M. (2008). Determinants of perceived web site interactivity. *J. Mark.* 72, 99–113. doi: 10.1509/jmkg.72.2.99

Sparks, B. A., and Browning, V. (2011). The impact of online reviews on hotel booking intentions and perceptions of trust. *Tour. Manag.* 32, 1310–1323. doi: 10.1016/j.tourman.2010.12.011

Sundar, S. S. (2020). Rise of machine agency: a framework for studying the psychology of human–AI interaction (HAII). *J. Comput. Mediat. Commun.* 25, 74–88. doi: 10.1093/jcmc/zmz026

Tinwell, A., Grimshaw, M., Nabi, D., and Williams, A. (2011). Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Comput. Hum. Behav.* 27, 741–749. doi: 10.1016/j.chb.2010.10.018

van Doorn, J., Mende, M., Noble, S., Hulland, J., Ostrom, A., Grewal, D., et al. (2017). Domo Arigato Mr. Roboto: emergence of automated social presence in organizational frontlines and customers' service experiences. *J. Serv. Res.* 20, 43–58. doi: 10.1177/1094670516679272

van Pinxteren, M., Wetzels, R., Rüger, J., and Wetzels, M. (2019). Trust in humanoid robots: implications for services marketing. *J. Serv. Mark.* 33, 507–518. doi: 10.1108/JSM-01-2018-0045

Venkatesh, V., and Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decis. Sci.* 39, 273–315. doi: 10.1111/j.1540-5915.2008.00192.x

Venkatesh, V., and Davis, F. D. (2000). A theoretical extension of the technology acceptance model: four longitudinal field studies. *Manage. Sci.* 46, 186–204. doi: 10.1287/mnsc.46.2.186.11926

Wang, Y. D., and Emurian, H. H. (2005). An overview of online trust: concepts, elements, and implications. *Comput. Hum. Behav.* 21, 105–125. doi: 10.1016/j.chb.2003.11.008

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., et al. (2018). Bravenew world: service robots in the frontline. *J. Serv. Manag.* 29, 907–931. doi: 10.1108/JOSM-04-2018-0119

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhang, Meng, Chen, Yang and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ERRATUM
published: 26 October 2021
doi: 10.3389/fpsyg.2021.790448

Erratum: Motivation, Social Emotion, and the Acceptance of Artificial Intelligence Virtual Assistants—Trust-Based Mediating Effects

*Frontiers Production Office**

Frontiers Media SA, Lausanne, Switzerland

Keywords: AI virtual assistant, motivation, social emotion, trust, acceptance, mediating effects, inverted U relationship

OPEN ACCESS

Approved by:
Frontiers Editorial Office, Frontiers Media SA, Switzerland

***Correspondence:**
Frontiers Production Office
production.office@frontiersin.org

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *06 October 2021*
Accepted: *06 October 2021*
Published: *26 October 2021*

Citation:
Frontiers Production Office (2021) Erratum: Motivation, Social Emotion, and the Acceptance of Artificial Intelligence Virtual Assistants—Trust-Based Mediating Effects. Front. Psychol. 12:790448. doi: 10.3389/fpsyg.2021.790448

An Erratum on

Motivation, Social Emotion, and the Acceptance of Artificial Intelligence Virtual Assistants—Trust-Based Mediating Effects
by Zhang, S., Meng, Z., Chen, B., Yang, X., and Zhao, X. (2021). Front. Psychol. 12:728495. doi: 10.3389/fpsyg.2021.728495

Due to a production error, an incorrect funding statement was inserted into the article.

A correction has been made to the **Funding** section.

"This work was supported by the Social Science Foundation of Heilongjiang Province of China (18TQC239), Guangxi Science and Technology Plan Project (Guangxi Science and Technology Base and Talent Special Project: AD20159069)."

The publisher apologizes for this mistake. The original version of this article has been updated.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Frontiers Production Office. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 19 August 2021
doi: 10.3389/fpsyg.2021.730066

Visual Expression of Emotion in Dynamic 3D Painting System Based on Emotion Synthesis Model

*Shenghe Cheng**

College of Art, Nanjing University, Nanjing, China

Emotion is a unique ability possessed by human beings as advanced creatures. Emotions give people a unique physical and mental experience. Assigning emotions to computer systems is one of the latest topics in artificial intelligence research. The purpose is to allow machines to achieve natural coordination between humans and computers. This article focuses on the visual expression of emotion in the dynamic three-dimensional painting system, creating an intelligent painting system and realizing a good user experience. In this paper, the discrete method is used to qualitatively analyze emotions, and the continuous method is used to quantify basic emotions, and emotional modeling and emotional quantitative analysis are proposed to realize quantitative analysis of emotions. Combining these two methods, a comprehensive method is proposed, which uses a continuous method to quantify the basic emotions of each discrete dimension, and finally superimposes them into a comprehensive emotional synthesis model. Emotion modeling is the basis of emotion visualization. Borrowing the relationship between emotion synthesis model and visual emotion elements, this article puts forward the concept of qualitative and quantitative visual emotion elements, and expounds that the multidimensional superposition of visual emotion elements makes dynamic three-dimensional painting system emotions. The experimental results in this article show that the emotional visualization scheme of 100 samples is tested by quantitative statistical methods to demonstrate its effectiveness. Starting from 5 points of concern, the emotion visualization method discussed in this article can indeed convey or suggest a certain positive emotion (the average value of experience, transitivity, and infectiousness $>$ 2.5, and the variance is close to 0), but we also found this recognition at the same time The degree is not high enough, and individual differences are large (mean value $<$ 2.5, variance close to 1). This can indicate that different subjects have different feelings and evaluations of this emotional visualization. As long as the difference is within a reasonable range, this emotional visualization also has practical value, and has the ability to convey or suggest emotions.

Keywords: emotion synthesis model, dynamic three-dimensional painting system, emotion visualization, discrete method, emotional synthesis model

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Wei Wen,
Hainan University, China
Jianbo Wang,
Central South University, China

***Correspondence:**
Shenghe Cheng
chengshenghe@jit.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *24 June 2021*
Accepted: *28 July 2021*
Published: *19 August 2021*

Citation:
Cheng S (2021) Visual Expression of Emotion in Dynamic 3D Painting System Based on Emotion Synthesis Model. Front. Psychol. 12:730066. doi: 10.3389/fpsyg.2021.730066

INTRODUCTION

Background

In recent years, computer networks have developed rapidly and have relatively mature functions, which have produced many amazing applications. However, current computers still have some shortcomings. For example, computers cannot make optimal decisions like humans. Although computers have huge computing power and storage capabilities, they lack the ability to perceive or express emotions, that is, a certain human emotional mechanism (Tang and Elhoseny, 2019). Therefore, in order to better realize the human-computer interaction, strengthen the emotional communication between users and the system in the dynamic 3D painting system, use the dynamic 3D painting system as the research object, and try to model abstract emotions through emotional calculations, effectively expressed in a dynamic three-dimensional painting system.

Significance

With the advancement of science and technology and the development of living materials, people's requirements for smart computers are getting higher and higher. If a computer has functions similar to human emotion mechanisms, then it will be smarter and more suitable for interacting with users. For example, a computer that can sense the user's drowsiness should not endlessly instill more information to the user. If computers can eliminate the estrangement between people and have a certain emotional mechanism like normal people, human-computer interaction will be another scene, just like a universal housekeeper. In practical applications, this is exactly what the research fields including artificial intelligence and pervasive computing want to achieve, and it is also what many website designs are trying to achieve today (Wang and Lu, 2019). By modeling emotions, computers can learn to analyze and express human emotions, have a certain emotional mechanism, eliminate obstacles between computers and humans, and provide better services. Therefore, a three-dimensional website with an emotional mechanism will greatly enhance the user experience of the website and enhance the superior experience of human-computer interaction.

Related Work

Emotion is considered to be the basic element of human-computer interaction performance. In expressive synthesized speech, it is important to generate emotional speech that reflects subtle and complex emotional states (Liu and Fu, 2021). However, the research on how to effectively synthesize emotional speech through intuitive control using different levels of emotional intensity is limited, and it is difficult to model effectively. Zhu and Xue (2020) explored an expressive speech synthesis model that can be used to generate speech with a variety of emotional strengths. Different from the previous study of encoding emotions into discrete codes, he proposed an embedded vector to continuously control the intensity of emotions. This is a data-driven method that can synthesize speech through fine control of emotions. Compared with models using retraining techniques or one-hot vectors, his model using embedding vectors can clearly learn high-level emotional intensity from low-level acoustic features. Therefore, he can control the emotional intensity of synthesized speech (Zhu and Xue, 2020). When people face internal or external stimuli, emotions are a subjective and conscious experience. Qiu and Zhao (2018) research solves the problem of emotional computing that is difficult to intuitively apply to practical areas in the real world due to unintuitive data representation, such as the diagnosis of emotional diseases. Given that people's ability to understand two-dimensional images is much higher than that of one-dimensional data, he uses Markov transition fields to visualize time series signals. The MTF image represents the first-order Markov transition probability along one dimension and the time dependence along another dimension. In addition, due to the limitations of experimental equipment and individual differences between volunteers, noise is inevitable. He applied AC-GAN to remove noisy pixels in high dimensions and obtain high-resolution images before classification. Then he used Tiled Convolutional Neural Networks on 2 real-world data sets to learn advanced features from MTF images. The classification results of other methods are competitive with the most advanced methods. This method makes it possible to recognize emotions based on visualization, which is conducive to the application of cognitive robots or the diagnosis of psychological problems such as depression in the medical field, and can help doctors and patients understand the condition more intuitively (Qiu and Zhao, 2018). Yu et al. (2017) believes that the emotional appeals that prevailed in charts and graphs in the late nineteenth century have basically been dormant since then and quickly reappeared in contemporary data visualization. This new form of data design changes the relationship between designers and users, and strengthens the emotional impact of the data by triggering emotions ranging from excitement and empathy to anxiety and fear. Yu et al. (2017) provided a historical overview of the tragic appeal in data design in the late nineteenth century and the transition to modernist minimalism in the twentieth century. Contemporary examples from companies, non-profit organizations, government agencies, and individual designers illustrate how data visualization can stimulate emotions, and the results of the research show that the emotional appeal of the nineteenth century is concentrated (Yu et al., 2017). Although the research perspective is forward-looking, there are still some shortcomings in research techniques and methods.

Innovation

The innovation of this article lies in (1) By acquiring emotional data, an emotional synthesis model is established to effectively analyze and express emotions. Calculate emotions qualitatively and quantitatively, and embody abstract emotions. It is measured by certain variables (such as prompts and incentives) and induces variables. (2) Based on the relationship between the emotion synthesis model and visual emotion elements, explore ways to visualize emotions. The concept of qualitative and quantitative visual emotion elements is put forward, and the multi-dimensional superposition of visual emotion elements enables the three-dimensional dynamic website space to express complex emotions.

RESEARCH FOUNDATION

The Mechanism of Emotion Generation

The key to the mechanism of emotion generation is the physiological and psychological mechanisms, including the neural stimulation layer, the emotional state layer and the psychological cognitive layer. The three are closely related, that is, under the influence of nerve stimulation and hormones, the corresponding emotional response is generated, and the body works together Produce emotion (Ridderinkhof, 2017). Through the study of the basic mechanism of human emotions, the understanding of human emotion mechanisms, and the use of direct emotional stimulation and information injection of mental cognitive abilities, will help us to further study how to concretize and model complex and abstract emotions. Can design a dynamic three-dimensional painting system that can stimulate people's emotions. Emotional audio-visual elements enable us to inject emotional information into user consciousness to achieve emotional visualization (Huxtable-Thomas et al., 2016). Visual and auditory stimulation is the most effective way to stimulate the cognitive level. For example, seeing a black room feels depressed, and hearing a joke feels joy. We can use emotional audio-visual elements to try to create a dynamic three-dimensional painting system that visualizes emotions. The mechanism of emotion generation is shown in **Figure 1**.

Emotion Synthesis Model

Modeling Process

Psychologists define the various possible factors that can cause changes in human emotions as evaluation factors (including the degree of expectation, relevance, adaptability, causes, etc.), and put forward the theory of emotional evaluation. It is believed that individuals can only produce emotions by cognitive evaluation of stimuli. There are also personality psychological structures, such as interest, motivation, and personality characteristics, which also have an impact on the generation of emotions (Topal and Ozsoyoglu, 2017). Individuals of different personalities have different evaluations of the same stimulus. The resulting emotions are also different. In addition, psychological experiments prove that when the emotional intensity reaches a certain level. Even if there are constant external stimuli, the intensity will not continue to increase, that is, emotions will reach saturation, especially after some basic emotions are activated, even if there is no external stimulus, they will not disappear immediately. It will gradually decay over time (Ren and Matsumoto, 2017). This article is based on the motivation theory and cognitive evaluation of human emotion. It not only considers the impact of cognitive factors on emotions, but also non-cognitive factors (such as physiological needs, personality factors, the characteristics of emotions themselves, the mutual influence between emotions, etc.). Impact on emotions. The construction process is shown in **Figure 2**.

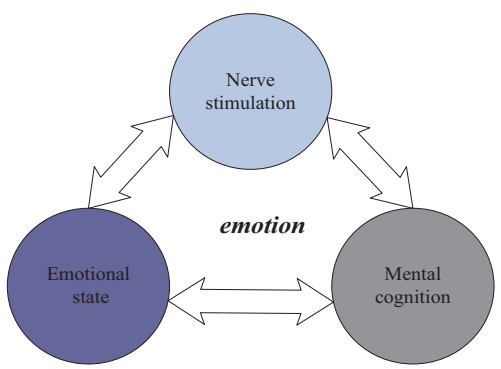


FIGURE 1 | The mechanism of emotion generation.

Basic Principles

Discrete method of emotional synthesis. There are many ways to divide emotions in psychology (Eyben et al., 2017). Researchers have different definitions of emotion category labels, and **Table 1** summarizes them.However, using discrete methods to construct complex emotions also has its limitations. On the premise of our six basic emotions, happiness, sorrow, sorrow, happiness, love, and evil, as shown in **Figure 3**, human emotions are divided into three dimensions: happiness, sorrow, joy, and love and evil, so as to achieve Qualitative analysis of emotions.

Assuming that it represents the emotional dimension of happiness and worry, when j = 1, 0, −1 represents the dimension of happiness, without the dimension of anger and the dimension of anger, s represents the dimension of sadness and happiness, and h represents the dimension of love and evil. The combination of

$$B = j \cup s \cup h, \left(j, s, h \in \{-1, 0, 1\}\right) \tag{1}$$

In the above formula, B represents emotion; j, s, h represent various emotional dimensions. However, it is impossible for human emotions to be interpreted using such a simple model. We know that human emotions are extremely rich and cannot be fully described by a simple arrangement of 27 emotions.

The dimensional method of emotion synthesis, also known as the continuous method of emotion synthesis, realizes the quantitative analysis of emotions (Radojevic, 2016). Due to the complexity of emotion generation and expression, many scholars have studied how to model emotions. Two three-dimensional emotional models are introduced here. The first is that the psychologist Wundt first proposed the theory of Emotion dimension. He researched and proposed that human emotions are composed of three dimensions, and different emotions will have two opposite directions in each dimension. Constantly making changes, namely excitement to inhibition (excitement-inhibition), pleasant to unpleasant (pleasure-displeasure), tension-relaxation (tension-relaxation) (Zhou et al., 2020). As shown in **Figure 4**.

The second type, in 1974, Mihrabian and Russell proposed a dimensional observation model-PAD three-dimensional emotional model. Take **Figure 5** as an example. The three-dimensional emotion model of PAD is a kind of pleasure Pleasure, activation degree Arousal, dominance degree, and it divides emotion into three dimensions. Pleasure degree

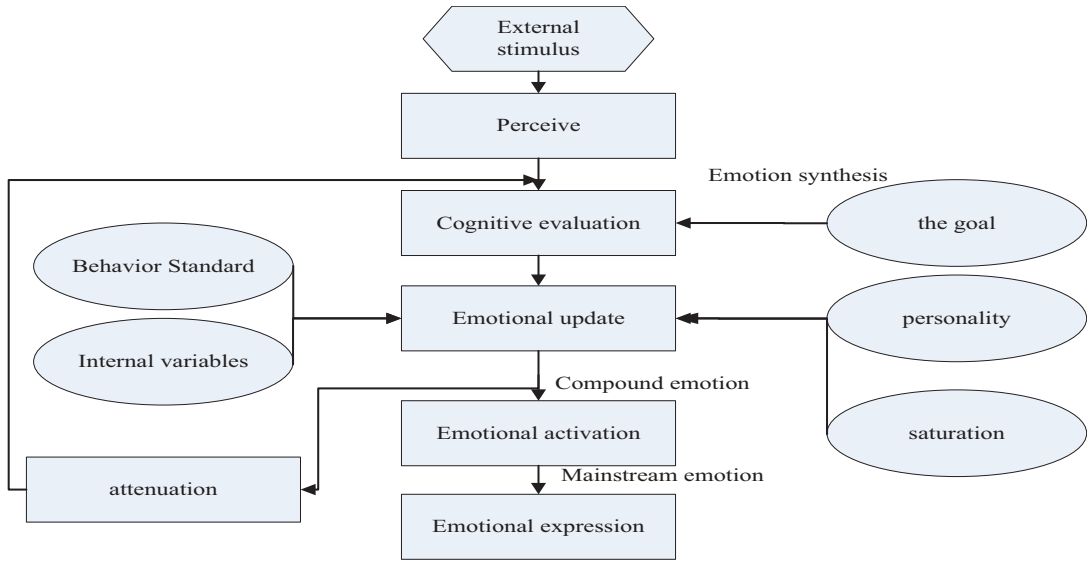


FIGURE 2 | Modeling process of emotion synthesis model.

P describes the positive and negative characteristics of an individual's emotional state, activation degree A describes the physiological activity of the individual's nerves, and specifically quantifies the calmness and excitement of emotions, while dominance D describes the individual's control over others and the situation, reflecting emotional control and control. The degree of Kim et al. (2016) and Hu et al. (2019).

Both discrete emotion model and dimensional emotion model have advantages and disadvantages. The two emotion synthesis models are compared (Ueoka and AlMutawa, 2018). As shown in **Table 2**:Comprehensive emotional synthesis. Combining the above two methods, a comprehensive method is proposed, which uses a continuous method to quantify the basic emotions of each discrete dimension, and finally superimposes it into an emotion synthesis model (Kostelnick, 2016; Volante et al., 2016). The formula of the comprehensive emotional model is as follows:

$$E = f_{a-b}(u_1, v_1) \cup f_{c-d}(u_2, v_2) \cup f_{e-g}(u_3, v_3),$$
$$-1 \le u \le 1, 0 \le v \le 1 \quad (2)$$

TABLE 1 | Different researchers' definitions of emotion category labels.

Researcher	Emotion category label
Ekman	Angry, disgusted, frustrated, happy, sad, surprised
Fridja	Desire, happiness, concern, surprise, curiosity, sorrow
Plutchik	Anger, identification, expectation, happiness, disgust, fear, sadness, surprise
Oatley	Anger, disgust, anxiety, happiness, sadness
Watson	Fear, love, anger
Graham	Happy, sad

The function formula under the joy and worry dimension:

$$f_{a-b}(u_1, v_1) \quad (3)$$

The function formula under the sorrow dimension:

$$f_{c-d}(u_2, v_2) \quad (4)$$

Function formula under the dimension of love and evil:

$$f_{e-g}(u_3, v_3) \quad (5)$$

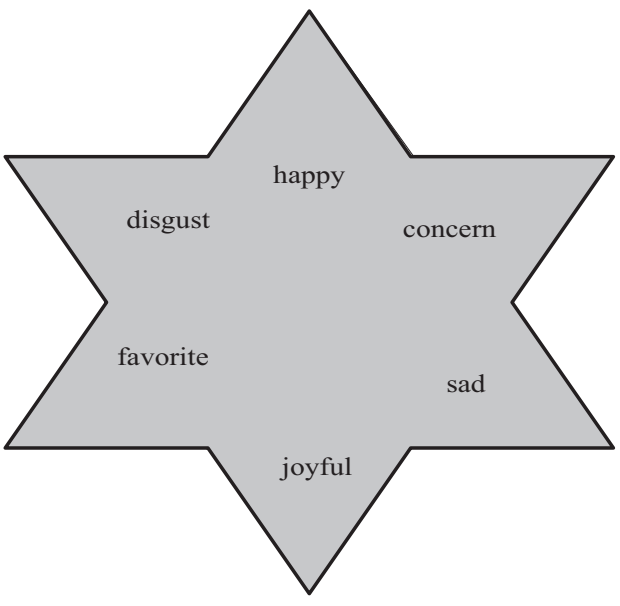


FIGURE 3 | Discrete emotion synthesis model.

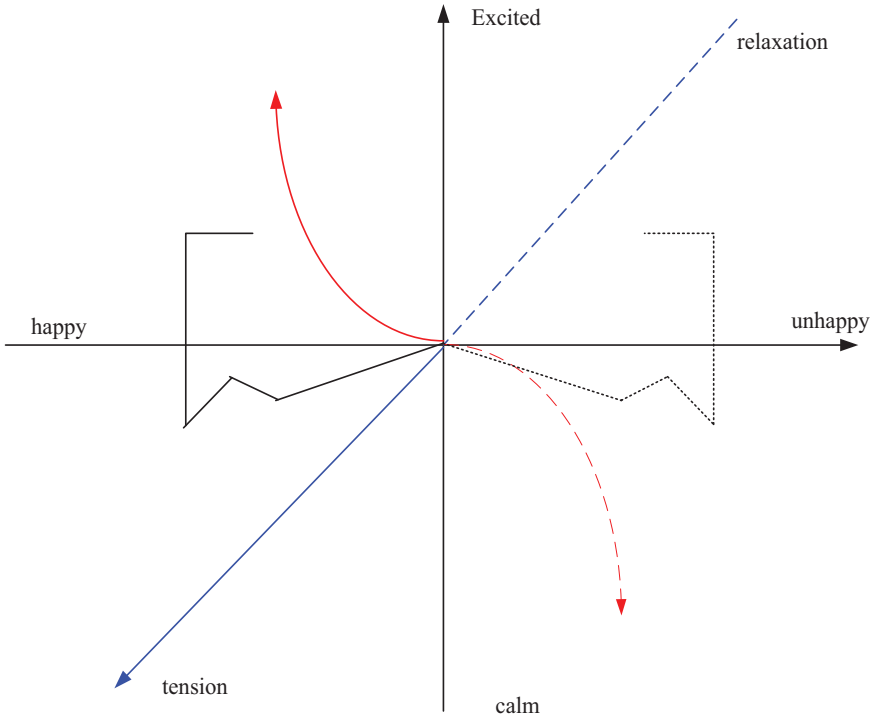


FIGURE 4 | Wundt's three-dimensional emotional model.

In the above formula, E represents emotion, U represents the incentive value of emotion, and V represents the induced value of emotion. The process of the value of u transitioning from −1 to 1 represents the process of transitioning from a negative emotion to a positive emotion, and the process of transitioning the value of v from 0 to 1 is also the process of the emotional activation degree rising to the full activation state (Huang, 2018; Wang et al., 2019).

The realization of the emotion synthesis model requires us to determine the basic attributes of each discrete dimension of emotion and how to combine the attributes of these discrete emotions. One of the most important steps to concretely realize this emotion model is that we need to set it based on the current visual emotion element under study. Here we set W1 as the value of the corresponding visual emotion element attribute on the basic emotion i1. At the same time, we define Si1 as a constant related to the basic emotion i1, which defaults to the value set during the emotional qualitative process (Kostelnick, 2016). We need to define some basic attributes of the emotional dimension,

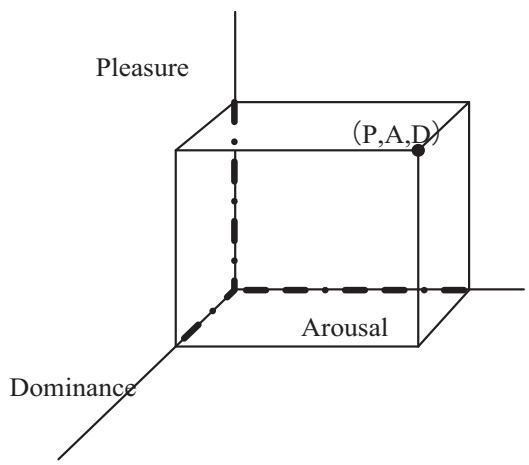


FIGURE 5 | PAD three-dimensional emotional model.

TABLE 2 | Comparison of two emotion synthesis models.

	Discrete emotion model	Continuous emotion model
Model complexity	Concise and easy to understand	More complicated
The way of emotional description	Independent adjective label	Coordinate point in space
The ability of emotional description	Single description with limitations	Arbitrarily complex emotion types
Advantage	Simple model, easy to start	Strong emotional description ability
Disadvantage	Limited emotional labels are difficult to satisfy	The process of quantifying complex emotions into specific values
	Description of mixed continuous emotions	Is arduous and the quality is difficult to guarantee

and the identification of these basic attributes is derived from the actual human emotions. Human emotions are composed of many basic emotions. However, the strength of each basic emotion is often different. That is, we need to assign a weight attribute to each basic emotion. Here we set the variable Zi1 as The weight attribute of the basic emotion. After defining many variables of the basic emotion of appeal, we need a way to mix emotions of various dimensions to obtain a model of complex emotions (Fortin and Cooperstock, 2017). For the basic emotions i1 and i2, then there are the following formulas:

$$i = f(s, u, v) = \begin{bmatrix} S_{i1}S_{i2} \\ a_{i1}a_{i2} \\ v_{i1}v_{i2} \end{bmatrix} \bullet \begin{bmatrix} \frac{Z_{i1}}{Z_{i1}+Z_{i2}} \\ \frac{Z_{i2}}{Z_{i1}+Z_{i2}} \end{bmatrix} \tag{6}$$

In the formula, i represents the composite emotion of two emotional dimensions; S represents the constant related to emotion i; Z represents the weight of emotion i. This kind of emotion synthesis method can balance the emotions of different dimensions according to their weights, thereby synthesizing the corresponding complex emotions. According to the synthesized value, we can determine the synthesized value of the corresponding emotion, and then apply it to the visual emotion elements. Above, we can set the attributes of a single visual emotion element deterministically (Bringmann et al., 2016). For synthetic emotion, its weight satisfies the following formula:

$$Z_i = Z_{i1} + Z_{i2} \tag{7}$$

Z represents the weight of emotion i, where the formula for calculating emotion intensity is as follows:

joy:

$$M_{JOY} = \left|1.7 \times D \times E^{0.5} - (0.7 \times D)\right| \tag{8}$$

sad:

$$M_{distress} = \left|2 \times D \times E^2 - D\right| \tag{9}$$

hope:

$$M_{hope} = \left|1.7 \times D \times E^{0.5} - (0.7 \times D)\right| \tag{10}$$

worry:

$$M_{fear} = \left|2 \times D \times E^2 - D\right| \tag{11}$$

relief:

$$M_{relief} = \left|fear \times D\right| \tag{12}$$

Disappointed:

$$M_{disappointment} = \left|hope \times D\right| \tag{13}$$

In the formula, M represents the intensity of emotion, D is desirability, the satisfaction value, E is expectation, the expected value, and the expected value of the event is affected by two factors, namely, the impact of the event on the target and the importance of the target. The impact of the event on the target uses a fuzzy set {Extremely negative impact, relatively negative impact, no impact, relatively positive impact, extremely positive impact}, the domain value is [−1, 1], where 1 represents extremely negative impact, and 1 represents extremely positive impact. Goal The importance of is also described by a fuzzy set {not important, slightly important, obviously important, extremely important, and the domain value is [0, 1], where 0 represents unimportant, and 1 represents the expected value of extremely important events. The same applies to fuzzy sets {Extremely Unexpected, Less Expected, Neutral, More Expected, Extremely Expected} to describe, the domain value is [−1, 1], where −1 represents extremely undesirable, and 1 represents extremely expected (Liu et al., 2016).

The emotional visualization process model is shown in **Figure 6**:

The first step is to collect information, and the second step is to preprocess the information to extract the multi-dimensional attributes and emotional characteristics of the dynamic 3D painting system data. The third step uses the computational model to conduct a scientific system analysis of the dynamic 3D painting system. The fourth step is to perform visualization operations based on a specific data model, and the fifth step is to visualize emotions, including showing the changing trends of group users' attitudes toward dynamic 3D painting systems under different characteristics for individuals and society.

The Mechanism of Visual Influence on Emotion

Emotional changes are produced in the process of interaction between visual elements and the subject. Therefore, the key to emotional visualization is that the visual expression elements and expression forms of emotion are derived from the influence of visual information on emotions. After a long period of human evolution, the visual system has evolved a variety of neural mechanisms to adapt to the surrounding environment, each of which can efficiently process specific visual information. These specific visual information include position, form, texture, and movement (Kim et al., 2016; Zhang X. et al., 2020). In addition, the processing of visual information has the characteristic of filtering, which is accompanied by the immediate focus of human attention. The processing of visual information is carried out as a whole, and the combination of all the characteristics of visual perception and the characteristics of visual filtering constructs a visual perception that conforms to human psychology and habits. Psychology believes that the physiological principle of emotion is a person's stressful body reaction directly caused by external stimuli, and this body reaction is directly produced by the nervous system. Not only that, vision is an important way for humans to obtain cognitive information and an important basis for human cognitive ability, which is also an important source of emotion. Therefore, the influence of visual information on emotion is obvious.

The current computer does not have the human emotion mechanism and the ability to think, and the construction of our three-dimensional website on the computer is also limited by this. Therefore, it is necessary to determine the basic emotional characteristics of these visual elements on the basis of the emotional synthesis model. What needs to be explained is that only when multiple visual emotional elements are combined

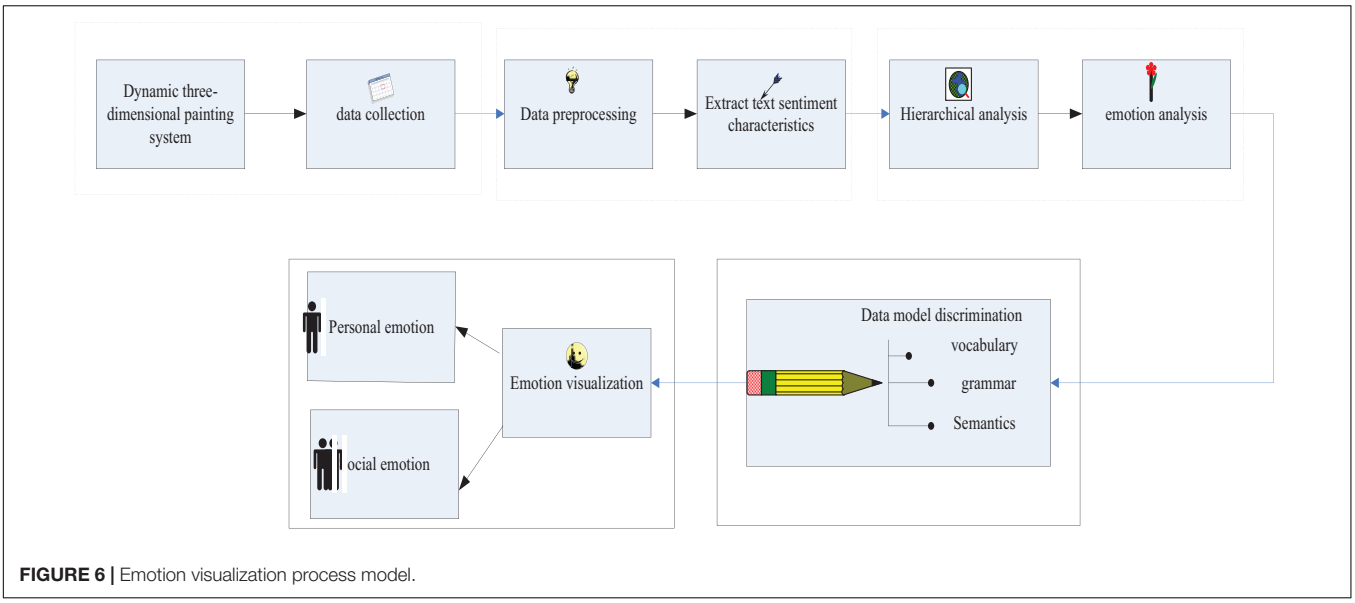


FIGURE 6 | Emotion visualization process model.

to express emotions can we get closer to expressing emotions. Such as material elements, sports elements, color elements, styling elements, language elements, etc. Before realizing the construction of visual emotional elements, we need to define the visual characteristics of emotions, that is, the qualitative visual characteristics of emotional elements (Andrew and Larceneux, 2019; Zhang Y. et al., 2020). For example, blue represents sad emotions, red represents love emotions, and cyan represents evil emotions. Secondly, it is necessary to determine the specific value of the emotional characteristics of visual elements under different inducing forces and incentive values, that is, the emotional quantification of visual elements. For example, setting blue is the qualitative color of the emotional element of sadness. However, it is impossible for all sadness to be pure blue. We must specifically combine the emotional synthesis model, so that we can more accurately determine the specificity of this blue. The color value may eventually be found to be biased toward cyan. After that, the visual emotion elements of all dimensions are multi-dimensionally superimposed and calculated to finally generate the final dynamic 3D painting system, which is the final synthesis after qualitative and quantitative emotions, and we obtain a 3D emotional space of complex emotions (Topal and Ozsoyoglu, 2017).

EXPERIMENTAL PROCESS

Purpose of the Experiment

The purpose of studying the emotional visualization expression method of the dynamic three-dimensional painting system is to convey emotions to the users of the system, and ultimately improve the user experience of the system, and propose corresponding solutions. Here, we specifically propose a visual expression method for emotions in this dynamic three-dimensional painting system. Verify that the method achieves our expected goals: First, does this emotional suggestion improve the user experience of the system? Second, whether the corresponding visual emotion element implies an emotional tendency.

Subjects

This experiment is the first optimization screening of the visual expression method of emotion in the dynamic three-dimensional painting system. Using the method of random sampling, 100 college students were recruited to participate in this experiment (51 boys and 49 girls), aged between 18 and 25 years old, in good health, no defects such as color blindness, no mental illness, and similar Has a knowledge background, has a full understanding of new things, and is willing to accept new ideas and things. All subjects used individual tests and were divided into four groups according to different grades. After the experiment, the data was checked on the spot to ensure the validity of the experimental data.

Experimental Method

The specific detection method is to first select multiple samples of the tested object, and experience some related states after experiencing this emotionalized 3D dynamic website. To achieve the purpose of emotion detection, we preset a number of detection concerns, including: the emotional state of the object,

TABLE 3 | Concerns and scores.

Number	Focus point	Score				
		1	2	3	4	5
1	Subject's emotional state					
2	Recognition of the system					
3	Experience of the system					
4	Communication of system emotions					
5	Infection of systemic emotions					

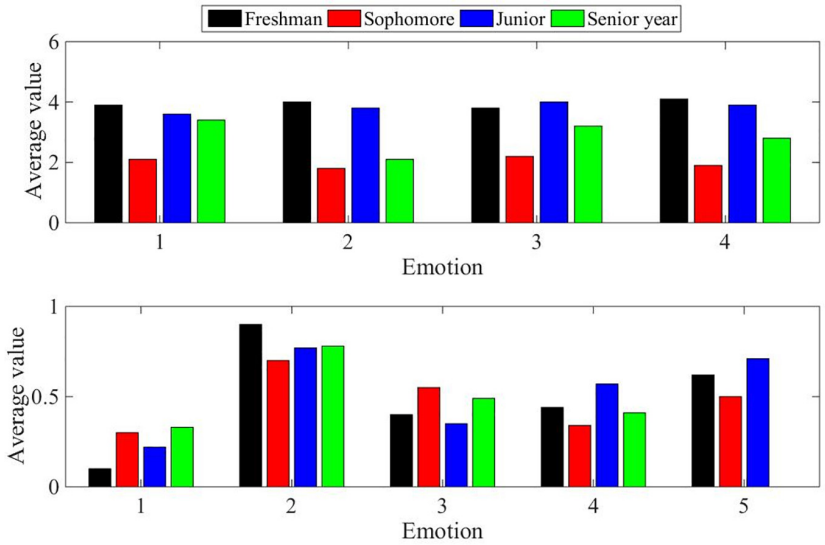


FIGURE 7 | Mean and variance of the overall statistics.

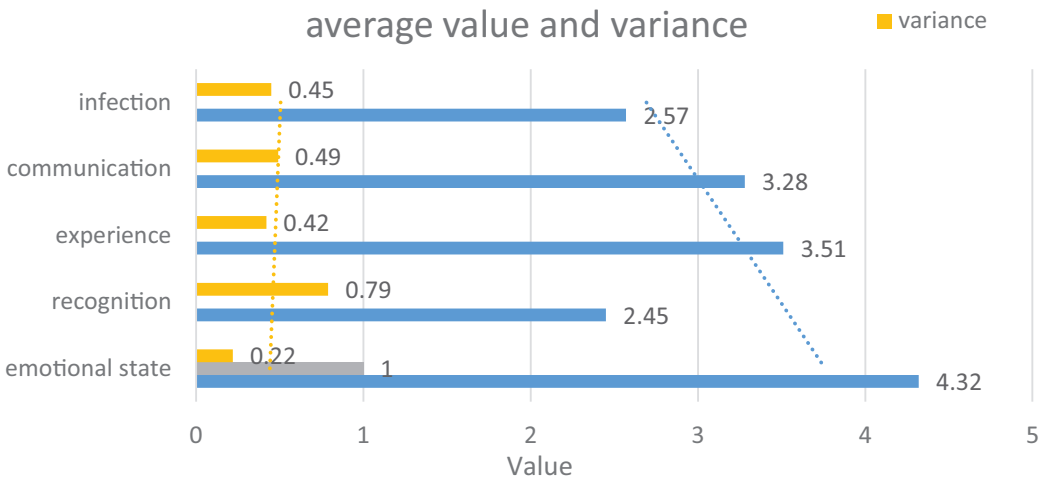


FIGURE 8 | Mean and variance of overall statistics.

the transmission of website emotions, the experience of the website, the contagiousness of website emotions, and the acceptance of three-dimensional websites. After the subject completes this emotional visualization experience, they will use questionnaire surveys to collect answers to related questions regarding these concerns. As shown in **Table 3**, for each point of interest, five values are used to scale the value of the corresponding point of interest. A smaller value means that the value of the corresponding point of interest is biased toward negative, and a larger value means that the value of the point of interest is biased toward positive. The detection process of a single sample is divided into 3 stages. The first stage introduces the purpose and meaning of this visual expression of emotion to the test subject; the second stage, the test subject answers the first question, that is, the current emotional state. Afterward, the test subjects experience the visual expression system for emotions; in the third stage, the test subjects give the other four points of attention corresponding to the scores based on the second stage experience.

EXPERIMENTAL RESULTS AND ANALYSIS

Experimental Results

After the completion of the test, a total of 100 valid questionnaires were collected. After the data was entered, the value of each focus point was counted, and the average value was calculated. At the same time, in order to verify the validity of the data, we

also calculated the variance of all the data, and plotted the final statistics as shown in **Figure 7**.

Experimental Analysis

First, the statistical test is performed on whether there is a statistical difference in the answer results of the grades to the questionnaire, and the result is $p > 0.05$, which shows that there is no statistical significance, so the subsequent analysis is no longer for each grade. The overall data result is shown in **Figure 8**:

From the data obtained from the detection, we can see that the overall emotional state of the tested object before the detection is good, and from the variance data, it can be seen that there is not much difference in the overall distribution. In such an emotional state, it can be said that the detection data from the subject has a certain degree of authenticity. After the testee is tested, first analyze the data in the average statistics chart. It can be seen that in general, the testee does not accept this form of 3D dynamic website. The cause of the problem may be due to 3D testing. Although dynamic websites have emotional mechanisms, they have not yet been applied. Therefore, subjects cannot intuitively feel this form of emotional visualization. We also found that the variance value of the recognition statistics is obviously too large, which proves that there is a gap in the understanding of this concern among different test subjects. I believe that as long as this method of visual expression of emotions can be applied to specific counterattacks, more users can understand. Third, we analyze data related to user experience. In terms of the average value, it is found that the detected object is also in a positive evaluation overall, but its value does not reach a very high level. The reason is that as we analyzed before, it may be because of what we present to the detected object. The website lacks a clear support environment, which leads to a not very high experience. At the same time, it can also be seen that the variance value of the experience statistical data is generally high, indicating that there is a certain gap in the user experience of each measured object. We then analyze the concerns of transitivity and infectiousness. It can be seen from the average statistical graph that the values of these two focus points are generally positive. In other words, this kind of emotional three-dimensional dynamic website space has a certain ability to convey emotions, and the measured objects generally accept this attitude. At the same time, we also found that the variance of these two concerns is also in a high state, which means that different users have different understandings on this. This is actually in line with the characteristics of human beings, because the differences in feelings and understanding of emotions will always be different for each person. Therefore, we believe that this method of visual expression of emotion has the ability to convey or suggest emotion. Finally, we can summarize this test. Generally speaking, the data we collected is effective, and the emotion visualization methods discussed in this article can indeed convey or imply certain emotions, but we also found that there is a phenomenon that this recognition is not high enough, and the reason is analyzed: lack A practical application environment makes this emotional visualization lack of practical significance. At the same time, it can also be seen that different tested subjects have different feelings and evaluations of this kind of emotion visualization. As long as this difference is within a reasonable range, this kind of emotion visualization also has practical value.

CONCLUSION

Human emotions have discrete and continuous characteristics. The construction of a comprehensive emotional synthesis model based on these two characteristics can describe human emotions more comprehensively. The qualitative and quantitative research methods of emotion in the comprehensive emotion synthesis model can model emotions more accurately to a certain extent. The visual expression based on the comprehensive emotional synthesis model makes the three-dimensional website present different forms, and the emotional interaction process between the website space and people is formed. Through the use of statistical analysis methods to detect the constructed emotional 3D dynamic website, the test results show that this emotion-based synthetic model can convey emotions to a certain extent, enhance the user experience of the dynamic 3D painting system, and enhance the superior experience of human-computer interaction.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

SC: editing data curation and writing–original draft preparation.

REFERENCES

Andrew, M., and Larceneux, F. (2019). The role of emotion in a housing purchase: an empirical analysis of the anatomy of satisfaction from off-plan apartment purchases in France. *Environ. Plan.* 51, 1370–1388. doi: 10.1177/0308518x18817539

Bringmann, L. F., Pe, M. L., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., et al. (2016). Assessing temporal emotion dynamics using networks. *Assessment* 23, 425–435. doi: 10.1177/1073191116645909

Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., Andre, E., Busso, C., et al. (2017). The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. *IEEE Trans. Affect. Comput.* 7, 190–202. doi: 10.1109/taffc.2015.2457417

Fortin, P. E., and Cooperstock, J. R. (2017). Laughter and tickles: toward novel approaches for emotion and behavior elicitation. *IEEE Trans. Affect. Comput.* 8, 508–521. doi: 10.1109/taffc.2017.2757491

Hu, X., Chen, J., Wang, F., and Zhang, D. (2019). Ten challenges for EEG-based affective computing. *Brain Sci. Adv.* 5, 1–20. doi: 10.1177/2096595819896200

Huang, A. (2018). A risk detection system of e-commerce: researches based on soft information extracted by affective computing web texts. *Electron. Commer. Res.* 18, 143–157. doi: 10.1007/s10660-017-9262-y

Huxtable-Thomas, L. A., Hannon, P. D., and Thomas, S. W. (2016). An investigation into the role of emotion in leadership development for

entrepreneurs: a four interface model. *Int. J. Entrep. Behav. Res.* 22, 510–530. doi: 10.1108/ijebr-12-2014-0227

Kim, D., Jeon, J., Cheong, E., Kim, D. J., Ryu, H., Seo, H., et al. (2016). Neuroanatomical Visualization of the Impaired Striatal Connectivity in Huntington's Disease Mouse Model. *Mol. Neurobiol.* 53, 2276–2286. doi: 10.1007/s12035-015-9214-2

Kostelnick, C. (2016). The re-emergence of emotional appeals in interactive data visualization. *Tech. Commun.* 63, 116–135.

Liu, H., Min, X., Wang, J., Rao, T., and Burnett, I. (2016). Improving Visual Saliency Computing With Emotion Intensity. *IEEE Trans. Neural Netw. Learn. Syst.* 27, 1201–1213. doi: 10.1109/tnnls.2016.2553579

Liu, Y., and Fu, G. (2021). Emotion recognition by deeply learned multi-channel textual and EEG features. *Future Gener. Comput. Syst.* 119, 1–6. doi: 10.1016/j.future.2021.01.010

Qiu, J. L., and Zhao, W. Y. (2018). Data Encoding Visualization Based Cognitive Emotion Recognition with AC-GAN Applied for Denoising// IEEE International Conference on Cognitive Informatics & Cognitive Computing. *IEEE Comput. Soc.* 7, 222–227.

Radojevic, N. (2016). Perspective, Cartography and Dynamic Notions: From the Plane Bozzetto to Solid Perspective in Two Examples in Tuscany. *Nexus Netw J.* 18, 669–695. doi: 10.1007/s00004-016-0300-1

Ren, F., and Matsumoto, K. (2017). Semi-automatic creation of youth slang corpus and its application to affective computing. *IEEE Trans. Affect. Comput.* 7, 176–189. doi: 10.1109/taffc.2015.2457915

Ridderinkhof, K. R. (2017). Emotion in Action: A Predictive Processing Perspective and Theoretical Synthesis. *Emot. Rev.* 9, 319–325. doi: 10.1177/1754073916661765

Tang, Y., and Elhoseny, M. (2019). Computer network security evaluation simulation model based on neural network. *J. Intell. Fuzzy Syst.* 37, 1–8. doi: 10.3233/JIFS-179121

Topal, K., and Ozsoyoglu, G. (2017). Emotional classification and visualization of movies based on their IMDb reviews. *Inf. Disc. Deliv.* 45, 149–158. doi: 10.1108/idd-05-2017-0045

Ueoka, R., and AlMutawa, A. (2018). "Emotion Hacking VR: Amplifying Scary VR Experience by Accelerating Actual Heart Rate," in *Human Interface and the Management of Information. Interaction, Visualization, and Analytics*, Vol. Vol. 10904, eds S. Yamamoto and H. Mori (Berlin: Springer Verlag), 436–445. doi: 10.1007/978-3-319-92043-6_37

Volante, M., Babu, S. V., Chaturvedi, H., Newsome, N., Ebrahimi, E., Roy, T., et al. (2016). Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. *IEEE Trans. Vis. Comput. Graph.* 22, 1326–1335. doi: 10.1109/tvcg.2016.2518158

Wang, Q., and Lu, P. (2019). Research on Application of Artificial Intelligence in Computer Network Technology. *Intern. J. Pattern Recognit. Artif. Intell.* 33:1959015. doi: 10.1142/s0218001419590158

Wang, Y., Segal, A., Klatzky, R., Keefe, D. F., Isenberg, P., Hurtienne, J., et al. (2019). An emotional response to the value of visualization. *IEEE Eng. Med. Biol. Mag.* 39, 8–17. doi: 10.1109/mcg.2019.2923483

Yu, J., Li, L., and Zou, J. (2017). Realistic emotion visualization by combining facial animation and hairstyle synthesis. *Multimed. Tools Appl.* 76, 14905–14919. doi: 10.1007/s11042-016-4239-8

Zhang, X., Liu, J., Shen, J., Li, K., Hou, B., Hu, J., et al. (2020). Emotion recognition from multimodal physiological signals using a regularized deep fusion of kernel machine. *IEEE Trans. Cybern.* 1–14. doi: 10.1109/tcyb.2020.2987575

Zhang, Y., Chen, J., Liu, B., Yang, Y., Li, H., Zheng, X., et al. (2020). COVID-19 public opinion and emotion monitoring system based on time series thermal new word mining. *Comput. Mater. Continua.* 64, 1415–1434. doi: 10.32604/cmc.2020.011316

Zhou, P., Li, K., Wei, W., Wang, Z., and Zhou, M. (2020). Fast generation method of 3D scene in Chinese landscape painting. *Multimed. Tools Appl.* 79, 16441–16457. doi: 10.1007/s11042-019-7476-9

Zhu, X., and Xue, L. (2020). Building a controllable expressive speech synthesis system with multiple emotion strengths. *Cogn. Syst. Res.* 59, 151–159. doi: 10.1016/j.cogsys.2019.09.009

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Cheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 19 August 2021
doi: 10.3389/fpsyg.2021.710122

New Technologies' Commercialization: The Roles of the Leader's Emotion and Incubation Support

Biaoan Shan[1], *Yi Pu*[1], *Biao Chen*[2]* *and Shan Lu*[1]

[1] School of Management, Jilin University, Changchun, China, [2] School of Business, Zhengzhou University, Zhengzhou, China

New technologies, such as brain-computer interfaces technology, advanced artificial intelligence, cloud computing, and virtual reality technology, have a strong influence on our daily activities. The application and commercialization of these technologies are prevailing globally, such as distance education, health monitoring, smart home devices, and robots. However, we still know little about the roles of individual emotion and the external environment on the commercialization of these new technologies. Therefore, we focus on the emotional factor of the leader, which is their passion for work, and discuss its effect on technology commercialization. We also analyzed the moderating role of incubation support in the relationship between the leader's emotion and technology commercialization. The results contribute to the application of emotion in improving the commercialization of new technologies.

Keywords: the leader emotion, passion for work, incubation support, technology commercialization, new venture

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Peng Huatao,
Wuhan University of Technology, China
Xin Wang,
Southwestern University of Finance and Economics, China

***Correspondence:**
Biao Chen
cb1512453@163.com

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *15 May 2021*
Accepted: *21 June 2021*
Published: *19 August 2021*

Citation:
Shan B, Pu Y, Chen B and Lu S (2021) New Technologies' Commercialization: The Roles of the Leader's Emotion and Incubation Support. Front. Psychol. 12:710122. doi: 10.3389/fpsyg.2021.710122

INTRODUCTION

In the digital economy era, the application and commercialization of new emerging technologies, such as brain-computer interfaces technology, advanced artificial intelligence, Cloud Computing, and Virtual Reality technology, are strongly influencing our daily life while creating significant amounts of innovation and entrepreneurship opportunities all over the world. Driven by the fast growth of these new technologies, the industries in distance education, health monitoring, smart home, and robots have grown quickly. Many new ventures in these industries have been created. Obviously, with the development of the digital economy and digital technologies, how to commercialize these new technologies has become an important research topic.

Prior studies have focused on various influencing factors of technology commercialization (TC), including external environmental factors (Li et al., 2008; Chen et al., 2011) and organizational factors, such as firm resources, strategic orientation, inter-departmental cooperation, and technological innovation capabilities (Zahra and Nielsen, 2002; Sohn and Moon, 2003; Park and Ryu, 2015; Jo and Park, 2017). Some scholars explored the moderating factors which may strengthen or weaken the efficiency of organizational elements, including formal and informal integration mechanisms, organizational culture, and environmental dynamic (Zahra and Nielsen, 2002; Park and Ryu, 2015). These studies have made important contributions to the understanding of TC. However, existing research ignores the role of leaders in TC, especially when the effect of the leader's emotion on TC is still unknown.

The organizational structure of new ventures is incomplete and informal (Fultz and Hmieleski, 2021). In the process of commercializing technologies, the leaders of new ventures are the core decision makers. However, they always face many challenges, such as a lack of knowledge and experiences, limited resources, and a rapidly changing environment (Chen, 2009; Sullivan and Marvel, 2011). In this situation, leaders need positive emotion to persist with their goals and conquer these challenges in TC (Baum and Locke, 2004; Connelly and Ruark, 2010). In this study, we focus on a salient emotion, a leader's passion for work, which influences leader's behaviors, including communication, learning, and decision making (Baum and Locke, 2004). It is a critical psychological characteristic of leaders (Huy and Zott, 2019; De Cock et al., 2020) and plays an important role in the application of new technologies. Leaders who are passionate about their work would take an active role and systematically participate in the knowledge management process associated with their work (De Clercq et al., 2013), which, in turn, would have a positive effect on the whole company.

In order to address this gap, this study considers the role of a leader's passion for work in commercializing new technologies in new ventures. Klaukien et al. (2013) argue that the role of emotional factors, especially a leader's passion for work in the innovation decision-making activities, needs to be further explored. To reply to this importance, we explore the role of the leaders, as the boundary spanner connecting customers/markets to technologies that significantly promote organizational technology commercialization, on the TC from an individual emotional characteristic perspective (Markman et al., 2008).

In addition, we suggest that the effect of passion for work on the TC is contingent on incubation support. In other words, we consider incubation support as an important moderating variable. TC involves a set of complex behaviors that result from a combination of organizational elements and support environments (Chen, 2009). However, only a few empirical studies explore the interactive role of a leader's emotional characteristics and support environment in TC. Recently, different types of business incubators have been created. These incubators are diverse in function. Incubators provide policy support, funding, necessary entrepreneurial counseling services, and technical support to new ventures (Bøllingtoft, 2012; Bruneel et al., 2012). Hence, incubators can build a support environment to carry out entrepreneurial activities and affect the TC of new ventures (Davidsson et al., 2006). In this study, we consider incubation support as an external support environment factor. We analyze the effect of the interaction between the leader's emotion and incubation support on technology commercialization.

This study mainly tries to answer the question "how do leaders' passion for work influence TC in new ventures?" and explores the moderating role of incubation support. By doing so, this study mainly makes two contributions to existing literature. First, this study recognizes an emotional characteristic of a leader (passion for work) and its impact on TC. The findings further deepen the research of technology innovation and commercialization and expand the research potential of emotion theories in the business field. Second, we consider incubation support as an external environment and examine its moderating impact on the relationship between a leader's passion for work and TC. Finally, researchers have tried to theorize the role of an incubator in the creation and growth of new ventures (Scillitoe and Chakrabarti, 2010; Wu et al., 2020a). In this study, we consider incubation support as an external environment factor and find that context differences in the strength of incubation support could strengthen the effect of the leader's emotion on TC. By doing so, we extend the theoretical research on the interaction between psychological characteristics and external environmental factors on TC.

We organize the remainder of this study as follows: Section Theoretical Background and Hypotheses proposes our hypotheses based on the literature review on the leader's emotion and TC. Section Methodology presents our research method, including the process of data collection, sample characteristics, measurement of variables, and sample validity. Section Results provides regression analyses of the hypotheses and reports the results. Section Discussion discusses the theoretical and managerial implications, the limitations, and the conclusions of this study.

THEORETICAL BACKGROUND AND HYPOTHESES

Leader's Positive Emotion: The Passion for Work

From a psychological perspective, an individual's behaviors and activities in their everyday life are accompanied by emotional experience (Russell, 2003). Passion, like sadness, happiness, or joy, is an important emotional experience (Cardon et al., 2005; Prayag et al., 2020). Identity centrality and positive feelings are the two basic elements of passion (Cardon et al., 2009). Studies show that individuals can determine their preferences by repeatedly participating in various activities and continuing to engage in related activities (Vallerand, 2008). In this process, an individual's preference gradually resonates with their self-identity and generates positive emotion (Murnieks et al., 2019; Wu et al., 2020b). Therefore, the leader's positive emotion is stable and sustainable which is different from sadness, happiness, and other emotions (Cardon et al., 2009).

A leader's passion for work is an important emotion that affects the organization (De Clercq et al., 2013; Ma et al., 2017). Passion for work represents an individual's feelings about work and reflects the extent to which an individual enjoys working and gains happiness from investing or participating in work-related activities (De Clercq et al., 2013). Passion for work comes from the internalization of an individual's self-identity, and it can be either active or passive (Cardon et al., 2009; Slemp et al., 2021). Active internalization creates a harmonious passion, while passive internalization creates an obsessive passion (Vallerand, 2008).

Regarding the theoretical research on the passion for work, many studies explore the passion for work of employees, such as the effect of passion for work on employee job satisfaction (Houlfort et al., 2013), employee's perception of work (Lavigne

et al., 2014), and employee's performance (Burke et al., 2015). A small number of scholars have focused on the passion of entrepreneurs and managers. Baum and Locke (2004) regard passion as a typical individual characteristic and discuss the passion of entrepreneurs in new ventures. With the promotion of passion, entrepreneurs will positively build their resource skills, shape their entrepreneurial cognition, and improve themselves to achieve growth goals. De Clercq et al. (2013) focused on the relationship between potential entrepreneur passion for work and entrepreneurial intention.

Most scholars believe that this type of positive emotion significantly influences individual cognition and behavioral tendencies. However, only a few studies have discussed the impact of passion for work on innovation activities in new ventures. Klaukien et al. (2013) confirmed that a manager's passion for work has a positive effect on their decision to use new product development opportunities. Strese et al. (2018) studied the passion for inventing in CEOs and argued that passion is beneficial for small and medium enterprises to carry out exploratory innovation. As an important psychological characteristic variable, we still know little about how leaders' passion for work impacts innovation in new ventures.

Technology Commercialization (TC)

Early studies viewed TC as a process to transform new technologies into products or services that meet market needs. The process includes designing, altering, producing, and marketing (Nevens, 1990; Mitchell and Singh, 1996). Originally the vision of implementing ideas, TC refers to the process of turning technology into profit or value (Siegel et al., 1995; Li et al., 2008). In the digital economy era, improving TC is a key means for new ventures to gain a sustained competitive advantage since new technologies, such as brain-computer interfaces technology, advanced artificial Intelligence, cloud computing, and virtual reality technology, constantly emerge and shape lots of opportunities for creating new ventures.

The capability perspective suggests that TC is an essential organizational level capability that enables a company to apply new technologies to develop products or services, increasing the company's revenues, profits, and competitiveness (Kim et al., 2011). Scholars use frequency, speed, novelty, and patents to measure TC (Zahra and Nielsen, 2002). Chen (2009) divides TC into three dimensions: TC speed, market scope, and technology breadth. TC speed refers to the degree of launching new products or services faster than competitors to obtain high profits. Market scope reflects the extent to which products or services are brought to different markets. The broader the scope, the more a firm can amortize the cost of R&D and marketing. Technology breadth refers to how many related technologies are used in a new product. The more technologies, the more likely a product is to satisfy diverse customer demands.

After exploring the concept and the process of TC (Nevens, 1990), more scholars focus on gaining a competitive advantage by promoting TC capability. Scholars in this stream revealed the roles of various factors in improving TC, such as the external environment (Sohn and Moon, 2003), the characteristics of technologies (Chen et al., 2011), the innovative capability and technological capability (Jiang et al., 2017; Jo and Park, 2017), the internal and external resources of the organization (Zahra and Nielsen, 2002), cooperation and knowledge creation (Lin et al., 2015), and the absorptive capacity (Jo and Park, 2017). They mainly focus on the roles of organizational factors (such as resources and capabilities) within a company and the contingent role of environmental factors. But the literature lacks concern about new ventures' commercialization and the effects of leaders' psychological characteristics in this process.

The Leader Passion for Work and TC in New Ventures

New ventures engage in opportunity development activities under a high resource-constrained environment (Senyard et al., 2014; Cai et al., 2017; Shan and Lu, 2020), where new ventures must cope with high uncertainty when exploring opportunities (Lu et al., 2021a,b). In this context, positive emotion, for example, passion for work, stimulates a leader's desire to succeed. Thus, they will be more enthusiastic and more actively respond to obstacles that may occur in the process of technology commercialization (Baum and Locke, 2004). Perrewé et al. (2014) found that passion for work reflects an ongoing emotion of an individual's desire to work. Therefore, passion for work would promote a leader to complete the work tasks and proactively seek solutions to overcome obstacles in their daily work. High passion for work drives a leader to cope with resource constraints and reduce uncertainty in technology development. This then improves the technology commercialization capability in new ventures.

Passion for work also triggers learning activities and promotes the acquisition of new knowledge (De Clercq and Pereira, 2020; Gong et al., 2020). For example, De Clercq et al. (2013) found that passion for work relates to an individual's cognitive learning process, making a leader actively participate in high-intensity and systematic knowledge processes. It is crucial for the new ventures that lack prior experiences (Shan and Lu, 2020). Driven by high passion for work, a new venture's leader would continuously update their knowledge systems. For example, the leaders would use creative bricolage to learn how to acquire resources to achieve the mission in the high-resources-limited context (Zott and Huy, 2007; Lurtz and Kreutzer, 2014; Murray et al., 2020).

Passion also strengthens a leader's perception of the ability to complete an important task (Cardon et al., 2009), thus improving a leader's motivation in their daily work (Donahue et al., 2012; Feng and Chen, 2020). Passion for work also increases a leader's self-efficacy, which is essential for the leader since high self-efficacy promotes investment in innovation activities and improves technology commercialization. Therefore, passion triggers a leader to choose appropriate strategies and an organizational structure that match innovative activities, promotes the application of innovative technologies, and strengthens the technological breadth of new products (Strese et al., 2018).

Based on the above arguments, we propose the following hypothesis:

H1. A leader's passion for work positively influences the capability of technology commercialization in the new venture.

The Moderating Role of Incubation Support

As an important form of the organization supporting entrepreneurial activities and promoting innovation, incubators provide shared office spaces and facilities to the leaders of the new ventures (Hackett and Dilts, 2004). They create a highly supportive environment for the new ventures (Bergek and Norrman, 2008; Luke et al., 2019), which will positively enhance the probability of new ventures' success (Chen, 2009; Scillitoe and Chakrabarti, 2010; Haugh, 2020). Incubators have various forms. There are incubators established by large companies (Evald and Bager, 2008), science and technology parks based on universities or research institutions (McAdam and McAdam, 2008), and business incubators established by private companies (entrepreneurs) or government agencies (Tello et al., 2012). Incubators can also be divided into diversified incubators and specialized incubators (Schwartz and Hornych, 2010).

In general, the incubation supports new ventures in three ways: providing basic services such as shared spaces and facilities (Grimaldi and Grandi, 2005); providing commercial services such as financial support, management consulting, and business contacts (Scillitoe and Chakrabarti, 2010); and providing technical services such as technical consulting, technology transfer, and intellectual property protection (Bergek and Norrman, 2008; Scillitoe and Chakrabarti, 2010). All the support can greatly reduce the transaction costs of commercializing new technologies. This then increases the survival rate and the success of a technology or a product innovation (Tello et al., 2012).

Incubators can be seen as an innovative system to provide a supportive environment to facilitate the acquisition of information and resources for new ventures (Schwartz and Hornych, 2010). The formal and informal social network inside and outside the incubators offers potential opportunities for new ventures (Nijssen and Michel, 2017). Incubation support becomes an important driving force for the creation and growth of new ventures and actively influences them to conduct innovative activities and exploit business ideas (Davidsson et al., 2006; Schwartz and Hornych, 2010). In this context, new ventures are more likely to succeed in technology commercialization with incubation support.

As mentioned above, a leader's passion for work plays an important role in the process of TC for new ventures. Incubation support provides a more friendly external environment for passionate leaders of new ventures to cope with their challenges (Bergek and Norrman, 2008). Environmental support has significant implications for leaders because passion for work is easily influenced by the external environment (Zigarmi et al., 2009). If one new venture is in an incubator, the leader has more channels to acquire resources, as well as basic services, and other commercial and technical support (Bøllingtoft and Ulhøi, 2005).

The key to technology commercialization is how to match technology and market requirement. A leader of a new venture always lacks market experience (Li and Zhang, 2007). Incubators guide these leaders of the new ventures to develop products or services to meet the needs of the customer. Many specialized incubators have a deep understanding of a specific industry with experience in the technical sectors, which can provide more specialized services in technology development and market operating (Schwartz and Hornych, 2010). These incubators can also provide the necessary technical advisory services (McAdam and McAdam, 2008) that allow the new ventures to meet the requirement of the market (Chen, 2009). Such a support environment greatly reduces the time and cost of commercialization. Therefore, incubation support promotes passionate leaders to focus more on the market-oriented development of technology and improve efficiency.

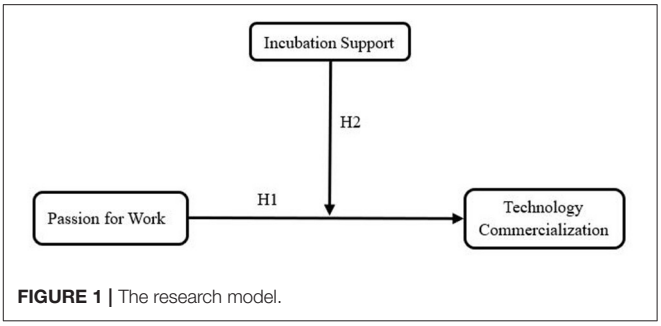


FIGURE 1 | The research model.

New technologies' commercialization is a complex process (Chen et al., 2011). The complex and innovative network system generated by an incubator promotes the flow of information and knowledge, which triggers the learning activities at different levels, including interorganizations learning, organizational learning, and employee learning. As Tötterman and Sten (2005) revealed, the incubators positively support new ventures acquiring basic knowledge in terms of technology commercialization processes. The leaders with a higher passion for work may more actively participate in relevant learning activities driven by the technology commercialization with the incubation support.

Therefore, we propose that:

H2. Incubation support strengthens the positive relationship between a leader's passion for work and the new venture's technology commercialization capability.

Thus, we propose our research model in **Figure 1**.

METHODOLOGY

Data Collection and Samples

In order to verify the theoretical hypotheses, we conducted a questionnaire survey in February 2018 in Beijing, Tianjin, Hangzhou, Shanghai, and other cities in the Yangtze River Delta regions, where entrepreneurial activities are relatively active in China. Also, the application of new emerging digital technologies is prevailing there. Driven by the new technologies, many new ventures in sectors such as smart healthcare, smart home devices, and online education are created. Therefore, we considered the new ventures in these regions of China as the sample firms.

Before the formal survey, the authors conducted a pilot test in several companies. According to the feedback of these interviews, we revised and improved the language and items of the core variables to avoid ambiguity in the survey. Subsequently, we randomly distributed questionnaires to companies in these areas mentioned above. We used different ways to collect data. Most samples were collected by email, while the other paper questionnaires were collected during the interview of research team members.

We sent about 500 questionnaires and received 243 questionnaires back. Our research focused on new ventures companies with a registration time within 10 years (Milanov and Fernhaber, 2009). Thus, we deleted the questionnaires from companies who had been registered for more than 10 years. At the same time, we also eliminated the questionnaires with a total missing value more than 25%. In the end, 140 valid samples remained and the characteristics are shown below 0.83 samples (59.28%) are from the Yangtze River Delta region, 57 samples come from other regions such as Beijing and Tianjin which account for 40.72%, 92 samples (65.71%) are from technological sectors such as computer science, new energy, new materials, and software industry, while 48 samples are from non-technological sectors which account for 34.29%. In terms of the number of employees, 84 samples have at least 50 employees and 56 samples (40%) have <50 employees.

Measures

The measures of the variables in this paper mainly draws on prior mature scales. Likert 7-point was used in the survey (1 is very low or extremely disagree, 7 is very high or extremely agree). In order to ensure the reliability of measurements, we mainly use the scales that are tested by other prior researchers. We translated all the items into Chinese and then translated them back to English independently by different professionals. This may ensure the accuracy of measures.

The Capability of Technology Commercialization (TC)

According to Chen (2009), we consider three dimensions: the speed of technology commercialization, market scope, and technology width. We use eight items to measure TC. These include, the capability to quickly develop products, the capability to quickly introduce products/services to the markets, and the capability to improve existing products/services to adapt to different markets.

A Leader's Passion for Work (PW)

This variable measurement was developed based on mature studies. We drew on the measurement of Baum and Locke (2004) and De Clercq et al. (2013). Four items were used to reflect passion for work. These included "I like to work hard" or "I am looking forward to returning to work when I am not working." The scale was used and well tested by prior studies.

Incubation Support (IS)

Following the study of Chen (2009), we set a virtual variable to measure incubation support. We asked the leader whether their firms are supported by an incubator, where 0 represents the company has not received any support from an incubator, while 1 means that the company has received support from an incubator.

TABLE 1 | The results of confirmatory factor analysis and Cronbach's alpha coefficients.

Variables	Items	Factor loading	Cronbach's alpha
TC	TC1	0.801	0.925
	TC2	0.895	
	TC3	0.841	
	TC4	0.869	
	TC5	0.826	
	TC6	0.815	
	TC7	0.857	
	TC8	0.859	
PW	PW1	0.918	0.870
	PW2	0.914	
	PW3	0.707	
	PW4	0.869	

Control Variables

We used five control variables in this study: the number of employees (1–20, 21–50, 51–200, and more than 200), the age of ventures, work experience of a leader, educational background of a leader (high school or below, junior college, bachelor's degree, and master or Ph.D. degree), and the industry. Among these, we set work experience variable based on the working years of a leader in the sector and we set a dummy variable for "industry" based on whether it belongs to the technology industry (0 means non-technology-based industry and 1 means technology-based industry).

We tested the validity of the sample using a confirmatory factor analysis. The results show that the value of factor loading for the variables is more than 0.7. We also calculated the Cronbach's alpha coefficient which exceeds 0.8. The results in **Table 1** show good reliability and validity. The Harman's one-factor test was conducted. The result showed that the largest factor only explained 33.76% of the entire variance. According to the view of Podsakoff and Organ (1986), the problem of common method bias is not significant.

RESULTS

The results of the descriptive statistics and the correlations are shown in **Table 2**. The coefficients of Pearson correlation show a significant correlation between passion for work and technology commercialization. We use multiple linear regression analysis to test the hypotheses. We use the VIFs (variance inflation factors) to test multicollinearity, and the results showed no obvious multicollinearity problem.

We built three regression models (see **Table 3**). Model 1 tests the effects of control variables on TC while model 2 examines the effect of a leaders' passion for work on TC. The results show that the coefficient of passion for work is 0.234 ($p < 0.05$, model 2). Therefore, Hypothesis 1 is supported. To test the moderating

TABLE 2 | Descriptive statistics and correlation analysis.

Variable	1	2	3	4	5	6
Number of employees	1					
Age	0.254**	1				
Work experience	0.144	0.183*	1			
Education background	0.115	−0.070	−0.210*	1		
PW	−0.086	−0.157	−0.203*	−0.098	1	
TC	0.117	0.046	0.217*	−0.030	0.230**	1
Mean	2.93	5.06	7.66	2.93	5.18	4.58
S. D.	0.871	2.861	6.258	0.892	1.071	1.238

*$p < 0.05$, **$p < 0.01$.
Dummy variables including industry and incubation support are not shown in the table.

TABLE 3 | The results of linear regression analysis.

	Dependent variable: TC		
	Model 1	Model 2	Model 3
Number of employees	0.055	0.061	0.048
Age	−0.103	−0.071	−0.016
Work experience	0.200	0.243*	0.164
Education background	0.080	0.114	0.063
Industry	−0.070	−0.080	−0.017
Independent variable			
PW		0.234*	0.205*
Interaction			
IS			−0.071
PW × IS			0.553***
R^2	0.056	0.107	0.397
Adj-R^2	0.001	0.043	0.339
F Value	1.013	1.678*	6.757***

*$p < 0.05$, ***$p < 0.001$.

effects of incubation support, this study built model 3. The regression analysis shows that the coefficient of passion for work is 0.205 ($p < 0.05$, model 3), and the coefficient of the interaction between passion for work and incubation support is 0.553 ($p < 0.001$, model 3). These results support Hypothesis 2.

DISCUSSION

TC requires a series of complex processes such as intuitive imagination, technology cultivation, design, and product testing (Lin et al., 2015). The findings show that a leader's passion for work is essential in improving TC, which is consistent with the study by Klaukien et al. (2013). To increase the success rate of TC, many scholars paid attention to exploring the effect factors of TC (Zahra and Nielsen, 2002; Sohn and Moon, 2003; Li et al., 2008; Chen et al., 2011; Park and Ryu, 2015; Jo and Park, 2017). However, they ignored the psychological characteristics of founders and the support environment of incubators. TC is a complex process, and many challenges exist in the process. The leader's passion for work is essential in improving TC.

Meanwhile, TC needs different types of resources. For new ventures, they generally lack these resources. Incubators provide a support environment for founders to cope with the challenge, which improves the TC. Hence, this research explored the effects of internal factor (the passion for work) and external factor (incubation support) on TC and found several valuable findings.

A leader with high passion for work promotes knowledge and skills enabling them to overcome potential difficulties (Baum and Locke, 2004; Perrewé et al., 2014). The flexible organizational structure of a new venture means that the leader is always close to other members. A leader's positive emotions can transmit to others through social comparison and emotional imitation (Breugst et al., 2012). A leader who is full of passion for work may encourage other members to work hard and enhance job satisfaction.

We also find that incubators provide a support environment for new ventures, which is a benefit for leaders to promote the capability of TC. Incubator, as a special form of organization, creates a support environment for the TC of new ventures by directly or indirectly providing various services or resources including basic services, commercial support, and technical support (Bøllingtoft and Ulhøi, 2005; Adlesic and Slavec, 2012). This put leaders in a friendly work environment, which stimulates their enthusiasm for participation in technological development and application.

Hence, leaders should flexibly use such incubation supports. For example, they can frequently contact the managers of incubators and utilize shared information and resources. Incubators exist as intermediaries connecting new ventures and resource owners. The leaders also have more opportunities to learn how to meet wider market demands based on the new technologies. It is obvious that the capability to commercialize technology will be strengthened by the support from incubators.

IMPLICATIONS

The findings of this study have several managerial implications and theoretical implications.

Theoretical Implications

First, existing research has ignored the role of leader's emotional characteristic in TC (Klaukien et al., 2013). We identify a leader's passion for work as a crucial internal emotional characteristic and examine its effect on TC of new venture, while previous studies have mainly focused on the effects of external

environmental factors and organizational factors on technology commercialization (e.g., Sohn and Moon, 2003; Lin et al., 2015; Kirchberger and Pohl, 2016). Based on the characteristics of a new venture, we identify a leader's passion for work as an important impactor and link it to TC. In doing so, we demonstrated that a leader's emotional characteristic plays an important role in the process of innovation, which has important implications for future research on individual emotional factors and innovation of new ventures.

Second, this study casts light on the contextual factor when explaining why some new ventures have more difficulty achieving achieving success than others in technology commercialization. In this regard, previous researchers remarked that the formal/informal integration of resources (Zahra and Nielsen, 2002), and technological turbulence (Li et al., 2008) create the differences. We identify incubation support as a new contextual factor and test its moderating effect on the relationship between a leader's passion for work and TC. Although the scholars have recognized the roles of incubators in the creation and growth of new ventures (e.g., Markman et al., 2008; Schwartz and Hornych, 2010), only a few scholars have used empirical research to analyze the role of interaction of emotional variables and incubation support in TC of new ventures. Our findings suggest that incubators not only influence entrepreneurial activities, but also new ventures' innovation and TC.

Managerial Implications

First, this research investigates a leader's positive emotion and its effect on TC of a new venture. The finding encourages a new ventures' managers or founders to foster and utilize their passion for work, especially for technology-based new ventures. It may improve the probability of success when the top managers or founders have high passion for work since high uncertainty exists in the process of TC. Therefore, whether an individual has passion for the work could be an evaluation standard to hire a new CEO or manager.

Second, the leaders of new ventures should seek the support of incubators. Different types of incubators, including state-owned incubators/private incubators or diversified incubators/specialized incubators (Schwartz and Hornych, 2010; Tello et al., 2012), provide various resources, knowledge, and services for new organizations located in them. It is an important decision for new ventures to start their business in incubator. In order to improve the speed and success rate of TC, leaders (managers or founders) need to frequently communicate with the managers of incubators to get support from them. For the new ventures that are not located in incubators, their leaders should create relations with some incubators near them. The business network of incubators is also beneficial for leaders of those organizations to facilitate technology commercialization.

Finally, more and more digital technologies are emerging and triggering new opportunities in various industries. Hence, we suggest the governments or their agencies encourage the creation of incubators by providing policy and financial support. Because of the important role of incubators in the application and commercialization of new technologies, more polices should be introduced and more financial funds should be created to develop incubators. Furthermore, government officials should connect with the managers of incubators and the leaders of new ventures directly to understand what is most needed. By doing so, the overall failure rate of TC may be reduced significantly.

LIMITATIONS AND FUTURE RESEARCH

This study also has limitations. We only explore the direct impact of a leader's passion for work on technology commercialization of the new venture. As a psychological factor, passion for work may affect individual cognition and thus have an indirect impact on organizational innovation. Therefore, future studies may focus on revealing these paths and mechanisms of how a leader's passion for work influences technology commercialization from a cognitive perspective. In addition, other emotional characteristics of leaders, such as happiness or sadness, may also have effects on technology commercialization of a new venture. Therefore, future research should explore the role of these emotional characteristics. Furthermore, considering the more detailed and professional support provided by different types of incubators, such as state-owned/private, specialized/diversified, and university science parks, future researchers can analyze whether different types of incubators could support TC in different ways.

CONCLUSION

This study explores the relationship between a leader's positive emotion and technology commercialization of new ventures. Based on a sample of 140 new ventures from China, we find that as a typical emotional characteristic, a leader's passion for work is an important driving force for TC of new ventures. The positive emotion promotes the successful rate of commercializing the emerging technologies such as brain-computer interfaces technology, advanced artificial intelligence, Cloud Computing, and Virtual Reality technology. We also find incubation support provides a friendly external environment for new ventures, and significantly strengthens the role of a leader's passion for work in TC.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Department of Technology Economy and Management, Jilin University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This research was supported by Construction Project of Excellent Scientific Research Team in Humanities and Social Sciences of Zhengzhou University (Grant No. 2020-QNTD-10).

REFERENCES

Adlesic, R. V., and Slavec, A. (2012). Social capital and business incubators performance: testing the structural model. *Econ. Bus. Rev. Central and South—Eastern Euro*. 14, 201–222.

Baum, J. R., and Locke, E. A. (2004). The relationship of entrepreneurial traits, skill, and motivation to subsequent venture growth. *J. Appl. Psychol.* 89, 587–598. doi: 10.1037/0021-9010.89.4.587

Bergek, A., and Norrman, C. (2008). Incubator best practice: a framework. *Technovation* 28, 20–28. doi: 10.1016/j.technovation.2007.07.008

Bøllingtoft, A. (2012). The bottom-up business incubator: leverage to networking and cooperation practices in a self-generated, entrepreneurial-enabled environment. *Technovation* 32, 304–315. doi: 10.1016/j.technovation.2011.11.005

Bøllingtoft, A., and Ulhøi, J. P. (2005). The networked business incubator—leveraging entrepreneurial agency? *Soc. Sci. Electron. Publish.* 20, 265–290. doi: 10.1016/j.jbusvent.2003.12.005

Breugst, N., Domurath, A., Patzelt, H., and Klaukien, A. (2012). Perceptions of entrepreneurial passion and employees' commitment to entrepreneurial ventures. *Entrep. Theory Pract.* 36, 171–192. doi: 10.1111/j.1540-6520.2011.00491.x

Bruneel, J., Ratinho, T., Clarysse, B., and Groen, A. (2012). The evolution of business incubators: comparing demand and supply of business incubation services across different incubator generations. *Technovation* 32, 110–121. doi: 10.1016/j.technovation.2011.11.003

Burke, R. J., Astakhova, M. N., and Hang, H. (2015). Work passion through the lens of culture: harmonious work passion, obsessive work passion, and work outcomes in Russia and China. *J. Bus. Psychol.* 30, 457–471. doi: 10.1007/s10869-014-9375-4

Cai, L., Chen, B., Chen, J., and Bruton, G. D. (2017). Dysfunctional competition and innovation strategy of new ventures as they mature. *J. Bus. Res.* 78, 111–118. doi: 10.1016/j.jbusres.2017.05.008

Cardon, M. S., Wincent, J., Singh, J., and Drnovesk, M. (2009). The nature and experience of entrepreneurial passion. *Acad. Ranage. Rev.* 34, 511–532. doi: 10.5465/amr.2009.40633190

Cardon, M. S., Zietsma, C., Saparito, P., Matherne, B. P., and Davis, C. (2005). A tale of passion: new insights into entrepreneurship from a parenthood metaphor. *J. Bus. Ventur.* 20, 23–45. doi: 10.1016/j.jbusvent.2004.01.002

Chen, C. J. (2009). Technology commercialization, incubator and venture capital, and new venture performance. *J. Bus. Res.* 62, 93–103. doi: 10.1016/j.jbusres.2008.01.003

Chen, C. J., Chang, C. C., and Hung, S. W. (2011). Influences of technological attributes and environmental factors on technology commercialization. *J. Bus. Ethics*. 104, 525–535. doi: 10.1007/s10551-011-0926-6

Connelly, S., and Ruark, G. (2010). Leadership style and activating potential moderators of the relationships among leader emotional displays and outcomes. *Leadersh. Q.* 21, 745–764. doi: 10.1016/j.leaqua.2010.07.005

Davidsson, P., Hunter, E., and Klofsten, M. (2006). Institutional forces: The invisible hand that shapes venture ideas?. *Int. Small Bus. J.* 24, 115–131. doi: 10.1177/0266242606061834

De Clercq, D., Honig, B., and Martin, B. (2013). The roles of learning orientation and passion for work in the formation of entrepreneurial intention. *Int. Small Bus. J.* 31, 652–676. doi: 10.1177/0266242611432360

De Clercq, D., and Pereira, R. (2020). Knowledge-sharing efforts and employee creative behavior: the invigorating roles of passion for work, time sufficiency and procedural justice. *J. Knowl. Manage.* 24, 1131–1155. doi: 10.1108/JKM-06-2019-0274

De Cock, R., Denoo, L., and Clarysse, B. (2020). Surviving the emotional rollercoaster called entrepreneurship: the role of emotion regulation. *J. Bus. Ventur.* 35:4. doi: 10.1016/j.jbusvent.2019.04.004

Donahue, E. G., Forest, J., Vallerand, R. J., Lemyre, P., Crevier-Braud, L., and Éliane Bergeron. (2012). Passion for work and emotional exhaustion: the mediating role of rumination and recovery. *Appl. Psychol. Health Well-Being.* 4, 341–368. doi: 10.1111/j.1758-0854.2012.01078.x

Evald, M. R., and Bager, T. (2008). Managing venture team relationships in corporate incubators: a case study of network dynamics and political rivalry in a high-tech incubator. *Int. Entrep. Manage. J.* 4, 349–364. doi: 10.1007/s11365-006-0024-x

Feng, B., and Chen, M. (2020). The Impact of entrepreneurial passion on psychology and behavior of entrepreneurs. *Front. Psychol.* 11:1733. doi: 10.3389/fpsyg.2020.01733

Fultz, A., and Hmieleski, K. M. (2021). The art of discovering and exploiting unexpected opportunities: the roles of organizational improvisation and serendipity in new venture performance. *J. Bus. Ventur.* 36, 106–121. doi: 10.1016/j.jbusvent.2021.106121

Gong, Z., Zhang, Y., Ma, J., Liu, Y., and Zhao, Y. (2020). Effects of work passion on turnover intention for Chinese government employees: the dualistic model of passion perspective. *J. Manage. Organiz.* 26, 502–518. doi: 10.1017/jmo.2017.71

Grimaldi, R., and Grandi, A. (2005). Business incubators and new venture creation: an assessment of incubating models. *Technovation* 25, 111–121. doi: 10.1016/S0166-4972(03)00076-2

Hackett, S. M., and Dilts, D. M. (2004). A systematic review of business incubation research. *J. Technol. Transf.* 29, 55–82. doi: 10.1023/B:JOTT.0000011181.11952.0f

Haugh, H. (2020). Call the midwife! Business incubators as entrepreneurial enablers in developing economies. *Entrepr. Region. Develop.* 32, 156–175. doi: 10.1080/08985626.2019.1640480

Houlfort, N., Philippe, F. L., Vallerand, R. J., and Ménard, J. (2013). On passion and heavy work investment: personal and organizational outcomes. *J. Manag. Psychol.* 29, 25–45. doi: 10.1108/JMP-06-2013-0155

Huy, Q., and Zott, C. (2019). Exploring the affective underpinnings of dynamic managerial capabilities: how managers' emotion regulation behaviors mobilize resources for their firms. *Strateg. Manage. J.* 40, 28–54. doi: 10.1002/smj.2971

Jiang, X., Yi, X., and Jiang, F. (2017). An ambidexterity view toward technology commercialization. *Acad. Manage. Ann. Meet. Proceed.* 2016:13340. doi: 10.5465/ambpp.2016.13340abstract

Jo, D. H., and Park, J. W. (2017). The determinants of technology commercialization performance of technology-based SMEs. *KSII Trans. Int. Inform. Syst.* 11, 4146–4161. doi: 10.3837/tiis.2017.08.023

Kim, S. K., Lee, B. G., Park, B. S., and Oh, K. S. (2011). The effect of RandD, technology commercialization capabilities and innovation performance. *Technol. Econ. Develop. Econ.* 17, 563–578. doi: 10.3846/20294913.2011.603481

Kirchberger, M. A., and Pohl, L. (2016). Technology commercialization: a literature review of success factors and antecedents across different contexts. *J. Technol. Transf.* 41, 1077–1112. doi: 10.1007/s10961-016-9486-3

Klaukien, A., Shepherd, D. A., and Patzelt, H. (2013). Passion for work, nonwork-related excitement, and innovation managers' decision to exploit new product opportunities. *J. Product Innov. Manage.* 30, 574–588. doi: 10.1111/jpim.12010

Lavigne, G. L., Forest, J., Fernet, C., and Crevier-Braud, L. (2014). Passion at work and workers' evaluations of job demands and resources: a longitudinal study. *J. Appl. Soc. Psychol.* 44, 255–265. doi: 10.1111/jasp.12209

Li, H., and Zhang, Y. (2007). The role of managers' political networking and functional experience in new venture performance: evidence from China's transition economy. *Strateg. Manage. J.* 28, 791–804. doi: 10.1002/smj.605

Li, Y., Guo, H., Liu, Y., and Li, M. (2008). Incentive mechanisms, entrepreneurial orientation, and technology commercialization: evidence from China's transitional economy. *J. Produc. Innov. Manage.* 25, 63–78. doi: 10.1111/j.1540-5885.2007.00283.x

Lin, Y., Wang, Y., and Kung, L. A. (2015). Influences of cross-functional collaboration and knowledge creation on technology commercialization: evidence from high-tech industries. *Industr. Market. Manage.* 49, 128–138. doi: 10.1016/j.indmarman.2015.04.002

Lu, S., Shan, B., and Fei, Y. (2021a). Exploring an inverted U-shaped relationship between entrepreneurial experience and Chinese new venture performance: the moderating role of environmental uncertainty. *Asia Pacific Bus. Rev*. 1-18. doi: 10.1080/13602381.2021.1899464

Lu, X., Shan, B., and Pu, Y. (2021b). Sustainability in the entrepreneurial ecosystem: operating mechanisms and enterprise growth. *Asia Pacific Bus. Rev*. 8941. doi: 10.1080/13602381.2021.1928941

Lukeš, M., Longo, M. C., and Zouhar, J. (2019). Do business incubators really enhance entrepreneurial growth? Evidence from a large sample of innovative Italian start-ups. *Technovation* 82, 25–34. doi: 10.1016/j.technovation.2018.07.008

Lurtz, K., and Kreutzer, K. (2014). What does your audience expect from you? How entrepreneurs acquire resources through storytelling. *Acad. Manage. Ann. Meet. Proceed*. 16008–16008. doi: 10.5465/ambpp.2014.85

Ma, C., Gu, J., and Liu, H. (2017). Entrepreneurs' passion and new venture performance in China. *Int. Entrepr. Manage. J.* 13, 1043–1068. doi: 10.1007/s11365-017-0435-x

Markman, G. D., Siegel, D. S., and Wright, M. (2008). Research and technology commercialization. *J. Manage. Stud.* 45, 1401–1423. doi: 10.1111/j.1467-6486.2008.00803.x

McAdam, M., and McAdam, R. (2008). High tech start-ups in University Science Park incubators: the relationship between the start-up's lifecycle progression and use of the incubator's resources. *Technovation* 28, 277–290. doi: 10.1016/j.technovation.2007.07.012

Milanov, H., and Fernhaber, S. A. (2009). The impact of early imprinting on the evolution of new venture networks. *J. Bus. Ventur.* 24, 46–61. doi: 10.1016/j.jbusvent.2007.11.001

Mitchell, W., and Singh, K. (1996). Survival of businesses using collaborative relationships to commercialize complex goods. *Strateg. Manag. J.* 17, 169–195. doi: 10.1016/S0737-6782(96)89828-6

Murnieks, C. Y., Mcmullen, J. S., and Cardon, M. S. (2019). Does congruence with an entrepreneur social identity encourage positive emotion under environmental dynamism? *J. Small Bus. Manage.* 57, 872–890. doi: 10.1111/jsbm.12335

Murray, A., Kotha, S., and And Fisher, G. (2020). Community-based resource mobilization: how entrepreneurs acquire resources from distributed non-professionals via crowdfunding. *Organiz. Sci.* 31, 1–30. doi: 10.1287/orsc.2019.1339

Nevens, T. M. (1990). Commercializing technology: what the best companies do. *Plann. Rev.* 18, 20–24. doi: 10.1108/eb054310

Nijssen, E. J., and Michel, V. (2017). Beyond the water cooler: using socialization to understand use and impact of networking services on collaboration in a business incubator. *RandD Manage.* 47, 443–457. doi: 10.1111/radm.12261

Park, T., and Ryu, D. (2015). Drivers of technology commercialization and performance in SMEs: the moderating effect of environmental dynamism. *Manage. Decisi.* 53, 338–353. doi: 10.1108/MD-03-2014-0143

Perrewé, P. L., Hochwarter, W. A., Ferris, G. R., McAllister, C. P., and Harris, J. N. (2014). Developing a passion for work passion: future directions on an emerging construct. *J. Organ. Behav.* 35, 145–150. doi: 10.1002/job.1902

Podsakoff, P. M., and Organ, D. W. (1986). Self-reports in organizational research: problems and prospects. *J. Manage.* 12, 531–544. doi: 10.1177/014920638601200408

Prayag, G., Mills, H., Lee, C., and Soscia, I. (2020). Team identification, discrete emotions, satisfaction, and event attachment: a social identity perspective. *J. Bus. Res.* 112, 373–384. doi: 10.1016/j.jbusres.2019.11.062

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychol. Rev.* 110, 145–172. doi: 10.1037/0033-295X.110.1.145

Schwartz, M., and Hornych, C. (2010). Cooperation patterns of incubator firms and the impact of incubator specialization: empirical evidence from Germany. *Technovation* 30, 485–495. doi: 10.1016/j.technovation.2010.05.001

Scillitoe, J. L., and Chakrabarti, A. K. (2010). The role of incubator interactions in assisting new ventures. *Technovation* 30, 155–167. doi: 10.1016/j.technovation.2009.12.002

Senyard, J., Baker, T., Steffens, P., and Davidsson, P. (2014). Bricolage as a path to innovativeness for resource-constrained new firms. *J. Produc. Innov. Manage.* 31, 211–230. doi: 10.1111/jpim.12091

Shan, B., and Lu, X. (2020). Founder's social ties, learning and entrepreneurial knowledge acquisition in China. *Asia Pacific Bus. Rev.* 26, 209–229. doi: 10.1080/13602381.2020.1718318

Siegel, R. A., Hansén, S. O., and Pellas, L. H. (1995). Accelerating the commercialization of technology: Commercialization through co-operation. *Industr. Manage. Data Syst.* 95, 18–26. doi: 10.1108/02635579510079425

Slemp, G. R., Zhao, Y., Hou, H., and Vallerand, R. J. (2021). Job crafting, leader autonomy support, and passion for work: testing a model in Australia and China. *Motiv. Emot.* 45, 60–74. doi: 10.1007/s11031-020-09850-6

Sohn, S. Y., and Moon, T. H. (2003). Structural equation model for predicting technology commercialization success index (TCSI). *Technol. Forecast. Soc. Change.* 70, 885–899. doi: 10.1016/S0040-1625(03)00004-0

Strese, S., Keller, M., Flatten, T. C., and Brettel, M. (2018). CEOs' passion for inventing and radical innovations in SMEs: the moderating effect of shared vision. *J. Small Bus. Manage.* 56, 435–452. doi: 10.1111/jsbm.12264

Sullivan, D. M., and Marvel, M. R. (2011). Knowledge acquisition, network reliance, and early-stage technology venture outcomes. *J. Manage. Stud.* 48, 1169–1193. doi: 10.1111/j.1467-6486.2010.00998.x

Tello, S., Yang, Y., and Latham, S. (2012). Nascent entrepreneurs access and use of network resources in a technology incubator. *J. Small Bus. Entrepr.* 25, 375–397. doi: 10.1080/08276331.2012.10593579

Tötterman, H., and Sten, J. (2005). Start-ups: Business incubation and social capital. *Int. Small Bus. J.* 23, 487–511. doi: 10.1177/0266242605055909

Vallerand, R. J. (2008). On the psychology of passion: In search of what makes people's lives most worth living. *Can. Psychol.* 49, 1–13. doi: 10.1037/0708-5591.49.1.1

Wu, C., Li, S., Cheng, F., Zhuang, L., Du, Y., He, S., et al. (2020a). Self-identity and career success of nurses in infectious disease department: The chain-mediating effects of cognitive emotion regulation and social support. *Front. Psychol.* 11:563558. doi: 10.3389/fpsyg.2020.563558

Wu, W., Wang, H., and Tsai, F. S. (2020b). Incubator networks and new venture performance: the roles of entrepreneurial orientation and environmental dynamism. *J. Small Bus. Enterpr. Develop.* 27, 727–747. doi: 10.1108/JSBED-10-2019-0325

Zahra, S. A., and Nielsen, A. P. (2002). Sources of capabilities, integration and technology commercialization. *Strat. Manage. J.* 23, 377–398. doi: 10.1002/smj.229

Zigarmi, D., Nimon, K., Houson, D., Witt, D., and Diehl, J. (2009). Beyond engagement: toward a framework and operational definition for employee work passion. *Hum. Resour. Develop. Rev.* 8, 300–326. doi: 10.1177/1534484309338171

Zott, C., and Huy, Q. (2007). How entrepreneurs use symbolic management to acquire resources. *Adm. Sci. Q.* 52, 70–105. doi: 10.2189/asqu.52.1.70

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Shan, Pu, Chen and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 31 August 2021
doi: 10.3389/fpsyg.2021.741928

Research on Migrant Works' Concern Recognition and Emotion Analysis Based on Web Text Data

Zhijie Dou[1], *Zixuan Cheng*[2] *and Dongmei Huang*[3]*

[1] Shool of Management, Changchun University, Changchun, China, [2] School of Electronic and Information Engineering, Changchun University, Changchun, China, [3] Department of Psychology, School of Philosophy and Sociology, Jilin University, Changchun, China

Based on the characteristics of convenience, autonomy, and equality, online self-media has become an important way for contemporary migrant workers to observe the world, understand society, examine themselves and express their demands. On the basis of the analysis of the domestic migrant works' concerns and their emotion analysis, we crawl data on Weibo about migrant works' topics as the basic corpus of migrant works' concerns, and then uses a combination of TF-IDF and Word2Vec methods to construct a recognition model of migrant workers' concerns. We found that wages, children's education, medical care and returning home are the main concerns of migrant workers. Meanwhile, further emotion analysis of the migrant works' concerns of using a deep learning model fused with Bi-LSTM and CNN was conducted. The results show that the proportion of negative emotion such as worries, complaints and impetuosity was significantly higher than that of other positive and neutral emotion like encourage and comfort. And the time when the negative emotion are concentrated is significantly related to the social events that occur in the corresponding time period. On the one hand, it shows that the concerns and emotion of migrant workers can be effectively observed and predicted through web text data. On the other hand, it also shows that the core well-being issues of migrant workers in the process of urban integration have not been effectively solved, and the government and relevant departments need to take targeted measures and give priority attention.

Keywords: migrant workers, concerns, emotion analysis, Word2Vec, Bi-LSTM, CNN

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Yuexin Li,
Hubei University, China
Xiaoqing Gu,
Changzhou University, China

****Correspondence:***
Dongmei Huang
zhaotaiyang20@jlu.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *15 July 2021*
Accepted: *03 August 2021*
Published: *31 August 2021*

Citation:
Dou Z, Cheng Z and Huang D (2021) Research on Migrant Works' Concern Recognition and Emotion Analysis Based on Web Text Data. Front. Psychol. 12:741928. doi: 10.3389/fpsyg.2021.741928

INTRODUCTION

Two billion farmers stand at the entrance of industrial civilization. This is the main problem posed to the social sciences in the second half of the 20th century. Since the mid to late 1980s, driven by marketization, industrialization and urbanization, a large number of rural labor forces in our country have transferred to cities, and a new social class called the migrant worker class has gradually formed. As the product of the era of the dual oppositional structure of urban and rural household registration, it has attracted widespread attention from the government, media, academia and even the society. It has become one of the important topics in the localization of social sciences and has achieved fruitful research results. However, most of the existing research focus on the four aspects of group characteristics summary (Looyestyn et al., 2018), urban integration

exploration (Mikulincer et al., 2005; Chou, 2009), mental health evaluation (Leavey et al., 2007; Li et al., 2007) and social identity (Zerubavel, 2000; Sb et al., 2020). The problems of migrant works' emotion have not yet obtained enough attention. As the key link between micro social reality and macro social reality, almost any aspect of human cognition, behavior and social organization is driven by emotion. Migrant workers are the emotional subjects with flesh and blood, joy and sorrow, and they also have spiritual demands and emotional needs. Therefore, to truly understand the living conditions and behavioral logic of migrant workers, especially migrant workers in the new era, emotion are just an indispensable link. In the study of migrant workers, if scholars ignore or simply pay attention to migrant works' emotion instead of digging into them, they cannot fully explain why migrant workers make some objectively apparently irrational decision-making actions, such as publicly beating leaders. At the same time, they cannot understand why migrant workers form certain mentalities and concepts, such as thinking that college students of the same age are naive. As a result, it is not conducive to grasping the interactive relationship between the micro-fortunes of migrant workers and the macro-structure of society. In addition, the process of migrant workers' integration into the city is the process by which farmers move from a rural society that characterizes traditional civilization to an urban society that characterizes modern civilization. It is also a resocialization mechanism. Existing studies have shown that social integration can be seen from two dimensions, one is behavior fusion and the other is emotion fusion (Kisar Koramaz, 2014). The close relationship among emotion, social structure and cultural environment makes the research of migrant workers' emotion complementary to urban integration and social identity. Therefore, the emotion adaptation efforts of migrant workers in the process of integrating into the city deserve an academic response. Especially the new generation of migrant workers, as a product of the internal structure iteration of the rural migrant labor group for more than 30 years, their emotion experience is more worthy of our attention. Compared to their parents who are based on survival rationality during their working careers and have limited levels and scope of emotional generation.

With the popularization and development of information technology, social media platforms such as Weibo, Moments, Zhihu, and Post Bar have become new ways for migrant workers to obtain information, express their opinions, express their emotion, and clarify their standpoints. According to public statistics, as of December 2020, the number of Chinese Internet users has reached 989 million. In the professional structure of netizens, laborers in agriculture, forestry, animal husbandry, fishery and rural migrant workers accounted for 15.3 and 11.4%, respectively, ranking the third and fourth. Agricultural laborers and migrant workers are important constituents of Chinese netizens. They will publish text comments with personal emotion on related events or topics, and these subjective texts often contain great value. By mining and analyzing the web text data, it is possible to identify the migrant works' concerns and emotion. This will help government departments to further understand issues of common concern for migrant worker and their mental status. In this way, targeted policies can be designed to better meet their demands. In view of this, the paper combines TF-IDF and Word2Vec to construct a recognition model of migrant workers' concerns, and mines the main concerns of migrant workers contained in the web text data. Then we use a deep learning model that combines bidirectional long and short-term memory network (Bi-LSTM) and convolutional neural network (CNN) to perform emotion analysis on the concerns of migrant workers.

Finally, The structure of the paper is as follows:

Part 1: The first part introduces the research background and value of the paper.

Part 2: Introduce the related research work of the thesis, and summarize the innovative points of the paper.

Part 3: Establish a model of migrant works' concerns recognition and emotion analysis based on web text data.

Part 4: Empirical results and analysis are conducted to prove the effectiveness of the model used.

BACKGROUNDS

Concerns Recognition

The migrant works' concerns can reflect the issues that migrant workers generally concern and their ideological status, which are an important basis for decision-making by the government and other relevant departments. In the past, most of the research on migrant workers was based on questionnaires, on-site interviews or discussions to collect data. It not only wasted time and energy, but the number of samples obtained was also very limited, which could not fully cover the concerns of migrant workers. According to Goffman's theater theory, our daily life and social activities can be compared to a dramatic performance. The performance area is divided into the front stage and the back stage. The front stage can be seen as a personal image based on the script, rather than a real personal image. In other words, in face-to-face communication, migrant workers may hide their true feelings, which makes it impossible for us to directly understand their concerns.

In recent years, with the help of new media represented by the mobile Internet, massive amounts of network data have shown an eruptive growth, and a large number of researches on the recognition of concerns on online social networks have emerged. Patil and Gupta (2016) and Chen et al. (2014) established a work process that combines qualitative analysis and large-scale data mining technology by comprehensively analyzing the literature on mining user social media data. Gou and Gaikwad (2016) dig out students' data from Twitter, and use qualitative analysis to find problems in students' daily learning life, such as heavy learning burden, lack of social skills, etc. In recent years, Patil and Kulkarni (2018) have further used Twitter API to obtain data on students' concerns, and classified the data into six major categories. And then the Memetic algorithm is used to classify the data. Compared with Bayes and ID3, the experimental results prove that the Memetic algorithm is more effective. At present, the topic recognition and tracking of online social network has become one of the research hotspots in the field of data mining, which has attracted the attention of more and more scholars, and has achieved fruitful research results (Vo and Ock, 2015; Bicalho

et al., 2017). This also provides a lot of reference for us to dig out the concerns of migrant workers.

Emotion Analysis

Text emotion analysis, also known as emotion analysis or opinion mining, is a research hotspot in the field of text analysis. It mainly uses the feature mining or learning of the contextual content information of the text to judge its emotional tendency, which is a text classification task that correlates emotion. The text to be analyzed can be content such as network comments, articles, and microblogs. In the current Internet era, news media platforms and various social media platforms generate a large number of user comments every day. Automatic emotion analysis of the data is helpful to understand the user's attitude toward specific events and topics, and to grasp public opinions so order to make corresponding decisions. Generally speaking, the methods of text emotion analysis can be divided into methods based on emotion dictionary, methods based on statistics, and methods based on deep learning.

The method based on the emotion dictionary needs to build an emotion dictionary composed of emotion words, and then use semantic rules to calculate the semantic similarity to judge the emotion tendency of the text. For example, Chesley et al. (2006) uses verbs and adjectives to construct emotion templates to obtain the emotion information of the text and achieve emotion classification. Joshi et al. (2011) uses the artificially annotated Twitter emotion information and traverses the emotion score in the text to judge the emotion polarity. Turney (2002) uses Pointwise Mutual Information (PMI) between words to classify text emotion. In summary, the methods based on emotion dictionary often relies heavily on the construction of emotion dictionaries and the design of rules, which has high labor costs and cannot identify text emotion information outside of the rules.

With the rapid development of machine learning, many scholars apply machine learning algorithms to text emotion analysis. Generally, emotion analysis methods based on machine learning mainly include support vector machines, decision trees, and naive Bayes. In emotion analysis tasks, the machine learning algorithms achieve a good classification effect by extracting information from the data set and constructing features. For example, in 2009, Boiy and Moens (2009) trained a variety of machine learning models and cascaded them into the final emotion classification model to solve the problem of multilingual emotion analysis. This method effectively compensates for the shortcomings of a single algorithm through the combination of different models. In 2013, Mikolov et al. (2013a) used a support vector machine (SVM) classification model that combines diversified features to complete the emotion classification task of Weibo text. The model expresses the characteristics of different types of words, which can make full use of the emotion information in the sentence and improve the effect of emotion classification. The above methods of text emotion analysis based on emotion dictionary and machine learning not only requires manual construction of high-quality emotion dictionary and feature engineering, but also consumes a lot of human resources, so it is difficult to be widely used in text emotion analysis tasks in open fields.

In recent years, due to the strong feature extraction capabilities of DNN, a variety of methods based on DNN have been widely used in emotion analysis and text classification. Unlike the traditional machine learning methods whose features are often sparse, deep neural network methods convert text into dense vectors and obtain high-level representations of text for classification. For example Vieira and Moura (2017) proposed a convolutional neural network based on sentence dimensions. This method sets different hyperparameters for different emotion types for training, so as to realize the task of text emotion analysis in sentence dimension. Since then, many learners have conducted more in-depth research on text emotion analysis methods. Yoon and Kim (2017) used the CNN-BiLSTM model of multi-channel word embedding for text emotion analysis. It can capture the high-level semantic relationships and long-term dependencies in the text, and has achieved good results on the emotion analysis data set provided by Twitter. Chen et al. (2018) proposed a Multi-Channel Information Crossing (MIX) model for text emotion analysis. MIX compares text fragments of different granularities to form a series of multi-channel similarity matrices, and cross-use with another set of carefully designed attention matrices to extract richer sentence features. The above text emotion analysis methods based on deep learning improve the accuracy of text emotion analysis by extracting semantic features of various sentence dimensions.

Based on the above research, we further study the concerns and emotion analysis of migrant workers based on the web text data. The related research work is as follows:

(1) First, construct a data set of the migrant works' concerns. At present, there is no public data set on the concerns of migrant workers or corresponding methods for mining the concerns of migrant workers in China. We use web crawler technology to capture the data related on Weibo, and then store these pre-processed data in SQLSever2008 after filtering out invalid comments such as advertisements and spam links. After that, technologies such as web page cleaning, word segmentation, and stop word removal are used to pre-process the acquired data, thereby saving a lot of calculation content and improving the accuracy of the topic.

(2) Second, establish a classification model of migrant workers' concerns. Through qualitative analysis of the pre-processed data, a multi-level classification framework is concluded. The first level includes three major classes. The second level is the refinement of each major class, with a total of nine sub-classes. And the third level is the emotion classification for each sub-class, namely positive, neutral and negative emotion. Then use TF-IDF and Word2Vec methods to construct a classification model of migrant workers' concerns, and identify the concerns of migrant workers.

(3) Finally, the emotion analysis is carried out on the concerns of migrant workers. Based on the classification of migrant works' concerns, the paper uses a deep learning model that combines Bi-LSTM and CNN to perform emotion analysis on migrant works' concerns. So as to understand the migrant workers' thoughts and expectations in time, which can help the relevant government departments to solve the problems in a more targeted manner.

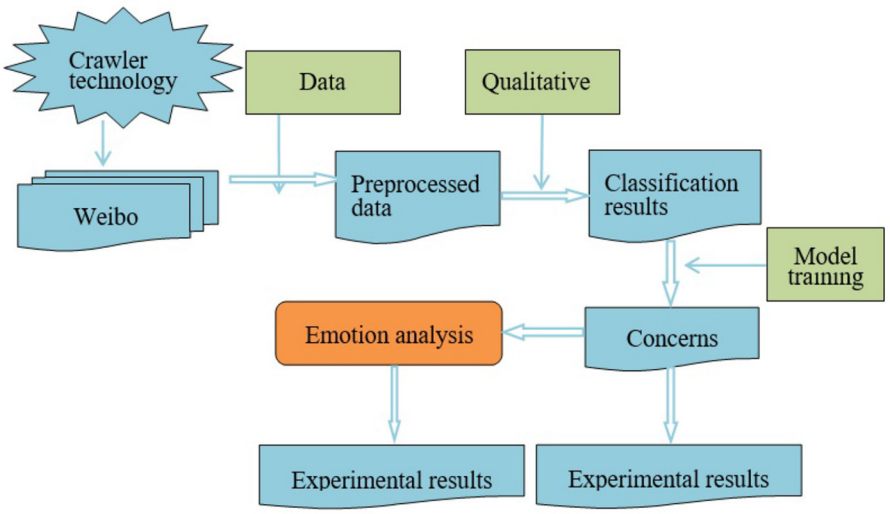


FIGURE 1 | The frame chart of migrant works' concerns recognition and emotion analysis.

THE MODEL FOR MIGRANT WORKS' CONCERNS RECOGNITION AND EMOTION ANALYSIS BASED ON WEB TEXT DATA

The research framework of the paper is shown in **Figure 1**, which is mainly divided into three parts. The first part is the collection and analysis of data sources. Firstly, we use crawler technology to obtain relevant data from social media and preprocess these data in turn to form a data source for migrant workers' concerns. The second part is to build a framework for migrant works' concerns. Based on the qualitative analysis of the acquired data, we found that migrant workers have different concerns at different stages, thus constructing a framework for migrant works' concerns. The third part is the construction of emotion analysis model. Different people will show different emotion tendencies toward the concerns. Based on the diversification of emotion, our paper establishes a method to analyze emotion tendencies. According to the results of emotion classification, we can clearly understand how migrant workers feel about different concerns.

The Model of Migrant Works' Concerns

In the paper, since the pre-processed data has no pre-defined classes, we conduct an inductive analysis of all the data. And then a three-layer concerns classification framework is established. The first layer is mainly divided into three classes: economic aspects, social aspects and cultural and educational aspects. Secondly, the three major classes are further subdivided into nine sub-classes. The economy class is divided into two sub-classes: wages and work. The social aspect is further divided into five sub-classes, namely, social security related, medical care, housing, household registration and homecoming. Finally, the cultural education category is divided into two sub-classes, namely identity and children's education. In summary, the concerns of migrant workers are divided into three major classes and nine sub-classes. And the third level is the emotion classification for each sub-class, namely positive, neutral and negative emotion. The three-layer concerns classification framework is shown in **Figure 2**. We categorize the data collected from Weibo into the classes listed in **Figure 2**.

Combined with the classification framework of the migrant works' concerns, we further use a method based on the combination of TF-IDF and Word2Vec to classify the data sources collected from Weibo, and identify the concerns of migrant workers. The method fully integrates the semantic information of the sentence with the word vector and similarity calculation, and effectively improves the accuracy of classification.

We use the TF-IDF method proposed by Salton in 1988 to calculate the vocabulary weights, and the calculation is shown in formula (1):

$$W_{ik} = \frac{tf(k,i) \times \log(N/n_k + 0.01)}{\sqrt{\sum_{k=1}^{t} (tf(k,i) \times \log(N/n_k + 0.01))^2}} \quad (1)$$

Where, W_{ik} represents the TF-IDF value of the kth word in the ith post, and $tf(k,i)$ is the frequency of the kth word in the ith post. N represents the total number of posts, n_{ik} is the number of posts containing the kth word, $\log(N/n_k + 0.01)$ is the IDF value of the kth word.

Then the Skip-gram model (Mikolov et al., 2013b) of the Word2Vec method is used to predict the context with a given input word, and its calculation formula is shown in equation (2).

$$P(w_o|w_i) = \frac{e^{U_o \cdot V_t}}{\sum_j e^{U_j \cdot V_t}} \quad (2)$$

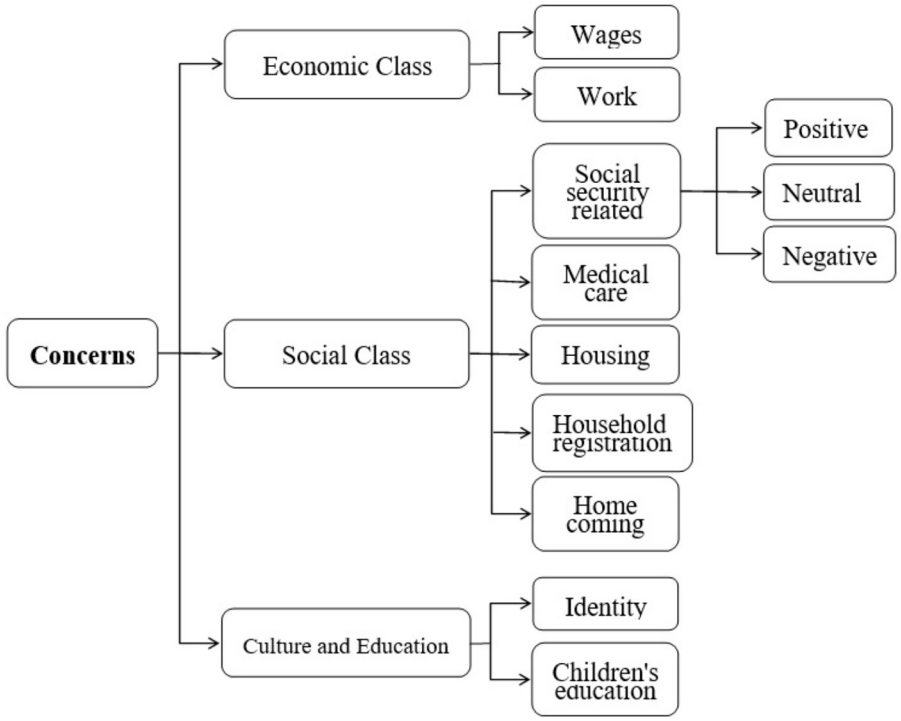


FIGURE 2 | Classification framework of migrant works' concerns.

Where V_t represents the input vector, and U_j represents the output vector.

Finally, suppose that after the Word2Vec, each post p_i can be represented by μ keywords, namely $p_i = \{wn_1, wn_2, \ldots, wn_\mu\} (1 \leq \mu \leq K \times M)$. Define the class set as $F = \{f_1, f_2, \ldots, f_9\}$, where $f_j (j = 1, 2, \ldots, 9)$ is the jth refined concern class. Define the posts of class f_j as $\left\{P_{f_j-1}, P_{f_j-2}, \ldots, P_{f_j-|f_j|}\right\}$, $|f_j|$ is the number of posts contained in class f_j. Then the formula for calculating the probability that the post p_i belongs to the class f_j is as follows:

$$P(p_i, f_j) = \frac{\sum_{i=1}^{\mu} \sum_{i=1}^{|f_j|} tf(wn_i, p_{f_j-i})}{\sum_{k=1}^{9} \sum_{i=1}^{\mu} \sum_{i=1}^{|f_k|} tf(wn_i, p_{f_k-i})} \quad (3)$$

The Framework of Migrant Works' Emotion Analysis

On the basis of recognizing the migrant works' concerns, we use a deep learning model combined with Bi-LSTM and CNN to analyze the emotion tendency of the migrant works' concerns. The network structure is shown in **Figure 3**.

As shown in the **Figure 3**, firstly, all posts of concerns will be represented by word vectors and input into the network after preprocessing. Secondly, the data will enter the convolutional layer and the memory layer, so as to make full use of CNN and Bi-LSTM, which is helpful for maintaining the text information and extracting features. Finally, the sofrmax classifier is used to obtain the classification results.

Word Vector Layer

First of all, the text of the input layer is segmented, and the word vector representation corresponding to each word in the text is obtained through Word2Vec (Lilleberg et al., 2015). Through the following formula, a word x_i can be converted into a word vector h_i:

$$h_i = E_w v^i \quad (4)$$

Where E_w represents the word vector matrix, and v^i indicates the size of the word vector. We use the method of randomly initializing the word vector, and it is constantly updated during the training process. Then the initial sentence will enter the convolutional layer in the form of word vector, namely $S_q = \{h_1, h_2, \cdots, h_n\}$.

Convolutional Layer

After the word vector layer, a sentence containing n words can be expressed as follows:

$$h_{0:n-1} = h_0 \oplus h_1 \oplus \cdots \oplus h_{n-1} \quad (5)$$

The symbol $\oplus$ is a connector between two adjacent words. Generally, the vector $h_{i:i+j}$ represents a series of word vectors

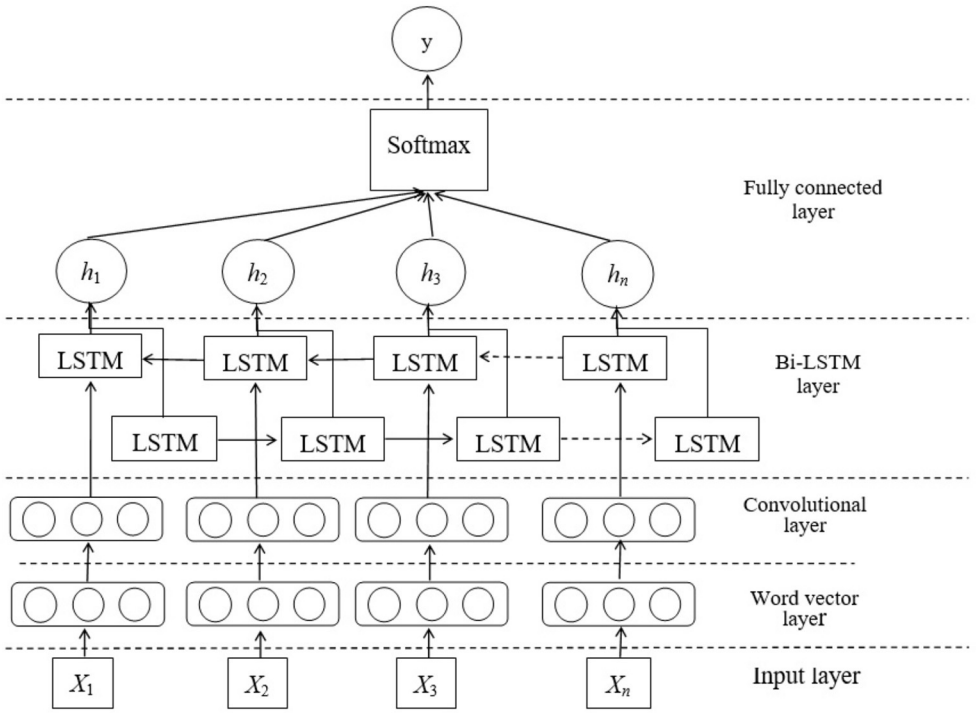


FIGURE 3 | The network model of emotion analysis based on Bi-LSTM and CNN.

$h_i, h_{i+1}, \cdots, h_{i+j}$. Each convolution operation contains a filter $w \in R^{md}$ (where d represents the dimension of the vector). It can generate a new feature through a window containing m words. For example, a feature c_i can be generated by window $h_{i:i+m-1}$:

$$c_i = f(w \cdot h_{i:i+m-1} + b) \tag{6}$$

Here $b \in R$ is a bias term, and f is a non-linear function similar to hyperbolic tangent. The input text can finally be expressed as:

$$c^* = [c_0, c_1, \cdots, c_{n-m}] \tag{7}$$

Bi-LSTM Layer

LSTM network was originally used to solve the problem of vanishing gradient, and then many variant network structures based on LSTM were proposed. In the paper, we adopt a variant structure proposed by Graves et al. (2009), which can increase the weight of the peephole connection on the same memory module.

In particular, a recurrent neural network based on long and short-term memory has 4 main components: An input gate i_t with the weight matrix of $W_{xi}, W_{hi}, W_{ci}, b_i$. A forgetting gate f_t with the weight matrix of $W_{xf}, W_{hf}, W_{cf}, b_f$. An output gate $o_t$with the weight matrix of $W_{xo}, W_{ho}, W_{co}, b_o$. All these dates will have a certain impact. Assuming that the current input x_i, h_{i-1} has been generated in the previous step. Then, the current state of unit c_{i-1} determines whether to use the input x_i, forgetting the previously stored memory. Finally output the generated state. These inferences can be proved by the following formulas:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \tag{8}$$
$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \tag{9}$$
$$g_t = \tanh(W_{xc}x_t + W_{hf}h_{t-1} + W_{cc}c_{t-1} + b_c) \tag{10}$$
$$c_t = i_t g_t + f_t c_{t-1} \tag{11}$$
$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o) \tag{12}$$
$$h_t = o_t \tanh(c_t) \tag{13}$$

Therefore, the generation of the current unit state c_t is determined by calculating the weight of the previous unit state and the current information generated by this unit. For many sentence-level processing tasks, it is necessary to consider contextual information. However, the standard LSTM network only considers the timing information when modeling the sentence and ignores the contextual information. The Bi-LSTM network expands the unidirectional LSTM network by introducing a second-layer network structure, and the hidden connections flow in the reverse time sequence. Therefore, Bi-LSTM can use the information of the context and ensure that the past and future information can be considered in the time series.

In the paper, we use the Bi-LSTM method to model the text sentences. As shown in **Figure 3**, the network contains two sub-networks to model text sentences before

and after. The output of the ith word is shown in the following formula:

$$h_i = \left[\overrightarrow{h_i} \oplus \overleftarrow{h_i}\right] \tag{14}$$

Classifier

In this layer of the network, we use a softmax classifier to predict the label y of the text sentence S from a set of discrete categories Y. The classifier uses the hidden state c^* as input:

$$\hat{p}(y|S) = soft\ \max(W^{(s)}c^* + b^{(s)}) \tag{15}$$

$$\hat{y} = \arg\max \hat{p}(y|S) \tag{16}$$

The loss function is as follows:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} t_i \log(y_i) + \lambda \|\theta\|_F^2 \tag{17}$$

Where $t \in R^m$ is the representation of one-hot, $y \in R^m$ represents the estimated probability of each class (m is the number of target classes), λ is the representation of L2 regularization parameter.

EXPERIMENTAL RESEARCH AND ANALYSIS

The empirical research of the paper mainly includes two parts: Firstly, we analyze the crawled data and construct a multi-level classification framework. Then we use TF-IDF and Word2Vec methods to identify the concerns of migrant workers. Secondly, we use the deep network structure method fused with Bi-LSTM and CNN to judge the emotion tendency of migrant workers' concerns.

Data Collection

As a social media and network platform for information sharing and exchange, Weibo is widely welcomed by migrant workers and has become the main platform for migrant workers in our country to share their experiences. They usually publish some articles or opinions in Weibo, which may be closely related to their work, life, family and social hot spots. Because of its social, freedom, anonymity and other characteristics, it is highly recognized by migrant workers, and it is also one of the important social platforms for migrant workers to show their emotion. In view of this, our paper chooses Weibo as the main source of data collection. Posts about migrant workers on Weibo cover a variety of topics, such as migrant workers' wages, recruitment, household registration and children's education. We use the Python technology to automatically retrieve the massive amount of scattered data on Weibo, and crawl the data as a corpus for concerns recognition and emotion analysis, with a total of about 20,000 lines. There are about 5,000 topic posts.

Data Pre-processing

Since the collected data contains the content that is not relevant to our research, such as advertising spam, etc. Meanwhile, a large amount of irrelevant information will affect the accuracy of the concerns recognition and easily cause the analysis results to deviate from the topic. Therefore, data pre-processing is necessary. We use web cleaning technology, word segmentation, stop word removal and other related technologies to pre-process the crawled data to obtain effective data related to the research. It will save a lot of computing space and improve the accuracy of concerns classification.

TABLE 1 | The classification results of migrant workers' concerns based on different methods (%).

Methods	Economic class	Social class	Cultural and educational class
LDA	0.55	0.53	0.51
Word2Vec	0.62	0.60	0.65
TF-IDF+Word2Vec	0.74	0.73	0.75

Experimental Results and Analysis

In the experiments, we use Precision which is commonly used in the text classification field, as the evaluation index. To show the effectiveness of the TF-IDF and Word2Vec methods used in the concerns classification, we also conducted comparative experiments with other classification methods. **Table 1** shows the classification results of migrant workers' concerns based on different methods.

Table 1 shows the classification results of migrant works' concerns in the three major classes of economy, society and culture. It can be seen from the results that compared with other methods, the classification accuracy of the TF-IDF+Word2Vec method is higher. It can also reflect the effectiveness of the method used in the paper.

In addition, we further visualized the data of migrant works' concerns. The histogram of the distribution of migrant works' concerns is shown in **Figure 4**. It should be noted that in the figure, the "other" class contains all posts that cannot be classified into any of the other nine classes. It can be seen from the figure that the migrant works' concerns mainly focus on wages, work, medical care, returning home during the Spring Festival and children's education, etc. It shows that the core well-being of migrant workers in the process of urban integration is still an important concern, and it is also a problem that relevant government departments and labor companies need to focus on and take effective measures to solve.

According to the results shown in **Figure 4**, the topic of migrant works' wages is one of their most concerned concerns. Therefore, taking the data of migrant works' wage as an example, we use the method fused Bi-LSTM and CNN to perform emotion analysis on them. The text whose judgment result is "-1" is negative emotion, the text whose judgment result is "1" is positive emotion, and the text whose judgment result is "0" is neutral emotion. Then we got 482 positive texts, 1,257 negative texts and 386 neutral texts. The results are visualized as shown in **Figure 5**.

In the **Figure 5**, the yellow part represents neutral emotion, the red part represents positive emotion, and the blue part

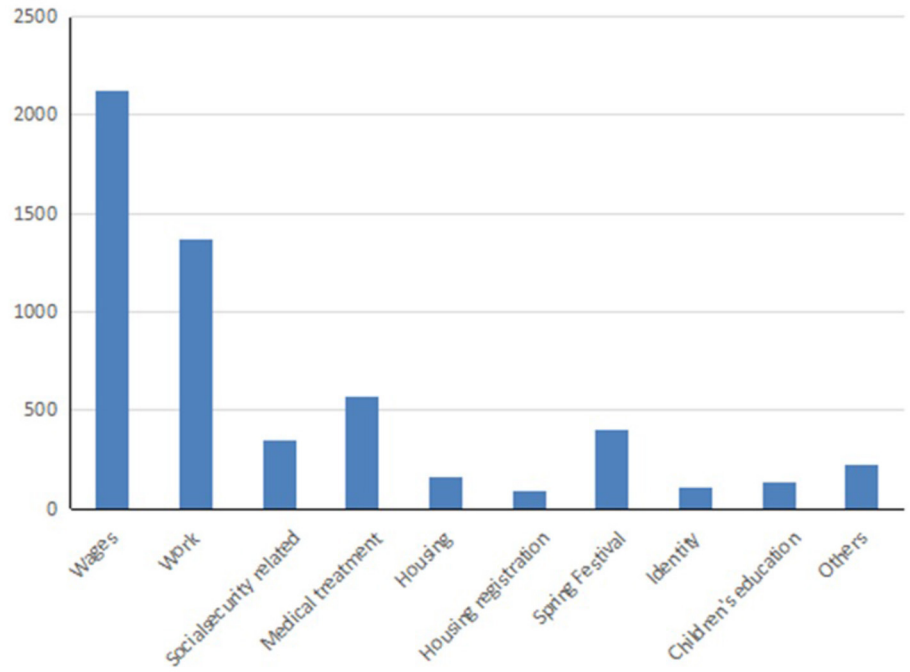


FIGURE 4 | The histogram of distribution of migrant works' concerns.

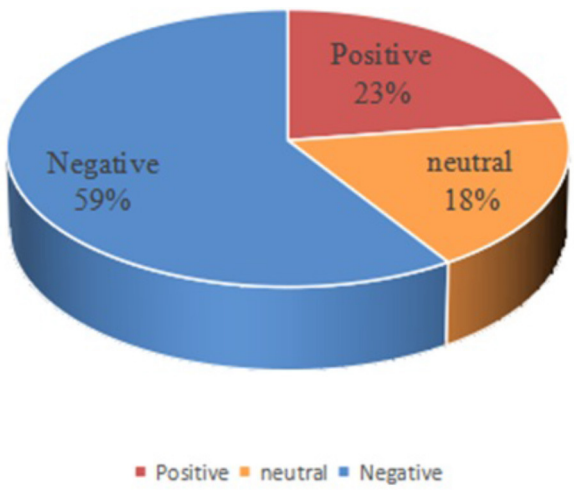


FIGURE 5 | Distribution of emotion analysis of migrant works' "wages".

represents negative emotion. It can be seen from the figure that the proportion of negative emotion is far greater than that of neutral and positive emotion. Combining the results of visualization and emotion analysis, we found that on the issue of "wages," migrant workers are generally concerned about topics such as "wage arrears, difficulty in asking for wages, and low wage income." They express their worries, complaints, impetuousness and other negative emotion on social platforms, expecting to receive the attention of society and related departments, and receive comfort from others. The second largest proportion is positive emotion, which includes encouragement and comfort among migrant workers.

Through the above analysis, we can draw the following conclusions: On the one hand, due to anonymity, social platforms (such as Weibo) are ideal spaces for people with the same experience to communicate, as well as places to seek help. Compared with positive emotion, migrant workers prefer to post negative emotion on social platforms. On the other hand, the government departments should pay attention to the data in time and provide necessary assistance to help the migrant workers adjust their psychological state.

In addition, as an example, we also analyzed the time series distribution of negative emotion texts of migrant workers from 2017 to 2018. Set the time as the horizontal axis and the number of negative emotion texts as the vertical axis. Then a time series distribution diagram of the negative emotion texts is shown in **Figure 6**. In the early 2017, December 2017 to January 2018, the number of negative emotion texts was relatively higher. Especially in December 2017, the number of negative emotion texts increased sharply. Therefore, we speculate that there is some correlation between the negative emotion of the research group and the social events that occurred in the corresponding time period.

As shown in the **Figure 6**, we analyze and discuss the negative text content according to the time period. In the beginning of 2017 and the end of 2017, the topics like "migrant workers ask for wages" and "wage arrears" appear more frequently in the negative emotion text. This shows that the research group pays

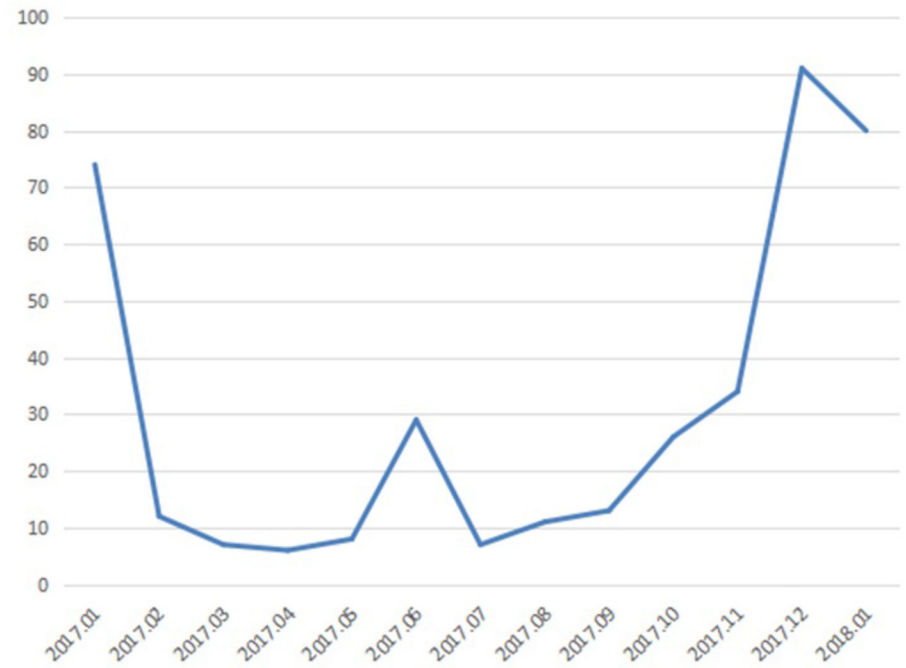


FIGURE 6 | The time series distribution diagram of the negative emotion texts.

more attention to "wages" in this time period. What's more, the three topics of "train ticket," "Spring Festival transport," and "return home" appeared frequently in the negative emotion texts. The results may predict that these three topics may be related to the negative emotion of the research group at this stage. In June 2017, the topics of "Hangzhou nanny arson case" appeared frequently. The topic basically coincides with the period of time when the negative emotion of migrant workers appear. Among them, "Hangzhou nanny arson case" is an emergency, and "wage arrears," "return home," and "Spring Festival travel" are topics of general concern for this group. This shows that in addition to daily life, this group also pay attention to social affairs. It also shows that the migrant works' emotion can be effectively intervened and guided.

CONCLUSION

Migrant workers are a unique social phenomenon in the process of industrialization and urbanization of our country. Their adaptation after entering the city is of great significance to the urbanization and modernization. However, as urban "migratory birds," the majority of migrant workers drift between urban and rural areas, encounter "marginalization" and rejection from all sides, they will inevitably face emotional dilemmas. In view of this, it is particularly important to understand the ideological trends of migrant workers in a timely manner, to grasp the concerns of migrant workers in an all-round way, and to effectively monitor and analyze their emotion. Only when the country, the government and the whole society timely understand the concerns of migrant workers, and detect signs and tendencies as early as possible, can they formulate relevant policies in a more targeted manner and resolve conflicts as soon as possible.

Based on the web text, the paper has conducted a comprehensive research on the concerns and emotion analysis of migrant workers and achieved good empirical results, but there are still the following aspects that need to be further improved. First of all, in terms of data collection, the paper only collected the data from Weibo. Although the data source after pre-processing is highly credible, the data range is relatively small. The current crawling methods in this paper are only for designated websites, and cannot crawl all the websites of the whole network. In other words, web page analysis is not intelligent enough and comprehensive. Secondly, to further improve the recognition rate of positive and negative emotion, the method of emotion analysis for migrant workers needs to be improved. However, this article only analyzes sentiment from the concerns of migrant workers and wages, which is not enough. Due to the limited information collected by the data set, it is impossible to analyze the emotions of migrant workers more comprehensively. In the future, we intend to collect and use more comprehensive and targeted data sets for emotion analysis.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

ZD conceived and designed the framework of the manuscript. From the determination of the research problem, the selection and implementation of the solution are all done independently by ZD. During the implementation of the plan, both ZD and ZC participated in model training, experimental data evaluation and analysis, and manuscript writing. DH was mainly responsible for controlling the entire process from conception to completion of the manuscript and making suggestions. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the Jilin Province Social Science Foundation Project under Grant 2020B094.

REFERENCES

Bicalho, P. V., Pitam, P. G., Lacerda, A., and Pappa, G. L. (2017). A general framework to expand short text for topic modeling. *Inf. Sci.* 393, 66–81. doi: 10.1016/j.ins.2017.02.007

Boiy, E., and Moens, M. F. (2009). A machine learning approach to sentiment analysis in multilingual Web texts. *Inf. Retr.* 12, 526–558. doi: 10.1007/s10791-008-9070-z

Chen, H., Niu, D., and Liu, D. (2018). "MIX: multi-channel information crossing for text matching," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery, Data Mining, Computing Machinery*, (New York, NY), 110–119.

Chen, X., Mihaela, V., and Krishna, P. C. M. (2014). Mining social media data for understanding students' learning experiencesp. *IEEE Trans. Learn. Technol.* 3, 246–259. doi: 10.1109/TLT.2013.2296520

Chesley, P., Vincent, B., Xu L., and Srihari, R. K. (2006). Using verbs and adjectives to automatically classify blog sentiment. *Training* 580, 233–235.

Chou, K. L. (2009). Pre-migration planning and depression among new migrants to Hong Kong: the moderating role of social support. *J. Affect. Disord.* 114, 85–93. doi: 10.1016/j.jad.2008.06.005

Gou, S. R., and Gaikwad, R. (2016). Social media data for understanding student's problem based on clustering techniques. *Int. J. Comput. Sci. Inform. Technol.* 7, 937–941.

Graves, A., Liwicki, M., Fernandez, S., Bertolami, R., Bunke, H., and Schmidhuber, J. (2009). A novel connectionist system for unconstrained handwriting recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 855–868. doi: 10.1109/TPAMI.2008.137

Joshi, A., Balamuralia, R., Bhattacharyya, P., and Mohanty, R. (2011). "C-Feel-It: a sentiment analyzer for micro-blogs," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Systems Demonstrations* (Stroudsburg, PA: Association for Computational Linguistics), 127–132.

Kisar Koramaz, E. (2014). The spatial context of social integration. *Soc. Indic. Res.* 119, 49–71. doi: 10.1007/s11205-013-0481-x

Leavey, G., Rozmovits, L., Ryan, L., and King, M. (2007). Explanations of depression among Irish migrants in Britain. *Soc. Sci. Med.* 65, 231–244. doi: 10.1016/j.socscimed.2007.03.022

Li, L., Wang, H. M., Ye, X. J., Jiang, M. M., Lou, Q. Y., and Hesketh, T. (2007). The mental health status of Chinese rural-urban migrant workers. *Soc. Psychiatry Psychiatr. Epidemiol.* 42, 716–722. doi: 10.1007/s00127-007-0221-0

Lilleberg, J., Zhu, Y., and Zhang, Y. (2015). "Support vector machines and Word2Vec for text classification with semantic features," in *IEEE International Conference on Cognitive Informatics and Cognitive Computing* (Beijing), 136–140.

Looyestyn, J., Kernot, J., Boshoff, K., and Maher, C. (2018). A web-based social networking beginners' running intervention for adults aged 18 to 50 years delivered via a facebook group: randomized controlled trial. *J. Med. Internet Res.* 20:e67. doi: 10.2196/jmir.7862

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *Comput. Sci.* 47–61.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013b). "Distributed representations of words and phrases and their compositionality," in *Proceedings of the 27th Advances in Neural Information Processing Systems* (Cambridge, MA: M IT Press), 3111–3119.

Mikulincer, M., Shaver, P. R., Gillath, O., and Nitzberg, R. A. (2005). Attachment, caregiving, and altruism: boosting attachment security increases compassion and helping. *J. Pers. Soc. Psychol.* 89, 817–839. doi: 10.1037/0022-3514.89.5.817

Patil, S., and Gupta, A. (2016). A survey on understanding students learning experiences by mining social media data. *Imp. J. Interdiscip. Res.* 2, 161–164.

Patil, S., and Kulkarni, S. (2018). Mining social media data for understanding students' learning experience using memetic algorithm. *Mater. Today Proc.* 5, 693–699. doi: 10.1016/j.matpr.2017.11.135

Sb, A., Pb, B., and Pu, C. (2020). The role of social identity in institutional work for sociotechnical transitions: the case of transport infrastructure in Berlin. *Technol. Forecast. Soc. Change* 162:120385. doi: 10.1016/j.techfore.2020.120385

Turney, P. D. (2002). "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," in *Meeting of the Association for Computational Linguistics* (Philadelphia, PA), 417–424.

Vieira, J., and Moura, R. S. (2017). "An analysis of convolutional neural networks for sentence classification," in *Proceedings of 2017 XLIII Latin American Computer Conference* (Cordoba), 1–5.

Vo, D., and Ock, C. (2015). Learning to classify short text from scientific documents using topic models with various types of knowledge. *Expert Syst. Appl.* 42, 1684–1698. doi: 10.1016/j.eswa.2014.09.031

Yoon, J., and Kim, H. (2017). "Multi-channel lexicon integrated CNN-BILSTM models for sentiment analysis," in *Proceedings of the 29th Conference on Computational Linguistics and Speech Processing* (Taipei), 244–253.

Zerubavel, E. (2000). Social identity and social descent: some sociological reflections on population genetics. *Sociol. Forum* 15, 361–366. doi: 10.1023/A:1007581827058

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Dou, Cheng and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 13 September 2021
doi: 10.3389/fpsyg.2021.734777

Relationships Among CEO Narcissism, Debt Financing and Firm Innovation Performance: Emotion Recognition Using Advanced Artificial Intelligence

Lan Zhang[1], *Biming Liang*[1], *Datian Bi*[2]*, *Yuan Zhou*[1] *and Xiaohan Yu*[1]

[1] School of Accounting, Jilin University of Finance and Economics, Changchun, Jilin, China, [2] School of Management, Jilin University, Changchun, Jilin, China

Psychological research shows that as the main component of enterprise decision-making, CEOs are not completely rational, cognitive and psychological biases often influence their decision-making process. CEO narcissism has gradually attracted academic attention. Based on upper echelon theory and subconscious theory, this paper uses advanced artificial intelligence technology to quantify CEO narcissism as a kind of emotional intelligence. Taking A-share listed companies in China from 2010 to 2019 as research objects, this paper empirically tests the impact of CEO narcissism on debt financing and innovation performance. The results show that CEO narcissism has a significant positive impact on firm innovation performance. Debt financing plays a mediating role in the relationship between CEO narcissism and firm innovation performance. CEO narcissism can have a positive impact on firm innovation performance through debt financing. Compared with non-SOEs, SOEs' CEO narcissism has a more significant positive effect on debt financing and enterprise innovation performance. The research in this paper enriches psychology and organizational management and provides a reference for an enterprise's management decisions and for investors' investment decisions.

Keywords: advanced artificial intelligence, emotion recognition, CEO narcissism, debt financing, innovation performance

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Xuefeng Tian,
China University of Petroleum (Huadong), China
Ping Zhou,
Guangdong University of Foreign Studies, China

***Correspondence:**
Datian Bi
bdt@jlu.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *01 July 2021*
Accepted: *03 August 2021*
Published: *13 September 2021*

Citation:
Zhang L, Liang B, Bi D, Zhou Y and Yu X (2021) Relationships Among CEO Narcissism, Debt Financing and Firm Innovation Performance: Emotion Recognition Using Advanced Artificial Intelligence. Front. Psychol. 12:734777. doi: 10.3389/fpsyg.2021.734777

INTRODUCTION

Corporate innovation is influenced by executives, especially narcissistic CEOs, such as Elon Musk, CEO of Tesla and SpaceX; Steve Jobs, former CEO of Apple; Bill Gates, former CEO of Microsoft; Mingzhu Dong, former CEO of Gree Air Conditioning Co.; Yueting Jia, former CEO of Letv; and other CEOs of famous enterprises, who showed different degrees of narcissistic personality tendencies. This phenomenon has attracted the attention of social groups and scholars to encourage enterprises to successfully break through technical barriers, improve their innovation performance and achieve success. According to the upper echelons theory, the management behavior of an enterprise is deeply influenced by the executive team (Hambrick and Mason, 1984; Hambrick, 2007). Executive traits, especially CEO traits (ability, risk-taking spirit, etc.), play a crucial role

in a company's investment decision, especially in enterprise innovation (Demerjian et al., 2012; Sunder et al., 2017). According to the subconscious theory, there may be emotional structures that can be converted into consciousness in the subconscious, and the core of the subconscious is the representation of instinct (Freud, 1915). Narcissism belongs to the id and is the content of the subconscious, but it is manifested in some forms in the pre-consciousness, such as arrogance and arrogance. Experimental research on psychology shows that individual decision-making cannot be completely rational, and CEOs' psychological and personality characteristics will affect the business decisions and behavior of enterprises (Chatterjee and Poolock, 2017), especially playing a crucial role in enterprise innovation (Demerjian et al., 2012; Sunder et al., 2017). Existing studies mainly study the influence of executive traits on enterprise innovation from the perspectives of agency theory and branding theory, among which CEO narcissism is also gradually attracting academic attention (Chatterjee and Poolock, 2017). The influence of CEO narcissism on enterprise investment, financing, and innovation is gradually emerging, especially in the actual work and decision-making process, and the leadership traits formed by different CEOs' cognition will lead to different decision-making performance. However, there are relatively few studies on how CEO narcissism affects enterprise innovation performance, especially exactly how performance is affected. The current relevant studies mainly focus on CEO narcissism itself while ignoring the subjective initiative of narcissistic CEOs. In particular, the influence mechanism between the degree of CEO narcissism and enterprise innovation performance is less involved.

For enterprises, innovation will bring long-term benefits and achieve long-term development, but it is also full of uncertainty, high risk, long payback period, and other characteristics. The relevant literature has explored the factors that determine firm innovation, mainly involving the size of the firm, the structure of the market, and the characteristics of the industry (Shefer and Frenkel, 2005). However, these studies still fail to explain the phenomenon that enterprises with similar scale and external environment show great differences in innovation output. This is mainly because innovation activities are often accompanied by large capital and human resource investment, a high failure rate, and difficult to measure innovation achievements. Although existing studies have examined from different perspectives that debt financing is an important source of enterprises' access to innovation funds, they have neglected the effect of debt financing on innovation (David et al., 2008). The research of David et al. (2008) on Japanese enterprises found that the performance of enterprises that adjusted their debt structure with R&D investment activities was significantly better than that of enterprises that did not adjust their debt structure. Debt financing can provide financial support for enterprises' production, investment, innovation, research and development, and other activities and influence enterprises' innovation decision-making and efficiency. In the process of enterprise development, debt financing plays a pivotal role. Previous studies on debt financing mainly focus on corporate governance, corporate characteristics, and the external macro-environment. Kaplan et al. (2012) and Bernile et al. (2017) found that executives' decision-making styles have significant differences in the signals they send to the capital market. There are few studies on the role of debt financing based on CEOs' personal traits. As the decision makers of the company, the top management team, especially the CEO, play a key role in the major decisions of the company. Then, will CEO narcissism affect the innovation performance of enterprises? Will there be an impact on debt financing? What is the effect mechanism of debt financing on CEO narcissism and firm innovation performance? Based on the existing research and the connotation of CEO narcissism and subconscious theory, this study opens the black box between CEO narcissism and innovation performance and explores the impact of corporate debt financing on the relationship between CEO narcissism and enterprise innovation performance.

The possible contributions of this paper are as follows. First, advanced artificial intelligence technology is used to quantify CEO narcissism as a kind of emotional intelligence. Psychophysiological computing is a research method in emotional intelligence research that studies the relationship between human psychological activities and physiological changes through the analysis and calculation of physiological signals. The research of this paper organically combines the advanced artificial intelligence technology of psychology with the actual needs of organizational management, which is a bold innovation. Second, based on the theory of narcissism, connotation, and the subconscious, CEO narcissism affects the performance of enterprise innovation and its mechanism of action. In the past performance of enterprise innovation related literature from an agency theory perspective focused on the influence of the top management team for enterprise innovation decision-making, but paid less attention to the CEO personality effects on the performance of enterprise innovation. Third, this paper examines the impact of CEO narcissism on enterprise innovation performance from the perspective of debt financing and finds that the larger the scale of debt financing is, the more narcissistic CEOs will increase their investment in enterprise innovation and take more active initiative to improve innovation output and performance. This study enriches the literature in the fields of psychological theories and methods, CEO narcissism and corporate financing and provides a useful reference for current listed companies to optimize corporate governance, build a core team of senior executives, and improve enterprise innovation performance.

THEORETICAL ANALYSIS AND RESEARCH HYPOTHESIS

In recent years, narcissism has gradually become a focus of scholars. The study of narcissistic personality traits began in psychology and is regarded as a dark personality trait and a personality disorder. Since the 1980s, as psychologists and sociologists have conducted a series of scientific and experimental studies on the narcissistic personality, later scholars tend to regard the narcissistic personality as a relatively stable personality trait, which is not commendatory

in itself (Wallace and Baumeister, 2002). In particular, with the development of the narcissistic personality inventory (NPI), the quantification of narcissistic personality traits has been realized. Since then, an increasing number of management scholars have incorporated it into their research field. Existing psychological studies have shown that personality has a significant impact on information processing and decision making, and individual decision making has difficulty achieving complete rationality. Narcissism is divided into explicit narcissism and implicit narcissism, but no matter what type of narcissism, the focus is on the self, desire, performance, and seeking the unique self.

Narcissism is a cognitive response (Lee et al., 2016). The main characteristics of narcissistic personality traits are as follows: First, narcissistic personality traits are relatively stable, basic, and deep-rooted (Campbell et al., 2004), usually endogenous, and less affected by external intervention and influence (Olsen, 2016). Second, narcissism is a combination of self-awareness and strong motivation, which affects external events from an internal perspective. Narcissists create a positive self-image by realizing their social status and self-worship. At the same time, narcissists have motivations such as yearning for their own rights and craving for the affirmation of others. Finally, Chatterjee and Hambrick (2007) pointed out that narcissistic personality traits include three important components: cognition, motivation, and behavior strategy. In terms of cognition, narcissistic CEOs have extremely inflated self-concepts and believe that they have privileges. They are always self-centered, believing that their decisions are always right and resisting others' disagreements (Campbell et al., 2011). In terms of motivation, narcissistic CEOs strongly pursue the power to control others and the praise of others to maintain their inflated self-concept (Chatterjee and Poolock, 2017). In terms of behavior, narcissistic CEOs usually adopt two strategies of "self-improvement" and "self-defense" to achieve their pursuit of power and the praise of others. On this basis, this paper summarizes the narcissistic personality characteristics as follows: narcissists are gifted, have extraordinary intelligence, and attraction, are good at sketching grand visions, and are very eager to exercise power. These characteristics urge narcissists to constantly seek recognition, affirmation and praise to maintain their sense of superiority.

The Impact of CEO Narcissism on Debt Financing

CEO narcissism often underestimates risk and overestimates earnings, leading to more radical risk taking. An "above average" bias in judging and making decisions by narcissistic CEOs (Larwood and Whittaker, 1977; Alicke, 1985) is mainly characterized by two aspects: (1) It is easy to overestimate the possibility of success of an event while ignoring its risk. (2) It is easy to attribute success to one's own ability and attribute failure to poor luck (Miller and Ross, 1975). CEO narcissism has an important impact on corporate financing strategy and capital structure. In recent years, scholars have carried out a series of theoretical and empirical studies on the impact of CEO narcissism on corporate financing preference and achieved certain results. Shefrin and Statman (1999) argues that narcissistic CEOs overestimate the company's future earnings which they do not want to share with new shareholders; they prefer to issue bonds rather than stocks when choosing means of financing. Heaton (2002) confirmed the pecking order financing theory from the perspective of over-optimism caused by CEO narcissism. Overly narcissistic CEOs will think that a company's value is underestimated by the market. If there is not sufficient cash, even if there are investment projects with positive net present value, CEOs will refuse because of the high financing cost. If companies have to choose external financing, CEOs will also prefer debt financing because stock prices are more sensitive to market expectations. Therefore, the financing order of narcissistic CEOs is endogenous financing-debt financing-equity financing. Hackbarth (2008) study shows that narcissistic CEOs overestimate the profitability of investment projects and the ability of enterprises to reduce risks and use higher debt ratios to issue debt more frequently, especially short-term debt. Hackbarth's model shows that narcissistic CEOs believe that their companies and investment projects can recover cash flow as soon as possible. Hackbarth also found that debt financing could raise corporate value in two ways: first, by limiting CEOs' arbitrary transfer of funds to reduce agency costs; second, by reducing conflicts between shareholders and creditors. The empirical study of Malmendier and Tate (2005) found that narcissistic CEOs are more cautious about external financing than ordinary leaders. When overseeing external financing, they are willing to choose more debt financing and less stock issuance. In addition, external institutions, especially the news media, tend to pay more attention to the risky decision-making of enterprises, and debt financing for R&D investment is a risky behavior that can bring more attention to enterprises and to some extent can satisfy the narcissistic CEO's desire for external attention and increase their sense of superiority. Based on these observations, the following assumptions are made:

H1: CEO narcissism is positively correlated with debt financing.

The Impact of CEO Narcissism on Corporate Innovation Performance

Innovation is the driving force of enterprise growth, and R&D investment can guarantee innovation results. The sunk costs, opportunity costs, and inherent risks of R&D investment once made many CEOs unwilling to invest more in R&D innovation. However, narcissistic CEOs are different. Driven by desire and achievement motivation, narcissistic CEOs' behavior has an obvious risk seeking tendency. In strategic decision-making, narcissistic CEOs are more inclined to "high risk and high yield" schemes. To seek opportunities to show their superiority, narcissistic CEOs are more willing to adopt cutting-edge technology. Overconfidence makes narcissistic CEOs overestimate returns, underestimate risks, and think that the probability of success of R&D innovation is greater (Campbell et al., 2004). Gerstner et al. (2013) found that narcissistic CEOs tend to adopt cutting-edge technology to obtain the stakeholders' general appreciation. Kumar (2019) points out that narcissistic leaders tend to be bold. In order to meet their own

needs, narcissistic leaders often take risks and accept challenges, and they have strong internal driving forces. Therefore, they will positively affect employees' innovative behavior and inject fresh vitality into organizational change. The characteristics and difficulties of innovation projects can often attract the attention of narcissistic CEOs. On the one hand, with high uncertainty, the risk of innovation projects is greater and more challenging. Once successful, it can often demonstrate the talents of CEOs, thus attracting more attention and praise. On the other hand, the success of innovation projects can often become a key breakthrough in the growth momentum of enterprises, increase the scientific and technological content and added value of products, accelerate the process of enterprise innovation and improve enterprise value. Therefore, narcissistic CEOs will increase the intensity of innovation investment.

Based on this motivation, narcissistic CEOs with their own capabilities will adopt a series of corporate decisions to promote the implementation of innovation projects and the improvement of innovation performance. CEO narcissism will pay more attention to innovation projects (Hirshleifer et al., 2012). Rosenthal and Pittinsky (2006) found in relevant studies that narcissistic CEOs will work harder than ordinary CEOs to maintain their sense of superiority. Narcissistic CEOs can significantly promote the innovation performance of knowledge workers and help to create a cultural atmosphere of innovation in enterprises to promote the innovation output performance of enterprises. Compared with non-narcissistic CEOs, narcissistic CEOs can better identify valuable risk projects (Gervais et al., 2011) to obtain innovation results. To better promote the improvement of enterprise innovation performance, narcissistic CEOs tend to modify the enterprise strategy, prompting enterprises to adopt more science and technology strategies (Galasso and Simcoe, 2011) so that the enterprise goals, profit model, budget management, organizational structure, and daily operation of departments are closely related to the science and technology strategy and provide comprehensive cooperation for the implementation of innovation projects. Narcissistic CEOs also tend to recruit more innovative employees (Steen, 2005), which is conducive to the team's efforts toward common goals and missions. The degree of organizational identity and organizational consistency is higher, which can guarantee the smooth implementation of innovative projects. Based on this, the following assumptions are made:

H2: CEO narcissism is positively correlated with corporate innovation performance.

The Mediating Role of Debt Financing on the Relationship Between CEO Narcissism and Corporate Innovation Performance

Enterprise innovation activity is a high-risk decision-making behavior. The initial stage of innovation activity requires a large amount of capital investment and the cooperation of various resources. The process is full of uncertainties. It may not be profitable for many years, or it may end in failure (Ham et al., 2018). Therefore, a CEO who can make innovation investment decisions must have certain power, dare to take risks, and have enough courage and the ability to take risks, and be willing to invest a large amount of debt financing in innovation and research of the enterprise. Narcissistic type CEOs have these characteristics. For example, Buyl and Boone (2017) found a positive correlation between CEO narcissism and corporate risk-taking. Because narcissistic CEOs underestimate the risk of failure of innovative projects, when the company is in a situation of high debt financing, they often adopt aggressive R&D investment strategies. The behavior of an enterprise to maintain its reputation and brand through innovation investment will be regarded by creditors as a signal for the enterprise to initiate active competitive projects. It is easy to obtain the approval of creditors and can reduce the company's financing costs. Narcissistic CEOs strongly pursue power and honor, and this vote of confidence will undoubtedly satisfy the psychological needs of narcissistic CEOs. At the same time, the information asymmetry, and asset substitution problems caused by external financing will also encourage narcissistic CEOs to engage in risk-taking behaviors and increase investment in innovation. On the other hand, under the constraints of financing, companies will strive to improve the efficiency of R&D and innovation. Based on the positive effects of narcissistic CEOs and corporate innovation performance described in the previous article, corporate innovation performance may be improved.

Based on the above analysis, in this article, it is proposed that the improvement in corporate innovation performance is the joint result of CEO narcissism and the increase in debt financing. CEOs with narcissistic characteristics can better optimize the allocation of resources invested and guarantee realization of the R&D investment from all aspects. With increased efficiency, CEO narcissism results in improved innovation performance. Narcissistic CEOs overestimate the profitability of companies or investment projects. They prefer lower-cost but higher-risk financing methods and debt financing, especially short-term debt financing. Debt financing also affects corporate innovation decisions through constraints and incentives on corporate managers. Narcissistic CEOs are more adventurous, and adventurous managers can alleviate the problem of underinvestment, which is conducive to enhancing corporate value (Thakor and Goel, 2008). Narcissistic CEOs are often more creative and can discover more opportunities for innovation. To better promote the improvement of corporate innovation performance, narcissistic CEOs will proactively acquire and integrate the information and resources needed for innovation, and by modifying corporate strategies, they will encourage companies to adopt more technology strategies. According to the paradigm of psychology, debt financing acts as a bridge, and facilitator in the relationship between CEO narcissism and corporate innovation performance based on the psychological characteristics of narcissism. Leaders' irrationality is transmitted to the company through the financing decisions they make and then affects the company's performance. Therefore, debt financing plays an intermediary role in the influence of CEO narcissism on the company's innovation performance. Based on this, the following hypothesis proposed:

H3: Debt financing plays a mediating role in the relationship between CEO narcissism and corporate innovation performance.

The Moderating Effect of the Nature of Enterprise Ownership

State-owned enterprises and non-state-owned enterprises have different structures, and the nature of corporate ownership may have different effects. This article speculates that state-owned enterprises have a significant positive moderating effect on the relationship between CEO narcissism, debt financing and corporate innovation performance. Based on the theory of resource dependence, state-owned enterprises are an important subject of national independent innovation and an indispensable part of the national innovation system. Compared with non-state-owned enterprises, state-owned enterprises shoulder more important social responsibilities to promote the high-quality development of the national economy through technological innovation. At the same time, state-owned ownership is regarded as a resource-rich external shareholder that can provide reliable resource support for enterprise technological innovation. State-owned enterprises have concentrated national high-quality R&D resources and possess strong innovation capabilities. They are the backbone of my country's innovation-driven development. CEOs with a high degree of narcissism will make full use of national innovation resources, carry out debt financing, increase R&D investment, and improve corporate innovation performance. State-owned enterprises can reduce the uncertain risks caused by system and policy changes and the uncertainty of the external environment, thereby promoting enterprise innovation (Choi et al., 2011). In recent years, the Chinese government has placed technological innovation at the core of national development and has made a series of major decisions and deployments for the implementation of an innovation-driven development strategy. Based on this, the narcissistic CEOs of state-owned enterprises actively cater to national policies to demonstrate their image as excellent leaders and then increase debt financing to promote R&D investment and improve corporate innovation performance. In addition, to obtain promotion and attention from the outside world, compared with non-state-owned enterprises, narcissistic CEOs of state-owned enterprises pay more attention to the improvement of the company's innovation performance. Innovation performance has become an important criterion for measuring the ability of enterprise managers, and it is also an important part of the promotion and assessment of state-owned enterprise managers. Narcissistic CEOs are more energetic and confident, make decisions more decisively and are more inclined to increase debt financing to increase corporate R&D investment to establish their image as a "visionary." Therefore, the following hypotheses are put forward:

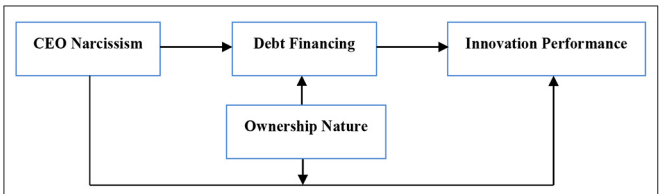


FIGURE 1 | Research framework.

H4a: Compared with non-state-owned enterprises, the CEO narcissism of state-owned enterprises has a more significant impact on debt financing.

H4b: Compared with non-state-owned enterprises, the CEO narcissism of state-owned enterprises has a more significant impact on innovation performance.

Based on the above analysis, the research framework constructed in this paper is shown in **Figure 1**.

RESEARCH DESIGN

Variable Definition and Measurement

CEO Narcissism

Researchers that studied artificial intelligence incorporated the theory of multiple intelligences and the concept of emotional intelligence; thus, a new direction of cognitive intelligence and emotional intelligence emerged. Emotional intelligence is based on the existing technology of artificial intelligence, and on this basis, it enhances the ability of artificial intelligence to understand emotion, generate emotion, and express emotion to construct human-like intelligence with emotion and empathy. The introduction of emotional intelligence makes artificial intelligence not only intelligent but also emotional. Emotional intelligence is a machine that finds and identifies human emotions by analyzing users' facial expressions, physiological signals, eye movements, voices, and content released on social media. It understands emotions, generates empathy, and expresses emotions. Emotional intelligence can be applied to medical treatment, education, production, commerce, entertainment, and other fields by giving machines emotion to achieve harmonious human-computer interaction. When quantifying CEO narcissism, this paper applies the new discovery of emotional intelligence to quantify the degree of CEO narcissism by identifying the content released by CEOs on social media.

CEO narcissism (Ceonar) refers to the growing self-consciousness of CEOs and their focus on self-concept (Campbell et al., 2011). At present, the most commonly used method to measure the degree of individual narcissism is the Narcissism Personality Inventory (NPI). However, CEOs of listed companies are usually not willing to take the sensitivity test of individual personality traits. Even if some CEOs are willing to take the test, the test results will also be affected by subjective biases caused by social expectations. Therefore, based on the study of Olsen et al. (2014) and Chatterjee and Hambrick (2011), this paper selects objective measurement indicators according to the four narcissism dimensions proposed by Emmons and Robert (1987). Combined with China's national conditions, individual indicators are excluded, and new indicators are added. Finally, four objective measurement indicators are used to measure the degree of CEO narcissism. The relationship between the selected indicators and the four sub-dimensions of CEO narcissism is shown in **Table 1**.

To further test the consistency of the four objective indicators, A-mos21.0 was used for confirmatory factor analysis (CFI = 0.91, EFI = 0.88, NNFI = 0.87, RMSEA = 0.08) and reliability analysis. The Cronbach's α value of the measurement item is

TABLE 1 | The relationship between the selected indicators and the four sub-dimensions of CEO narcissism.

Four subdimensions of narcissism (Emmons and Robert, 1987)				
	Leadership/authority	**Arrogance/self-worship**	**Superiority/arrogance**	**Utilitarian/power desire**
Typical items in Narcissistic Personality Inventory (NPI)	People always recognize my authority. I was born a leader	I often admire myself in the mirror. I am an excellent person.	I love being the focus of attention. I like to be above others	I must get others awe and respect. I envy others' achievements
The correspondence between indicator meaning and narcissism subdimensions				
Indicators selected in this article				
Proportion of CEO news reports on company home page (sub1)	I am the core of the company.	I like the high attention CEO brings me		My site, I 'm the owner, I 'm the company news
The proportion of CEOs using the first person singular (sub2)	Corporate performance is my own achievement, not my team		I am the company, the company is me	My importance should be highlighted.
Proportion of original social media works (sub3)		My point is right. I am the embodiment of truth.	I am the example of others, they should imitate me, follow my steps	

0.68, which is higher than the minimum acceptable value of 0.65, indicating that the CEO narcissism measurement index has good reliability and validity. Based on the above tests, the average value of the four indicators is finally standardized as the CEO narcissism index.

Debt Financing

Debt financing (Debt) represents the scale of corporate debt financing, which is the intermediary variable of this study. As China's current financial system and banking system are relatively stable, banks remain dominant in corporate external financing. In this paper, the total amount of bank loan, namely the balance of long-term loan and short-term loan, is used to measure debt financing. The total amount of bank loan reflects the total ability of enterprises to borrow (Lu et al., 2012). So in light of Kim and Lee (2008) and other researchers, the ratio of bank loans to total corporate assets is used to measure the scale of corporate debt financing, which is the sum of long-term and short-term borrowings.

Innovation Performance

The dependent variable innovation performance (IP) includes innovation input and innovation output. Innovation input is measured by R&D expenditure/operating income, and innovation output is measured by the number of patent applications. Enterprises generally apply for patents in the year when they invent new technologies, and the year of patent authorization may lag behind the year of patent application. In addition, compared with patents, the value of new products must be tempered by the fact that enterprises do not apply for patents anonymously; thus, patent applications can measure innovation output more comprehensively, and can better reflect the return of enterprise innovation by reflecting the degree of market acceptance of new products (Wang and Kafouros, 2009). Therefore, it is more objective to measure innovation output performance by the number of patent applications. The greater the number of invention patent applications is, the higher the innovation performance of enterprises is.

Nature of Ownership

Ownership nature, according to the nature of the actual controller, is divided into state-owned (SOE) and non-state-owned enterprises, state-owned enterprises for 1, and non-state-owned enterprises for 0.

Control Variables

Drawing on previous studies on CEO narcissism, debt financing and innovation performance, control variables are mainly introduced at the enterprise level, corporate governance level and other levels. The control variables at the firm level include firm size (Size) and firm return on total assets (Roa). The control variables of corporate governance include CEO age (Ceoage), CEO education (Education), board size (Bsize) and the proportion of independent directors (Indir). The above variables and explanations are shown in **Table 2**.

Model Construction

To test the relationships among CEO narcissism, corporate debt financing and corporate innovation performance, this paper establishes the following equations:

$$\begin{aligned} \text{Debt} &= \phi_0 + \phi_1 \text{Ceonar} + \phi_2 \text{Size} + \phi_3 \text{Roa} + \phi_4 \text{Ceoage} \\ &+ \phi_5 \text{Education} + \phi_6 \text{Bsize} + \phi_7 \text{Indir} + \varepsilon \quad (1) \\ \text{IP} &= \beta_0 + \beta_1 \text{Ceonar} + \beta_2 \text{Size} + \beta_3 \text{Roa} + \beta_4 \text{Ceoage} \\ &+ \beta_5 \text{Education} + \beta_6 \text{Bsize} + \beta_7 \text{Indir} + \mu \quad (2) \\ \text{IP} &= \lambda_0 + \lambda_1 \text{Ceonar} + \lambda_2 \text{Debt} + \lambda_3 \text{Size} + \lambda_4 \text{Roa} \\ &+ \lambda_5 \text{Ceoage} + \lambda_6 \text{Education} + \lambda_7 \text{Bsize} + \lambda_8 \text{Indir} \\ &+ \sigma \quad (3) \end{aligned}$$

TABLE 2 | Definition and measurement of variables.

Variable type	Variable name	Symbol	Measurement method
Dependent variable	Innovation performance	IP	Number of patent applications
Explanatory variable	CEO narcissism	Ceonar	CEO narcissism index
Mediating variable	Debt financing	Debt	(Short-term loans + long-term loans) / total assets
Moderating variable	Nature of ownership	SOE	dummy variable, state-owned enterprise value 1, non-state-owned enterprise value 0
Control variables	Enterprise size	Size	Ln (Total assets)
	Return on total assets	Roa	Net profit / total assets
	CEO age	Ceoage	Actual age of the year (age)
	CEO degree	Education	≤specialist 1, undergraduate 2, master 3, doctor 4
	Number of board of directors	Bsize	Number of formal members of the board of directors
	Proportion of independent directors	Indir	Proportion of independent directors

Model (1) tests the relationship between CEO narcissism and debt financing; if ϕ_1 is significantly positive, then it confirms hypothesis H1. Through model (2) testing the relationship between CEO narcissism and enterprise innovation performance, if β_1 is significantly positive, then hypothesis H2 is confirmed. Model (3) tests the mediating effect of debt financing on the relationship between CEO narcissism and enterprise innovation performance. To test the moderating effect of the nature of enterprise ownership, this paper uses the above three models to test the grouping of state-owned enterprises and non-state-owned enterprises.

Sample Selection

This paper selects 2010–2019 Shanghai and Shenzhen A-share listed companies as the initial research sample. Sample selection follows the following principles: (1) Financial and insurance listed companies are excluded. (2) The listed companies whose transaction statuses are ST and *ST are excluded. (3) Companies whose listing year is <3 years or whose CEO is <3 years are excluded. (4) Incomplete financial indicator data on CEO narcissism, debt financing, innovation performance or missing listed companies are excluded. After the above steps, a total of 402 sample enterprises and a total of 2,332 observations constitute the company-year data sample; among them, 103 state-owned enterprises, a total of 598 sample points and 299 non-state-owned enterprises, a total of 1,734 sample points. CEO narcissism data are obtained through a large amount of text mining through financial websites, Baidu search engines, microblog and blog network communication platforms, and corporate homepage news. The sample enterprise financial data are obtained from the CSMAR and Wind databases.

EMPIRICAL RESEARCH

Descriptive Statistics

Table 3 shows the descriptive statistical results of the relevant variables. As shown in **Table 3**, the minimum value of innovation performance (IP) is 0, the maximum value is 824, and the standard deviation is 118.91, indicating that different enterprises have substantial differences in innovation performance. The minimum value of the CEO narcissism index (Ceonar) is 0.020, and the maximum value is 0.589. The mean value is 0.112, to a certain extent, indicating that the degree of narcissism of CEOs varies greatly, and the majority of people have a low degree of narcissism. The minimum value of debt financing (Debt) is 0, and the maximum value is more than 50%. There is a large difference between the two values. The average difference is 0.181, which indicates that the total debt financing of listed companies in China mainly consists of bank loans that amount to more than 18%. The standard deviation is 0.146, which indicates that the debt financing scale obtained by different enterprises differs greatly. The descriptive statistical analysis of other variables is shown in **Table 3** and will not be repeated.

TABLE 3 | Descriptive statistical results of variables.

The variable name	Mean value	Standard deviation	Minimum value	Maximum value
IP	45.73	118.91	0.00	824.00
Ceonar	0.112	0.208	0.020	0.589
Debt	0.181	0.146	0.000	0.569
Size	24.126	1.351	19.425	26.536
Roa	0.052	0.063	−0.183	0.192
SOE	0.412	0.498	0.000	1.000
Ceoage	46.892	5.726	30.000	68.000
Education	3.655	0.729	1	4
Bsize	8.725	1.826	5.000	15.000
Indir	0.384	0.061	0.095	0.800

Correlation Analysis

Table 4 carries out Pearson test on the correlation of main variables. The results show that CEO narcissism, debt financing and innovation performance pass Pearson correlation test. In addition, the absolute value of the maximum pairwise correlation coefficient between variables is 0.379, indicating that there is no serious multicollinearity between variables.

TABLE 4 | Correlation analysis results.

变量	1	2	3	4	5	6	7	8	9	10
1. IP	1									
2. Ceo_Nar	0.065**	1								
3. Debt	0.084***	0.112***	1							
4. Size	0.379***	0.181**	0.198**	1						
5. Roa	−0.051*	−0.034	−0.177**	0.336***	1					
6. SOE	0.127***	0.265**	0.063**	0.246**	0.017*	1				
7. Ceoage	0.037	−0.128**	0.092*	0.037	0.014*	−0.019	1			
8. Education	0.122***	0.009	−0.005**	−0.017***	0.088*	0.004	−0.012	1		
9. Bsize	0.021*	0.018	0.155**	−0.002**	0.013*	−0.034	0.018	0.011	1	
10. Indir	0.256	0.023*	−0.009	−0.025	−0.030	0.041*	0.072	0.071*	0.056	1

*$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$.

TABLE 5 | Regression analysis results.

Variables		The dependent variable (Debt)		The dependent variable (IP)		
		Model 1	Model 2	Model 3	Model 4	Model 5
Control variables	Size	0.081 (74.32)	0.313*** (5.24)	−0.230*** (−3.78)	−0.155*** (−2.36)	−0.117** (−1.92)
	Roa	0.126** 2.11	0.123** (2.01)	0.002 (3.21)	0.003 (3.12)	0.002 (3.19)
	Ceoage	0.142** (3.95)	0.136** (3.62)	−0.021* (−2.31)	−0.016* (−1.68)	−0.013* (−1.78)
	Education	0.026* (1.15)	0.017* (1.11)	0.152** (2.78)	0.191*** (3.12)	0.175*** (2.78)
	Bsize	0.002*** (3.19)	0.0014* (2.25)	0.098*** (1.67)	0.084*** (1.72)	0.080*** (1.21)
	Indir	0.051* (3.69)	0.005 (3.35)	0.51 (1.745)	0.63 (1.911)	0.57 (3.79)
	Year	Control	Control	Control	Control	Control
	Industy	Control	Control	Control	Control	Control
The independent variable	Ceonar		0.282*** (5.26)		0.166** (2.58)	0.096* (1.58)
Intervening variable	Debt					0.134* (2.16)
Adjusted R2		0.501	0.338	0.847	0.470	0.261
F		405.2	206.46	86.65***	48.892***	10.25***

*$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$.

Hypothesis Testing

Main Effect Test

This paper uses regression analysis. First, the impact of CEO narcissism on debt financing is analyzed. Second, the influence of CEO narcissism on innovation performance is investigated. Finally, the mediating effect of debt financing on the relationship between CEO narcissism and enterprise innovation performance is tested in the next part. The results are shown in **Table 5**.

Model (1) and Model (2) examine the relationship between CEO narcissism and debt financing. Model (1) is an estimate containing only control variables, Model (2) takes CEO narcissism as the explanatory variable and debt financing as the explained variable. The results show that H1 passes the test ($\beta = 0.282$, $P < 0.001$). Model (3) and Model (4) examine the relationship between CEO narcissism and innovation performance. Model (3) is an estimated result containing only control variables, Model (4) takes CEO narcissism as the explanatory variable and innovation performance as the explained variable. The results show that H2 is tested ($\beta = 0.166$, $p < 0.005$).

Mediating Effect Test

For the mediating effect, this paper will use the 3 steps described by Baron et al. to verify its significance and use the Sobel, Aroian, and Goodman tests to further verify its significance. According to Baron (1986), there are three regression steps for mediator variables: step 1 is the measurement of the relationship between the predictor variable and the dependent

TABLE 6 | Measurement steps of mediating variables.

Steps			β value	Establishment condition
Step1	The independent variable	The dependent variable	β1	β1 should be significant
Step2	The independent variable	Intervening variable	β2	β2 should be significant
	The independent variable	The dependent variable	β3	β4 should be significant, β1> β3
Step3	Mediating variable	The dependent variable	β4	β3 is completely true without significance: The significance is partially true

TABLE 7 | Mediating effect test results.

Step	Explanatory variable	Explained variable	β value	Hypothesis condition
Step1	The independent variable	The dependent variable	β1	β1 should be significant
	Ceonar	IP	0.166**	significant
Step2	The independent variable	Mediating variable	β2	β2 should be significant
	Ceonar	Debt	0.282***	significant
Step3	The independent variable	The dependent variable	β3	β3 should be significant
	Ceona	IP	0.096*	significant,0.166>0.096,partial mediating effect holds
	Mediating variable		β4	β4 should be significant
	Debt		0.134*	significant

*$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$.

variable, and its β value should be significant. Step 2 is the measurement of the relationship between the independent variable and intermediate variable, and its β value should also be significant. Step 3 considers independent variables and intermediary variables simultaneously and measures the relationship between them and dependent variables. At this point, the β value between the independent variable and the dependent variable is lower than the β value in Step 1, and it is completely true for those that are not significant, while it is partially true for those that are significant. However, the relationship between the mediator variable and the dependent variable is still significant, as shown in **Table 6**.

This paper uses the mediated regression method to study the relationship between CEO narcissism, debt financing and innovation performance. The specific test results are shown in **Table 7**.

The results in **Table 7** show that both Step 1 and Step 2 meet the conditions. In Step 3, the partial mediating effect of debt financing on the relationship between CEO narcissism and innovation performance is established, since 0.166 > 0.096 meets the principle of partial mediating effect described by Baron et al. Therefore, H3 is partially supported. **Figure 2** shows the mediating effect results of debt financing.

To further verify the mediating effect, the Sobel method and the bootstrap method were selected for retesting, **Table 8** shows the mediating effect test results. These two methods could reduce the impact of the second type of error on the research results. Bootstrapping is currently the most ideal, and the latest inspection intermediary effect method has been applied in different disciplines. The method of samples takes back the whole sampling method for repeated sampling. Each time, the parameters obtained from the sample average are the final results,

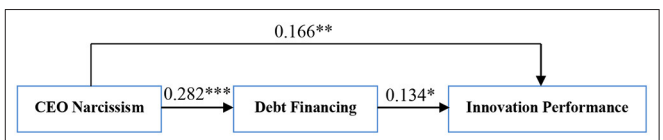


FIGURE 2 | Debt financing mediating effect results. *$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$.

TABLE 8 | Mediating effect test results.

Sobel test	Bootstrap test (95% confidence interval)	
Z value	The lower limit	Ceiling
1.69***	0.0005	0.002

***$p < 0.01$.

have high effectiveness in statistical results, and provides more reliable results. If the Sobel test results show a significant Z value, the confidence interval of the bootstrap test results does not contain 0, indicating that the mediation test has been passed. **Table 6** shows the results of the Sobel test and bootstrap test. The Z-value is 1.69, which is significant at the level of 1%, passing the Sobel test. The confidence interval of the bootstrap test is 0.0005 to 0.002, excluding 0, and the result is robust after the test.

Test of the Moderating Effect of the Nature of Enterprise Ownership

According to the nature of ownership, the samples are divided into state-owned enterprises and non-state-owned enterprises for regression. **Table 9** shows the regression results. The results of Model 1 and Model 2 show that in SOEs, CEO narcissism is

TABLE 9 | Moderating effect test results.

Variable	Debt		IP	
	Model 1	Model 2	Model 3	Model 4
	State-owned	Non-state	State-owned	Non-state
Size	0.286*** (4.79)	0.178*** (2.95)	−0.174*** (−2.19)	−0.109* (−1.76)
Roa	0.214* (3.47)	0.112* (1.86)	0.005 (1.12)	0.001 (1.03)
Ceoage	0.125* (1.96)	0.347* (1.978)	−0.081* (−1.28)	−0.098* (−1.69)
Education	0.013* (0.522)	0.096** (1.575)	0.265*** (4.34)	0.123* (2.12)
Bsize	0.107 (1.75)	0.031* (0.513)	0.064* (1.01)	0.104*** (1.79)
Indir	−0.026 (−0.29)	0.013 (0.211)	0.251 (3.97)	0.302 (4.19)
Year	Control	Control	Control	Control
Industy	Control	Control	Control	Control
Ceonar	0.385*** (5.26)	0.164*** (2.581)	0.284** (4.76)	0.133** (2.12)

*$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$.

significantly positively correlated with debt financing. Compared with non-SOEs, CEO narcissism in SOEs has a greater impact on debt financing ($0.385 > 0.164$). H4a is verified. The results of Models 3 and 4 showed that CEO narcissism is significantly positively correlated with innovation performance. Compared with non-state-owned enterprises, the CEO narcissism of state-owned enterprises has a greater impact on innovation performance ($0.284 > 0.133$). H4b is verified.

Robustness Test

To ensure the reliability of the research conclusions, we carried out a series of robustness tests. For the CEO signature size data used in this paper, the approach of Ham et al. (2017, 2018) was used to measure the degree of narcissism of CEOs. The greater the value is, the higher the degree of narcissism of the CEO. IPO prospectuses and prospectuses[1] of listed companies from 2010 to 2019 from Juchao Information Network, including 2,663 prospectuses and 1,043 prospectuses, a total of 3,706 documents[2] were downloaded. Then, Python deep learning software was used to collect the CEO signature size. The specific steps are as follows: First, the CEO name coordinates are manually determined. The program then returns the number of pixels that record the size of the signature and the area it occupies. Then, the data are standardized to obtain the area of the signature. Finally, the data obtained by dividing the signature area by the number of words in the name were taken as the proxy variable of CEO narcissism, which was substituted into the model for regression analysis, and the research conclusion remained unchanged.

[1]The source of the CEO's signature is the IPO prospectus and offering prospectus of the listed company. The reason is that on May 17, 2006, the 180th Chairman's Office Meeting of China Securities Regulatory Commission approved the Measures for the Administration of Initial Public Offering and Listing. Article 43 of the Measures stipulates that the issuer and all of its directors, supervisors and senior managers shall sign and seal the prospectus/offering prospectus to ensure that the contents of the prospectus are true, accurate and complete. As the chairman of the sponsor or general manager to bear the corresponding legal responsibility, there will be no high-definition pictures affixed with the signature. The prospectus and the prospectus have a unified format, which facilitates the unified quantification of the signature size of the CEO in this paper.

[2]This paper takes the signature part of the prospectus as the main data source and makes a detailed comparison of the overlap with the CEO's signature in the prospectus and finds that the signature size of the two is basically the same. For this reason, the part about overlap is the prospectus.

CONCLUSION AND LIMITATION

Research Conclusion

Although existing studies have explored the impact of CEO narcissism on enterprise innovation performance from different perspectives, the research is not perfect, and there are few studies from the perspective of debt financing. Based on the sample of A-share listed companies in China from 2010 to 2019, this paper incorporates CEO narcissism, corporate innovation performance and debt financing into a framework for analysis. Debt financing is used as an intermediary variable to test its role in the relationship between CEO narcissism and corporate innovation performance. The empirical results show that CEO narcissism is positively correlated with debt financing. The higher the degree of CEO narcissism is, the larger the scale of debt financing is. CEO narcissism is positively correlated with enterprise innovation performance. The higher the degree of CEO narcissism is, the higher the enterprise innovation performance is. Debt financing plays a mediating role between CEO narcissism and corporate innovation performance. Compared with non-state-owned enterprises, CEO narcissism has a more significant impact on debt financing and innovation performance. This shows that the improvement of corporate innovation performance is the common result of CEO narcissism and the increase of debt financing. CEOs with narcissistic characteristics can choose debt financing to alleviate the problem of insufficient investment capital of enterprises, optimize the allocation of resources invested, and ensure the improvement of transformative efficiency of R&D investment in all aspects. Therefore, debt financing will improve the innovation performance of narcissistic CEOs through the role of bridges and intermediaries.

The research conclusions provide a new behavioral perspective for corporate innovation motivation and enrich the research on the economic consequences of CEO narcissism. This paper expands the analytical framework of the relationship between CEO narcissism and corporate innovation performance by introducing corporate debt financing. This provides a reference for enterprises to obtain debt financing, for creditors to make loan decisions and some reference ideas and a theoretical basis for future research. Through empirical research, this paper finds that CEO narcissism is of great significance to enterprise innovation and corporate debt financing behavior. It provides more specific guidelines, mainly reflected in the following: it provides effective reference value for enterprises to appoint executives, provides more effective guidance in practice, and provides effective ideas for China to realize the development of

its innovative companies. For an enterprise, an excellent CEO is reflected not only in his/her ability to deal with various affairs but also in his/her personality traits and self-awareness. Personal traits and psychological risk preferences should be matched with the company's image, positioning and overall strategy. Therefore, it is necessary to pay attention to the selection of CEOs, not only to examine their personal skills but also to integrate the CEO's own personality and psychological characteristics into the assessment system and pay attention to the level of narcissism, not only to avoid the negative impact of excessive narcissism on their personality and effective supervision of CEOs but also to pay attention to the positive role of narcissistic CEOs in enterprises. For creditors, attention should be given to the personal characteristics of the CEO of the debt enterprise to measure the CEO's ability to protect the creditor's rights and collect repayment from debtor companies.

Although some foreign studies have focused on leader narcissism as a starting point to study R&D investment behavior, China's economic, social and cultural environment is not the same as that of Western countries, so whether foreign scholars' conclusions can be further confirmed in China remains to be further explored. In the background of economic transformation and supply-side structural reform in contemporary China, China's enterprise innovation should not only focus on innovation input but also pay attention to performance output to avoid ineffective allocation and waste of resources. By comparing the differences between state-owned and non-state-owned enterprises in R&D investment, it is found that state-owned enterprises are characterized by high R&D investment and low innovation efficiency. It can be seen that state-owned enterprises have obvious advantages in innovation input, but what needs to be solved urgently is how to solve the problem of low innovation efficiency and alleviate the principal-agent conflict. Debt financing can provide financial support for enterprises' production, investment, innovation and research and development activities, and affect enterprises' innovation decisions and efficiency. The conclusion of this paper will help optimize the debt structure of state-owned enterprises and promote the reform of state-owned enterprises from the perspective of innovation-driven development. One is that Chinese listed companies should improve their articles of association and increase the terms of supervision over CEOs. The second is that enterprises should do a better job of separation of power and responsibility, improve corporate governance, and pay attention to enterprise R&D innovation. However, the decision-making mechanism should avoid the irrational investment and financing behavior of narcissistic CEOs caused by excessive concentration of power, and a benign balance between CEOs and executive teams should be formed. Third, we can consider granting certain benefits to the CEO for the achievement of results in which the CEO is consistent with the interests of the enterprise. Fourth, enterprises should strengthen internal audits, follow-up attention to corporate debt financing, timely correction of deviations, and control of corporate financial risks. Fifth, to create a healthy growth environment for narcissistic CEOs, the government should promote a good atmosphere of innovation and development, and strengthen the guidance of narcissistic CEOs in terms of the long-term interests of enterprise-oriented decision-making.

RESEARCH LIMITATIONS

There may be some deficiencies in this study. First, since narcissism is an extremely complex psychological and personality feature, the objective indicators collected manually in this paper have certain errors, which cannot cover all the characteristics of narcissism and can only partially reflect some aspects of the narcissistic personality. Second, this paper mainly discusses the influence mechanism between CEO narcissism and enterprise innovation performance from the perspective of CEOs' own characteristics and debt financing. However, due to the limitations of the research, there may be other influencing mechanisms, and the influence of CEO narcissism on enterprise innovation is not fully investigated. To better clarify the impact of CEO narcissism on corporate innovation behavior, more boundaries, factors and paths can be considered to enrich the existing theoretical results. Since CEO narcissism is a complex personality trait, it is difficult to obtain the relevant data. The measurement data are mainly collected and collated manually, which may have some errors and need further improvement. This paper only explores the impact of CEO narcissism on enterprise innovation performance.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

LZ and BL: writing. DB: providing revised advice. YZ and XY: processing data. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Science and Technology Development Program of Jilin Province(China): Research on Problems and Countermeasures of Jilin Province's Economic Development Driven by Innovation of Science and Technology Service System; Doctoral Foundation of Jilin University of Finance and Economics: Research on the relationship between CEO power and firm growth of Gem listed companies in China.

REFERENCES

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *J. Pers. Soc. Psychol.* 49:1621. doi: 10.1037/0022-3514.49.6.1621

Baron, R, and Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic and statistical consideration. *J. Pers. Soc. Psychol.* 51, 1173–1182. doi: 10.1037/0022-3514.51.6.1173

Bernile, G., Bhagwat, V., and Rau, P. R. (2017). What doesn't kill you will only make you more risk-loving: early-life disasters and CEO behavior. *J. Financ.* 72, 167–206. doi: 10.1111/jofi.12432

Buyl, T., and Boone, C. (2017). CEO narcissism, risk-taking, and resilience: an empirical analysis in U.S. commercial banks. *J. Manage.* 45:7699521doi: 10.1177/0149206317699521

Campbell, W. K., Goodie, A. S., and Foster, J. D. (2004). Narcissism, confidence, and risk attitude. *J. Behav. Decis. Mak.* 17, 297–311. doi: 10.1002/bdm.475

Campbell, W. K., Hoffman, B. J., and Campbell, S. M. (2011). Narcissism in organizational contexts. *Hum. Resourc. Manage. Rev.* 21, 268–284. doi: 10.1016/j.hrmr.2010.10.007

Chatterjee, A., and Hambrick, D. C. (2007). It's all about me: narcissistic chief executive officers and their effects on company strategy and performance. *Adm. Sci. Q.* 52, 351–386. doi: 10.2189/asqu.52.3.351

Chatterjee, A., and Hambrick, D. C. (2011). Executive personality, capability cues, and risk taking: how narcissistic CEOs react to their successes and stumbles. *Adm. Sci. Q.* 56, 202–237. doi: 10.1177/0001839211427534

Chatterjee, A., and Poolock, T. (2017). Master of puppets: how narcissistic CEOs construct their professional worlds. *Acad. Manage. Rev.* 42, 703–725. doi: 10.5465/amr.2015.0224

Choi, S. B., Lee, S. H., and Williams, C. (2011). Ownership and firm innovation in a transition economy: evidence from China. *Res. Policy.* 40, 441–452. doi: 10.1016/j.respol.2011.01.004

David, P., O'Brien, JP., and Yosikawa, T. (2008). The implications of debt heterogeneity for RandD investment and firm performance. *Acad. Manage. J.* 51, 165–181. doi: 10.5465/amj.2008.30772877

Demerjian, P., Lev, B., and Mcvay, S. (2012). Quantifying managerial ability. *Manage. Sci.* 3, 1229–1248. doi: 10.1287/mnsc.1110.1487

Emmons, R. A., and Robert, A. (1987). Narcissism: theory and measurement. *J. Pers. Soc. Psychol.* 52, 11–17. doi: 10.1037/0022-3514.52.1.11

Freud, S. (1915). "On narcissism: an introduction," in *The Standard Edition of the Complete Works of Sigmund Freud* (Vol. XIV). London: Hogarth Press.

Galasso, A., and Simcoe, T.S. (2011). CEO overconfidence and innovation. *Manage. Sci.* 57, 1469–1484. doi: 10.1287/mnsc.1110.1374

Gerstner, W. C., Koenig, A., Enders, A., and Hambrick, D. C. (2013). CEO narcissism, audience engagement, and organizational adoption of technological discontinuities. *Adm. Sci. Q.* 58, 257–291. doi: 10.1177/0001839213488773

Gervais, S., Heaton, J.B., and O dean, T. (2011). Overconfidence, compensation contracts, and capital budgeting. *J. Financ.* 66, 1735–1777. doi: 10.1111/j.1540-6261.2011.01686.x

Hackbarth, D. (2008). Managerial traits and capital structure decisions. *J. Financ. Quant. Anal.* 43, 843–882. doi: 10.1017/S002210900001437X

Ham, C., Lang, M., Seybert, N., and Wang, S. (2017). CFO narcissism and financial reporting quality. *J. Account. Res.* 55, 1089–1135. doi: 10.1111/1475-679X.12176

Ham, C., Seybert, N., and Wang, S. (2018). Narcissism is a bad sign: CEO signature size, investment, and performance. *Rev. Account. Stud.* 23, 1–31. doi: 10.1007/s11142-017-9427-x

Hambrick, D.C. (2007). Upper echelons theory: an update. *Acad. Manage. Rev.* 32, 334–343. doi: 10.5465/amr.2007.24345254

Hambrick, D.C., and Mason, P.A. (1984). Upper echelons: the organization as a reflection of its top managers. *Acad. Manage. Rev.* 9, 193–207. doi: 10.5465/amr.1984.4277628

Heaton, J. B. (2002). Managerial optimism and corporate finance. *Financ. Manage.* 31, 33–45. 3666221doi: 10.2307/3666221

Hirshleifer, D., Low, A., and Teoh, S. H. (2012). Are overconfident CEOs better innovators? *J. Financ.* 67, 1457–1498. doi: 10.1111/j.1540-6261.2012.01753.x

Kaplan, S., Klebanov, M., and Sorensen, M. (2012). Which CEO characteristics and abilities matter? *NBER Work. Pap.* 67, 973–1003. doi: 10.1111/j.1540-6261.2012.01739.x

Kim, H., and Lee, K. (2008). Corporate governance ownership structure and the relationship between financial slack and RandD investments: evidence from Korean firms. *Organiz. Sci.* 19, 404–418. doi: 10.1287/orsc.1080.0360

Kumar, V, D. (2019). Multiple faces of narcissistic leadership in medical education. *J. Adv. Med. Educ. Prof.* 7, 103–105. doi: 10.30476/JAMP.2019.44705

Larwood, L., and Whittaker, W. (1977). Managerial myopia: self-serving biases in organizational planning. *J. Appl. Psychol.* 62, 194–198. doi: 10.1037/0021-9010.62.2.194

Lee, J. M., Hwang, B. H., and Chen, H. (2016). Are founder CEOs more overconfident than professional CEOs? evidence from SandP 1500 companies. *Strateg. Manage. J.* 38, 751–769. doi: 10.1002/smj.2519

Lu, Z., Zhu, J., and Zhang, W. (2012). Bank discrimination, holding bank ownership, and economic consequences: evidence from china. *J. Bank. Financ.* 36, 341–354. doi: 10.1016/j.jbankfin.2011.07.012

Malmendier, U., and Tate, G. (2005). CEO overconfidence and corporate investment. *Finance* 60, 2661–2700. doi: 10.1111/j.1540-6261.2005.00813.x

Miller, D., and Ross, M. (1975). Self-serving biases in the attribution of causality: fact or fiction? *Psychol. Bull.* 82, 213–225. doi: 10.1037/h0076486

Olsen, K. J, and Stekelberg, J. (2016). CEO narcissism and corporate tax sheltering. *The J. Am. Luxat. Assoc.* 38, 1–22. doi: 10.2308/atax-51251

Olsen, K. J., Dworkis, K. K., and Young, S. M. (2014). CEO narcissism and accounting: a picture of profits. *J. Manage. Account. Res.* 26:2. doi: 10.2308/jmar-50638

Rosenthal, S. A., and Pittinsky, T. L. (2006). Narcissistic leadership. *Leader. Q.* 17, 617–633. doi: 10.1016/j.leaqua.2006.10.005

Shefer, D., and Frenkel, A. (2005). RandD, firm size and innovation: an empirical analysis. *Technovation* 25, 25–32. doi: 10.1016/S0166-4972(03)00152-4

Shefrin, H., and Statman, M. (1999). Making sense of beta, size and book-to-market: investor expectations are consistent–but wrong. *J. Portfolio Manage.* 21, 26–34. doi: 10.3905/jpm.1995.409506

Steen, E. (2005). Organizational beliefs and managerial vision. *Econ. Organ.* 21, 256–283. doi: 10.1093/jleo/ewi011

Sunder, J., Sunder, S. V., and Zhang, J. (2017). Pilot CEOs and corporate innovation. *J. Financ. Econ.* 123, 209–224. doi: 10.1016/j.jfineco.2016.11.002

Thakor, A. V., and Goel, A. M. (2008). Overconfidence, CEO selection, and corporate governance. *J. Financ.* 63, 2737–2784. doi: 10.1111/j.1540-6261.2008.01412.x

Wallace, H., and Baumeister, R. (2002). The performance of narcissists rises and falls with perceived opportunity for glory. *J. Pers. Soc. Psychol.* 82, 819–834. doi: 10.1037/0022-3514.82.5.819

Wang, C., and Kafouros, M I. (2009). What factors determine innovation performance in emerging economies? *Int. Bus. Rev.* 18, 606–616. doi: 10.1016/j.ibusrev.2009.07.009

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhang, Liang, Bi, Zhou and Yu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 20 September 2021
doi: 10.3389/fpsyg.2021.725168

Dynamic Evolution Mechanism of Digital Entrepreneurship Ecosystem Based on Text Sentiment Computing Analysis

Jiahui Li and Meifang Yao**

School of Management, Jilin University, Changchun, China

To solve the limitations of the current entrepreneurial ecosystem, the research on the digital entrepreneurial ecosystem is more meaningful. This article aims to study the dynamic evolution mechanism of the digital entrepreneurship ecosystem based on text sentiment computing analysis. It proposes an improved Bi-directional long short-term memory (Bi-LSTM) model, which uses a multilayer neural network to deal with classification problems. It has a higher accuracy rate, recall rate, and F1 value than the traditional LSTM model and can better perform sentiment analysis on text. The algorithm uses the optimized Naive Bayes algorithm, which is based on Euclidean distance weighting and can assign different weights to the final classification results according to different attributes. Compared with the general Bayes algorithm, it improves the calculation efficiency and can better match the digital entrepreneurial ecosystem, which is evolving dynamically, predicting and analyzing its future development. The experimental results in this article show that the improved Bi-LSTM is better than the traditional Bi-LSTM model in terms of accuracy and F1 value. The accuracy rate is increased by 1.1%, the F1 value is increased by 0.6%, and the recall rate is only <0.2%. Running on the Spark platform, although 3% accuracy is sacrificed, the running time is increased by 320%. Compared with the traditional cellular neural network (CNN) algorithm, the accuracy rate is increased by 4%, the recall rate is increased by 14%, and the F1 value is increased by 9%, which proves that it has a strong non-linear fitting ability. The performance improvement brought by the huge data set is very huge, which fully proves the feasibility of the digital entrepreneurship ecosystem.

Keywords: text sentiment analysis, deep learning, digital entrepreneurship ecosystem, dynamic evolution, computing analysis

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Li Ming,
Jiangxi University of Finance and Economics, China
Sang-Bing Tsai,
University of Electronic Science and Technology of China, China

****Correspondence:***
Jiahui Li
jiahuil18@mails.jlu.edu.cn
Meifang Yao
yaomf@jlu.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *15 June 2021*
Accepted: *28 July 2021*
Published: *20 September 2021*

Citation:
Li J and Yao M (2021) Dynamic Evolution Mechanism of Digital Entrepreneurship Ecosystem Based on Text Sentiment Computing Analysis. Front. Psychol. 12:725168. doi: 10.3389/fpsyg.2021.725168

INTRODUCTION

Since 2015, supply-side reforms have come into being, emphasizing that innovation should be the first engine driving economic development, stimulating innovation potential and optimizing industrial structure. Entrepreneurship activities are very important for improving regional economic development, talent gathering, and promoting the sustainable development of industries. However, the technological content of entrepreneurship of Chinais still low, and the strong scientific and technological strengths have not been absorbed by entrepreneurial enterprises. Under different technological systems and market environments, issues such as the allocation of

scientific and technological resources and the efficiency of resource utilization have a significant impact on the sustainable development of enterprises. The entrepreneurial practice of Chinese enterprises shows that the "national system" type of entrepreneurial support has not made the national entrepreneurial system out of the "island phenomenon" in technological innovation. The gestation and growth processes of entrepreneurship are nested in a series of complex systems that influence each other. In these systems, each subsystem interacts with other systems and individuals, affecting all aspects of entrepreneurial activities. In recent years, the construction of an entrepreneurial ecosystem has become an inevitable way to cultivate unique regional competitive advantages. The self-organization and self-regulation capabilities of the entrepreneurial ecosystem play an important role in the sustainable development of their internal enterprises. Due to the lack of a scientific evaluation mechanism, the construction of entrepreneurial ecosystem of China is still in the exploratory stage, and the promotion of typical cases established by following the trend can easily lead to the waste of resources, such as repeated construction. "No Measurement, No Improvement," how to use a reasonable evaluation system to provide targeted and operable suggestions for the optimization of the entrepreneurial ecosystem is worthy of further research. As a diversified, mutually beneficial, and win-win innovation and entrepreneurship system, the business ecosystem is of great significance to the formulation of innovation and entrepreneurship policies by analyzing the impact of different factors on its operation.

The entrepreneurial ecosystem is composed of three aspects: entrepreneurial enterprises, stakeholders, and the entrepreneurial environment. The entrepreneurial entities are connected by formal or informal relationships. Different communities influence and depend on each other, and evolve together over time, and they have obvious regional characteristics. On the one hand, the existence of the entrepreneurial ecosystem can save excessive transaction costs and organization costs. On the other hand, it can provide a more advantageous development environment for new start-ups in an unstable market environment, and improve the competitiveness of start-ups, which is a guarantee keyway to the quality and vitality of new ventures. However, the effective operation of the entrepreneurial ecosystem is inseparable from the coevolution and mutual synergy between the constituent communities and elements. Therefore, through efficiency evaluation and analysis of influencing factors, the various elements within the entrepreneurial ecosystem, the linkages between the elements, and the driving force of the linkages are discussed. It has an important guiding role in the formation of a real-time, creative, and orderly entrepreneurial ecosystem.

At present, the Internet is developing rapidly, and many people are paying attention to the digital entrepreneurial ecosystem. The following people have unique insights on this. Singh SK proposed a new framework for a spell-check system that extracts user reactions, emotions, and opinions from social media text (SMT). User opinions are extracted from their written texts on social media and based on SMT sentiment scores, using dictionary-based methods and binary classifiers, which are classified as positive or negative opinions. The dictionary-based method uses the opinion verb dictionary (OVD) to extract the sentiment of opinion verbs that appear in SMT. This OVD contains only opinion verbs and their sentiment scores. The various steps of the framework have been discussed, such as lowercase conversion, tokenization, spell-checking, part-of-speech tagging, stop word elimination, stemming, sentiment score calculation, and SMT classification. Although Singh introduced the new concept of threshold negative parameters to facilitate the extraction of text information data, the efficiency of this extraction method is slightly low. If the optimized Bayes algorithm is used, the efficiency can be greatly improved (Singh and Sachan, 2019). Le X proposed that the rapid popularization of digital technology with new functions has profoundly changed the competitive environment and reshaped traditional business strategies and processes. Such technologies have also spawned new ways of collaboration, using resources, service design, development, and deployment based on open standards and shared technologies. At the micro-level, digital technology has also reshaped the mindset of entrepreneurs, thereby affecting their decision-making process. Digital entrepreneurship includes venture capital and the transformation of existing enterprises through the creation of new digital technologies and new uses of these technologies. At present, many countries regard digital entrepreneurship as an important pillar for the development of the digital economy, and it is necessary to have a detailed understanding of digital entrepreneurship. Although Le has formulated the relevant content of digital entrepreneurship, he has not made an in-depth dynamic evolution of the digital entrepreneurship system, which makes it impossible for him to estimate the future development of digital entrepreneurship (Le et al., 2019). Allenby CE discovered that digital health entrepreneurship is a topic widely discussed in non-professional and professional media. Innovators in this field continue to raise large amounts of funds to develop applications and devices that will change the way health care is managed in the United States. As many as one-third of American consumers use wearable devices and mobile health applications, which provide important opportunities for putting health care in the hands of consumers. However, it remains to be seen who will play the greatest role in this field. Technology giants, biopharmaceutical companies, and start-ups all have the opportunity to take advantage of new areas and change the way health care interacts with patients. Although entrepreneurs are in a leading position in the technology industry, they lack the regulatory and healthcare expertise of large pharmaceutical companies. However, as start-ups and large technology companies continue to promote innovation in this field, the pharmaceutical industry will not be able to continue to pack lightly. Companies that decide to take stock in this market will have to decide on how to interact with the health care sector. He discovered that digital entrepreneurship has great potential for development. He focused his attention on the development of medical and health care, but this development relied on large local technology companies and biopharmaceutical companies. He did not take advantage of the convenience of text sentiment calculation and analysis. If based on text sentiment calculation

analysis to dynamically evolve digital entrepreneurship, then he can develop on his own (Allenby et al., 2018).

The innovations of this article are the following: (1) Using text sentiment calculation and analysis methods, combined with deep learning algorithms, the traditional Bayesian algorithm is improved and optimized, and the efficiency is improved. (2) Putting forward the theory of the digital entrepreneurship ecosystem, discovering the huge potential for future development, and analyzing the important components of the system and its influencing factors. (3) Comparing the algorithm with other models, a better Bi-LSTM model is obtained. It is found that the running time under the Spark platform is much shorter than the running time on the native model.

The first section of this article introduces the background and significance of the current domestic and foreign entrepreneurial ecosystems, and cites several references that have important results in emotional computing, and also describes the innovations of this article. The second section briefly introduces the creation of ecological theory, core competitiveness theory, text sentiment calculation analysis, naive Bayes algorithm, and entrepreneurial process theory. In the introduction of naive Bayes algorithm, it proposes its improvement and optimization. In the later Bayesian algorithm theory. The third section introduces the experimental purpose, experimental procedures, and experimental methods of this article. The fourth section focuses on the analysis of the influencing factors of sentiment computing in the dynamic evolution of the digital entrepreneurship ecosystem, compares the performance of various traditional algorithms, and then analyzes and introduces the optimized and improved text sentiment computing algorithms in this article. The fifth section is a summary of this article and puts forward the shortcomings in the article and the outlook for the future.

DYNAMIC EVOLUTION MECHANISM OF DIGITAL ENTREPRENEURSHIP ECOSYSTEM BASED ON TEXT SENTIMENT CALCULATION ANALYSIS

Entrepreneurship Ecosystem

The concept of an entrepreneurial ecosystem originates from the combination of entrepreneurial theory and ecological theory. To be precise, it is a reexamination of entrepreneurial theory from an ecological perspective. With the advancement of science and technology and the shortening of product life cycles, the knowledge required for product innovation is distributed in fragments, and the process of technological innovation shows more non-linear characteristics. The effect of research and development (R&D) investment on R&D performance also depends on the collaboration of various entrepreneurial entities. From the beginning of the establishment to the final demise, start-ups have always been in dynamic communication with the environment in which they are located (the inflow of key elements and product output). The competition and cooperation between different entrepreneurs are complicated, but they have always maintained the entire entrepreneurial system. These characteristics make the entrepreneurial process naturally have an analysis function of ecological significance (Shen et al., 2018). Moore (1996) first proposed the concept of the business ecosystem and analyzed the relationship between entrepreneurial enterprises and their related multiple entrepreneurial entities from an ecological perspective, setting a precedent for entrepreneurial ecological research. After that, many scholars have conducted more in-depth research on entrepreneurial ecology (Gabison, 2019).

The ecosystem needs the inflow and outflow of material and energy to maintain the normal operation of the system. For business ecology, this means the need to study the creation and development of new businesses in various fields related to the business environment and business resources. Gnyawali and Fogel (1994) divided the entrepreneurial environment into five dimensions: financial support, non-financial environment, policy regulations, socioeconomic support, and entrepreneurship and management skills, and the impact of these factors on entrepreneurial performance in the entrepreneurial ecosystem perform analysis (Chen, 2021). Li et al. (2011) has very in-depth research in the field of entrepreneurship. He has conducted multidimensional analysis on the relationship between entrepreneurial networks and entrepreneurial performance. In the entrepreneurial environment, entrepreneurial resources have a positive impact on the growth and performance of new ventures. Verify shows (Hutson, 2018). Only when organisms exchange material and energy with the environment can they ensure their development, and entrepreneurial companies can only ensure their sustainable development if they fully utilize and integrate resources and the environment. The entrepreneurial ecosystem has the characteristics of regionality, diversity, symbiosis, external spillover, self-organization, and self-regulation. Many literature studies have conducted relevant research on its regionality and diversity, but the research on its self-organization and self-regulation is still relatively scarce. At present, with the further development of the entrepreneurial ecological theory, the research on the entrepreneurial ecosystem has gradually shifted from its conceptual framework to the analysis of its dynamic evolution. The method of metaphor alone cannot fully play the role of ecology in the entrepreneurial field. Think about the integration of the two from a deeper level (Nuseir, 2018).

Core Competitiveness Theory

The core competitiveness theory focuses on the ability to integrate related technologies and knowledge, and at the same time points out that the core competence is transferable between organizations and can be transferred to all members of the enterprise in accordance with the organizational structure (Ishmukhametov et al., 2020). Domestic experts and scholars have carried out rich and in-depth research on this theory. The earliest domestic research on core competitiveness was done by Professor Jin Bei of the Chinese Academy of Social Sciences, which can make enterprises stand out from the competition (Miao and Chonghui, 2018). The core competitiveness has the following characteristics: user value, user evaluation, and

feedback are the most objective and direct standards to measure the competitiveness of an enterprise. The core competitiveness can be improved only by continuously improving the quality of products and services and by increasing user satisfaction. Occupancy, the company has a long-term occupation of certain necessary development resources, which makes itself have a unique advantage different from other competitors. This resource is difficult to obtain by other companies and will be occupied by the company for a long time; malleability, the core competitiveness of the company can be extended from ideas or technologies to products and services and maintained market advantages. Uniqueness, core competence can give enterprises an advantage in a series of products or services. There are differences in the core competitiveness of different enterprises. This kind of competitiveness is gradually formed in the long-term development of enterprises and has a high degree of imitation (Liu et al., 2018).

Text Sentiment Calculation Analysis

As shown in **Figure 1**, the text sentiment calculation analysis model is built on the LSTM artificial neural network. The word vector is extracted from the top text source, and the input layer is the input word vector for sentiment calculation analysis, like the hidden layer input data, training, and filtering. After that, select the data that suit the requirements for output (Frank, 2018). In recent years, sentiment analysis has flourished in computer science, and a large number of articles on sentiment analysis of text have also appeared in major well-known natural language processing-related conferences. According to the granularity of the text, sentiment analysis can be divided into word-level, sentence-level, and text-level. The first two are fine-grained, and the latter are coarse-grained (Zeng et al., 2021). Because an article often has multiple sentences expressing different emotions, and words are the smallest unit of expressing sentence emotions, which are likely to be affected by other vocabulary or sentence patterns, most research are more inclined to use sentences as the granularity. Taking into account complicated situations such as the omission of evaluation objects in emotional expression, experiments have proved that this work can provide a certain basis for emotional analysis based on supervised learning. In terms of sentiment classification, methods based on rules and machine learning are more common. Among unsupervised classification methods, sentiment dictionaries and their construction work also play an important role. A sentiment dictionary is constructed by integrating three parts of psychology domain vocabulary, WordNet ontology dictionary, and commonly used slang words on the Internet. Then, based on this dictionary, it analyzes the incentives and laws of emotions of people in three different cities over time (Chang and Huo, 2018). It can also combine information increment, support vector machine, decision tree, and naive Bayes algorithm for feature selection and sentiment classification, and use positive and negative feature mean to evaluate sentiment similarity; construct a method based on graph learning and semi-supervision, a classifier of Weibo sentiment polarity is constructed. This classifier takes into account the community relationship of Weibo users and the similarity of Weibo text. Experiments have shown that this classifier can replace part of the manual annotation work (Hamdi et al., 2018). After finishing the work of removing stop words and feature extraction on Twitter text, the improved positive feedback neural network method is used to classify the sentiment tendency of the text. Although satisfactory results are finally achieved, the algorithm is trained when the text is too large. The usefulness of the memory needs to be optimized; it can be found that long text and short text have a high degree of overlap in terms. Therefore, based on the characteristics of short text and traditional feature filtering methods, the long text features and short text common features are left, and the superior information is added to the short text classification process (Nam and Lee, 2019). An improved conditional random field model is also proposed, which treats the tendency analysis of phrase sentiment as sequence labeling (Hahn et al., 2020). A method for calculating the similarity of comments posted by users based on cosine similarity is proposed, and each text is assigned a value of emotional orientation strength from -2 (negative) to $+2$ (positive). The result is similar to the traditional shell, where maximum entropy and support vector machines are all improved (Li et al., 2020); linguistic knowledge is integrated into the traditional sentiment classification, a large number of rules are expanded, and the improved N-grams model is used to solve the language morphology where there are more complicated sentence patterns in characteristics (Kotelnikov, 2018). However, the purely rule-based method requires domain experts to define a large number of rules, which is costly and has limitations. In recent research results, concept-level sentiment analysis is often be mentioned; a concept-level sentiment analysis system is constructed, which integrates semantic analysis, subjectivity detection, irony syntax detection, topic positioning, and other multidimensional, experimental. It is proved that this method can solve the performance defects of some supervised sentiment analysis algorithms in different situations (Singh and Sachan, 2021); through the use of the conceptual hierarchical model to extract text emotional features, extract the semantics of the text, and finally explore some microblog text and natural semantic relevance of language text; it proves that the performance of sentiment analysis at the concept level is better than the performance of sentiment analysis at the lexical level. In addition, in the sentiment analysis corpus, the corpus that does not contain sentimental tendencies is also of practical value. There is a map-reduce-based e-commerce user opinion recognition framework, and new dictionary-based technology is used to mine the products. Neutral evaluation to correct the results of these opinions being classified as positive or negative, to reduce the error of sentiment analysis (Bhagat and Mane, 2019; Schmidt et al., 2020).

Naive Bayes Algorithm

Conditional probability refers to the occurrence of event B as the premise. In this case, the probability of occurrence of event A is recorded as P(A|B).

Let A and B be two events in the sample space Ω, if P(B) >0, then it is called as shown in Equation 1:

$$P(A|B) = \frac{P(AB)}{P(B)} \tag{1}$$

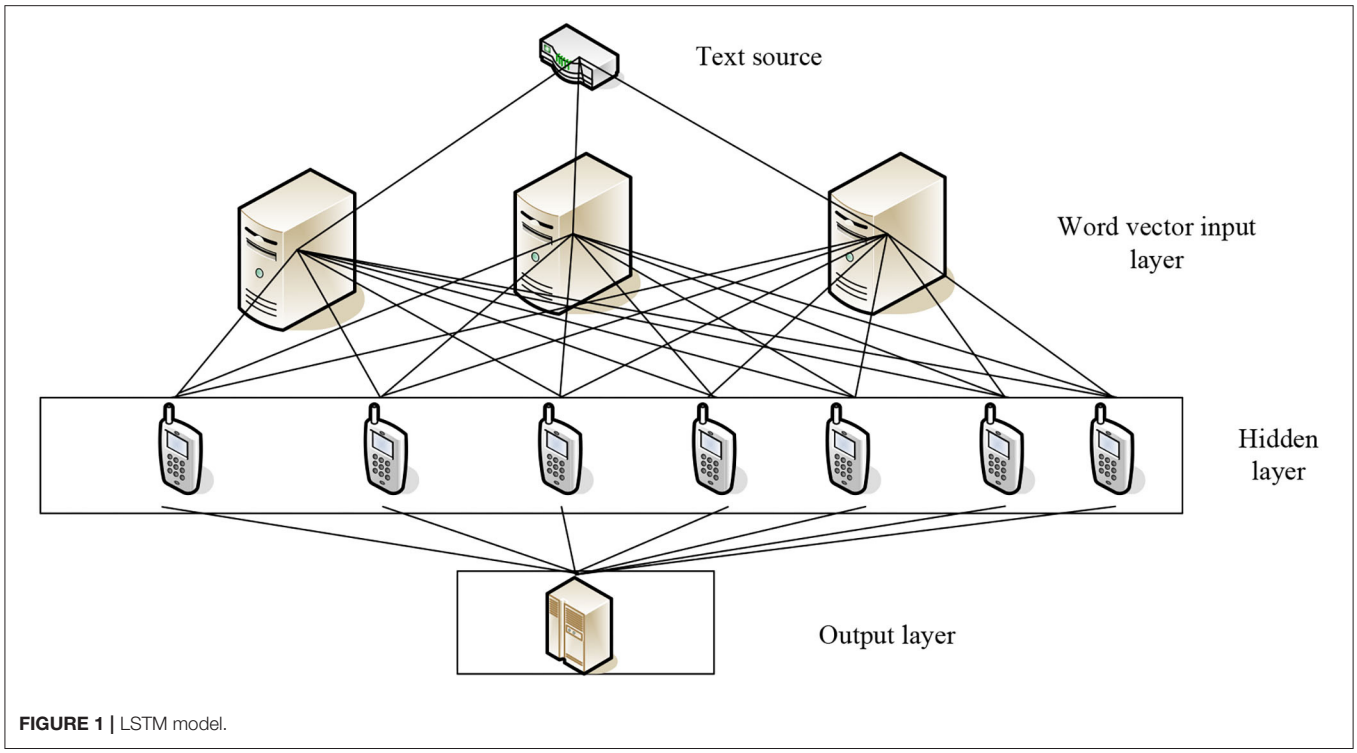


FIGURE 1 | LSTM model.

Event B occurs first. On this basis, the probability that event A will occur. Among them, the probability P(B) of event B is called the prior probability, P(A|B) is the probability that event A will occur based on the occurrence of event B, this probability is called the posterior probability of event A, P (AB) represents the probability that both events A and B will occur. Let A and B be events in the sample space Ω, if P(B) > 0, then equation 2 is described as follows:

$$P(AB) = P(B)P(A|B) \tag{2}$$

If $P(A_1A_2A_3 \ldots A_{n-1}) > 0$, then

$$P(A_1A_2A_3...A_n) = P(A_1)P(A_2|A_1)...P(A_n|A_1A_n) \tag{3}$$

Assuming that $B_1, B_2, \ldots B_n$ is a self of the sample space Ω, then $B_1, B_2, \ldots B_n$ will be mutually exclusive, and $\bigcup_{j=1}^{n} B_j = \Omega$, if P(B) > 0, then any event A has

$$\mathrm{P(A)} = \sum_{\mathrm{j=1}}^{n} P(B_j)P(A|B_j) \tag{4}$$

From the above, we can integrate the two formulas into one formula, which is the formula of Bayes' theorem:

$$P(B_j|A) = \frac{P(AB_j)}{P(A)} = \frac{P(B_j)P(A|B_j)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)} \tag{5}$$

The most likely hypothesis $F_j \in F$ for a given sample C is selected from a set of candidate hypotheses f. This hypothesis or category is called the maximal posterior hypothesis, namely

$$f_{MAP} = \arg\max_{f_j \in F} P(f|C) \tag{6}$$

$$= \arg\max_{f_j \in F} \frac{P(C|f)P(f)}{P(C)} \tag{7}$$

$$= \arg\max_{f_j \in F} P(C|f)P(f) \tag{8}$$

Under certain circumstances, assuming that the prior probability of each hypothesis in F is the same, just consider P(C|f) to find the most probable hypothesis, so P(C|f) is called when f The likelihood of data C, where the hypothesis that maximizes P(C|f) is called the maximum likelihood hypothesis (Lutfullaeva et al., 2018):

$$f_{ML} = \arg\max_{f_j \in F} P(C|f) \tag{9}$$

It can be seen from the Bayesian formula that the posterior probability P(f|C) depends on the product of P(C|f) P(f), which is also the most important idea of the Bayesian classification algorithm. According to Bayes' theorem:

$$P(d_i|y_1, y_2, ...y_n) = \frac{P(y_1, y_2, ...y_n|d_i)P(d_i)}{P(y_1, y_2, ...y_n)} \tag{10}$$

Among them, P_i is the probability that the text belongs to the category d_i in the training text. In the partially selected

categories, Bayesian classification assumes that the attributes in the classification are independent of each other, then Equation 11 is given as follows:

$$P(y_1, y_2, ...y_n|d_i) = \prod_{k=1}^{n} P(y_k|d_i) \quad (11)$$

Substituting the prior probability $P(y_1, d_i), P(y_2, d_i), ...P(y_n, d_i)$, calculated from the training set into the above formula, we can get the following Equation 12:

$$P(d_i|Y) = \frac{\prod_{K=1}^{n} P(y_k|d_i)P(d_i)}{P(y_1, y_2, ...y_n)} \quad (12)$$

The denominator in the formula is a constant, and the Naive Bayes model formula is a maximum value of the calculation formula:

$$d_{NB}(Y) = \arg\max_{f_j \in F} P(d_i) \prod_{K=1}^{n} P(y_k|d_i) \quad (13)$$

In the above formula, because it is necessary to calculate a large number of product operations, which brings difficulties to the calculation, the formula is modified as follows:

$$d_{NB}(Y) = \arg\max_{f_j \in F} [\log P(d_i) + \sum_{K=1}^{n} P(y_k|d_i)] \quad (14)$$

The naive Bayes classification algorithm simplifies the estimation of the prior probability when the assumption of independence of distribution is established, without considering the correlation between different attributes. The process of predicting the sample to be predicted is simple, the classification complexity is low, the speed is fast, and the cost in the calculation process is relatively low.

$$P(D_j|Y) = P(D_j) \prod P^{w_k}(y_k|D_j) \quad (15)$$

Among them, the weight of the attribute A is w_k, that is, the weight of the $P(Y|D_j)$. The larger w_k, the greater the importance of the attribute to the classification result, which affects the classification result.

Entrepreneurship Process Theory

Many scholars have conducted in-depth research on entrepreneurial-related theories, such as entrepreneurial characteristics theory, entrepreneurial process theory, entrepreneurial cycle theory, entrepreneurial network theory, etc., discussing the factors of entrepreneurial success and failure from multiple aspects. The first to emerge is the entrepreneurial characteristics theory, which believes that the personal characteristics of an entrepreneur (enterprise willingness, risk tolerance, ability to formulate strategies, etc.) play a decisive role in the success of entrepreneurship. With the deepening of entrepreneurial research, scholars have

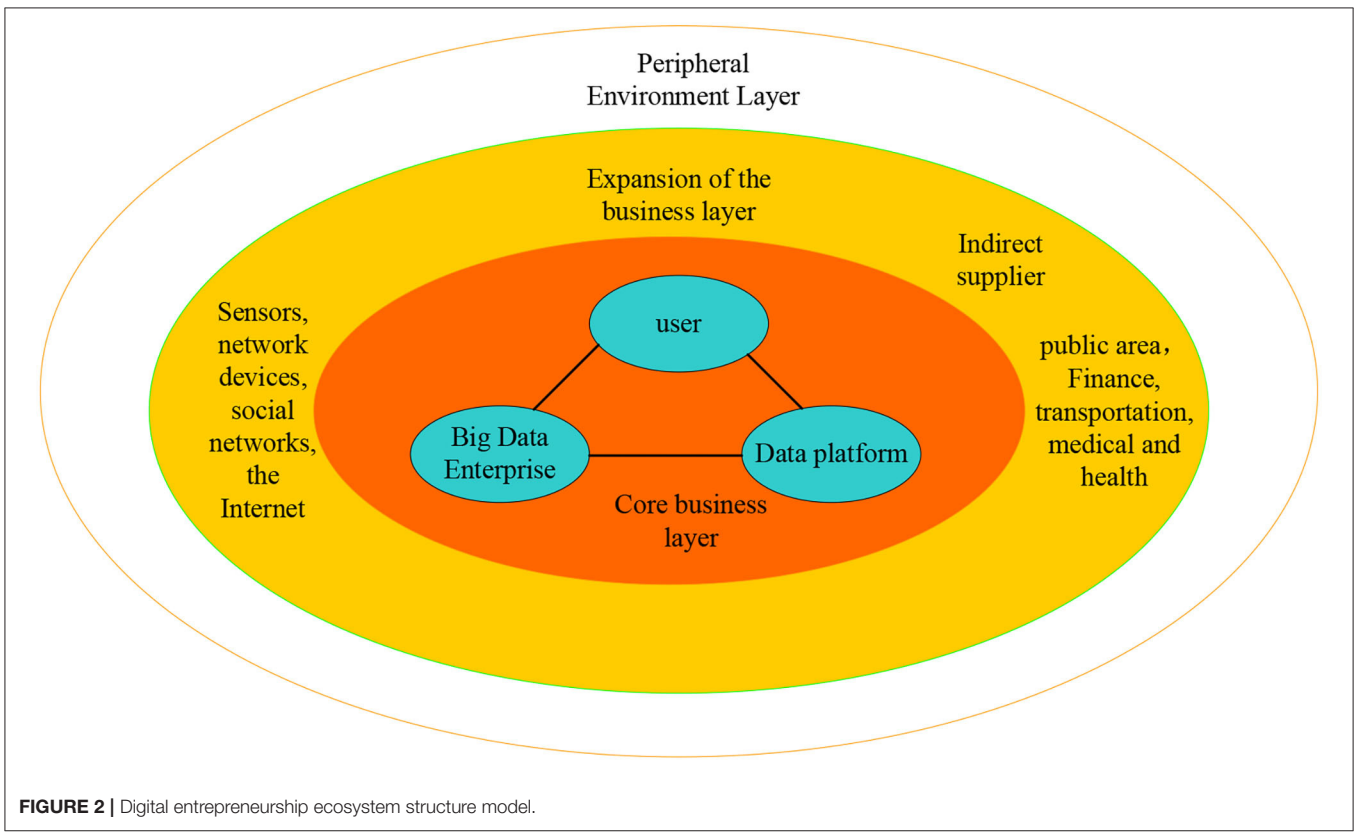


FIGURE 2 | Digital entrepreneurship ecosystem structure model.

begun to pay attention to the influence of more traits of non-entrepreneurs on entrepreneurial activities, and the research on the entrepreneurial process has gradually attracted the attention of relevant scholars.

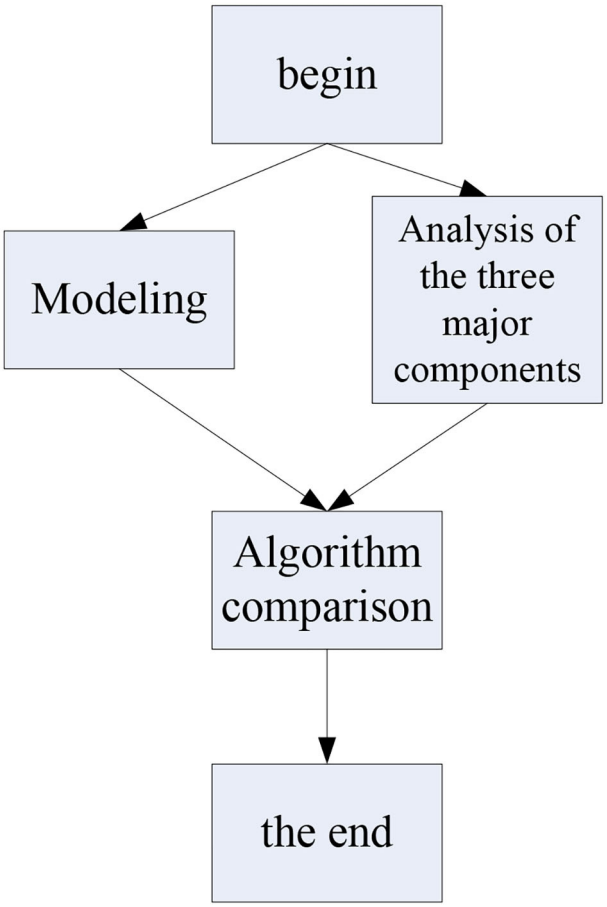


FIGURE 3 | Experimental flowchart.

Gartner (1985) first proposed that the success of entrepreneurship is not only related to the characteristics of entrepreneurs. The entrepreneurial process is a dynamic, complex, and diverse process. It is a synthesis of many factors in the entrepreneurial process. He compared these related to entrepreneurship. The elements are summarized into four aspects, namely entrepreneurs, stakeholders, entrepreneurial environment, and entrepreneurial practices. Morris et al. (1994) combined previous entrepreneurial research results, regarded the entrepreneurial process as a dynamic process of resource input-product output, and constructed an input-output integration model based on the entrepreneurial process. Xuebing and Gang (2009) believe that entrepreneurship is a process by which entrepreneurs transform new products and new patents into economic benefits by integrating entrepreneurial resources and identifying market opportunities in technological innovation. Entrepreneurship process theory is a very important theory in the field of entrepreneurship. It makes entrepreneurial activity no longer a "black box." People can analyze the different stages

TABLE 1 | Combination analysis of sufficient conditions for efficient operation of an entrepreneurial ecosystem.

Variable Name	M1	M2	M3	M4
Regional economic development	•	•	○	○
Technology market maturity	○	○	•	○
Government support	○	○	○	○
Industry-university-research cooperation degree	○	○	○	•
Entrepreneur traits		○		•
Entrepreneurship support organization scale	○		•	○
Raw coverage	0.423	0.415	0.431	0.122
Consistency	0.815	0.785	0.776	0.910

"•" means that the condition exists, "○" means that the condition does not exist, and blank means that the condition may or may not exist.

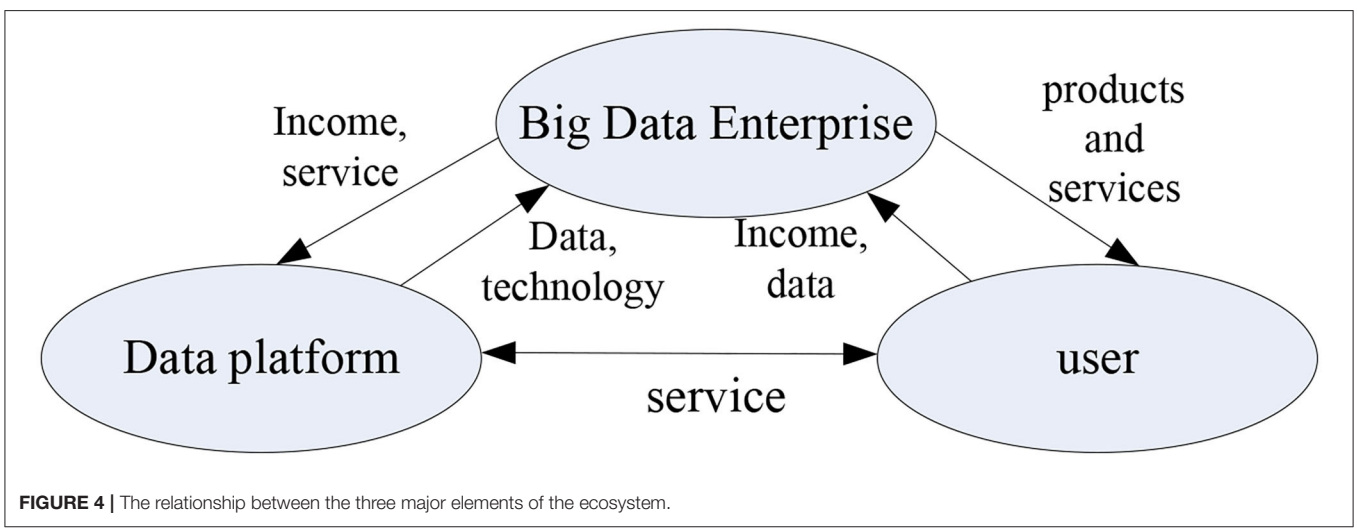


FIGURE 4 | The relationship between the three major elements of the ecosystem.

of the entrepreneurial process and explore the impact of different stages on entrepreneurial performance, making the part and the whole static Take into account the dynamics (Lian et al., 2020; Sodhar et al., 2020).

EXPERIMENT ON THE DYNAMIC EVOLUTION MECHANISM OF DIGITAL ENTREPRENEURSHIP ECOSYSTEM BASED ON TEXT SENTIMENT CALCULATION ANALYSIS

The key to understanding the digital entrepreneurship ecosystem is to clearly define the core role of the system, which is also the prerequisite for the construction of the system structure model. By combining and analyzing the existing literature, it is found that there is good consistency in the division of the roles of the digital entrepreneurship ecosystem in the existing research, which has laid a good foundation for the following experimental parts.

Subjects

The digital entrepreneurship ecosystem is a big data ecosystem based on text sentiment calculation analysis. The digital entrepreneurship ecosystem can be divided from a micro and macro perspective. From a micro perspective, the core roles in the big data ecosystem are divided into users and providers and data. From a macro perspective, the system can be divided into a core layer, an extension layer, and an environment layer. The microscopic digital entrepreneurship system is the main research object of this experiment.

TABLE 2 | Comparative data of different models.

Model	Accuracy	Recall Rate	F1 Value
CNN	60.2%	49.7%	54.3%
LSTM	57.4%	63.1%	59.6%
Bi-LSTM	63.2%	64.0%	63.4%
Improved Bi-LSTM	64.2%	63.7%	63.8%

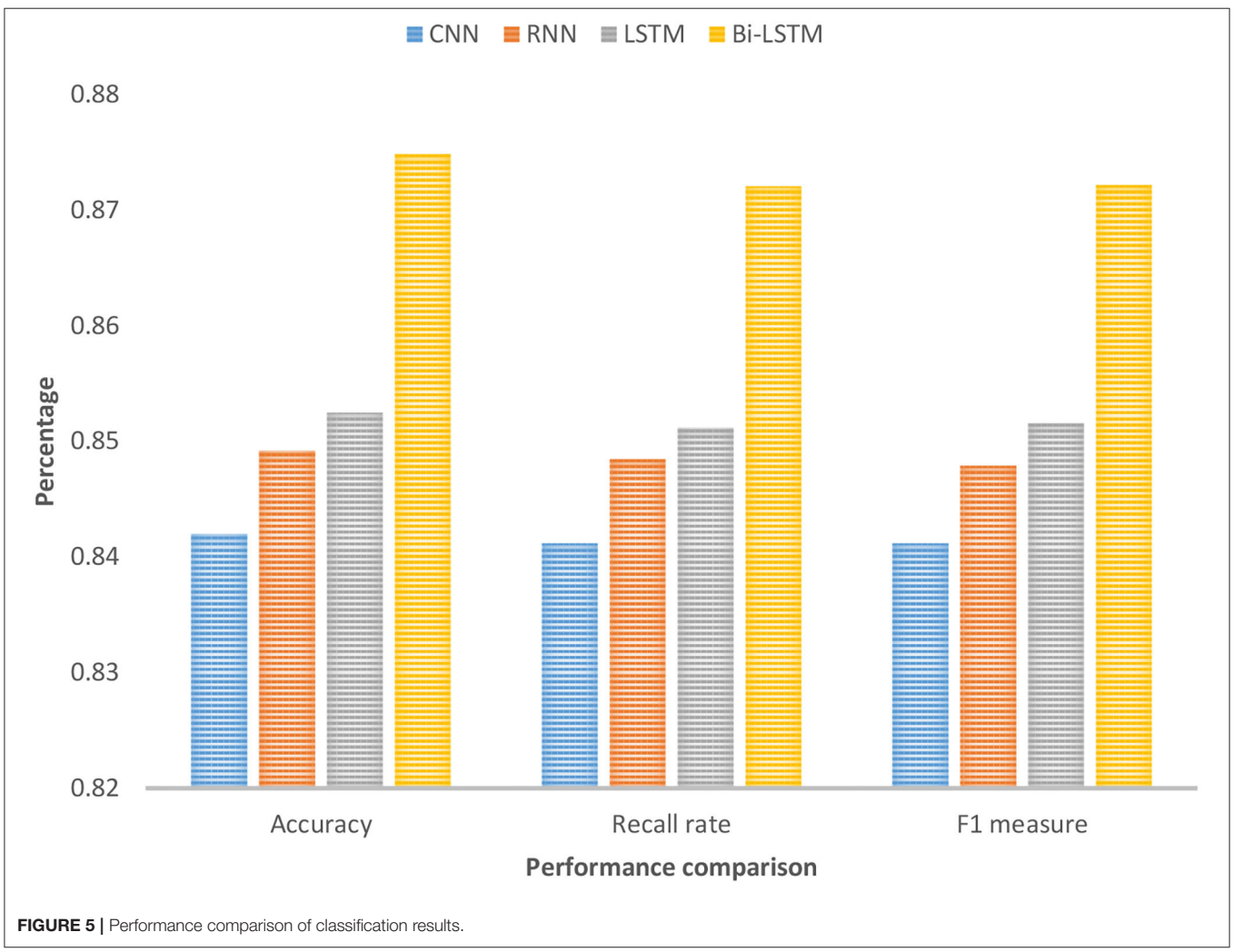


FIGURE 5 | Performance comparison of classification results.

As shown in **Figure 2**, the digital entrepreneurship ecosystem is divided into three levels, namely the core entrepreneurial layer, the extended entrepreneurial layer, and the peripheral environment layer. The core entrepreneurial layer is composed of big data companies, data platforms, and users. Big data companies can be divided into big data core companies and big data niche companies based on the entrepreneurial ecological perspective.

Purpose of the Experiment

The purpose of the experiment in this article is to analyze the elements of the digital entrepreneurship ecosystem model and then analyze the pros and cons of the algorithm by comparing text-based sentiment computing with other algorithms.

Experimental Procedure

As shown in **Figure 3**, the first step of this experiment is to build a digital entrepreneurship ecosystem model, then analyze the construction elements of the digital entrepreneurship model, analyze the key points of its elements and the difference between the manufacturing ecological models, and then compare the algorithms. The cross-experiment method will perform performance screening based on text sentiment computing and other artificial intelligence algorithms, and finally, the experiment ends.

DYNAMIC EVOLUTION MECHANISM OF DIGITAL ENTREPRENEURSHIP ECOSYSTEM BASED ON TEXT SENTIMENT COMPUTING ANALYSIS

Influencing Factors

The relationship between the three core elements of the digital entrepreneurship ecosystem is shown in **Figure 4**. The relationship between the elements can be understood from two levels. One is from the macro-level of system composition. The second is from the internal components of each level. Among them, big data companies are the main force driving the formation of the digital entrepreneurial ecosystem, which provides users with a wealth of data products and services; the data platform, as an exchange and circulation platform for data resources and a related technical service platform, continues to provide the development of the big data industry. "Raw materials" and technology and service support; users are consumers of data products and services, and the ultimate channel for product value to be realized. Their huge demand

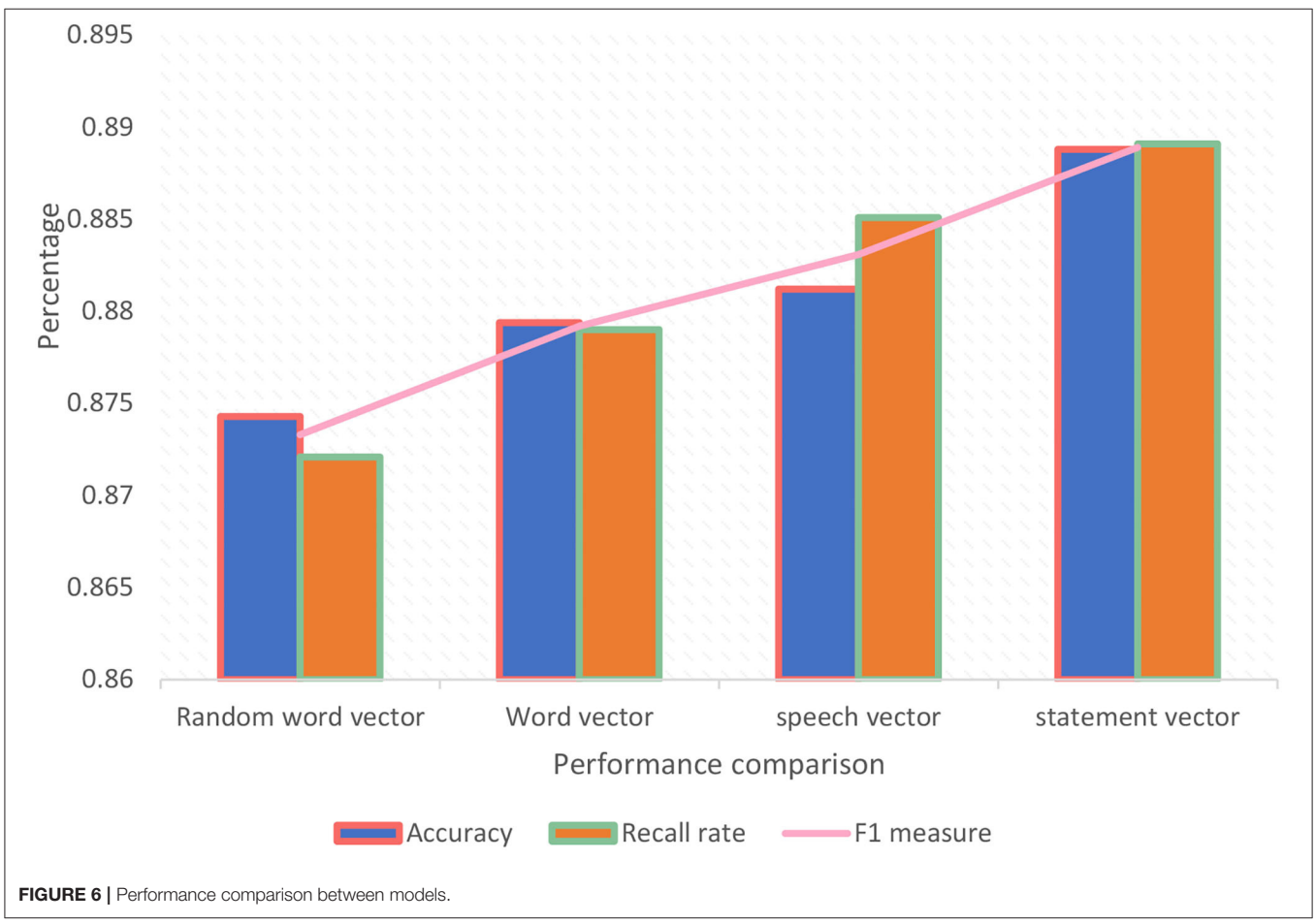


FIGURE 6 | Performance comparison between models.

for data products and services is the driving force behind the development of the big data industry.

As shown in **Table 1**, one is the efficient operation of the entrepreneurial ecosystem (M1, M2) brought about by good regional economic conditions, and the other is the operation of the entrepreneurial ecosystem due to the interaction between the maturity of the technology market and the entrepreneurial support institutions. The high efficiency obtained (M3) is the high efficiency brought about by the combined effect of the degree of cooperation between industry, university, and research institute and the characteristics of entrepreneurs (M4). It can be seen that the progress of entrepreneurial activities is closely related to the entrepreneurial environment in which they are located. A relaxed entrepreneurial atmosphere and an active market environment are the keys to the smooth progress of entrepreneurial activities.

Performance Comparison of Algorithm Classification Results

As shown in **Figure 5**, the overall effect of the Bi-LSTM algorithm is better than other comparative experimental methods, indicating that more effective semantic information can be obtained after considering the contextual information of the input data to improve the accuracy of the results. Bi-LSTM uses a randomly initialized vector to represent the input text data, and then uses a two-way long and short-term memory network structure for text classification. The network selected for the feature expression process is Bi-LSTM, and the number of hidden layer nodes is set to 256. The learning rate is set to 0.01, and the stochastic gradient descent algorithm is used to update and iterate the weight to obtain a stable model. Bi-LSTM is one level higher than the traditional artificial intelligence algorithm in terms of accuracy, recall, or F1 measurement.

As shown in **Table 2**, the improved Bi-LSTM is superior to other models in terms of accuracy and F1 value, and the recall rate is only lower than the highest 0.2%, which fully proves that this model is reasonable. In addition, it can be seen that the accuracy of the traditional artificial intelligence CNN algorithm is significantly higher than that of LSTM, but the recall rate is lower, which proves that strong non-linear fitting ability of LSTM guarantees its high recall rate.

As shown in **Figure 6**, it is found that the result of using word vector training is better than the training result of a random word vector. The reason is that in the training process, the context

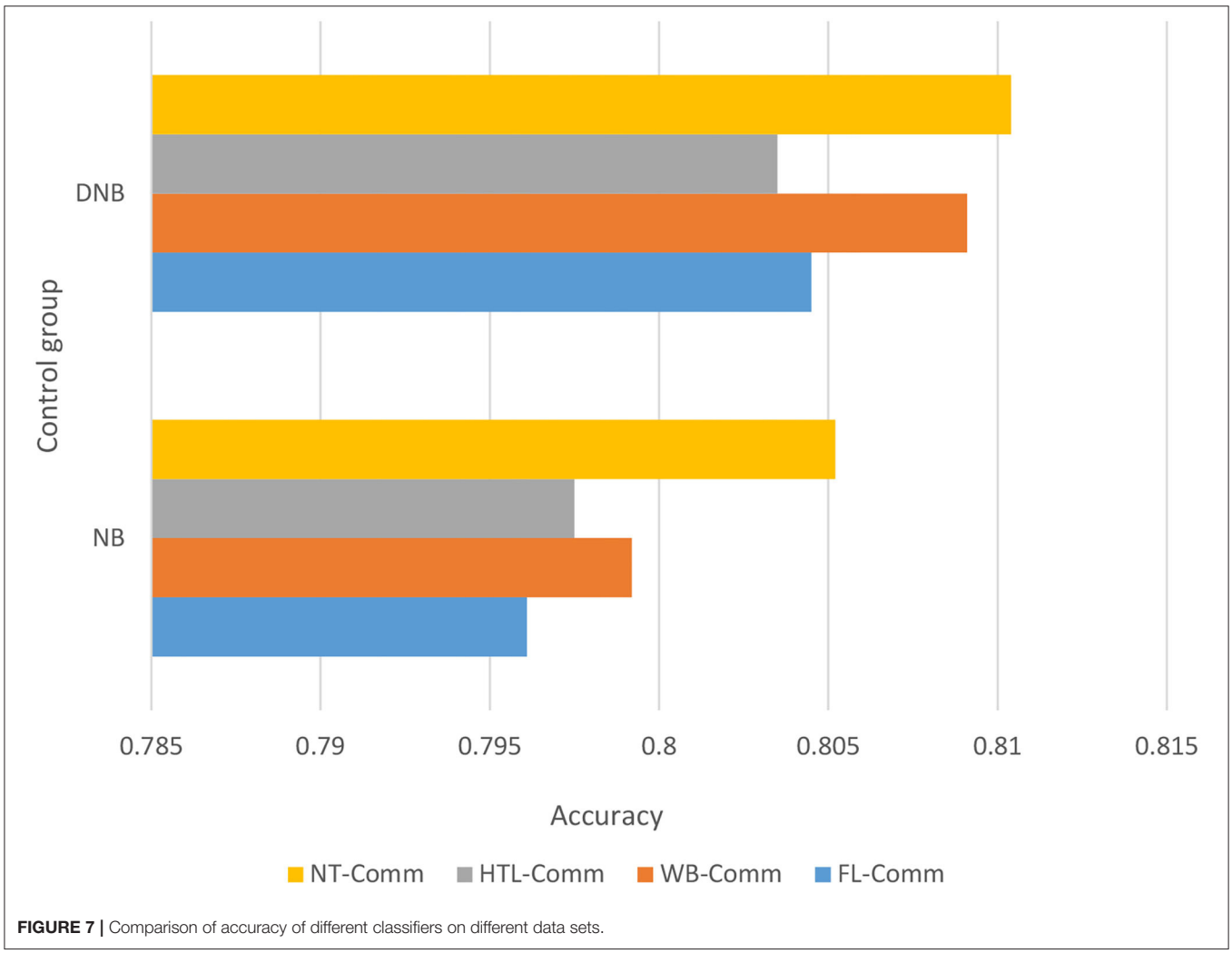


FIGURE 7 | Comparison of accuracy of different classifiers on different data sets.

information between the words can be considered to get a better representation, thereby improving the text sentiment analysis, the accuracy, and other effects of the comparison; the training result of the comparison word vector is better than the training result of the word vector. The reason is that in the Chinese language, compared to the word, the word is a more effective collection of semantic information. This training method can be more effective. It effectively represents the semantic information of the data, thereby obtaining more meaningful relevant features in the text sentence, and improving the effectiveness of text sentiment analysis; the result of word vector training is better than the result of word vector, because for Chinese, in the training process of the word vector, the words in the context are also introduced, so that in the training process, the context space of the word vector can contain the semantic information of the words in the data, which can make the semantic information of the context richer and obtain better word representation, thereby improving the effect of text sentiment analysis.

Algorithm Analysis Based on Improved Text Sentiment Calculation

In the text-based sentiment calculation analysis, the improved naive Bayes algorithm is used. In this article, the 10-fold crossover method is selected to evaluate the performance of the classifier, and the classification accuracy results are averaged to the final classification results. The higher the average, the better the accuracy.

As shown in **Figure 7**, the improved naive Bayes algorithm has a higher accuracy rate than the traditional Bayes algorithm. The improved Naive Bayes algorithm, based on Euclidean distance weighting, can give different weights to the final classification results according to different attributes to improve the classification performance of the algorithm, thereby improving the classification accuracy of the algorithm. Finally, it can be concluded that the improved naive Bayes algorithm is more effective. The optimized Bayesian algorithm uses the existing data information, and the prior probability obtained by weighting the attributes replaces the traditional naive Bayes prior probability, which improves the classification accuracy. Then, transplant the improved weighted naive Bayes classifier to the Spark platform, and use caching mechanism of Spark and the

TABLE 3 | Comparison of root mean square error and model training time result comparison.

Model	RMSE
Single machine running time	517.381
Spark running time	173.805
Single machine + Spark running time	209.712

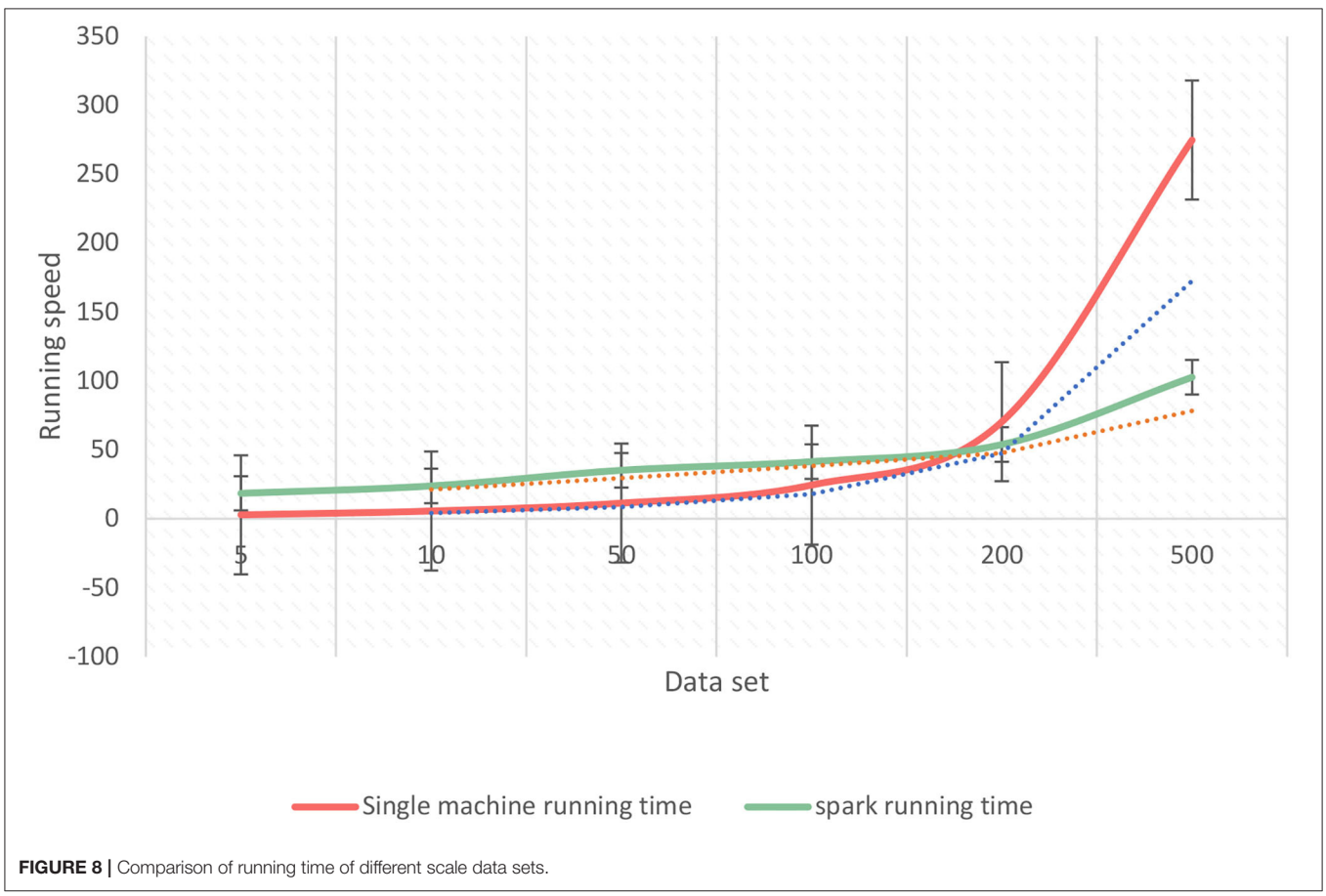


FIGURE 8 | Comparison of running time of different scale data sets.

operation of the resilient distributed datasets (RDD) conversion operator to greatly speed up the execution time, reflecting the advantages of parallelization of the program running in the Spark cluster.

As shown in **Figure 8**, under different data scales, the running speed of a single machine and the running speed under the Spark platform have their own merits. On a scale of <100,000, the running speed of a single machine is faster, because, between nodes under the Spark platform, communication will consume time and will consume more time in communication and scheduling. On the scale of more than 100,000, the running speed of the Spark platform will be faster than that of a single machine, because the Spark platform will divide the data set into many small pieces, then divide it into different tasks according to different types, and then distribute according to the principle of data locality. Therefore, the traditional algorithm will affect the classification ability of Naive Bayes on the assumption of independence, and the optimized Naive Bayes improves the classification accuracy.

As shown in **Table 3**, Spark has achieved better results in the selected theme emotional trend sequence. The Spark model + native model proposed in this article is slightly lower than the Spark model but much higher than the native model, which illustrates this article proposes the validity of the model.

As shown in **Figure 9**, when comparing three sets of experiments, the training time of the Spark model is much lower than the other two models, and the training time of the native model is the longest. In summary, although the Spark model loses part of its accuracy, it greatly shortens the model training time and is more in line with the scenario of public opinion trend prediction. When the data set is increased to a certain extent, memory overflow and other problems will occur in a stand-alone environment, and Spark will greatly accelerate the execution speed because of the caching mechanism and the operation of the RDD conversion operator, which reflects the advantages of parallelization of the program running in the Spark cluster. Therefore, the experiment shows that the classifier based on the Spark platform has an advantage in the running speed of the algorithm over the classifier on a stand-alone machine.

CONCLUSIONS

With the abundance of Internet resources and the vigorous development of mobile Internet. People's participation in the

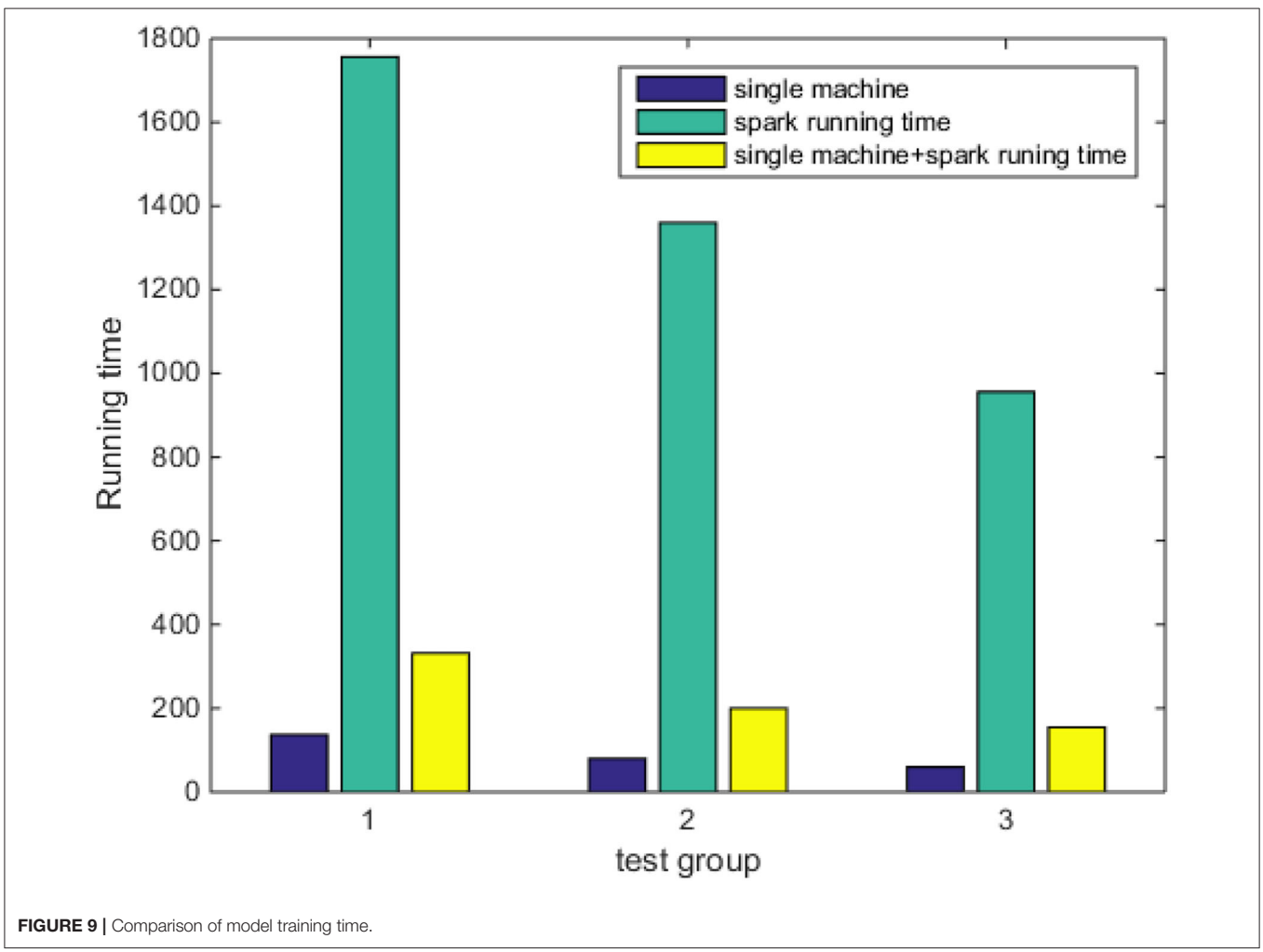


FIGURE 9 | Comparison of model training time.

Internet is increasing. Due to the active participation of people, the digital entrepreneurial ecosystem has great development potential, but because these data exist in an unstructured way. Guidance lacks the ability of text sentiment analysis. Therefore, if emotional information can be extracted from these unstructured data, it will greatly promote the ability of text sentiment analysis, which has great value in scientific research and practical applications. With the rapid development of artificial intelligence technology and the maturity of large-scale data computing frameworks, digital entrepreneurship ecosystems have also become possible. This article mainly studies the ability of text emotional word extraction and trend prediction technology based on the digital entrepreneurial ecosystem and models and analyzes and predicts text data. The optimized Bayesian algorithm is used, which has a higher accuracy rate than the traditional Bayesian algorithm. Using the optimized Bi-LSTM model, the accuracy, recall, and F1 values have been slightly improved. On the Spark platform running down, the training time is greatly reduced. In the digital entrepreneurial ecosystem designed in this article, many functions are not perfect due to the short time. In the next step, it is anticipated to improve the functions of the system.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

JL: writing–editing. MY: data analysis. Both authors contributed to the article and approved the submitted version.

REFERENCES

Allenby, C. E., Babiash, E. S., Blank, P. N., Marco, D. C., Isabelle, G. L., Robin, B. L., et al. (2018). Analysis of the market, regulatory landscape, and current state of clinical trials pertaining to digital health. *Technol. Transf. Entrepreneurship* 5, 21–34. doi: 10.2174/2213809905666180816115016

Bhagat, C., and Mane, D. (2019). Survey on text categorization using sentiment analysis. *J. Sci. Technol. Res.* 8, 1189–1195.

Chang, G., and Huo, H. (2018). A method of fine-grained short text sentiment analysis based on machine learning. *Neural Netw. World* 28, 325–344. doi: 10.14311/NNW.2018.28.019

Chen, Y. S. (2021). Sustainability innovation enabled by digital entrepreneurship in Franchise organizations. *Int. J. E-Entrepreneurship Innov*. 11, 71–85. doi: 10.4018/IJEEI.2021010105

Frank, L. A. V. (2018). Text Sentiment Analysis. *MSDN Mag.* 33, 6–8.

Gabison, G. A. (2019). Government-sponsored patent monetizing entities. *J. Bus. Entrepreneurship Law* 12, 9–9. doi: 10.2139/ssrn.3279544

Gartner, W. B. (1985). A conceptual framework for describing the phenomenon of new venture creation[J]. *Acad. Manage. Rev.* 10, 696–706.

Gnyawali, D. R., and Fogel, D. S. (1994). Environments for entrepreneurship development: key dimensions and research implications. *Entrep. Theory Pract.* 18, 43–62.

Hahn, A., Klug, K., and Riedmüller, F. (2020). Digital empathy: kombinierte erfassung über affective computing und eye tracking. *Digitale Welt* 4, 24–27. doi: 10.1007/s42354-020-0278-4

Hamdi, A., Shaban, K., and Zainal, A. (2018). CLASENTI: a class-specific sentiment analysis framework. *ACM Trans. Asian Lang. Inf. Process.* 17, 32.1–32.28. doi: 10.1145/3209885

Hutson, M. (2018). Artificial intelligence faces reproducibility crisis. *Science* 359, 725–726. doi: 10.1126/science.359.6377.725

Ishmukhametov, E. M., Khisaeva, A. I., and Gaysina, R. (2020). Institutional trends of entrepreneurship development in the republic of Bashkortostan. *Bull. USPTU Sci. Educ. Econ. Ser. Econ*. 2, 92–100. doi: 10.17122/2541-8904-2020-2-32-92-100

Kotelnikov, E. V. (2018). TextJSM: text sentiment analysis method. *Autom. Doc. Math. Linguist*. 52, 24–34. doi: 10.3103/S0005105518010089

Le, X., Chu, J., Deng, S., Jiao, Q., Pei, J., Zhu, L., et al. (2019). CiteOpinion: evidence-based evaluation tool for academic contributions of research papers based on citing sentences. *J. Data Inf. Sci*. 4, 26–41. doi: 10.2478/jdis-2019-0019

Li, C., Biao'an, S., Zhao, L., and Hongqing, G. (2011). Research on the impact of entrepreneurial network on new enterprise performance – the intermediary role of organizational learning. *Sci. Res.* 28, 1592–1600.

Li, N., Jin, X., and Li, Y. (2020). Identification of key customer requirements based on online reviews. *J. Intel. Fuzzy Syst*. 39, 1–14. doi: 10.3233/JIFS-200057

Lian, Z., Li, Y., Tao, J. H., Huang, J., and Niu, M. (2020). Expression analysis based on face regions in read-world conditions. *Int. J. Autom. Comput*. 17, 96–107. doi: 10.1007/s11633-019-1176-9

Liu, J., Nandhakumar, J., and Zachariadis, M. (2018). When guanxi meets structural holes: Exploring the guanxi networks of Chinese entrepreneurs on digital platforms. *J. Strat. Inf. Syst*. 27, 311–334 doi: 10.1016/j.jsis.2018.10.003

Lutfullaeva, M., Medvedeva, M., Komotskiy, E., and Spasov, K. (2018). Optimization of sentiment analysis methods for classifying text comments of bank customers - ScienceDirect. *IFAC-PapersOnLine* 51, 55–60. doi: 10.1016/j.ifacol.2018.11.353

Miao, Y., and Chonghui, G. (2018). Using news to predict Chinese medicinal material price index movements. *Ind. Manage. Data Syst*. 118, 998–1017. doi: 10.1108/IMDS-06-2017-0287

Moore, J. F. (1996). The advent of business ecosystems. *Upside* 7, 30.

Morris, M. H., Davis, D. L., and Allen, J. W. (1994). Fostering corporate entrepreneurship: cross-cultural comparisons of the importance of individualism versus collectivism. *J. Int. Bus. Stud.* 25, 65–89.

Nam, S., and Lee, H. C. (2019). A text analytics-based importance performance analysis and its application to airline service. *Sustainability* 11, 61–78. doi: 10.3390/su11216153

Nuseir, M. T. Al. (2018). Digital media impact on smes performance in the UAE. *Acad. Entrepreneurship J*., 24, 1–13.

Schmidt, T., Schlindwein, M., Lichtner, K., and Wolff, C. (2020). Investigating the relationship between emotion recognition software and usability metrics. *i-com* 19, 139–151. doi: 10.1515/icom-2020-0009

Shen, K. N., Lindsay, V., and Xu, Y. C. (2018). Digital entrepreneurship. *Inf. Syst. J.* 28, 1125–1128. doi: 10.1111/isj.12219

Singh, S. K., and Sachan, M. (2019). SentiVerb system: classification of social media text using sentiment analysis. *Multimed. Tools Appl.* 78, 32109–32136. doi: 10.1007/s11042-019-07995-2

Singh, S. K., and Sachan, M. K. (2021). Classification of code-mixed bilingual phonetic text using sentiment analysis. *Int. J. Semantic Web Inf. Syst*. 17, 59–78. doi: 10.4018/IJSWIS.2021040104

Sodhar, I. N., Jalbani, A. H., Buller, A. H., Channa, M., and Hakro, D. (2020). Sentiment analysis of romanized sindhi text. *J. Intel. Fuzzy Syst*. 38, 1–7. doi: 10.3233/JIFS-179675

Xuebing, P., and Gang, Z. (2009). Research on technological entrepreneurship and technological innovation. *Sci. Technol. Prog. Countermeas.* 27, 15–19.

Zeng, Q., Zhao, X., Hu, X., Duan, H., Zhao, Z., and Li, C. (2021). Learning emotional word embeddings for sentiment analysis. *J. Intel. Fuzzy Syst.* 40, 1–13. doi: 10.3233/JIFS-201993

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Li and Yao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 27 September 2021
doi: 10.3389/fpsyg.2021.759485

Facial Expression Emotion Recognition Model Integrating Philosophy and Machine Learning Theory

*Zhenjie Song**

School of Humanities and Social Sciences, Xi'an Jiaotong University, Xi'an, China

Facial expression emotion recognition is an intuitive reflection of a person's mental state, which contains rich emotional information, and is one of the most important forms of interpersonal communication. It can be used in various fields, including psychology. As a celebrity in ancient China, Zeng Guofan's wisdom involves facial emotion recognition techniques. His book Bing Jian summarizes eight methods on how to identify people, especially how to choose the right one, which means "look at the eyes and nose for evil and righteousness, the lips for truth and falsehood; the temperament for success and fame, the spirit for wealth and fortune; the fingers and claws for ideas, the hamstrings for setback; if you want to know his consecution, you can focus on what he has said." It is said that a person's personality, mind, goodness, and badness can be showed by his face. However, due to the complexity and variability of human facial expression emotion features, traditional facial expression emotion recognition technology has the disadvantages of insufficient feature extraction and susceptibility to external environmental influences. Therefore, this article proposes a novel feature fusion dual-channel expression recognition algorithm based on machine learning theory and philosophical thinking. Specifically, the feature extracted using convolutional neural network (CNN) ignores the problem of subtle changes in facial expressions. The first path of the proposed algorithm takes the Gabor feature of the ROI area as input. In order to make full use of the detailed features of the active facial expression emotion area, first segment the active facial expression emotion area from the original face image, and use the Gabor transform to extract the emotion features of the area. Focus on the detailed description of the local area. The second path proposes an efficient channel attention network based on depth separable convolution to improve linear bottleneck structure, reduce network complexity, and prevent overfitting by designing an efficient attention module that combines the depth of the feature map with spatial information. It focuses more on extracting important features, improves emotion recognition accuracy, and outperforms the competition on the FER2013 dataset.

Keywords: facial expression, emotion recognition, philosophy, machine learning, neural networks

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Weiwei Cai,
Northern Arizona University, United States
Lijun Xu,
Nanjing Institute of Technology (NJIT), China

***Correspondence:**
Zhenjie Song
xuezengguofan@163.com

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *16 August 2021*
Accepted: *06 September 2021*
Published: *27 September 2021*

Citation:
Song Z (2021) Facial Expression Emotion Recognition Model Integrating Philosophy and Machine Learning Theory. Front. Psychol. 12:759485. doi: 10.3389/fpsyg.2021.759485

INTRODUCTION

As an ancient Chinese celebrity, Zeng Guofanno wisdom involves the skill of facial emotion recognition. His book Bing Jian summarizes eight methods on how to identify people, especially how to choose the right one, which means "look at the eyes and nose for evil and righteousness, the lips for truth and falsehood; the temperament for success and fame, the spirit for wealth and fortune; the fingers and claws for ideas, the hamstrings for setback; if you want to know his consecution, you can focus on what he has said." It is said that a person's personality, mind, goodness and badness can be showed by his face. That is to say, complete bones are not as neat as skin, clean skin is not as dignified as facial features. Eyes can reflect a person's good and evil, in contrast to the people who behave indecently, the one with pure and unbiased minds has bright eyes. Therefore, eyes are an important part of facial emotion recognition. Because facial representation (Dubuisson et al., 2002; Sariyanidi et al., 2014; Sun et al., 2018) is an intuitive reflection of human mental state, it contains rich emotional information (D'Aniello et al., 2018; Maydych et al., 2018; Suslow et al., 2020), and can intuitively reflect a person's true thoughts. In daily human communication, we cannot only convey information through language and text, but we also use some movements and facial expressions to complete the communication between people, and research shows that expressions and movements are often more effective than words. Deliver key messages. Facial expression emotion is a common form of non-verbal communication that can effectively communicate personal emotions and intentions. We can observe other people's facial expressions with our eyes, and the brain will analyze the data to determine their mental state, completing the expression and communication of emotions between people. Facial expressions can give language emotions, and facial expressions can clearly show a person's true emotions, which is more accurate than language, in the course of people's communication.

In social situations, humans will naturally express their personal emotions. An accurate understanding of each other's emotions will help build mutual understanding and trust. The expression and understanding of emotions is an essential skill for humans. We mainly convey personal emotions in three ways, namely language, voice and facial expressions. Scholars have found that facial expressions are the most important way of expressing human emotion information (Cai and Wei, 2020; Zhang et al., 2021). Facial expression information accounts for about 55 percent of the information transmitted by the experimenters, voice information for 38 percent, and language information accounts for only 7% of the total information. It's clear that, when compared to language and sound, facial expression information is more important for emotional comprehension. Naturally, researchers concentrate on facial expressions in order to gain a better understanding of human inner emotional activities.

In recent years, as computers have increasingly powerful computing power and huge data sets continue to emerge, machine learning algorithms (Domínguez-Jiménez et al., 2020; Zhang et al., 2020; Cai et al., 2021a) have developed vigorously. Compared with traditional methods, the machine learning algorithm integrates the two processes of feature extraction (Koduru et al., 2020) and classification (Oberländer and Klinger, 2018), reduces the operation process, and can automatically extract the internal features of the sample data, has powerful feature extraction capabilities, and is related to computer vision (CV) (Schmøkel and Bossetta, 2021; Cai et al., 2021b). The performance in various competitions is very good. Among them, Convolutional Neural Network (CNN) (Santamaria-Granados et al., 2018; Ghosal et al., 2019; Gao et al., 2021) is one of the most common machine learning algorithms, and the classification effect of images is excellent, and the recognition accuracy rate of the ImageNet database has not been updated. Therefore, many researchers have begun to use neural networks (Chu et al., 2021; Tong et al., 2021) to solve the recognition problem of facial expressions. However, the facial expression images collected in real life are very uncontrollable, and these uncontrollability increase facial expressions. How to design the structure of the CNN to efficiently and accurately recognize facial expressions still needs continuous exploration.

Based on the above observations, I found that how the CNN can efficiently and accurately recognize facial expressions is the focus of this article. Therefore, I propose a dual-channel emotion recognition algorithm. The first path of the proposed algorithm uses the Gabor feature of the ROI area as input. In order to make full use of the detailed features of the active facial expression area, first segment the active facial expression area from the original face image, use Gabor transform to extract the features of this area, and focus more on the detailed description of the local area. The second path proposes an efficient channel attention network based on deep separable convolution to achieve linear bottleneck structure improvement, reduce network complexity and prevent overfitting, and improve the accuracy of emotion recognition.

The main contributions of this paper are as follows:

(1) This paper proposes a novel feature fusion dual-channel expression recognition algorithm based on machine learning theory and philosophical thinking. It has achieved competitive performance on the FER2013 data set and has a positive significance in promoting the recognition and employment of people.
(2) Aiming at the problem that the features extracted using CNNs ignore the subtle changes in the active areas of facial expressions, the first pass of the proposed algorithm takes the Gabor features of the ROI area as input. In order to make full use of the detailed features of the active areas of facial expressions, start the original face image is segmented into the active area of expression, the Gabor transform is used to extract the features of this area, and the detailed description of the local area is focused.

(3) An efficient attention module is designed to combine the depth of the feature map with the spatial information, focus on the extraction of important features, and use the joint loss function to make the network have a better feature discrimination effect, reduce the difference of the same facial expressions within the class, expand the feature spacing between different facial expressions, and ensure the accuracy of classification.

The rest of this paper is arranged as follows. In section "Related Work," we introduce relevant work, in section "Methodology," we describe the algorithm in this paper, and in section "Experiments and Results," we give experiments and experimental results. Section "Conclusion" presents the research conclusions of this paper.

RELATED WORK

Emotion Recognition Based on Facial Expressions

The process of human communication is inextricably linked to the fluctuation of various emotions. When people are experiencing basic emotions, their faces will display a variety of expression patterns, each with its own set of characteristics and distribution scale. Facial expression recognition is a crucial part of human-computer interaction that allows computers to understand facial expressions based on human thinking. According to the processing of facial expression recognition process can be divided into three important face detection, feature extraction and classification module, face detection as the key technology of face recognition (Adjabi et al., 2020; Zhang et al., 2021) with its rapid development has basic mature, which can effectively extracted from the original face image of excellent characteristics and the characteristics of correct classification becomes key factor affecting the recognition result. For example, Gao and Ma (2020) obtained facial expression attributes from facial images so as to predict emotional states according to facial expression changes.

Speech-Based Emotion Recognition

Language is another way for human beings to express emotions. The speech signals expressed by human beings in different emotional states have different characteristics and rules, such as speed, pitch, duration, etc. The emotion recognition method based on speech is to identify and judge the emotional information of the speaker at this time by studying and analyzing the physical characteristics of the speaker's speech in different emotional states. Ton-That and Cao (2019) applied speech signals to emotion recognition and achieved good results on the voice emotion database. However, on the one hand, individual differences will lead to great differences in speech signals, which requires the establishment of a large phonetic database, which will bring some difficulties to recognition. On the other hand, the noisy environment will affect the sound quality of speech, thus affecting the emotion recognition, so the acquisition of speech signal has a high requirement on the surrounding environment.

Emotion Recognition Based on Physiological Signals

The basis of emotion recognition based on physiological signals is that humans will produce different responses under different stimuli. For example, physiological signals such as brain electricity, electrocardiogram, pulse, and skin electrical response can all reflect emotions. Momennezhad (2018) used EEG signals for emotion recognition, extracting features from the time domain and frequency domain of EEG signals. Although the changes of physiological signals are not controlled by humans, they can most objectively reflect human emotional conditions.

Emotion Recognition Based on Gestures

People will involuntarily undergo some posture changes in different environmental states and moods, and judge human emotions based on physical information such as the time and frequency of these posture changes, according to gesture-based emotion recognition. Ajili et al. (2019) used human movement analysis to identify motion, then evaluated and evaluated the emotions expressed by human motion posture. However, the single use of human gestures for emotion recognition has certain limitations, because many gestures do not have emotional significance or the same gestures have different emotional meanings in different background environments, so human gestures are usually different from others. The modalities (such as expressions, speech, etc.) are combined for emotion recognition.

Expressions are the most intuitive way to convey emotions among several ways to express human emotion information, such as facial expressions, voices, physiological signals, and gestures, and expression information is relatively easy to obtain in most environments, so I use them. Human emotional states are studied using facial expressions as objects.

METHODOLOGY

The Philosophical Definition of Emotion

Psychology defines emotion as: "a special form of reflection of human beings on objective reality is the experience of human attitudes toward whether objective things meet human needs." It can be understood from this definition that emotion is a subjective experience, subjective attitude, or subjective reflection, which belongs to the category of subjective consciousness, not the category of objective existence. Dialectical materialism believes that any subjective consciousness is a reflection of a person's objective existence. Emotion is a special subjective consciousness and must correspond to a special objective existence. The key to the problem lies in whether such a special objective existence can be found. It is not difficult to find that "whether a person's objective things meet people's needs" is actually a typical value judgment problem, "meeting people's needs" is the value characteristic of things, an objective existence, "attitude" and "experience" Both are the way people recognize or reflect the value characteristics of things. In this way, the psychological definition of emotion can be expressed as: "Emotion is the subjective reflection of people on the value characteristics

of things." The objective existence corresponding to emotion should be the value characteristic of things. From this I can get the philosophical definition of emotion: emotion is the subjective reflection of human beings on the value relationship of objective things.

Two-Channel Emotion Recognition Model

This section will elaborate on the proposed dual-channel emotion recognition model from the ROI area division and Gabor feature of the first path, and the efficient channel attention network of the second path.

Division of ROI Area

Let P_{el} and P_{er} denote the positions of the left and right eyes, P_n denote the key points of the tip of the nose, P_{ml} and P_{mr} denote the key points of the left and right corners of the mouth, X_1 and X_2 denote the horizontal coordinates of the left and right borders of the face, and Y_1 and Y_2 denote the vertical positions of the upper and lower edges, respectively, coordinate. Assuming that the position of the eyebrow area is calculated as an example, only the height and width of the area need to be calculated. The calculation equation for the height of the eyebrow area is as follows:

$$\mathrm{H_{eye}} = \begin{cases} \frac{|\mathrm{P_{el}:y-P_n:y}|}{2} + \left|\mathrm{Y_2 - P_{er}:\,y}\right|, \mathrm{P_{el}:\,y \le P_{er}:\,y} \\ \left|\mathrm{Y_2 - P_{el}:\,y}\right| + \frac{|\mathrm{P_{er}:y-P_n:y}|}{2}, \mathrm{P_{el}:\,y \ge P_{er}:\,y} \end{cases} \quad (1)$$

I use W_{eye} to represent the width of the eyebrow region. In order to make the extracted human eye region not only include the eye, but also the part of the information around the corner of the eye, the calculation of W_{eye} directly takes the distance between the left and right edges. The calculation equation is as follows:

$$W_{\mathrm{eye}} = |X_2 - X_1| \quad (2)$$

Similarly, the calculation of the position of the mouth in the face image is as follows, using H_{mouth} and W_{mouth} to represent the height and width of the region, respectively.

$$\mathrm{H_{mouth}} = \begin{cases} \frac{|\mathrm{P_n:y-P_{ml}:y}|}{2} + \left|\mathrm{P_{mr}:\,y - Y_1}\right|, \mathrm{P_{ml}:\,y \le P_{mr}:\,y} \\ \left|\mathrm{P_{ml}:\,y - Y_1}\right| + \frac{|\mathrm{P_n:y-P_{mr}:y}|}{2}, \mathrm{P_{ml}:\,y \ge P_{mr}:\,y} \end{cases} \quad (3)$$

$$W_{\mathrm{mouth}} = |X_2 - X_1| \quad (4)$$

Finally, the rectangular clipping areas of the eyebrow, eye, and mouth in the face image are determined using the above calculation method. The interference of non-key parts of the face can theoretically be reduced, and the calculation cost can be reduced, by cutting out these regions with large facial expression changes.

Gabor Filter

Compared with other wavelets, Gabor transform has a unique biological background. The Gabor filter is similar to the frequency and direction representation of the human visual system in terms of frequency and direction, and can extract local information of different frequencies, spatial positions and directions of the image. A special advantage of Gabor filters is their invariance to scale, rotation and translation. The reason why Gaboe wavelet can be used for facial expression recognition is that when expression changes occur, the key parts of the face such as eyes, mouth, and eyebrows will undergo great changes due to muscle changes. These parts are reflected in the image as gray-scale changes. Severe, the real and imaginary parts of the wavelet will fluctuate at this time, so the amplitude response of the Gabor filter in these parts will be very obvious, so it is very suitable for extracting local features of expressions. In the field of image processing, two-dimensional Gabor filtering is generally used to process images. The kernel function of the two-dimensional Gabor wavelet can be written as:

$$\psi_{\mathrm{uv}}(\mathrm{z}) = \frac{\left|\left|\mathrm{k_{uv}}\right|\right|^2}{\sigma^2} \times \mathrm{e}^{\frac{||\mathrm{k_{uv}}||^2||\mathrm{z}||^2}{2\sigma^2}} \times \left(\mathrm{e}^{\mathrm{ik_{uv}z}} - \mathrm{e}^{\frac{\sigma^2}{2}}\right) \quad (5)$$

where u and v represent the direction and frequency of the Gabor wavelet kernel, $z=(x,y)$ represents the position of a certain pixel in the image, σ represents the filter bandwidth, and $|\mathrm{k_{uv}}|^2/\sigma^2$ is used to compensate for the attenuation of the energy spectrum determined by the frequency.

The Gabor feature of the facial expression image can be obtained by convolving the facial expression image and the Gabor wavelet kernel. Assuming that the gray value of the (x, y) point in the facial expression image is set to k, the calculation equation for the Gabor feature is as follows :

$$G_{uv}(x, y) = I(x, y) * \psi_{uv}(x, y) \quad (6)$$

where $G_{uv}(x,y)$ represents the Gabor feature of the extracted image, $\psi_{uv}(x,y)$ represents the kernel function of the two-dimensional Gabor wavelet, and $*$ represents the convolution operation.

Feature Fusion

To make full use of the features of the key areas of facial expressions and make up for the lack of global representation in the local features extracted by Gabor, the features extracted by CNN and the local features of the ROI region extracted by Gabor are feature-fused. Simply put, feature fusion is to combine multiple different features extracted by different algorithms into a new feature with stronger characterization capabilities through a certain fusion method. The process of feature fusion is shown in **Figure 1**.

There are three ways of feature fusion, namely summation, product and splicing. Among the three feature fusion methods, the method of splicing and fusion is simple in calculation and less computationally expensive. Therefore, this article uses splicing and fusion to compare Gabor features and CNN. Feature fusion is performed, and the local features of the ROI region extracted by Gabor and the features extracted by CNN are fused in the fully connected layer of the CNN. Suppose two feature vectors with the same dimension are defined as $X=(x_1,x_2,\cdots,x_n)$ and Y = 1, then the calculation equation for feature stitching is as follows:

$$Z = (x_1, x_2, \cdots, x_n, y_1, y_2, \cdots, y_n) \quad (7)$$

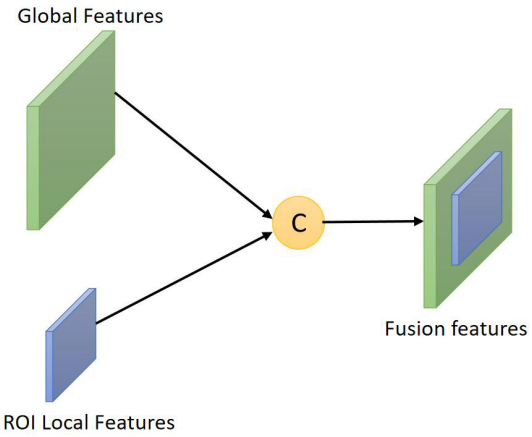


FIGURE 1 | Schematic diagram of feature fusion.

Channel Attention Model

To extract more core expression features from the facial expression feature map, I introduces the ultra-lightweight attention module ECA-Net (Wang et al., 2020) to weight the attention of the improved linear bottleneck structure, and give greater weight to the core features, Which makes the network pay more attention to the core features of expressions. This structure contains only one parameter k, but the performance improvement it brings is very obvious. The main function of this module is to generate weights for each channel and learn the correlation between features, just like humans always selectively ignore non-critical information, but instead focus on information that is useful to us. This module The purpose is to let the network model ignore some non-core features, increase the emphasis on core features, and the module only adds a very small amount of additional parameters.

As shown in **Figure 2**, the distribution of features in the spatial dimension is compressed and extracted from a two-dimensional matrix to a single value, and this value obtains the feature information in this space. Then through a fully connected layer to complete the channel dimension reduction, and then through the second fully connected layer to complete the channel dimensionality, so as to obtain the relevant dependencies between the various channels, generate weights for each feature channel, and comprehensively obtain one of the feature channels. For the correlation between the channels, the important channel features of facial expressions are generated with larger weights, and on the contrary, smaller weights are generated, that is, the attention mechanism is introduced. It generates channel weights through local one-dimensional convolution in high dimensions, and obtains the correlation dependency between each channel. The side effect of channel dimension reduction on the direct correspondence between channels and weights is avoided, and obtaining appropriate cross-channel correlation dependencies is more efficient and accurate for establishing the channel attention mechanism. The last two attention modules both multiply the weights of the generated channels to the original input feature map, and merge the features weighted by attention with the original features to complete the feature attention weighting in the channel space.

TABLE 1 | Comparison of experimental results with different methods.

Methods	Acc
InceptionV4 (Szegedy et al., 2017)	0.7080
DNNRL (Kim et al., 2016)	0.7082
ICL (Liu and Zhou, 2020)	0.7215
ABP (Liu et al., 2019)	0.7316
MobileNetV3	0.7189
Ours	**0.7400**

The bold values represent the best values of the adopted methods.

Overall Structure

Figure 3 shows the overall frame diagram of this algorithm. Firstly, the feature extraction module is composed of two different CNN branches: The first CNN branch takes the Gabor feature as the input. In order to make full use of the features of the regions with obvious facial expression changes and rich facial information, the original face image should be preprocessed and the emotion-related ROI region should be trimmed out, and then Gabor wavelet changes should be used to extract the ROI feature. Since the extracted features still have a high dimension, feature mapping processing of Gabor features is required before feature fusion. This CNN is composed of two convolution layers, which reduces the size of Gabor features and facilitates subsequent feature fusion. In the second path, an efficient channel attention network based on deep separable convolution is proposed to improve the linear bottleneck structure, reduce network complexity and prevent overfitting. By designing an efficient attention module, the depth of the feature map is combined with spatial information, focusing more on the extraction of important features, and improving the accuracy of emotion recognition. Finally, the feature classification module classifies the fused features through Softmax layer.

EXPERIMENTS AND RESULTS

Experimental Setup

The research in this paper is carried out on the PC platform, and the experiment is carried out on the ubuntu18.04 operating system. The experiment is based on the Pytorch deep learning framework, and the programming language uses Python 3.6.5. The hardware platform is: Intel Core i7-9700k CPU, 16GB memory, GPU is GTX1080TI, video memory is 11GB. In order to ensure the fairness of the experiment on the improved network and the comparison network, the training parameters used in the experiment are exactly the same. All model training strategies adopt the learning rate attenuation strategy. The initial learning rate is 0.01 and the attenuation coefficient is 0.0001, the batch size is 128. After the model training is completed, all images in the training data set are called an epoch, and 150 epochs are set in the experiment. In order to optimize the network faster, I use the Adam optimization algorithm.

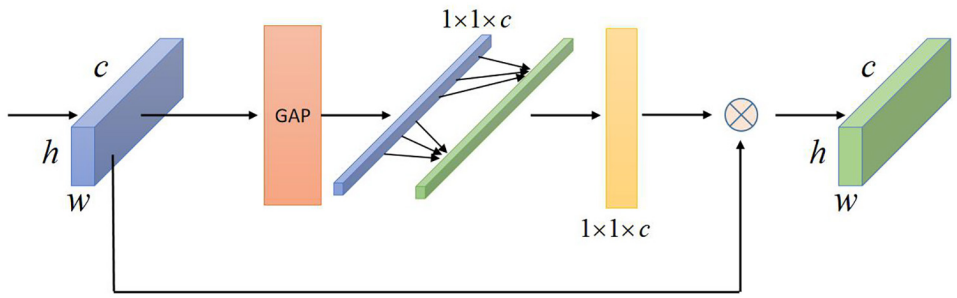


FIGURE 2 | Schematic diagram of the channel attention model.

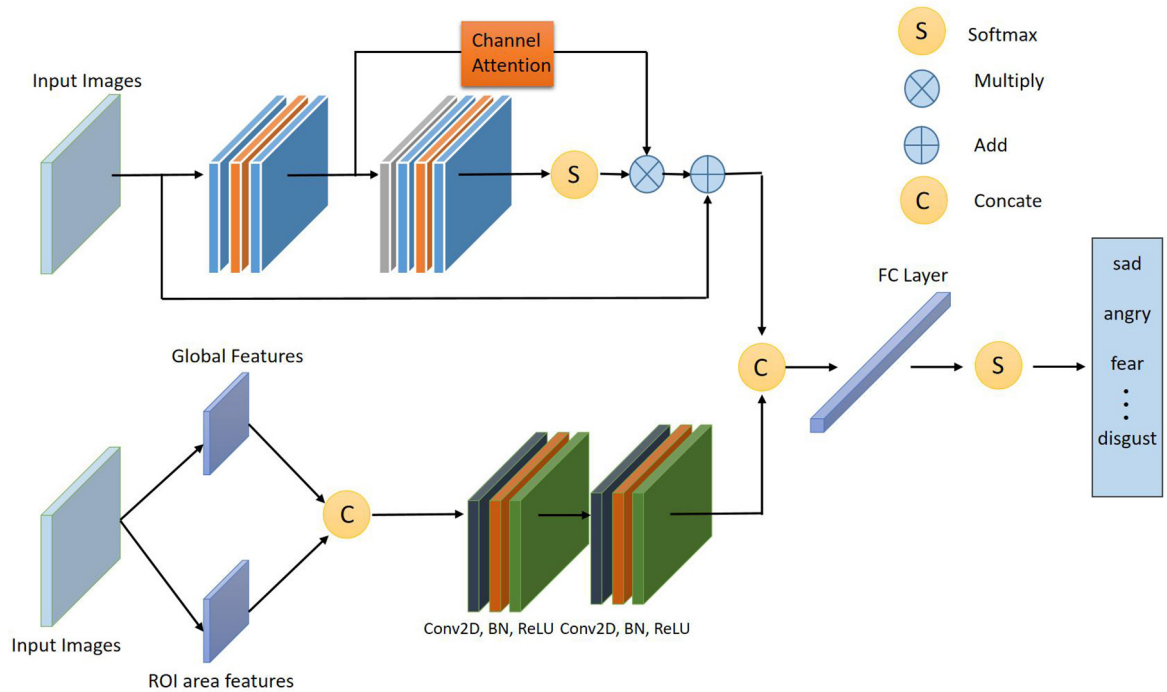


FIGURE 3 | Schematic diagram of the overall framework of the proposed algorithm.

Experimental Data Set

The data set used in this article is FER-2013. Due to the small amount of data in the original facial expression data set, it is far from enough for data-driven deep learning, so data augmentation is a very important operation. In the network training stage, in order to prevent the network from overfitting, I first do a series of random transformations, including flipping, rotating, cutting, etc., and then transform the data image size to 104 × 104 size, and then randomly cut into 96 × 96 size Image, and then randomly rotate the image between 0 and 15° and perform horizontal mirroring operation, and then send it to the network for training. In the network test stage, we cut the four corners and the center of the image to obtain five 96 × 96 images, and then perform the horizontal mirroring operation, respectively, which is equivalent to amplifying the data by 10 times. I input the amplified picture into the network for its recognition, and average the results, and finally the output classification with the highest score is the corresponding expression. This method can expand the size of the data set, make the trained network model more generalized and robust, and further improve the accuracy of recognition.

Evaluation Method

The overall accuracy rate is used as the evaluation index of this study, and its calculation formula is as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

where TP represents the positive samples predicted by the model as positive, TN represents the negative samples predicted by the model as negative, FP represents the negative samples predicted by the model as positive, and FN represents the positive samples predicted by the model as negative.

TABLE 2 | Confusion matrix of recognition rate on FER-2013 dataset.

	Anger	Fear	Disgust	Happy	Sad	Surprised	Normal
Anger	**0.66**	0.01	0.11	0.04	0.11	0.01	0.06
Fear	0.20	**0.69**	0.04	0.05	0.00	0.00	0.02
Disgust	0.09	0.00	**0.64**	0.02	0.12	0.05	0.09
Happy	0.01	0.00	0.02	**0.91**	0.02	0.01	0.03
Sad	0.05	0.00	0.11	0.05	**0.68**	0.00	0.13
Surprised	0.02	0.00	0.06	0.05	0.02	**0.83**	0.02
Normal	0.03	0.00	0.03	0.05	0.12	0.01	**0.77**
Acc				**0.74**			

The bold values represent the best values of the adopted methods.

Experimental Results

In order to verify the reliability of the overall algorithm in this paper, this paper has carried out comparative experiments on the FER-2013 data set with the current advanced expression recognition network to evaluate the performance of the algorithm in this paper. The experimental results are shown in **Table 1**.

Table 1 is a comparison of the recognition rates of different methods. This article uses a lightweight network structure. When compared with MobileNetV3 and Inception, which are also lightweight networks, the accuracy of the FER-2013 data set is improved with fewer model parameters. Increased by 3.3 and 4.7%. Compared with the current mainstream methods on the FER-2013 data set, DNNRL (Kim et al., 2016) proposed combining multiple CNNs and using weighted joint decision-making methods, and ICL (Liu and Zhou, 2020) proposed clustering to obtain the center distance of expression classes and continuously Adding the method of difficult sample training and the method of combining bilinear pooling and attention mechanism proposed by ABP (Liu et al., 2019), the method in this paper has achieved superior performance and achieved a higher recognition rate of 74.00%.

In order to avoid misjudgment of model performance when only the overall recognition rate is used as the evaluation index, we conduct detailed experiments on the recognition results of each type of expression through the confusion matrix. The confusion matrix is also called the error matrix. Each row represents the expression prediction label, and each column represents the actual expression label. Using the confusion matrix can clearly observe the recognition of each type of data, and from the recognition accuracy of each type of expression, we can analyze the performance of the network model in more detail.

Table 2 is the confusion matrix of the high-efficiency channel attention model proposed in this paper for the recognition results of the FER-2013 test set. The data in bold on the diagonal line in the table represents the recognition accuracy of each type of expression correctly classified, and the remaining data are expression errors. The proportion of classification, the last line is the average recognition accuracy of all expressions that are correctly classified. For example, the neutral expression recognition accuracy in the lower right corner of the diagonal of the confusion matrix is 0.77, which indicates that 76% of the expression samples in the expression data set are correctly predicted. It can be seen that the recognition accuracy of happy and surprised expressions is high, with accuracy rates of 0.91 and 0.83, respectively, while the recognition accuracy of disgusted and angry expressions is low, with accuracy rates of only 0.64 and 0.66. By observing the samples, it is found that there are indeed many similarities between the facial morphology of fear and surprise, anger and disgust, and the number of samples is far smaller than the number of happy and surprised, which leads to the model's insufficient learning of its features, so the recognition rate is low. Finally, the average recognition rate of the model in the FER-2013 test set reached 0.74.

TABLE 3 | Results of ablation experiments with feature fusion.

Methods	Acc
Add	0.7285
Mul	0.7195
C	**0.7400**

The bold values represent the best values of the adopted methods.

TABLE 4 | Results of ablation experiments with attention.

Methods	Acc
SA	0.7395
CA	**0.7400**

The bold values represent the best values of the adopted methods.

Ablation Experiment for Feature Fusion

To verify the influence of the feature fusion strategy on the performance of the proposed algorithm, an ablation experiment is set up in this section, where add represents the feature addition strategy, mul represents the feature multiplication strategy, and C represents the feature concat strategy. The results of the ablation experiment are shown in **Table 3**.

It can be seen from **Table 3** that the feature concat strategy has achieved the best results. In addition, the feature addition strategy is better than the feature multiplication strategy. Therefore, this proves that the proposed algorithm is effective in adopting the feature concat strategy.

Ablation Experiment for Attention Model

To verify the influence of the channel attention mechanism on the performance of the proposed algorithm, an ablation experiment is set up in this section. CA stands for the channel attention

mechanism and SA stands for the spatial attention mechanism. The results of the ablation experiment are shown in **Table 4**.

It can be seen from **Table 4** that the channel attention mechanism has achieved better performance, which proves the superiority of CA in facial emotion recognition.

CONCLUSION

In this paper, I propose a novel feature fusion dual-channel expression recognition algorithm based on machine learning theory and emotional philosophy. Because features extracted using CNNs ignore subtle changes in the active regions of facial expressions, the proposed algorithm's first path takes the Gabor feature of the ROI region as input. The active facial expression region is first segmented from the original face image, and the features of this region are extracted using Gabor transform, focusing more on the detail description of the local region, in order to make full use of the detail feature of the active facial expression region. To improve the linear bottleneck structure, reduce network complexity, and avoid overfitting, a channel attention network based on deep separable convolution is proposed in the second path. The depth of the feature map is combined with spatial information by designing an efficient attention module, focusing more on the extraction of important features and improving the accuracy of emotion recognition. On the FER2013 data sets, competitive performance was achieved. Furthermore, this research will serve as a guide for promoting people selection, and it also confirms that Zeng Guofan's philosophy of employing people is effective.

In future work, we will investigate the feasibility of real-time face recognition, and will use the Internet of Things technology to collect faces in real time for emotion recognition.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data/.

AUTHOR CONTRIBUTIONS

ZS was responsible for designing the framework of the entire manuscript, from topic selection to solution to experimental verification.

REFERENCES

Adjabi, I., Ouahabi, A., Benzaoui, A., and Taleb-Ahmed, A. (2020). Past, present, and future of face recognition: a review. *Electronics* 9:1188. doi: 10.3390/electronics9081188

Ajili, I., Mallem, M., and Didier, J. Y. (2019). Human motions and emotions recognition inspired by LMA qualities. *Vis. Comput.* 35, 1411–1426. doi: 10.1007/s00371-018-01619-w

Cai, W., Song, Y., and Wei, Z. (2021a). Multimodal data guided spatial feature fusion and grouping strategy for E-commerce commodity demand forecasting. *Mob. Inf. Syst.* 2021:5568208. doi: 10.1155/2021/5568208

Cai, W., and Wei, Z. (2020). PiiGAN: generative adversarial networks for pluralistic image inpainting. *IEEE Access* 8, 48451–48463. doi: 10.1109/ACCESS.2020.2979348

Cai, W., Wei, Z., Song, Y., Li, M., and Yang, X. (2021b). Residual-capsule networks with threshold convolution for segmentation of wheat plantation rows in UAV images. *Multimed. Tools Appl.* 1–17. doi: 10.1007/s11042-021-11203-5

Chu, Z., Hu, M., and Chen, X. (2021). Robotic grasp detection using a novel two-stage approach. *ASP Trans. Internet Things* 1, 19–29. doi: 10.52810/TIOT.2021.100031

D'Aniello, B., Semin, G. R., Alterisio, A., Aria, M., and Scandurra, A. (2018). Interspecies transmission of emotional information via chemosignals: from humans to dogs (*Canis lupus* familiaris). *Anim. Cogn.* 21, 67–78. doi: 10.1007/s10071-017-1139-x

Domínguez-Jiménez, J. A., Campo-Landines, K. C., Martínez-Santos, J. C., Delahoz, E. J., and Contreras-Ortiz, S. H. (2020). A machine learning model for emotion recognition from physiological signals. *Biomed. Signal Process. Control* 55:101646. doi: 10.1016/j.bspc.2019.101646

Dubuisson, S., Davoine, F., and Masson, M. (2002). A solution for facial expression representation and recognition. *Signal Process. Image Commun.* 17, 657–673. doi: 10.1016/S0923-5965(02)00076-0

Gao, H., and Ma, B. (2020). A robust improved network for facial expression recognition. *Front. Signal Process.* 4:4. doi: 10.22606/fsp.2020.44001

Gao, M., Cai, W., and Liu, R. (2021). AGTH-net: attention-based graph convolution-guided third-order hourglass network for sports video classification. *J. Healthc. Eng.* 2021:8517161. doi: 10.1155/2021/8517161

Ghosal, D., Majumder, N., Poria, S., Chhaya, N., and Gelbukh, A. (2019). Dialoguegcn: a graph convolutional neural network for emotion recognition in conversation. *arXiv* [Preprint]. arXiv:1908.11540 doi: 10.18653/v1/D19-1015

Kim, B. K., Roh, J., Dong, S. Y., and Lee, S. Y. (2016). Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. *J. Multimodal User Interfaces* 10, 173–189. doi: 10.1007/s12193-015-0209-0

Koduru, A., Valiveti, H. B., and Budati, A. K. (2020). Feature extraction algorithms to improve the speech emotion recognition rate. *Int. J. Speech Technol.* 23, 45–55. doi: 10.1007/s10772-020-09672-4

Liu, L., Zhang, L., and Jia, S. (2019). "Attention bilinear pooling for fine-grained facial expression recognition," in *Proceedings of the International Symposium on Cyberspace Safety and Security*, (Cham: Springer), 535–542. doi: 10.1007/978-3-030-37352-8_47

Liu, X., and Zhou, F. (2020). Improved curriculum learning using SSM for facial expression recognition. *Vis. Comput.* 36, 1635–1649. doi: 10.1007/s00371-019-01759-7

Maydych, V., Claus, M., Watzl, C., and Kleinsorge, T. (2018). Attention to emotional information is associated with cytokine responses to psychological stress. *Front. Neurosci.* 12:687.

Momennezhad, A. (2018). EEG-based emotion recognition utilizing wavelet coefficients. *Multimed. Tools Appl.* 77, 27089–27106. doi: 10.1007/s11042-018-5906-8

Oberländer, L. A. M., and Klinger, R. (2018). "An analysis of annotated corpora for emotion classification in text," in *Proceedings of the 27th International Conference on Computational Linguistics*, (Santa Fe, NM: Association for Computational Linguistics), 2104–2119.

Santamaria-Granados, L., Munoz-Organero, M., Ramirez-Gonzalez, G., Abdulhay, E., and Arunkumar, N. J. I. A. (2018). Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS). *IEEE Access* 7, 57–67. doi: 10.1109/ACCESS.2018.2883213

Sariyanidi, E., Gunes, H., and Cavallaro, A. (2014). Automatic analysis of facial affect: a survey of registration, representation, and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 1113–1133. doi: 10.1109/TPAMI.2014.2366127

Schmøkel, R., and Bossetta, M. (2021). FBAdLibrarian and Pykognition: open science tools for the collection and emotion detection of images in Facebook political ads with computer vision. *J. Inf. Technol. Polit.* 1–11. doi: 10.1080/19331681.2021.1928579

Sun, W., Zhao, H., and Jin, Z. (2018). A complementary facial representation extracting method based on deep learning. *Neurocomputing* 306, 246–259. doi: 10.1016/j.neucom.2018.04.063

Suslow, T., Hußlack, A., Kersting, A., and Bodenschatz, C. M. (2020). Attentional biases to emotional information in clinical depression: a systematic and meta-analytic review of eye tracking findings. *J. Affect. Disord.* 274, 632–642. doi: 10.1016/j.jad.2020.05.140

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. New York: Association for Computing Machinery (ACM)

Tong, Y., Yu, L., Li, S., Liu, J., Qin, H., and Li, W. (2021). Polynomial fitting algorithm based on neural network. *ASP Trans. Pattern Recognit. Intell. Syst.* 1, 32–39. doi: 10.52810/TPRIS.2021.100019

Ton-That, A. H., and Cao, N. T. (2019). Speech emotion recognition using a fuzzy approach. *J. Intell. Fuzzy Syst.* 36, 1587–1597. doi: 10.3233/JIFS-18594

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., and Hu, Q. (2020). "ECA-Net: efficient channel attention for deep convolutional neural networks," in *Proceedings of the 2020 IEEE CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Piscataway, NJ: IEEE). doi: 10.1109/CVPR42600.2020.01155

Zhang, J., Yin, Z., Chen, P., and Nichele, S. (2020). Emotion recognition using multi-modal data and machine learning techniques: a tutorial and review. *Inf. Fusion* 59, 103–126. doi: 10.1016/j.inffus.2020.01.011

Zhang, L., Sun, L., Yu, L., Dong, X., Chen, J., Cai, W., et al. (2021). "ARFace: attention-aware and regularization for face recognition with reinforcement learning," in *IEEE Transactions on Biometrics, Behavior, and Identity Science*, (Piscataway, NJ: IEEE). doi: 10.1109/TBIOM.2021.3104014

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Song. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 28 September 2021
doi: 10.3389/fpsyg.2021.758967

Sentiment Classification of News Text Data Using Intelligent Model

*Shitao Zhang**

School of Network Communication, Zhejiang Yuexiu University, Shaoxing, China

Text sentiment classification is a fundamental sub-area in natural language processing. The sentiment classification algorithm is highly domain-dependent. For example, the phrase "traffic jam" expresses negative sentiment in the sentence "I was stuck in a traffic jam on the elevated for 2 h." But in the domain of transportation, the phrase "traffic jam" in the sentence "Bread and water are essential terms in traffic jams" is without any sentiment. The most common method is to use the domain-specific data samples to classify the text in this domain. However, text sentiment analysis based on machine learning relies on sufficient labeled training data. Aiming at the problem of sentiment classification of news text data with insufficient label news data and the domain adaptation of text sentiment classifiers, an intelligent model, i.e., transfer learning discriminative dictionary learning algorithm (TLDDL) is proposed for cross-domain text sentiment classification. Based on the framework of dictionary learning, the samples from the different domains are projected into a subspace, and a domain-invariant dictionary is built to connect two different domains. To improve the discriminative performance of the proposed algorithm, the discrimination information preserved term and principal component analysis (PCA) term are combined into the objective function. The experiments are performed on three public text datasets. The experimental results show that the proposed algorithm improves the sentiment classification performance of texts in the target domain.

Keywords: transfer learning, text sentiment, cross-domain, intelligent model, sentiment classification

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Jing Xue,
Wuxi People's Hospital Affiliated to Nanjing Medical University, China
Tongguang Ni,
Changzhou University, China

***Correspondence:**
Shitao Zhang
wisepro1982@163.com

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *15 August 2021*
Accepted: *08 September 2021*
Published: *28 September 2021*

Citation:
Zhang S (2021) Sentiment Classification of News Text Data Using Intelligent Model. Front. Psychol. 12:758967. doi: 10.3389/fpsyg.2021.758967

INTRODUCTION

News media platforms and various social media platforms produce a large amount of content every day, including a large number of comments generated by network users. These views and opinions often potentially express their feelings or sentiments. How to exploit users' sentimental information and analyze their sentimental tendency from these massive comment data has potential application value in many aspects. For example, government departments analyze users' attitudes toward specific events and topics, grasp public opinions, and keep abreast of public opinion to make appropriate decisions. Commercial enterprises can keep abreast of the market's reputation by evaluating the relevant content of goods and services, to further improve the quality of their products and services.

Sentiment analysis, also known as sentiment classification, is the field of study that analyses the opinions, views, and sentimental attitudes that people display in texts (Wen et al., 2020;

Birjali et al., 2021). Conventional sentiment classification methods are roughly grouped into sentiment dictionary methods and machine learning methods (Zhang et al., 2018; Ahuja et al., 2019). The classification method based on sentiment dictionary mainly uses a manually collated and constructed sentiment dictionary library. The sentiment score of the text is then calculated according to the predetermined rules. Finally, the results of sentiment classification can be obtained based on the obtained sentiment score. There are several obvious shortcomings in the classification method based on sentiment dictionaries. The accuracy of classification results depends on the size of the whole sentiment dictionary resources. When the number of dictionary resources is not large enough, the accuracy of classification results is often not high. In addition, such methods are also difficult to deal with the implied sentimental content. The core of the machine learning methods is the effective feature extraction and classifier. By constructing the feature set, the classifier is trained on the feature set of the training set, and the sentiment label is output for the unlabeled text. In recent years, the commonly used algorithms for text sentiment analysis and text classification include deep learning (Chen et al., 2020; Zhang et al., 2020), decision tree (Almunirawi and Maghari, 2016), support vector machine (Ahmad et al., 2017), sparse representation (Unnikrishnan et al., 2019; Gu et al., 2020), KNN classifier (Bozkurt et al., 2019), Naïve Bayes (Alshamsi et al., 2020), fuzzy logic (Chaturvedi et al., 2019), and extreme learning machine (Lauren et al., 2018; Waheeb et al., 2020).

However, machine learning methods depend on a large number of high-quality label texts, and the manually labeled samples obviously cannot fully meet the needs of sentiment analysis. The expression of sentiment in a text is closely related to the described semantic concept (domain), and the description of sentiment in different domains differs significantly. The distribution of data characteristics in different fields is also different. In this case, directly using the classifiers trained in other domains for text sentiment analysis will lead to the poor adaptability problem. For example, the word "unpredictable" expresses the positive sentiment in literary and artworks, such as "This movie is both exciting and unpredictable! The film is worth seeing!" but in the field of electronic, the word "unpredictable" expresses the negative sentiment, such as "Even if the power is maintained, it is unpredictable. It's terrible!" As another example, the words "excellent" and "horrible" appear in both domains of electronic and literary works, but "high resolution" and "run fast" rarely appear in the field of literary works, and "printed well" rarely appears in the field of electronic products. Therefore, it is inappropriate to directly use the sentiment classifier trained in the electronic domain to predict the sentiment of comments in the literary works domain.

Recently, the research on cross-domain sentiment classifiers using transfer learning technology becomes a research hotspot (Meng et al., 2019). Transfer learning is a machine learning technology that extracts knowledge from the source domain (the similar but different domain) and applies it to the target domain (the current domain) (Jiang et al., 2020, 2021). The common transfer learning algorithms include sample-based algorithms and feature space-based algorithms (Fu and Liu, 2021; Zhao et al., 2021). The former includes feature selection and feature projection. The sample-based algorithm mainly selects the samples that are valuable for the classification of the target domain from the source domain. To reduce the impact of negative transfer, Gui et al. (2015) developed a negative transfer detection algorithm. This algorithm detected multiple quality levels and retained the high-quality samples. Chen et al. (2010) developed a combined feature level and instance level transfer learning algorithm. This algorithm combined the samples from different domains by evaluating the classification performance in the target domain. Tian et al. (2019) developed an instance-based algorithm for imbalanced cross-linguistic viewpoint analysis. This algorithm translated other relevant markup datasets into the target domain as the supplementary training data.

Feature space-based transfer learning methods include feature selection and feature projection (Xia et al., 2013). The purpose of feature selection is to find the shared features between different domains through some strategies; and then use these features for knowledge transfer. The feature projection strategy maps the features of each domain into a shared feature representation space and establishes the association between the features of each domain. The difference between feature projection and feature selection is that the projected features are not among the original features, but new features. Wu and Tan (2011) developed a two-stage transfer learning sentiment classification algorithm. In the "bridge-building" stage, a bridge between different domains was built to obtain some of the most reliable labeled texts in the target domain. In the "structure following" stage, the internal structure was adopted to label the texts by using the obtained reliable texts. Du et al. (2020) developed a Wasserstein-based transfer network. This algorithm also used a recurrent neural network to capture useful features. López et al. (2019) developed an evolutionary ensemble algorithm for text sentiment analysis. This algorithm was built on a set of evolutionary algorithms. The optimization of the model was achieved by optimizing each sub-algorithm. Wang et al. (2018) developed a measure index called sentiment-related index to measure associations between different domains, and then used this index as a bridge between different domains of features. Tang et al. (2021) developed a graph domain adversarial transfer network for text sentiment classification. This algorithm used a gradient reversal layer to obtain the domain-invariant text features and adopted a projection mechanism to obtain the domain-independent feature representations. Fei et al. (2020) developed a deep learning structure-based transfer learning algorithm, which combined the cross-entropy and weighted for word into the recurrent neural network framework.

Dictionary learning plays a key role in sparse representation and low-rank modeling. The basic operation in sparse representation is called "sparse coding," which involves reconstructing the data representation using a sparse set of constructive blocks (called "atoms"), and the atoms are clustered in a structure called a "dictionary" (Gu et al., 2021; Ni et al., 2021). Therefore, the dictionary learning algorithm has data self-adaptability, which aims to learn a series of basic atoms

from the dataset to linearly approximate the given data. Since dictionary learning learns a set of atoms to obtain discriminative feature representation, it performs well in classification tasks. In this study, I develop a transfer learning discriminative dictionary learning (TLDDL) algorithm for cross-domain text sentiment classification. Considering the distribution difference between different domains, I adopt the subspace technology to project all samples into a subspace. In the common subspace, a domain-invariant dictionary is built to connect different domains. Moreover, the discrimination information preserved term and PCA term are embedded into the objective function, so that the performance of TLDDL for classification tasks across domains is enhanced. TLDDL learns all parameters in an alternating iteration manner. By comparing several non-transfer learning and transfer learning algorithms on toutiao-text (Zhang et al., 2021), Chinese Weibo (Bai et al., 2020), and Amazon review (Pan et al., 2010) datasets, one can draw that the TLDDL algorithm is efficient.

The rest of this paper is organized as follows: the next section presents dictionary learning. The proposed TLDDL algorithm and its optimization are described in section "Transfer Learning Discriminative Dictionary Learning Algorithm". Experimental results are given in section "Experiment." Finally, section "Conclusion" concludes the paper.

BACKGROUNDS

Given a data matrix $\mathbf{Y}$, the dictionary learning algorithm attempts to learn the dictionary $\mathbf{D}$ so that the data $\mathbf{Y}$ can be approximately linearly expressed as $\mathbf{Y} \approx \mathbf{DS}$. Dictionary learning can be represented as,

$$\begin{aligned} &\arg\min_{D,S} \ \| \mathbf{Y}-\mathbf{DS} \|_F^2, \\ &s.t. \ \forall i, \| s_i \|_0 \leq T_i, \end{aligned} \tag{1}$$

where T_i is the given threshold. $\mathbf{S}$ is the sparse coding matrix.

The constraints are used to constrain the complexity of the dictionary matrix. Dictionary learning can also be represented as the following optimization problem, where the constraints are replaced as well as the objective function,

$$\begin{aligned} &\min_{D,S} \ \| \mathbf{s}_i \|_0, i = 1, ..., n \\ &s.t. \ \| \mathbf{Y}-\mathbf{DS} \|_F^2 < \varepsilon, \\ &\quad \| d_j \|_2^2 < 1, j = 1, ..., K \end{aligned} \tag{2}$$

Where $\| \cdot \|_0$ represents the 0 norm of the vector. n and K are the number of training samples and atoms, respectively. It is difficult to optimize the 0 norm in Equations 1, 2, so the 1 norm is commonly used in dictionary learning. For the above optimization problem, after the transformation of the Lagrange equation, Equation 1 can be represented as follows,

$$\begin{aligned} &\min_{D,S} \ \| \mathbf{Y}-\mathbf{DS} \|_F^2 + \lambda \| s_i \|_1, i = 1, ..., n \\ &s.t. \ \| d_j \|_2^2 < 1, j = 1, ..., K \end{aligned} \tag{3}$$

It is easily can be seen that dictionary learning minimizes the reconstruction error $\| \mathbf{Y}-\mathbf{DS} \|_F^2$, while the 1 norm of sparse coding $\mathbf{s}_i$ is as small as possible, that is, the sparsity of $\mathbf{S}$ is as strong as possible. λ is the trade-off parameter. The optimization process of dictionary learning generally includes two alternating iterative steps, one is to solve sparse coding $\mathbf{S}$, and the other is to solve dictionary $\mathbf{D}$. The optimization problem of sparse coding $\mathbf{S}$ is

$$\min_{D,S} \ \| \mathbf{Y}-\mathbf{DS} \|_F^2 + \lambda \| s_i \|_1, i = 1, ..., n \tag{4}$$

And the optimization problem of dictionary $\mathbf{D}$ is

$$\begin{aligned} &\min_{D,S} \ \| \mathbf{Y}-\mathbf{DS} \|_F^2, \\ &s.t. \ \| d_j \|_2^2 < 1, j = 1, ..., K \end{aligned} \tag{5}$$

By alternately optimizing the above problems until convergence, the optimal $\mathbf{D}$ and $\mathbf{S}$ can be obtained. For the sample $\mathbf{y}$ on the test set, the corresponding sparse coding can be obtained by solving Equation 4, and then the classification is performed using the obtained sparse coding.

TRANSFER LEARNING DISCRIMINATIVE DICTIONARY LEARNING ALGORITHM

Objective Function

It is known that when there are few domain-invariant features between different domains, the cross-domain classification performance will decline. To extract as sufficient domain-invariant features as possible and reduce their differences between different domains, the TLDDL algorithm uses the dictionary learning model to learn a domain-invariant dictionary as a bridge to realize the association of the two domains. The source domain is $\mathbf{Y}_s = \{\mathbf{y}_{s,1}, \mathbf{y}_{s,2}, ..., \mathbf{y}_{s,n_s}\}$, and the target domain is $\mathbf{Y}_t = \{\mathbf{y}_{t,1}, \mathbf{y}_{t,2}, ..., \mathbf{y}_{t,n_t}\}$, where n_s and n_t are the number of the source domain and the target domain, respectively.

Firstly, I use the k-nearest neighbor knowledge to represent the local structure and label information of the original data. I think that if the sample $\mathbf{y}_j$ is in the k-nearest neighbor of its same-class sample $\mathbf{y}_i$, then the corresponding sparse coding coefficients $\mathbf{s}_i$ and $\mathbf{s}_j$ for $\mathbf{y}_i$ and $\mathbf{y}_j$ should also be closer to each other. On the other hand, it is necessary to minimize the within-class variance and maximize the between-class separability for the sparse coding coefficients.

Given the training samples $\mathbf{y}_i$ and $\mathbf{y}_j$, the following within-class graph matrix E^{wit} and between-class graph matrix $\mathbf{E}_s^{bet}$ are defined as

$$e_{ij}^{\text{wit}} = \begin{cases} e(\mathbf{y}_i, \mathbf{y}_j), & if \ \mathbf{y}_i \in N_{wit}(\mathbf{y}_j) \ or \ \mathbf{y}_j \in N_{wit}(\mathbf{y}_i) \\ 0, & otherwise \end{cases} \tag{6}$$

$$e_{ij}^{bet} = \begin{cases} -e(\mathbf{y}_i, \mathbf{y}_j), & if \ \mathbf{y}_i \in N_{bet}(\mathbf{y}_j) \ or \ \mathbf{y}_j \in N_{bet}(\mathbf{y}_i) \\ 0, & otherwise \end{cases} \tag{7}$$

where $e(\mathbf{y}_i, \mathbf{y}_j) = \exp(- \| \mathbf{y}_i - \mathbf{y}_j \|^2 / \theta)$, θ is the tune parameter. $N_{wit}(\mathbf{y}_i)$ is the k-nearest neighbors of $\mathbf{y}_i$ belonging to the same class, and $N_{bet}(\mathbf{y}_i)$ is the k-nearest neighbors of $\mathbf{y}_i$ belonging to the different class.

Considering the within-class minimizing and between-class maximizing of sparse coding coefficients, I define the following discrimination information preserved term $F_1(\mathbf{S}_t)$ in the source domain,

$$\begin{aligned} F_1(\mathbf{S}_s) &= \min_{\mathbf{S}_s} \ \frac{1}{2}\sum_{i=1}^{n_s}\sum_{j=1}^{n_s} \parallel \mathbf{s}_{s,i} - \mathbf{s}_{s,j} \parallel_2^2 \ (\mathbf{E}_s^{wit} - \mathbf{E}_s^{bet}) \\ &= Tr(\mathbf{S}_s\mathbf{L}_s^{wit}\mathbf{S}_s^T) - Tr(\mathbf{S}_s\mathbf{L}_s^{bet}\mathbf{S}_s^T) \\ &= Tr(\mathbf{S}_s\mathbf{Q}_s\mathbf{S}_s^T), \end{aligned} \tag{8}$$

where $\mathbf{Q}_s = \mathbf{L}_s^{wit} - \mathbf{L}_s^{bet}$, $Tr(\cdot)$ is the trace operator.

Similarly, I can obtain the discrimination information preserved term $F_1(S_t)$ in the target domain,

$$\begin{aligned} F_1(\mathbf{S}_t) &= \min_{\mathbf{S}_t} \ \frac{1}{2}\sum_{i=1}^{n_t}\sum_{j=1}^{n_t} \parallel \mathbf{s}_{t,i} - \mathbf{s}_{t,j} \parallel_2^2 \ (\mathbf{E}_t^{wit} - \mathbf{E}_t^{bet}) \\ &= Tr(\mathbf{S}_t\mathbf{L}_t^{wit}\mathbf{S}_t^T) - Tr(\mathbf{S}_t\mathbf{L}_t^{bet}\mathbf{S}_t^T) \\ &= Tr(\mathbf{S}_t\mathbf{Q}_t\mathbf{S}_t^T), \end{aligned} \tag{9}$$

where $\mathbf{Q}_t = \mathbf{L}_t^{wit} - \mathbf{L}_t^{bet}$.

When reconstructing the projection space, the proposed transfer learning algorithm not only needs to establish potential connections between multiple domains in the projection space, but also transfers the domain-invariant information from the source domain to the target domain. Meanwhile, TLDDL should also retain the identification information of label samples and maintain the discriminative information in the projection space. To achieve this goal, I establish the principal component analysis (PCA) term of the projection matrix. Based on the PCA criterion, the discrimination information in the original space will be retained in the projection space. The PCA terms of the projection matrix on the source and target domains are represented as

$$F_2(\mathbf{P}_s) = \max_{\mathbf{P}_s} \ Tr(\mathbf{P}_s\mathbf{Y}_s\mathbf{Y}_s^T\mathbf{P}_s^T), \tag{10}$$

$$F_2(\mathbf{P}_t) = \max_{\mathbf{P}_t} \ Tr(\mathbf{P}_t\mathbf{Y}_t\mathbf{Y}_t^T\mathbf{P}_t^T), \tag{11}$$

where $\mathbf{P}_s$, $\mathbf{P}_t \in \mathbf{R}^{p\times d}$, p and d are the dimensions of the projection subspace and original feature space, respectively.

Based on the traditional dictionary learning algorithm, TLDDL combined the discriminative information preserved terms and PCA terms of the projection matrix together, and builds a domain-invariant dictionary between different domains. The objective function of TLDDL is

$$\begin{aligned} \min_{D,P_s,P_t,S_s,S_t} & \parallel P_sY_s^T - DS_s \parallel_2^2 + \parallel P_tY_t^T - DS_t \parallel_2^2 + \delta Tr(S_sQ_sS_s^T) \\ & + \delta Tr(S_tQ_tS_t^T) \\ & - \gamma Tr(P_sY_sY_s^TP_s^T) - \gamma Tr(P_tY_tY_t^TP_t^T) + \beta \parallel S_s \parallel_2^2 \\ & + \beta \parallel S_t \parallel_2^2, \\ s.t. \ & \parallel d_i \parallel_2^2 \leq 1, \\ & P_s^TP_s = I, \\ & P_t^TP_t = I, \end{aligned} \tag{12}$$

where δ, γ and β are trade off parameters.

For Equation 12, the first and second terms inherit the dictionary learning algorithm and are used to reconstruct the data in the source and target domains. These two terms are also domain-invariant dictionary learning terms to realize the connection of knowledge between different domains. The third and fourth terms are discriminative information preserved terms. The fifth and sixth terms are PCA terms of the projection matrices. The last two terms are the regularization terms of the sparse coding matrices.

Optimization

Let $\mathbf{P} = [\mathbf{P}_s, \mathbf{P}_t]$, $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_s^T & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_t^T \end{bmatrix}$, $\mathbf{S} = \begin{bmatrix} \mathbf{S}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_t \end{bmatrix}$, $\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_t \end{bmatrix}$, Equation 12 can be simplified as,

$$\begin{aligned} & \min_{\mathbf{D},\mathbf{P},\mathbf{S}} \parallel \mathbf{PY}^T - \mathbf{DS} \parallel_F^2 + Tr(\mathbf{S}(\delta\mathbf{Q} + \beta\mathbf{I})\mathbf{S}^T) - \gamma Tr(\mathbf{PYY}^T\mathbf{P}^T), \\ & s.t. \ \parallel \mathbf{d}_i \parallel_2^2 \leq 1, \ \forall i \\ & \quad \mathbf{P}^T\mathbf{P} = \mathbf{I}, \end{aligned} \tag{13}$$

The alternating iteration method is used to optimize variables {**P, S, D**}. In the optimization process, other variables are fixed and only one variable is optimized.

TABLE 1 | Accuracy results for each cross-domain task on Chinese corpus.

Tasks	K-SVD	ARTL	DMTTL	SMITL	WAAR	SFA	TLDDL
Sto→Cul	62.31	66.18	66.64	66.35	68.89	68.63	**70.64**
Sto→Ent	77.69	80.02	81.01	81.14	83.03	83.59	**85.58**
Cul→Ent	77.05	80.54	80.90	81.55	82.03	83.06	**85.05**
Cul→Edu	78.72	82.37	81.39	83.85	84.87	84.79	**86.76**
Ent→Spo	85.07	89.34	88.40	90.99	91.85	91.98	**93.89**
Ent→Edu	77.15	80.13	81.08	81.91	83.99	83.59	**85.51**
Spo→Wor	64.00	67.62	67.64	68.08	69.06	70.01	**72.01**
Spo→Gam	82.04	85.46	85.02	86.37	87.86	88.02	**90.03**
Fin→Car	85.08	88.59	88.59	89.94	90.15	91.12	**93.02**
Fin→Agr	78.33	81.89	82.08	81.76	83.22	84.40	**86.42**
Hou→Wor	63.11	67.29	67.67	67.15	69.13	69.72	**71.77**
Hou→Fin	58.04	60.29	61.13	62.08	63.71	63.93	**66.05**
Car→Tec	79.18	83.13	83.40	84.84	85.00	85.82	**87.87**
Car→Gam	82.28	85.78	86.82	87.61	87.33	88.22	**90.21**
Edu→Tec	79.02	82.14	82.89	83.78	84.04	85.34	**87.32**
Edu→Sto	61.29	65.54	65.91	65.00	67.70	67.45	**69.43**
Tec→Wor	63.30	67.32	67.56	68.43	69.47	69.82	**71.87**
Tec→Mil	78.07	81.65	81.09	82.97	83.28	84.10	**86.02**
Mil→Sto	60.19	64.39	64.85	65.04	66.93	66.85	**68.97**
Mil→Agr	75.10	79.58	79.28	80.94	81.16	82.05	**84.05**
Wor→Mil	78.60	81.21	82.92	83.97	84.67	84.67	**86.66**
Wor→Cul	63.06	66.65	66.12	67.65	68.60	69.00	**71.03**
Agr→Fin	58.29	62.91	62.06	63.74	64.18	64.76	**66.78**
Agr→Sto	60.29	64.39	64.18	65.98	66.10	66.87	**68.88**
Gam→Spo	85.05	89.56	89.93	89.10	91.72	91.43	**93.47**
Gam→Car	84.17	88.07	88.85	89.80	90.94	90.92	**92.79**
Mean	72.94	76.62	76.82	77.69	78.81	79.24	**81.23**

The best classification results are indicated in bold in Table.

Fixed{**S, D**}, the optimization problem of parameter **P** can be written as,

$$\min_{P} \parallel \mathbf{PY}^T - \mathbf{DS} \parallel_F^2 - \gamma Tr(\mathbf{PYY}^T\mathbf{P}^T), \quad s.t.\ \mathbf{PP}^T = \mathbf{I}. \tag{14}$$

Following the Proposition (2) in Shekhar et al. (2013), let

$$\mathbf{P} = (\mathbf{YW})^T, \tag{15}$$

$$\mathbf{D} = \mathbf{PYC}, \tag{16}$$

where $\mathbf{W} \in \mathbf{R}^{(n_s+n_t)\times p}$, $\mathbf{C} \in \mathbf{R}^{(n_s+n_t)\times K}$.

Equation 14 can be written as,

$$\min_{\mathbf{W}} \parallel \mathbf{W}^T\mathbf{M}(\mathbf{I} - \mathbf{CS}) \parallel_{\mathbf{F}}^2 - \gamma Tr(\mathbf{W}^T\mathbf{MM}^T\mathbf{W}), \quad s.t.\ \mathbf{WMW}^T = \mathbf{I}. \tag{17}$$

where $\mathbf{M} = \mathbf{Y}^T\mathbf{Y}$.

Then **W** has a closed-form solution as,

$$\mathbf{W} = \Theta\mathbf{V}^{-1/2}\mathbf{U}^T, \tag{18}$$

TABLE 2 | Precision results for each cross-domain task on Chinese corpus.

Tasks	K-SVD	ARTL	DMTTL	SMITL	WAAR	SFA	TLDDL
Stor→Cul	68.13	72.23	72.69	72.26	75.03	74.64	**76.68**
Stor→Ent	77.74	80.09	81.05	81.34	82.84	83.58	**85.54**
Cul→Ent	76.81	80.31	80.81	81.51	82.34	83.17	**85.37**
Cul→Edu	76.67	80.54	79.51	81.92	82.64	82.64	**84.44**
Ent→Spo	85.15	89.10	88.47	90.90	91.70	92.04	**94.10**
Ent→Edu	75.09	78.09	79.16	79.89	81.99	81.54	**83.41**
Spo→Wor	64.07	67.61	67.38	68.02	69.14	69.92	**71.96**
Spo→Gam	84.25	87.46	87.17	88.46	90.04	90.09	**92.35**
Fin→Car	85.29	88.53	88.48	90.06	90.15	91.01	**93.22**
Fin→Agr	78.34	81.98	82.04	81.87	83.25	84.66	**86.38**
Hou→Wor	63.03	67.21	67.41	67.21	69.11	69.79	**71.81**
Hou→Fin	57.99	60.35	61.15	62.14	63.57	63.92	**66.03**
Car→Tec	79.01	83.19	83.41	84.59	85.10	85.94	**87.70**
Car→Gam	84.34	87.67	88.94	89.78	89.48	90.36	**92.15**
Edu→Tec	78.95	82.21	82.98	83.75	84.01	85.56	**87.50**
Edu→Stor	59.32	63.46	63.67	62.76	65.85	65.47	**67.20**
Tec→Wor	63.17	67.02	67.58	68.28	69.52	70.01	**71.71**
Tec→Mil	81.03	84.71	84.17	85.94	86.34	87.14	**89.17**
Mil→Sto	58.34	62.22	62.95	63.25	65.15	64.84	**67.03**
Mil→Agr	75.16	79.26	78.96	80.94	80.95	82.03	**83.87**
Wor→Mil	81.47	84.09	85.92	87.06	87.61	87.65	**89.88**
Wor→Cul	62.96	66.71	66.30	67.53	68.87	68.95	**71.28**
Agr→Fin	58.43	62.96	61.95	63.67	64.40	64.94	**67.02**
Agr→Sto	60.14	64.39	64.14	65.99	65.94	66.74	**68.88**
Gam→Spo	85.05	89.45	89.71	89.01	91.83	91.61	**93.27**
Gam→Car	84.10	88.10	88.87	89.82	90.86	90.81	**92.96**
Mean	73.23	76.88	77.11	78.00	79.14	79.58	**81.57**

The best classification results are indicated in bold in Table.

where $\mathbf{M} = \Theta\mathbf{V}\Theta^T$, and **U** can be obtained as,

$$\min_{\mathbf{U}} Tr(\mathbf{U}^T\mathbf{H}^T U), \quad s.t.\ \mathbf{U}^T\mathbf{U} = \mathbf{I}, \tag{19}$$

where $\mathbf{H} = \mathbf{V}^{1/2}\Theta^T((\mathbf{I–CS})(\mathbf{I–CS})^T - \gamma\mathbf{I})\Theta\mathbf{V}^{1/2}$. Due to the orthonormality condition on **U**, Equation 19 has a closed-form solution. Therefore, **P** can then be updated by Equation 15.

Fixed{**P, S**}, the optimization problem of parameter **D** can be written as,

$$\min_{D} \parallel \mathbf{PY} - \mathbf{DS} \parallel_{\mathbf{F}}^2, \quad s.t.\ \parallel \mathbf{d}_i \parallel_2^2 \leq 1. \tag{20}$$

Using the Lagrange dual method, **D** can be obtained as,

$$\mathbf{D} = (\mathbf{PYS}^T)(\mathbf{SS}^T + \Delta)^{-1}, \tag{21}$$

where Δ is a diagonal matrix.

Fixed {**P, D**}, the optimization problem of parameter **S** can be written as,

$$\min_{S} \parallel \mathbf{PY}^T - \mathbf{DS} \parallel_{\mathbf{F}}^2 + Tr(\mathbf{S}(\delta\mathbf{Q} + \beta\mathbf{I})\mathbf{S}^T), \tag{22}$$

By setting the derivative of **S** to zero, I get,

$$\mathbf{D}^T\mathbf{DS} + \mathbf{S}(\delta\mathbf{Q} + \beta\mathbf{I}) = \mathbf{D}^T\mathbf{PY}. \tag{23}$$

In this study, Equation 23 is solved by the Bartels-Stewart method (Kleinman and Rao, 2003).

The above optimization steps are summarized in **Algorithm 1**.

ALGORITHM 1 | Algorithm 1 the TLDDL algorithm.

Input: Training data $\mathbf{Y}_s$ and $\mathbf{Y}_t$;
Output: dictionary **D**, projection matrix $\mathbf{P}_s$ and $\mathbf{P}_t$
Step 1. Initialize **D** and **S** using the K-SVD algorithm (Aharon et al., 2006);
Step 2. Compute the matrices $\mathbf{E}^{wit}$ and $\mathbf{E}_s^{bet}$ *via* Equation 6, 7;
Step 3. Learn **P** with fixed {**S**, **D**} *via* Equations 14–19;
Step 4. Learn **D** with fixed {**P**, **S**} *via* Equation 21;
Step 5. Learn **S** with fixed {**P**, **D**} *via* Equation 23;
Step 6. Go to Step 3 until the convergence or reaching the maximum number of iterations
Step 7. Output the dictionary **D**, projection matrix $\mathbf{P}_s$ and $\mathbf{P}_t$.

Test

For an unlabeled test text $\mathbf{y}^{test}$, according to the obtained optimal projection matrix $\mathbf{P}_t$ and dictionary **D**, its sparse coding coefficient $\mathbf{s}^*$ can be solved by the following problem,

$$\min_{\mathbf{s}^*} \parallel \mathbf{P}_t(\mathbf{y}^{test})^T - \mathbf{Ds}^* \parallel_2^2 + \beta \parallel \mathbf{s}^* \parallel_2^2 . \tag{24}$$

I can obtain $\mathbf{s}^*$ as follows,

$$\mathbf{s}^* = (\mathbf{DD}^T + \beta\mathbf{I})^{-1}\mathbf{D}^T\mathbf{P}_t. \tag{25}$$

Then, I compute the reconstruction error of $\mathbf{y}^{test}$ on the dictionary $D = [\mathbf{D}_1, \mathbf{D}_2, ..., \mathbf{D}_J]$, where J is the number of sample

classes. The reconstruction error of $\mathbf{y}^{test}$ on the jth sub dictionary $\mathbf{D}_j$ can be computed as follows,

$$\mathbf{r}_j(\mathbf{y}^{test}) = \| \mathbf{y}^{test} - \mathbf{D}_j\mathbf{s}_j^* \|_2^2, \quad (26)$$

where $\mathbf{s}_j^*$ is the sparse coding coefficient of $\mathbf{y}^{test}$ on the D_j. Finally, the class label of $\mathbf{y}^{test}$ is the class with the smallest reconstruction error, i.e.,

$$label(\mathbf{y}^{test}) = \min_j \mathbf{r}_j(\mathbf{y}^{test}) . \quad (27)$$

EXPERIMENT

Datasets and Experiment Setup

The experiment adopts the Chinese corpus (including toutiao-text and Chinese Weibo datasets) to verify the proposed TLDDL algorithm. I select 13 categories in Chinese corpus includes: Story (Sto), Culture (Cul), Entertainment (Ent), Sports (Spo), Finance (Fin), House (Hou), Car, Education (Edu), Technology (Tec), Military (Mil), World (Wor), Agriculture (Agr), and Game (Gam). Each category represents a domain. In addition, the experiment adopts the English multi-domain dataset Amazon review dataset, which is often used in cross-domain sentiment classification. Amazon review dataset includes reviews of DVD, book (Boo), electronic (Ele), kitchen and household appliance (Kit). Each product also represents a domain. To facilitate comparison with existing methods, 2,000 comments are selected respectively, including 1,000 positive comments and 1,000 negative comments. I use 80% texts in the source domain and 10% texts in the target domain as the training dataset, and the rest of the texts in the target domain are used for testing.

I investigate the performance of TLDDL compared with several algorithms on two text corpora. For comparison, the K-SVD algorithm (Aharon et al., 2006) is used as the baseline

TABLE 3 | Recall results for each cross-domain task on Chinese corpus.

Tasks	K-SVD	ARTL	DMTTL	SMITL	WAAR	SFA	TLDDL
Stor→Cul	76.10	79.95	80.70	80.53	82.62	82.69	**84.61**
Stor→Ent	76.63	79.09	80.00	80.12	82.04	82.56	**84.43**
Cul→Ent	75.99	79.63	79.75	80.56	80.74	81.87	**84.28**
Cul→Edu	77.62	81.38	80.41	82.59	83.92	83.91	**85.58**
Ent→Spo	83.88	88.55	87.18	89.76	90.74	91.10	**92.86**
Ent→Edu	76.28	79.21	80.30	80.84	82.79	82.62	**84.49**
Spo→Wor	61.09	64.63	64.43	65.19	66.20	66.85	**69.25**
Spo→Gam	77.01	80.38	80.06	81.25	83.13	82.95	**85.16**
Fin→Car	84.25	87.40	87.54	89.15	88.98	90.17	**92.06**
Fin→Agr	77.36	80.86	81.11	80.76	82.14	83.56	**85.46**
Hou→Wor	62.23	66.47	66.83	66.16	68.21	68.91	**70.53**
Hou→Fin	57.76	60.59	61.32	62.32	63.70	64.13	**66.11**
Car→Tec	77.91	82.43	82.39	83.81	83.97	84.95	**86.62**
Car→Gam	77.01	80.72	81.74	82.68	82.43	82.99	**85.04**
Edu→Tec	77.97	81.22	81.87	82.57	82.90	84.04	**86.39**
Edu→Stor	63.28	67.53	67.81	67.26	69.59	69.59	**71.54**
Tec→Wor	60.30	64.38	64.74	65.32	66.53	66.89	**68.70**
Tec→Mil	74.04	77.75	77.00	79.16	79.47	80.37	**81.97**
Mil→Sto	61.94	66.26	66.68	67.17	68.95	68.81	**71.18**
Mil→Agr	74.02	78.74	78.24	79.96	80.04	81.32	**82.78**
Wor→Mil	74.45	77.42	79.09	80.00	80.47	80.61	**82.75**
Wor→Cul	77.13	80.56	80.10	81.83	82.70	83.04	**85.02**
Agr→Fin	58.52	63.11	61.89	63.92	64.17	64.64	**66.82**
Agr→Sto	59.41	63.66	63.29	64.99	65.33	65.84	**67.77**
Gam→Spo	83.90	88.56	88.85	87.92	90.78	90.26	**92.36**
Gam→Car	83.24	87.01	87.65	88.86	90.07	89.83	**91.87**
Mean	72.67	76.44	76.58	77.49	78.56	79.02	**80.99**

The best classification results are indicated in bold in Table.

TABLE 4 | F1-score results for each cross-domain task on Chinese corpus.

Tasks	K-SVD	ARTL	DMTTL	SMITL	WAAR	SFA	TLDDL
Stor→Cul	72.50	75.91	76.44	76.41	79.09	78.62	**80.55**
Stor→Ent	77.41	79.44	80.54	80.62	82.76	83.06	**84.76**
Cul→Ent	76.23	79.90	80.44	80.93	82.03	82.48	**84.70**
Cul→Edu	77.26	80.97	80.19	82.34	83.26	83.34	**84.66**
Ent→Spo	84.53	88.76	87.83	89.94	91.34	91.97	**93.48**
Ent→Edu	75.83	78.98	79.86	80.40	82.17	81.77	**84.22**
Spo→Wor	62.63	66.18	65.75	66.31	67.89	68.52	**70.85**
Spo→Gam	80.71	83.54	83.34	84.83	86.72	86.40	**88.45**
Fin→Car	84.83	87.72	87.91	89.61	89.69	90.53	**92.68**
Fin→Agr	78.08	81.22	81.50	81.19	82.59	84.54	**86.02**
Hou→Wor	62.93	67.18	66.99	66.93	68.53	69.51	**71.01**
Hou→Fin	57.92	60.16	61.20	62.57	63.61	63.70	**66.05**
Car→Tec	78.47	83.20	83.05	84.10	84.37	85.68	**86.75**
Car→Gam	80.37	84.15	85.17	86.03	86.00	86.85	**88.81**
Edu→Tec	78.88	81.77	82.27	82.99	83.17	84.90	**86.93**
Edu→Stor	61.08	65.37	65.34	64.68	67.43	67.60	**69.38**
Tec→Wor	61.85	65.72	65.87	66.72	68.24	68.40	**70.33**
Tec→Mil	77.28	81.46	80.43	82.39	82.87	83.94	**85.73**
Mil→Sto	60.14	64.29	64.76	64.87	66.86	66.75	**69.22**
Mil→Agr	74.49	78.99	78.66	80.76	80.56	81.64	**83.01**
Wor→Mil	77.81	80.48	82.72	83.75	83.89	84.24	**86.41**
Wor→Cul	69.84	73.58	72.86	74.60	75.94	75.60	**78.25**
Agr→Fin	58.47	63.05	61.76	63.93	64.05	64.87	**67.03**
Agr→Sto	59.67	64.21	63.51	65.59	65.88	66.34	**68.44**
Gam→Spo	84.18	89.08	88.98	88.52	91.51	90.65	**93.07**
Gam→Car	83.66	87.58	88.21	89.37	90.50	90.26	**92.22**
Mean	72.96	76.65	76.75	77.71	78.88	79.31	**81.27**

The best classification results are indicated in bold in Table.

TABLE 5 | Accuracy results for each cross-domain task on English corpus.

Tasks	K-SVD	ARTL	DMTTL	SMITL	WAAR	SFA	TLDDL
DVD→Boo	74.44	76.76	77.18	78.00	79.95	81.05	**82.98**
DVD→Ele	70.28	72.44	73.30	73.68	74.65	76.01	**76.78**
DVD→Kit	71.09	73.70	73.79	74.92	75.03	75.67	**76.09**
Boo→DVD	76.32	79.02	79.20	81.02	82.01	82.94	**84.74**
Boo→Ele	71.32	73.41	74.12	75.52	76.29	77.28	**78.07**
Boo→Kit	73.64	76.36	76.64	78.13	77.43	77.67	**78.12**
Ele→DVD	73.04	75.78	75.96	77.56	79.02	80.54	**81.63**
Ele→Boo	70.88	73.88	73.34	75.68	76.67	78.66	**80.33**
Ele→Kit	83.23	85.61	86.32	87.71	85.39	86.37	**87.54**
Kit→DVD	74.73	77.44	77.04	78.77	79.27	80.14	**80.78**
Kit→Boo	82.07	84.53	85.11	86.03	86.02	86.90	**87.28**
Kit→Ele	73.65	75.95	76.59	77.57	78.88	79.67	**80.76**
Mean	74.56	77.07	77.38	78.72	79.22	80.24	**81.26**

The best classification results are indicated in bold in Table.

TABLE 6 | Precision results for each cross-domain task on English corpus.

Tasks	K-SVD	ARTL	DMTTL	SMITL	WAAR	SFA	TLDDL
DVD→Boo	74.75	77.65	77.45	78.57	80.26	81.36	**83.25**
DVD→Ele	71.03	73.02	73.82	73.96	75.28	76.16	**77.12**
DVD→Kit	71.85	73.84	73.89	75.13	75.68	76.32	**76.43**
Boo→DVD	76.51	79.88	79.45	81.02	82.76	83.71	**85.42**
Boo→Ele	72.13	73.93	74.75	75.58	77.15	77.49	**78.40**
Boo→Kit	73.96	76.83	77.02	78.26	77.82	78.57	**78.61**
Ele→DVD	73.15	75.80	76.16	77.77	79.85	81.16	**81.80**
Ele→Boo	71.60	74.16	73.75	75.98	77.42	78.97	**80.90**
Ele→Kit	83.78	85.68	86.84	88.21	85.73	87.06	**88.39**
Kit→DVD	75.53	77.81	77.28	79.26	79.39	80.85	**81.30**
Kit→Boo	82.58	84.90	85.62	86.77	86.54	87.26	**87.67**
Kit→Ele	73.80	76.73	77.05	78.31	79.60	79.86	**80.77**
Mean	75.06	77.52	77.76	79.07	79.79	80.73	**81.67**

The best classification results are indicated in bold in Table.

TABLE 7 | Recall results for each cross-domain task on English corpus.

Tasks	K-SVD	ARTL	DMTTL	SMITL	WAAR	SFA	TLDDL
DVD→Boo	73.86	76.23	77.14	77.30	79.93	80.44	**82.69**
DVD→Ele	69.48	71.69	72.82	73.51	73.98	75.38	**76.14**
DVD→Kit	70.87	72.95	73.70	74.20	74.43	75.03	**76.04**
Boo→DVD	75.78	78.39	78.54	80.40	81.75	82.20	**84.05**
Boo→Ele	71.18	73.21	73.93	74.77	76.05	77.06	**77.57**
Boo→Kit	73.43	75.61	76.07	78.12	76.68	77.36	**77.39**
Ele→DVD	72.35	75.77	75.81	77.02	78.35	80.02	**81.06**
Ele→Boo	70.87	73.36	73.30	75.05	76.17	77.97	**79.97**
Ele→Kit	82.91	85.48	85.55	87.35	84.78	86.15	**87.48**
Kit→DVD	73.98	76.96	76.55	78.70	78.94	79.41	**80.76**
Kit→Boo	82.03	83.78	85.08	85.29	85.85	86.78	**86.68**
Kit→Ele	73.63	75.50	76.07	77.35	78.09	79.66	**80.64**
Mean	74.20	76.58	77.05	78.26	78.75	79.79	**80.87**

The best classification results are indicated in bold in Table.

TABLE 8 | F1-score results for each cross-domain task on English corpus.

Tasks	K-SVD	ARTL	DMTTL	SMITL	WAAR	SFA	TLDDL
DVD→Boo	74.08	76.90	77.49	78.03	80.13	80.85	**82.89**
DVD→Ele	70.08	72.18	73.25	73.96	74.70	75.72	**76.58**
DVD→Kit	71.54	73.13	73.66	75.07	74.81	75.71	**76.56**
Boo→DVD	76.05	79.53	78.79	80.89	82.42	82.84	**84.83**
Boo→Ele	71.25	73.94	74.43	75.42	76.43	77.37	**77.97**
Boo→Kit	73.69	75.84	76.58	78.37	77.18	77.68	**77.98**
Ele→DVD	72.62	75.97	75.92	77.24	79.11	80.51	**81.40**
Ele→Boo	71.16	73.39	73.77	75.17	76.50	78.57	**79.96**
Ele→Kit	83.05	85.39	86.34	87.63	84.83	86.56	**88.09**
Kit→DVD	74.77	77.37	76.77	78.98	79.37	80.09	**80.96**
Kit→Boo	81.86	84.19	85.30	86.36	86.36	87.12	**87.29**
Kit→Ele	74.06	76.26	76.83	77.50	79.01	79.94	**80.50**
Mean	74.52	77.01	77.43	78.72	79.24	80.25	**81.25**

The best classification results are indicated in bold in Table.

of the proposed algorithm. The size of dictionary in K-SVD is set to 300. In addition, six transfer learning algorithms are used, including: ARTL (Long et al., 2014), DMTTL (Zheng et al., 2019), SMITL (Liu et al., 2018), WAAR (Jia et al., 2018), and SFA (Pan et al., 2010). In ARTL, the parameter λ is set in the grid $\{10^{-2}, 10^{-1}, ..., 10^{2}\}$, the parameter γ is set in the grid {0.01, 0.05, 0.1, 0.5, 1.5, 10}, and the parameter σ is set in the grid {0.01, 0.02, 0.1, 0.5, 1, 2, 5}. In DMTTL, the parameters λ, λ_s, λ_t are set in the grid $\{10^{-4}, 10^{-3}, ..., 10^{3}\}$, γ is set in the grid $\{10^{2}, 5 \times 10^{2}, ..., 2 \times 10^{4}\}$. In SMITL, the Gaussian kernel is used, and the kernel and penalty parameters are set in the grid $\{10^{-4}, 10^{-3}, ..., 10^{3}\}$. In WAAR, the parameter l is set to 600, min-support = 0.014, min-confidence = 0.08, and ε = 0.005. In SFA, the parameters l, k and γ are set to 500, 100 and 0.6, respectively. In the proposed TLDDL, the subspace dimension and the size of dictionary are set to 500 and 300, respectively. The parameters δ_s, δ_t, γ_s, γ_t, β_s and $\beta_t$are set in the grid {0.01, 0.05,..., 2}. In the experiments, the dimension of the text word vectors is set to 300. I take the first 100 text units in each text. Following (Meng et al., 2019), a convolutional neural network composed of five layers is used to extract the text features. The mini-batch size is set to 16, and the number of convolution kernels is set to 256. Finally, these vectors are set to 768 dimensions when I treat them into the training model. All the experiments are executed in Matlab 2018a environment. I repeat the experiments five times. The performances of experimental results are generally evaluated based on accuracy, precision, recall, and F1-score.

Experiments on Chinese Corpus

Tables 1–4 show the classification results on Chinese corpus based on accuracy, precision, recall, and F1-score, respectively. For example, task "Stor→Cul" indicates that the source domain is "Story" and the target domain is "Culture." In the respect of classification accuracy, the best average accuracy obtained by

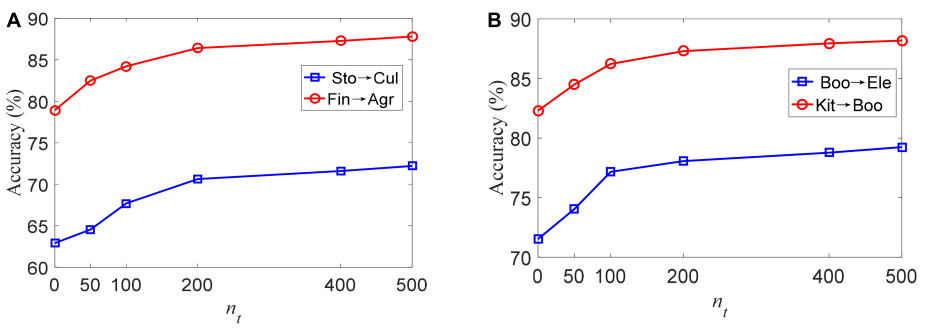


FIGURE 1 | The accuracy of TLDDL with different size of n_t, **(A)** Chinese corpus; **(B)** English corpus.

TLDDL is 81.23%, followed by SFA, and its average accuracy is 79.24%. The non-transfer learning algorithm K-SVD is the lowest. In the respect of precision, recall, and F1-score, all transfer learning algorithms are always better than the non-transfer learning algorithm K-SVD in each classification task. The proposed TLDDL algorithm also obtains the best performance. Compared with the second-best algorithm, the average precision, recall, and F1 score exceed 1.99, 1.97, and 1.96%, respectively. Therefore, the generalization ability of the proposed TLDDL algorithm is higher. The experimental results indicate that under the framework of the dictionary learning algorithm, TLDDL uses the projection technology to reduce the differences between different domains, and builds a domain-invariant dictionary to establish a bridge between the related domains. The transferring discriminative information of the source domain to the target domain can improve the cross-domain classification performance.

Experiments on English Corpus

Tables 5–8 show the classification results on the English corpus based on accuracy, precision, recall, and F1-score, respectively. I can see that all transfer learning algorithms outperform the baseline algorithm K-SVD. Some tasks have high classification performance, such as DVD→Boo and Boo→DVD. The reason for the high classification performance is that the differences between the source domain and the target domain are similar. Some tasks have low classification performance, such as DVD→Ele and DVD→Kit. The reason is that the domain-invariant information transferring from the source domain to the target domain is insufficient due to the great differences between domains. The results in **Tables 5–8** show that the TLDDL algorithm obtains the best performance. TLDDL obtains the classification performance 81.26, 81.67, 80.87, and 81.25% in the respect of accuracy, precision, recall and F1-score, respectively. TLDDL is 1.02, 0.94, 0.98, and 1.00% higher than the second-best algorithm in four evaluation indexes. The results indicate that joint learning of projection technology and dictionary learning is an efficient strategy in cross-domain text classification tasks. In addition, the discrimination information preserved and PCA terms are helping to improve the generalization of the classifier.

Experiments With Different Training Samples in the Target Domain

To show the influence of the number of training samples in the target domain on the proposed TLDDL algorithm, I compare its classification accuracy on the Chinese and English corpora in **Figure 1**. The size of n_t is set to be 0, 50, 100, 200, 400, and 500, respectively. Since the size of n_t should be less than that of n_s, the maximum size of n_t is set to 500. The results of Sto→Cul and Fin→Agr on the Chinese corpus are shown in **Figure 1A**. When the size of n_t is 0, the classification accuracy of the two tasks is the lowest. With the increase of the size of n_t, the classification accuracy is gradually improved. When the size of n_t exceeds 200, the classification accuracy reaches a relatively stable state. The results of Boo→Ele and Kit→Boo on the English corpus are shown in **Figure 1B**. I can see that the classification accuracy of TLDDL increases rapidly after adding the training samples in n_t. When n_t reaches 200, the classification accuracy is stable and the growth rate of accuracy is small. Considering the practical application scenarios of transfer learning, 200 texts selected in the target domain for training are reasonable.

CONCLUSION

In this study, I have developed a transfer learning classification algorithm for cross-domain text sentiment classification. Inspired by the advantage of dictionary learning in knowledge reconstruction and sparse representation, I proposed to employ subspace projection and transfer learning into the framework of dictionary learning. Considering the within-class minimizing and between-class maximizing of sparse coding coefficients, I define the following discrimination information preserved term in the objective function; meanwhile, I adopt the PCA term in the objective function to retain the discrimination knowledge. In such an algorithm, a domain-invariant dictionary is built to establish a connection between different domains. Experimental results indicate that the TLDDL algorithm achieves good classification performance. In the future, from the perspective of multiple learning strategies on dictionary learning, I will consider how to realize transfer learning of multi-source and heterogeneous data in the proposed algorithm. In addition, from the perspective of extracting features, I will investigate how to automatically implement feature selection and building a classifier in a model framework.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The toutiao-text dataset is addressed as: https://github.com/aceimnorstuvwxz/toutiao-multilevel-text-classfication-dataset. The Chinese Weibo dataset is addressed as: http://tcci.ccf.org.cn/conference/2014/pages/page04_sam.html. The Amazon review dataset is addressed as: http://jmcauley.ucsd.edu/data/amazon/.

AUTHOR CONTRIBUTIONS

SZ conceived and developed the theoretical framework of the manuscript, carried out experiment and data process, and drafted the manuscript.

REFERENCES

Aharon, M., Elad, M., and Bruckstein, A. (2006). K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* 54, 4311–4322. doi: 10.1109/TSP.2006.881199

Ahmad, M., Aftab, S., and Ali, I. (2017). Sentiment analysis of tweets using SVM. *Int. J. Comput. Appl.* 177, 975–8887. doi: 10.5120/ijca2017915758

Ahuja, R., Chug, A., Kohli, S., Gupta, S., and Ahuja, P. (2019). The impact of features extraction on the sentiment analysis. *Procedia Comput. Sci.* 152, 341–348. doi: 10.1016/j.procs.2019.05.008

Almunirawi, K. M., and Maghari, A. (2016). A comparative study on serial decision tree classification algorithms in text mining. *Int. J. Intell. Comput. Res.* 7, 754–760. doi: 10.20533/ijicr.2042.4655.2016.0093

Alshamsi, A., Bayari, R., and Salloum, S. (2020). Sentiment analysis in English texts. *Adv. Sci. Technol. Eng. Syst. J.* 5, 1683–1689. doi: 10.25046/aj0506200

Bai, H., Yu, H., Yu, G., and Huang, X. (2020). A novel emergency situation awareness machine learning approach to assess flood disaster risk based on Chinese Weibo. *Neural Comput. Appl.* doi: 10.1007/s00521-020-05487-1 [Epub ahead of print].

Birjali, M., Kasri, M., and Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: approaches, challenges and trends. *Knowl. Based Syst.* 226:107134. doi: 10.1016/j.knosys.2021.107134

Bozkurt, F., Çoban, O., Gunay, F., and Yucel Altay, S. (2019). High performance twitter sentiment analysis using CUDA based distance kernel on GPUs. *Tehnicki Vjesnik-Technical Gazette* 26, 1218–1227.

Chaturvedi, I., Satapathy, R., Cavallari, S., and Cambria, E. (2019). Fuzzy commonsense reasoning for multimodal sentiment analysis. *Pattern Recognit. Lett.* 125, 264–270. doi: 10.1016/j.patrec.2019.04.024

Chen, D., Xiong, Y., Yan, J., Xue, G. R., Wang, G., and Zheng, C. (2010). Knowledge transfer for cross domain learning to rank. *Inf. Retrieval* 13, 236–253.

Chen, L. C., Lee, C. M., and Chen, M. Y. (2020). Exploration of social media for sentiment analysis using deep learning. *Soft Comput.* 24, 8187–8197. doi: 10.1007/s00500-019-04402-8

Du, Y., He, M., Wang, L., and Zhang, H. (2020). Wasserstein based transfer network for cross-domain sentiment classification. *Knowl. Based Syst.* 204:106162. doi: 10.1016/j.knosys.2020.106162

Fei, R., Yao, Q., Zhu, Y., Xu, Q., and Hu, B. (2020). Deep learning structure for cross-domain sentiment classification based on improved cross entropy and weight. *Sci. Program.* 2020:3810261. doi: 10.1155/2020/3810261

Fu, Y., and Liu, Y. (2021). Cross-domain sentiment classification based on key pivot and non-pivot extraction. *Knowl. Based Syst.* 228:107280.

Gu, X., Fan, Y., Zhou, J., and Zhu, J. (2021). Optimized projection and fisher discriminative dictionary learning for EEG emotion recognition. *Front. Psychol.* 12:705528. doi: 10.3389/fpsyg.2021.705528

Gu, X., Lu, L., Qiu, S., Zou, Q., and Yang, Z. (2020). Sentiment key frame extraction in user-generated micro-videos via low-rank and sparse representation. *Neurocomputing* 410, 441–453. doi: 10.1016/j.neucom.2020.05.026

Gui, L., Lu, Q., Xu, R., Wei, Q., and Cao, Y. (2015). "Improving transfer learning in cross lingual opinion analysis through negative transfer detection," in *International Conference on Knowledge Science, Engineering and Management*, (China: Springer). doi: 10.1007/978-3-319-25159-2_36

Jia, X., Jin, Y., Li, N., and Su, X. (2018). Words alignment based on association rules for cross-domain sentiment classification. *Front. Inf. Technol. Electronic.Eng.* 19:260–272. doi: 10.1631/FITEE.1601679

Jiang, Y., Gu, X., Ji, D., and Jian, P. (2020). Smart diagnosis: a multiple-source transfer TSK fuzzy system for EEG seizure identification. *ACM Trans. Multimed. Comput. Commun. Appl.* 59, 1–21. doi: 10.1145/3340240

Jiang, Y., Gu, X., Wu, D., Hang, W., Xue, J., Qiu, S., et al. (2021). A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 40–52. doi: 10.1109/TCBB.2019.2963873

Kleinman, D. L., and Rao, P. K. (2003). Extensions to the Bartels-Stewart algorithm for linear matrix equations. *IEEE Trans. Autom. Control* 23, 85–87. doi: 10.1109/TAC.1978.1101681

Lauren, P., Qu, G., Yang, J., Watta, P., Huang, G. B., and Lendasse, A. (2018). Generating word embeddings from an extreme learning machine for sentiment analysis and sequence labeling tasks. *Cognit. Comput.* 10, 625–638. doi: 10.1007/s12559-018-9548-y

Liu, B., Xiao, Y., and Hao, Z. (2018). A selective multiple instance transfer learning method for text categorization problems. *Knowl. Based Syst.* 141, 178–187. doi: 10.1016/j.knosys.2017.11.019

Long, M., Wang, J., Ding, G., Pan, S. J., and Yu, P. S. (2014). Adaptation regularization: a general framework for transfer learning. *IEEE Trans. Knowl. Data Eng.* 26, 1076–1089. doi: 10.1109/TKDE.2013.111

López, M., Valdivia, A., Martínez-Cámara, E., Luzón, M. V., and Herrera, F. (2019). E2sam: evolutionary ensemble of sentiment analysis methods for domain adaptation. *Inf. Sci.* 480, 273–286. doi: 10.1016/j.ins.2018.12.038

Meng, J., Long, Y., Yu, Y., Zhao, D., and Liu, S. (2019). Cross-domain text sentiment analysis based on CNN_FT method. *Information* 10:162.

Ni, T., Ni, Y., Xue, J., and Wang, S. (2021). A domain adaptation sparse representation classifier for cross-domain electroencephalogram-based emotion classification. *Front. Psychol.* 12:721266. doi: 10.3389/fpsyg.2021.721266

Pan, S. J., Ni, X., Sun, J. T., Yang, Q., and Chen, Z. (2010). "Cross-domain sentiment classification via spectral feature alignment," in *Proceedings of the 19th International Conference on World Wide Web*, (Carolina: ACM Digital Library), 26–30. doi: 10.1145/1772690.1772767

Shekhar, S., Patel, V. M., Nguyen, H. V., and Chellappa, R. (2013). "Generalized domain-adaptive dictionaries," in *IEEE Conference on Computer Vision and Pattern Recognition*, (Portland: IEEE), 361–368. doi: 10.1109/CVPR.2013.53

Tang, H., Mi, Y., Xue, F., and Cao, Y. (2021). Graph Domain Adversarial Transfer Network for Cross-Domain Sentiment Classification. *IEEE Access* 9, 33051–33060. doi: 10.1109/ACCESS.2021.3061139

Tian, F., Wu, F., Fei, X., Shah, N., and Wang, Y. (2019). Improving generalization ability of instance transfer-based imbalanced sentiment classification of turn-level interactive Chinese texts. *Serv. Oriented Comput. Appl.* 13, 155–167. doi: 10.1007/s11761-019-00264-y

Unnikrishnan, P., Govindan, V. K., and Madhu Kumar, S. (2019). Enhanced sparse representation classifier for text classification. *Expert Syst. Appl.* 129, 260–272. doi: 10.1016/j.eswa.2019.04.003

Waheeb, S. A., Ahmed Khan, N., Chen, B., and Shang, X. (2020). Machine Learning Based Sentiment Text Classification for Evaluating Treatment Quality of Discharge Summary. *Information* 11:281. doi: 10.3390/info11050281

Wang, L., Niu, J., Song, H., and Atiquzzaman, M. (2018). SentiRelated: a cross-domain sentiment classification algorithm for short texts through sentiment related index. *J. Netw. Comput. Appl.* 101, 111–119. doi: 10.1016/j.jnca.2017.11.001

Wen, J., Zhang, G., Zhang, H., Yin, W., and Ma, J. (2020). Speculative text mining for document-level sentiment classification. *Neurocomputing* 412, 52–62. doi: 10.1016/j.neucom.2020.06.024

Wu, Q., and Tan, S. (2011). A two-stage framework for cross-domain sentiment classification. *Expert Syst. Appl.* 38, 14269–14275. doi: 10.1016/j.eswa.2011.04.240

Xia, R., Zong, C., Hu, X., and Cambria, E. (2013). Feature ensemble plus sample selection: domain adaptation for sentiment classification. *IEEE Intell. Syst.* 28, 10–18. doi: 10.1109/MIS.2013.27

Zhang, H., Sun, S., Hu, Y., Liu, J., and Guo, Y. (2020). Sentiment classification for Chinese text based on interactive multitask learning. *IEEE Access* 8, 129626–129635. doi: 10.1109/ACCESS.2020.3007889

Zhang, K., Hei, X., Fei, R., Guo, Y., and Jiao, R. (2021). "Cross-domain text classification based on BERT model," in *International Conference on Database Systems for Advanced Applications*, (Taipei: Springer), doi: 10.1007/978-3-030-73216-5_14

Zhang, S., Wei, Z., Yin, W., and Tao, L. (2018). Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary. *Future Gener. Comput. Syst.* 81, 395–403. doi: 10.1016/j.future.2017.09.048

Zhao, C., Wang, S., Li, D., Liu, X., Yang, X., and Liu, J. (2021). Cross-domain sentiment classification via parameter transferring and attention sharing mechanism. *Inf. Sci.* 578, 281–296. doi: 10.1016/j.ins.2021.07.001

Zheng, X., Lin, L., Liu, B., Xiao, Y., and Xiong, X. (2019). A multi-task transfer learning method with dictionary learning. *Knowl. Based Syst.* 191:105233.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 28 September 2021
doi: 10.3389/fpsyg.2021.760060

A Novel Music Emotion Recognition Model Using Neural Network Technology

*Jing Yang**

Zhejiang Conservatory of Music, Hangzhou, China

Music plays an extremely important role in people's production and life. The amount of music is growing rapidly. At the same time, the demand for music organization, classification, and retrieval is also increasing. Paying more attention to the emotional expression of creators and the psychological characteristics of music are also indispensable personalized needs of users. The existing music emotion recognition (MER) methods have the following two challenges. First, the emotional color conveyed by the first music is constantly changing with the playback of the music, and it is difficult to accurately express the ups and downs of music emotion based on the analysis of the entire music. Second, it is difficult to analyze music emotions based on the pitch, length, and intensity of the notes, which can hardly reflect the soul and connotation of music. In this paper, an improved back propagation (BP) algorithm neural network is used to analyze music data. Because the traditional BP network tends to fall into local solutions, the selection of initial weights and thresholds directly affects the training effect. This paper introduces artificial bee colony (ABC) algorithm to improve the structure of BP neural network. The output value of the ABC algorithm is used as the weight and threshold of the BP neural network. The ABC algorithm is responsible for adjusting the weights and thresholds, and feeds back the optimal weights and thresholds to the BP neural network system. BP neural network with ABC algorithm can improve the global search ability of the BP network, while reducing the probability of the BP network falling into the local optimal solution, and the convergence speed is faster. Through experiments on public music data sets, the experimental results show that compared with other comparative models, the MER method used in this paper has better recognition effect and faster recognition speed.

Keywords: emotion recognition, music, BP neural network, ABC algorithm, MediaEval Emotion in Music data set

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Weiwei Cai,
Northern Arizona University, United States
Runmin Liu,
Wuhan Sports University, China

***Correspondence:**
Jing Yang
yangjing@zjcm.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *17 August 2021*
Accepted: *06 September 2021*
Published: *28 September 2021*

Citation:
Yang J (2021) A Novel Music Emotion Recognition Model Using Neural Network Technology. Front. Psychol. 12:760060. doi: 10.3389/fpsyg.2021.760060

INTRODUCTION

Music appeared earlier than language, and human beings are born with music to express feelings. Music plays an important role in the history of mankind, and music is integrated into all aspects of human life. With the development of science and technology, the creation, storage, and dissemination of music have all undergone tremendous changes. The changes in creation, storage, dissemination, and technology have made music play an increasingly important role in human social life. Stores often play dynamic music to arouse customers' desire to buy. Playing rhythmic

music during exercise can increase the enthusiasm of exercise and reduce exercise fatigue. Playing soft music during a break can soothe the nerves. Exciting and rhythmic music is played during the party, which can set off the joyful atmosphere of the party. During the medical process, appropriate music will be selected to provide psychological counseling to the patient. Music has been integrated into all aspects of life including waking up, eating, shopping, learning, sports, driving, education, medical care and so on. With the rapid increase of music file data, how to use computers to complete fast and effective music information retrieval has become a basic demand of people in the current society. Traditional music retrieval methods are based on classification tags, such as retrieving songs by song name, artist name, and album name. This method is currently the most commonly used method. The traditional methods of retrieving music information can no longer meet people's needs for intelligent and personalized music retrieval (Casey et al., 2008). For these needs, many music websites have also launched music recommendation services. Recommend similar songs in the music library according to the needs of users and the songs that users often listen to. In order to obtain a better user experience, in recent years, various listening platforms have begun to provide song recommendation services with different moods. There are relatively few intelligent classification of music emotions and emotion-based search services. Therefore, emotion-based music retrieval is an important part of meeting people's personalized music retrieval needs, and it is also an important development direction of current music retrieval.

To achieve emotion-based music classification and retrieval, it is necessary to label music works with emotions. Emotional annotation of massive music works based on artificial methods is not only a huge workload, but also the quality cannot be guaranteed. Therefore, it is inevitable to study music emotion automatic recognition technology and realize automatic emotion labeling of music works. Music emotion recognition (MER) refers to the construction of a calculation model based on music audio data and other related information to realize automatic music discrimination (Vidas et al., 2020). Agarwal and Om (2021) constructed a MER system to estimate the evaluation value of each musical work, and used regression methods to detect emotional changes in music. Lu et al. (2006) proposed a hierarchical MER system. The system uses Gaussian mixture model (GMM) and Bayesian classifier to classify music emotions. The identification object is popular music. This research is the first time that the GMM has been applied to music sentiment classification. Torres et al. (2007) identifies music emotions through music lyrics information. Sharaj et al. (2019) identified music emotions by analyzing auxiliary information such as lyrics in the music. The research also compares the recognition effects of three widely used recognition models: support vector machine (SVM), K-Nearest Neighbor Method (KNN), and GMM. Liu et al. (2006) uses the Psy Sound2 system they developed to extract music features from music waveform files. The classification algorithm is used to simulate the feature classifier and evaluate each feature to improve the reliability and robustness of the system, thereby obtaining music emotion. Wu and Jeng (2008) used the probability distribution model to measure the similarity between the tested music and the sample music based on the Hevner emotion model, so as to estimate the emotion type of the tested music. Anders (2011) expressed various common emotions by controlling the combination of multiple feature quantities such as the rhythm and tone of music. The relationship between feature quantity and music emotion was given by multiple music experts through manual annotation. Shan et al. (2009) applied MER to film music, and its goal was to judge the type of film based on the type of music emotion. Yang and Chen (2011) studied a music ranking algorithm based on a two-dimensional emotional model. Lin et al. (2011) uses online genre-labeled vocabulary for sentiment classification. Sentiment classification is performed after genre classification. Zhang and Sun (2012) conducts emotion recognition on web music and classifies music into a certain type of emotional label. The above studies are based on machine learning algorithms (Jiang et al., 2019, 2021), and the recognition effect can reach more than 60%.

In this study, back propagation (BP) neural network (Rumelhart et al., 1986) is selected as the basic recognition model. Aiming at the problems of BP network which are easy to fall into local minima, slow convergence speed and poor generalization ability, this paper uses the final value obtained by artificial bee colony (ABC) algorithm iteration as the weight and threshold of BP network. The optimized BP network greatly improves its global search ability, while reducing the probability of falling into a local optimal solution, and the convergence speed is faster. The optimized BP network has better music recognition capabilities. The main contributions of this article are:

(1) After comparative analysis, the BP neural network in machine learning is selected as the basic model of MER. After the BP neural network selects a large amount of random data to pass the characteristic function, it can backpropagate the output result with larger error. The characteristic function is sent through continuous feedback, and the cyclic operation can be performed again. After a large number of cyclic operations, the ideal value is output. This is the adaptive learning process of BP neural network. The BP neural network has become the most widely used artificial neural network model due to its strong non-linear mapping ability, high fault tolerance, self-adaptation and self-learning performance.
(2) Aiming at the problems of BP network that are easy to fall into local minima, slow convergence speed, and poor generalization ability, this paper uses the output obtained by the ABC algorithm as the weight and threshold of the BP network. This optimization method greatly improves the global search capability of the BP network, while reducing the probability of falling into a local optimal solution, and the network converges faster.
(3) To verify the effectiveness of the network used, based on the public music data set, four methods of SVM, KNN, GMM, and BP are used for experimental comparison. The experimental results show that the used network is better than other comparison models in terms of recognition rate and generalization.

THEORETICAL KNOWLEDGE OF MUSIC EMOTION RECOGNITION

Music Emotion Recognition Process

The process of MER is shown in **Figure 1**. Almost all existing MER algorithms are based on supervised learning, so it is necessary to first establish a learning library, that is, a music data set with emotion annotations. Then extract the features of the music to form a feature vector, use the feature vector and emotion annotations for training, and get the recognition model. Then perform feature extraction on the unknown music, and input the extracted feature vector into the trained recognition model for recognition, so as to obtain the emotion classification result of the unknown music.

Musical Emotion Model

The premise of automatic recognition of music emotion is to establish a reasonable emotional psychological model. Due to the subjectivity of emotions, the modeling methods of musical emotion mental models are also different. Up to now, psychologists have proposed several classic emotional psychological models by studying the relationship between music and emotion. Commonly used are discrete emotion models and continuous emotion models. The discrete emotion modeling method divides several typical different emotion categories, and each category can contain a group of similar emotion descriptions. The continuous dimension model represents human emotional state as points in a two-dimensional or three-dimensional continuous space, which is a continuous emotional psychological description method. In the discrete emotion model, different adjectives are used to describe some basic emotions. Among them, the most famous is the Hevner emotional ring model (Hevner, 1936, 1937), which is shown in **Figure 2**. The Hevner model consists of 8 groups of 67 emotional adjectives arranged in circles. Among them, the adjectives in each large group express similar emotions, the emotions expressed by the adjacent large groups are closer and gradually progress, while the emotions expressed by the large groups in the relative position are just the opposite.

Compared with the discrete emotion model, the dimensional emotion space represents emotion through a small number of emotion dimensions. Dimensions are intended to correspond to the emotional representations within human beings. Continuous dimensional models usually represent human emotions as points in a multi-dimensional continuous space. Generally speaking, the higher the dimensionality, the more detailed the expression of emotions. But it is not that the higher the dimension, the better the emotion. If the dimensionality is too high, it will become very complicated, so the general dimensional model is two-dimensional or three-dimensional. The dimensional model does not simply classify human emotions, but describes the subtle differences between emotional states. It is a fine-grained emotional model. The disadvantage of the dimensional model is that there is a gap with people's common cognitive methods, the interaction is not good, and it is very difficult to describe the emotions in detail. The widely used continuous dimensional model is the VA model.

The VA model is proposed by Russell (Kanimozhi and Raj, 2015), as shown in **Figure 3**. This model abstracts emotions into a two-dimensional space based on arousal and valence. Valence's term in psychology is called "inducing force," which is a measure of whether people feel happy or not. The term arousal in psychology is called "activity," which is a measure of

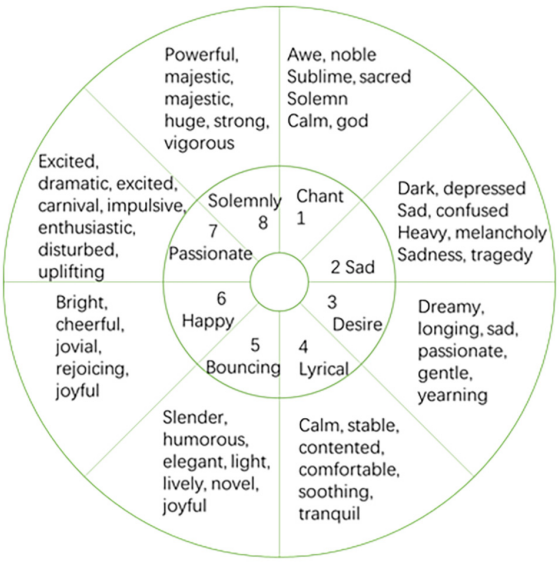


FIGURE 2 | Hevner emotional model.

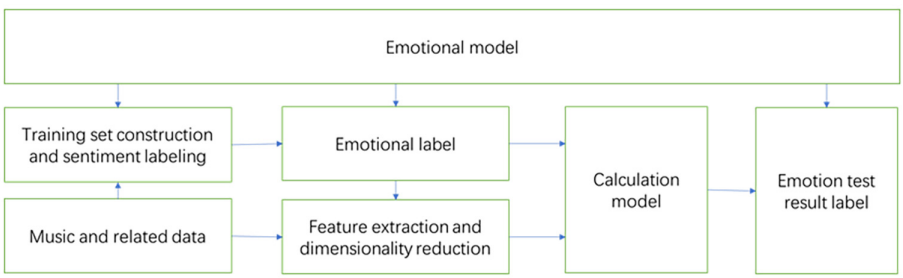


FIGURE 1 | Music emotion recognition process.

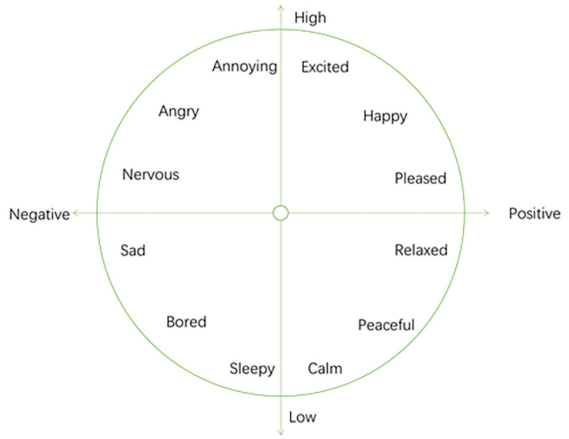


FIGURE 3 | VA emotional model.

whether people's emotions are high. Russell found that the two variables valence and arousal can represent all emotional changes. She pointed out that people's emotional preference is directly related to valence, but valence is not the only determining factor. Activation and people's emotional preferences show an "inverted U-shape." Her discovery can be summarized as the following three laws: (1) Valence has a linear relationship with people's emotional preference. (2) Activation and people's emotional preference are in an inverted U shape. (3) Valence interacts with Activation. When people are happy, there will be a higher activation value, and when people are unhappy, there will be a lower activation value. The positive and negative values of arousal and valence can be used to express four types of emotions: Positive valence and positive arousal means happiness and excitement. Negative valence and positive arousal means anger and anxiety. Negative valence and negative arousal means sadness and disappointment. Positive valence and negative arousal means relaxation and quietness.

Music Data Set

In MER, currently commonly used public data sets mainly include Computer Audio Lab (CAL) (Hu et al., 2008), Music Information Retrieval Evaluation eXchange (MIRER) (Downie, 2008), MediaEval Emotion in Music (MEM) (Aljanaki et al., 2017), and AMG1608 (Chen et al., 2017).

Computer Audio Lab contains 500 Western pop music from different musicians. The emotion model used in this data set is a discrete emotion model. However, the 18 emotions included in the model are not completely mutually exclusive, and each emotion concept is marked as an integer value between 1 and 5. For the CAL data set, if the value corresponding to each emotion concept is binarized, a discrete model of emotion can be established. If the value corresponding to the emotion is regarded as a continuous value, then a dimensional model of the emotion can be established. It can be seen that a certain conversion can be carried out between the discrete emotion model and the dimensional emotion model.

Both MIRER 2007 AMC and MIRER 2013 K-POP AMC are derived from the MIREX evaluation activity and use the same emotional representation model. The evaluation activity began to classify music emotions in 2007. The data obtained through this evaluation activity has greatly promoted the development of MER technology.

MediaEval Emotion in Music is a dynamic MER algorithm evaluation. The data set used in this evaluation comes from the research and development results of Mohammad Soleymani and others. The data set used for the evaluation contains 1744 pieces of music, all of which are 45 s segments. Each segment is marked with a segment-level static VA value and a set of dynamic VA values with an interval of 0.5 s. The music sentiment labeling of this data set is done using Amazon Mechanical Turk, and each song has at least 10 people labeling. The dynamic VA value is marked in a continuous time mode and can be under-sampled as required. The track, audio, and emotion annotations of this dataset are completely public.

AMG1608 adopts the VA dimension model, and annotates a pair of VA values for each music segment. The labeling task of this data set is also completed by Amazon Mechanical Turk. This data set contains 1608 pieces of contemporary Western music, and each piece of music selects its most emotionally representative 30 s segment as the original audio data.

MUSIC EMOTION RECOGNITION BASED ON IMPROVED BACK PROPAGATION NETWORK

Improved Back Propagation Network

The introduction of the ABC algorithm (Zhou et al., 2021) improves the original BP neural network structure. The target output value of the ABC algorithm is used as the weight and threshold of the BP neural network. It can be seen that the ABC algorithm is mainly responsible for adjusting the weights and thresholds of the BP network, and feeds back the optimal weights and thresholds to the BP network. The BP neural network after the introduction of the ABC algorithm can greatly improve its global search ability, while reducing the probability of falling into a local optimal solution, and the convergence speed is faster. The execution process of the improved BP network is shown in **Figure 4**.

By introducing the ABC algorithm, the global search function of the BP neural network is greatly enhanced. In the BP network weight and threshold training, the ABC algorithm is used to match the optimal weight and threshold to the network connection of each neuron in the BP network. The original BP neural network uses gradient descent to train the weights and thresholds of the network. The method used in this article is to complete the adjustment of the weights and thresholds of the network with the cooperation of hire bees, follow bees and detective bees.

The ABC algorithm has a memory function. The original BP neural network is prone to redundant operations such as repeated operations, so the BP neural network is unstable. After each iteration of the improved BP algorithm, the algorithm will calculate the fitness of the solution. Calculate the fitness evaluation value, sort, and record the fitness

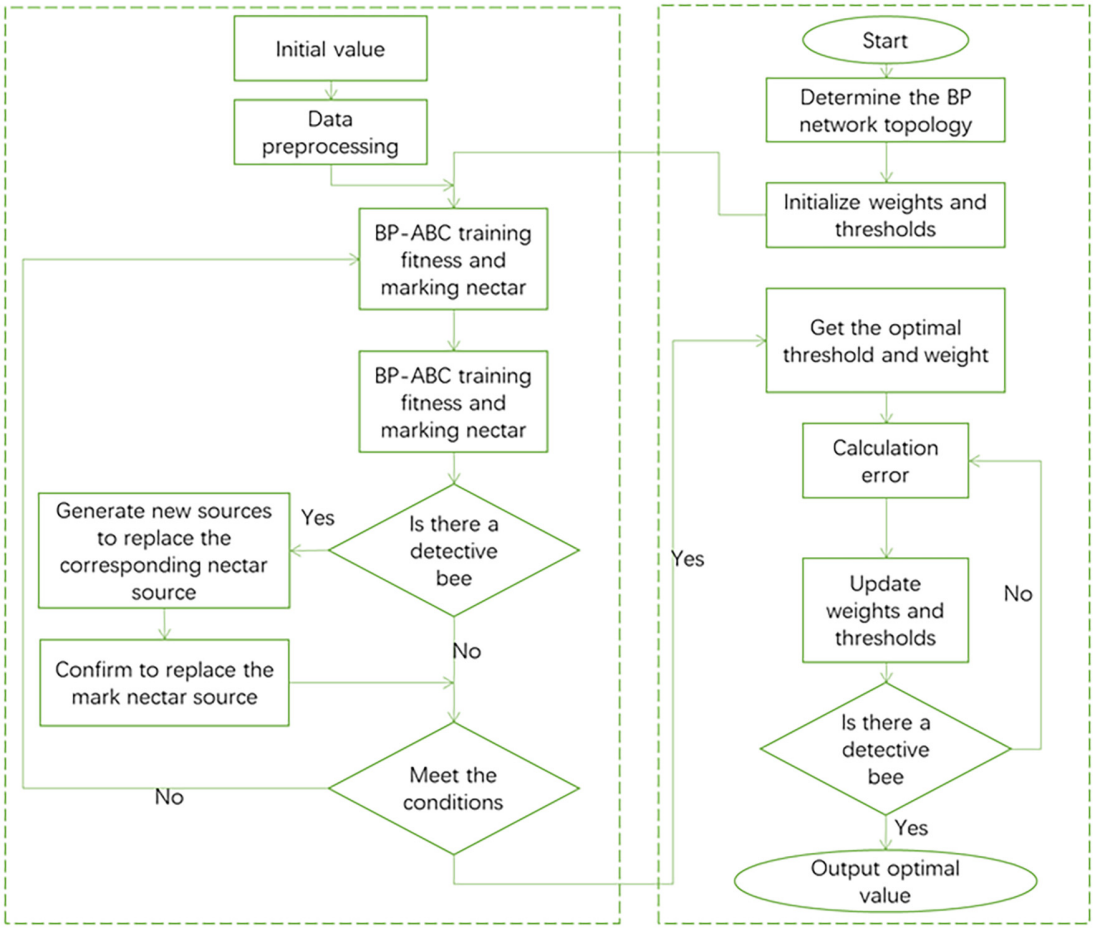


FIGURE 4 | Improved BP network flow chart.

value. After each iteration is completed, the output solution will be compared with the previous optimal solution. Take better solutions as records and save them to get the global optimal solution. In this way, problems such as repeated training caused by the unstable memory of the BP neural network are avoided.

The ABC algorithm adopts the survival of the fittest mechanism to solve the problem of handling superior solutions and inferior solutions. The ranking is performed according to the fitness value of the solution. The higher the fitness value of the solution, the higher the ranking level and the higher the probability of being selected. Finally, two iterations are selected to form the next generation group. Eliminate inferior solutions and combine high-quality solutions into the next-generation iteration group to speed up the algorithm's convergence.

The algorithm implementation steps can be summarized as the following:

(1) Set the basic parameters of the BP neural network, such as the number of input nodes, the number of output nodes, the number of hidden layers, and the number of hidden layer nodes.

(2) Set the basic parameters of the ABC algorithm, such as the value of the nectar source, the size of the bee colony, and the maximum number of iterations.

(3) Randomly assign nectar sources to the hired bees, and the hired bees begin to collect nectar source information.

(4) Calculate the fitness of the nectar source and evaluate the fitness value of the nectar source.

(5) The operation follows the bee phase. Follow bees select nectar sources according to the rule of survival of the fittest and search for high-quality nectar sources in the neighborhood.

(6) The output optimal value is compared and judged with the original value, the optimal solution is retained and the number of finding solutions is updated. If the number of search times reaches the preset upper limit and the optimal value has not been updated, the search for nectar will not continue. The follower bee turns into a scout bee and starts searching for new sources of nectar.

(7) The reconnaissance bee operation stage. Determine whether the number of iterations is the maximum. When the maximum number of iterations is not reached, jump to (4) to continue execution.

(8) Pass the global optimal solution obtained by the ABC algorithm to the BP neural network as the optimal weight and threshold solution. BP neural network uses the obtained optimal weights and thresholds for data training.

Music Emotion Recognition Process Based on Improved Back Propagation Network

First, select the samples needed for the experiment in this article, and divide the samples into two parts: training samples and test samples. The training samples are used as initial data to train the MER model. Then use test samples to detect the correctness of the training process, so as to realize MER. The experimental process is shown in **Figure 5**.

ANALYSIS OF RESULTS

Experimental Data

The experimental data set of this study is the MEM data set, 80% of the entire data set is selected as the training set, and the remaining 20% is used as the test set. The music feature data is extracted and combined to construct training and testing of MER. In order to find a better combination of eigenvalues for MER, this experiment tries to combine the eigenvalues of

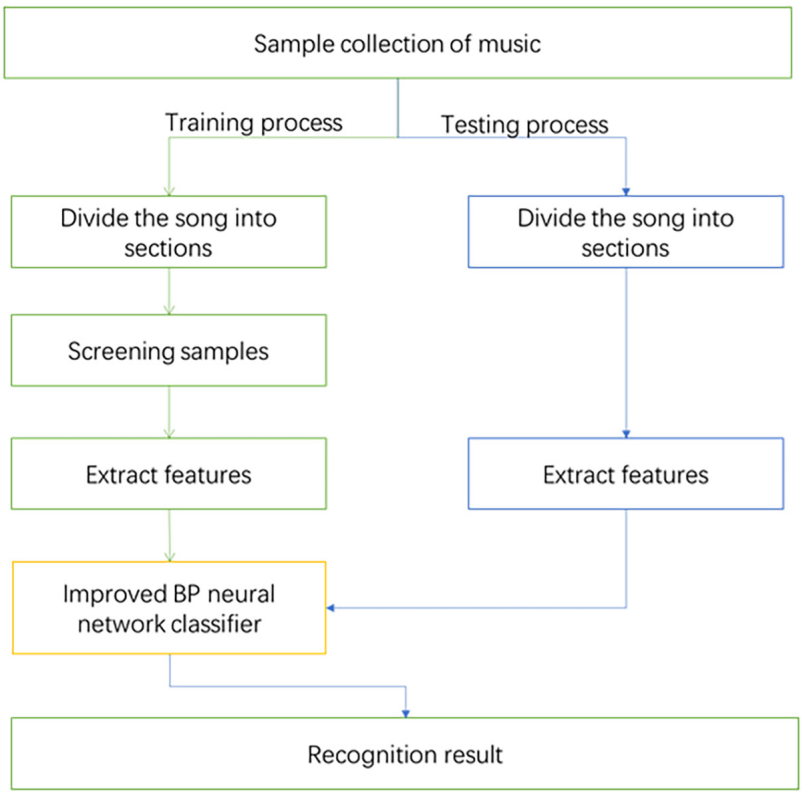


FIGURE 5 | Identification process.

TABLE 1 | Details of feature combinations.

No.	Features included in the combination
1	Short-term energy, short-term average amplitude, short-term autocorrelation function, frequency spectrum, amplitude spectrum, phase spectrum, and complex cepstrum
2	Short-term energy, short-term average amplitude, short-term autocorrelation function, short-term zero-crossing rate, frequency spectrum, amplitude spectrum, and phase spectrum
3	Short-term energy, short-term average amplitude, short-term autocorrelation function, frequency spectrum, amplitude spectrum, phase spectrum, complex cepstrum, and cepstrum
4	Short-term autocorrelation function, frequency spectrum, amplitude spectrum, phase spectrum, and cepstrum
5	Complex cepstrum and cepstrum
6	Short-term energy, short-term average amplitude, short-term autocorrelation function, short-term zero-crossing rate, complex cepstrum, and cepstrum

TABLE 2 | Music recognition rate of various feature combinations based on BP algorithm.

Combination number	1	2	3	4	5	6
Acc(%)	77.89	83.83	76.48	74.20	69.42	73.91

TABLE 3 | Experimental results of each recognition model.

Algorithm		MAE	RMSE	R^2
SVM	Valence	1.1253	0.1392	0.4374
	Arousal	1.3091	0.1405	0.5663
KNN	Valence	1.0974	0.2988	0.4403
	Arousal	1.1686	0.3265	0.5881
GMM	Valence	1.2342	0.3072	0.4236
	Arousal	1.4105	0.3136	0.5712
BP	Valence	1.1066	0.1190	0.4576
	Arousal	1.1987	0.1284	0.6284
Proposed	Valence	0.8872	0.1066	0.4606
	Arousal	0.9156	0.1322	0.6687

the time domain, frequency domain, and cepstrum domain into multiple combinations to perform MER through the improved BP algorithm. **Table 1** shows Different combinations of features.

Evaluation Index

The contrast model is SVM (Duan et al., 2012), K Nearest Neighbor Algorithm (KNN) (Bressan and Azevedo, 2018), GMM (Wang et al., 2015), BP Neural Network (Di and Wu, 2015). The evaluation indicators of each recognition model are mean absolute error (MAE), root mean square error (RMSE), and square coefficient (R^2). MAE and RMSE are error functions. The smaller the value of the function, the better the performance of the model. R^2 represents the fit of the model. The fit is the overall curve between the predicted value and the actual value of the model, which represents the fitness of the regression model. The larger the R^2 value, the closer the predicted curve is to the actual data curve, and the better the model performance.

$$MAE = \frac{\sum_{i=1}^{N} |x_i - y_i|}{N} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (x_i - y_i)^2}{N}} \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (\hat{z}_i - z_i)^2}{\sum_{i=1}^{N} (z_i - \bar{z}_i)^2} \tag{3}$$

$$Acc = \frac{TP + TN}{TP + FN + FP + TN} \tag{4}$$

where, x_i and y_i are the predicted label and actual label of sample i. $\hat{z}_i$ and z_i are the predicted label and the actual label of the test sample i, and $\bar{z}$ is the average of the actual label of the test set. TP indicates that the positive sample is judged to be a positive class. TN indicates that the negative sample is judged as negative. FP indicates that the negative samples are judged as positive. FN indicates that the positive sample is judged as a negative class.

The processor of the computer used in the experiment is Intel® Core™ i5-6200U CPU @ 2.40 GHz, the memory is 12G, and the operating system is win10 64-bit. The software environment used is MATLAB 2019.

Analysis of Results

Realize the task of MER as a classification task. The continuous emotions in VA emotional space are divided into four discrete categories, which are happy, sad, nervous and calm. Since the labels of the music clips in the data set are the points marked in the VA space, to map the emotional value to the emotional category, the emotional value needs to be divided. In this paper, the VA space is divided into four parts, and the four emotions are, respectively, corresponded to the VA space, and then the sample data is processed by classification tasks. In order to analyze which feature combination obtains the best recognition effect, the BP

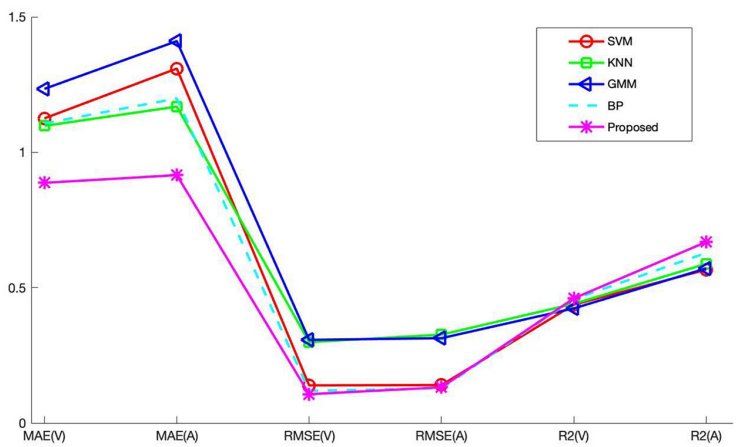


FIGURE 6 | Comparison chart of recognition results.

algorithm is first used to compare various combination features. **Table 2** gives the experimental results of recognition rate.

It can be seen from **Table 2** that in the MER experiment based on BP, the feature combination with better recognition effect is composed of short-term energy, short-term average amplitude, short-term autocorrelation function, short-term zero-crossing rate, frequency spectrum, and amplitude spectrum. The correct rate is 83.83% for the feature combination composed of phase spectrum. The accuracy of the feature combination composed of short-term energy, short-term average amplitude, short-term autocorrelation function, frequency spectrum, amplitude spectrum, phase spectrum and complex cepstrum is 77.89%. The accuracy of the feature combination composed of short-term energy, short-term average amplitude, short-term autocorrelation function, frequency spectrum, amplitude spectrum, phase spectrum, complex cepstrum, and cepstrum features is 76.48%. It can be seen from the experimental results that temporal features have a good influence on MER.

Through the above experiment, it can be analyzed that the recognition result obtained by the feature data of Combination 2 is the best. Therefore, the music recognition rate of each model pair is compared based on the feature data of Combination 2. The recognition results of each recognition model on music data are shown in **Table 3**. Since the experimental results obtained by the recognition models are not much different, the changes are not obvious. In order to visually compare the experimental results, the experimental results are displayed in a line graph. **Figure 6** is a comparison diagram of experimental results.

From the experimental results in **Table 3** and **Figure 6**, the following conclusions can be drawn:

(1) The two index values of MAE and RMSE of each comparison model show that the value on the Valence dimension is lower than the value on the Arousal dimension. This shows that the results obtained in the Valence dimension are better. The value of the Valence dimension in the R^2 indicator is lower than the value of the Arousal dimension. Since the larger the R^2 parameter, the better the experimental result, so the Arousal dimension has a better effect on R^2.

(2) The emotion recognition results obtained by the BP model are better than those of SVM, KNN, and GMM, which is why this article chooses BP as the basic recognition model. There are still some gaps between the recognition results of BP network and the proposed algorithm. This shows that the introduction of the ABC algorithm to optimize the BP network has a significant effect. The optimized BP algorithm can effectively improve the music recognition rate.

(3) The proposed method has little difference between the two dimensions of Valence and Arousal in MAE and RMSE index values. The difference between MAE and RMSE is 0.284 and 0.0256, respectively, indicating that the overall predicted value of the method used is very close to the actual value.

CONCLUSION

The continuous development of modern information technology has also promoted audio digitization research. The use of computer-related technology for MER has gradually become a research hotspot. In order to improve the recognition rate of music emotion, this research uses an improved BP network to recognize music data. This research first classifies the acoustic features of music in a combined form for emotion classification, and analyzes the most suitable feature data for emotion recognition. The experimental results show that the classification of music features based on short-term energy, short-term average amplitude, short-term autocorrelation function, short-term zero-crossing rate, frequency spectrum, amplitude spectrum and phase spectrum is better. Secondly, a music sentiment classifier was constructed using the BP network optimized by the ABC algorithm and compared with the experimental results of other classifiers. Based on the experimental results, it can be seen that the network used has a better recognition effect. The improved BP network used can complete the emotion recognition of music. In the recognition process, it can also be found that the selection of music features is also an important factor that affects the recognition effect. Reasonable extraction of music features is also an important research content of emotion recognition. In the future, this research intends to apply deep learning methods to MER. Deep learning is a learning method based on feature hierarchical structure that can learn unsupervised features. It has the feature learning ability of artificial neural network. The learned features describe the data to a higher degree. Mass music is used for feature learning, so that the machine can independently select better music features to describe the relationship between music and emotion.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://cvml.unige.ch/databases/emoMusic/.

AUTHOR CONTRIBUTIONS

JY completed the design of the experiment and the writing of the manuscript.

REFERENCES

Agarwal, G., and Om, H. (2021). An efficient supervised framework for music mood recognition using autoencoder-based optimised support vector regression model. *IET Signal Process.* 15, 98–121. doi: 10.1049/sil2.12015

Aljanaki, A., Yang, Y. H., and Soleymani, M. (2017). Developing a benchmark for emotional analysis of music. *PLoS One* 12:e0173392. doi: 10.1371/journal.pone.0173392

Anders, R. (2011). Emolion rendering in music range and characteristic values of seven musical variables. *Cortex* 47, 1068–1081.

Bressan, G. M., and Azevedo, B. C. F. (2018). A comparison of classifiers for musical genres classification and music emotion recognition. *Adv. Math. Sci. Appl.* 27, 241–262.

Casey, M. A., Veltkamp, R., and Goto, M. (2008). Content-based music information retrieval: current directions and future challenges. *Proc. IEEE* 96, 668–696. doi: 10.1109/JPROC.2008.916370

Chen, Y. A., Wang, J. C., Yang, Y. H., and Chen, H. H. (2017). Component tying for mixture model adaptation in personalization of music emotion recognition. *IEEE/ACM Trans. Audio Speech Lang. Process.* 25, 1409–1420. doi: 10.1109/TASLP.2017.2693565

Di, G. Q., and Wu, S. X. (2015). Emotion recognition from sound stimuli based on back-propagation neural networks and electroencephalograms. *J. Acoust. Soc. Am.* 138, 994–1002. doi: 10.1121/1.4927693

Downie, J. S. (2008). The music information retrieval evaluation exchange (2005-2007): a window into music information retrieval research. *Acoust. Sci. Technol.* 29, 247–255. doi: 10.1250/ast.29.247

Duan, R., Wang, X., and Lü, B. (2012). EEG-based emotion recognition in listening music by using support vector machine and linear dynamic system. *Lect. Notes Comput. Sci.* 7666, 1611–3349.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *Am. J. Psychol.* 48, 246–268. doi: 10.2307/1415746

Hevner, K. (1937). The affective value of pitch and tempo in music. *Am. J. Psychol.* 49, 621–630. doi: 10.2307/1416385

Hu, X., Downie, J. S., and Laurier, C. (2008). "The 2007 MIREX audio mood classification task: ISMIR 2008," in *Proceedings of the 9th International Conference on Music Information Retrieval, Drexel University, September 14-18* (Philadelphia, PA).

Jiang, Y., Gu, X., Wu, D., Hang, W., Xue, J., Qiu, S., et al. (2021). A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 40–52.

Jiang, Y., Zhao, K., Xia, K., Xue, J., Zhou, L., Ding, Y., et al. (2019). A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation. *J. Med. Syst.* 43:118. doi: 10.1007/s10916-019-1245-1

Kanimozhi, A., and Raj, V. C. (2015). A cognitive e-learning system using arousal valence emotional model. *J. Theor. Appl. Inf. Technol.* 78, 431–440.

Lin, Y. C., Yang, Y. H., and Homer, H. (2011). Exploiting online music tags for music emotion classification. *ACM Trans. Multimed. Comput. Commun. Appl.* 7, 1–16. doi: 10.1145/2037676.2037683

Liu, C. C., Yang, Y. H., Wu, P. H., and Chen, H. H. (2006). Detecting and classifying emotion in popular music. *Natl. Taiwan Univ.* 2006, 158–167. doi: 10.1109/TSA.2005.860344

Lu, L., Liu, D., and Zhang, H. J. (2006). Automatic mood detection and tracking of music audio signals. *IEEE Trans. Audio Speech Lang. Process.* 14, 5–18.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagation errors. *Nature* 323, 533–536. doi: 10.1038/323533a0

Shan, M. K., Kuo, F. F., and Chiang, M. F. (2009). Emotion-based music recommendation by affinity discovery from film music. *Expert Syst. Appl. Int. J.* 36, 7666–7674. doi: 10.1016/j.eswa.2008.09.042

Sharaj, P., Mehdi, R., Paul, R., and Raymond, C. K. K. (2019). Are you emotional or depressed? Learning about your emotional state from your music using machine learning. *J. Supercomput.* 75, 2986–3009. doi: 10.1007/s11227-018-2499-y

Torres, D. A., Turnbull, D., Barrington, L., and Lanckriet, G. R. G. (2007). "Identifying words that are musically meaningful," in *Proceedings of the 8th International Conference on Music Information Retrieval, ISMIR 2007, September 23-27* (Vienna).

Vidas, D., Calligeros, R., Nelson, N. L., and Dingle, G. A. (2020). Development of emotion recognition in popular music and vocal bursts. *Cogn. Emot.* 34, 906–919. doi: 10.1080/02699931.2019.1700482

Wang, J. C., Yang, Y. H., Wang, H. M., and Jeng, S. K. (2015). Modeling the affective content of music with a gaussian mixture model. *IEEE Trans. Affect. Comput.* 6, 56–68. doi: 10.1109/TAFFC.2015.2397457

Wu, T. L., and Jeng, S. K. (2008). Probabilistic estimation of a novel music emotion model. *Lect. Notes Comput. Sci.* 4903, 487–497.

Yang, Y. H., and Chen, H. H. (2011). Ranking-based emotion recognition for music organization and retrieval. *IEEE Trans. Audio Speech Lang. Process.* 19, 762–774. doi: 10.1109/TASL.2010.2064164

Zhang, K. J., and Sun, S. Q. (2012). Web music emotion recognition based on higher effective gene expression programming. *Neuro Comput.* 2012, 1–7.

Zhou, X., Lu, J., Huang, J., Zhong, M., and Wang, M. (2021). Enhancing artificial bee colony algorithm with multi-elite guidance. *Inf. Sci.* 543, 242–258. doi: 10.1016/j.ins.2020.07.037

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 18 October 2021
doi: 10.3389/fpsyg.2021.738776

Research on Sentiment Analysis and Satisfaction Evaluation of Online Teaching in Universities During Epidemic Prevention

*Jing Zhang**

School of Mechanical Technology, Wuxi Institute of Technology, Wuxi, China

Sentiment analysis of online and offline integrated teaching in universities is being paid more and more attention. Many universities have carried out online teaching activities. However, due to the lack of face-to-face teaching, the lack of emotional communication is the key problem affecting the quality of online teaching. We analyze the relations from the perspectives of the change of teaching mode, the reconstruction of teacher-student relationship, and the transmission of emotional attitude of teachers and students in this paper. Then based on the Bayesian network (BN) theory, the satisfaction of online teaching can be evaluated from the aspects of emotion analysis, learning investment, and teaching interaction. Further, some suggestions are put forward to improve the satisfaction of online teaching.

Keywords: sentiment analysis, online teaching, satisfaction evaluation, fuzzy Bayesian theory, evaluation index system of online teaching

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Wenlong Hang,
Nanjing Tech University, China
Shuang Liang,
Nanjing University of Posts and Telecommunications, China

***Correspondence:**
Jing Zhang
zhangjing@wxit.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *09 July 2021*
Accepted: *20 September 2021*
Published: *18 October 2021*

Citation:
Zhang J (2021) Research on Sentiment Analysis and Satisfaction Evaluation of Online Teaching in Universities During Epidemic Prevention. Front. Psychol. 12:738776. doi: 10.3389/fpsyg.2021.738776

INTRODUCTION

In 2020, the COVID-19 epidemic hit the world. It affected the normal life of people everywhere and changed the teaching mode of traditional higher education seriously. China's Ministry of Education had encouraged schools at all levels to make use of various online platforms to conduct online teaching. Subsequently, Chinese universities, primary and secondary schools have launched online education before other countries.

Online teaching is the main way for students to study independently during the epidemic period. As the subject of education, students have had to quickly adapt to the new teaching mode. The change from the traditional offline teaching mode requires teachers to combine online and offline flexibly and realize the seamless transition between teaching and learning, which creates high requirements for the comprehensive teaching ability of teachers. In online teaching, because teachers teach across the screen, students can't directly feel the presence of teachers. This mode leads to a lack of good emotional communication with teachers, which reduces the investment and participation of students. This has a great impact on the online teaching. In addition, the teaching platform and its upgrading will affect the satisfaction evaluation of online teaching.

Therefore, online teaching satisfaction is evaluated from the aspects of emotion analysis, learning investment, and teaching interaction. Then based on the Bayesian network (BN) model, some suggestions with strong operability are proposed.

Scholars have achieved a lot of research results on sentiment analysis and the satisfaction of online teaching. A comprehensive evaluation of online teaching satisfaction was evaluated based on the text mining method (Zhang et al., 2019). A complete mental health evaluation system is constructed based on information technology. This system can help us to understand the attitudes, opinions, and emotions of students. And we can observe their mental health state under the change of emotion (Zhou et al., 2020). In order to analyze the public opinions of "Weibo" users on online learning, web crawler software was used to crawl relevant blog data on Weibo. A text analysis method was applied to the obtained public opinion information to explore the emotional tendency of the communication of online public opinions (Liu and Yue, 2019; Guo and Min, 2020). The harmonious relationships between the teachers and the students can help to promote positive classroom interactions. A good teacher-student relationship can improve the effectiveness of classroom teaching (Nie, 2008). In order to find out the affective tendency in student teaching evaluation and provide a scientific basis for improving teaching quality, a deep learning model was proposed, based on a double-channel deep memory network for sentiment analysis in student teaching evaluation (Wang et al., 2021). A new method based on a krill swarm optimization support vector machine (KH-SVM) was proposed to evaluate the performance of college students, which was affected by the penalty coefficient and kernel function parameters (Li and Wang, 2021).

SENTIMENT ANALYSIS OF ONLINE TEACHING

The Significance of Emotional Resonance in Teaching

The teaching modes of universities were mostly offline teaching before the COVID-19 epidemic. Teaching processes in which teachers create a harmonious teaching environment are advantageous to the students. Teachers and students can properly handle the relationship between emotion and cognition in the teaching process, and teaching modes in face-to-face teaching give full play to the positive role of emotional factors and enhance students' positive emotional experience by emotional communication. So it can cultivate and develop students' emotions, then stimulate their thirst for knowledge and spirit of exploration. This teaching method urges them to form independent and sound personalities and characteristics. Teachers can shorten the distance with students in language, psychology, and space in classroom teaching. They deal with, adjust, guide, and control the psychological relationship between teachers and students correctly so that students can have a kind, pleasant, and positive emotional experience. So a teaching mode of face-to-face teaching is conducive to the formation of a good and harmonious education environment full of emotions and promotes the virtuous cycle of teaching and learning.

Lack of Emotional Input in Offline Teaching

Man is different from all other creatures because people have emotions. As far as the present situation is concerned, the channels and ways that most live broadcasting platforms provide for teachers and students to interact with each other are limited. But that is not an excuse for teachers to ignore the emotional aspects of online teaching. In online learning, teachers and students are usually separated by a computer screen. Students can see the teacher, but the teacher can't see all the students at once. With no audience in sight, the teacher is like a soloist standing in the wilderness. They easily lose the passion for teaching and feel that time stands still: their mind becomes stagnant, and they may even become incoherent. The teachers' performance is likely to directly affect the students' interest in listening to the class and thereby form a vicious circle.

In addition, there are many differences in the teaching platform, network environment, and teaching methods. During online teaching across the screen, students suffer from a lack of emotional communication with the teacher. They cannot raise awareness of their participation and cooperation and can only accept knowledge mechanically. In the network classroom, the students quietly lose interest. If truancy occurs in large numbers, then the teacher's efficacy will be greatly reduced, so that teachers and students cannot achieve emotional resonance in teaching activities. In such an atmosphere or teaching situation, the emotions of teachers and students cannot be highly consistent in mutual feeling, interaction, and mutual motivation.

In the context of online teaching, the selection of teaching content needs to be informed by science, especially in ensuring that the presentation form of the content is suitable for the online teaching mode. Because of their digital nature, network classrooms are a carrier with infinite memory ability. All the language and behavior of the teacher, as well as the subtle emotional reactions of the audience, can be recorded on the network. For a teacher, a casual mistake may be magnified on the internet, or even fermented into a storm of public opinion, which will bring great pressure or even negative influence to the teacher and the school.

People's normal and well-adapted lives had been disrupted by COVID-19, and many are facing mental health challenges. Mental health criteria are observed, including cognitive health, emotional health, personality health, social function, and psychological adjustment. During the COVID-19 outbreak, students' mental health often manifest anxiety, fear, depression, despair, anger, etc., and their emotions are more volatile than usual. Teachers' mental health problems are reflected in the disruption of life and work rhythms. There are insufficient boundaries between work and home for teachers. In addition, they have less experience with online tools from before the COVID-19 outbreak. They worry about the poor

network connection, incorrect operation, and students' lack of concentration. In online classes, teachers face not only students but also parents and even more people, which increases the pressure on teachers.

In conclusion, emphasis on scientific and emotional elements in online teaching is very important for all teachers and universities. The scientific design of online teaching will enable students to feel the beauty of science and emotion so that online teaching of science can become an art and a kind of enjoyment, and online teaching can become a new teaching method that complements offline teaching.

ONLINE TEACHING EVALUATION RESEARCH

Compared with traditional offline teaching, online teaching is not just a change of location. In online teaching, the learning environment, learning methods, teacher-student relationships, and teaching management methods have all changed. Online teaching satisfaction is closely related to students' learning ability, teachers' comprehensive quality, the network environment and hardware equipment, etc. Therefore, the evaluation of online teaching satisfaction needs to be comprehensively analyzed from multiple perspectives.

Data Sources

In order to fully understand the actual situation and satisfaction of online teaching during the epidemic period, 1,180 students from some colleges and universities in Jiangsu are surveyed by questionnaire. The survey subjects include higher vocational colleges, vocational colleges, and undergraduate universities. Survey data show that teachers choose a variety of online teaching platforms and teaching methods. At the same time, in the investigation and research of this paper, the satisfaction of some college students in Jiangsu Province with regard to online teaching during the epidemic period is also counted.

Online Teaching Satisfaction Analysis

(1) Students' online learning status is poor. According to the survey data, only 47% of students are satisfied with their learning status. The learning environments in students' homes are not as good as in school, and their psychologies are affected by the epidemic, so their learning status is negatively affected. Their satisfaction with online teaching is affected by these problems.

(2) Students have high satisfaction with teachers' teaching attitudes. In the survey, 81% of the students have a good satisfaction with teachers' teaching attitudes. Students know clearly whether the teachers are serious about their education.

(3) Students lack of self-planning. Scientific self-learning planning is conducive to students' self-learning, but the survey data show that more than half of the students lack self-learning planning. Only 47% of the students are satisfied with their self-learning planning. Without a clear self-learning plan, the learning efficacy will be reduced to some extent, thus affecting the satisfaction of online teaching.

(4) Students' low mastery of knowledge. According to the statistics, only 33% of the students have excellent satisfaction with their knowledge mastery. Horizontal analysis shows that students' satisfaction with the online learning environment, learning state, and self-planning are low. These influencing factors all directly affect students' knowledge mastery, so most students' knowledge mastery is not high.

Based on survey data of online teaching satisfaction, we built a multi-index system for evaluating online teaching satisfaction. Using the method of fuzzy empowerment, the probability values of the top indices are accurately computed by combining them with fuzzy Bayesian theory. The proposed method can further infer the weak links of online teaching during the epidemic period and provide a useful reference for the development of online teaching in a future emergency. Finally, some reasonable improvement measures of online teaching are put forward.

Online Teaching Satisfaction Evaluation Based on Improved Fuzzy Bayesian Network

A BN is a graph network model used to describe the uncertain causal relationship between variables. It is a directed acyclic graph composed of nodes, directed line segments that represent the causal relationship between connected nodes, and probability tables of nodes. A BN combines a directed acyclic graph and probability theory organically and has strong bi-directional reasoning ability. Bayesian networks play a great role in risk analysis and multi-state uncertainty reasoning.

Based on fuzzy BNs and fuzzy weighting methods, this paper improves the fuzzy BN. The object of weighting is the basic-level indexes. After the weighting of basic-level indexes, it plays a role in adjusting the influence of basic-level indexes on middle-level indexes when calculating middle-level indexes, to improve the accuracy of top-level indexes. The weighting method is the expert weighting method. The unified weights are obtained after the experts study the relationships of the basic-level indexes. The evaluation index system after weighting is more appropriate to the actual situation, so the calculated results are more accurate.

ONLINE TEACHING SATISFACTION EVALUATION BASED ON IMPROVED FUZZY BAYESIAN NETWORK

Build an Evaluation Index System for Online Teaching Satisfaction

The top-level index T of online teaching satisfaction is determined, and then the influencing factors are extended and analyzed from the three levels including the learning quality *A*, the teaching quality *B*, and the platform management *C*, respectively. The evaluation index system of online teaching satisfaction is obtained, as shown in **Table 1**.

TABLE 1 | Evaluation index system of online teaching.

The top-level index	The middle-level index		The basic-level index
Online Teaching Satisfaction T	Learning quality A	Learning conditions A_1	Learning environment A_{11}
			Learning facilities A_{12}
		Learning efficiency A_2	Learning state A_{21}
			Study plan A_{22}
		Study achievement A_3	Knowledge mastery A_{31}
			Degree of completion A_{32}
	Quality of teaching B	Resource quality B_1	Timeliness B_{11}
			Comprehensiveness B_{12}
		Course design B_2	Teaching method B_{21}
			Classroom instruction quality B_{22}
			Quality of classroom interaction B_{23}
		Teaching management B_3	Attendance management B_{31}
			Work published B_{32}
			Examinations B_{33}
	Platform management C	System Quality C_1	Account security C_{11}
			Platform stability C_{12}
		Quality of service C_2	The quality of interaction C_{21}
			Teaching support C_{22}

Determine the Satisfaction Evaluation Level

Online teaching satisfaction evaluation level is defined as $V = \{v_1, v_2, v_3, v_4\}$={*Verysatisfied*, *satisfied*, *general*, *dissatisfied*}, When $P(T) \in [0.90, 1.00]$, *it is very satisfied* v_1; when $P(T) \in [0.80, 0.90]$, *it is satisfied* v_2; when $P(T) \in [0.70, 0.80]$, *it is general* v_3; when $P(T) \in [0.00, 0.70]$, *it is dissatisfied* v_4. The online teaching satisfaction questionnaires were analyzed comprehensively. The 9-degree evaluation method evaluates m basic-level indexes in combination with the opinions of n experts to produce a weighted score. For each basic-level index obtained, the score and the evaluation matrix are shown in Equation 1:

$$D = (d_{ij})_{m \times n} \tag{1}$$

In Equation 1, d_{ij} is the evaluation of the ith index by the jth expert.

Index Probability Calculation

The BN starts from the basic-level index, and proceeds to the middle-level to get the probability of the top-level index.

(1) Prior probability calculation of the basic-level indexes. The calculation formula for the prior probability d_i of the ith basic-level index is shown in Equation 2:

$$d_i = \sum_{j=1}^{n} d_{ij}/9n \tag{2}$$

(2) Probability calculation of the middle-level indexes. According to Equation 2, the prior probability vector W of the basic-level index contained in each middle-level index can be calculated, as shown in Equation 3:

$$W = (W_1,\ W_2, ..., W_i, ..., W_n)^T \tag{3}$$

Where W_i is the prior probability vector element obtained after weighting and normalization of the basic-level indexes by Equations 4, 5:

$$p_i = d_i \times u_i \tag{4}$$

$$W_i = p_i / \sum_{t=1}^{k} p_t \tag{5}$$

Where u_i is the weight of the ith basic-level index that constitutes the middle-level index. k is the number of basic-level indexes that constitute the middle-level index, and Equation 6 is given:

$$\sum_{i=1}^{k} u_i = 1 \tag{6}$$

A tree Bayesian network, constructed by path selection and prior probability, has a reverse logical causality, and the prior probability of the middle-level index is obtained by the logical operation of its prior probability vector. The logical relation includes two kinds of relation: "AND gate" and "OR gate." For different relations, the prior probability calculation formulas are also different, as shown in Equations 7, 8:

A priori probability calculation of the relationship between "AND gate":

$$P(M = 1 \mid N_1, N_2, ..., N_n = 1) = \prod_{i=1}^{n} P(N_i = 1) \tag{7}$$

A priori probability calculation of "OR gate" relation:

$$P(M = 1 \mid N_1, N_2, ..., N_n = 1) = 1 - \prod_{i=1}^{n} P(N_i = 0) \tag{8}$$

In Equations 7, 8, M represents the middle-level index, and Ni represents the basic-level level index. $P(N_i = 1)$ represents the prior probability value of the corresponding basic-level level index as W_i, while $P(Ni = 0)$ represents the prior probability value of the corresponding basic-level level index as 1- W_i.

(3) Probability calculation of top-level indexes. By calculating the prior probability of the middle-level indexes and judging the corresponding logical relationship of the middle-level indexes, the top-level probability is calculated according to Equation 7,

8. By referring to the defined satisfaction evaluation set V, the interval of the top-level probability is found, and the online teaching satisfaction evaluation results can be obtained.

CASE STUDY

Take a university as an example, the whole university plans to open a total of 2,464 online courses, selected as a typical representative of online teaching.

Construction of Evaluation Matrix

The online teaching satisfaction questionnaire was issued to the students, and then experts were employed to comprehensively evaluate and score the basic-level indexes, and the basic-level indexes under the same middle-level indexes were weighted to obtain their weighted scores.

According to the scoring results in **Table 2**, Equation 2 can be used to calculate the prior probability of each basic-level index. Taking learning state A_{21} and learning plan A_{22} as examples, the prior probability calculation results are $d_1 = 0.5242$ and $d_2 = 0.6889$.

The calculated results of the prior probabilities *P1* and *P2* after weighting in Equation 4 are as follows:

$$p_1 = d_1 \times u_1 = 0.5242 \times 0.7 = 0.36694$$

$$p_2 = d_2 \times u_2 = 0.6648 \times 0.3 = 0.19944$$

The calculated results of the normalized prior probabilities W_{A21} and W_{A22} in Equation 5 are as follows:

$$W_{A_{21}} = \frac{p_1}{\sum_{t=1}^{2} p_t} = 0.36694 \div (0.366940 + 19944) = 0.6479$$

$$W_{A_{22}} = \frac{p_2}{\sum_{t=1}^{2} p_t} = 0.19944 \div (0.366940 + 19944) = 0.3521$$

Similarly, according to the calculated prior probabilities of other basic-level indexes, the weighted and normalized prior probabilities of each basic-level index are shown in **Table 2**.

Take learning quality A_2 as an example. The calculation result of the prior probability vector W_{A_2} is showed as follows:

$$W_{A_2} = \left(W_{A_{21}}, W_{A_{22}}\right) = (0.6479, 0.3521)$$

TABLE 2 | Prior probability of evaluation index.

Indexes	Prior probability	Indexes	Prior probability
A_{11}	0.4625	B_{22}	0.5206
A_{12}	0.5375	B_{23}	0.1859
A_{21}	0.6479	B_{31}	0.4608
A_{22}	0.3521	B_{32}	0.3226
A_{31}	0.4986	B_{33}	0.2166
A_{32}	0.5014	C_{11}	0.3266
B_{11}	0.4878	C_{12}	0.6734
B_{12}	0.5122	C_{21}	0.4656
B_{21}	0.2935	C_{22}	0.5344

Since there is an "AND gate" relationship between A_{21} and A_{22}, the prior probability of A_2 calculated by Equation 7 is 0.2281.

$$P(A_2) = \prod_{i=1}^{2} P(N_i = 1) = 0.6479 \times 0.3521 = 0.2281$$

Similarly, the prior probabilities of A_1,A_3, B_1, B_2B_3, C_1, and C_2 are, respectively, 0.2486, 0.2500, 0.2499, 0.00284, 0.0322, 0.2199, and 0.2488. The prior probabilities of indices *A*, *B,* and *C* were further calculated as 0.5650, 0.2946, and 0.4140.

Online Teaching Satisfaction Evaluation

The middle-level indexes *A*, *B,* and *C* corresponding to the top-level index form have an "OR gate" relationship. Therefore, Equation 8 is used to calculate the probability of the top-level index, and the calculated result is 0.8205. According to the definition of the satisfaction rating, 0.8202 is in the satisfactory V_2 rating range. Therefore, it is concluded that the degree of online teaching satisfaction at this university is "satisfied". The result is consistent with the actual satisfaction rating of online teaching. The result shows that the proposed model based on an improved fuzzy BN is effective in online teaching satisfaction evaluation. Therefore, the evaluation results can be used as the basis for the improvement of online teaching.

In view of the online teaching satisfaction evaluation index system, software is used to analyze the influence of the average score of different basic-level indexes on the probability size of top-level indexes. Taking learning environment A_{11}, learning planning A_{22}, teaching method B_{21}, and account security C_{11} as examples, the analysis was carried out.

It was found that the influence of the four indices on the top probability increases monotonically with the average score, but their influence on the top probability is different. Among them, the average score of the index learning environment A_{11}, learning planning A_{22}, and account security C_{11} has a significant impact on the top-level index of online teaching satisfaction T, while the low subsection of learning environment A_{11} and learning planning A_{22} with the average score interval of [0,5] has a severe impact on online teaching satisfaction T. In the average interval of [5,9], the influence is relatively gentle. This indicates that in the low subsections of [0,5] because of the students' poor learning environment and unscientific learning plan, there are serious restrictions on the improvement of online teaching satisfaction. Therefore, in this case, other aspects can be considered for improvement, so as to further improve online teaching satisfaction. The influence of account security C_{11} on online teaching satisfaction T has been increasing. It can be seen that account security has an important influence on the rapid improvement of online teaching satisfaction. The more secure the account, the higher the satisfaction of online teaching. However, teaching method B_{21} has a gentle influence on the top-level index T of online teaching satisfaction. Students can make efficient self-adjustment to adapt to their teachers' teaching methods and have a high acceptance of their teachers' teaching methods. Therefore, in order to improve the satisfaction of online teaching, we can focus on improving

students' learning environment, learning state, and the security of their accounts.

Analysis of Online Teaching System

After calculation, the probability value of the top-level index *T* of online teaching satisfaction is 0.8205, which belongs to the degree of "*satisfied*." However, there are still defects and deficiencies in online teaching. Based on the above research results, in order to improve online teaching satisfaction, the following suggestions are put forward.

(1) Improving students' learning environment. A good learning environment is the basis for teachers and students to conduct online teaching. Schools and platforms should invest more in hardware construction. More backup equipment and solutions should be available in case of network congestion. It is suggested that universities should arrange and adjust the online teaching time and implement online teaching at different peaks. According to the needs of different courses and features of the platform, different online teaching strategies are selected. According to the survey data, the performance of Tencent meeting, pinning, and learning pass are stable. Students and teachers have high satisfaction when using them. At the same time, the teachers should prepare for the online teaching environment in advance, such as live broadcasting and recording. In order to ensure normal condition online teaching, the teachers should adjust the teaching content according to the software and hardware conditions. Online teaching satisfaction is naturally improved when the learning environment is reliable.

(2) A clear study plan. Students' independent learning ability is required to be higher with online teaching. Students should make a reasonable study plan firstly. Then they must study hard and review in time. They should arrange learning time reasonably and strive to improve their learning efficiency. According to the teaching objectives, teachers should scientifically formulate the course learning outline. They should provide the relevant preview materials, thinking questions, and related online course links to the students in advance, so as to guide the students to realize their self-oriented deep learning. Clarifying learning planning is a key step to achieving satisfactory learning outcomes. It is an important index to measure the satisfaction of online teaching. Therefore, teachers should help their students to make a scientific learning plan, which can effectively improve the satisfaction of online teaching.

(3) Improve classroom interaction. Effective online teaching should be a learning process that involves the participation of all students and the interactions between teachers and students. So the interactions among teachers and students, and between students and students can form multiple interactive modes of online teaching. Teachers can discuss with students with the help of cloud class, nail, learning pass, etc., after online class. They can also set time-limited online tests to increase real-time comment results. If some course assistants can be hired, they collect and feedback students' problems in the process of online teaching. And they will cooperate with the teachers to better complete online teaching. In addition, students' independent ability would be constantly cultivated through lively classroom interaction. Participatory teaching is good for improving students' initiative and interest in learning. Teachers and students will have a better experience of online teaching with good classroom interactions. From the perspective of emotion analysis, the teachers communicate the useful course information to students in time by bullet screen. This method will build a bridge of interaction between teachers and students. The emotional connection between teachers and students will be further deepened, so that the satisfaction of online teaching is improved.

(4) Develop a new assessment mechanism. Online teaching assessment mechanisms are an important link in the evaluation of online teaching management. The scientific and reasonable construction of online teaching assessment systems is conducive to improving teachers' initiative in teaching. The final examination plays a decisive role in the assessment of students' grades. They study hard only for a few weeks before the final exam. This kind of cramming learning method does not help to master their knowledge. As a result, most students are slack in their studies. Based on the above problems, a new assessment mechanism is proposed by changing the composition of grades and defining their total score as "the average score accounts for 60%, the final score accounts for 40%" (or "the average score accounts for 70%, the final score accounts for 30%"). The average score has a large impact on the total score, which means the assessment mechanism can run through the whole process of online teaching. This kind of comprehensive evaluation focuses on the evaluation of students' learning process including the students' participation in class, the status of homework completion, the number of discussion posts, and the final exam results. This new assessment mechanism can avoid students from slacking off in class for most of the semester, and only make a brief effort at the end of the semester. It requires students to take online classes seriously, actively perform, and earnestly complete learning tasks. This new assessment mechanism can better play the role of the teachers, so as to improve the satisfaction of online teaching.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Wuxi Institute of Technology. Written informed consent to participate in this study was provided by the participants.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work was supported in part by the Jiangsu Province education system party building research key project (Grant No. 2019JSJYDJ01018); the Jiangsu University philosophy and social science research project (Grant Nos. 2019SJB281 and 2021SJB1394); the Jiangsu Social Science Application Research Project (Special Topic of Ideological and Political Education in Colleges): (Grant No. 21SZB-013).

REFERENCES

Guo, R. R., and Min, S. Q. (2020). "Online learning" public opinion analysis and online teaching improvement strategies. *J. Commun. Univ. China Sci. Technol.* 27, 48–54, 75.

Li, L., and Wang, X. (2021). Mental health evaluation of college students based on KH-SVM. *Microcomput. Appl.* 37, 45–51.

Liu, L., and Yue, Y. W. (2019). Cross-grained sentiment analysis oriented to college student microblog. *Appl. Res. Comput.* 36, 1619–1622.

Nie, Y. (2008). An analysis on the effect of new teacher- students relationship in university inter active teaching. *J. Guangdong Polytech. Normal Univ.* 1, 96–98.

Wang, B. H., Xiong, Y., Yao, Y., Chu W., and Lv, Y. (2021). Sentiment analysis of student teaching evaluation based on deep learning. *Res. Audio Vis. Educ.* 4, 101–107.

Zhang, D. M., Li, Y. P., and Jin, M. (2019). Satisfaction evaluation of online education platform of higher education in China. *Educ. Modern.* 6, 175–180.

Zhou, T. H., Gao, K. T., Zhang, C. F., and Wang, L. (2020). Emotional analysis model of college students based on big data public-opinion. *Wirel. Internet Technol.* 9, 115–116.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 22 October 2021
doi: 10.3389/fpsyg.2021.741665

Emotion Recognition of Chinese Paintings at the Thirteenth National Exhibition of Fines Arts in China Based on Advanced Affective Computing

Jing Li[1], Dongliang Chen[2], Ning Yu[1], Ziping Zhao[1] and Zhihan Lv[3]*

[1] College of Art, Qingdao Agricultural University, Qingdao, China, [2] College of Computer Science and Technology, Qingdao University, Qingdao, China, [3] Faculty of Arts, Uppsala University, Uppsala, Sweden

Today, with the rapid development of economic level, people's esthetic requirements are also rising, they have a deeper emotional understanding of art, and the voice of their traditional art and culture is becoming higher. The study expects to explore the performance of advanced affective computing in the recognition and analysis of emotional features of Chinese paintings at the 13th National Exhibition of Fines Arts. Aiming at the problem of "semantic gap" in the emotion recognition task of images such as traditional Chinese painting, the study selects the AlexNet algorithm based on convolutional neural network (CNN), and further improves the AlexNet algorithm. Meanwhile, the study adds chi square test to solve the problems of data redundancy and noise in various modes such as Chinese painting. Moreover, the study designs a multimodal emotion recognition model of Chinese painting based on improved AlexNet neural network and chi square test. Finally, the performance of the model is verified by simulation with Chinese painting in the 13th National Exhibition of Fines Arts as the data source. The proposed algorithm is compared with Long Short-Term Memory (LSTM), CNN, Recurrent Neural Network (RNN), AlexNet, and Deep Neural Network (DNN) algorithms from the training set and test set, respectively, The emotion recognition accuracy of the proposed algorithm reaches 92.23 and 97.11% in the training set and test set, respectively, the training time is stable at about 54.97 s, and the test time is stable at about 23.74 s. In addition, the analysis of the acceleration efficiency of each algorithm shows that the improved AlexNet algorithm is suitable for processing a large amount of brain image data, and the acceleration ratio is also higher than other algorithms. And the efficiency in the test set scenario is slightly better than that in the training set scenario. On the premise of ensuring the error, the multimodal emotion recognition model of Chinese painting can achieve high accuracy and obvious acceleration effect. More importantly, the emotion recognition and analysis effect of traditional Chinese painting is the best, which can provide an experimental basis for the digital understanding and management of emotion of quintessence.

Keywords: advanced affective computing, Chinese paintings, deep learning, the Thirteenth National Exhibition of Fines Arts in China, emotion recognition

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Mary Nina Wang,
University of Malaya, Malaysia
Yuan Fu,
Tianjin University, China

****Correspondence:***
Ning Yu
yuningluck@126.com

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *15 July 2021*
Accepted: *24 September 2021*
Published: *22 October 2021*

Citation:
Li J, Chen D, Yu N, Zhao Z and Lv Z (2021) Emotion Recognition of Chinese Paintings at the Thirteenth National Exhibition of Fines Arts in China Based on Advanced Affective Computing. Front. Psychol. 12:741665. doi: 10.3389/fpsyg.2021.741665

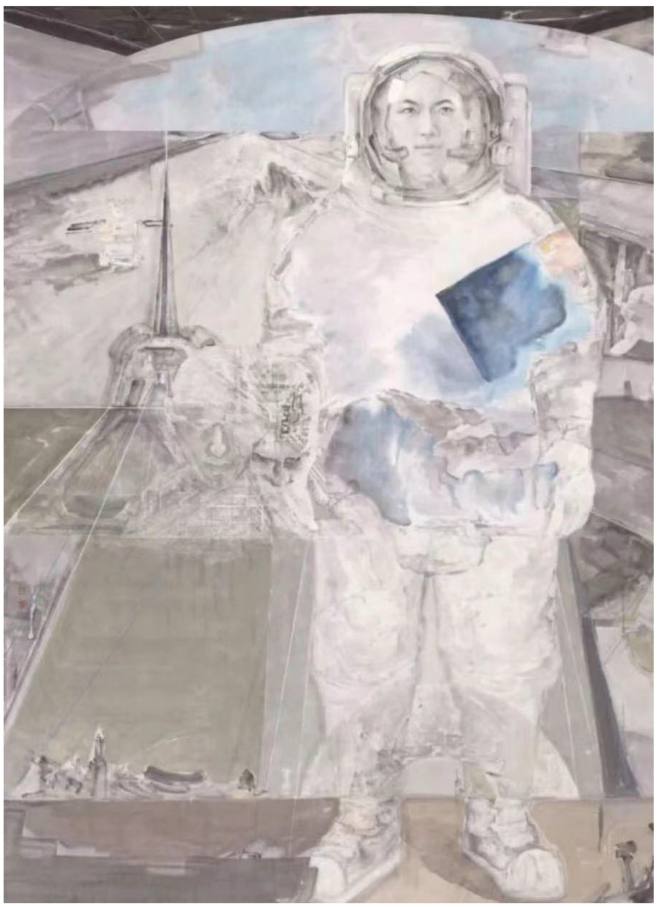
INTRODUCTION

Now that the world comes to the big data era, Artificial Intelligence (AI) has impacted people's lives profoundly; the traditional interpersonal interactions begin to shift to human-computer emotional interactions. In the meantime, people put forward higher requirements for esthetic appreciations than in the past. The performance of simple painting techniques has been unable to meet the emotional needs of artists, more and more artists focus on the emotional expression of works. In this case, how to understand the emotional expression of works of art is particularly important. More and more researchers are committed to the emotional expression of works of art.

The works of the Thirteenth National Exhibition of Fines Arts are close to the daily life of the Chinese people. It reflects the new changes in the development of modern society and the achievements of China's development in the new era, and shapes and depicts the spirit of the Chinese nation in the new era. The Chinese paintings in this exhibition have the following emotional descriptions: Li Encheng's *Fanghua* shows the posture of natural mountains and fields, praising the unknown and selfless workers in all walks of life. *Mission* embodies the patriotic spirit of heroism and model practitioners—fire fighters. *Pursuing the Dream of Space* shows the astronauts' exploration of the secrets of the universe, the endless Chinese spirit, and Chinese power. The selected works of Chinese painting in the Thirteenth National Exhibition of Fines Arts reveal that, the themes consciously highlight the spirit of the times, depict all kinds of people's livelihood, the Chinese mountains and rivers are thriving, and the painting themes show a new fashion of the times.

Fanghua by Li Encheng

Mission by Li Yuwang



Pursuing the Dream of Space by Sun Chunlong

It may be challenging to capture the emotions expressed in these paintings accurately. Therefore, using Advanced Affective Computing to analyze the emotions expressed in these paintings can offer an opportunity for the general public to appreciate art. However, the result may be contrary to the user's actual emotion applying the traditional single model for Chinese painting sentiment analysis. Regarding the complementary associations of various modality information, multimodal emotion recognition has gradually received widespread attention to thoroughly utilize the present modality data to capture information in and between the modalities (He and Zhang, 2018). Users pose new challenges to emotion recognition while they employ different modalities to portray their real-time emotions and make their current emotions vivid.

As for image data such as Chinese paintings, features such as SIFT (Scale-Invariant Feature Transform) operator vectors can be obtained. Original feature data of different modalities have their unique structural characteristics belonging to different feature spaces. Hence, noise modality may also be mixed into them during feature extraction (Zhang et al., 2020). Traditional emotion recognition algorithms may increase the calculation time and space cost, resulting in disturbed recognition outcomes. As of now, many large-scale, high-quality datasets are proposed, such as ImageNet, Deep Learning (DL) approaches have made breakthrough achievements in image recognition. Compared with artificially designed features, the unique multi-layer structure of DL enables it to learn the in-depth features that gradually transits from universal, low-level visual features (such as edges and textures) to high-level semantic representations (such as the torso and head); the higher the level, the stronger the expression (Wang et al., 2019; Chen et al., 2020). The hierarchical nature of the deep features provides a practical means to bridge the semantic gap and understand human affects in paintings (Liu et al., 2021) .

To sum up, under the trend of the rapid development of AI, using advanced affective technology to identify the emotion of Chinese painting with multiple feature spaces has become the internal driving force for people to understand artistic emotion, which has great practical significance for the inheritance of Chinese traditional culture. Therefore, the innovation is to solve the problem of "semantic gap" in the emotion recognition task of images such as traditional Chinese painting, select and improve the AlexNet algorithm on the basis of deep learning theory to further improve its performance. The study also adds chi square test to solve the data redundancy and noise problems in various modes such as Chinese painting. Additionally, a multimodal emotion recognition model of Chinese painting based on improved AlexNet neural network and chi square test is designed to provide theoretical support for the digital understanding and development of emotion in Chinese quintessence.

RELATED WORKS

Trend of Intelligent Emotion Recognition

Emotion capture is the foundation of daily communication. Multi-party cooperation in all professions and trades is inseparable from emotion analysis and emotion recognition. Jain et al. (2018) proposed a hybrid DNN (Deep Neural Network) to recognize emotions in face images. They tested this hybrid DNN on two public datasets and harvested satisfactory experimental results (Jain et al., 2018). Avots et al. (2019) captured and classified facial expressions using SVMs (Support Vector Machines). They utilized Viola-Jones facial recognition algorithm and CNN (AlexNet) emotion classification algorithm to find human faces in keyframes. Ultimately, the proposed algorithm was validated (Avots et al., 2019). Hwang et al. (2020) put forward a novel emotion recognition approach based on CNN (Convolutional Neural Network) while preventing local information loss. Lastly, visualization proved that this approach outperformed other Electroencephalography (EEG) feature representation models based on standard features, proving its effectiveness (Hwang et al., 2020). Wei et al. (2021) designed an algorithm to extract keyframes from videos based on emotion saliency estimation regarding the current status of video emotion recognition. Keyframes could be extracted to avoid the influence of the emotion-independent frames on the result by estimating the emotional saliency of the video frame. Through simulation, the designed algorithm could improve the performance of video emotion recognition, outperforming the present user-generated video emotion recognition (Wei et al., 2021).

Advanced Affective Computing for Multimodal Analysis

Hammad et al. (2018) put forward a secure multimodal biometric system based on CNN and a QG-MSVM (Q-Gaussian Multi-Support Vector Machine) based on different fusion levels. They applied this internal fusion system to combine the biological characteristics to generate a biological characteristic template. Results demonstrated that the proposed system could provide higher efficiency, robustness, and reliability than the present multimodal authentication system (Hammad et al., 2018). Ćosić et al. (2019) enhanced the process of ATC (Air Traffic Controller) selection based on the traditional ATC psychophysiological data measurement, including complicated physiological, eye volume, and voice measurements and appropriate metrics, as well as the ability to measure the multimodality in particular stimulus tasks. They comprehensively analyzed the multimodal features of this method under different experimental conditions and found it pretty advantageous (Ćosić et al., 2019). Yuan et al. (2020) designed and implemented the MSNVRFL (Multimodal Sensing Navigation Virtual and Real Fusion Laboratory) regarding the importance of virtual experiments in HCI (Human-Computer Interaction). MSNVRFL was equipped with a set of experimental devices with cognitive functions, in which a multimodal chemical experiment fusion algorithm was explored, validated, and applied (Yuan et al., 2020). Walia and Rohilla (2021) reviewed the various biological activity detection technologies based on multimodal biometric systems. After analysis and discussion, they proposed a new classification approach and finally proved its effectiveness (Walia and Rohilla, 2021).

Apparently, most works about image sentiment analysis are based on face images or sequences; in contrast, the sentiment analysis for images such as Chinese paintings is still a great challenge. Works about intelligent Chinese painting recognition have been reported; nevertheless, they focused on identifying the author of the painting and analyzing the work style, with very little research on the emotions shown by Chinese paintings. Intelligent algorithms, such as DL, have not played a significant role in emotion recognition. In this case, DL approaches are employed in the present work to identify and analyze the emotions expressed in Chinese paintings, which is of great significance for the in-depth understanding of the emotions of Chinese paintings subsequently.

MULTIMODAL EMOTION RECOGNITION AND ANALYSIS OF CHINESE PAINTINGS AT THE THIRTEENTH NATIONAL EXHIBITION OF FINES ARTS IN CHINA BASED ON DL APPROACH

The Thirteenth National Exhibition of Fines Arts vividly shows the spirit of the Chinese nation in the new era and the happy life of the people. A DL-based multimodal emotion recognition model for Chinese paintings is proposed based on image data of Chinese paintings at the Thirteenth National Exhibition of Fines Arts in China. Subsequently, its performance is validated through comparative simulation experiments.

Demand for Chinese Painting Emotion Recognition

Chinese paintings pay attention to image modeling. It cannot blindly imitate the objective reality, nor can it be purely fabricated, but pursues the artistic state of "like and not like." Based on respecting the objective images, painters integrate their feelings into the objective things. In this way, they can not only express their feelings in the works, but also make the images have emotion and vitality. Emotion is the driving force of Chinese painting creation (Bi et al., 2019; Yang et al., 2021). Traditional algorithms to recognize human affects in Chinese paintings often combine art theory and computer vision, practicing artificially designed features and statistical ML (Machine Learning) approaches to recognize the emotional responses evoked by Chinese paintings. The emotional response of Chinese paintings is presented in **Figure 1**. However, the large-scale, high-quality labeled datasets, such as ImageNet, make traditional ML approaches incapable of training deeper DL models, leading to severe overfitting.

Regularly, the large-scale dataset distribution used for pre-training is very different from the dataset distribution of the target task. For instance, the ImageNet dataset is dominated by natural images (such as animals and natural scenes). In contrast, image sentiment classification datasets such as Chinese paintings are completely composed of painting elements. Thus, if the pre-training weights of the ImageNet dataset are directly transferred, under-fitting caused by cross-domain transfer will occur. Meanwhile, a large-scale, image emotion recognition dataset has not been established yet. In this case, how to resolve under-fitting becomes a hot topic.

An emotion recognition model for images such as Chinese paintings is established based on DL approaches to bridge the gap between the universal, low-level visual features and high-level emotional semantics and address the small sample issue of emotion recognition datasets. Afterward, DL overfitting in the case of a small dataset gets improved based on TL (Transfer Learning). In the meantime, a two-layer TL strategy is proposed for emotion recognition of Chinese paintings to resolve under-fitting during cross-domain transfer.

DL Approaches for Chinese Painting Emotion Recognition

While appreciating Chinese paintings, people often evaluate art pieces regarding artistic conception, charm, and drawing techniques. If emotions in Chinese paintings can be decomposed and analyzed quantitatively, people will understand the way Chinese paintings express emotions and thus better appreciate the beauty of Chinese paintings. Besides, emotion visualization can intuitively demonstrate the digital laws of emotions behind Chinese paintings, greatly enlarging the effect and quality of cultural and artistic learning, analysis, and training (Wang et al., 2017). CNN, RNN (Recurrent Neural Network), and LSTM (Long Short-Term Memory) are standard image recognition algorithms. In the present work, CNN is selected for emotion recognition. CNN is a feedforward neural network, which often contains multiple layers in different types, such as convolutional layers, fully connected layers, and pooling layers (Haberl et al., 2018; Kim et al., 2020). **Figure 2** visualizes the procedures of CNN processing and recognizing Chinese paintings.

Normally, CNN performs convolution operations in multiple dimensions. Suppose the input is a 2D matrix I; in that case, there is a 2D kernel K satisfying the following equation:

$$S(i,j) = (I \cdot K)(i,j) = \sum_m \sum_n I(m,n)\, K(i-m, j-n) \quad (1)$$

In (1), *(i, j)* describes the matrix dimension, and *(m, n)* refers to the matrix order. Convolution can be exchanged and equivalently written as the following equation:

$$S(i,j) = (I \cdot K)(i,j) = \sum_m \sum_n I(i-m, j-n) K(m,n) \quad (2)$$

The flipped convolution kernel, as opposed to the input, gives interchangeability features to the convolution operation; the input index is increasing, while the kernel index is decreasing. The only goal of kernel flipping is to achieve exchangeability; although this feature is helpful in proofs, it is not that significant to neural networks. Instead, many neural network libraries possess a function called CC (Cross-Correlation) (Lin et al., 2018), almost the same as the convolution operation but cannot flip the kernel:

$$S(i,j) = (I \cdot K)(i,j) = \sum_m \sum_n I(i+m, j+n) K(m,n) \quad (3)$$

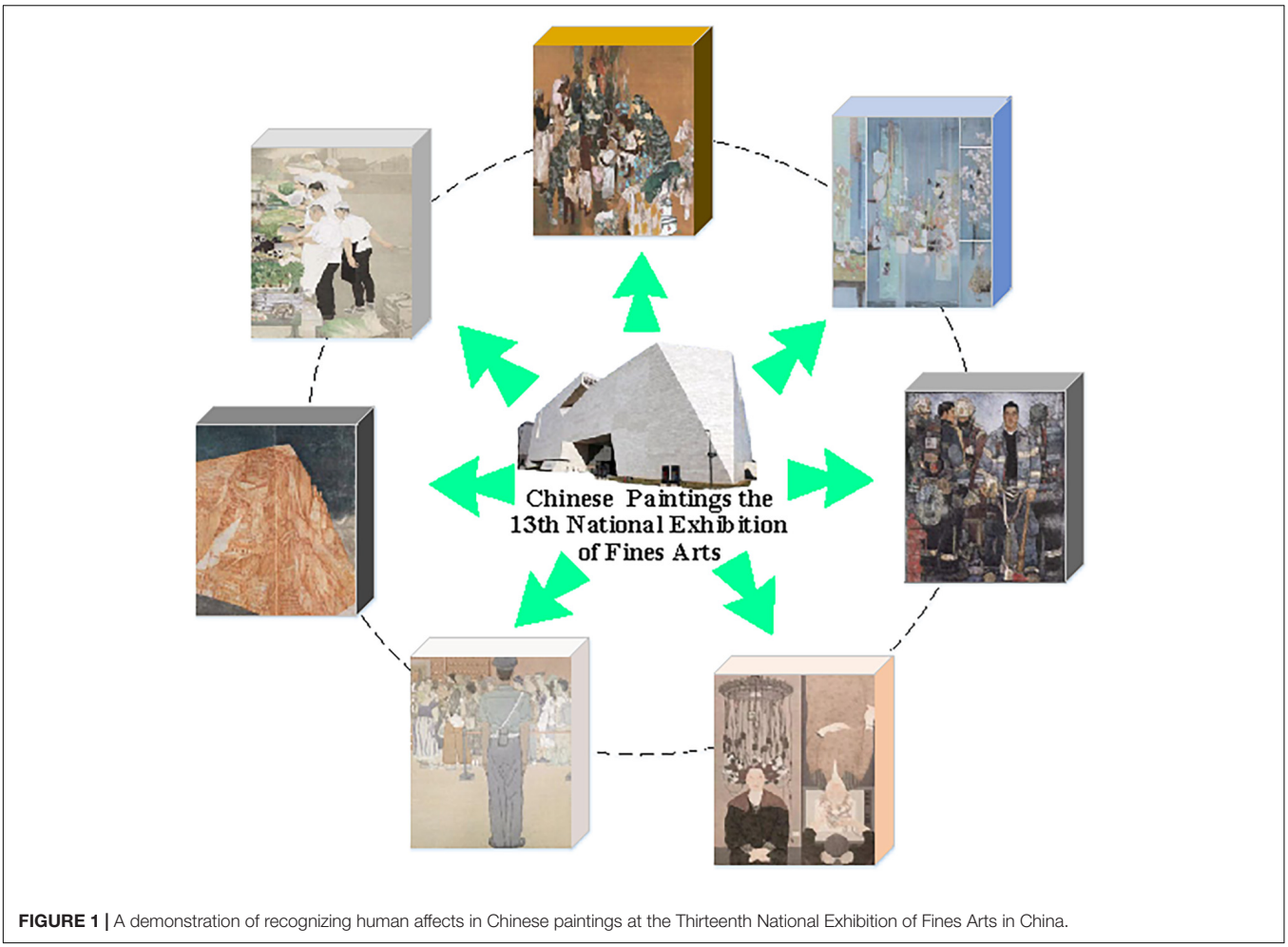


FIGURE 1 | A demonstration of recognizing human affects in Chinese paintings at the Thirteenth National Exhibition of Fines Arts in China.

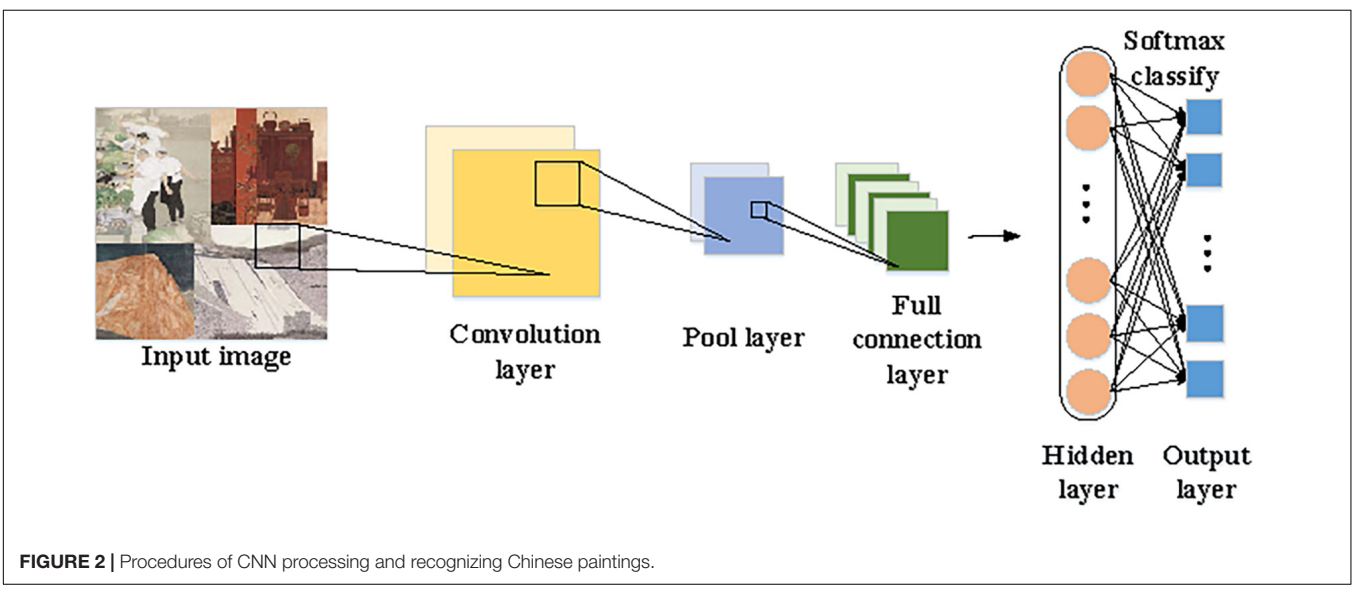


FIGURE 2 | Procedures of CNN processing and recognizing Chinese paintings.

Convolutional neural network classifies pixels of the original Chinese paintings before the up-sampling operation that continuously scales the image to the original size. The final output is the up-sampled image, so that each pixel in the output image can be forecasted. In particular, the maximal image pixel is found in all the images finally obtained. AlexNet has more layers and

stronger learning capability among various CNNs; thus, it is selected in the present work to reduce the calculation amount and strengthen the generalization performance of the algorithm (Jia et al., 2019). Suppose that $A_i^{(l)}$ denotes the output of the l-th convolutional layer in CNN, $A^{(0)}$ describes the Chinese painting input; in that case, its $i\left(1 \le i \le M^{(l)}\right)$-th feature map can be calculated through:

$$A_i^{(l)} = \sigma\left(\sum_{j=1}^{M^{(l-1)}} A_j^{(l-1)} \otimes W_{i,j}^{(l)} + b_i^{(l)}\right) \tag{4}$$

In (4), $M^{(l)}$ refers to the total number of feature maps of the l-th convolutional layer, W represents the weight of the convolution kernel, b is the bias, $\otimes$ signifies the convolution operation, and $\sigma(\cdot)$ demonstrates ReLU (Rectified Linear Unit) (Gu et al., 2019). ReLU can fix vanishing and exploding gradient and enhance the network's expression. In the pooling layer, standard pooling strategies are Max-Pooling and Average-Pooling. AlexNet often employs the Max-Pooling strategy. After the convolutional layer and the pooling layer map the original data to the hidden layer characteristic space, multiple fully connected layers are connected to map the learned feature representation to the sample's label space. AlexNet has three fully connected layers; their equations are:

$$A_i^{(6)} = \sigma\left(\sum_{j=1}^{n^{(5)}} W_{i,j}^{(6)} \times F_j\left(A^{(5)}\right) + b_i^{(6)}\right) \tag{5}$$

$$A_i^{(7)} = \sigma\left(\sum_{j=1}^{n^{(6)}} W_{i,j}^{(7)} \times A^{(6)} + b_i^{(7)}\right) \tag{6}$$

$$A_i^{(8)} = \varphi\left(\sum_{j=1}^{n^{(7)}} W_{i,j}^{(8)} \times A^{(7)} + b_i^{(8)}\right) \tag{7}$$

In (5) $\sim$ (7), $n^{(l)}$ refers to the number of neurons in the l-th layer, $F(\cdot)$ refers to the tiling operation in which the result of the last convolutional layer is expanded into a 1D vector, $\varphi(\cdot)$ signifies SoftMax that predicts the probability that the input vector belongs to each emotion category, and its equation is:

$$\varphi_j(X) = \frac{e^{X_j}}{\sum_k e^{X_k}} \tag{8}$$

In (8), j represents the j-th element of the input vector X, and k denotes the total number of sample categories. Furthermore, AlexNet's convolutional layer is improved. The operation of "local normalization before pooling" is advanced to "pooling before local normalization." This improvement brings two benefits. First, the generalization ability of AlexNet can be enhanced while the over-fitting can be weakened, which greatly shortens the training time. Second, pooling before local normalization can retain useful information, weaken redundant information, and accelerate the convergence of the proposed algorithm, highlighting the superiority of overlapping Max-Pooling. **Figure 3** presents the procedures of the improved AlexNet processing and recognizing Chinese paintings.

Removing redundant information from feature data is a necessary step before an emotion recognition task, especially when the amount of original modality data is particularly huge. Traditionally, text information and image information can be extracted from images for emotion recognition. However, emotion recognition of Chinese paintings is a binary classification task. Only using image information can accurately determine the negativity and positivity of the user's current emotion. Hence, image information occupies more weight than text information. If text information appears in this case, it will be regarded as subordinate redundant information. A process to remove redundant information in the modality feature, retain the weighted feature in the emotion recognition task, and reduce the calculation cost and resource wastes is extremely meaningful and remarkable. There are three well-known feature selection methods: packaging, embedding, and filtering. The packaging method comprises a learning algorithm that can evaluate the accuracy performance of a subset of candidate features, providing better solutions than the other two methods. SVM-RFE (Support Vector Machine-Recursive Feature Elimination) is a standard packaging method (Spoorthi et al., 2018; Nalepa et al., 2019; Sukumarana and Mb, 2019). Suppose a given training sample set as $\{x_i, y_i\}, x_i \in R^d, y_i \in \{-1, 1\}, i = 1, \ldots, n$; in that case, the decision function of linear SVM is:

$$f(x) = w * x + b \tag{9}$$

In (9), w represents the weight, and b denotes the bias term. The boundary M is proved to be only $2/|w|$. Hence, the maximum boundary is equivalent to minimizing $||w||^2$ under constraints. The dual form of the LaGrange problem can be expressed as:

$$L_D = \sum_{i=1}^{n} \partial_i - \frac{1}{2}\sum_{j=1}^{n} \partial_i \partial_j y_i y_j x_i x_j \tag{10}$$

In (10), ∂_i refers to the LaGrange multiplier. The solution to ∂_i can be calculated by maximizing L_D under constraints $\partial_i \ge 0$ and $\sum_{i=1}^{n} \partial_i y_i = 0$. Samples corresponding to non-zero ∂ are called support vectors. The weight vector w can be obtained by the following equation:

$$w = \sum_{i=1}^{n} \partial_i y_i x_i \tag{11}$$

Then, the ranking criterion of the pixel feature k of Chinese paintings is the square of the k-th element of w:

$$J(k) = w_k^2 \tag{12}$$

SVM-RFE trains all SVM features to calculate their weights, which is pretty cumbersome. Chi Square Test and information gain are two widespread filtering methods. During feature extraction, Chi Square Test measures the dependency between features and categories. The higher the score, the more dependent the categories are on the given features. Thus, features with lower scores have less information and should be deleted. Suppose

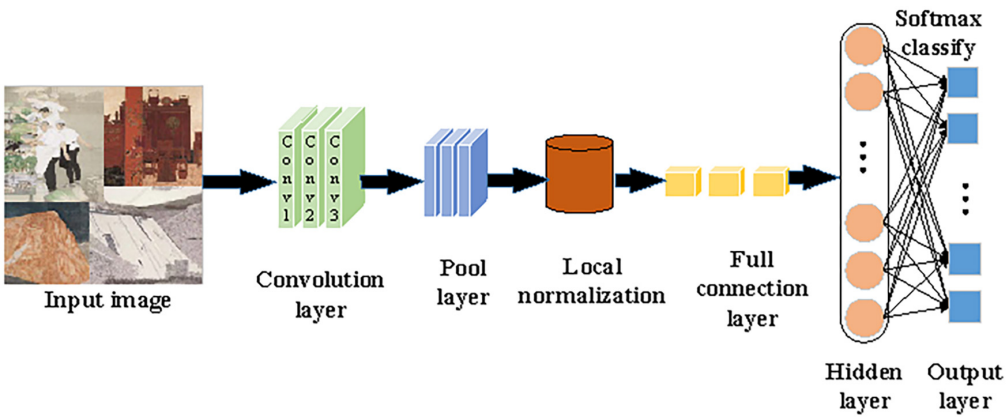


FIGURE 3 | Procedures of the improved AlexNet processing and recognizing Chinese paintings.

that a feature in a Chinese painting is independent of the final classification; in that case, the Chi Square Test can be defined as:

$$CHI(t, c_i) = \frac{N \times (AD - BC)^2}{(A + C) \times (B + D) \times (A + B) \times (C + D)} \quad (13)$$

$$CHI_{\max}(t) = \max\left(CHI(t, c_i)\right) \quad (14)$$

In (13) and (14), A refers to the number of documents with feature t and belonging to category c_i , B describes the number of documents with feature t but not belonging to category c_i , C refers to the number of occurrences of c_i without t, D signifies the frequency where neither c_i nor t appears, and N refers to the total number of instances in the document set, $N = A+B+C+D$. If t and c_i are independent, the *Chi* statistic will be zero.

Information gain measures the information obtained after the feature values in the document are known. The higher the information gain, the better the ability to distinguish different categories. The information can be calculated by capturing the uncertainty entropy of the probability distribution of a given category. Suppose there are m categories: $C = \{c_1, c_2, \cdots, c_m\}$; in that case, the following equation can be deduced based on entropy:

$$H(C) = -\sum_{i=1}^{m} p\left(c_i\right) \log_2 p\left(c_i\right) \quad (15)$$

In (15), $p\left(c_i\right)$ denotes the probability of how many documents are in category c_i. Suppose that the attribute A in Chinese paintings has n different values: $A = \{a_1, a_2, \cdots, a_m\}$; in that case, the entropy after examining attribute A can be defined as follows:

$$H(C|A) = \sum_{j=1}^{m} \left(-P\left(a_j\right) \sum_{i=1}^{m} p\left(c_i|a_j\right) \log_2 p\left(c_i|a_j\right) \right) \quad (16)$$

In (16), $P\left(a_j\right)$ suggests the probability of how many documents contain the attribute value a_j , and $p\left(c_i|a_j\right)$ describes the probability of the number of documents containing the attribute value a_j in category c_i . Based on the above definition, the information gain of an attribute is exactly the difference between the entropy values before and after examining the attribute, which can be expressed as:

$$IG(A) = H(C) - H(C|A) \quad (17)$$

Therefore, during sentiment analysis, the filtering method can select a subset of features from the original features, use statistical measures to rank the available features, and automatically filter out those features whose scores are lower than a predetermined threshold. More importantly, it can assess feature subsets without learning algorithms. In short, this method is easy to design and does not require substantial computing resources, showing significant advantages for large datasets.

The Multimodal Emotion Recognition Algorithm for Chinese Paintings at the Thirteenth National Exhibition of Fines Arts in China Based on Chi Square Test and Improved AlexNet

A multimodal emotion recognition algorithm for Chinese paintings is designed based on improved AlexNet and Chi Square Test. In particular, because it has more layers and stronger learning capability than other CNNs, AlexNet is chosen in the present work and gets improved to solve the "semantic gaps" in recognizing emotions from images such as Chinese paintings. Furthermore, Chi Square Test removes data redundancy and data noise in each modality of Chinese paintings and captures the internal connections among the modalities. This algorithm not only addresses the above concerns but also saves calculation costs and improves emotion recognition accuracy. **Figure 4** below explains its procedures.

The proposes algorithm obeys the following two principles: (1) thoroughly utilizing different modality data, avoiding any wastes, and (2) choosing the most relevant features in high-dimensional spaces to discover the internal connections among features. It can provide more excellent classification accuracy than other algorithms.

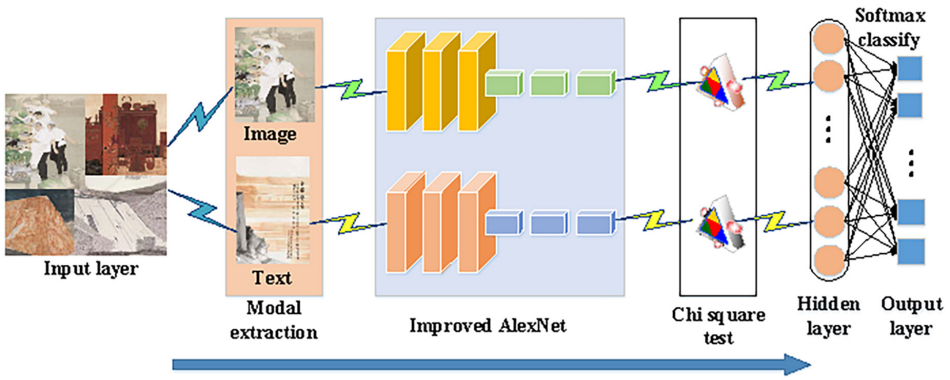


FIGURE 4 | Procedures of the proposed multimodal emotion recognition algorithm based on Chi Square Test and improved AlexNet processing Chinese paintings at the Thirteenth National Exhibition of Fines Arts in China.

During training, one local normalization layer is attached after AlexNet's pooling layer to standardize the feature map $c_t^l(i,j)$:

$$c_t^l(i,j) = a_t^l(i,j) / \left(k + \alpha \sum_{\max(0,t-m/2)}^{\min(N-1,t+m/2)} \left(a_t^l(i,j) \right)^2 \right)^{\beta} \quad (18)$$

In (18), k, α, β, m are hyperparameters valuing 2, 0.78, 10^{-4}, and 7, respectively, and N is the total number of convolution kernels in the l-th convolutional layer. To prevent "gradient dispersion" (Xu and Li, 2021) ReLU is employed to activate the convolution output $S_t^l(i,j)$:

$$y_t^l(i,j) = f\left(S_t^l(i,j)\right) = \max\left\{0, S_t^l(i,j)\right\} \quad (19)$$

In (19), $f()$ represents ReLU. To avoid over-fitting in the fully connected layer, the dropout parameter is set to 0.5.

While improving the generalization ability, neurons C^{ll} in the fully connected layer are discarded and output, $r_j^l \sim bernoulli(dp)$, $\tilde{C}^l = r^l C^l$; in that case, the i-th neuron input in the fully connected layer Z_i^{l+1} is $W_i^{l+1}\tilde{C}^l + b_i^{l+1}$, where the i-th neuron input in the next fully connected layer C_i^l is $f\left(Z_i^l\right)$, namely $\max\left\{0, Z_i^l\right\}$. Eventually, the input q^i of the i-th neuron in the fully connected layer can be obtained:

$$q^i = soft\max(Z_i^8) = \frac{e^{Z_i^8}}{\sum_{j=1}^{12} e^{Z_i^8}} \quad (20)$$

Here, the cross-entropy loss function suitable for classification is taken as the algorithm's error function, and the equation is:

$$Loss = \sum_{i=1}^{K} y_i \cdot \log\left(p_i\right) \quad (21)$$

$$p_i = \frac{\exp\left(\tilde{y}_i\right)}{\sum_{i=1}^{K} \exp\left(\tilde{y}_j\right)} \quad (22)$$

In (21) and (22), N signifies the number of categories, y_i represents the actual category distribution of samples, $\tilde{y}_i$ describes the network output, and p_i denotes the result after the SoftMax classifier. SoftMax's input is an N-dimensional real number vector, denoted as x. Its equation is:

$$\xi(x)_i = \frac{e^{x_i}}{\sum_{n=1}^{N} e^{x_i}}, i = 1, 2, ..., N \quad (23)$$

Essentially, SoftMax can map an M-dimensional arbitrary real number vector to an M-dimensional vector whose values all fall in the range of (0,1), thereby normalizing the vector. To reduce the computational amount, the output data volume is reduced to 2^8 through μ companding conversion, that is, $\mu = 255$, thereby improving the algorithm's forecasting efficiency.

$$f(x_t) = sign(x_t) \frac{\ln(1+\mu|x_t|)}{\ln(1+\mu)}, |x_t| < 1 \quad (24)$$

The proposed algorithm is trained through learning rate updating using the polynomial decay approach (Poly) (Cai et al., 2020). The equation is:

$$init_lr \times \left(1 - \frac{epoch}{\max_epoch}\right)^{power} \quad (25)$$

In (26), the initial learning rate $init_lr$ is 0.0005 (or $5e^{-4}$), and power is set to 0.9. The Weighted Cross-Entropy (WCE) is accepted as the cost function to optimize the algorithm training process. Suppose that $z_k(x,\theta)$ describes the unnormalized logarithmic probability of pixel x in the k-th category under the given network parameter θ. In that case, the SoftMax function $p_k(x,\theta)$ is defined as:

$$p_k(x,\theta) = \frac{\exp\{z_k(x,\theta)\}}{\sum_{k'}^{K} \exp\{z_{k'}(x,\theta)\}} \quad (26)$$

In (27), K represents the total number of image categories. During forecasting, once equation (26) reaches the maximum, pixel x will be labeled as the k-th category, namely $k* = \arg\max\{P_k(x,\theta)\}$. A semantic segmentation task needs to sum the pixel data loss in each input mini-batch (Li et al., 2018; Guo et al., 2019). Suppose that N denotes the total number of

pixels in the training batch of image data, y_i refers to the real semantic annotation of the pixel x_i , and $p_k(x_i, \theta)$ describes the forecasted probability of pixel x_i belonging to the k-th semantic category, that is, the log-normalized probability, abbreviated as p_{ik} . In that case, the training process aims to find the optimal network parameter $\theta*$ by minimizing the WCE loss function $\ell(x, \theta)$, denoted as $\theta* = \min_{\theta} \ell(x, \theta)$. Training samples with unbalanced categories in brain images usually make the network emphasize some easily distinguishable categories, resulting in poor recognition on some more difficult samples. In this regard, the Online Hard Example Mining (OHEM) strategy (Wen et al., 2019) is adopted to optimize the network training process. The improved loss function is:

$$\ell(x, \theta) = -\frac{1}{\sum_{i=1}^{N}\sum_{k=1}^{K}\delta\left(y_i = k, p_{ik} < \eta\right)} \sum_{i=1}^{N}\sum_{k=1}^{K}\delta\left(y_i = k, p_{ik} < \eta\right)\log p_{ik} \quad (27)$$

In (27), $\eta \in (0, 1]$ refers to the predefined threshold, and $\delta(\cdot)$ describes the symbolic function, which will value 1 if the condition is met and 0 otherwise. The weighted loss function for brain image fusion is defined as:

$$\ell(x, \theta) = -\sum_{i=1}^{N}\sum_{k=1}^{K} w_{ik} q_{ik} \log p_{ik} \quad (28)$$

In (28), $q_{ik} = q(y_i = k|x_i)$ denotes the true label distribution of the k-th category of the pixel x_i , and w_{ik} refers to the weighting coefficient. The following strategy is employed during training:

$$w_{ik} = \frac{1}{\ln(c + p_{ik})} \quad (29)$$

In (29), c is an additional hyperparameter, set to 1.10 based on experience in the simulation experiments of the present work.

Procedures of the proposed algorithm are illustrated in **Figure 5**.

Start
Input: Unimodal features matrix *Xi*
Output : Unimodal features matrix *Zi* after Improved AlexNet
$Z_i \leftarrow \varnothing$
for $t:\left[1,\cdots,L_i\right], i = 1,2,...,N$ **do**
"Poly" learning rate adjustment:
$init_lr \times \left(1 - \frac{epoch}{\max_epoch}\right)^{power}$
$init_lr \leftarrow 5e^{-4}$
$power \leftarrow 0.9$
Searching for the optimal network parameter θ^* by cross entropy loss function $\ell(x, \theta)$;
$\theta^* \leftarrow \min_{\theta} \ell(x, \theta)$
$\ell(x, \theta) \leftarrow -\sum_{i=1}^{N}\sum_{k=1}^{K} w_{ik} q_{ik} \log p_{ik}$
$q_{ik} \leftarrow q(y_i = k \mid x_i)$
$w_{ik} \leftarrow \frac{1}{\ln(c + p_{ik})}$
$c \leftarrow 1.10$
$z_t \leftarrow \mathrm{Re}lu\left(W_z h_t + b_z\right)$//Fully connected layer
$prediction \leftarrow soft\max\left(W_{sft} z_t + b_{sft}\right)$ // Emotion prediction
$Z_i \leftarrow Z_i \cup z_t$
return Z_i
end

FIGURE 5 | Procedures of the proposed algorithm based on Chi Square Test and improved AlexNet.

Simulation Experiments

MATLAB is adopted in the present work to validate the performance of the proposed algorithm. Image data utilized in the simulation experiments come from Chinese paintings at the Thirteenth National Exhibition of Fines Arts in China; they are divided into a training dataset and a test dataset in 7:3. The ratio of each data type in the two datasets shall be consistent. Hyperparameters of the neural network are set as follows: 0.5 dropout, 300 fully connected layer, 120 iterations, and 40 mini-batch to avoid over-fitting. Some state-of-art algorithms are included for performance comparison, including LSTM (Wen et al., 2019), CNN (Ul Haq et al., 2019), RNN (Zhang et al., 2018), AlexNet (Sajjad and Kwon, 2020), and DNN (Jain et al., 2020). Experimental environment configuration includes software and hardware. Regarding software, the operating system is Linux 64bit, the Python version is 3.6.1, and the development platform is PyCharm. Regarding hardware, the Central Processing Unit (CPU) is Intel Core i7-7700@4.2GHz 8 Cores, the internal memory is Kingston DDR4 2,400 MHz 16G, and the Graphics Processing Unit (GPU) is NVIDIA GeForce 1060 8G.

Performance evaluation metrics include Accuracy, Precision, Recall, and F1 score, calculated through the following equations:

$$Acc = \frac{\sum_{i=1}^{l} \frac{TP_i + TN_i}{TP_i + FP_i + TN_i + FN_i}}{l} \quad (30)$$

$$\mathrm{Pr}ecision = \frac{\sum_{i=1}^{l} \frac{TP_i}{TP_i + FP_i}}{l} \quad (31)$$

$$Recall = \frac{\sum_{i=1}^{l} \frac{TP_i}{TP_i + FN_i}}{l} \quad (32)$$

$$F1 = \frac{2\mathrm{Pr}ecision \cdot Recall}{\mathrm{Pr}ecision + Recall} \quad (33)$$

In (30) ∼ (33), *TP* represents the number of positive samples forecasted to be positive, *FP* represents the number of negative samples forecasted to be positive, *FN* represents the number of positive samples forecasted to be negative,

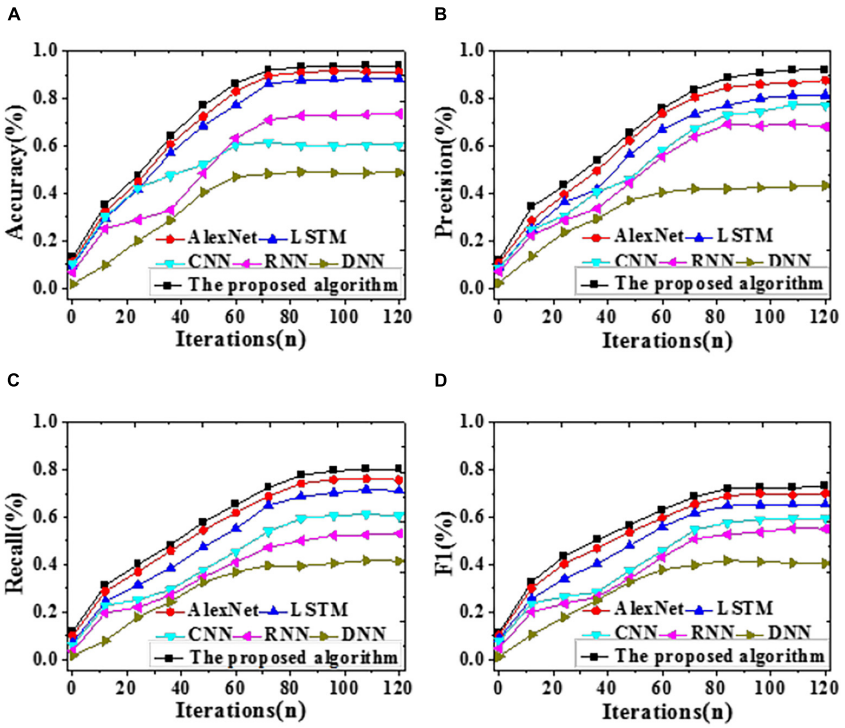


FIGURE 6 | The recognition accuracy of different algorithms with iteration on the training dataset [**(A)** Accuracy; **(B)** Precision; **(C)** Recall; **(D)** F1 score].

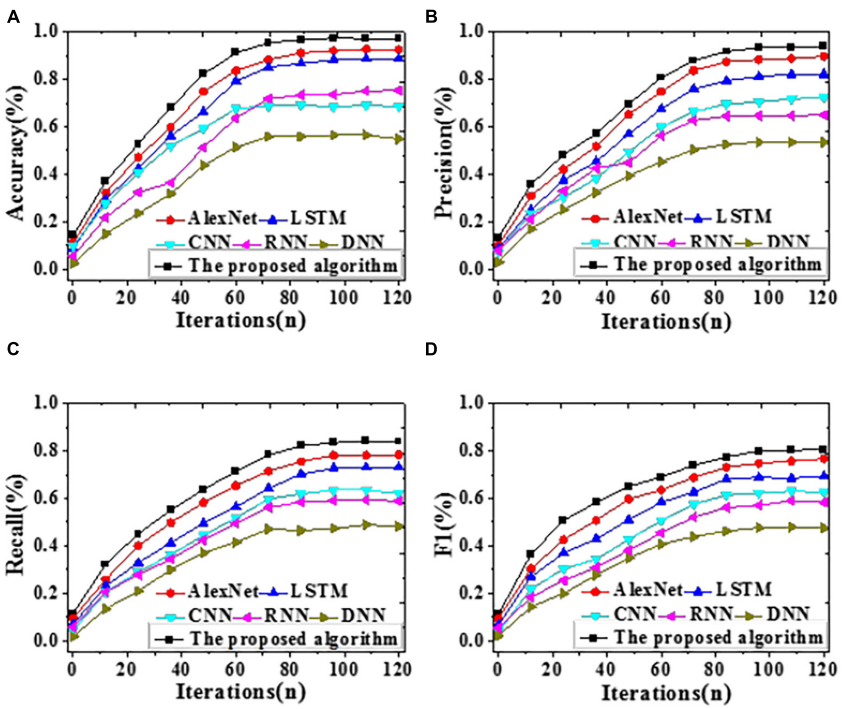


FIGURE 7 | The recognition accuracy of different algorithms with iteration on the test dataset [**(A)** Accuracy; **(B)** Precision; **(C)** Recall; **(D)** F1 score].

and *TN* represents the number of negative samples forecasted to be negative. Accuracy measures the overall classification accuracy, that is, the proportion of samples that forecasted correctly. Recall measures the coverage of positive samples, that is, the proportion of correctly classified positive samples to the total number of positive samples. Precision represents the proportion of examples classified as positive examples to actual positive examples. The most commonly used method is F1 score, which is the weighted harmonic average of precision and recall.

Meanwhile, the emotion recognition efficiency of the algorithm is analyzed from the perspective of SpeedUp. SpeedUp is the ratio of time consumed by the same task running in single processor system and parallel processor system, used to measure the performance and effect of parallel system or program parallelization.

RESULTS AND DISCUSSION

Analysis of Forecasting Performance

To analyze the forecasting performance, the proposed algorithm, LSTM, CNN, RNN, AlexNet, and DNN are included for comparative simulation. **Figures 6**, **7** visualize the results of forecasting accuracy (Accuracy, Precision, Recall, and F1 score). **Tables 1**, **2** illustrate the results of time durations required for training and testing.

As shown in **Figure 6**, on the training dataset, the proposed algorithm can provide an Accuracy of 92.23%, reaching an improvement of at least 4.56% over other algorithms. Remarkably, its Precision, Recall, and F1 score are also the best, at least 3.03% higher than others. To sum up, the proposed algorithm can provide a notably better forecasting performance than DNN, LSTM, AlexNet, RNN, and CNN on the training dataset.

According to **Figure 7**, on the test dataset, the proposed algorithm can provide an Accuracy of 97.11%, reaching an improvement of at least 4.66% over other algorithms. Remarkably, its Precision, Recall, and F1 score are also the best, at least 0.44% higher than others. Hence, the proposed algorithm can provide a remarkably better forecasting performance than DNN, LSTM, AlexNet, RNN, and CNN on the test dataset. **Tables 1**, **2** below illustrate the results of time durations required on the training and test datasets.

Time durations required by all algorithms first decrease sharply then tend to stabilize with epochs; that is, the algorithms converge. In particular, the proposed algorithm requires a training duration of 54.97 s and a testing duration of 23.74 s, remarkably shorter than other algorithms. The proposed algorithm takes less time to make forecasts because of its enhanced generalization ability and accelerated algorithm convergence. In the meantime, the Chi Square Test is specific to emotion recognition, which reduces the time required again. To sum up, the proposed algorithm can achieve higher forecasting accuracy more quickly than other simulated algorithms.

TABLE 1 | Time duration required by different algorithms on the training dataset(s).

	Epochs		
	1.00	**60.00**	**120.00**
The proposed algorithm	90.50	57.91	54.97
AlexNet	90.50	63.44	60.09
LSTM	90.25	67.97	65.87
CNN	91.00	71.49	70.90
RNN	91.51	75.51	75.42
DNN	90.50	80.28	80.20

TABLE 2 | Time duration required by different algorithms on the test dataset(s).

	Epochs		
	1	**60**	**120**
The proposed algorithm	57.81	25.04	23.74
AlexNet	58.15	26.13	24.44
LSTM	57.31	31.03	27.90
CNN	57.65	34.89	29.92
RNN	58.49	39.26	36.81
DNN	59.66	44.64	42.69

Analysis of Acceleration Efficiency

The acceleration efficiency is tested on the training and the test datasets, respectively. Results of time delay error for Chinese painting emotion recognition are visualized in **Figure 8**. The time required and speedup ratio of different algorithms under different data volumes are presented in **Figures 9**, **10**.

As shown in **Figure 8**, errors gradually reduce with iterations on both the training and the test datasets. DNN provides the longest time delay, reaching 455.91 and 356.21 ms, respectively. In contrast, the proposed algorithm provides a time delay approaching zero, the smallest among all simulated algorithms.

According to **Figures 9**, **10**, the proposed algorithm is less sensitive to data growth than other algorithms. Hence, it is suitable to process massive data; the larger the data volume, the higher the speedup ratio, and the greater the acceleration ratio. All algorithms provide slightly better acceleration efficiencies on the test dataset than the training dataset probably because of the absent emotion recognition and analysis path for Chinese paintings during training. Through the adaptive learning, neural networks can process massive amounts of emotion data in Chinese paintings, which remarkably increases the efficiency. To sum up, the proposed algorithm can better recognize and classify human affects in Chinese paintings than other algorithms.

Aiming at the abstract phenomenon of emotion expressed in the field of Chinese painting, the study proposes a multimodal emotion recognition model of Chinese painting based on improved AlexNet neural network and chi square test. The simulation analysis reveals that the recognition and prediction accuracy is significantly higher than that of LSTM, CNN, RNN, AlexNet and DNN model algorithms proposed by scholars in related fields. Among them, the accuracy of the proposed algorithm in the test set reaches 97.11%, and the required time

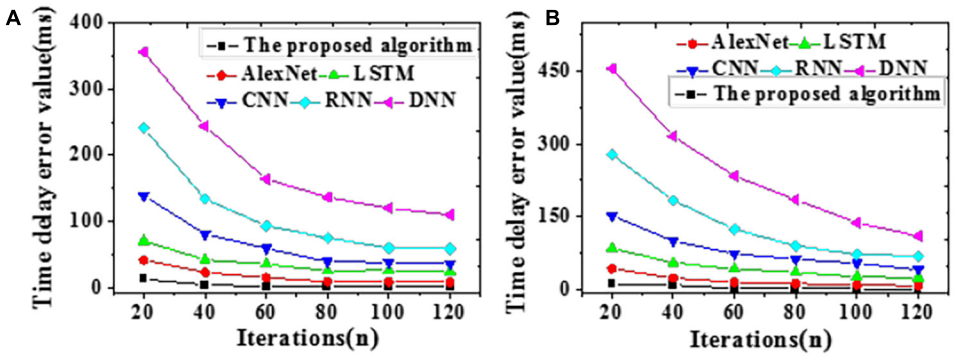


FIGURE 8 | The time delay error of different algorithms for Chinese painting emotion recognition [**(A)** The test dataset; **(B)** The training dataset].

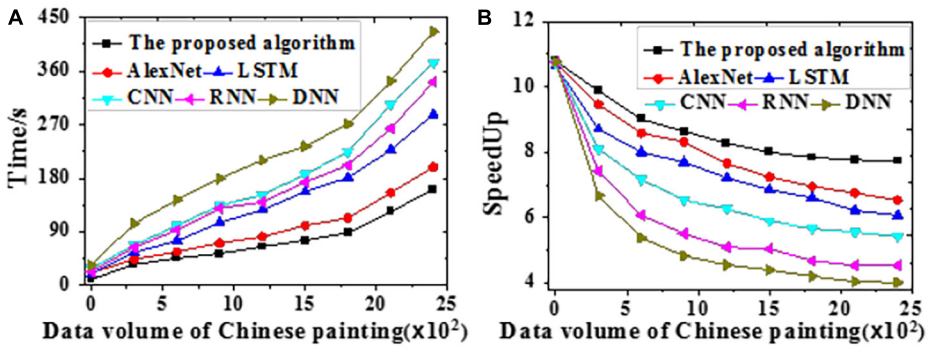


FIGURE 9 | The time required and speedup ratio of different models under different data volumes on the training dataset [**(A)** Time required; **(B)** Speedup ratio].

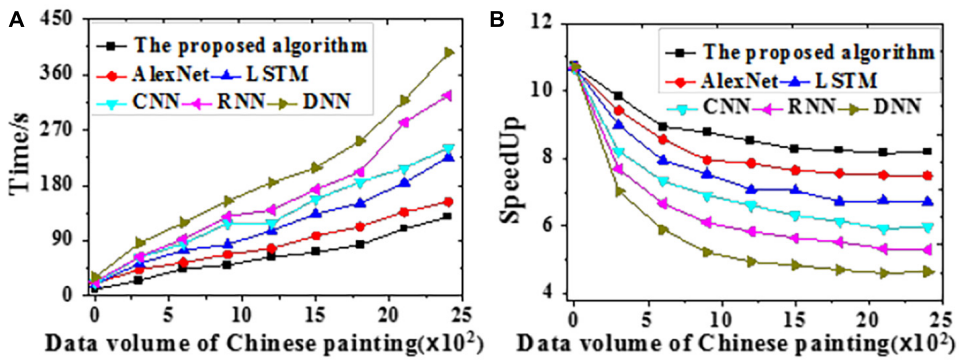


FIGURE 10 | The time required and speedup ratio of different models under different data volumes on the test dataset [**(A)** Time required; **(B)** Speedup ratio].

is only 23.74 s. This may be because the improved AlexNet neural network proposed not only enhances the generalization ability, but also accelerates the convergence rate of the model training process. Moreover, the chi square test is targeted for emotion recognition, which reduces the time required for emotion recognition again. Regarding the recognition efficiency, it is also obvious that the acceleration ratio of the proposed algorithm is higher. This may be because the model has not formed an emotion recognition and analysis path corresponding to Chinese painting in the training process. After autonomous learning in the training process, the neural network can better analyze many emotions of Chinese painting, and the efficiency has been significantly improved. Therefore, the algorithm proposed has a good effect on emotion recognition of Chinese painting.

CONCLUSION

Art such as Chinese painting aims to express the esthetics and emotion of works through visual art elements such as

color, line, and shape. Thus, emotion recognition of images is not only conducive to the management of art information, but also can promote the popularization and promotion of art. To solve "semantic gap" in the emotion recognition task of images such as traditional Chinese painting, this study constructs a multimodal emotion recognition model of Chinese painting based on improved AlexNet neural network and chi square test. The simulation reflects that the emotion recognition accuracy of the improved AlexNet neural network combined with chi square test algorithm model proposed reaches 92.23 and 97.11% in the training set and test set, respectively. The training and test time are stable at about 54.97 and 23.74 s, and the acceleration efficiency is obviously better than other algorithms, which can provide experimental reference for the management, popularization, and promotion of art information in the later stage. However, there are still some deficiencies. First, due to the great differences in the characteristics of traditional Chinese painting in different techniques and content categories, it is difficult to use the same algorithm to identify emotion. In the future, the emotion recognition algorithm for various categories of traditional Chinese painting will be further discussed. Additionally, the mathematical model of traditional Chinese painting emotion will be established, and the traditional Chinese painting emotion will be calculated and processed more accurately to create greater value in the digital management and protection of national quintessence. Second, this study models and analyzes the emotion of images such as Chinese painting in the 13th National Exhibition of Fine Arts. But it is not clear to what extent the proposed method and its theory can be applied to other types of painting and natural images. Therefore, future work will focus on evaluating the effectiveness of this method except for the field of various categories of art.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

JL and DC designed the multi-modal emotion recognition algorithm of Chinese paintings. ZZ and NY conducted the evaluation and analysis of experimental data, compared other algorithms include LSTM, CNN, RNN, AlexNet, and DNN under the guidance of ZL. All authors participated in the process of writing and revising the manuscript.

FUNDING

This work was the concluding outcome of the Key Subject of Shandong Art Science under grant number QN202008253, Shandong Art Education Special Project under grant number ZY20201335, and Shandong Social Science Planning Project under grant number 21CWYJ16.

REFERENCES

Avots, E., Sapiński, T., Bachmann, M., and Kamińska, D. (2019). Audiovisual emotion recognition in wild. *Mach. Vis. Appl.* 30, 975–985. doi: 10.1007/s00138-018-0960-9

Bi, H., Xu, F., Wei, Z., Xue, Y., and Xu, Z. (2019). An active deep learning approach for minimally supervised polsar image classification. *IEEE Trans. Geosci. Remote Sens.* 57, 9378–9395. doi: 10.1109/tgrs.2019.2926434

Cai, H., Qu, Z., Li, Z., Zhang, Y., Hu, X., and Hu, B. (2020). Feature-level fusion approaches based on multimodal EEG data for depression recognition. *Inf. Fusion* 59, 127–138. doi: 10.1016/j.inffus.2020.01.008

Chen, M., Jiang, Y., Cao, Y., and Zomaya, A. Y. (2020). CreativeBioMan: a Brain- and Body-Wearable, Computing-Based, Creative Gaming System. *IEEE Syst. Man Cybern. Mag.* 6, 14–22. doi: 10.1109/msmc.2019.2929312

Ćosić, K., Popović, S., Šarlija, M., Mijić, I., Kokot, M., Kesedžić, I., et al. (2019). New tools and methods in selection of air traffic controllers based on multimodal psychophysiological measurements. *IEEE Access* 7, 174873–174888. doi: 10.1109/access.2019.2957357

Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., et al. (2019). Ce-net: context encoder network for 2d medical image segmentation. *IEEE Trans. Med. Imaging* 38, 2281–2292. doi: 10.1109/tmi.2019.2903562

Guo, Z., Li, X., Huang, H., Guo, N., and Li, Q. (2019). Deep learning-based image segmentation on multimodal medical imaging. *IEEE Trans. Radiat. Plasma Med. Sci.* 3, 162–169.

Haberl, M. G., Christopher, C., Lucas, T., Daniela, B., Sébastien, P., Bushong, E. A., et al. (2018). Cdeep3m—plug-and-play cloud-based deep learning for image segmentation. *Nat. Methods* 15, 677–680. doi: 10.1038/s41592-018-0106-z

Hammad, M., Liu, Y., and Wang, K. (2018). Multimodal biometric authentication systems using convolution neural network based on different level fusion of ECG and fingerprint. *IEEE Access* 7, 26527–26542. doi: 10.1109/access.2018.2886573

He, X., and Zhang, W. (2018). Emotion recognition by assisted learning with convolutional neural networks. *Neurocomputing* 291, 187–194. doi: 10.1016/j.neucom.2018.02.073

Hwang, S., Hong, K., Son, G., and Byun, H. (2020). Learning CNN features from DE features for EEG-based emotion recognition. *Pattern Anal. Appl.* 23, 1323–1335. doi: 10.1007/s10044-019-00860-w

Jain, D. K., Zhang, Z., and Huang, K. (2020). Multi angle optimal pattern-based deep learning for automatic facial expression recognition. *Pattern Recognit. Lett.* 139, 157–165. doi: 10.1016/j.patrec.2017.06.025

Jain, N., Kumar, S., Kumar, A., Shamsolmoali, P., and Zareapoor, M. (2018). Hybrid deep neural networks for face emotion recognition. *Pattern Recognit. Lett.* 115, 101–106. doi: 10.1016/j.patrec.2018.04.010

Jia, H., Peng, X., Song, W., Lang, C., and Sun, K. (2019). Multiverse optimization algorithm based on lévy flight improvement for multithreshold color image segmentation. *IEEE Access* 7, 32805–32844. doi: 10.1109/access.2019.2903345

Kim, W., Kanezaki, A., and Tanaka, M. (2020). Unsupervised learning of image segmentation based on differentiable feature clustering. *IEEE Trans. Image Process.* 29, 8055–8068. doi: 10.1109/tip.2020.3011269

Li, L., Fu, H., and Tai, C. L. (2018). Fast sketch segmentation and labeling with deep learning. *IEEE Comput. Graph. Appl.* 39, 38–51. doi: 10.1109/mcg.2018.2884192

Lin, Z., Chen, Y., Ghamisi, P., and Benediktsson, J. A. (2018). Generative adversarial networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 56, 5046–5063.

Liu, S., Yang, J., Agaian, S. S., and Yuan, C. (2021). Novel features for art movement classification of portrait paintings. *Image Vis. Comput.* 108:104121. doi: 10.1016/j.imavis.2021.104121

Nalepa, J., Myller, M., and Kawulok, M. (2019). Validating hyperspectral image segmentation. *IEEE Geosci. Remote Sens. Lett.* 16, 1264–1268. doi: 10.1109/lgrs.2019.2895697

Sajjad, M., and Kwon, S. (2020). Clustering-based speech emotion recognition by incorporating learned features and deep BiLSTM. *IEEE Access* 8, 79861–79875. doi: 10.1109/access.2020.2990405

Spoorthi, G. E., Gorthi, S., and Gorthi, R. (2018). Phasenet: a deep convolutional neural network for two-dimensional phase unwrapping. *IEEE Signal Process. Lett.* 26, 54–58. doi: 10.1109/lsp.2018.2879184

Sukumarana, A., and Mb, A. (2019). A Brief Review of Conventional and Deep Learning Approaches in Facial Emotion Recognition. *Artificial Intelligence for Internet of Things* 101.

Ul Haq, I., Ullah, A., Muhammad, K., Lee, M. Y., and Baik, S. W. (2019). Personalized movie summarization using deep cnn-assisted facial expression recognition. *Complexity* 2019, 1–10. doi: 10.1155/2019/3581419

Walia, G. S., and Rohilla, R. (2021). A Contemporary Survey of Multimodal Presentation Attack Detection Techniques: challenges and Opportunities. *SN Comput. Sci.* 2:49.

Wang, G., Li, W., Zuluaga, M. A., Pratt, R., Patel, P. A., Aertsen, M., et al. (2017). Interactive medical image segmentation using deep learning with image-specific fine-tuning. *IEEE Trans. Med. Imaging* 37, 1562–1573. doi: 10.1109/tmi.2018.2791721

Wang, Z., Lian, J., Song, C., Zhang, Z., Zheng, W., Yue, S., et al. (2019). SAS: painting Detection and Recognition via Smart Art System With Mobile Devices. *IEEE Access* 7, 135563–135572. doi: 10.1109/access.2019.2941239

Wei, J., Yang, X., and Dong, Y. (2021). User-generated video emotion recognition based on key frames. *Multimed. Tools Appl.* 80, 14343–14361. doi: 10.1007/s11042-020-10203-1

Wen, S., Wei, H., Yang, Y., Guo, Z., Zeng, Z., Huang, T., et al. (2019). Memristive LSTM network for sentiment analysis. *IEEE Trans. Syst. Man Cybern. Syst.* 51, 1794–1804.

Xu, Y., and Li, Q. (2021). Music Classification and Detection of Location Factors of Feature Words in Complex Noise Environment. *Complexity* 2021, 1–12. doi: 10.1155/2021/5518967

Yang, D., Ye, X., and Guo, B. (2021). Application of Multitask Joint Sparse Representation Algorithm in Chinese Painting Image Classification. *Complexity* 2021, 1–11. doi: 10.1155/2021/5546338

Yuan, J., Feng, Z., Dong, D., Meng, X., Meng, J., and Kong, D. (2020). Research on Multimodal Perceptual Navigational Virtual and Real Fusion Intelligent Experiment Equipment and Algorithm. *IEEE Access* 8, 43375–43390. doi: 10.1109/access.2020.2978089

Zhang, J., Miao, Y., Zhang, J., and Yu, J. (2020). Inkthetics: a Comprehensive Computational Model for Aesthetic Evaluation of Chinese Ink Paintings. *IEEE Access* 8, 225857–225871. doi: 10.1109/access.2020.3044573

Zhang, T., Zheng, W., Cui, Z., Zong, Y., and Li, Y. (2018). Spatial–temporal recurrent neural network for emotion recognition. *IEEE Trans. Cybern.* 49, 839–847. doi: 10.1109/tcyb.2017.2788081

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Li, Chen, Yu, Zhao and Lv. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 22 October 2021
doi: 10.3389/fpsyg.2021.762795

MIFAD-Net: Multi-Layer Interactive Feature Fusion Network With Angular Distance Loss for Face Emotion Recognition

Weiwei Cai[1,2], *Ming Gao*[3], *Runmin Liu*[1] *and Jie Mao*[1]*

[1] College of Sports Engineering and Information Technology, Wuhan Sports University, Wuhan, China, [2] School of Logistics and Transportation, Central South University of Forestry and Technology, Changsha, China, [3] College of Sports Science and Technology, Wuhan Sports University, Wuhan, China

Understanding human emotions and psychology is a critical step toward realizing artificial intelligence, and correct recognition of facial expressions is essential for judging emotions. However, the differences caused by changes in facial expression are very subtle, and different expression features are less distinguishable, making it difficult for computers to recognize human facial emotions accurately. Therefore, this paper proposes a novel multi-layer interactive feature fusion network model with angular distance loss. To begin, a multi-layer and multi-scale module is designed to extract global and local features of facial emotions in order to capture part of the feature relationships between different scales, thereby improving the model's ability to discriminate subtle features of facial emotions. Second, a hierarchical interactive feature fusion module is designed to address the issue of loss of useful feature information caused by layer-by-layer convolution and pooling of convolutional neural networks. In addition, the attention mechanism is also used between convolutional layers at different levels. Improve the neural network's discriminative ability by increasing the saliency of information about different features on the layers and suppressing irrelevant information. Finally, we use the angular distance loss function to improve the proposed model's inter-class feature separation and intra-class feature clustering capabilities, addressing the issues of large intra-class differences and high inter-class similarity in facial emotion recognition. We conducted comparison and ablation experiments on the FER2013 dataset. The results illustrate that the performance of the proposed MIFAD-Net is 1.02–4.53% better than the compared methods, and it has strong competitiveness.

Keywords: face emotion, emotion recognition, multi-layer interactive, feature fusion, deep learning, neural networks

OPEN ACCESS

Edited by:
Yaoru Sun,
Tongji University, China

Reviewed by:
Yuanpeng Zhang,
Nantong University, China
Yufeng Yao,
Changshu Institute of Technology, China

***Correspondence:**
Jie Mao
maojie@whsu.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *22 August 2021*
Accepted: *23 September 2021*
Published: *22 October 2021*

Citation:
Cai W, Gao M, Liu R and Mao J (2021) MIFAD-Net: Multi-Layer Interactive Feature Fusion Network With Angular Distance Loss for Face Emotion Recognition. Front. Psychol. 12:762795. doi: 10.3389/fpsyg.2021.762795

INTRODUCTION

Emotions are extremely important in everyday life. It is often necessary to accompany the correct understanding of other people's emotions in the process of human daily communication and behavior judgment, and facial expressions contain a lot of information about emotions and mental states. Therefore, it is possible to say that recognizing facial expressions (Crivelli et al., 2017; Chengeta and Viriri, 2019; González-Lozoya et al., 2020) is the key to understanding

emotions. According to psychologists' research, only 7% of information in the process of human communication comes from pure language expression, 38% from sound information such as speech pitch, and 55% from visuals such as facial emotions. The content has been communicated. As a result, accurate recognition of facial expressions is critical for understanding information in human communication.

Facial emotion recognition (Sreedharan et al., 2018; Jain et al., 2019) can be used in a variety of situations. In terms of human-computer interaction, accurate facial expression recognition to determine human emotions can make machines more appropriate, accurate, and effective in interacting with humans, resulting in a more natural interaction. Interaction and exchange with humans. In terms of security scenarios, it is possible to effectively identify suspects with criminal intent in public by accurately identifying facial expressions and subtle expressions. In terms of transportation, it is possible to better judge whether a driver is fatigued by recognizing the facial expressions of drivers of vehicles such as vehicles (Theagarajan et al., 2017; Zepf et al., 2020). Furthermore, facial expression recognition has gotten a lot of attention in the advertising (Hamelin et al., 2017) and marketing, automation, and communications fields.

In recent years, facial emotion recognition based on deep learning technology (Cai and Wei, 2020; Cai et al., 2021; Zhang et al., 2021) has made great progress, but there are still many problems to be solved. For example, the recognition accuracy in real scenes is still not ideal. Among the basic emotion categories of human faces, negative emotions, including angry, disgust, disappointment, etc., have no relatively uniform standard for facial expressions, and feature differences are minimal, which are not conducive to computer feature learning and are often difficult to correctly recognize. Furthermore, because the face area occupies a relatively small area in an image, the data used for facial emotion recognition model training has a small input size. The current convolutional neural network (CNN) model (Bendjoudi et al., 2020; Kollias and Zafeiriou, 2020; Kwon, 2021) necessitates a relatively large image size as input. Excessive use of interpolation and other methods to increase image size results in more calculations. On the contrary, the recognition effect has not improved significantly.

Based on the above observations, in this paper, a multi-layer and multi-scale module is designed to extract the global and local features of facial expressions to capture part of the feature relationships between different scales, thereby enhancing the model's ability to discriminate subtle features of facial expressions. Secondly, in view of the problem of loss of useful feature information due to layer-by-layer convolution and pooling of CNNs, a hierarchical interactive feature fusion module is designed. The attention mechanism (Gao et al., 2021; Liu et al., 2021) is used between convolutional layers at different levels to control the network. Strengthen the saliency information of different characteristics in the Internet and suppress irrelevant information, thereby improving the discriminative ability of the network. Finally, for the problem of large intra-class differences and high inter-class similarity in facial expression recognition, we use the angular distance loss function to improve the capabilities of the proposed algorithm for feature separation between classes and clustering of features within classes.

The main innovations of this paper are as follows:

(1) To address the problem of subtle differences in facial emotion causing difficulty in classification, we designed a multi-layer and multi-scale module to extract global and local facial emotion features to capture partial feature relationships between different scales, thereby improving the model's ability to discriminate subtle facial emotion features.
(2) To address the issue of loss of useful feature information caused by layer-by-layer convolution and pooling of convolutional neural networks, we created a hierarchical interactive feature fusion module that controls the network using the attention mechanism between convolutional layers at different levels. The importance of various characteristics is increased, while irrelevant information is reduced.
(3) We use the angular distance loss function to improve the capabilities of the proposed algorithm for feature separation between classes and clustering of features within classes, with the goal of addressing the problem of large intra-class differences and high inter-class similarity in face emotion recognition.

The remainder of this article is organized as follows. We introduce relevant work in section Related work, describe the proposed algorithm in section Methodology, and present the experimental results in section Experiments and Results. This paper's research conclusions are presented in section Conclusion.

RELATED WORK

Emotion Recognition Based on Traditional Machine Learning

The emotion recognition method based on traditional machine learning (Bota et al., 2019; Kerkeni et al., 2019; Domínguez-Jiménez et al., 2020) is mainly to manually extract the emotion image features and then use the appropriate classification algorithm to classify the emotion. The specific method is to manually select some appropriate feature extraction operators to extract facial features, and then do appropriate dimension reduction processing on the extracted facial features, and finally select a classifier to classify the facial features after dimension reduction. Kumar et al. (2016) believed that features in different regions of the face contribute to expression recognition to different extent, and important locations have important feature information, such as mouth and eyes. Therefore, they proposed a weighted projection LBP feature extraction algorithm for different information regions, and improved the accuracy of expression recognition by cascading the weighted features of different regions. As a linear filter, Gabor is robust to light changes, and it can also change the frequency and direction to analyze texture features. Zhang et al. (2014) proposed an emotion recognition method in the case of occlusion. This method adopted Monte Carlo algorithm to extract features based on Gabor template in the image, and the features obtained were robust for occlusion. Harit et al. (2018) proposed an

automatic expression recognition algorithm that constructed an expression feature space using multiple Gabor filters at reference points before sending it to a neural network for classification. Wang et al. (2017) developed a multi-scale geometric feature extraction method, which mapped the original expression information to geometric feature functions, and then used the feature functions for further analysis. Tarannum et al. (2016) took the Euclidean distance between various facial regions as a feature and then used Canberra distance to classify the features.

After the expression feature extraction is completed, the classifier can specifically classify the expression feature into a certain expression category. Common facial expression classifiers include SVM algorithm (Xu et al., 2012), KNN algorithm and so on. Liew and Yairi (2015) proposed to use SVM as a classifier to classify the Hog expression features extracted in the early stage. This method has achieved good results on the JAFFE data set. Ouellet (2014) replaced the sofhnax classification layer of the Alexnet network with SVM Multi-classifiers, and achieved better recognition results on the CK+ expression library. Rieger et al. (2014) used a pattern recognition paradigm with spectral feature extraction and a set of KNN classifiers to investigate speech-based emotion recognition, and found that using two KNNs yielded the best results.

Emotion Recognition Based on Deep Learning

CNNs, in contrast to traditional machine learning methods, can automatically extract deep-level features of facial expressions by constructing multiple convolutional layers. On the one hand, it avoids errors caused by artificial feature extraction, on the other hand, it has strong robustness and generalization ability, so it has gradually become the mainstream method. Researchers are beginning to study applying deep learning to facial expression recognition tasks. Mollahosseini et al. (2016) proposed a 7-layer convolutional neural network that combined AlexNet and GoogleNet models and then verified them using seven public expression data sets, which was faster and more accurate than a traditional convolutional neural network. Zhang et al. (2019) used a stacked hybrid autoencoder to recognize expression. Three encoders were used in the network structure: a denoising autoencoder, a sparse autoencoder, and an autoencoder. The feature extraction was done with the denoising encoder, and the sparse autoencoder was utilized for cascading to extract more abstract sparse features. Tang (2013) combined CNNs and svm, and used the hinge loss instead of the common cross-entropy loss function in convolutional neural networks. On the FER2013 data set, the detection rate was 71.2 percent, and the team won the Kaggle facial expression recognition competition in 2013. After analyzing the network structure of expression recognition based on deep CNNs, Pramerdorfer and Kampel (2016) improved the classic ResNet and input a single face image to extract facial expression features, achieving an average recognition rate of 72.4% on the FER2013 data set. Lee et al. (2019) proposed a deep network for context-aware emotion recognition, which not only uses human facial expressions, but also uses context information in a joint and enhanced manner, which effectively improves the performance of emotion recognition.

Although deep learning technology has achieved excellent results in face emotion recognition tasks, the differences caused by facial expression changes are very subtle, and different facial expression features are not distinguishable, resulting in low face emotion recognition accuracy.

METHODOLOGY

The MIFAD-Net model of this article is shown in **Figure 1**. The three columns show that networks with different thickness scales using convolutional scale kernels of 7, 5, and 3 can extract more refined facial emotion features. Each column of the network has six convolutional layers and five BN layers. The three-column network has a common facial emotion image input. After that, the feature maps of the last three convolutional layers of the same depth position of the three-column network are interacted through the feature splicing strategy to integrate different cross-layer features of the same network and different networks to capture the deep connection between different levels to facilitate subsequent facial emotions Feature classification. In addition, the three-column network also interacts with the collection of features through the addition strategy, and uses the attention mechanism to focus on the effective features. Finally, we are employing the angular distance loss function to improve the model's capacity to segregate features between classes and cluster features within classes, which is a challenge with big intra-class variances in facial expression recognition and high similarity between classes. Then, using Softmax, establish the face expression category. After that, we'll go over the proposed model in more detail.

Multi-Scale and Multi-Layer Interactive Feature Fusion

This paper proposes a multi-scale and multi-layer interactive feature fusion module, as shown in **Figure 2**, to make better use of the different scale features of facial emotion images. The module first merges the feature maps 3*3, 5*5, and 7*7 of the three coarse and fine scale networks through 3×3. One branch is activated by Sigmoid to generate feature weights and then multiplied by the three feature map elements to obtain the re-calibrated features. Figure, and finally get the final output through the feature splicing strategy. The module can self-update learning according to back propagation, and automatically select the multi-scale features that each branch needs to be fused.

Multi-Scale Convolution

The use of a multi-scale convolution kernel has two major advantages, as discussed in this article. First and foremost, the multi-scale convolution kernel has the advantage of allowing different-sized convolution kernels to extract multiple scales of facial emotion picture data, allowing the filter to extract and learn richer high-dimensional features. Second, the convolutional neural network trains the model by learning the filter's

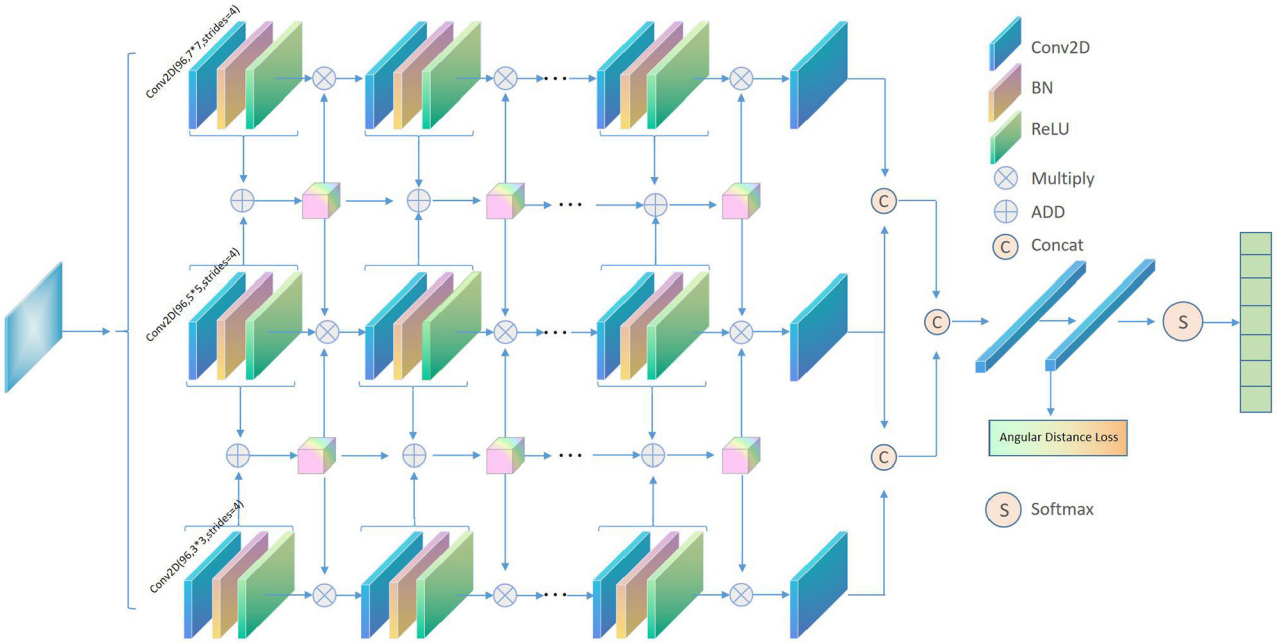


FIGURE 1 | Schematic diagram of the proposed MIFAD-Net algorithm.

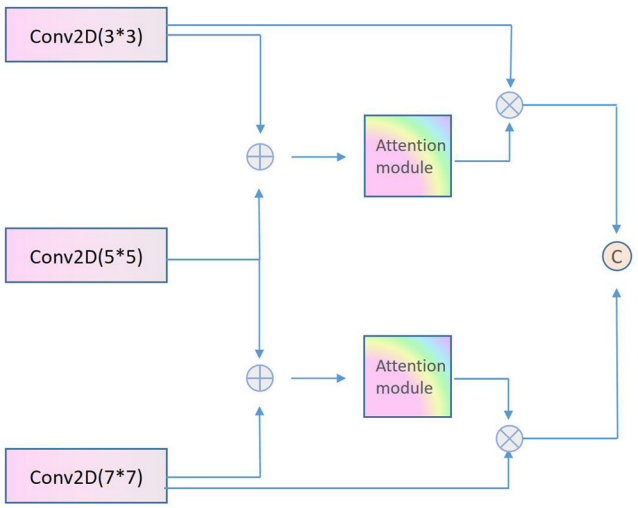


FIGURE 2 | Schematic diagram of multi-scale and multi-layer interactive feature fusion.

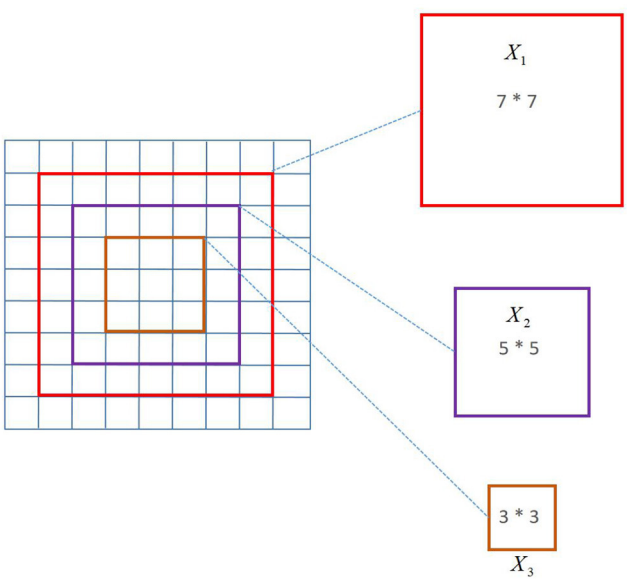


FIGURE 3 | Schematic diagram of multi-scale convolution.

parameters (weight and offset), i.e., continuously learning the filter's parameters to acquire the ideal value closest to the label. This article employs a multi-scale convolution kernel with the goal of allowing a single convolution layer to have several filters, so diversifying the weight and bias learning, and thereby extracting and learning the semantic aspects of facial emotion photos fully and efficiently. A schematic diagram of multi-scale convolution is shown in **Figure 3**.

To achieve the best results, multi-scale inference approaches are commonly used in computer vision models. Fine details are better predicted at larger sizes, larger objects are better predicted at smaller sizes, and the network's receiving field can interpret the scene better at smaller sizes. The 3*3, 5*5, and 7*7 scale convolution kernels were employed in this article's multi-scale

convolution. The following is the calculating formula:

$$Y^{H\times W} = \phi\left\{\sum_{i=0}^{n} w_{ij}^{H\times W} * x_i^{H\times W} + b_j\right\} \quad (1)$$

where $H \times W$ represents the size of the convolution kernel.

Attention Mechanism

Different areas in facial emotion images have different weights for different tasks. The higher the relevance to the task, the more important the field is. In this article, the attention module we designed is composed of cascaded channel attention and spatial attention. The schematic diagram of the channel attention mechanism is shown in **Figure 4**.

The given input feature $F_{in} \in R^{C\times H\times W}$ is fed into the proposed channel attention module. First, global average pooling (GAP) and max-pooling are used to compress the feature map along the spatial axis in parallel to generate two $C \times 1 \times 1$ dimensional feature vectors F_a^c And F_m^c, and then perform element-wise summation to obtain the aggregate characteristic F_s^c of all features. After that, go through a convolutional layer with a kernel size of 1×1, and then execute PReLU and BatchNorm to get the middle feature map F_p^c, then:

$$F_p^c = \mathrm{Conv}\left(F_a^c \oplus F_m^c\right) = \phi^{1\times 1}\left(F_a^c \oplus F_m^c\right) \quad (2)$$

where $\oplus$ represents the element summation, and ϕ represents the convolution operation. Then, the F_p^c is deformed and transposed to obtain two feature maps with dimensions $C \times 1$ and $1 \times C$, and then matrix multiplication and softmax operations are performed to obtain the channel attention matrix A_c. The calculation equation is as follows:

$$A_c = soft\max\left(F_p^c \otimes \left(F_p^c\right)^T\right) \quad (3)$$

where $\otimes$ represents matrix multiplication, and the following equation can be obtained:

$$A_{c_{i,j}} = \frac{\exp\left(\left(F_p^c\right)_i \left(F_p^c\right)_j^T\right)}{\sum_{i=1}^{C}\exp\left(\left(F_p^c\right)_i \left(F_p^c\right)_j^T\right)} \quad (4)$$

where $A_{c_{i,j}}$ represents the influence of the i-th channel on the j-th channel. Finally, the input feature F_{in} is multiplied by the channel attention matrix A_c, and then the refined feature $F_C \in R^{C\times H\times W}$ of the channel is obtained through learning.

$$F_C = F_{in} \oplus (\alpha\,(A_c \otimes F_{in})) \quad (5)$$

where α is a learnable parameter, and generally the initial value is set to 0 to reduce the difficulty of the convergence process of the first few training cycles. In this way, the channel attention matrix A_c can be regarded as a kernel selector to select the filter used to describe the emotional characteristics of the face.

In addition to channel attention, we also cascade a spatial attention module to learn the relationship between the spatial structure of the intermediate feature maps. The spatial attention module, which can be used in conjunction with the channel attention module, generates a spatial attention matrix to focus attention on the part that best represents facial feature information. Apply average pooling and maximum pooling along the channel axis, cascade them to get an effective feature descriptor, and then use matrix calculation and softmax layer to perform convolution operation to get the final Note the matrix of space, following the same strategy as the channel attention module. The spatial attention module is depicted schematically in **Figure 5**.

Given a channel refinement feature $F_C \in R^{C\times H\times W}$, first pass the spatial attention module, use GAP and max-pooling to compress the feature map along the channel axis in parallel, and then obtain two feature vectors F_a^s and F_m^s with a dimension of $1 \times H \times W$. Then the channel cascade is performed to merge the aggregation characteristics of the F_s^s. After the channel cascade, the convolutional layer with the kernel size of 3×3 is performed first, and then the PReLU and BatchNorm operations are performed to obtain the intermediate feature map F_p^s. During the convolution process, the step size is set to 1, and the filling value is also 1. In order to ensure that the size of the feature map remains unchanged, then:

$$F_p^s = Conv\left(F_a^s; F_m^s\right) = \varphi^{3\times 3}\left(F_a^s; F_m^s\right) \quad (6)$$

where φ represents the convolution operation. Transform and transpose the middle F_p^s to obtain two feature maps of $HW \times 1$ and $1 \times HW$, and then perform matrix multiplication and softmax operations to obtain the spatial attention matrix A_s, and perform the softmax operation on each row of the spatial matrix, then:

$$A_s = soft\max\left(F_p^s \otimes \left(F_p^s\right)^T\right) \quad (7)$$

where $\otimes$ represents matrix multiplication, and the following equation can be obtained:

$$A_{s_{i,j}} = \frac{\exp\left(\left(F_p^s\right)_i \left(F_p^s\right)_j^T\right)}{\sum_{i=1}^{HW}\exp\left(\left(F_p^s\right)_i \left(F_p^s\right)_j^T\right)} \quad (8)$$

where $A_{s_{i,j}}$ represents the influence of the i-th channel on the j-th channel. Finally, the channel refined F_C is multiplied by the channel attention matrix A_s, and then the refined feature $F_S \in R^{C\times H\times W}$ is obtained through learning.

$$F_S = F_C \oplus (\beta\,(A_s \otimes F_C)) \quad (9)$$

where β is a learnable parameter, and generally the initial value is set to 0 to reduce the difficulty of the convergence process of the first few training cycles. The spatial attention matrix A_s can be regarded as a position mask to focus on describing the most important part of the facial feature map. Therefore, the final structure of the attention module in the proposed algorithm is shown in **Figure 6**.

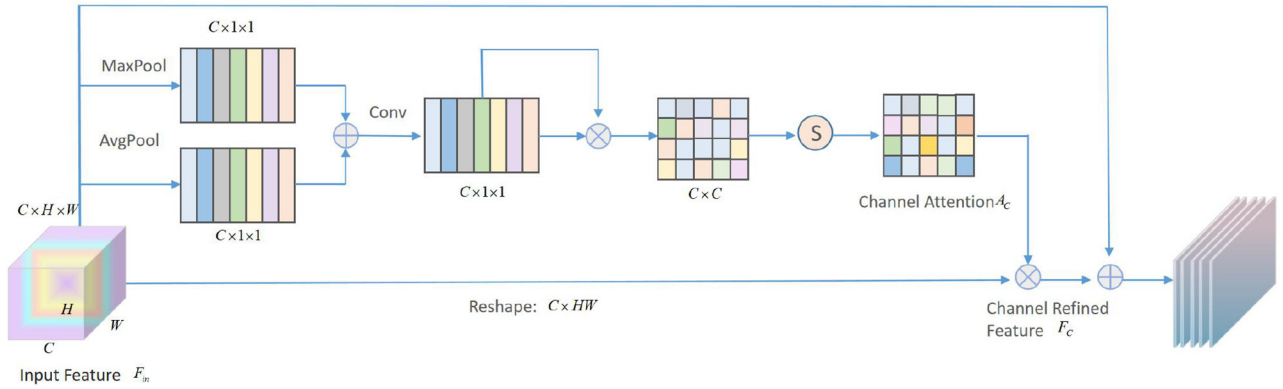


FIGURE 4 | Schematic diagram of Channel attention module.

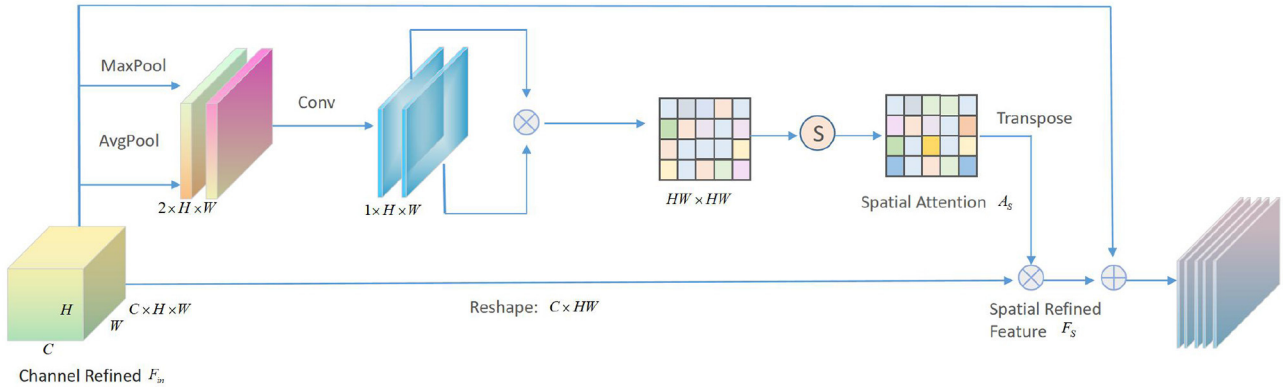


FIGURE 5 | Schematic diagram of spatial attention module.

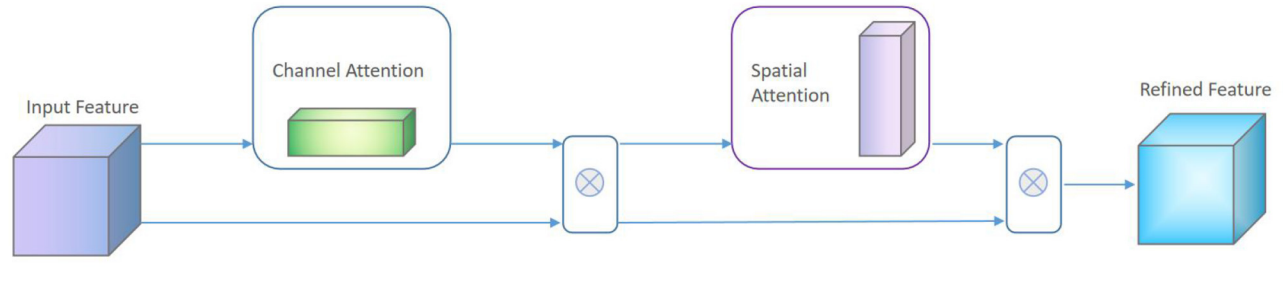


FIGURE 6 | Schematic diagram of attention module.

Interaction Mechanism

Figure 7 depicts a schematic representation of multi-layer feature interaction and flow. The orange features are retrieved by the 7*7 convolution kernel, the blue features are extracted by the 5*5 convolution kernel, and the green features are extracted by the 3*3 convolution kernel. The multi-layer feature interaction module can capture the feature information between layers of different scales, and through the spatial attention mechanism, it can also extract the feature relationships between layers of different scales.

Angular Distance Loss

In order to further deal with the problems of large intra-class differences and high similarity between classes in facial emotion image recognition, we use the angular distance loss function to improve the capabilities of the proposed algorithm for feature

separation between classes and clustering of features within classes. The calculation equation of angular distance loss is as follows:

$$L_{ads} = -\frac{1}{m}\sum_{i=1}^{m}\log\frac{\exp[s \times \cos(\theta_{y_i} + m)]}{\exp[s \times \cos(\theta_{y_i} + m)] + \sum_{j=1}^{n}\exp(s \times \cos\theta_j)} \quad (10)$$

where s is the scaling factor, $\cos(\theta_{y_i} + m)$ is the angular distance, and m determines the size of the distance. The decision boundary of the softmax and the angular distance loss function in the case of two classifications is shown in **Figure 8**. The blue dashed line represents the classification decision boundary. Softmax classifies by angle, and the angle distance loss directly controls the distance of the classification decision boundary in the angle space through the decision margin m, thereby increasing the distance between classes, which is conducive to classification decision.

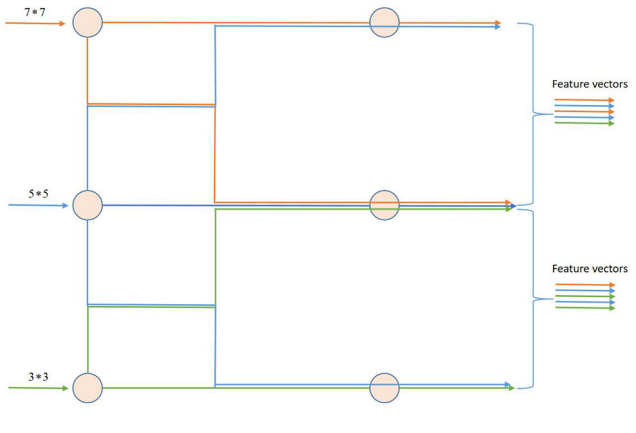


FIGURE 7 | Schematic diagram of multi-layer feature interaction and feature flow.

EXPERIMENTS AND RESULTS

Experimental Setup

All experiments in this section are run on the same server to ensure a fair evaluation of the proposed algorithm. The server's specific configuration is as follows: the operating system is Windows 10, the GPU is NVIDIA GTX1080 (11G), the memory is 16G, and the CPU is AMD Ryzen 7 1700X; the deep learning development framework is Keras 2.1.5, install CUDA9+cudnn7, and the programming language is Python 3.6.5, Adam is the optimizer, and batch size = 16, learning rate = 0.001, Epochs = 300.

Experimental Data Set

The FER2013 data set is the official data set for Kaggle's facial expression recognition competition in 2013. Because the majority of the images are downloaded from web crawlers, they will be compared because they comprise images of various ages, angles, and partially obscured images, among other things. There will be some inaccuracies, but it will be near to natural facial emotions. There are 35,887 photos in all in FER2013. The training set, public test set, and private test set are the three elements of the data set. The training set contains 28,709 photographs. Both the public and private test sets contain 3,589 images. **Table 1** shows the data distribution in this data set, with the tags corresponding to the seven expressions numbered 0–6. Because the majority of

TABLE 1 | Data distribution of FER2013 dataset.

	Angry	Fear	Disgust	Happy	Sad	Surprise	Neutral	Total
Label	0	1	2	3	4	5	6	–
Train set	3,995	4,097	436	7,215	4,830	3,171	4,965	28,709
Validation set	467	496	56	895	653	415	607	3,589
Test set	491	528	55	879	594	416	626	3,589
Total	4,953	5,121	547	8,989	6,077	4,002	6,198	35,887

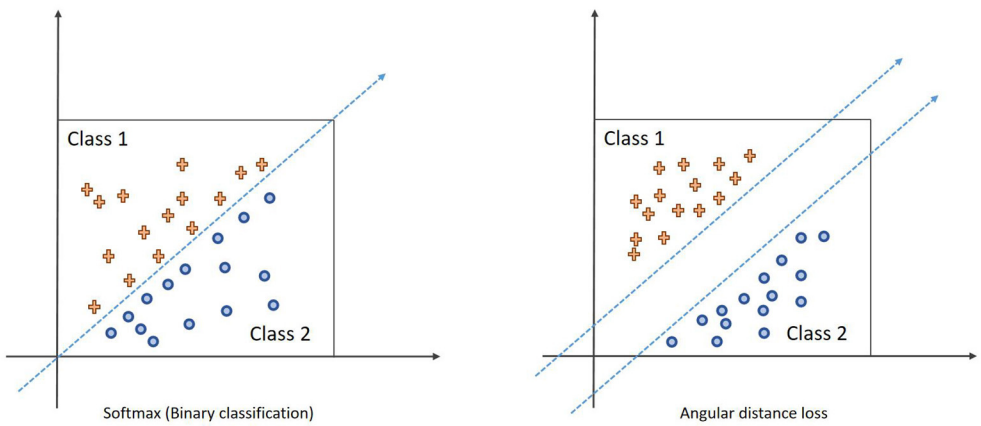


FIGURE 8 | Schematic diagram of angular distance loss.

the FER2013 data set comes from web crawlers, the background is more complex, making identification difficult.

Evaluation Method

Facial emotion recognition research mainly uses accuracy and confusion matrix as the evaluation indicators of the model. The accuracy rate represents the ratio of the number of correctly identified samples to the total number of samples, which can reveal the overall recognition ability of the model. The calculation equation is as follows:

$$Accuracy = \frac{\sum_{i=1}^{C} TP^i}{N} \tag{11}$$

where TP^i represents the number of correctly classified samples in the i-th category, C represents the number of categories, and N represents the total number of samples.

The confusion matrix is a square matrix of size (Z, Z), in which the true label provided by the element CP_{ij} in the i-th row and j-column is the probability of the i-th category and the predicted label is the j-th category. The calculation equation is as follows:

$$CP_{ij} = \frac{n^{ij}}{n^i} \tag{12}$$

where n^{ij} represents the true label is the i-th class and the predicted label is the number of samples in the j-th class, and n^i represents the total number of samples in the i-th class. By analyzing the confusion matrix, the accuracy performance of the model in each category can be measured.

Experimental Results

We compared several well-known methods on the FER2013 data set to better evaluate the effectiveness of the proposed algorithm, and the results are shown in **Table 2**. Turan et al. (2018) proposed Soft Locality Preserving Map, a new and more effective manifold learning method that aims to control the diffusion level of different classes, effectively reducing the dimensionality of feature vectors and enhancing the extracted features. The improvement effect is not ideal for facial expression recognition distinguishing ability. Yang et al. (2018) proposed a facial expression recognition method based on residual expressions. Residual error learning is used to generate the residuals of the middle layer of the model. The residuals contain the expression components of any generated model of the input expression image, but the feature connection between the levels is not captured, and the classification effect is not good. The expression recognition rate of Shao and Qian (2019) in the two data sets is not high, and there is a problem that the recognition rate is low due to insufficient expression feature extraction. In addition, we also compared with InceptionV4, DNNRL, Multi-scale CNN, and Hybrid CNN-SIFT aggregator.

It can be seen from **Table 2** that the proposed algorithm has achieved the best classification effect on the FER2013 data set in a complex environment. This is because the method in this paper makes full use of the multi-scale features of the multi-layer interactive feature fusion network, and integrates the cross-layer deep feature representation, captures the subtle changes in the deep level of expression, and restores the expression image gradually through multi-layer feature fusion. The useful feature information lost in the layer transfer process solves the problem of interaction between model layers and multi-layer feature fusion, and improves the network's ability to distinguish facial expressions caused by subtle changes in the corners of the mouth, eyebrows, and eyes. In addition, we propose Using the angular distance loss function effectively alleviates the problems of large intra-class differences and high inter-class similarity in facial expression recognition, and is more suitable for subtle facial expression classification. We also show the confusion matrix of the proposed algorithm on the test set in **Figure 9**. In addition, as shown in **Figure 10**, although the training data is unbalanced, the proposed

TABLE 2 | Comparison of recognition rates of different algorithms on FER2013 dataset.

Methods	Acc
InceptionV4 (Szegedy et al., 2017)	0.7080
DNNRL (Kim et al., 2016)	0.7082
Multi-scale CNN (Wang and Yuan, 2016)	0.7282
SLPM (Turan et al., 2018)	0.7091
DeRL (Yang et al., 2018)	0.7264
Shao (Shao and Qian, 2019)	0.7114
Hybrid CNN-SIFT aggregator (Connie et al., 2017)	0.7340
Ours	**0.7416**

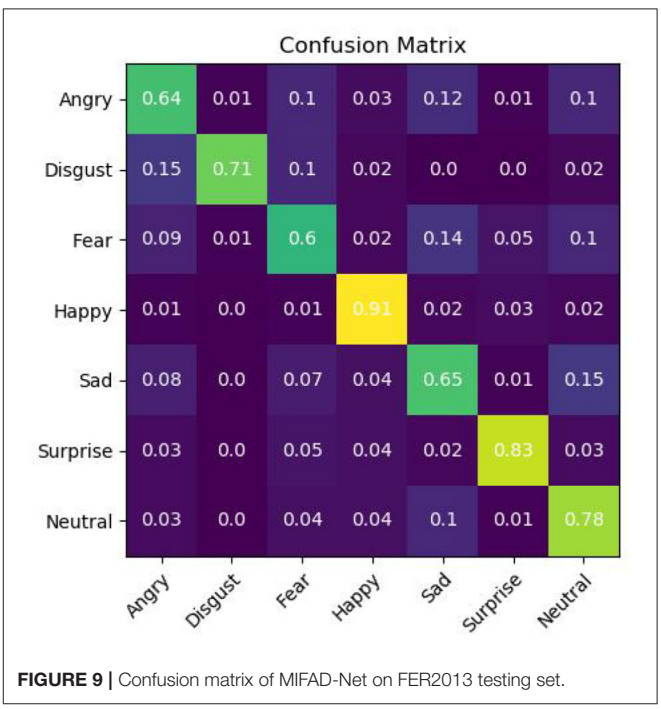


FIGURE 9 | Confusion matrix of MIFAD-Net on FER2013 testing set.

FIGURE 10 | FER2013 training data distribution.

algorithm overcomes this problem and achieves a competitive classification performance.

Ablation Experiment for Different Loss Functions

To further verify the effectiveness of the angular distance loss function in the proposed algorithm, ablation experiments are set up in this section. We introduce Island loss and center loss for ablation studies. In addition, for fair comparison, all experiments are performed in the same environment. The results of the ablation experiment are shown in **Table 3**.

It can be seen from **Table 3** that the best classification performance is obtained by using the angular distance loss. In addition, we find that the Island loss is better than the center loss, and the softmax performance alone is the worst. Because Softmax predicts the probability of each category, it does not optimize the distance between the classes and the class, which leads to the lack of distinction between features. In order to reduce the difference in features within the class, the authors (Cai et al., 2018) proposed the optimization and improvement of Center and Island Loss. Island Loss increases the constraints of facial expression features to make the distance between classes larger, thereby improving the classification performance.

Ablation Experiment for Multi-Layer and Multi-Scale

A parallel three-branch network is used in the proposed algorithm. An ablation experiment is set up in this section to further prove its effectiveness. For comparison, we add two-branch and four-branch networks. The ablation experiment's results are shown in **Table 4**.

Table 4 shows that using a two-branch network reduces the model's classification performance significantly, whereas using a four-branch network does not improve classification performance. As a result, the proposed algorithm is proven to be effective.

TABLE 3 | Results of ablation experiments for different loss functions on FER2013 dataset.

Methods	Acc
Softmax	0.7211
Island+Softmax	0.7385
Center loss+Softmax	0.7294
Ours	**0.7416**

TABLE 4 | Results of ablation experiments for multi-layer and multi-scale on FER2013 dataset.

Methods	Acc
Two-branch	0.7105
Four-branch	0.7391
Ours	**0.7416**

TABLE 5 | Results of ablation experiments for attention mechanism on FER2013 dataset.

Methods	Acc
No-attention	0.7262
Ours	**0.7416**

TABLE 6 | Results of ablation experiments for feature fusion strategy on FER2013 dataset.

Methods	Acc
Add	0.7325
Mul	0.7298
Ours (C)	**0.7416**

Ablation Experiment for Attention Mechanism

This section sets up an ablation experiment to test the effect of the proposed algorithm's attention mechanism on classification performance in order to verify its effectiveness. The term "No-attention" refers to the lack of use of the attention mechanism. **Table 5** shows the results of the ablation experiment.

It can be seen from **Table 5** that if the attention mechanism is not used, the classification performance of the model will be reduced by 1.97%. Because facial expressions consist of muscle movements in specific parts of the face. The features produced by these local areas contain the information that best describes expressions. Therefore, using the attention mechanism to quantify the importance of each spatial position in the feature map and focusing on the areas with rich emotional information is beneficial to the recognition task.

Ablation Experiment for Feature Fusion Strategy

To further verify the influence of the feature fusion strategy on the experimental results, an ablation experiment was carried

out in this section. "Add" stands for addition strategy, "C" stands for concat strategy, and "Mul" stands for multiplication strategy. The results of the ablation experiment are shown in **Table 6**.

It can be clearly seen from **Table 6** that the Concat strategy used by the proposed algorithm achieves the best results. Secondly, the addition and multiplication is better than the multiplication strategy, which proves that the extraction of multi-scale features effectively improves the classification performance of the proposed algorithm, and also further prove the superiority of the proposed algorithm.

CONCLUSION

In this paper, we propose a novel multi-layer interactive feature fusion network model with angular distance loss. First, a multi-layer and multi-scale module is designed to extract the global and local features of facial expressions to capture part of the feature relationships between different scales, thereby enhancing the model's ability to discriminate subtle features of facial expressions. Secondly, in view of the problem of loss of useful feature information due to layer-by-layer convolution and pooling of convolutional neural networks, a hierarchical interactive feature fusion module is designed. The attention mechanism is used between convolutional layers at different levels to control the network. Strengthen the saliency information of different characteristics in the Internet and suppress irrelevant information, thereby improving the discriminative ability of the network. Finally, for the problem of large intra-class differences and high similarity between classes in facial expression recognition, we use the angular distance loss function to improve the capabilities of the proposed algorithm for feature separation between classes and clustering of features within classes. We conducted comparison and ablation experiments on the FER2013 data set. The results illustrate that the proposed MIFAD-Net outperforms a number of well-known methods and is highly competitive.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data/.

AUTHOR CONTRIBUTIONS

WC: conceptualization, methodology, software, and writing. JM: investigation. MG: data curation, software, and validation. RL: data curation and investigation. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Scientific Research Program of Education Department of Hubei Province, China (D20184101), Higher Education Reform Project of Hubei Province, China (201707), East Lake Scholar of Wuhan Sport University Fund, China and Hubei Provincial University Specialty subject group construction Special fund, China.

REFERENCES

Bendjoudi, I., Vanderhaegen, F., Hamad, D., and Dornaika, F. (2020). Multi-label, multi-task CNN approach for context-based emotion recognition. *Inf. Fusion*. 76, 422–428. doi: 10.1016/j.inffus.2020.11.007

Bota, P. J., Wang, C., Fred, A. L., and Da Silva, H. P. (2019). A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals. *IEEE Access* 7, 140990–141020. doi: 10.1109/ACCESS.2019.2944001

Cai, J., Meng, Z., Khan, A. S., Li, Z., O'Reilly, J. and Tong, Y. (2018). "Island loss for learning discriminative features in facial expression recognition," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (Xi'an), 302–309. doi: 10.1109/FG.2018.00051

Cai, W., Song, Y., and Wei, Z. (2021). Multimodal data guided spatial feature fusion and grouping strategy for E-commerce commodity demand forecasting. *Mobile Inf. Syst.* (2021) 2021:5541298. doi: 10.1155/2021/5568208

Cai, W., and Wei, Z. (2020). PiiGAN: generative adversarial networks for pluralistic image inpainting. *IEEE Access* 8, 48451–48463. doi: 10.1109/ACCESS.2020.2979348

Chengeta, K., and Viriri, S. (2019). "A review of local, holistic and deep learning approaches in facial expressions Recognition," in *2019 Conference on Information Communications Technology and Society (ICTAS)* (Durban: IEEE), 1–7.

Connie, T., Al-Shabi, M., Cheah, W. P., and Goh, M. (2017). "Facial expression recognition using a hybrid CNN-SIFT aggregator," in *International Workshop on Multi-disciplinary Trends in Artificial Intelligence* (Cham: Springer), 139–149.

Crivelli, C., Russell, J. A., Jarillo, S., and Fernández-Dols, J. M. (2017). Recognizing spontaneous facial expressions of emotion in a small-scale society of Papua New Guinea. *Emotion* 17:337. doi: 10.1037/emo0000236

Domínguez-Jiménez, J. A., Campo-Landines, K. C., Martínez-Santos, J. C., Delahoz, E. J., and Contreras-Ortiz, S. H. (2020). A machine learning model for emotion recognition from physiological signals. *Biomed. Signal Process. Control* 55:101646. doi: 10.1016/j.bspc.2019.101646

Gao, M., Cai, W., and Liu, R. (2021). AGTH-Net: attention-based graph convolution-guided third-order hourglass network for sports video classification. *J. Healthc Eng.* (2021) 2021:8517161. doi: 10.1155/2021/8517161

González-Lozoya, S. M., de la Calleja, J., Pellegrin, L., Escalante, H. J., Medina, M. A., and Benitez-Ruiz, A. (2020). Recognition of facial expressions based on CNN features. *Multimed. Tools Appl.* 79, 13987–14007. doi: 10.1007/s11042-020-08681-4

Hamelin, N., El Moujahid, O., and Thaichon, P. (2017). Emotion and advertising effectiveness: a novel facial expression analysis approach. *J. Retail. Consum. Serv.* 36, 103–111. doi: 10.1016/j.jretconser.2017.01.001

Harit, A., Joshi, J. C., and Gupta, K. K. (2018). Facial emotions recognition using gabor transform and facial animation parameters with neural networks. *IOP Conf. Ser. Mater. Sci. Eng.* 331:012013. doi: 10.1088/1757-899X/331/1/012013

Jain, D. K., Shamsolmoali, P., and Sehdev, P. (2019). Extended deep neural network for facial emotion recognition. *Pattern Recognit. Lett.* 120, 69–74. doi: 10.1016/j.patrec.2019.01.008

Kerkeni, L., Serrestou, Y., Raoof, K., Mbarki, M., Mahjoub, M. A., and Cleder, C. (2019). Automatic speech emotion recognition using an optimal combination of features based on EMD-TKEO. *Speech Commun.* 114, 22–35. doi: 10.1016/j.specom.2019.09.002

Kim, B. K., Roh, J., Dong, S. Y., and Lee, S. Y. (2016). Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. *J. Multimodal User Interf.* 10, 173–189. doi: 10.1007/s12193-015-0209-0

Kollias, D., and Zafeiriou, S. P. (2020). Exploiting multi-cnn features in cnn-rnn based dimensional emotion recognition on the omg in-the-wild dataset. *IEEE Trans. Affect. Comput.* 12, 595–606. doi: 10.1109/TAFFC.2020.3014171

Kumar, S., Bhuyan, M. K., and Chakraborty, B. K. (2016). Extraction of informative regions of a face for facial expression recognition. *IET Comput. Vis.* 10, 567–576. doi: 10.1049/iet-cvi.2015.0273

Kwon, S. (2021). MLT-DNet: speech emotion recognition using 1D dilated CNN based on multi-learning trick approach. *Expert Syst. Appl.* 167:114177. doi: 10.1016/j.eswa.2020.114177

Lee, J., Kim, S., Kim, S., Park, J., and Sohn, K. (2019). "Context-aware emotion recognition networks, in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (Seoul), 10143–10152.

Liew, C. F., and Yairi, T. (2015). Facial expression recognition and analysis: a comparison study of feature descriptors. *IPSJ Trans. Comput. Vis. Appl.* 7, 104–120. doi: 10.2197/ipsjtcva.7.104

Liu, R., Ning, X., Cai, W., and Li, G. (2021). Multiscale dense cross-attention mechanism with covariance pooling for hyperspectral image scene classification. *Mobile Inf. Syst.* (2021) 2021:9962057. doi: 10.1155/2021/9962057

Mollahosseini, A., Chan, D., and Mahoor, M. H. (2016). "Going deeper in facial expression recognition using deep neural networks," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (Lake Placid, NY: IEEE), 1–10.

Ouellet, S. (2014). Real-time emotion recognition for gaming using deep convolutional network features. *arXiv:1408.3750 [arXiv preprint]*.

Pramerdorfer, C., and Kampel, M. (2016). Facial expression recognition using convolutional neural networks: state of the art. *arXiv:1612.02903 [arXiv preprint]*.

Rieger, S. A., Muraleedharan, R., and Ramachandran, R. P. (2014). "Speech based emotion recognition using spectral feature extraction and an ensemble of kNN classifiers," in *The 9th International Symposium on Chinese Spoken Language Processing* (Singapore: IEEE), 589–593.

Shao, J., and Qian, Y. (2019). Three convolutional neural network models for facial expression recognition in the wild. *Neurocomputing* 355, 82–92. doi: 10.1016/j.neucom.2019.05.005

Sreedharan, N. P. N., Ganesan, B., Raveendran, R., Sarala, P., and Dennis, B. (2018). Grey Wolf optimisation-based feature selection and classification for facial emotion recognition. *IET Biometr.* 7, 490–499. doi: 10.1049/iet-bmt.2017.0160

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-First AAAI Conference on Artificial Intelligence* (San Francisco, CA).

Tang, Y. (2013). Deep learning using linear support vector machines. *arXiv:1306.0239 [arXiv preprint]*.

Tarannum, T., Paul, A., and Talukder, K. H. (2016). "Human expression recognition based on facial features," in *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)* (Dhaka: IEEE), 990–994.

Theagarajan, R., Bhanu, B., Cruz, A., Le, B., and Tambo, A. (2017). "Novel representation for driver emotion recognition in motor vehicle videos," in *2017 IEEE International Conference on Image Processing (ICIP)* (Beijing: IEEE), 810–814.

Turan, C., Lam, K. M., and He, X. (2018). Soft locality preserving map (SLPM) for facial expression recognition. *arXiv:1801.03754 [arXiv preprint]*.

Wang, J., and Yuan, C. (2016). "Facial expression recognition with multi-scale convolution neural network," in *Pacific Rim Conference on Multimedia* (Cham: Springer), 376–385.

Wang, Q., Lu, X., Li, P., Gao, Z., and Piao, Y. (2017). An information geometry-based distance between high-dimensional covariances for scalable classification. *IEEE Trans. Circ. Syst. Video Technol.* 28, 2449–2459. doi: 10.1109/TCSVT.2017.2712704

Xu, C., Dong, C., Feng, Z., and Cao, T. (2012). "Facial expression pervasive analysis based on haar-like features and svm," in *International Conference on E-business Technology and Strategy* (Berlin; Heidelberg: Springer), 521–529.

Yang, H., Ciftci, U., and Yin, L. (2018). "Facial expression recognition by de-expression residue learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE), 2168–2177.

Zepf, S., Hernandez, J., Schmitt, A., Minker, W., and Picard, R. W. (2020). Driver emotion recognition for intelligent vehicles: a survey. *ACM Compu. Surv.* 53, 1–30. doi: 10.1145/3388790

Zhang, L., Sun, L., Yu, L., Dong, X., Chen, J., Cai, W., et al. (2021). "ARFace: attention-aware and regularization for Face Recognition with Reinforcement Learning," in *IEEE Transactions on Biometrics, Behavior, and Identity Science* (Piscataway, NJ: IEEE). doi: 10.1109/TBIOM.2021.3104014

Zhang, L., Tjondronegoro, D., and Chandran, V. (2014). Random Gabor based templates for facial expression recognition in images with facial occlusion. *Neurocomputing* 145, 451–464. doi: 10.1016/j.neucom.2014.05.008

Zhang, Z. Y., Wang, R. Q., and Wei, M. M. (2019). Stack hybrid self-encoder facial expression recognition method. *Comput. Eng. Appl.* 55, 1–8. doi: 10.3778/j.issn.1002-8331.1803-0398

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Cai, Gao, Liu and Mao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

SYSTEMATIC REVIEW
published: 01 November 2021
doi: 10.3389/fpsyg.2021.771591

Research on the Emotions Based on Brain-Computer Technology: A Bibliometric Analysis and Research Agenda

Wei Yan, Xiaoju Liu, Biaoan Shan, Xiangxian Zhang and Yi Pu*

School of Management, Jilin University, Changchun, China

This study conducts a scientific analysis of 249 literature on the application of brain-computer technology in emotion research. We find that existing researches mainly focus on engineering, computer science, neurosciences neurology and psychology. PR China, United States, and Germany have the largest number of publications. Authors can be divided into four groups: real-time functional magnetic resonance imaging (rtfMRI) research group, brain-computer interface (BCI) impact factors analysis group, brain-computer music interfacing (BCMI) group, and user status research group. Clustering results can be divided into five categories, including external stimulus and event-related potential (ERP), electroencephalography (EEG), and information collection, support vector machine (SVM) and information processing, deep learning and emotion recognition, neurofeedback, and self-regulation. Based on prior researches, this study points out that individual differences, privacy risk, the extended study of BCI application scenarios and others deserve further research.

Keywords: emotion, brain-computer, digital technology, bibliometric analysis, research agenda

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Gaohui Cao,
Central China Normal University, China
Xingqun Xue,
Dalian University of Technology, China

****Correspondence:***
Biaoan Shan
shanbiaoan@163.com

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *06 September 2021*
Accepted: *07 October 2021*
Published: *01 November 2021*

Citation:
Yan W, Liu X, Shan B, Zhang X and Pu Y (2021) Research on the Emotions Based on Brain-Computer Technology: A Bibliometric Analysis and Research Agenda. Front. Psychol. 12:771591. doi: 10.3389/fpsyg.2021.771591

INTRODUCTION

In the area of organizational psychology and individual psychology, human emotion is an essential research topic and is concerning by many scholars. The definition of emotion is not widely agreed among scholars. According to existing researches, emotion consists of various structures (including joy, love, shame, affect, mood), and could be divided into different dimensions, e.g., approach-oriented or withdrawal-oriented emotion, pleasure or arousal emotion, and self-caused or other-caused emotion, basic or social emotion (e.g., Ashforth and Humphrey, 1995; Butt et al., 2005). It is considered as a key antecedent of cognition and perception, and strongly influences behaviors of actors.

Prior studies in this area have focused on two aspects. Firstly, the conceptual study of emotions, including definition, dimension, and measurement (e.g., Ashforth and Humphrey, 1995; Barrett et al., 2007). Secondly, some literature has revealed that emotions play important roles in organizational and individual behaviors, including decision making, negotiation strategy, work performance, innovation, and entrepreneurship (e.g., Ashforth and Humphrey, 1995; Butt et al., 2005; Roundy, 2014). Besides, scholars have analyzed emotions of individuals from various organizations, such as companies, social organizations, hospitals. These studies have significantly improved the research field of psychology and have highly contributed to the understanding of emotional process.

Recently, the progress of digital technologies is promoting the development of emotion research. The rapid development of emerging digital technologies such as artificial intelligence (AI), virtual reality, human-computer interaction, and machine learning has enhanced their application in psychology (Shan et al., 2021). Especially, researches on human emotion states via brain-computer technology have increased dramatically in the past several years. However, prior researches mainly explore the concept of emotions and their influences on individuals or organizations. We still know little about how to analyze emotions by using advanced technologies.

Especially, brain-computer technology strongly enhances the ability to recognize human emotion. It provides a new tool for traditional emotion research. The emotion analysis via brain-computer technology plays a key role in designing communication systems (Atkinson and Campos, 2016). Considering that the study on this area is still in its infancy stage, it is not surprising that we know little about the research contents and progress in this field. As Wang et al. (2014) pointed out, the relationship between emotion and new technology, such as emotion recognition based on BCI, is an important subject that needs to be developed. In order to address this gap, this study reviews literatures on emotions based on brain-computer technology. We try to systematically introduce existing researches in this area and provide future research directions.

The contribution of this study mainly includes the following two aspects. On the one hand, we use scientific bibliometric analysis method to visually analyze the researches in the field of affective BCI. We use VOSviewer software to build a research topic framework to help researchers have a more comprehensive understanding of the field. On the other hand, in the part of future research, we point out the gaps in existing research and possible advanced research directions, laying a foundation for subsequent research.

This study is structured as follows. Firstly, the process of methodology is introduced. And then, analysis results of publication years, research areas, countries, authors, and cooperation networks, and the most cited papers are provided. Follow that, theme clustering analysis is completed and relevant results are produced. Finally, future researches and conclusions are provided.

METHODOLOGY

We study the application of BCI in the field of emotion research through using the literature bibliometric technology being widely used increasingly and choosing the bibliometric software VOSviewer. Furthermore, we also try to explore the publication situation, the distribution of research fields and intellectual networks, the most influential publications, existing research streams, and future research directions of this topic in recent years, to promote further research and attention of academia on this research topic. Bibliometric technology is a quantitative method based on bibliometric data. Compared with traditional literature analysis methods, it is more systematic and transparent (Wörfel, 2019), and is more suitable for analyzing the distribution of intellectual networks (Velt et al., 2020), structure and dynamic trends (Baier-Fuentes et al., 2019), focus and future research directions (Zaheer et al., 2019), and others of a given research field. Specific process of bibliometric analysis is as follows.

Question Formulation

Our research aims to comprehensively demonstrate current academic researches on the application of BCI in the field of emotion research through a bibliometric analysis of existing related publications. We try to discuss and solve following research questions:

- How has the academia's attention to this research topic changed in recent years?
- How are the research areas of existing researches distributed?
- Which countries are most interested in this topic?
- Who are the most productive authors in this topic?
- What are the most influential publications about this topic?
- What types of existing researches about this topic can be divided into? What are the foci of each category?
- What are the possible future research directions about this topic?

Literature Selection

To determine publications most relevant to the application of brain computer interaction in the field of emotion research, we first use *((TS = ("emotion*" and "brain-computer*")) AND LA = (English)) AND DT = (Article)* as a search term to retrieve in the Core Collection Database of Web of Science, the most commonly used database in bibliometrics. English is chosen to ensure the quality and understandability of publications, and peer-reviewed articles are chosen to ensure the scientificity and authority of publications. And 270 publications are obtained (the retrieval was conducted on August 5, 2021). Next, to eliminate publications that are not related to our research topic (e.g., art, physics), research team members read title, abstract, author keywords, and keywords plus® of these publications, respectively, and further read introduction or full text when necessary. After integrating opinions of various researchers, we finally get 249 publications as our research objects.

Bibliometric Analysis

With the help of the Web of Science database and VOSviewer software, we conduct a bibliometric analysis on the selected literature set. We first analyze publication years, research areas, countries, authors, and cooperation networks, and most cited papers in BCI and emotion research. Through these analyses, scholars and relevant researchers can have a clearer understanding of the research status of BCI and emotion research, so that they can find research questions they are interested in.

Thematic Clustering Analysis

To identify research clusters of our research topic and find out current foci and future research directions, we continue to use

VOSviewer software to conduct thematic clustering analysis on selected publications. This part will be discussed separately.

RESULTS

Publication Years

In 2005, Pun et al. (2006) shared a paper titled *Brain-Computer Interaction Research at The Computer Vision and Multimedia Laboratory* on 3rd International Meeting on Brain-Computer Interface Technology in University of Geneva. This paper is related to the work of the Multi-modal Interaction Group at the University of Geneva in the field of BCI. The team has three main research topics: brain source activity recognition, BCI protocol and performance research, and assessment of users' emotion states. This paper first focused on the application of BCI in emotion field and was published in IEEE Transactions on Neural Systems and Rehabilitation Engineering in 2006.

In recent years, scholars have paid more and more attention to the application of BCI in emotion field. In this study, the number of annual publications about this topic is counted (**Figure 1**). In 2006, the first paper on the application of BCI in emotion field was published, but it failed to attract widespread attention of scholars. In 2011, 7 related articles were published successively, exceeding the total number of articles published in the past 5 years. The progress of BCI technology has greatly promoted research interests of scholars. The number of articles published in 2018 reached 38, about twice the number published in 2017. With the increase in the number of applied studies of BCI in emotion field, it has become a very important work to sort out and analyze these documents with the help of scientific software systems.

Research Areas

Figure 2 shows the research field distribution of the 249 articles we retrieved. The same article may belong to multiple fields at the same time. The top three areas of this study are Engineering, Computer Science, and Neurosciences Neurology, respectively with 98, 81, and 76 papers. This is because BCI is an interdisciplinary research field involving multiple disciplines including engineering, computer science, and neuroscience, aiming to study and develop the ability to identify, interpret, process, and model human brain activity. Human emotion recognition is closely related to Neurosciences Neurology and Psychology, which have laid the foundation for the application of BCI in the identification of human emotions.

Countries

In order to understand the attention of each country on BCI and emotion research, the number of published papers in each country is statistically analyzed (**Figure 3**). Countries with more publications are more interested in the topic. Peoples R China ranks first with 45 articles, accounting for 18.072% of the total articles, followed by the United States with 41 articles and Germany with 30 articles. In addition, India, England, Italy, Japan, and South Korea also have good research results.

Authors and Cooperation Networks

The following is a study on the number of authors' publications (**Table 1**). Birbaumer is the most productive scholar in the field of affective BCI, with 10 articles published from 2007 to 2017, accounting for 4.016% of the total literature. Sitaram, Bajaj, Cichocki, Kubler, Placidi, and Spezialetti are prolific authors

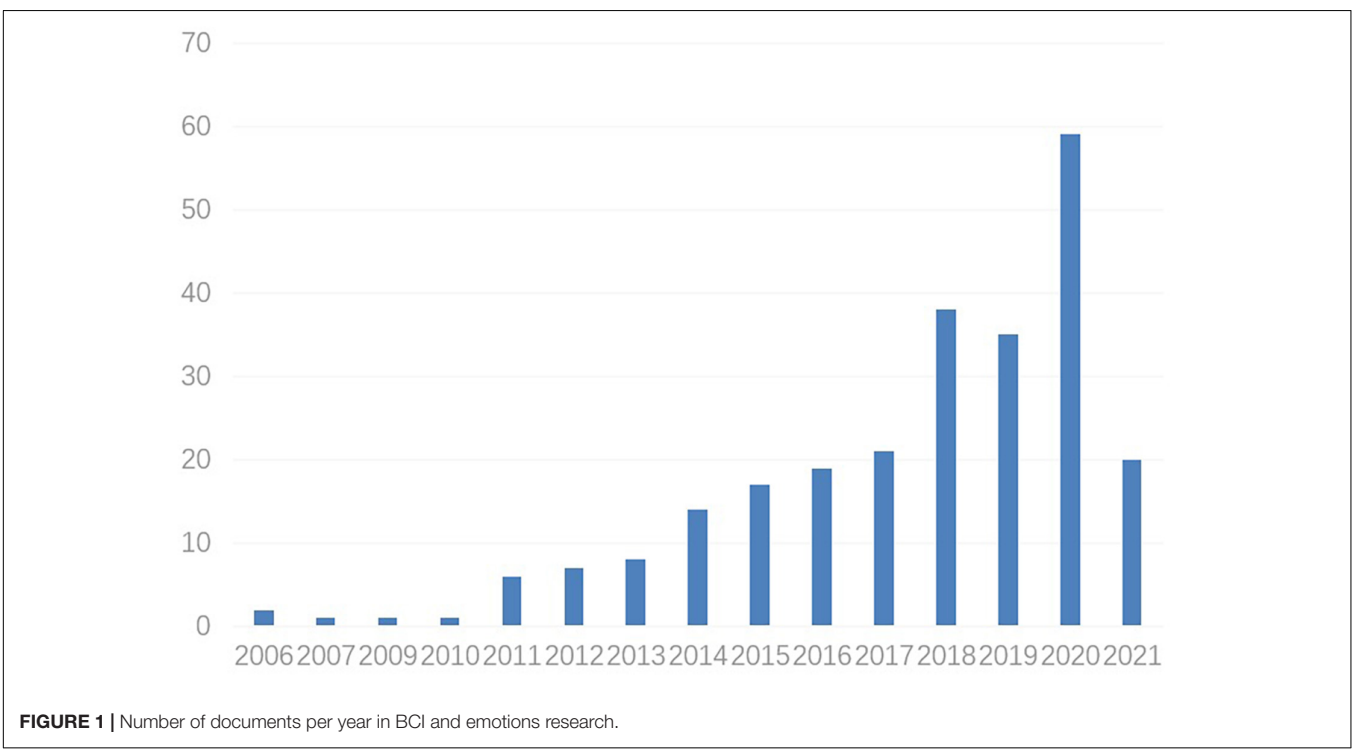


FIGURE 1 | Number of documents per year in BCI and emotions research.

FIGURE 2 | Number of documents by research areas in BCI and emotions research.

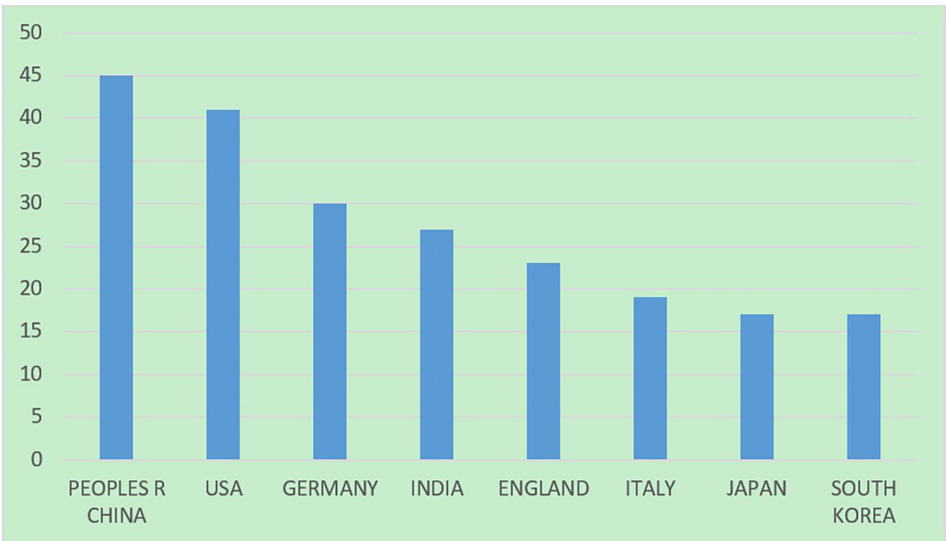


FIGURE 3 | Number of documents by countries in BCI and emotions research.

in the field of emotion research based on BCI, with the first two authors publishing 7 and 6 articles, respectively, and the remaining authors publishing 5 articles.

Further, in the field of BCI and emotion research, scholars have carried out extensive cooperation. A total of 869 authors has co-published 249 articles. Only authors who have published more than two articles in collaboration with others will be included in the collaboration network. In this way, 125 eligible authors are selected. Then we chose the largest set of 36 authors for our study because it is the most widely collaborative network of authors. VOSviewer classifies authors according to the authors' cooperation, and colors the different groups. The collaboration network is shown in **Figure 4**. Authors are divided into four categories.

Group 1: Real-Time Functional Magnetic Resonance Imaging Research Group

Authors in the first category in **Figure 4** are in red group, include Niels Birbaumer, Ranganatha Sitaram, Ralf Veit, Andrea Caria, Antonino Raffone, etc. By reading articles of the first type authors, we find that they pay more attention to the effect of functional magnetic resonance imaging (fMRI) on the insula state. In 2007, Caria et al. (2007) used real-time functional magnetic

TABLE 1 | The most productive authors in BCI and emotion research.

	Authors	Record count	% of 249	Typical article
1	Birbaumer N	10	4.016	Regulation of anterior insular cortex activity using real-time fMRI
2	Sitaram R	7	2.811	Regulation of anterior insular cortex activity using real-time fMRI
3	Bajaj V	6	2.41	Emotion recognition from single-channel EEG signals using a two-stage correlation and instantaneous frequency-based filtering method
4	Cichocki A	5	2.008	The changing face of P300 BCIs: A Comparison of stimulus changes in a P300 BCI involving faces, emotion, and movement
4	Kubler A	5	2.008	The Changing face of P300 BCIs: A comparison of stimulus changes in a P300 BCI involving faces, emotion, and movement
4	Placidi G	5	2.008	A real-time classification algorithm for EEG-based BCI driven by self-induced emotions
4	Spezialetti M	5	2.008	A real-time classification algorithm for EEG-based BCI driven by self-induced emotions
7	Chau T	4	1.606	Classification of prefrontal activity due to mental arithmetic and music imagery using hidden Markov models and frequency domain near-infrared spectroscopy
7	Daly I	4	1.606	Affective brain-computer music interfacing
7	Falk TH	4	1.606	Classification of prefrontal activity due to mental arithmetic and music imagery using hidden Markov models and frequency domain near-infrared spectroscopy

resonance imaging (rtfMRI) for the first time to train and study the willpower control of brain regions related to emotion and confirmed that rtfMRI may enable local brain activity to achieve self-regulation. Distinguishing brain response to different stimuli is a prerequisite for communication through BCI. In the context of fMRI, Van der Heiden et al. (2014) found that insula and inferior frontal triangularis activations could distinguish brain responses under different emotional sound stimuli, which laid a foundation for the applied research of BCI. In addition, researchers have investigated the volitional control of the anterior insula in criminal psychopaths (Sitaram et al., 2014) and the application of BCI in Alzheimer's disease patients (Liberati et al., 2012). According to their research content, we name the first group rtfMRI Research Group.

Group 2: Brain-Computer Interface Impact Factors Analysis Group

We refer to the green group of authors as the BCI Impact Factors Analysis Group, including Andrzej Cichocki, Andrea Kuebler, Xingyu Wang, Baoliang Lu, etc. Above authors focused on the effects of emotional face, background music, and emotion recognition pathways on BCI performance. Bakardjian et al. (2011) found that emotional faces can enhance the steady-state visual responses for BCI and improve the reliability of results. Zhou et al. (2016) explored the influence of background music on BCI performance when users used auditory BCI. The study showed that background music would not affect the effect of auditory BCI, but it is recommended to add background music when designing auditory BCI since users preferred BCI with it. Combining brain waves and eye movements, Zheng et al. (2019) proposed a multi-modal emotion recognition framework named EmotionMeter, which was better than the single mode for human emotion recognition. Studies have found that brain waves are good at identifying happy emotions, while eye movements have advantages in identifying fear emotions.

Group 3: Brain-Computer Music Interfacing Group

In the blue group, authors conducted a series of studies on brain-computer music interfacing (BCMI). Therefore, this study names it BCMI group, including Eduardo Miranda, Ian Daly, James Weaver, Duncan Williams, Asad Malik, etc. Eaton et al. (2015) proposed a multi-user affective BCMI for performers and audiences to measure their emotional changes in concert scene. The multi-user affective BCMI provides a platform for personalized design of music by matching music and emotion. Daly et al. (2016) developed an affective brain-computer music interfacing that can regulate users' emotions, which has the potential to be used in music therapy and entertainment. Daly et al. (2020) further proposed a personalized method that could be used for BCMI emotional state detection. Compared with population-based detection method, the accuracy of personalized emotion detection was significantly improved.

Group 4: User Status Research Group

Sebastian Grissmann, Josef Pieces, Thorsten O Zander, Peter Gerjets, Martin Spueler et al., shown in the yellow group in **Figure 4**, belong to the fourth category of authors. This group's name is the User Status Research Group. Zander and Kothe (2011) proposed passive BCI, which is an extension method of BCI. It integrates cognitive detection into BCI technology and provides valuable information such as users' emotional intentions and states for the technical system. Grissmann et al. (2017) argued that the previous BCI approach focused on detecting a single aspect of user states in EEG was limited and we also need to consider the effect of unknown mental states on BCI performance. It is found that adding user state information can improve the application of BCI in real-world scenarios. Grissmann et al. (2020) believed that when BCI was applied in actual situation, user state recognition would be affected by users' psychological states. Their results showed that when researchers classified working memory load under affective valence, the classification accuracy was significantly improved.

In this paper, authors and their cooperation networks are classified, and the main research contents of each group are analyzed in detail. The main purpose of this part is to help

beginners find articles to study and to provide a guide for beginners in the field. For example, if beginners are interested in the study of affective BCMI, they can carry out a targeted study on articles of BCMI Group.

The Most Cited Papers

The number of citations of an article can reflect the recognition degree of the academic community to this article. The greater the total citation number, the more classic and influential the article. **Table 2** lists the top 10 articles with the most cited numbers in the affective BCI field. *Toward passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general* was cited 369 times, ranking first in **Table 2**. This article was published by Zander and Kothe in 2011 on Iop Publishing Ltd. This paper focused on the study of passive BCI, which is a BCI extension method integrating BCI and cognitive monitoring. Through reading these 10 articles, we find that these articles mainly study the regulation of brain activity by real-time fMRI, the classification of EEG data by machine learning and the identification of emotions according to EEG and so on. These articles laid a foundation for subsequent research on affective BCI.

THEME CLUSTERING ANALYSIS

Keywords can help readers quickly understand the content of an article. With the help of VOSviewer software, the co-occurrence analysis of keywords in 249 literatures is carried out. The classification of articles according to keywords helps us construct the framework of BCI applied research in emotion field. In 249 literatures, there are 791 author keywords. We limit keyword occurrence to at least 3 times and finally obtain 58 keywords for keyword co-occurrence analysis. In order to get accurate classification results, we also combine synonymous keywords. For example, we replace "brain computer interfaces," "brain-computer interface," and "brain-computer interfaces" with "brain computer interface." The clustering results of keywords are shown in **Figure 5**. From **Figure 5**, we can intuitively find that brain computer interface, electroencephalography, and emotion recognition are keywords appeared most frequently, which are related to our literature screening method and also reflect research topics in this field. Specifically, the study of the affective BCI field can be divided into five parts.

Cluster 1: External Stimulus and Event-Related Potential

The yellow part in **Figure 5** is cluster 1, which is referred to as external stimulus and event-related potential (ERP). Keywords in cluster 1 include event-related potential, p300, attention, computer interface, emotional state, motivation, etc.

ERP is a special kind of brain evoked potential, also known as "cognitive potential." ERP reflects electrical activities of our brain in the cognitive process, and is a psychological response to external stimuli (such as vision, hearing, touch, etc.). Classic ERP mainly includes P1, N1, P2, N2, P300, N400, etc. P300 is one of the most used evoked potentials for BCI. P300 is an endogenous component that is not affected by physical properties of the stimulus, and is related to cognitive functions such as attention, recognition and memory (Polich, 2007).

Cluster 1 focuses on external stimulus and cognitive potential, especially on external stimulus and P300. P300-based BCI can be used to identify the influence of external stimulus on the tester's emotional state, and can accurately identify the focus of attention of the tester. Pan et al. (2018) developed a BCI system for emotion recognition in patients with consciousness disorders. Subjects received a random stimulus to evoke the P300 potential. BCI monitors P300 potential and feedback results in time. The results show that the BCI system can be used to conduct cognitive experiments on patients with consciousness disorders. Ganin et al. (2019) took patients with anorexia nervosa as the research object and observed their EEG responses when they received different emotional stimuli. The study found that images of thin people's body parts caused higher ERP amplitudes than images of food. The study also confirmed that P300-based BCI can effectively identify covert emotional foci of attention. Kleih and Kubler (2013) focused on effects of motivation for helping and empathy on P300-BCI performance. It was found that helping motivation did not affect P300 amplitude and BCI performance. However, the P300 amplitudes of high empathy testers were significantly lower than those of low empathy testers. The reason may be that people with high empathy are better able to understand external emotional stimuli and get more emotionally involved. As a result, their attention assigned to BCI tasks was reduced.

Cluster 2: Electroencephalography and Information Collection

Cluster 2 (green) mainly consists of electroencephalography, affective brain-computer interface, affective states, arousal, classification, cognition, emotion, learning, transfer learning, etc. By reading corresponding articles, we find that cluster 2 research is related to electroencephalography (EEG) and information collection. Therefore, we call the cluster 2 EEG and information collection.

BCI records EEG signals, which are then processed and analyzed in a series of ways, and then converted into output signals that can be controlled by external devices. EEG is signal recording changes of scalp potential, which reflects the activity of the cerebral cortex to a certain extent. It has advantages of low cost, simple collection and high time resolution (Alarcao and Fonseca, 2019). The decoding of EEG signal is the key technology for BCI.

Cluster 2 revolves around EEG and information collection. Specifically, Cluster 2 focuses on how EEG signals are collected and analyzed when using emotional brain-computer interfaces. Scholars agree that emotional changes are closely linked to electrical signals in brain. Mowla et al. (2020) used laboratory data and DEAP database to evaluate the effectiveness of EEG recordings in identifying emotional states. Atkinson and Campos (2016) proposed a feature-based emotion recognition model for EEG and confirmed the effectiveness of the model. Charland et al. (2015) believed that emotion recognition should not be limited

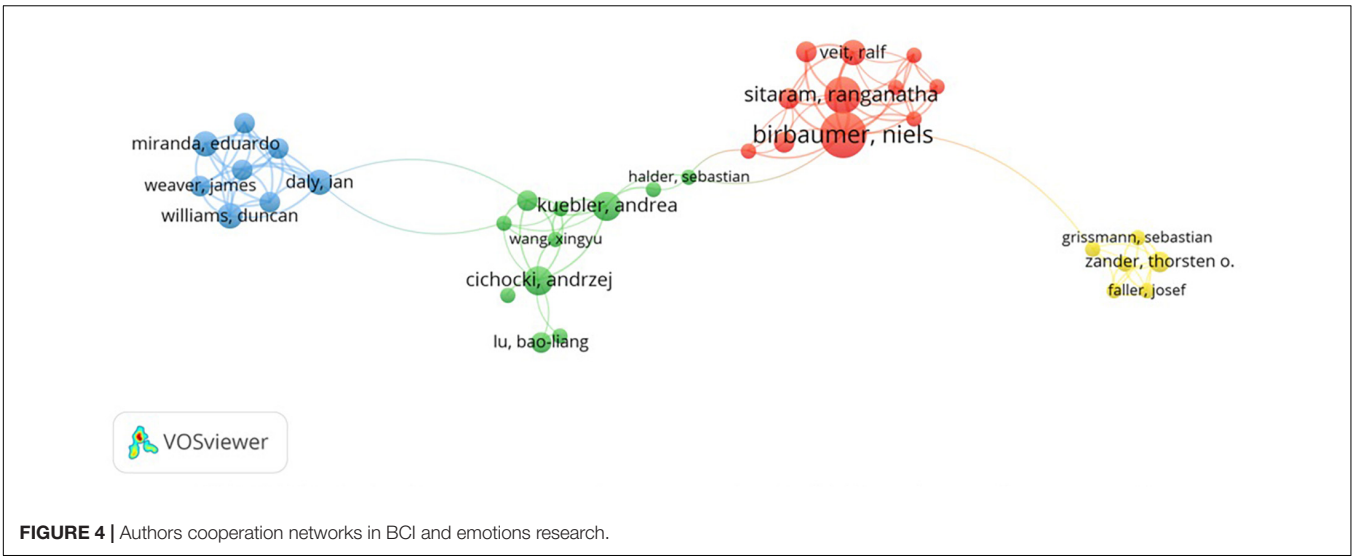


FIGURE 4 | Authors cooperation networks in BCI and emotions research.

TABLE 2 | The most cited papers in BCI and emotion research.

	Article title	Authors	Year	Publisher	Time cited
1	Toward passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general	Zander, Thorsten O.; Kothe, Christian	2011	Iop Publishing Ltd.	369
2	Emotional state classification from EEG data using machine learning approach	Wang, Xiao-Wei; Nie, Dan; Lu, Bao-Liang	2014	Elsevier	277
3	Regulation of anterior insular cortex activity using real-time fMRI	Caria, Andrea; Veit, Ralf; Sitaram, Ranganatha et al.	2007	Academic Press Inc., Elsevier Science	234
4	Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers	Atkinson, John; Campos, Daniel	2016	Pergamon-Elsevier Science Ltd.	148
5	Toward an EEG-based recognition of music liking using time-frequency analysis	Hadjidimitriou, Stelios K.; Hadjileontiadis, Leontios J.	2012	Ieee-Inst Electrical Electronics Engineers Inc.	114
6	Brain-computer interface-based communication in the completely locked-in state	Chaudhary, Ujwal; Xia, Bin; Silvoni, Stefano et al.	2017	Public Library Science	109
7	Classification of prefrontal activity due to mental arithmetic and music imagery using hidden Markov models and frequency domain near-infrared spectroscopy	Power, Sarah D.; Falk, Tiago H.; Chau, Tom	2010	Iop Publishing Ltd.	104
8	Single-trial classification of NIRS signals during emotional induction tasks: toward a corporeal machine interface	Tai, Kelly; Chau, Tom	2009	Bmc	94
9	Identifying stable patterns over time for emotion recognition from EEG	Zheng, Wei-Long; Zhu, Jia-Yi; Lu, Bao-Liang	2019	Ieee-Inst Electrical Electronics Engineers Inc.	91
10	The changing face of P300 BCIs: A comparison of stimulus changes in a P300 BCI involving faces, emotion, and movement	Jin, Jing; Allison, Brendan Z.; Kaufmann, Tobias et al.	2012	Public Library Science	81

to EEG, but should collect data information through multiple channels. Therefore, authors proposed a method for collecting and synchronizing multidimensional data to simultaneously analyze electroencephalography, electrodermal, eye-tracking, and facial emotion recognition data. Lin and Jung (2017) also focused on the application of transfer learning in EEG emotion classification system. They proposed that conditional transfer learning framework can effectively improve the classification of individual default emotions.

Cluster 3: Support Vector Machine and Information Processing

When exploring the third group support vector machine (SVM) and information processing cluster (purple), there are items like SVM, feature extraction, machine learning, brain modeling, electrodes, etc.

SVM classifies data in binary mode according to supervised learning. It is a machine learning method to solve classification problems and an important tool to solve data classification (Cortes and Vapnik, 1995; Sexton et al., 2017).

Cluster 3 focuses on SVM and information processing. Identifying emotions from EEG signals can help people with emotional expression disorders to better communicate with people and environment around them. Emotion recognition requires information processing of EEG signals. Information processing includes information preprocessing, feature extraction and selection, and emotion classification. Khare and Bajaj (2021) believed that for multi-components EEG

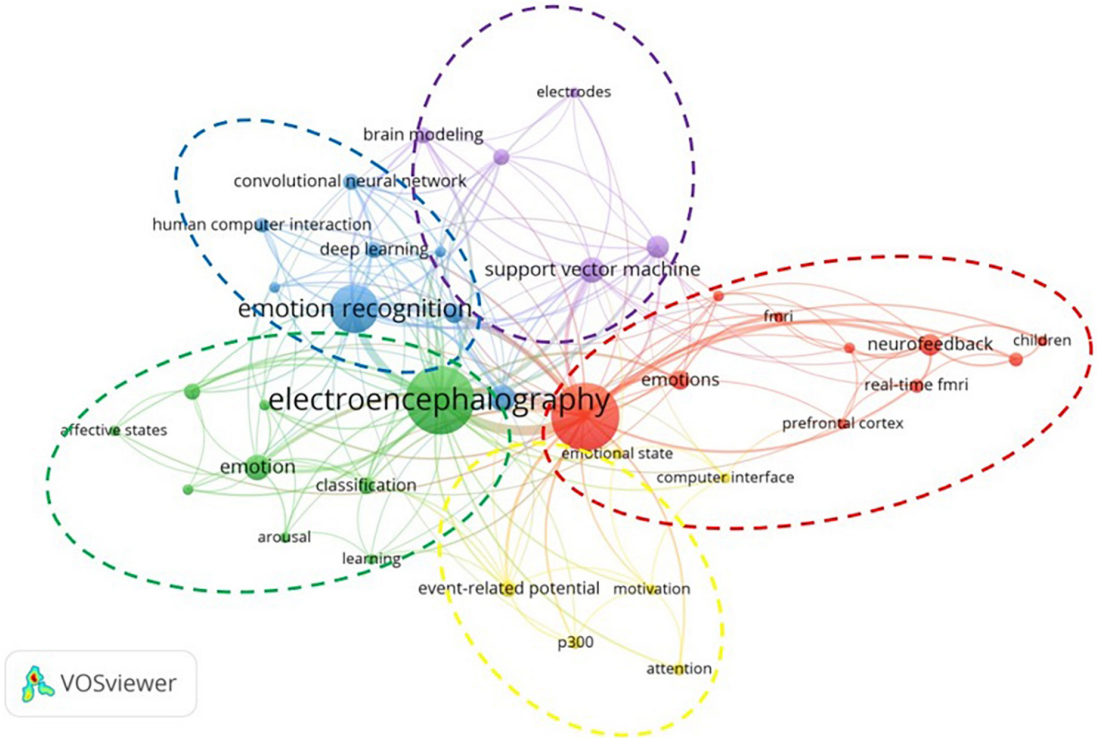


FIGURE 5 | Keywords co-occurrence analysis in BCI and emotions research.

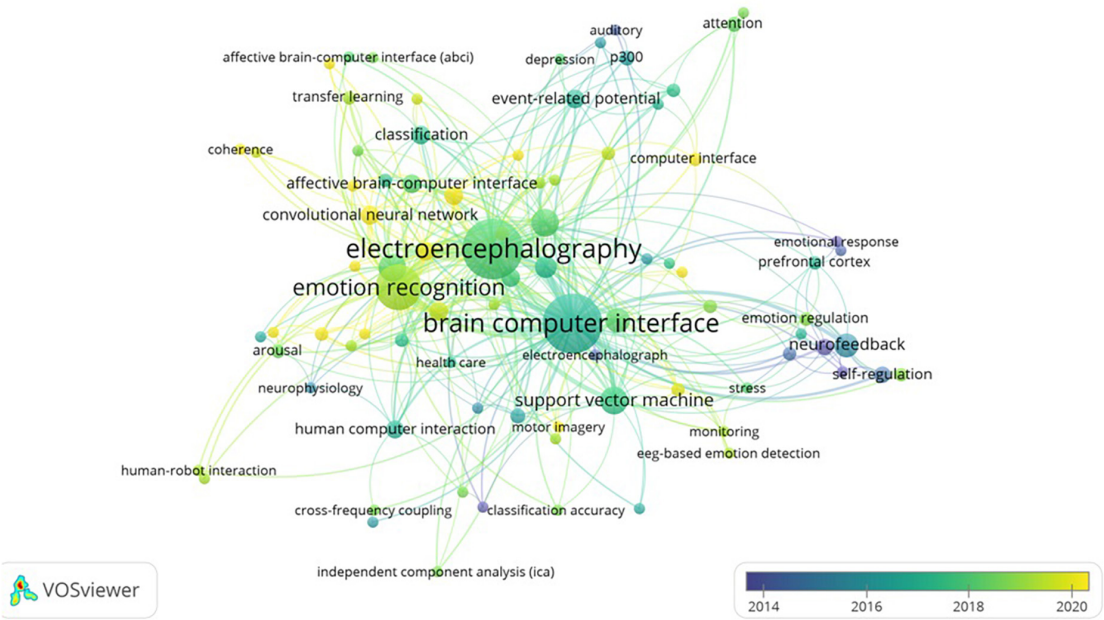


FIGURE 6 | Map of research trends.

signal, it is necessary to decompose it into component sets before extracting hidden information. Authors also proposed optimized variational mode decomposition. The method uses the eigenvector centrality method to select dominant channels. Arpaia et al. (2020) proposed a wearable single-channel instrument that can monitor human stress states by

analyzing changes of EEG signals in real time. This instrument evaluates EEG pressure features by analyzing the frontal asymmetry and uses machine learning classifiers such as SVM to classify EEG, with high classification accuracy. Gannouni et al. (2020) proposed VAD model to extract, select and classify EEG signals from three dimensions of valence, arousal, and dominance, respectively.

Cluster 4: Deep Learning and Emotion Recognition

Deep learning and emotion recognition cluster (blue) dominate by keywords such as emotion recognition, emotion classification, deep learning, affective computing, convolutional neural network (CNN), human computer interaction, medical signal processing, signal processing, etc.

Deep learning is a deep machine learning model, which is often used in a variety of supervised pattern recognition problems. Deep learning mainly includes CNN (Lecun et al., 1998), deep belief nets (DBN) (Hinton et al., 2006), recurrent neural network (RNN) (Jieun et al., 2020), etc. Different types of deep learning apply to different scenarios. For example, CNN is mostly used in image processing, and RNN is often used in natural language processing.

Cluster 4 focuses on deep learning and emotion recognition, especially on CNN and emotion recognition. CNN is a feedforward neural network, which carries out convolution and pooling operations on the basis of the neural network. CNN is widely used in the field of emotion recognition. Chen et al. (2021) proposed data augmentation method borderline-synthetic minority oversampling technique based on CNN. Firstly, EEG signals were preprocessed, and then frequency domain features were extracted by data enhancement algorithm. Finally, the valence and arousal dimensions of emotions were classified by CNN. Hwang et al. (2020) proposed a new emotion recognition method using CNN, which can ensure the comprehensiveness of information. Authors test the method on SEED data set, and their results show that the method is effective and accurate. Li et al. (2018) believed that traditional machine learning has the problem of overfitting in the process of emotion recognition based on EEG. Therefore, they proposed to use hierarchical convolutional neural network (HCNN) for emotion classification, and proved through experiments that HCNN is better at emotion recognition than traditional shallow models (such as SVM, KNN), especially for the recognition of Beta wave and Gamma wave.

Cluster 5: Neurofeedback and Self-Regulation

Cluster 5 is the red part shown in **Figure 5**, which is named neurofeedback and self-regulation. Cluster 5 mainly includes brain computer interface, fMRI, real-time fMRI, neurofeedback, prefrontal cortex, self-regulation, emotions, virtual reality, children, etc.

Neurofeedback technology can feedback brain activities to individuals in real time, requiring individuals to autonomously regulate feedback signals, so as to realize self-management of brain activities and ultimately regulate individual behaviors (Sitaram et al., 2017). Neurofeedback is a channel for BCI to consciously access neural activity (Paret et al., 2018), which makes it possible to realize the automatic regulation of brain activity and has great application potential in enhancing and restoring emotion and cognitive (Lorenzetti et al., 2018).

Cluster 5 focuses on neurofeedback and self-regulation. Compared with the early EEG based neurofeedback technology, Cluster 1 pays more attention to the new real-time functional magnetic resonance imaging neurofeedback (rtfMRI-NFB) technology. This new neurofeedback technique requires participants to modulate brain activity during training by changing the oxygen-dependent signal index, which in turn achieves goals such as emotional and cognitive regulation (Herwig et al., 2019). Lorenzetti et al. (2018) proposed a multi-mode VR/rtfMR-NFB protocol as a brain activity intervention tool. It can provide virtual environments as BCI, play different music to stimulate users' emotions and use personalized strategies to enhance the emotional intensity. Neurofeedback training is a clinical application of BCI and is widely used in children (such as children with developmental trauma and poor children) to regulate negative emotions, improve behavior, and ability. Rogel et al. (2020) suggested that neurofeedback training could reduce symptoms of post-traumatic stress disorder in children with developmental trauma and confirmed this idea using a randomized controlled study method. Antle et al. (2018) took poor children as the research object and proposed a neurofeedback system called Mind-Full to help poor children regulate negative emotions and focus their attention.

We use VOSviewer software to conduct the co-occurrence analysis of keywords, and analyze main research contents of each cluster in detail. Through reading articles on BCI and emotion research, we find that BCI is a communication mode that does not depend on peripheral nerves and muscles, but is completed by our brain to control the output of commands (Ash and Benson, 2018). The workflow of emotional brain computer interface generally includes five steps: emotion elicitation, signal acquisition, signal preprocessing, feature extraction, and classification, feedback and regulation. The five clusters we studied were divided according to the affective BCI process and the technology used.

Cluster 1 (External stimulus and ERP) corresponds to emotion elicitation, and the use of ERP can monitor people's psychological responses to external stimulus. Cluster 2 (EEG and information collection) is related to the signal acquisition process, and EEG is used to record brain activity. Cluster 3 (SVM and information processing) and cluster 4 (Deep learning and emotion recognition) both correspond to information processing. Cluster 3 focuses on the application of SVM in processing EEG signals, and cluster 4 focuses on the application of deep learning in emotion recognition. Cluster 5 (Neurofeedback and self-regulation) is related to feedback and regulation process. Neurofeedback training can help subjects to autonomously regulate brain activities. The purpose of

keyword cluster analysis is to construct the research framework of affective BCI and help readers better understand research contents in this field.

RESEARCH AGENDA

In this part, we use VOSviewer software to represent the time sequence of keywords by using graphical method. The color of the node represents the first time the keyword appeared, gradually transitioning from cool color to warm color. Blue represents the earliest year, and the research content is more classic. Yellow represents the latest year, and the research content is more cutting-edge. **Figure 6** shows the time sequence of keywords in the 249 articles, which can help researchers understand research trends of BCI's application in the field of emotion more clearly and intuitively.

By observing the newly emerging keywords and corresponding articles, we find that the existing researches are devoted to developing new technologies or new models to improve the accuracy of emotion recognition and have achieved good results, but there are still some shortcomings. For example, the article containing the keyword (transfer learning) hopes to solve the influence of individual differences on emotion recognition, because individual differences hinder the large-scale promotion of BCI. Although this problem has attracted the attention of scholars, there is still a lack of effective methods to solve this problem. This study proposes four future research directions, which will be elaborated below.

Individual Differences Research

Due to the physiological differences between different users, for the same emotional stimulus, different individuals will induce different emotions, and physiological signals. The existence of individual differences creates a barrier to mass distribution of BCI. In order to solve the problem of individual differences, researchers use transfer learning to improve the generalization ability of emotional computing model (Wu et al., 2020). There are two methods of transfer learning: domain adaptation and domain generalization, both of which have their advantages and disadvantages. How to balance application domain adaptation and domain generalization to improve emotion recognition performance is a problem worth discussing. At present, there are still few studies on transfer learning, which is worth further study by scholars.

The Extended Research of Brain-Computer Interface Application Scenarios

At present, BCI is mainly applied in medical field, game and entertainment field, learning and education field and military field (Vlek et al., 2012). With the increasing maturity of BCI technology, BCI will be applied to more fields in the future. In the future, social requirements for intelligent robots are becoming higher and higher, and BCI robots will have excellent development prospects. BCI smart home will be a direction for the future development of smart home. BCI smart home can meet needs of users more easily and conveniently, provided that BCI technology is mature enough. Existing studies have confirmed that using BCI to monitor users' emotions in entertainment scenes such as music and movies can provide real-time and personalized services for users. In the future, with user permission, BCI can be applied to more entertainment scenarios.

Privacy Risk Research

The protection of user information should be considered during the use of BCI. In the use of BCI, the disclosure of user information will lead to the illegal acquisition and utilization of users' ideas, and even the disclosure of users' thinking. For example, businesses may analyze users' preferences and spending power based on their brain signals and induce unfair consumption (Jebari, 2013). Facing different application scenarios, privacy risks of BCI applications are different. We need to search possible countermeasures for different problems (Li et al., 2015). At present, researches on ethical issues in BCI have just begun. How to develop new technologies to protect user information is worthy of scholars' attention and research (Agarwal et al., 2019). For different communication modes, such as brain to brain and brain to internet, future research needs to find new coping methods combined with BCI life cycle (Bernal et al., 2021).

Others

The current application of BCI in the emotion field is limited to a certain technology or a certain model. The application of BCI should be a complete process, which requires researchers to conduct integrated research from the perspective of the whole process. Therefore, researchers should carry out more extensive cooperation and actively carry out interdisciplinary communication.

CONCLUSION

In this study, literature related to affective BCI were collected and sorted out from WOS database. With the help of VOSviewer software, we conducted a scientific analysis of 249 literatures and systematically and comprehensively constructed the affective BCI research framework.

First, we found that the number of publications on emotional BCI is increasing year by year. Researches on affective BCI mainly focus on engineering, computer science, neurosciences neurology and psychology. Peoples R China, United States, and Germany have the largest number of publications. This study listed the top 10 authors with most publications and analyzed collaboration networks among authors. The top 10 most cited articles were also listed in this study.

Second, it is worth noting that we used the keyword co-occurrence function of VOSviewer software to conduct theme

clustering analysis on keywords of 249 literatures. Clustering results were divided into 5 categories, which were named as external stimulus and ERP, EEG and information collection, SVM and information processing, deep learning and emotion recognition, neurofeedback and self-regulation, respectively. These five clusters are closely related to the BCI workflow.

Third, in the future research part, we pointed out individual differences research, privacy risk research, the extended research of BCI application scenarios and others deserve further study.

This study visually analyzes the application of BCI in the field of emotion and builds a framework for existing research. The results of this study reveal the shortcomings of current affective research based on BCI. The conclusions are helpful to guide the follow-up practitioners to actively use the new technology for emotion analysis.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This research was partly supported by the National Natural Science Foundation of China (NNSFC) (Grant No. 72172052).

REFERENCES

Agarwal, A., Dowsley, R., Mckinney, N. D., Wu, D. R., Lin, C. T., De Cock, M., et al. (2019). Protecting privacy of users in brain-computer interface applications. *IEEE Transac. Neural Syst. Rehabil. Eng.* 27, 1546–1555. doi: 10.1109/TNSRE.2019.2926965

Alarcao, S. M., and Fonseca, M. J. (2019). Emotions recognition using EEG signals: a survey. *IEEE Transac. Affect. Comput.* 10, 374–393. doi: 10.1109/TAFFC.2017.2714671

Antle, A. N., Chesick, L., Sridharan, S. K., and Cramer, E. (2018). East meets west: a mobile brain-computer system that helps children living in poverty learn to self-regulate. *Personal and Ubiquitous Comput.* 22, 839–866. doi: 10.1007/s00779-018-1166-x

Arpaia, P., Moccaldi, N., Prevete, R., Sannino, I., and Tedesco, A. (2020). A wearable EEG instrument for real-time frontal asymmetry monitoring in worker stress analysis. *IEEE Transac. Instrument. Measur.* 69, 8335–8343. doi: 10.1109/TIM.2020.2988744

Ash, C., and Benson, P. J. (2018). Decoding brain-computer interfaces. *Science* 360, 615–617. doi: 10.1126/science.360.6389.615-h

Ashforth, B. E., and Humphrey, R. H. (1995). Emotion in the workplace: a reappraisal. *Hum. Rel.* 48, 97–125. doi: 10.1177/001872679504800201

Atkinson, J., and Campos, D. (2016). Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Exp. Syst. Appl.* 47, 35–41. doi: 10.1016/j.eswa.2015.10.049

Baier-Fuentes, H., Merigó, J. M., Amorós, J. E., and Gaviria-Marin, M. (2019). International entrepreneurship: a bibliometric overview. *Int. Entrepr. Manag. J.* 15, 385–429. doi: 10.1007/s11365-017-0487-y

Bakardjian, H., Tanaka, T., and Cichocki, A. (2011). Emotional faces boost up steady-state visual responses for brain-computer interface. *Neuroreport* 22, 121–5. doi: 10.1097/WNR.0b013e32834308b0

Barrett, L. F., Mesquita, B., Ochsner, K. N., and Gross, J. J. (2007). The experience of emotion. *Annu. Rev. Psychol.* 58, 373–403. doi: 10.1146/annurev.psych.58.110405.085709

Bernal, S. L., Celdran, A. H., Perez, G. M., Barros, M. T., and Balasubramaniam, S. (2021). Security in brain-computer interfaces: state-of-the-art, opportunities, and future challenges. *ACM Comput. Surv.* 54, 1–35. doi: 10.1145/3427376

Butt, A. N., Choi, J. N., and Jaeger, A. M. (2005). The effects of self-emotion, counterpart emotion, and counterpart behavior on negotiator behavior: a comparison of individual-level and dyad-level dynamics. *J. Organiz. Behav.* 26, 681–704. doi: 10.1002/job.328

Caria, A., Veit, R., Sitaram, R., Lotze, M., Weiskopf, N., Grodd, W., et al. (2007). Regulation of anterior insular cortex activity using real-time fMRI. *Neuroimage* 35, 1238–1246. doi: 10.1016/j.neuroimage.2007.01.018

Charland, P., Leger, P. M., Senecal, S., Courtemanche, F., Mercier, J., Skelling, Y., et al. (2015). Assessing the multiple dimensions of engagement to characterize learning: a neurophysiological perspective. *J. Vis. Exp.* 2015:e52627. doi: 10.3791/52627

Chen, Y., Chang, R., and Guo, J. (2021). Effects of data augmentation method Borderline-SMOTE on emotion recognition of EEG signals based on convolutional neural network. *IEEE Access* 9, 47491–47502. doi: 10.1109/ACCESS.2021.3068316

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Lear.* 20, 273–297. doi: 10.1007/BF00994018

Daly, I., Williams, D., Kirke, A., Weaver, J., Malik, A., Hwang, F., et al. (2016). Affective brain–computer music interfacing. *J. Neural Eng.* 13:046022. doi: 10.1088/1741-2560/13/4/046022

Daly, I., Williams, D., Malik, A., Weaver, J., Kirke, A., Hwang, F., et al. (2020). Personalised, multi-modal, affective state detection for hybrid brain-computer music interfacing. *IEEE Transac. Affect. Comput.* 11, 111–124. doi: 10.1109/TAFFC.2018.2801811

Eaton, J., Williams, D., and Miranda, E. (2015). The space between us: evaluating a multi-user affective brain-computer music interface. *Brain Comput. Interf.* 2, 103–116. doi: 10.1080/2326263X.2015.1101922

Ganin, I. P., Kosichenko, E. A., Sokolov, A. V., Ioannisyanc, O. M., Arefev, I. M., Basova, A. Y., et al. (2019). Adapting the p300 brain-computer interface technology to assess condition of anorexia nervosa patients. *Bull. Russ. State Med. Univ.* 8, 32–38. doi: 10.24075/brsmu.2019.022

Gannouni, S., Aledaily, A., Belwafi, K., and Aboalsamh, H. (2020). Adaptive emotion detection using the valence-arousal-dominance model and EEG brain rhythmic activity changes in relevant brain lobes. *IEEE Access* 8, 67444–67455. doi: 10.1109/ACCESS.2020.2986504

Grissmann, S., Spuler, M., Faller, J., Krumpe, T., Zander, T. O., Kelava, A., et al. (2020). Context sensitivity of EEG-based workload classification under different affective valence. *IEEE Transac. Affect. Comput.* 11, 327–334.

Grissmann, S., Zander, T. O., Faller, J., Bronstrup, J., Kelava, A., Gramann, K., et al. (2017). Affective aspects of perceived loss of control and potential implications for brain-computer interfaces. *Front. Hum. Neurosci.* 11:370. doi: 10.3389/fnhum.2017.00370

Herwig, U., Lutz, J., Scherpiet, S., Scheerer, H., Kohlberg, J., Opialla, S., et al. (2019). Training emotion regulation through real-time fMRI neurofeedback of amygdala activity. *Neuroimage* 184, 687–696. doi: 10.1016/j.neuroimage.2018.09.068

Hinton, G. E., Osindero, S., and Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Comput.* 18, 1527–1554. doi: 10.1162/neco.2006.18.7.1527

Hwang, S., Hong, K., Son, G., and Byun, H. (2020). Learning CNN features from DE features for EEG-based emotion recognition. *Patt. Anal. Appl.* 23, 1–13. doi: 10.1007/s10044-019-00860-w

Jebari, K. (2013). Brain machine interface and human enhancement: an ethical review. *Neuroethics* 6, 617–625. doi: 10.1007/s12152-012-9176-2

Jieun, P., Dokkyun, Y., and Sangmin, J. (2020). Analysis of recurrent neural network and predictions. *Symmetry* 12:615. doi: 10.3390/sym12040615

Khare, S. K., and Bajaj, V. (2021). An evolutionary optimized variational mode decomposition for emotion recognition. *IEEE Sens. J.* 21, 2035–2042. doi: 10.1109/JSEN.2020.3020915

Kleih, S. C., and Kubler, A. (2013). Empathy, motivation, and P300 BCI performance. *Front. Hum. Neurosci.* 7:642. doi: 10.3389/fnhum.2013.00642

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791

Li, J. P., Zhang, Z. X., and He, H. G. (2018). Hierarchical convolutional neural networks for EEG-based emotion recognition. *Cogn. Comput.* 10, 368–380. doi: 10.1007/s12559-017-9533-x

Li, Q. Q., Ding, D., and Conti, M. (2015). "Brain-computer interface applications: security and privacy challenges" *IEEE Conference on Communications and Network Security*. (France: IEEE). 663–666.

Liberati, G., Rocha, J., Heiden, L., Raffone, A., Birbaumer, N., Belardinelli, M. O., et al. (2012). Toward a brain-computer interface for Alzheimer's disease patients by combining classical conditioning and brain state classification. *J. Alzheimers Dis.* 31, S211–S220. doi: 10.3233/JAD-2012-112129

Lin, Y. P., and Jung, T. P. (2017). Improving EEG-based emotion classification using conditional transfer learning. *Front. Hum. Neurosci.* 11:334. doi: 10.3389/fnhum.2017.00334

Lorenzetti, V., Melo, B., Basilio, R., Suo, C., Yucel, M., Tierra-Criollo, C. J., et al. (2018). Emotion regulation using virtual environments and real-time fMRI neurofeedback. *Front. Neurol.* 9:390. doi: 10.3389/fneur.2018.00390

Mowla, R., Cano, R. I., Dhuyvetter, K. J., and Thompson, D. E. (2020). Affective brain-computer interfaces: choosing a meaningful performance measuring metric. *Comput. Biol. Med.* 126:104001. doi: 10.1016/j.compbiomed.2020.104001

Pan, J. H., Xie, Q. Y., Huang, H. Y., He, Y. B., Sun, Y. P., Yu, R. H., et al. (2018). Emotion-related consciousness detection in patients with disorders of consciousness through an EEG-based BCI system. *Front. Hum. Neurosci.* 12:198. doi: 10.3389/fnhum.2018.00198

Paret, C., Zahringer, J., Ruf, M., Gerchen, M. F., Mall, S., Hendler, T., et al. (2018). Monitoring and control of amygdala neurofeedback involves distributed information processing in the human brain. *Hum. Brain Mapp.* 39, 3018–3031. doi: 10.1002/hbm.24057

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Pun, T., Alecu, T. I., Chanel, G., Kronegg, J., and Voloshynovskiy, S. (2006). Brain-computer interaction research at the computer vision and multimedia laboratory, University of Geneva. *IEEE Transac. Neural Syst. Rehabil. Eng.* 14, 210–213. doi: 10.1109/TNSRE.2006.875544

Rogel, A., Loomis, A. M., Hamlin, E., Hodgdon, H., Spinazzola, J., and van der Kolk, B. (2020). The impact of neurofeedback training on children with developmental trauma: a randomized controlled study. *Psychol. Trauma Theor. Res. Pract. Policy* 12, 918–929. doi: 10.1037/tra0000648

Roundy, P. (2014). Doing good by telling stories: emotion in social entrepreneurship communication. *J. Small Bus. Strat.* 24, 41–68. doi: 10.5465/ambpp.2014.10318abstract

Sexton, T., Brundage, M. P., Hoffman, M., and Morris, K. C. (2017). "Hybrid datafication of maintenance logs from AI-assisted human tags" *IEEE International Conference on Big Data*. (Boston: IEEE). 1769–1777. doi: 10.1109/BigData.2017.8258120

Shan, B., Pu, Y., Chen, B., and Lu, S. (2021). New technologies' commercialization: the roles of the leader's emotion and incubation support. *Front. Psychol.* 12:710122. doi: 10.3389/fpsyg.2021.710122

Sitaram, R., Caria, A., Veit, R., Gaber, T., Ruiz, S., and Birbaumer, N. (2014). Volitional control of the anterior insula in criminal psychopaths using real-time fMRI neurofeedback: a pilot study. *Front. Behav. Neurosci.* 8:344. doi: 10.3389/fnbeh.2014.00344

Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., et al. (2017). Closed-loop brain training: the science of neurofeedback. *Nat. Rev. Neurosci.* 18, 86–100. doi: 10.1038/nrn.2016.164

Van der Heiden, L., Liberati, G., Sitaram, R., Kim, S., Jaskowski, P., Raffone, A., et al. (2014). Insula and inferior frontal triangularis activations distinguish between conditioned brain responses using emotional sounds for basic BCI communication. *Front. Behav. Neurosci.* 8:247. doi: 10.3389/fnbeh.2014.00247

Velt, H., Torkkeli, L., and Laine, I. (2020). Entrepreneurial Ecosystem Research: Bibliometric Mapping of the Domain. *J. Bus. Ecosyst.* 1, 43–83. doi: 10.4018/JBE.20200701.oa1

Vlek, R. J., Steines, D., Szibbo, D., Kubler, A., Schneider, M. J., Haselager, P., et al. (2012). Ethical issues in brain-computer interface research, development, and dissemination. *J. Neurol. Phys. Ther.* 36, 94–99. doi: 10.1097/NPT.0b013e31825064cc

Wang, X. W., Nie, D., and Lu, B. L. (2014). Emotional state classification from EEG data using machine learning approach. *Neurocomputing* 129, 94–106. doi: 10.1016/j.neucom.2013.06.046

Wörfel, P. (2019). Unravelling the intellectual discourse of implicit consumer cognition: A bibliometric review. *J. Retail. Consum. Serv.* 61:101960. doi: 10.1016/j.jretconser.2019.101960

Wu, D., Xu, Y., and Lu, B. L. (2020). Transfer learning for EEG-based brain-computer interfaces: a review of progress made since 2016. *IEEE Transac. Cogn. Dev. Syst.* 2020, 1–1. doi: 10.1109/TCDS.2020.3007453

Zaheer, H., Breyer, Y., and Dumay, J. (2019). Digital entrepreneurship: An interdisciplinary structured literature review and research agenda. *Technol. Forecast. Soc. Change* 148:119735. doi: 10.1016/j.techfore.2019.119735

Zander, T. O., and Kothe, C. (2011). Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *J. Neural Eng.* 8:025005. doi: 10.1088/1741-2560/8/2/025005

Zheng, W. L., Liu, W., Lu, Y. F., Lu, B. L., and Cichocki, A. (2019). EmotionMeter: a multimodal framework for recognizing human emotions. *IEEE Transac. Cybernet.* 49, 1110–1122. doi: 10.1109/TCYB.2018.2797176

Zhou, S. J., Allison, B. Z., Kubler, A., Cichocki, A., Wang, X. Y., and Jin, J. (2016). Effects of background music on objective and subjective performance measures in an auditory BCI. *Front. Comput. Neurosci.* 10:105. doi: 10.3389/fncom.2016.00105

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Yan, Liu, Shan, Zhang and Pu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 03 November 2021
doi: 10.3389/fpsyg.2021.782379

Foreign Language Teachers' Emotion Recognition in College Oral English Classroom Teaching

*Yanyun Dai**

School of Western Languages, Anhui International Studies University, Hefei, China

One of the significant courses in Chinese universities is English. This course is usually taught by a foreign language instructor. There will, however, necessarily be some communication hurdles between "foreign language teachers" and "native students." This research presents an emotion recognition method for foreign language teachers in order to eliminate communication barriers between teachers and students and improve student learning efficiency. We discovered four factors of emotion recognition through literature analysis: smile, eye contact, gesture, and tone. We believe that differences in foreign language teachers' performance in these four areas will have an impact on students' emotion recognition and, as a result, on their learning efficiency. The influence of the foreign language teacher's eye contact and gestures is larger (the weight of a single variable accounted for 30% or more) in the decision whether can improve the students' classroom learning efficiency, according to 43 of the questionnaire data analysis. The second is the tone and smile (the weight of a single variable accounted for between 10 and 20%). Our research contributes to the body of knowledge on emotion recognition in university foreign language teachers by presenting a practical method for recognizing emotion in foreign language teachers. We recommend that college foreign language teachers pay attention to eye and gesture communication with students in English classrooms based on the findings. By enriching the style of emotional expression in class, college language teachers, particularly foreign language teachers, can improve communication, and connection with students.

Keywords: emotion recognition, college oral English, foreign language teacher, teaching, emotion recognition in college oral English

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Yufeng Yao,
Changshu Institute of Technology, China
Tongguang Ni,
Changzhou University, China

***Correspondence:**
Yanyun Dai
daisydyy@126.com

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *24 September 2021*
Accepted: *15 October 2021*
Published: *03 November 2021*

Citation:
Dai Y (2021) Foreign Language Teachers' Emotion Recognition in College Oral English Classroom Teaching. Front. Psychol. 12:782379. doi: 10.3389/fpsyg.2021.782379

INTRODUCTION

In China, the majority of universities offer English language courses. Moreover, these courses are all taught by foreign native English speakers. There are many advantages to employing foreign language teachers as lecturers of college English courses. For example, as a native speaker, their pronunciation is very pure, the language logic is very authentic, and the description of transactions, and scenes can be very appropriate (Chen and Lee, 2011). However, it is inevitable for Chinese students to encounter some obstacles in their English learning because the teacher is a foreigner. For example, students with weak English foundation may find it difficult to keep up with the speed of the teacher's speech and understand the main points of the teacher's lecture. In China, where Mandarin is the official language, English is not used much as a second language.

In fact, it is an effective way to increase the understanding of foreign language teachers by identifying their emotions (Lorette and Dewaele, 2015). The teacher's smile, body movements, intonation, etc., can all become the reference way for students to judge context and understand semantics (Gong and Rina, 2021). Emotion recognition refers to that the computer analyses and processes the signals collected from the sensor, so as to get the emotional state of the other party (person) (Cowie et al., 2001). At present, there are two methods for emotion recognition, one is to detect physiological signals such as respiration, heart rhythm, and body temperature, and the other is to detect emotional behaviors such as facial feature expression recognition, voice emotion recognition, and posture recognition (Kwon et al., 2003). In college English class, students can recognize the emotion by observing the facial expression, pronunciation, and posture of foreign language teachers, so as to help them understand the course content.

Many scholars have studied the emotion recognition of foreign language teachers in English classrooms. For example, (Chen and Lee, 2011) proposed an embedded human emotion recognition system to help teachers reduce individual learners' language learning anxiety. Lorette and Dewaele (2019) studied the relationship between visual, sound, speech, and emotion recognition ability, and multilingual ability in foreign language teaching. Barcelos and Ruohotie-Lyhty (2018) explore the relationship between emotion and belief and how this understanding can be used to support the development of language teachers. The above examples are all studies of different regions. However, in mainland China, emotion recognition is particularly important for foreign language teachers in English classes due to the widespread Chinese learning. And there's not much research in this area. Therefore, the research question for this study is: What factors are employed to identify emotion recognition in foreign English teachers? What is the relative importance of these variables? In view of this, this paper mainly does the following work. First, we propose the factors that can be used for emotion recognition of foreign language teachers. Secondly, we analyze the influence degree of each factor by R software, nnet package and garson algorithm. Our research has enriched the literature on the field of emotion recognition for foreign language teachers and, in practice, has helped to guide students to assist classroom learning by identifying teachers' emotions.

The rest of the paper is arranged as follows: the second section describes the work related to emotion recognition of foreign language teachers; the third section presents the analytical method and experimental results; the results are discussed in the fourth section; in the fifth and last section, we put forward the research conclusions, limitations, and possible future research directions.

RELATED WORK

Emotion Recognition and Its Role

Professor Rosalind W. Picard from the Media Lab of Massachusetts Institute of Technology first proposed to create a computer system that can sense, recognize, and understand human emotions and make intelligent, sensitive and friendly responses by recognizing emotional signals of human body (Picard, 1997). At present, many studies on emotion recognition focus on text [e.g., (Yoon et al., 2018; Batbaatar et al., 2019)], facial expressions [e.g., (Liliana, 2019; Wang et al., 2020)], and speech recognition [e.g., (Alu et al., 2017; Tao et al., 2018)]. The emergence of these researches, especially the development and application of emotion recognition intelligent system, makes machines have "emotion." We also make use of scholars' research results to carry out emotional recognition on foreign language teachers' teaching posture, facial expressions and other aspects.

Emotion recognition is widely used in human life (Cai et al., 2021). For example, the application of emotion recognition in vehicle driving (Wu et al., 2021) can monitor the state of drivers and identify emotions from their voice, body language and facial expressions, so as to help reduce the occurrence of traffic accidents (Vögel et al., 2018). The application of emotion recognition in the field of elderly care can increase verbal communication and interaction with the elderly and relieve the loneliness of the elderly group (Wang et al., 2018). In addition, (Estrada et al., 2020) introduced emotion recognition into students' language teaching classes to identify students' expression skills. It can be seen that the application of emotion recognition in all walks of life has a broad basis. This provides a literature basis for this paper to evaluate the performance of foreign language teachers by using emotion recognition.

Emotion Recognition Factors for College Foreign Language Teachers

Based on different perspectives and environments, scholars have conducted diversified studies on the influencing factors of emotion recognition of college foreign language teachers. For example, (Unsworth and Mills, 2020) proposed that facial expressions and gestures could be used as evaluation factors for emotion recognition. Similarly, (Tonguç and Ozkara, 2020), (Rieffe and Wiefferink, 2017) also emphasize the important role of facial expressions in emotion recognition. In addition to the above factors, (Haq and Jackson, 2011) proposed that audio and vision should also be used as important reference aspects of emotion recognition.

According to the above literature analysis, emotion recognition is widely used in human life, work, and study. At present, in college language courses, emotion recognition technology has become an important way to assist teachers (Liang et al., 2021) to understand students' performance [e.g., (Unsworth and Mills, 2020)]. However, previous studies have paid little attention to how students identify emotions through different aspects of foreign language teachers. The research on this issue is very necessary, because it helps students to better understand the teaching content of foreign language teachers.

METHODS AND RESULTS

The research process in this paper mainly consists of four steps, as shown in **Figure 1**. The first step is questionnaire

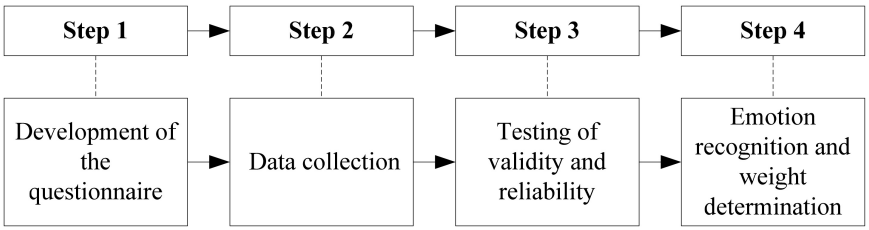


FIGURE 1 | Research steps.

development, the second step is data collection, the third step is the reliability and validity test of the scale and questionnaire, and the fourth step is emotion identification and weight calculation of independent variables.

Questionnaire Development

Based on the above analysis of the emotion recognition factors of college foreign language teachers, we determined four factors as independent variables of this paper: smile, eye contact, gesture, and tone. We hypothesize that students can effectively identify the emotion of foreign language teachers through the judgment of these four factors, so as to improve their classroom learning efficiency. Therefore, we take students' classroom learning efficiency as the dependent variable. Accordingly, we developed the scale shown in **Table 1**. It should be noted that the four independent variables are numerical variables, and students can rate the teacher's performance in smile, eye contact, gesture, and tone. The score ranges from 1 to 7, and the higher the score, the better the performance of foreign language teachers in this aspect, and vice versa. Whether it is helpful to improve the classroom learning efficiency is classified value, where 1 represents helpful to improve the classroom learning efficiency, 0 represents no help to improve the classroom learning efficiency.

Data Collection

After the questionnaire was designed, we determined the respondents of this study. The students in this survey need to meet the following two requirements: (1) have studied college English courses in the past 1 year; (2) their teacher of this course is a foreigner (native English speaker). Our survey was carried out by filling in offline. A total of 43 students were surveyed during a week (from April 23, 2021 to April 29, 2021). Of the 43 students, 52% are freshmen and 48% are sophomores. The male to female ratio is 4:6. The age is mainly between 19 and 22 years old. 25 students thought that their foreign language teachers performed better in these four variables, which improved their classroom learning efficiency, while 18 students held the opposite view.

TABLE 1 | Measurement scale.

No.	Variables	Abbreviations	Mark
1	Smile	SM	1–7
2	Eye contact	EC	1–7
3	Gesture	GE	1–7
4	Tone	TO	1–7
5	The improvement of classroom learning efficiency	ICLE	1, 0

Reliability and Validity Tests

We conducted reliability and validity tests on the scale designed in this paper and the collected questionnaires (Brink, 1991; Roberts and Priest, 2006), and the results are shown in **Tables 2**, **3**. It can be seen from **Table 2** that Cronbach's Alpha coefficient is greater than 0.8, indicating that the scale and questionnaire data have high reliability. The data in **Table 3** show that both Kaiser-Meyer-Olkin (KMO) and Bartlett's test meet relevant thresholds, indicating that our scale and questionnaire data have good validity (Brink, 1991).

Emotion Recognition and Weight Calculation

In this paper, we use R software and nnet package to carry out the emotion recognition of foreign language teachers and the weight measurement of the variables. R is a language and operating environment used for statistical analysis and mapping (Ihaka and Gentleman, 1996), as well as a free and open source software belonging to GNU system (Maronna et al., 2019).

Nnet packets are feedforward neural networks and multinomial log-linear models that can be used to process forward and back propagation data until convergence (Ripley et al., 2016). In this paper, we set the parameters of the model to: size = 5, rang = 0.1, decay = 5e−2, and maxit = 5,000.

TABLE 2 | Reliability statistics.

Cronbach's alpha	Cronbach's alpha based on standardized items	N of items
0.893	0.930	5

TABLE 3 | Kaiser-Meyer-Olkin (KMO) and Bartlett's test.

Kaiser-Meyer-Olkin measure of sampling adequacy.		**0.846**
Bartlett's test of sphericity	Approx. Chi-Square	177.400
	df	10
	Sig.	0.000

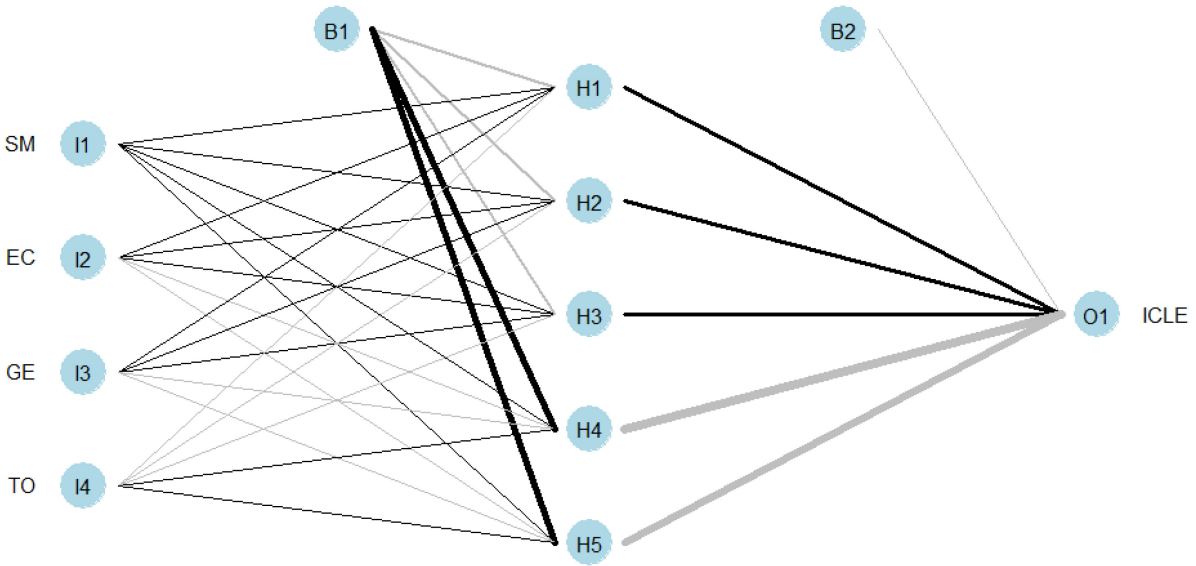


FIGURE 2 | Calculation model.

By drawing the model in R language, we get the calculation model shown in **Figure 2**. **Figure 2** shows that there are five nodes in a single hidden layer.

Then, we used the NeuralNetTools package (Beck, 2018) to obtain the relative importance (weight) of input variables (smile, eye contact, gesture, and tone) in the neural network (Zhao et al., 2020; Chu et al., 2021; You et al., 2021; Zhang et al., 2021) through garson algorithm (Ghanizadeh et al., 2020). As shown in **Figure 3**, it can be found that the foreign language teacher's eye contact and gesture have a greater influence on the decision of whether to improve students' classroom learning efficiency (the weight of each variable is above 30%), followed by tone and smile (the weight of each variable is between 10 and 20%).

DISCUSSION

In this paper, we analyze the factors that can be used for emotion recognition through literature review. We then identify the specific factors of emotion recognition by university foreign language teachers, which are smile, eye contact, gesture, and tone of voice, respectively. Furthermore, we use the nnet package of R and garson algorithm to draw the emotion recognition model of college foreign language teachers and determine the weight of each variable. Our study found that eye contact and gestures have a large impact on English teaching by college foreign language teachers (each weighing more than 30% for a single variable), while tone and smile have a relatively small impact (each weighing between 10 and 20% for a single variable). Based on the above findings, we propose the following implications.

Theoretical Implications

From the perspective of literature, research content, and methods, our research has three theoretical implications. Firstly, this paper enrichis the literature on emotion recognition of foreign language teachers in the university. Mandarin is the official language in China. English, as a widely used language in the world, is the second language of most people in China. Hence, in many Chinese universities, more and more foreign language teachers come from english-speaking countries. In addition to the positive effects, it will inevitably bring students a certain degree of language learning disabilities. This paper aims to improve students' learning efficiency from four aspects: smile, eye contact, gesture, and tone. In addition, the majority of the previous studies are based on the perspective of teachers to identify students' emotions (e.g., Estrada et al., 2020; Unsworth and Mills, 2020; Liang et al., 2021). However, in line with (Lorette and Dewaele, 2015) and (Gong and Rina, 2021), we believe that, from the perspective of students, identifying the lecturer's emotion can assist students better understand the subject delivered in English class by the teacher. This paper tries to identify the emotion of foreign language teachers from the perspective of students. Therefore, this paper enriches the literature on emotion recognition for foreign language teachers.

Second, we propose specific factors that can be used to identify foreign teachers' emotion. Emotion recognition is uncertain and may be disturbed by different factors in different environments. Based on the context of college English courses, this paper takes foreign language teachers as research subjects and identifies four factors (smile, eyes contact, gesture, and tone) that can be used to identify their emotions. This conclusion has certain theoretical reference value for the peer research. But there is no denying that our conclusions are suggestive and instructive. We look forward to exploring different emotion recognition factors in different courses, different environments, and different objects.

Thirdly, this paper introduces nnet package and garson algorithm into the research of emotion recognition of university foreign language teachers, and provides a feasible method of

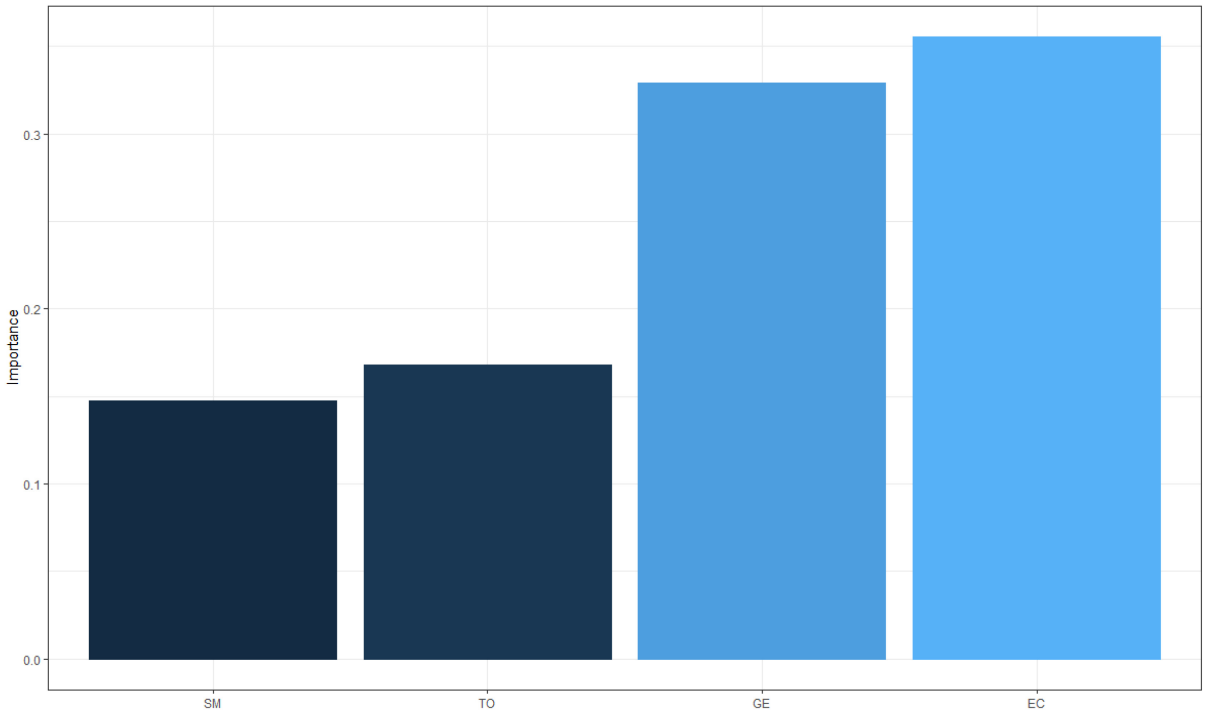


FIGURE 3 | Weight of independent variables.

emotion recognition. This method is easy to operate in R software and is applicable to both small sample and large sample data. It is worth mentioning that we can clearly calculate the weights of different independent variables. Therefore, this paper extends the application of nnet package and agrson algorithm in emotion recognition of college foreign language teachers.

Practical Implications

Our findings are likely to benefit teachers and students in college English classes. On the one hand, foreign English teachers can adjust their body movements or language habits in class to improve teaching efficiency based on our research findings; on the other hand, students can assist in understanding course content based on foreign English teachers' performance in the four factors examined in this paper. Specifically, there are two practical implications of this paper. First of all, college foreign language teachers should pay attention to eye contact and gesture communication with students in English classes. Our study found that eye contact and gestures of foreign language teachers have a significant influence on the decision of whether to improve students' classroom learning efficiency (each variable accounts for more than 30% of the weight). This suggests that rich and varied eye and gesture communication can help students deepen their understanding of the knowledge in English class and improve their learning efficiency. The majority of foreign language teachers come from native English-speaking countries, so students' expectations on tone are not very strong. The reason may be that the vast majority of students think that having good pronunciation and oral English is the basic skills that foreign language teachers should have. The eye contact and gestures, as the auxiliary language expressions of foreign language teachers, have a great influence on students.

Second, in college language teaching, especially foreign language teachers can enhance communication and interaction with classmates by enriching the way of emotional expression in class. This is conducive to the improvement of students' learning efficiency. In this paper, the four factors of emotion recognition, smile, eye contact, gesture, and tone pair, all have an impact on the improvement of students' learning efficiency. Therefore, college language teachers need to recognize the importance of emotional communication in the classroom and assist teaching through smiles, eye contact, gestures, and tones of voice.

CONCLUSION

One of the essential courses in Chinese colleges is English, which is usually taught by a foreign language teacher. However, contact between "foreign language teachers" and "native Chinese students" will necessarily be hampered. This research presents an emotion recognition method for foreign language teachers in order to eliminate communication barriers between teachers and students and improve student learning efficiency. We discovered four factors of emotion recognition through literature analysis: smile, eye contact, gesture, and tone. We believe that disparities in foreign language teachers' performance in these four aspects will have an impact on students' emotion recognition and, as a result, on their learning efficiency. The influence of the foreign language

teacher's eye contact and gestures is larger (the weight of a single variable accounted for 30% or more) in the decision whether can improve the students' classroom learning efficiency, according to 43 of the questionnaire data analysis. The second is the tone and smile (the weight of a single variable accounted for between 10 and 20%). Our research contributes to the body of knowledge on emotion recognition in university foreign language teachers by presenting a practical method for recognizing emotion in foreign language teachers.

We recommend that college foreign language teachers pay attention to eye and gesture communication with students in English classrooms based on the findings. By enriching the style of emotional expression in class, college language teachers, particularly foreign language teachers, can improve communication, and connection with students.

To be clear, our research has a lot of flaws. For example, because our study's sample size is rather small, we can expand it in the future. Second, whether there are disparities in foreign language teachers' emotional performance in the classroom among universities and localities, and how students react to such differences. These are intriguing questions that should be investigated further. Third, this article simply evaluates the foreign language teacher's smile, eye contact, gestures, and tone. Dress and teaching posture (standing or sitting) are also worth investigating further. Fourth, many English teachers at Chinese universities are actually Chinese. Do their issues with emotion recognition differ from those of foreign language instructors? This is something that needs to be looked at more. Finally, there are numerous approaches for recognizing emotions. We can test multiple approaches for identifying foreign language teachers' emotions and analyze the benefits and drawbacks of each method as well as the disparities in conclusions.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/appdevcode/Emotion_Teaching.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Anhui International Studies University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YD: conceptualization, methodology, software, writing, data curation, and investigation.

FUNDING

This research was funded by the second batch of industry-university cooperation collaborative education project of the Ministry of Education in 2018 (201802084038), the major online teaching reform research project of provincial quality engineering in Anhui Province's colleges and universities (2020zdxsjg148), and the funding project for the cultivation of outstanding top talents in colleges and universities in Anhui Province (gxbjZD61) funding.

REFERENCES

Alu, D., Zoltan, E., and Stoica, I. C. (2017). Voice based emotion recognition with convolutional neural networks for companion robots. *Sci. Technol.* 20, 222–240.

Barcelos, A. M. F., and Ruohotie-Lyhty, M. (2018). "Teachers' emotions and beliefs in second language teaching: implications for teacher education," in *Emotions in Second Language Teaching*, ed. J. Martínez Agudo (Cham: Springer), 109–124. doi: 10.1007/978-3-319-75438-3_7

Batbaatar, E., Li, M., and Ryu, K. H. (2019). Semantic-emotion neural network for emotion recognition from text. *IEEE Access* 7, 111866–111878. doi: 10.1109/ACCESS.2019.2934529

Beck, M. W. (2018). NeuralNetTools: visualization and analysis tools for neural networks. *J. Stat. Softw.* 85:1. doi: 10.18637/jss.v085.i11

Brink, P. J. (1991). "Issues of reliability and validity," in *Qualitative Nursing Research: A Contemporary Dialogue*, ed. J. M. Morse (London: Sage), 164–186. doi: 10.4135/9781483349015.n20

Cai, W., Song, Y., and Wei, Z. (2021). Multimodal data guided spatial feature fusion and grouping strategy for e-commerce commodity demand forecasting. *Mobile Inform. Syst.* 2021, 1–14. doi: 10.1155/2021/5568208

Chen, C. M., and Lee, T. H. (2011). Emotion recognition and communication for reducing second−language speaking anxiety in a web−based one−to−one synchronous learning environment. *Br. J. Educ. Technol.* 42, 417–440. doi: 10.1111/j.1467-8535.2009.01035.x

Chu, W., Ho, P. S., and Li, W. (2021). An adaptive machine learning method based on finite element analysis for ultra low-k chip package design. *IEEE Trans. Comp. Packag. Manufact. Technol.* 11, 1435–1441. doi: 10.1109/TCPMT.2021.3102891

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., and Votsis, G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Process. Mag.* 18, 32–80. doi: 10.1109/79.911197

Estrada, M. L. B., Cabada, R. Z., Bustillos, R. O., and Graff, M. (2020). Opinion mining and emotion recognition applied to learning environments. *Expert Syst. Applic.* 150:113265. doi: 10.1016/j.eswa.2020.113265

Ghanizadeh, A. R., Heidarabadizadeh, N., and Jalali, F. (2020). Artificial neural network back-calculation of flexible pavements with sensitivity analysis using Garson's and connection weights algorithms. *Innov. Infrastruct. Solut.* 5, 1–19. doi: 10.1007/s41062-020-00312-z

Gong, Y., and Rina, S. (2021). Autonomous learning of foreign language based on facial emotion recognition and cloud computing. *J. Intell. Fuzzy Syst.* 40, 1–12. doi: 10.3233/JIFS-189564

Haq, S., and Jackson, P. J. B. (2011). "Multimodal emotion recognition," in *Machine Audition: Principles, Algorithms and Systems*, ed. W. Wang (Pennsylvania: IGI Global Press), 398–423. doi: 10.4018/978-1-61520-919-4.ch017

Ihaka, R., and Gentleman, R. (1996). R: a language for data analysis and graphics. *J. Comp. Graph. Stat.* 5, 299–314. doi: 10.1080/10618600.1996.10474713

Kwon, O. W., Chan, K., Hao, J., and Lee, T. W. (2003). "Emotion recognition by speech signals," in *Proceedings of the Eighth European Conference on Speech Communication and Technology*, Geneva.

Liang, L., Yin, Q., and Shi, C. (2021). Exploring proper names online and its application in English teaching in university. *ASP Trans. Comp.* 1, 24–29. doi: 10.52810/TC.2021.100024

Liliana, D. Y. (2019). "Emotion recognition from facial expression using deep convolutional neural network," in *Proceedings of the Journal of Physics: Conference Series. 2018 International Conference of Computer and Informatics Engineering 12–13 September 2018* (Bogor: IOP Publishing).

Lorette, P., and Dewaele, J. M. (2015). Emotion recognition ability in English among L1 and LX users of English. *Int. J. Lang. Cult.* 2, 62–86. doi: 10.1075/ijolc.2.1.03lor

Lorette, P., and Dewaele, J. M. (2019). The relationship between bi/multilingualism, nativeness, proficiency and multimodal emotion recognition ability. *Int. J. Biling.* 23, 1502–1516. doi: 10.1177/1367006918808042

Maronna, R. A., Martin, R. D., Yohai, V. J., and Salibián-Barrera, M. (2019). *Robust Statistics: Theory and Methods (with R)[M]*. Hoboken, NJ: John Wiley & Sons. doi: 10.1002/9781119214656

Picard, R. W. (1997). *Affective Computing*. Cambridge: MIT Press. doi: 10.1037/e526112012-054

Rieffe, C., and Wiefferink, C. H. (2017). Happy faces, sad faces: emotion understanding in toddlers and preschoolers with language impairments. *Res. Dev. Disabil.* 62, 40–49. doi: 10.1016/j.ridd.2016.12.018

Ripley, B., Venables, W., and Ripley, M. B. (2016). Package 'nnet'. *R Package Version* 7:700.

Roberts, P., and Priest, H. (2006). Reliability and validity in research. *Nurs. Standard* 20, 41–46. doi: 10.7748/ns.20.36.41.s58

Tao, F., Liu, G., and Zhao, Q. (2018). "An ensemble framework of voice-based emotion recognition system for films and TV programs," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Calgary, AB: IEEE), 6209–6213. doi: 10.1109/ICASSP.2018.8461617

Tonguç, G., and Ozkara, B. O. (2020). Automatic recognition of student emotions from facial expressions during a lecture. *Comp. Educ.* 148:103797. doi: 10.1016/j.compedu.2019.103797

Unsworth, L., and Mills, K. A. (2020). English language teaching of attitude and emotion in digital multimodal composition. *J. Second Lang. Writ.* 47:100712. doi: 10.1016/j.jslw.2020.100712

Vögel, H. J., Süß, C., Hubregtsen, T., Ghaderi, V., Chadowitz, R., André, E., et al. (2018). "Emotion-awareness for intelligent vehicle assistants: a research agenda," in *Proceedings of the 2018 IEEE/ACM 1st International Workshop on Software Engineering for AI in Autonomous Systems (SEFAIAS)* (New York, NY: IEEE), 11–15. doi: 10.1145/3194085.3194094

Wang, K., Zhu, Z., Zhang, J., and Chen, L. (2018). Speech emotion recognition of Chinese elderly people. *Web Intell.* 16, 149–157. doi: 10.3233/WEB-180382

Wang, X., Chen, X., and Cao, C. (2020). Human emotion recognition by optimally fusing facial expression and speech feature. *Signal Process. Image Commun.* 84:115831. doi: 10.1016/j.image.2020.115831

Wu, H., Chen, Y., and Qin, H. (2021). MPC based trajectory tracking for an automonous deep-sea tracked mining vehicle. *ASP Trans. Internet Things* 1, 1–13. doi: 10.52810/TIOT.2021.100062

Yoon, S., Byun, S., and Jung, K. (2018). "Multimodal speech emotion recognition using audio and text," in *Proceedings of the 2018 IEEE Spoken Language Technology Workshop (SLT)* (Athens: IEEE), 112–118. doi: 10.1109/SLT.2018.8639583

You, H., Yu, L., Tian, S., Ma, X., Xing, Y., Xin, N., et al. (2021). MC-Net: Multiple max-pooling integration module and cross multi-scale deconvolution network. *Chem. Rev.* 107, 2411–2502. doi: 10.1016/j.knosys.2021.107456

Zhang, L., Sun, L., Yu, L., Dong, X., Chen, J., Cai, W., et al. (2021). "ARFace: Attention-aware and regularization for face recognition with reinforcement learning," in *Proceedings of the IEEE Transactions on Biometrics, Behavior, and Identity Science* (Piscataway, NJ: IEEE). doi: 10.1109/TBIOM.2021.3104014

Zhao, M., Chang, C. H., Xie, W., Xie, Z., and Hu, J. (2020). Cloud shape classification system based on multi-channel cnn and improved fdm. *IEEE Access* 8, 44111–44124. doi: 10.1109/ACCESS.2020.2978090

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Dai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 08 November 2021
doi: 10.3389/fpsyg.2021.769509

Multimodal Art Pose Recognition and Interaction With Human Intelligence Enhancement

Chengming Ma, Qian Liu and Yaqi Dang*

College of Communication, Northwest Normal University, Lanzhou, China

This paper provides an in-depth study and analysis of human artistic poses through intelligently enhanced multimodal artistic pose recognition. A complementary network model architecture of multimodal information based on motion energy proposed. The network exploits both the rich information of appearance features provided by RGB data and the depth information provided by depth data as well as the characteristics of robustness to luminance and observation angle. The multimodal fusion is accomplished by the complementary information characteristics of the two modalities. Moreover, to better model the long-range temporal structure while considering action classes with sub-action sharing phenomena, an energy-guided video segmentation method is employed. And in the feature fusion stage, a cross-modal cross-fusion approach is proposed, which enables the convolutional network to share local features of two modalities not only in the shallow layer but also to obtain the fusion of global features in the deep convolutional layer by connecting the feature maps of multiple convolutional layers. Firstly, the Kinect camera is used to acquire the color image data of the human body, the depth image data, and the 3D coordinate data of the skeletal points using the Open pose open-source framework. Then, the action automatically extracted from keyframes based on the distance between the hand and the head, and the relative distance features are extracted from the keyframes to describe the action, the local occupancy pattern features and HSV color space features are extracted to describe the object, and finally, the feature fusion is performed and the complex action recognition task is completed. To solve the consistency problem of virtual-reality fusion, the mapping relationship between hand joint point coordinates and the virtual scene is determined in the augmented reality scene, and the coordinate consistency model of natural hand and virtual model is established; finally, the real-time interaction between hand gesture and virtual model is realized, and the average correct rate of its hand gesture reaches 99.04%, which improves the robustness and real-time interaction of hand gesture recognition.

Keywords: intelligent augmentation, multimodality, human ART, pose recognition, interaction

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Leilei Jiang,
Anhui Radio and TV University, China
Yili Sun,
Nanjing University, China
Linlin Guo,
Henan University, China

****Correspondence:***
Chengming Ma
xmgmcm@163.com

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *02 September 2021*
Accepted: *01 October 2021*
Published: *08 November 2021*

Citation:
Ma C, Liu Q and Dang Y (2021) Multimodal Art Pose Recognition and Interaction With Human Intelligence Enhancement. Front. Psychol. 12:769509. doi: 10.3389/fpsyg.2021.769509

INTRODUCTION

With the development and application of artificial intelligence, machine vision, human-computer interaction, and other technologies, computers are rapidly integrated into our lives and gradually change our lifestyles, in which human-computer interaction (HCI) technology, which allows close contact between humans and machines, continues to emerge as innovative and novel, which

serves as a special language for communication between humans and computers, such as body movements, facial expressions, etc., to express specific interaction intentions and achieve the exchange of information between humans and computers (Abdullah et al., 2021). This kind of interaction not only restricts people's actions, but also affects the interaction between people and computers. Furthermore, it gradually derives to the operation of autonomous actions by people, such as gestures, voice, body posture, facial expressions and so on. Initially, the traditional human-computer interaction is operated by external forces such as the mouse and keyboard, which not only constrains human behavior and action but also affects the interaction between humans and computers. Further, it is gradually derived to the operation of behavioral actions issued by humans autonomously, such as gestures, voice, body gestures, facial expressions, etc. This way makes the human-computer interaction interface change from traditional to intelligent, from graphical user interface and network user interface to multimodal and multimedia intelligent interaction interface, from unilateral human-computer interaction to the stage of cognitive understanding of human, from satisfying basic functions to attaching importance to user experience, making the intelligent human-computer interaction interface more natural and diverse (Ludl et al., 2020). Ignoring the students' hands-on operation ability during the experiment, most of the virtual experiments are inseparable from the teacher's guidance during the operation, resulting in the students' lack of independent research ability. The virtual teaching method gradually adds augmented reality (AR) technology based on simulation technology, which realizes the transformation from two-dimensional virtual space to three-dimensional real space, which is conducive to improving students' interest and enriching the experience (Islam and Iqbal, 2021). The virtual experiment classroom, taking the chemistry experiment classroom as an example, not only has the problems of difficult observation of students' experimental results, tedious experimental steps, and difficult experimental operations, but also ignores the students' hands-on ability in the experimental process, and most of the virtual experiments cannot be operated without the teacher's guidance, leading to the lack of students' ability to conduct independent research (Rastgoo et al., 2021). The main task of traditional human action recognition is to design a complete and effective algorithm so that the computer can determine the category of human action from a series of pictures or a video. The realization process mainly includes two parts: feature extraction and classifier classification. These problems cannot be solved by unimodal interaction in AR alone, and making the virtual-real interaction in virtual experiments more efficient and enhancing students' feelings of the authenticity of experiments is the main goal at present. Therefore, for domestic secondary school students' virtual chemistry experiments, there is an urgent need for a new experimental system with intelligent interaction, which not only has the characteristics of the normal operation of the experimental process, but also needs to meet the characteristics of experimental ease of operation, experimental ease of observation, and low cost, which is an intelligent experimental system (Jain et al., 2021).

Action recognition also plays a big role in sports, art performance, etc. In sports, analyzing the basic posture movements of athletes and comparing them with the standard movements, can help athletes to check the deficiencies and fill in the gaps to a great extent, and assist in training; while in art performance, especially in dance performance, the standard movements largely determine the effect of training and the success of the performance, human motion recognition can help performers to complete training and performance better through motion analysis. The main task of traditional human action recognition is to design a complete and effective algorithm to make the computer determine the human action category from a series of pictures or a video (Jegham et al., 2020). The implementation process mainly consists of two major parts: feature extraction and classifier classification. Although the traditional manually designed feature representation can accomplish the given recognition task better, due to its excessive reliance on the database itself, the designed features can only satisfy the recognition task of a single database, and cannot be effectively applied to multiple databases, lacking generalization ability (Xue et al., 2020). At the same time, in the context of information explosion and big data caused by the improvement of hardware performance and computing power, the traditional manual design feature method is even more unable to meet the task requirements because of its inability to handle massive data and then faces elimination.

To solve the problem of human action recognition, this paper focuses on three aspects of multimodal human action recognition, object recognition, and human-object interaction action recognition. Through theoretical demonstration and experimental results comparing multi-modal data and single-modal data, the validity and feasibility of the multi-modal action recognition algorithm are verified. The focus is on exploring the effect of multimodal fusion on the improvement of complex action recognition accuracy, the effect of constructing a subset of object shape features on the improvement of object recognition accuracy, and the effect of fusing different subsets of features from different modalities on the improvement of human-object interaction action recognition. In addition, the application of multimodal data in classical action recognition algorithms is reproduced. Firstly, a theoretical introduction is made for two classical action recognition algorithms. Later action recognition algorithms have been improved in terms of data processing and network structure, but the basic model theoretical basis remains unchanged. Secondly, the effectiveness and feasibility of multimodal action recognition algorithms are verified through theoretical arguments and experimental results comparing multimodal data and unimodal data. Finally, through visualization experiments, the defects of the traditional action recognition algorithm are found, and the foundation for further improvement of the algorithm is laid. After the initial recognition of the action by the action recognition network, the network fusion strategy module is also designed to give different fusion methods for various situations that may occur during the fusion

of the two networks, avoiding the misjudgment of the target object or other redundant information due to the target detection network, which may affect the overall accuracy of the action recognition algorithm.

CURRENT STATUS OF RESEARCH

Multimodal recognition includes recognition of modalities such as gesture, speech, somatic, and sensing devices, where sensing devices are more accurate in detecting human behavior and are relatively easy to interact with, and are generally used in combination with other modalities, including sensors that detect various user behaviors, Kinect, data gloves, stereo helmets, etc. (Rasipuram and Jayagopi, 2020). Gesture recognition and speech recognition are common areas of research and have become important components of perceptual user interfaces in accomplishing intelligent human-computer interaction functions that are highly structured, which makes gestures and speech an essential interaction modality in the field of human-computer interaction (Pandeya and Lee, 2021). In human-computer interaction, it is found that gesture recognition is not only affected by different contexts, multiple interpretations, and spatial and temporal changes, but also has unsolved problems until now due to the complex non-rigid nature of the human hand, while speech recognition is more susceptible to the influence of surrounding environmental factors and human factors, and has great challenges (Halfon et al., 2021). However, it is found that gesture recognition, speech recognition, and sensor sensing can complement each other, and the use of multi-modal interaction can better reduce the user's operational burden and improve the efficiency of interaction.

Therefore, if a real breakthrough in intelligence is to be achieved and the problem of the excessive operational load of unimodal interaction is solved, there is an urgent need to apply the combination of gestures, speech, and sensors to complement each other to achieve a human-computer interaction environment with more natural and intelligent features. Multimodal interaction is a very dynamic and extensive research area, and many researchers have proposed multimodal information fusion methods in recent years, mostly through multimodal integration strategies or fusion models (Kunda, 2020). Overall, multimodal fusion methods are mainly classified by level into fusion at the data layer, model (feature) layer, and decision layer (Yoon et al., 2020). Fusion at the data layer is the fusion of the entire multimodal information data into a single feature vector, which is characterized by centralized analysis and fusion of unpreprocessed data. The fusion of the model layer is the feature extraction of multimodal information and then comprehensive analysis and processing of the extracted feature information, which is characterized by the compression of sizable information, real-time processing of feature information, and the flexibility to choose the location of fusion (Poibrenski et al., 2021). Compared with video and image action recognition, skeletal point data is more robust to lighting conditions, viewpoint transformations, and changes in human form. Moreover, in complex pure action recognition (action recognition without the influence of other objects), skeleton point data has an advantage that video data cannot match (Morales et al., 2021). Regarding the action recognition of skeletal points, the traditional method of manually designed features usually represents the motion of several human skeletal nodes as the action of the whole human body. Based on the 3D coordinate position information of these skeletal nodes, the motion information is characterized and extracted from the skeletal joints using algorithms such as the K nearest neighbor classification algorithm (KNN) and dynamic time regularization (DTW).

It is eventually fed into a designed implicit Markov model (HMM) or a support vector machine (SVM) for classification. However, due to the limitations and inefficiencies of traditional machine learning algorithms and the unreliability of early data sources. Skeletal point-based action recognition algorithms have not received much attention from researchers for a long time. Complex scenes include a range of uncontrollable scenes such as the background in which the person is in the video, the environment of the camera, etc. Complex scenes can lead to reduced recognition effectiveness and decreased recognition accuracy of existing action recognition algorithms. Common complex scenes include the intensity of lighting, changes in shooting perspective, object occlusion, and camera shake itself. It is this variety of complex factors that cause the action recognition technology to not achieve the theoretical results in the practical application process. Recognition algorithms that are robust to such complex environments are in urgent need of development. More effective recognition of similar actions with object influence is performed by incorporating a target detection network and a target association module. After the initial recognition of the action by the action recognition network, a network fusion strategy module is designed to give different fusion methods for various situations that may occur during the fusion of the two networks to avoid misjudgment of the target object or other redundant information due to the target detection network, which may affect the overall accuracy of the action recognition algorithm.

INTELLIGENT AUGMENTED MULTIMODAL HUMAN ART POSE RECOGNITION AND INTERACTION ANALYSIS

Intelligent Augmented Multimodal Human Art Pose Recognition Algorithm Design

Video-based action recognition algorithms have also entered a rapid development path. The data type of action recognition has changed from the original single RGB image data to the current multi-modal data (Ou et al., 2020), including but not limited to RGB video data, depth video data, and infrared video data. Including but not limited to RGB video data, depth video data and infrared video data, this requires that the existing motion recognition algorithms not only be compatible with different basic network model structures, to find the best network model

to achieve the expected recognition accuracy, but also It must also be able to process data of different modalities, and use the complementarity of the feature information of multi-modal data to complete the recognition task. This requires existing action recognition algorithms to be compatible not only with different basic network model structures to facilitate finding the best network model to achieve the expected recognition accuracy, but also to be able to handle data from different modalities and utilize the complementary nature of the feature information of multimodal data to accomplish the recognition task.

$$\begin{cases} V_{(i,f)} = \frac{f_i^f + f_i^{f-1}}{T} \\ V_i = \left(V_{(i,i)}, \ldots, V_{(i,f)}\right) \end{cases} \tag{1}$$

$$L_{(emg,k)} = \sum_{f(emg,k) \in A_{(emg,k)}} \left(f_{(emg,n)} + C_{(emg,k)}\right) \tag{2}$$

In this section, two classical action recognition algorithms and their principles, from which most of the existing algorithms derive their theoretical basis, are described in detail, while both algorithms can handle multimodal data well and possess the effect of complementing multimodal information. Finally, the shortcomings in the classical algorithms are also found and improvements are proposed through experimental replication. In addition, the theory of the algorithm proposed in this paper is also based on the classical algorithm and is optimized and improved on this basis to achieve better results. The principle of an action recognition algorithm based on the dual-stream network is fundamentally different from the C3D algorithm, which extracts Spatio-temporal features directly from video data by integrating spatial features and temporal features. The latter directly extracts spatio-temporal features from video data by integrating spatial and temporal features. The former designed two isomorphic networks to extract the appearance and motion characteristics of the data, respectively, and the clues of the motion characteristics are provided by the timing information of the video.

$$\begin{cases} f(x) = sign\left(\sum_{i=1}^{N} a_i y_i k(x_i, x_j) - b\right) \\ sign(x) = \begin{cases} -x > 0 \\ x = 0 \\ x \le 0 \end{cases} \end{cases} \tag{3}$$

$$\min \sum_{i=1}^{N} a_i + \frac{1}{2} \sum_{i=1}^{N} \sum_{j}^{N} a_i y_j a_j y_i k(x_i^2, x_j^2) \tag{4}$$

In the spatial stream, the spatial network receives the single RGB image after frame splitting and outputs the judged action category by extracting the spatial information features of the image. Spatial information contains a variety of information such as complex background information, texture information, and shape information of the video, and doing information extraction for the pictures helps to recognize actions with the help of appearance information features of the human body and video background features. The recognition of visual modal is mainly divided into the establishment of gesture recognition model and gesture interaction stage. The recognition of voice modal is mainly keyword recognition. In addition, the spatial network based on RGB pictures can also effectively recognize various kinds of actions with object influence, for example, the action of playing basketball, RGB pictures can improve the recognition accuracy through the recognition of basketball, and then assist in the recognition of the overall action. Among them, the recognition of visual modality is mainly divided into the establishment of gesture recognition model and gesture interaction stage, and the recognition of speech modality is mainly keyword recognition. Finally, a prototype of an intelligent chemistry experiment based on an augmented reality environment is realized by fusing visual, auditory, and haptic information, and its framework diagram is shown in **Figure 1**.

During gesture recognition, there are errors in skeletal tracking due to differences in human hand size and hand movement in different positions, which affects the accuracy of gesture recognition. Because the magnitude of the hand wave is relatively small, analysis with a single frame is more difficult and less stable, while problems such as self-obscuration can occur. In the feature extraction stage, after obtaining the position of the object and the circumscribed rectangular frame, first perform feature extraction on the object in the circumscribed frame area. The existing frameworks for object recognition can be divided into manual feature-based object recognition framework and deep learning-based object recognition framework in the feature extraction phase (Mohammed and Hassan, 2020). Manual feature-based object recognition framework contains two phases, namely feature extraction and object classification. In the feature extraction phase, after obtaining the location of the object and the external rectangular box, feature extraction is first performed for the objects in the external box region. Shape features are intuitive features that describe objects. Without considering color features and texture features, people can distinguish a variety of objects based on their shape alone. Common object shape features include shape context features, chain code features, area, and horizontal Aspect ratio etc. The quality of the features is crucial to the accuracy of object recognition. Gradient-based HOG features are widely used in object recognition. HOG features divide the region inside the rectangular box into multiple sub-regions and calculate the gradient histogram for each region, and the final feature vector is the combination of the gradient histogram values of multiple sub-regions. In addition to gradient features, object color and object shape are also commonly used to describe objects. Color features mainly describe the color information of a region in an image, where features such as color histogram features and color moments are commonly used. Shape features are intuitive features that describe objects, and without considering color features and texture features, a person can distinguish multiple objects based on object shape alone, and common object shape features include shape context features, chain code features, area, and aspect ratio.

Different modal data can provide different information features, while each has advantages for different kinds of recognition. Since depth data is presented as grayscale images and

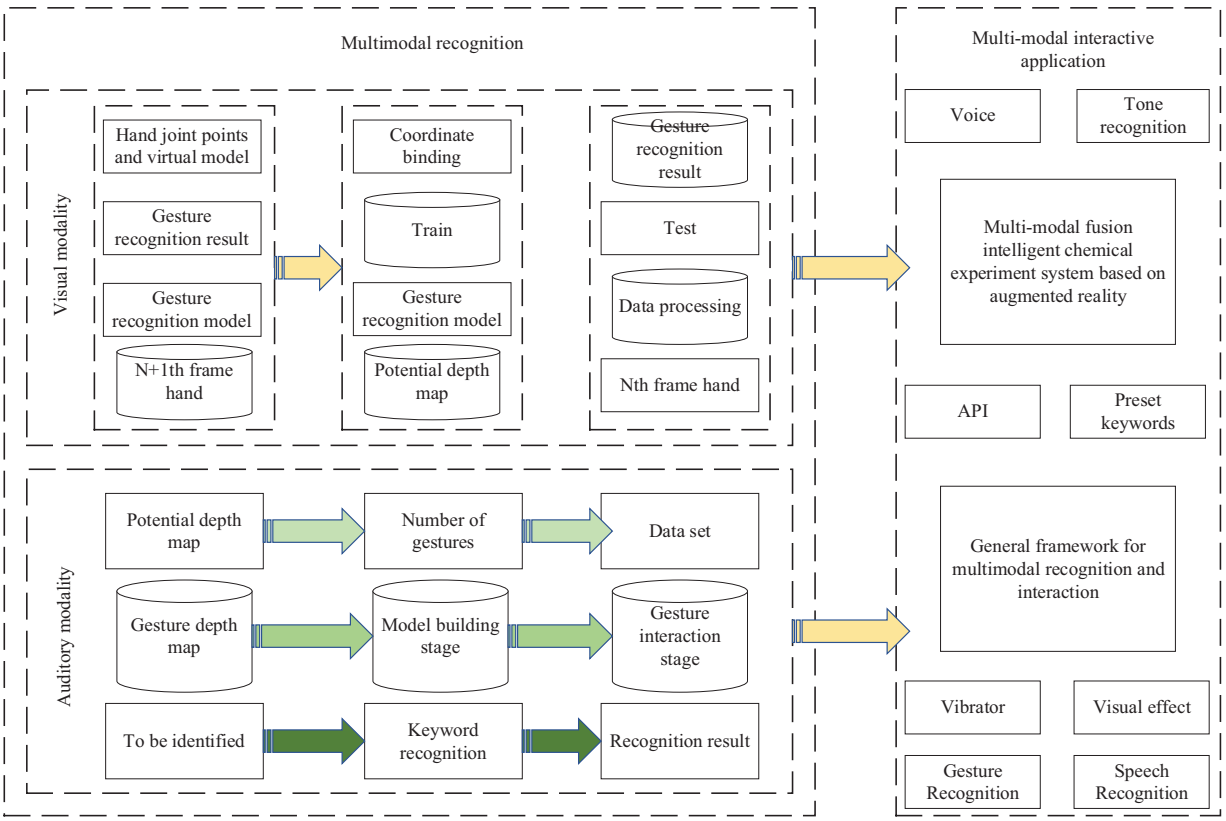


FIGURE 1 | General framework for multimodal recognition and interaction.

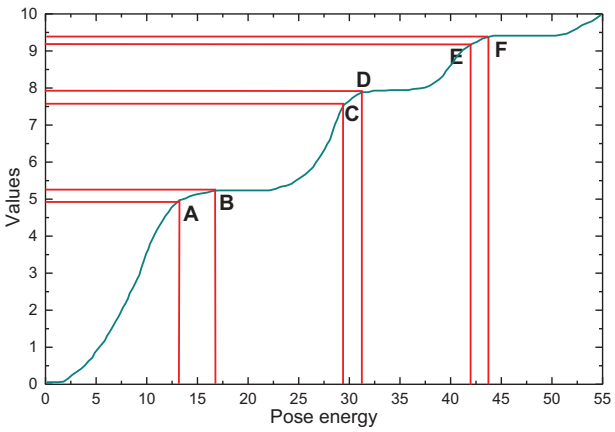


FIGURE 2 | Video segmentation method for artistic pose energy.

contains the depth information of the video. Therefore, its overall recognition is better than RGB data when recognizing complex actions with less interaction with objects, such as hugging others and shaking hands due to the reason that depth data itself has one more dimensional information than RGB data and possesses the property of robustness to video light and dark, angle transformation, and the absence of complex scene information for that action category. Therefore, in the actual recognition process, this method is likely to reduce the accuracy due to the lack of information between modalities or the redundancy of information. The RGB data, on the other hand, highlights the appearance features of the video data, including features such as color, texture, and shape, as well as information about the human surroundings, as shown in **Figure 2**.

A video data is divided into three segments by an energy-guided video segmentation method, and each video segment is subsequently randomly sampled to obtain short segments. Where the three video segments are denoted as (y_i, y_2, y_3), and the short film segments are denoted as(x_i, x_2, x_3). The sampled short segments then fed into the multimodal information complementary network separately to obtain the initial action category scores. The scores of the three short segments are finally fused by weighted fusion to obtain the action category consensus. For the modeling of the short clip segments as shown in Equation (5):

$$f(x_i, x_2, x_3) = G\,(F(x_1;\, W)\, G\,(F(x_2;\, W)\, G\,(F(x_3;\, W) \quad (5)$$

$$\mathrm{L(y,\, g)} = \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{y_i}(\mathrm{g_i} + \ln \sum_{\mathrm{j=1}}^{\mathrm{n}} \exp \mathrm{g_i}) \quad (6)$$

But multimodal data not only have information that complements each other but there is also a lot of redundant

TABLE 1 | Overview of species actions.

Action number	Arm	Gesture	Action number	Arm	Gesture
1	Hold high	Clenched	8	Squat down and raise your hand	Pick up the book
2	Hold high	Right swing	9	Squat down and raise your hand	Pick up the CD
3	Hold high	Scissor hand	10	Squat down and raise your hand	Snap
4	Flat pendulum	Left swing	11	Bend over	Pick up the ball
5	Flat pendulum	Open palm	12	Bend over	Pick up the pen
6	Flat pendulum	Like	13	Bend over	Shoe shine
7	Swing to the chest	Pen cap	14	Sit down	Open the book

information between modalities. The feature mapping stitching alone cannot effectively distinguish which features are shared in the feature mapping and which are special features unique to each modality. Therefore, this method is likely to lead to a decrease in accuracy in the actual recognition process due to missing information between modalities or information redundancy (Chen et al., 2020). The eight channels of EMG are used for classification. The subject's movements are composed of body and gestures. Each movement of the body is before the gesture and the movement is behind. All movements are performed only once in a sample, among which bending movements include bending over and getting up. In addition, in multimodal networks, both low-level features and high semantic features between modalities have complementarity and redundancy, and the pre-fusion strategy cannot effectively share both levels of features at the same time, as shown in **Table 1**.

Due to the variability of individual subjects, i.e., the subjects themselves have different personal habits, the total duration of each sample varied, and it is not possible to describe the duration of each sample in detail here, but the duration of each action sample did not exceed 15 s. Each action performed three times during the experiment, and for both skeletal and EMG data, the entire segment of each sample was used for classification, and all eight channels of EMG were used for classification. Subjects' movements were composed of both body and gesture components, with the body movements preceding the gesture movements for each movement. Decrease by 0.55 and 1.48%, respectively. Under the condition of 10-fold cross-validation, the accuracy of MKL with five different initializations for K cluster and the accuracy of MKL with only one initialization for K cluster are both 98.70%. All movements were performed only once in a sample, where bending movements included bending and rising while squatting and sitting movements did not include rising movements.

Experiments in Human Art Pose Interaction

The most important thing about virtual reality is the construction of virtual environment, i.e., virtual world, which not only pays attention to multiple media elements such as text, image, sound, language, animation, etc. but more importantly, emphasizes the synthesis of various media elements to constitute a 3D sensory world with high approximation to the real environment (Qasim et al., 2021). To obtain a better immersion, this chapter implements the construction of a virtual space configuration system in the Unity3D engine based on a splicing screen, Kinect, and a camera. The virtual reality environment contains both hardware and software environments, and the combination of hardware and software environments builds a highly simulated realistic environment that provides a system guarantee for natural interaction. The output of the experimental environment in this paper is a spliced screen solution, which consists of eight screens placed in an arc to ensure immersion when experiencing the virtual environment and portability when navigating the perspective, as shown in **Figure 3**.

In addition to the use of multimodality and MKL, VLAD features are also helpful. VLAD features encode spatial and temporal information for a given time series using a K-clustering algorithm, which allows features extracted from samples of different lengths to have the same number of dimensions. The dimensionality of the VLAD features extracted from the s EMG signal is 96 and the dimensionality of the VLAD features extracted from the skeletal sequence is 612. This section first analyzes the user's intention through the expression of modal information. Since there are many different types of virtual experiments, the same type of experiment must be specified when studying the expression of intention. Therefore, we combine the structure and function of the smart device to focus on the intelligent the equipment determines the virtual chemical experiment under a specific interactive situation. In the condition using 10-fold cross-validation, not using intra-cluster principal component analysis and power-law normalization during VLAD extraction leads to a decrease in accuracy of 0.55 and 1.48%, respectively. In the 10-fold cross-validation condition, the accuracy of MKL with five different initializations for K-clusters was 98.70% compared to MKL with only one initialization for K-clusters.

Although more complementary features can be extracted using a K clustering algorithm with multiple initializations, MKL can use weights to measure the contribution of features, which can improve accuracy even with fewer features.

Given the better descriptive power of multi-grain shape descriptors, we combine 17 simple shape features into a robust and accurate shape descriptor that contains both local and global features, reflecting the strong complementarity between multi-grain features, i.e., global features can distinguish widely varying

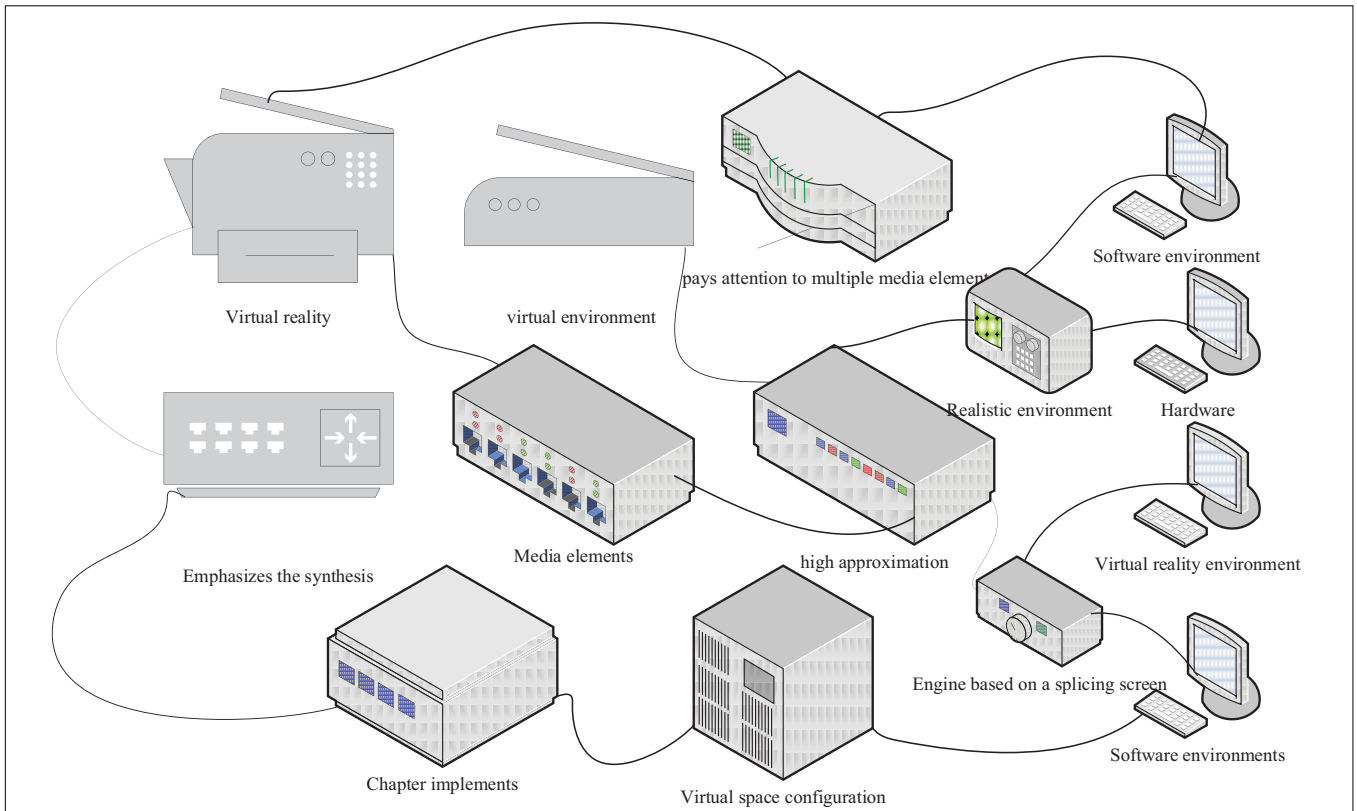


FIGURE 3 | Composition of the experimental environment.

shapes, while detailed differences can be further identified by local features.

$$f_1 = \left|\left|p_i + g\right|\right|_2 \tag{6}$$

$$f_{11} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2, \frac{1}{n}\sum_{i=1}^{n} y_i^2, \right) \tag{7}$$

Wraparound-based feature selection methods can achieve good performance and are easy to implement. We use multiple different wraparound feature selections on the original feature set to filter out several different feature subsets (Ahmad and Khan, 2020). Specifically, to obtain differentiated feature subsets, we subset the feature set using wraparound feature selection methods with different base classifiers. From this perspective, by training on the differentiated feature subsets, differentiated classifier models can be obtained, and then the output of integrating these differentiated models can be used to improve the classification performance in the classifier fusion phase. Then understand the operations performed by the user from this information. In the chemical experiment, the operation requirements and steps are very important. According to the structure, function and gesture types of the smart device designed in this article, combined with the chemical experiment requirements and principles.

Since the wrapped feature selection based on the base classifier does not consider the internal relationship between global and local features, feature selection should be used separately on the global and local feature sets to avoid uninterpretability between global and local features in the filtered feature subset. In addition, for each local feature, feature selection should also be used separately for chain code local features and shape context local features. Specifically, five base classifiers trained by using decision trees, K-nearest neighbors, probabilistic neural networks, fuzzy rules, and random forests can be used to select the five best feature subsets based on the performance of each base classifier. Here the feature selection uses a forward feature search strategy to determine the starting features. After feature selection, the five optimal feature subsets corresponding to the five base classifiers can be obtained.

$$f_{16} = \{C_1, C_2, ..., C_{10}\} \tag{8}$$

$$F_{16} = \{f_1, f_2, ..., f_{10}\} \tag{9}$$

Through technological innovation and design interventions, intelligent machines are made more efficient in recognizing the subtle emotions expressed by users and getting quick understanding. Focus on how the user interacts with the product throughout the design process and apply the principles of good communication to create the desired user emotional experience. Focus on human-centeredness and attention to detail in design

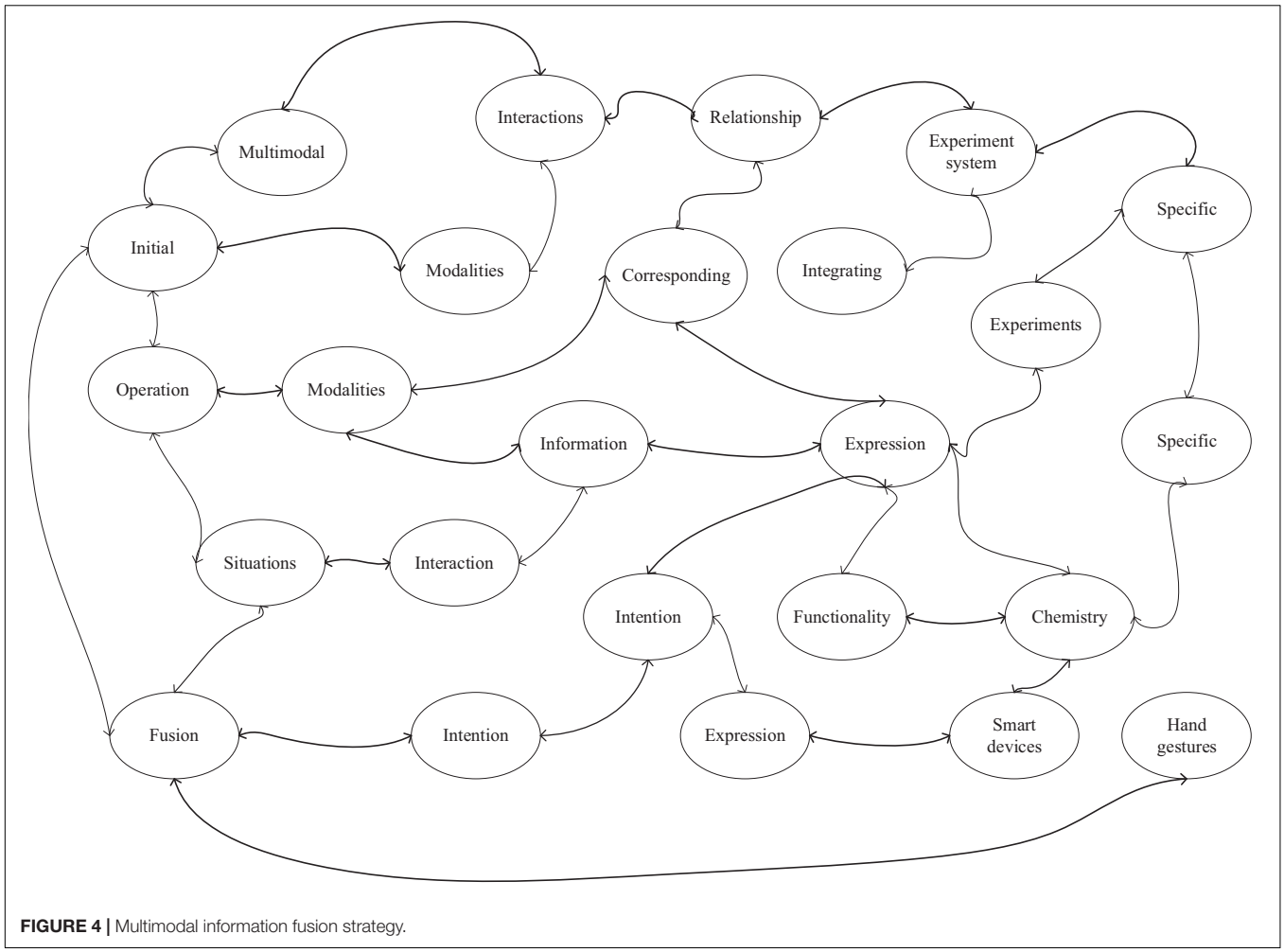


FIGURE 4 | Multimodal information fusion strategy.

to improve the depth and quality of interaction. Combine the context to generate or retrieve the answer, and then synthesize the voice to answer the user through TTS. Design personas and create great experiences through immediate feedback and an emotional engagement approach.

In multimodal interactions, analyzing the relationship between the different modalities corresponding to user actions requires integrating information from all three modalities. Each action behavior of the user is represented as each interaction intention of the user, i.e., the user's intention is expressed in the interaction behavior (Li et al., 2020). This section first analyzes the user's intention through the expression of modal information, and since there are many kinds of virtual experiments that exist and must be specific to the same type of experiment when studying intention expression, we combine the structure and functionality of the smart device to identify virtual chemistry experiments for the smart device in a specific interaction scenario.

The multimodal fusion referred to in this paper is to acquire visual, auditory, and tactile information within a certain period, and then understand the operation performed by the user from this information. The GBT in this experiment did not use any data sampling method, that is, did not use the bootstrap sampling method in the experimental setting, which means that each tree is trained on the same set of samples. In the chemical experiment's, its operational requirements and steps are very important, according to the structure-function and gesture types of the intelligent devices designed in this paper, combined with the chemical experiment requirements and principles, to accurately simulate the experiment to achieve the effect of the real experiment, and at the same time to meet the user's intelligent interaction needs, so for the intelligent chemical experiment system, this paper proposes a multimodal information fusion strategy based on directed graphs, we use inputting information from different channels and judging the interaction semantics under different trigger conditions, as shown in **Figure 4**.

Human-computer voice interaction is landing in increased scenarios, giving rise to a brand-new incremental market. Voice interaction is essentially an interaction mode that uses voice input to understand user needs and provide feedback to them (Saktheeswaran et al., 2020). It first converts the user's conversational speech into text through ASR, then uses the NLP part to understand the text, generates or retrieves the answer in combination with the context, and then answers the user

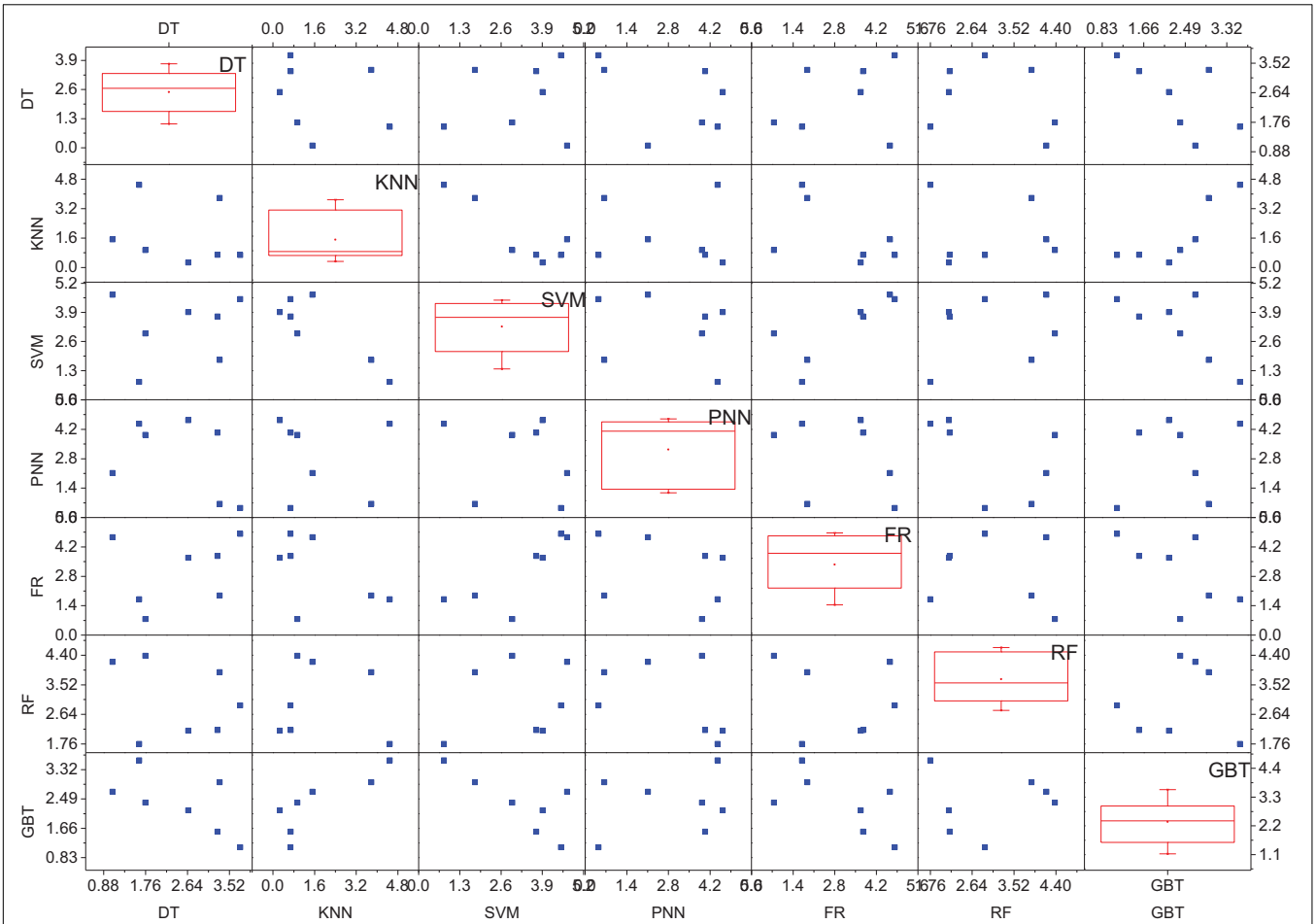


FIGURE 5 | Recognition rate of training classifiers on feature subsets.

by synthesizing the voice through TTS. This also shows the superiority of using integrated learning. Although local fusion can effectively improve the recognition rate, the recognition rate after fusion is still lower than the recognition rate using only the RF classifier. At the core of the speech processing module used by Takenuma Intelligence is a mixture of retrieval and generation models. By using the platform provided by the product, enterprise-level and individual-level users generate many conversations and data, annotate general information with specific information protection and confidentiality, and continuously improve the speech and text libraries, allowing users to get increasingly accurate answers in multiple subsequent rounds of conversations.

ANALYSIS OF RESULTS

Results of Intelligently Enhanced Multimodal Human Art Pose Recognition

Here, we give the setup of each algorithm for each phase. In the feature selection phase, different feature subsets are obtained by wraparound based feature selection methods, which have five different base classifiers, and for the DT classifier, the internal feature selection uses the Gini coefficient and prunes the decision tree to avoid overfitting using the Reduced Error Pruning (REP) method. However, when a feedback mechanism is added, because the mechanism can effectively filter the output results of the target detection network, it adaptively filters out objects and their corresponding tags that are helpful for identifying certain types of actions and corrects the score matrix. The minimum number of records per node is set to 2, and the average split point is selected from the general options. For FR nodes, activation is chosen to use the Min/Max parametrization and the volume border-based shrink function is chosen to reduce the rules to avoid conflicts. For the PNN classifier, theta minus and theta plus are set to default values of 0.2 and 0.4, respectively. For the RF classifier, the information gain ratio is selected as the splitting criterion in the tree option and the integration tree size is set to 100.

For the GBT classifier, the tree depth is set to 10, the number of models is set to 20, and the learning rate is set to 0.1. XGboost is used to handle missing values when dealing with problems for instances belonging to none class. For integrated learning algorithms, data sampling and attribute sampling are very critical. In this experiment, the self-service sampling method is used in

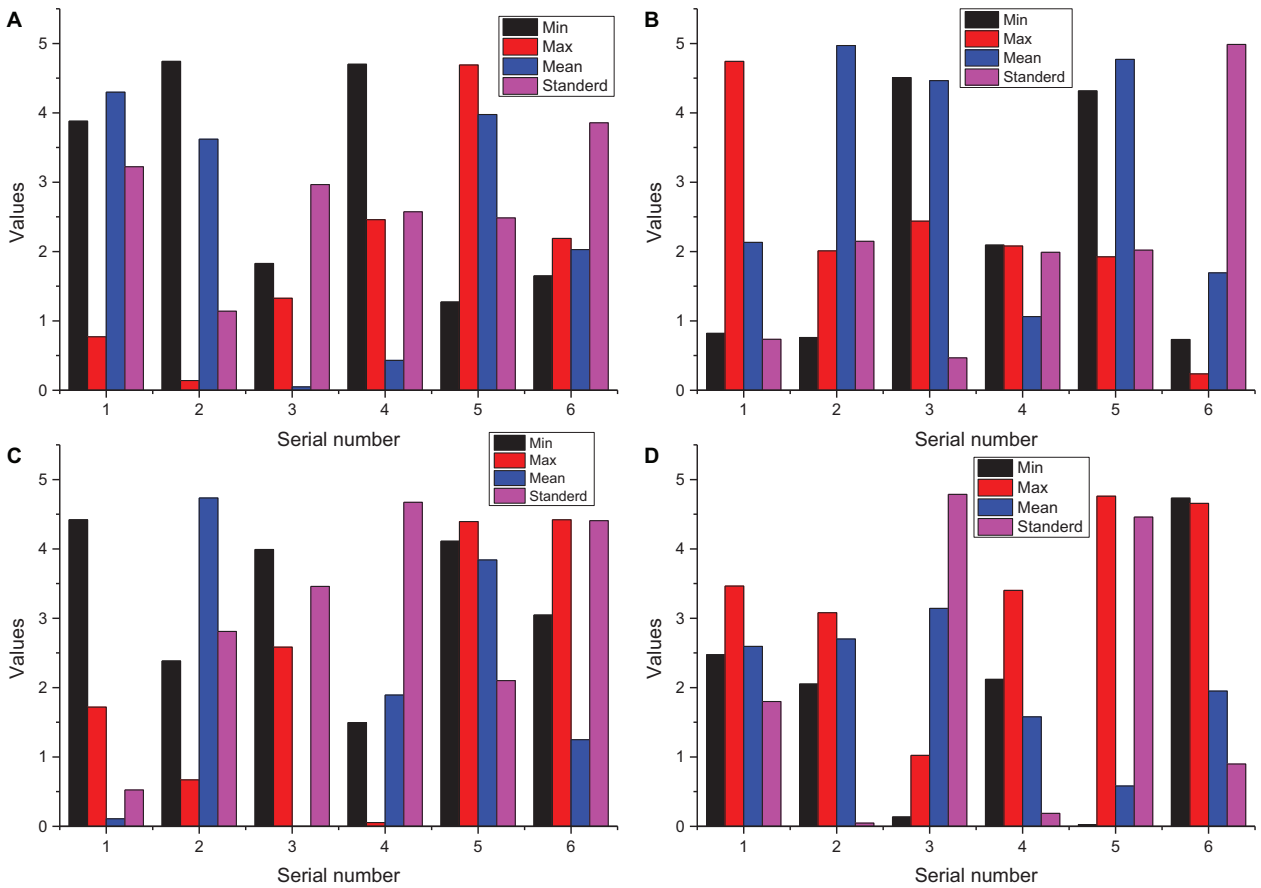


FIGURE 6 | Statistics on measurement questions.

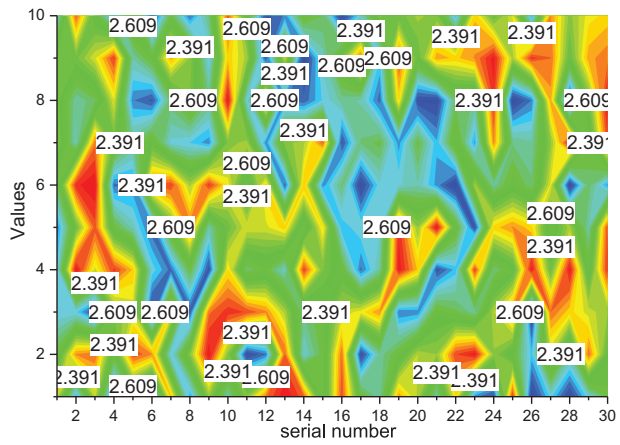


FIGURE 7 | Confusion matrix based on multimodal fusion.

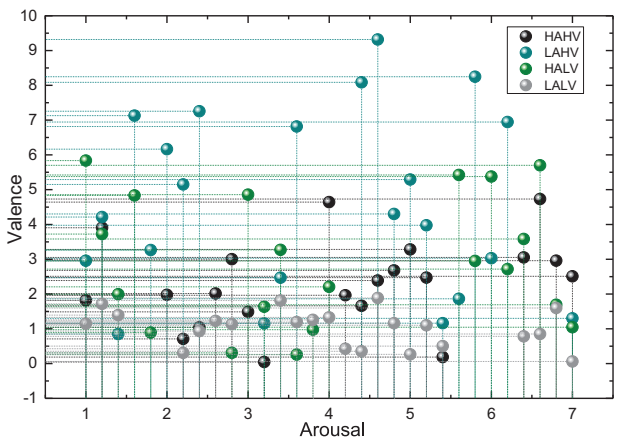


FIGURE 8 | Spatial distribution of subjective emotion ratings of all subjects.

the random forest classifier, i.e., the data sampling mode should be set in the experimental setup to be set to random replacement with the same number of instances before and after sampling. The GBT in this experiment does not use any data sampling method, i.e., the self-sampling method is not used in the experimental setup, which means that each tree is trained on the same set of samples. For attribute sampling and selection, each tree in the RF classifier and the GBT classifier uses a different subset of

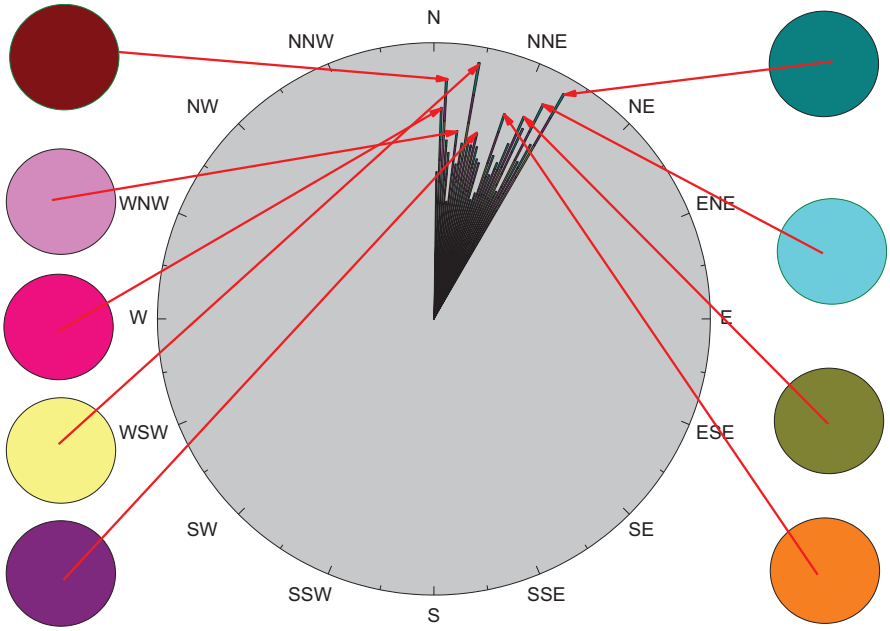


FIGURE 9 | Combinatorial analysis between features.

features, with the size of each feature subset being the square root of the total number of attributes. The algorithm has achieved a recognition rate of 91.40% in the MPEG-7 shape data set. First, the key frame of the time series sequence is extracted from the bone data, and then the relative position feature of the bone data is extracted at the key frame to describe the spatiality of the action. For the fusion of all classifiers, the mean value rule is used and each classifier weights 1 as shown in **Figure 5**.

The recognition rates after local fusion by using a subset of differentiated features to train the classifiers were all improved compared to the recognition rates of locally fused individual classifiers. The improvement in recognition rate after local fusion was much more pronounced when DT was used as the base classifier for local fusion (10.2% improvement). For the seven classifiers selected, RF has the highest recognition rate, and it is worth mentioning that RF itself is an integrated learning algorithm, which also indicates the superiority of using integrated learning. Although local fusion can effectively improve the recognition rate, the recognition rate after fusion is still lower than that of using only RF classifiers. There may be several reasons for the above phenomenon: some instances are correctly classified by RF only, and the sum of confidence of other classifiers for that instance is higher than the confidence of RF for that instance during the fusion, thus leading to misclassification of those instances by classifier fusion. Based on the above analysis, if instances that are not correctly classified by the RF classifier are still not correctly classified after classifier fusion, this will lead to an increase in misclassified instances, i.e., a decrease in the overall classification recognition rate, as shown in **Figure 6**.

As for the action recognition algorithm module, still taking the cross-topic (CS) NTU data as an example, after training the network with 40,320 training samples, a scoring matrix of size 16,560 $\times$ 60 is output for 16,560 test samples, which has 16,560 rows representing the number of video samples, while 60 columns represent the 60 action category scores output, as shown in **Figure 7**. Finally, the score matrix output by the eye detection network and the score matrix output by the action recognition network proposed by the network fusion strategy are fused and the final fused score matrix is obtained. This fusion score matrix is the final action category discrimination result.

As can be seen in **Figure 7**, the addition of a feedback mechanism to the network fusion strategy module can effectively improve the accuracy of the action recognition algorithm. However, only converting the output results of the target detection network into the generation matrix without being adjusted by the feedback mechanism will instead backfire on the discrimination of the action recognition algorithm due to the false detection of the target detection network, causing the overall accuracy of the action recognition algorithm to decrease. However, when the feedback mechanism added, because the machine can effectively filter the output results of the target detection network, the objects and their corresponding labels that help recognize a certain type of action are adaptively filtered out and the score matrix is corrected. And when objects that do not help judge a certain type of action are detected, the generation matrix is automatically corrected, which in turn sums up the sum of the two matrices. Traditional action recognition algorithms have low accuracy in recognizing action classes in complex scenes, action classes with sub-action sharing phenomena, and similar action classes with the influence of other objects. The proposed cross-modal cross-fusion strategy, energy-guided video segmentation, and target detection network-assisted algorithms effectively solve the above problems. As can be seen from the

figure, the recognition rate of similar action categories in many complex scenes by the conventional network is only around 60–70%.

Results of the Human Art Pose Interaction Experiment

Feature selection and emotion classification were performed for the collected EEG signals and peripheral physiological signals. The subjective emotion ratings of all subjects were used as the reference standard, as shown in **Figure 8**.

First, in terms of data validity, the performance clips that participated in the experiment carefully selected in this study to elicit the subjects' emotions as successfully as possible. To prove that the subjects' emotional experiences were consistent with the emotional ratings of the performance clips, the scores of the 46 subjects' subjective evaluations were analyzed and compared with the emotional ratings of the performance clips themselves, and the results showed that the subjects' subjective evaluations were mostly consistent with the emotional ratings of the performance clips, indicating that the emotion-evoking materials were selected more accurately and the subjects could be successfully evoked with the corresponding emotions.

The LOP area feature counts the number of points falling into the cube, which can roughly describe the area of the object, the LOP orientation feature uses PCA to find the principal direction of all the points in the cube, which can roughly describe the orientation of the object, and the HSV color histogram counts the distribution of the color of the object in the HSV color histogram can count the distribution of the color of the object in HSV space, i.e., describe the color information of the object, and the three features complement each other, as shown in **Figure 9**.

The number of trees in the random forest will not perform well if the value is too small. As the number of trees increases, the performance of the random forest stabilizes at about 65% after the number is 60 and reaches the highest recognition rate when the number is 120, so this experiment sets the random forest to generate 120 trees for integration. The confusion matrix of the results of this experiment is shown in **Figure 8**, the recognition results of action 2 and action 5 need to be improved, where the number of correct recognitions of action 1, action 3, action 4, action 6, and action 7 is above 10. Firstly, 66-dimensional features are extracted for the object shape, including global features and local features. Secondly, five different wraparound feature selection methods are used to select five optimal feature subsets from the original feature set, respectively, from the grain computing perspective, and then five homogeneous classifiers are trained on these five optimal feature subsets to merge into a local set, and a total of seven heterogeneous classification algorithms are used to construct seven local sets to merge into a global setting, and the final result output is generated by each classifier voting. The algorithms achieved a recognition rate of 91.40% in the MPEG-7 shape dataset. Keyframe extraction first performed on the temporal sequence by skeletal data, then relative position features are extracted from the skeletal data at keyframes to describe the spatiality of the action, followed by HSV color histogram from RGB data and LOP features from depth data to describe the interacting objects, respectively. Finally, the random forest algorithm used for training and the algorithm was validated on a public dataset.

CONCLUSION

This paper proposes a gesture recognition and interaction method based on augmented reality, firstly, the collected gesture map is pre-processed by segmentation, and the gesture recognition model is trained by the deep learning method. Then, through the ARG algorithm, the mapping relationship between hand joint point coordinates and the virtual scene is determined to achieve the effect of virtual-reality fusion between the natural hand and the virtual model. This algorithm solves the coordinate relationship problem between real space and virtual space based on gesture recognition. Its average correct gesture rate reaches 99.04%, which improves the robustness of gesture recognition and proves the effectiveness of this algorithm and the flexibility of gesture interaction. For the recognition and perception methods of the three-modal information, the multimodal fusion method and the overall framework are proposed at the decision level, including the fusion of gesture, speech, and sensor input information. The information of the three modalities is subjected to data set construction and intention analysis, and multimodal fusion is performed by intersecting and merging functions of multimodal information and independent functions to construct a multimodal information fusion interaction strategy to infer the user's intention and propose a MIDI algorithm. The final validation experiment has a success rate of 92% at normal operation speed, which effectively verifies the feasibility of the method. On the other hand, it demonstrates that multimodality reduces the average operational load by 36% compared to unimodal interaction, i.e., it shows that multimodal interaction reduces the operational load of the user.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

CM conceived of the presented idea. CM wrote and revised the article. QL and YD provided the suggestions. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the National Language Commission's project "Research on the path and mode of language poverty alleviation in typical northwest areas in the post poverty era" (Project No.YB135-139) and the liberal arts project of Northwest Normal University "Research on the path and mode of typical deep poverty language poverty alleviation in the post poverty era". We would like to express our sincere thank for supporting this study.

REFERENCES

Abdullah, S. M. S. A., Ameen, S. Y. A., and Sadeeq, M. A. M. (2021). Multimodal emotion recognition using deep learning. *J. Appl. Sci. Technol. Trends* 2, 52–58. doi: 10.38094/jastt20291

Ahmad, Z., and Khan, N. (2020). CNN-based multistage gated average fusion (MGAF) for human action recognition using depth and inertial sensors. *IEEE Sensors J.* 21, 3623–3634. doi: 10.1109/JSEN.2020.3028561

Chen, L., Wang, K., Wu, M., Pedrycz, W., and Hirota, K. (2020). K-Means clustering-based Kernel canonical correlation analysis for multimodal emotion recognition. *IFAC-PapersOnLine* 53, 10250–10254. doi: 10.1016/j.ifacol.2020.12.2756

Halfon, S., Doyran, M., Türkmen, B., Oktay, E. A., and Salah, A. A. (2021). Multimodal affect analysis of psychodynamic play therapy. *Psychother. Res.* 31, 313–328. doi: 10.1080/10503307.2020.1839141

Islam, M. M., and Iqbal, T. (2021). Multi-Gat: a graphical attention-based hierarchical multimodal representation learning approach for human activity recognition. *IEEE Robot. Automation Lett.* 6, 1729–1736. doi: 10.1109/LRA.2021.3059624

Jain, S., Rustagi, A., Saurav, S., Saini, R., and Singh, S. (2021). Three-dimensional CNN-inspired deep learning architecture for Yoga pose recognition in the real-world environment. *Neural Comput. Applications* 33, 6427–6441. doi: 10.1007/s00521-020-05405-5

Jegham, I., Khalifa, A. B., Alouani, I., and Ali Mahjoub, M. (2020). Soft spatial attention-based multimodal driver action recognition using deep learning. *IEEE Sensors J.* 21, 1918–1925. doi: 10.1109/JSEN.2020.3019258

Kunda, M. (2020). AI, visual imagery, and a case study on the challenges posed by human intelligence tests. *Proc. Natl. Acad. Sci. U.S.A.* 117, 29390–29397. doi: 10.1073/pnas.1912335117

Li, H., Fang, L., and Zhang, T. (2020). Guest editorial introduction to the special section on intelligent visual content analysis and understanding. *IEEE Trans. Circ. Syst. Video Technol.* 30, 4405–4408. doi: 10.1109/TCSVT.2020.3031416

Ludl, D., Gulde, T., and Curio, C. (2020). Enhancing data-driven algorithms for human pose estimation and action recognition through simulation. *IEEE Trans. Intelligent Transportation Syst.* 21, 3990–3999. doi: 10.1109/TITS.2020.2988504

Mohammed, S. N., and Hassan, A. K. A. A. (2020). Survey on emotion recognition for human robot interaction. *J. Comput. Inform. Technol.* 28, 125–146. doi: 10.20532/cit.2020.1004841

Morales, J., Vázquez-Martín, R., Mandow, A., Morilla-Cabello, D., and García-Cerezo, A. (2021). The UMA-SAR Dataset: multimodal data collection from a ground vehicle during outdoor disaster response training exercises. *Int. J. Robot. Res.* 40, 835–847. doi: 10.1177/02783649211004959

Ou, Y., Chen, Z., and Wu, F. (2020). Multimodal local-global attention network for affective video content analysis. *IEEE Trans. Circ. Syst. Video Technol.* 31, 1901–1914. doi: 10.1109/TCSVT.2020.3014889

Pandeya, Y. R., and Lee, J. (2021). Deep learning-based late fusion of multimodal information for emotion classification of music video. *Multimed. Tools Appl.* 80, 2887–2905. doi: 10.1007/s11042-020-08836-3

Poibrenski, A., Klusch, M., and Vozniak, I. (2021). Multimodal multi-pedestrian path prediction for autonomous cars. *ACM SIGAPP Appl. Comput. Rev.* 20, 5–17. doi: 10.1145/3447332.3447333

Qasim, M., Zia, H. B., Athar, A., Habib, T., and Raza, A. A. (2021). Personalized weather information for low-literate farmers using multimodal dialog systems. *Int. J. Speech Technol.* 24, 455–471. doi: 10.1007/s10772-021-09806-2

Rasipuram, S., and Jayagopi, D. B. (2020). Automatic multimodal assessment of soft skills in social interactions: a review. *Multimed. Tools Appl.* 79, 13037–13060. doi: 10.1007/s11042-019-08561-6

Rastgoo, R., Kiani, K., and Escalera, S. (2021). Hand pose aware multimodal isolated sign language recognition. *Multimed. Tools Appl.* 80, 127–163. doi: 10.1007/s11042-020-09700-0

Saktheeswaran, A., Srinivasan, A., and Stasko, J. (2020). Touch? Speech? or Touch and Speech? Investigating multimodal interaction for visual network exploration and analysis. *IEEE Trans. Visualization Comput. Graphics* 26, 2168–2179. doi: 10.1109/TVCG.2020.2970512

Xue, T., Wang, W., Ma, J., and Liu, W. (2020). Progress and prospects of multimodal fusion methods in physical human–robot interaction: a review. *IEEE Sensors J.* 20, 10355–10370. doi: 10.1109/JSEN.2020.2995271

Yoon, Y., Cha, B., and Lee, J. H. (2020). Speech gesture generation from the trimodal context of text, audio, and speaker identity. *ACM Trans. Graphics (TOG)* 39, 1–16. doi: 10.1145/3414685.3417838

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Ma, Liu and Dang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 11 November 2021
doi: 10.3389/fpsyg.2021.788552

Emotion Recognition of Foreign Language Teachers in College English Classroom Teaching

*Lei Li**

School of Foreign Languages, Xuchang University, Xuchang, China

In order to improve students' learning effect, more and more universities favor foreign language teachers who are native speakers of English. Based on the analysis and summary of the research status of emotion recognition, this paper proposes that, in college English classroom teaching, foreign language teachers can reduce the communication barriers with Chinese students through emotion recognition. Based on literature review and actual situation investigation, this study identified four influencing factors on emotion recognition of foreign language teachers, namely, interactive action, facial expression, vocal emotion, and body posture. In our opinion, in the teaching process, teachers can adjust the four factors of emotion recognition to achieve better teaching effect. Further, improve students' learning efficiency. Analytic Hierarchy Process (AHP) is chosen as the research method in this study. After building the analysis model, we collected the questionnaire using the Questionnaire Star, and finally got 12 valid data. After determining the importance of different factors by pairwise comparison, we draw the following conclusions: the influence degree of emotion recognition factors of foreign language teachers is in descending order, interactive action (43%), facial expression (28%), vocal emotion (21%), and body posture (9%). Our research adds to the body of knowledge on emotion recognition among college English teachers. Furthermore, this research assists students in improving their grasp of course content based on the emotions of foreign English lecturers. Based on the findings, we recommend that foreign language teachers in college English classrooms alter their interactive behaviors, facial expressions, and vocal emotions in response to various instructional materials and emphases.

Keywords: emotion recognition, college English, foreign language teachers, analytic hierarchy process, factors

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Qiusheng Shen,
Changshu Affiliated Hospital of Nanjing University of Traditional Chinese Medicine, China
Weiwei Cai,
Northern Arizona University, United States

***Correspondence:**
Lei Li
lilei@xcu.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *02 October 2021*
Accepted: *25 October 2021*
Published: *11 November 2021*

Citation:
Li L (2021) Emotion Recognition of Foreign Language Teachers in College English Classroom Teaching. Front. Psychol. 12:788552. doi: 10.3389/fpsyg.2021.788552

INTRODUCTION

From the perspective of internationalization, the exchanges in economy, culture, and other aspects of the world are deepening. Therefore, there are higher requirements for Chinese people's intercultural and intercultural communication skills. English is the most common language in the international market. In China, the majority of universities set English as a compulsory course for students' future development in the world, so as to improve students' English level and enhance their ability to communicate in English. As an additional language course, English requires students to spend a lot of time learning basic vocabulary and sentences

and applying them in daily life communication. However, it is challenging for students to express more specialized and special usages through word selection, sentence combination, and the use of the culture hidden behind the language (Australian Curriculum, Assessment and Reporting Authority, 2016). Foreign teachers who speak English as their mother tongue have certain advantages in vocabulary use and cultural understanding. Therefore, more and more universities choose to hire them as English course teachers. However, in the process of learning a second language, especially a foreign language, students will inevitably have anxiety in the face of knowledge they do not understand (Chen and Lee, 2011). Moreover, anxiety is an important reason for students' poor learning effect (Woodrow, 2006).

Emotion recognition (Cai et al., 2021) is considered to be an effective method to promote classroom teaching, reduce students' anxiety, and enhance students' learning effect (Chen and Lee, 2011; Xiao et al., 2021). Generally speaking, emotion recognition includes facial expression analysis (Yu, 2020), speech analysis (Ando et al., 2021), physiological signs analysis (Kort et al., 2002), observable behavior analysis (Vicente and Pain, 2002), and so on (Cai and Wei, 2021). It will make the English course more attractive to students by constructing the relationship between language, pronunciation, gesture, and other meaningful ways of expression.

With the vigorous development of information technology (Miao et al., 2021; Wu and Chu, 2021; Zhao et al., 2021), there is a growing tendency for universities to employ foreigners as English teachers. More and more scholars have begun to study the emotion recognition of foreign language teachers in English curriculum teaching. For example, Yu (2020) pointed out that emotion recognition is helpful for English course teaching quality assessment. In terms of English writing, and Unsworth and Mills (2020) pointed out that creating an appropriate body context through the combination of multiple modes such as body language and sight line can effectively help students to expand and use language ability. However, most of the research on Chinese university curriculum focuses on mandarin teaching. Therefore, the emotion recognition of foreign language teachers in English courses has not received much attention. Therefore, this study tries to answer the following questions: (1) what factors can affect the emotion recognition of foreign language teachers? (2) to what extent these factors can affect the emotion recognition of foreign language teachers? (3) how should foreign language teachers adjust the emotion recognition factors to help students to improve their comprehensive English level? We first establish a research model that can be used to analyze the emotion recognition of college English teachers. Then the analytic hierarchy process (AHP) is used to analyze the collected data and determine the weight of each factor used to represent emotion recognition. Our research adds to the body of knowledge on emotion recognition among college English teachers. Furthermore, this research assists students in improving their grasp of course content based on the emotions of foreign English lecturers.

The rest of the paper is arranged as follows: the second part describes the work related to emotion recognition of foreign language teachers; the third part presents the analytical method and experimental results; the fourth part discusses the results; and, in the fifth and last part, we put forward the research conclusions, limitations and possible future research directions.

RELATED WORK

Emotion Recognition in Teaching

In the field of teaching, emotion recognition is often used to test students' learning effect. For example, in the virtual teaching environment, Yang et al. (2018) pointed out that teachers can pay attention to students' state through emotion recognition, and then adjust teaching strategies to achieve better teaching effects. Similarly, Zhang and Zhang (2020) believe that emotion recognition has advantages in improving the effect of O2O English teaching. In real class, teachers can also monitor students' English learning emotions in real time through facial emotion recognition machines to switch teaching content and teaching focus (Cui et al., 2021). Alternatively, long-term collection of facial images can be used for emotion recognition and analysis to find patterns (Oo et al., 2019). In addition, emotion recognition has also been applied in the training of pre-service teachers. Through the virtual situation teaching simulation system, the teaching ability of pre-service teachers is cultivated in order to achieve better learning in the future (Park and Ryu, 2018).

Factors of Emotion Recognition

Voice is an important part of emotion recognition. Neumann and Vu (2018) took English and French as examples to construct a cross-lingual and multilingual voice emotion recognition classification model. Ando et al. (2021) studied the differences in the perception of voice emotions by different listeners and considered as the individuality of voice emotion recognition. In addition, facial expressions (Zhang et al., 2021) play an important role in cross-cultural communication. In addition to voice, people tend to recognize emotions through visual appearance (Miyasato and Nakatsu, 1997). Facial expression is an indispensable part of emotion recognition (Park and Ryu, 2018; Oo et al., 2019). Some scholars also pay attention to the influence of posture and behavior on emotion recognition. Human behavior often carries more emotional information than language (Cui et al., 2021). Brone and Jeroen (2017) point out that it is necessary for teachers to guide students to construct English comprehension schema by using the body mechanism in the teaching process to help students to understand language. In English courses, teachers' use of language, vision, gesture, and other communication methods is conducive to reducing students' learning anxiety and improving their application ability of complex words and sentences (Unsworth and Mills, 2020).

Teaching emotion recognition is an important research field in emotion recognition. According to the above literatures, teachers can adjust teaching methods and strategies by identifying students' emotions. This provides theoretical reference for our research. Accordingly, can teachers adjust their teaching methods

according to students' identification and evaluation of teachers' emotions? However, there are relatively few studies from this perspective. And, this is a question worth thinking about in college English courses. Because when students are faced with difficulties in learning a second language and communication barriers with foreign language teachers, foreign language teachers can help students to obtain better learning results by adjusting the expression forms of emotion recognition.

METHODS AND RESULTS

Identification of Emotion Recognition Factors

Combined with literature research and practical investigation, we identify four factors of emotion recognition in college English classroom teaching. They are facial expression, voice mood, physical gestures, and body movement, as shown in **Table 1**. We believe that in English class teaching, teachers can adjust the content identified by students' emotion by adjusting these four factors, and further, let students obtain better learning effect.

Selection of Methods

In order to explore the influence of different factors on emotion recognition of foreign language teachers, we choose AHP as the main research method. AHP is a decision-making and selection method, which combines qualitative and quantitative research. This method is generally used to solve complex multi-objective decision issues. It takes the form of expert scoring by comparing different factors in pairs. Then calculate the influence weight of each factor according to the collected data.

TABLE 1 | Factors of foreign language teachers' emotion recognition.

Factors	Introduction of factors
Facial expressions	Facial muscles and features, including anger, disgust, fear, happiness, sadness, and surprise (Krause et al., 2021).
Voice mood	Emotions contained in voice, such as anger, happiness, and sadness (Tsiourti et al., 2019).
Physical gestures	Teachers' physical states in the teaching process, such as relaxation and tension, etc. (Matsumoto et al., 2000).
Body movements	It contains the teacher's actions and gestures (Unsworth and Mills, 2020).

At present, the AHP is widely used in the research on the weight analysis of influential factors. For example, Su et al. (2015) used AHP to determine the influencing factors of inter-task dependence. Park and Lee (2020) chose AHP to analyze the influencing factors of airport passenger station efficiency. Akbar et al. (2020) consider this approach to be an effective tool in identifying and prioritizing DevOps success factors. It can be seen that this method has been widely used in different fields in judging the weight of influencing factors. Therefore, AHP is a feasible and effective way to determine the weight of influencing factors of foreign language teachers' emotion recognition.

Data Collection and Conformance Check

Based on the influencing factors of foreign language teachers' emotion recognition identified above and the principle of AHP, we constructed an emotion recognition model of foreign language teachers in college English classroom teaching, as shown in **Figure 1**.

After the model was built, YAANP software was used as a research tool and a questionnaire was designed, as shown in **Table 2**. The purpose of this questionnaire is to determine the relative weight of each influencing factor of emotion recognition. The questionnaire was designed according to the method of AHP. This method compares the importance of influencing factors in pairs at the same level. The measurement scale is divided into five grades, among which the values of 5, 4, 3, 2, and 1 correspond to absolute importance, very important, relatively important, slightly important and equally important, respectively. The hierarchical cells on the left indicate that the left-column factors are more important than the right-column factors. If you select the right rank cell, it means that the right column is more important than the left column.

Students that match the following criteria are chosen to participate in the survey: (a) students who have taken 1-year college English courses and (b) the students' foreign language teacher is a foreigner who speaks English as a first language. Questionnaires were sent by mail to students who met the requirements. The subject of the email is to evaluate the influence of four factors of foreign language teachers' emotion recognition on the improvement of students' learning effect in college English class. Finally, after half a month, 14 questionnaires were collected, in which the names and schools were replaced by numbers in the questionnaires.

We imported the collected questionnaires into YAANP software. Furthermore, the consistency of the questionnaire

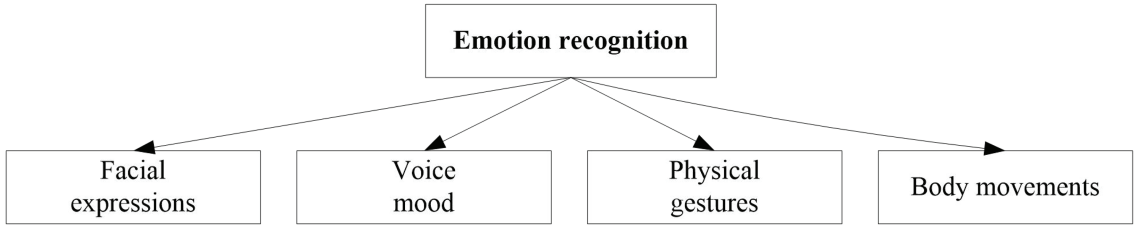


FIGURE 1 | Emotion recognition model.

data is tested. The results showed that two of the 14 questionnaires collected failed the consistency test (CR > 0.1). Thus, we have a total of 12 valid data sets (CR < 0.1). The original data we obtained are shown in **Table 3**.

Data Analysis

In YAANP software, we used arithmetic average to process the 12 valid questionnaires collected. The judgment matrix of the target layer and criterion layer of emotion recognition of foreign language teachers in college English classroom teaching calculated by us is shown in **Table 4**. Among them, the target layer is the influence of foreign language teachers' emotion recognition on improving students' learning efficiency, and the criterion layer is the influence of physical gesture, facial expression, voice mood, and body movement.

According to the judgment matrix data in **Table 3**, we calculate the matrix $\lambda max = 4$, $CR = 0$. This shows that the judgment matrix meets the consistency requirements. Therefore, the aggregate data collected are valid. Furthermore, we use YAANP software to calculate group decision-making and finally determine the weight of the four influencing factors of foreign language teachers' emotion recognition in college English classroom teaching. The results showed that facial expressions were weighted at 28%, vocal emotions at 21%, body postures at 8% and interactive movements at 43%, as shown in **Figure 2**. Therefore, we find that foreign language teachers' body movements and facial expressions have a great influence on students' learning effect, with a total weight of more than 70%.

DISCUSSION

The vast majority of human communication is closely related to emotion recognition. For example, facial expressions, gestures, body posture, and tone of voice can reflect people's emotions to a certain extent, thus affecting communication. More and more scholars are attaching importance to the research of emotion recognition in teaching application.

Through literature analysis, this study identified four specific factors affecting the emotion recognition of foreign language teachers in college English classroom teaching, which are facial expression, voice mood, body movement and physical gestures. Then, we use the AHP and YAANP software's group decision calculation method to determine the weight of the influence of each factor of foreign language teachers' emotion recognition on students' learning effect. Data analysis results show that when foreign language teachers take body movements in college English classroom teaching, they have the greatest influence on students' learning effect, accounting for 43% of the emotion recognition. Facial expressions are much easier to regulate and control than voice mood. In this study, the facial expressions of foreign

TABLE 4 | Judgment matrix table of target layer and criterion layer.

	Facial expression	Voice mood	Physical gesture	Body movement
Facial expression	1	1.377764	3.464610	0.667052
Voice mood	0.725814	1	2.514661	0.484156
Physical gesture	0.288633	0.397668	1	0.192533
Body movement	1.499133	2.065452	5.193911	1

TABLE 2 | Scale for comparison of factors in pairs.

Items	5	4	3	2	1	2	3	4	5	Items
Facial expressions										Voice mood
Facial expressions										Physical gestures
Facial expressions										Body movements
Voice mood										Physical gestures
Voice mood										Body movements
Physical gestures										Body movements

TABLE 3 | Original data of valid samples.

Comparison of importance of relevant indicators	1	2	3	4	5	6	7	8	9	10	11	12
Facial expressions vs. voice mood	3	3	2	3	2	2	3	3	1	3	2	2
Facial expressions vs. physical gestures	4	4	4	3	3	3	4	4	4	4	3	3
Facial expressions vs. body movements	1/3	1/2	1	1/2	1/2	1/3	1/3	1/2	1/3	1/2	1/2	1/2
Voice mood vs. physical gestures	5	5	4	2	3	4	3	3	3	4	3	4
Voice mood vs. body movements	1/3	1	1/2	1/3	1/2	1/2	1/3	1/3	2	1/3	1/3	1/2
Physical gestures vs. body mobements	1/4	1/3	1/4	1/4	1/4	1/3	1/3	1/4	1/3	1/3	1/3	1/4

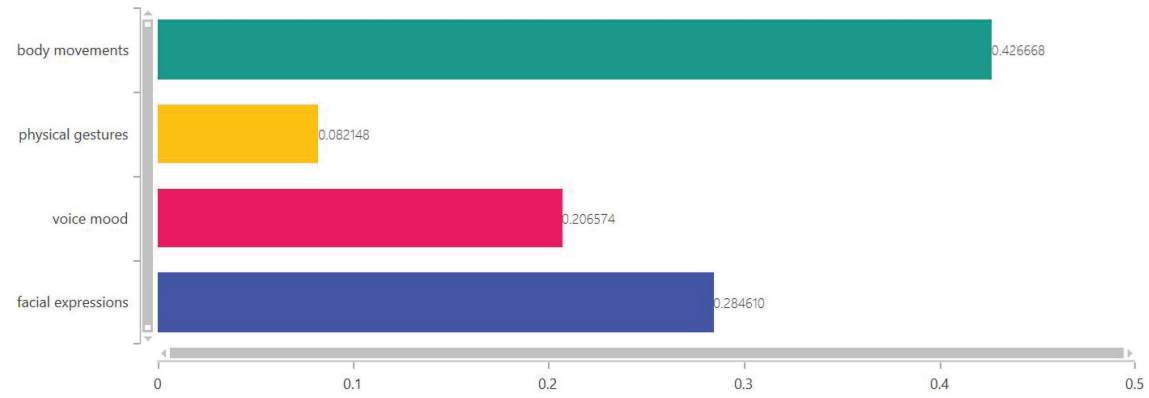


FIGURE 2 | Weight of each factor. The abscissa is the weight of each factor.

language teachers have 28% impact on students' learning effect during the teaching process. The influence degree of voice mood on students' learning effect was 21%. Finally, the teacher's physical gesture had the least impact on students' learning outcomes, at just 8%. The findings add to the body of knowledge on emotion recognition among college English teachers. In addtion, this study can assist students in improving their grasp of course content based on the emotions of foreign English lecturers.

In order to ensure a high level of college English classroom teaching and improve the quality of teaching, foreign language teachers need to reasonably adjust the factors of emotion recognition to provide students with a better English classroom experience. First, from the above data analysis results, it is not difficult to find that body movements play an important role in the emotion recognition of foreign language teachers. Teachers should reasonably use interactive actions to create body context and guide students to understand complex language and the culture contained in language according to the teaching content and focus in class. This can help students to better understand the knowledge in class. For example, when explaining travel-related knowledge points, foreign language teachers can describe the differences between Chinese and foreign countries in taking taxis through body language. Chinese hail taxis by waving their hands. However, in parts of the United States, the thumbs-up is the correct gesture for getting a taxi. This teaching method can help students to enter the teaching situation faster and to better understand the teaching content.

Second, we suggest that foreign language teachers interact with students through expressions such as eyes and smiles, so as to reduce the sense of distance with students. In the classroom, a teacher with a serious face tends to increase the sense of distance with students. Further, it will have a negative impact on the initiative of communication between students and teachers. A teacher who smiles and often makes eye contact with students will be more popular. In this teaching environment, students' interest in learning and love for college English courses will be improved.

Finally, voice mood can help students to establish associations (Park and Ryu, 2018). For the complex sentences that are difficult to understand, it is helpful for students to understand the knowledge points better and faster by driving students into the specific context through the voice mood. In the process of teaching, foreign language teachers change the voice mood according to the teaching content. For example, the teachers will use happy, sad and depressed pronunciation to teach. Furthermore, by guiding the learners' emotional state, they can be attracted to participate in learning better.

CONCLUSION

English is a compulsory course for most Chinese college students in their first and second years. In order to improve the teaching quality, more and more schools choose to hire foreign teachers to teach English courses. However, the differences in teaching methods, teaching environment and classroom teaching habits between China and foreign countries inevitably cause difficulties for students in communicating with teachers and receiving knowledge. In order to reduce the barriers between teaching and learning, this paper studies the factors of emotion recognition of foreign language teachers in college English teaching. Through literature analysis, we believe that the influencing factors of foreign language teachers in emotion recognition include interactive action, facial expression, voice emotion, and body posture. In our opinion, in the teaching process, teachers can adjust the four factors of emotion recognition to achieve better teaching effect. The data analysis results show that the weight proportion of the four factors on students' learning effect is body movements (43%), facial expression (28%), voice mood (21%) and physical gestures (9%). English courses play an important role in college courses. How to make students learn English better and improve their English scores is the focus of many colleges and universities. Therefore, we suggest that foreign language teachers should reasonably adjust their interactive behaviors, facial expressions and vocal emotions in college English classes according to different teaching contents and teaching emphases. This can improve students' sense of immersion in class, so that students are better engaged in learning.

It is worth noting that part of our research has not been fully considered, and there are many aspects that need to be further explored and analyzed. First, in terms of the influence of foreign language teachers' emotion recognition on students' knowledge mastery, our study ignores the differences between English majors' learning in compulsory English courses and students' learning in general English courses. Whether there is a difference between the two and how big the difference is, these are questions worthy of further study. Second, we consider four factors of foreign language teachers' facial expressions, voice mood, body movements, and physical gestures, but the subdivision of each factor is also important. For example, sound includes speed, pitch, amplitude, etc., and it is also worth further studying to what extent these sub-factors influence foreign language teachers in college English classroom teaching. Third, our overall sample size is small, so we can increase the sample survey in the future. Finally, the interaction between the factors and the relationship between them and teacher's emotion recognition is also worth discussing.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LL was responsible for designing the framework of the entire manuscript from topic selection to solution to experimental verification.

REFERENCES

Akbar, M. A., Mahmood, S., Shafiq, M., Alsanad, A., Abdul-Aziz Alsanad, A., and Gumaei, A. (2020). Identification and prioritization of DevOps success factors using fuzzy-AHP approach. *Soft. Comput.* 1–25. doi: 10.1007/s00500-020-05150-w

Ando, A., Mori, T., Kobashikawa, S., and Toda, T. (2021). Speech emotion recognition based on listener-dependent emotion perception models. *APSIPA Trans. Signal Inf. Process.* 10, 1–11. doi: 10.1017/ATSIP.2021.7

Australian Curriculum, Assessment and Reporting Authority (2016). The Australian curriculum: English. 8.2. Available at: http://www.australiancurriculum.edu.au/ (Accessed October 30, 2021).

Brone, G., and Jeroen, V. (2017). *Cognitive Poetics: Goals, Gains, Andgaps*. Shanghai: Shanghai Foreign Language Education Press.

Cai, W., Gao, M., Liu, R., and Mao, J. (2021). MIFAD-net: multi-layer interactive feature fusion network with angular distance loss for face emotion recognition. *Front. Psychol.* 12:762795. doi: 10.3389/fpsyg.2021.762795

Cai, W., and Wei, Z. (2021). PiiGAN: generative adversarial networks for pluralistic image inpainting. *IEEE Access* 8, 48451–48463. doi: 10.1109/ACCESS.2020.2979348

Chen, C. M., and Lee, T. H. (2011). Emotion recognition and communication for reducing second-language speaking anxiety in a web-based one-to-one synchronous learning environment. *Br. J. Educ. Technol.* 42, 417–440. doi: 10.1111/j.1467-8535.2009.01035.x

Cui, Y., Zhao, R., and Wang, S. (2021). Machine learning-based student emotion recognition for business English class. *Int. J. Emerg. Technol. Learn.* 16, 94–104. doi: 10.3991/ijet.v16i12.23313

Kort, B., Reilly, R., and Picard, R. W. (2002). "An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion." in *Proceedings IEEE International Conference on Advanced Learning Technologies*; August 6–8, 2001. 43–46.

Krause, F. C., Linardatos, E., Fresco, D. M., and Moore, M. T. (2021). Facial emotion recognition in major depressive disorder: a meta-analytic review. *J. Affect. Disord.* 293, 320–328. doi: 10.1016/j.jad.2021.06.053

Matsumoto, D., Leroux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., et al. (2000). A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian brief affect recognition test (JACBART). *J. Nonverbal Behav.* 24, 179–209. doi: 10.1023/A:1006668120583

Miao, M., Cai, W., and Li, X. (2021). Parameter estimation of gamma-gamma fading with generalized pointing errors in FSO systems. *Wirel. Commun. Mob. Comput.* 2021:1301878. doi: 10.1155/2021/1301878

Miyasato, T., and Nakatsu, R. (1997). "Facial emotion recognition using multi-modal information, communications and signal processing." in *ICICS Proceedings of 1997 International Conference*; September 12–12, 1997.

Neumann, M., and Vu, N. T. (2018). "Cross-lingual and multilingual speech emotion recognition on English and French." in *ICASSP*; April 15–20, 2018. 5769–5773.

Oo, T. T., Boonroungrut, C., and One, K. (2019). Exploring classroom emotion with cloud-based facial recognizer in the Chinese beginning class: a preliminary study. *Int. J. Instr.* 12, 947–958. doi: 10.29333/iji.2019.12161a

Park, S. U., and Lee, J. D. (2020). Analyzing factors that influence the efficiency of airport passenger terminal by using AHP (analytic hierarchy process) method (case study of Juanda airport – Surabaya, Indonesia). *KSCE J. Civ. Eng.* 24, 3856–3865. doi: 10.1007/s12205-020-0880-x

Park, S., and Ryu, J. (2018). Exploring preservice teachers' emotional experiences in an immersive virtual teaching simulation through facial expression recognition. *Int. J-Hum-Comput. Interact.* 35, 1–13. doi: 10.1080/10447318.2018.1469710

Su, X., Mahadevan, S., Xu, P., and Deng, Y. (2015). Dependence assessment in human reliability analysis using evidence theory and AHP. *Risk Anal.* 35, 1296–1316. doi: 10.1111/risa.12347

Tsiourti, C., Weiss, A., and Wac, K. (2019). Multimodal integration of emotional signals from voice, body, and context: effects of (In) congruence on emotion recognition and attitudes towards robots. *Int. J. Soc. Robot.* 7, 555–573. doi: 10.1007/s12369-019-00524-z

Unsworth, L., and Mills, K. A. (2020). English language teaching of attitude and emotion in digital multimodal composition. *J. Second. Lang. Writ.* 47, 1–17. doi: 10.1016/j.jslw.2020.100712

Vicente, A. D., and Pain, H. (2002). Informing the detection of the students' motivational state: an empirical study. *Lect. Notes Comput. Sci*, 79–86. doi: 10.1007/3-540-47987-2_93

Woodrow, L. (2006). Anxiety and speaking English as a second language. *RELC J.* 37, 308–328. doi: 10.1177/0033688206071315

Wu, Z., and Chu, W. (2021). "Sampling strategy analysis of machine learning models for energy consumption prediction." in *2021 IEEE 9th International Conference on Smart Energy Grid Engineering (SEGE)*; August 11–13, 2021. 77–81.

Xiao, T., Zhang, P., Zhang, Y., Li, D., and Shen, J. (2021). A research on the application of college students' physique data mining based on logistic regression algorithm. *ASP Trans. Comput.* 1, 12–18. doi: 10.52810/TC.2021.100042

Yang, D., Alsadoon, A., Prasad, P., Singh, A. K., and Elchouemi, A. (2018). An emotion recognition model based on facial recognition in virtual

learning environment. *Procedia Comput. Sci.* 125, 2–10. doi: 10.1016/j.procs.2017.12.003

Yu, H. (2020). Online teaching quality evaluation based on emotion recognition and improved AprioriTid algorithm. *J. Intell. Fuzzy Syst.* 40, 1–11. doi: 10.3233/JIFS-189534

Zhang, L., Sun, L., Yu, L., Dong, X., Chen, J., Cai, W., et al. (2021). "ARFace: attention-aware and regularization for face recognition with reinforcement learning." in *IEEE Transactions on Biometrics, Behavior, and Identity Science*; August 11, 2021.

Zhang, M., and Zhang, L. (2020). Cross-cultural O2O English teaching based on AI emotion recognition and neural network algorithm. *J. Intell. Fuzzy Syst.* 40, 1–12. doi: 10.3233/JIFS-189546

Zhao, M., Jha, A., Liu, Q., Millis, B. A., Mahadevan-Jansen, A., Lu, L., et al. (2021). Faster mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking. *Med. Image Anal.* 71:102048. doi: 10.1016/j.media.2021.102048

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 November 2021
doi: 10.3389/fpsyg.2021.767844

A Study on the Construction of Emotion Recognition Based on Multimodal Information Fusion in English Learning Cooperative and Competitive Mode

*Haihua Tu**

College of Foreign Languages, Shaoxing University, Shaoxing, China

With the development of science and education, English learning has become increasingly important. In the past, English learning was mainly based on missionaries, and students were not very motivated to learn. The purpose of this article is to use the English cooperative model to improve the enthusiasm and initiative of students in learning, and to improve the efficiency of students in learning English. A team learning model based on the game is proposed. This article constructs a cooperative and competitive model of English learning based on multimodal information fusion. The main manifestation is that students form groups in small groups, and there is a competitive relationship between the groups. The competition among students in learning is the common interest of the entire group, so that the overall interests of each student will be more competitive. This article refers to the main body association model in the literature to adjust English grammar, vocabulary, and language perception ability: learn together in team communication to improve students' multifaceted abilities. Finally, a questionnaire was designed. The results show that after changing the English team learning mode and optimizing the English team learning support system of the students' English learning team, the English learning cooperation and competition model based on multimode information fusion proposed in this article can improve the learning effect by 55%-60%. In all English teaching, the two dimensions of professional knowledge and English ability training are not mutually orthogonal and mutually exclusive, but mutually supportive and interdependent. To form an effective teaching model of "student-centered and teacher-led," active and rich communication and feedback in the classroom are the keys, and they also help to form a gradual teaching and learning cycle.

Keywords: multi-mode information fusion, English learning, cooperation and competition, team learning model, emotion recognition

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Yi Liang,
Hebei GEO University, China
Xueping Liu,
Shenyang Aerospace University, China
Kangjie Sun,
Shanghai Jiaotong University, China

***Correspondence:**
Haihua Tu
tuhaihua0628@usx.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *31 August 2021*
Accepted: *11 October 2021*
Published: *24 November 2021*

Citation:
Tu H (2021) A Study on the Construction of Emotion Recognition Based on Multimodal Information Fusion in English Learning Cooperative and Competitive Mode. Front. Psychol. 12:767844. doi: 10.3389/fpsyg.2021.767844

INTRODUCTION

Cooperative learning in English teaching requires the cooperation of teachers and students, mutual help, and teamwork, while integrating the competitive model into cooperative learning, to stimulate students' enthusiasm for learning and enhance the learning effect. The traditional English learning mode is mainly based on teacher's preaching.

Teachers become the main body of the classroom, especially for language subjects, students with poor basic ability, it is difficult to integrate into the classroom, and students often feel boring. The game-based English teaching model has emerged and achieved good results. Although the quality of scientific research output is discussed in this article, it is reasonable to produce different results because of the differences in the selection of research objects and indicators. Inspired by the research, we can continue to explore whether the U-shaped relationship between interdisciplinary and team output quality depends on other factors, such as team member recognition or team communication.

Through literature review and empirical analysis, Eknc and others believe that unplanned management and team member conflict are the main causes of research failure (Eknc, 2020). Siddiqui's research points out that the lack of face-to-face communication will lead to members who are geographically far away from the core of the team to feel confused about team goals and tasks (Siddiqui and Georgiadis, 2020). Freedberg believes that game design in the teaching process can trigger users' learning motivation, generate learning achievements, enrich learning experience, and correct learning attitudes (Freedberg et al., 2019). Liao first put forward the word "Gamification," but the use of gamification design thinking to improve the experience of Internet products has gradually become popular since 2010 (Liao et al., 2018). Corell proposed that the game design is a "people-oriented design," which means that in the non-game context, game thinking and interesting elements are used to stimulate user behavior and meet human psychological motivation and needs (Corell et al., 2018).

In terms of educational methods, Qinchen found that proper autonomous learning (such as video and game interaction) can effectively improve students' performance by changing teaching strategies and comparing students' performance changes before and after learning, and can change the way of autonomous learning according to different learning objectives (Cao et al., 2021). Cano found that although there was no significant difference between the game group and the traditional group in learning environment and curriculum attitude, the game group was better than the traditional group in group cohesion, score, and team evaluation (Cano and Villón, 2018). Lobov believes that informal institutional differences will hinder communication and exchange between the two sides of cooperation, which will have a negative impact on transnational cooperation and innovation (Lobov et al., 2020). Li W believes that appropriate informal institutional distance will bring differentiated complementary knowledge and provides impetus for cooperative innovation, and the relationship between informal institutional distance and Transnational Cooperative Innovation is inverted U-shaped (Li and Gu, 2020). Through empirical analysis, Zidan et al. found that the impact of informal institutional distance on Transnational Cooperative Innovation is not significant (Zidan et al., 2019). The above studies all recognize the disadvantages of the traditional non-team learning model, but most of them evaluate the learning situation from the perspective of English learners' behavior and do not form a complete set of English learning models, so it is the lack of research value for the wide applicability of learners.

This article constructs a cooperative and competitive model of English learning based on multimodal information fusion. The main manifestation is that students form groups in small groups, and there is a competitive relationship between the groups, it can stimulate students' interest in learning, give play to their enthusiasm for learning, and enhance the interaction in learning. The competition among students in learning is the common interest of the entire group, so that the overall interests of each student will be more competitive. This article refers to the main body association model in the literature to adjust English grammar, vocabulary, and language perception ability: learn together in team communication to improve students' multifaceted abilities. Finally, a questionnaire was designed.

ENGLISH LEARNING STATUS AND TEAM LEARNING EVALUATION MODEL

Current Situation of Traditional English Classroom

The traditional formative assessment of the classroom is facing many difficulties in the process of implementation. On the one hand, the traditional formative assessment techniques and strategies put forward high requirements for teachers, who need to have evidence thinking, and observe and record students' process performance while teaching. This increases the difficulty of implementing formative assessment (Mohammed et al., 2018). On the other hand, traditional formative assessment techniques such as classroom observation records, activity records, and portfolio use, paper materials, and the process of filling, receiving, and sorting is very time-consuming, and it is difficult to store assessment materials such as learning products (Wariyo, 2019). In recent years, although some classroom teaching systems or learning management systems have created conditions for recording process data in electronic form, they have not fundamentally solved the difficulties in the implementation of formative assessment. Because these evaluation techniques only focus on the data collection link of evaluation, they fail to systematically solve the problem of teachers' implementation of formative evaluation (Lobov et al., 2020). With the development of information technology in the classroom, flipped classroom, problem-solving learning, project-based learning, and other new teaching modes have emerged one after another, which are student-centered and mainly aimed at cultivating students' high-level abilities (such as collaborative inquiry abilities, innovation abilities, etc.) (Hamzah and Nasri, 2020). The learning tasks of these classrooms are oriented to the formal modeling and intelligent calculation of classroom teaching evaluation. With the transformation of classroom teaching from solidifying single teacher instruction to emphasizing group cooperation and participation, how to implement formative assessment in the classroom with cooperative learning as the basic feature has become an urgent problem to be solved in teaching evaluation reform (Liao et al., 2019). With the support of intelligent technology, formal modeling can

deconstruct the complex and changeable classroom teaching process to form a mathematical model; intelligent computing can evaluate students' learning state through algorithms and generate teaching auxiliary information according to teaching principles (Telaumbanua et al., 2020). The combination of the two can promote the effective integration of human intelligence and machine intelligence, and form a classroom evaluation mechanism of human-computer cooperation (Ghodbane and Achachi, 2019). The general framework of formal modeling and intelligent computing for classroom teaching evaluation consists of four parts: perception and storage of teaching and learning behavior, construction of teaching and learning behavior evaluation model, intelligent computing of teaching and learning status, and generation of teaching auxiliary information. The first two parts focus on the representation of educational situations and problems, which is the key step of formal modeling; the latter two parts focus on the realization of the specific technical routes, which is the embodiment of the specific process and function of intelligent computing (Emerson et al., 2018). Based on the perception and storage of teaching and learning behavior, the whole system determines the output of the model by constructing the evaluation model; then, it introduces an intelligent algorithm to calculate the model to achieve the evaluation of teaching and learning status; finally, according to the corresponding teaching principles, it automatically generates the information to assist teachers in classroom teaching evaluation (Mustafa, 2018). The further improvement of the general architecture, human-computer collaborative teaching, and evaluation mechanism requires researchers to work with teachers to "design together." And activities are often complex and difficult, so group cooperation to solve problems and jointly create learning products are common ways (Vellayan et al., 2020). At the same time, the development of such complex teaching activities requires teachers to adjust teaching strategies in time according to the students' learning progress. All of these put forward higher requirements for the intelligence of classroom teaching evaluation. How to implement formative assessment in the classroom characterized by collaborative learning has become an urgent problem to be solved in the reform of classroom teaching evaluation (Nievecela and Ortega, 2019). Some researchers propose to use formal modeling and intelligent computing to help and support teachers more systematically and to form a man-machine collaborative classroom evaluation mechanism (Carrillo et al., 2019). Formal modeling methods, originated from software engineering, aim at modeling representation based on combing the software development process and realizing the standardization of software design, development, and verification. The formal modeling of classroom teaching needs to collect data according to software and hardware equipment, establish a data set according to certain standards, and on this basis, make clear the input and output status of the model according to the actual classroom needs (Nelli and Hartati, 2018). Intelligent computing is the process of introducing machine learning, deep learning, and other intelligent algorithms to calculate the input data after modeling, get the output state represented by mathematics, and use this to analyze learners' potential learning characteristics and rules (Juliati, 2019). In classroom teaching evaluation, formal modeling is to make a symbolic representations of classroom teaching and learning behavior to form a mathematical model; intelligent computing is to analyze and calculate the represented teaching and learning behavior to assist teachers in classroom teaching evaluation (Ghufron and Ermawati, 2018). The linkage of the two can not only realize automation in the process of evidence collection and explanation of classroom teaching evaluation, but also give teachers support in the process of implementing teaching behavior, so it has become an important direction of classroom teaching evaluation reform. The way of cooperation between teachers and students is shown in **Figure 1**.

Typical Classroom Teaching Evaluation System and Architecture

With the support of intelligent technology, the classroom management system develops rapidly. Considering that the classroom management system to provide classroom teaching evaluation must have the basic functions of real-time analysis, evaluation of learning process, and providing feedback information for teachers, this article summarizes the fact intelligent teaching management system through research. The fact system takes students' operation behavior and answer records as input data and provides information and guidance for teachers in cognition, learning status, and cooperation with others through certain calculation methods. On the computing path, these systems follow the ideas of formal modeling and intelligent computing, and not only focus on students' individual learning behaviors and group cooperative learning behavior in the classroom, but also provide support for the transformation of individual, group, and class activities. Input data are as follows: student's operation behavior, answer record student's operation behavior, answer record student's real-time code, group learning product output state, student's personal cognitive state and abnormal state, student's personal cognitive state, problem-solving state, and cooperation state, students' individual cognitive state, students' individual problem-solving progress and learning engagement state, group problem-solving progress, cooperation state, group cognitive state, problem-solving progress, and cooperation state. In the actual classroom, students may switch between individual learning and cooperative learning, so the intelligent system for classroom teaching evaluation needs to pay attention to the two learning processes at the same time. The teacher's teaching framework is shown in **Figure 2**.

Deduction and Evaluation Model of English Competitive Learning

This article proposes an overall prediction model for cooperative prediction in English learning. First, through the cooperative network of English learning participants, the embedded vector representation of English learning participants is obtained by using the network representation learning method. The English competitive learning model is shown in **Figure 3**.

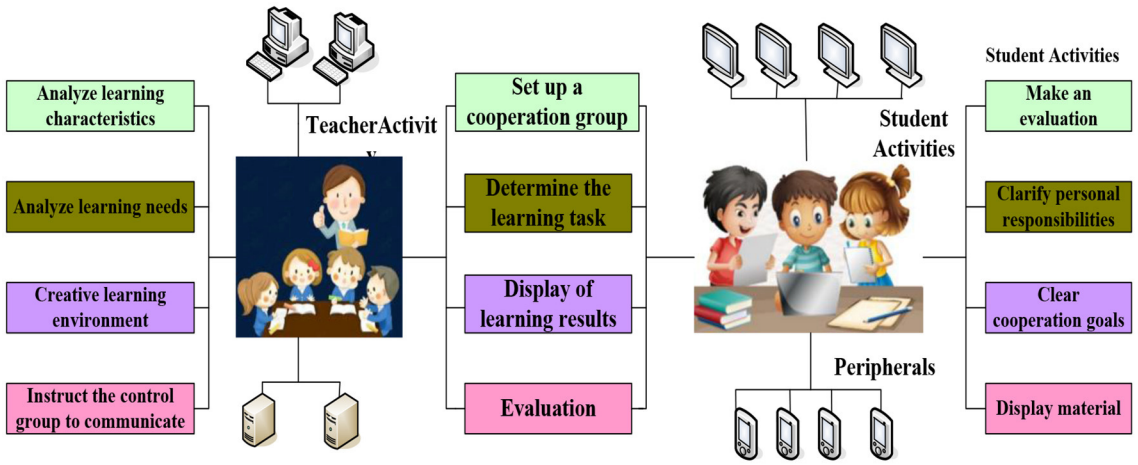


FIGURE 1 | The way of cooperation between teachers and students.

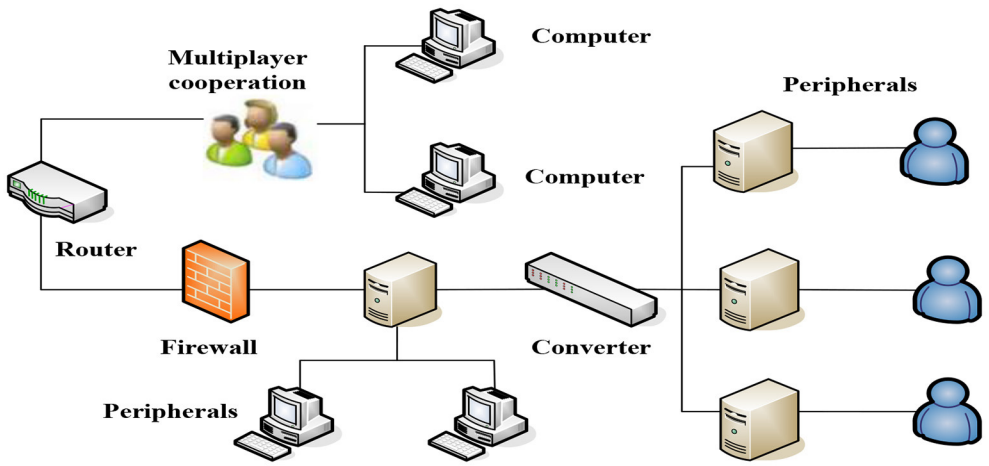


FIGURE 2 | The teacher's teaching framework.

$$A_{\lim} = f(x) = \sum_{j \subset Q} c_j \frac{x_j}{\sigma(X'_j)} - p \tag{1}$$

$$N\left(d_i, w_j\right) = P\left(d_i\right) P\left(w_j \left| d_i\right.\right); P\left(w_j \left| d_i\right.\right) = \sum_{k=1}^{K} P\left(w_j \left| z_k\right.\right) P\left(z_k \left| d_i\right.\right) \tag{2}$$

The structure similarity and content similarity of English learning participants are linearly fused to obtain the node pair similarity discriminant matrix, and the threshold segmentation method is used for cooperative prediction. Next, this article introduces the core module of the cooperative prediction model of English learning participants based on the fusion of representation learning and content features

$$Sd_{gain}(Y) = \frac{\sigma(Y) - avg(\sigma(Y'), \sigma(Y'))}{\sigma(Y)} \tag{3}$$

$$K\left(Y_1, Y_2, \ldots, Y_m\right) = -\sum_{j=1}^{m} M\left(K_j\right) \log_2\left(P\left(K_j\right)\right) \tag{4}$$

$$F\left(X_f\right) = \sum_{x=1}^{N} \frac{N_{1x} + N_{2x} + \ldots + N_{mx}}{T} \tag{5}$$

In addition:

$$I\left(M_{1x}, \ldots, M_{mx}\right) = -\varphi \sum_{j=1}^{M} K_{jx} \log_2\left(K_{jx}\right) \tag{6}$$

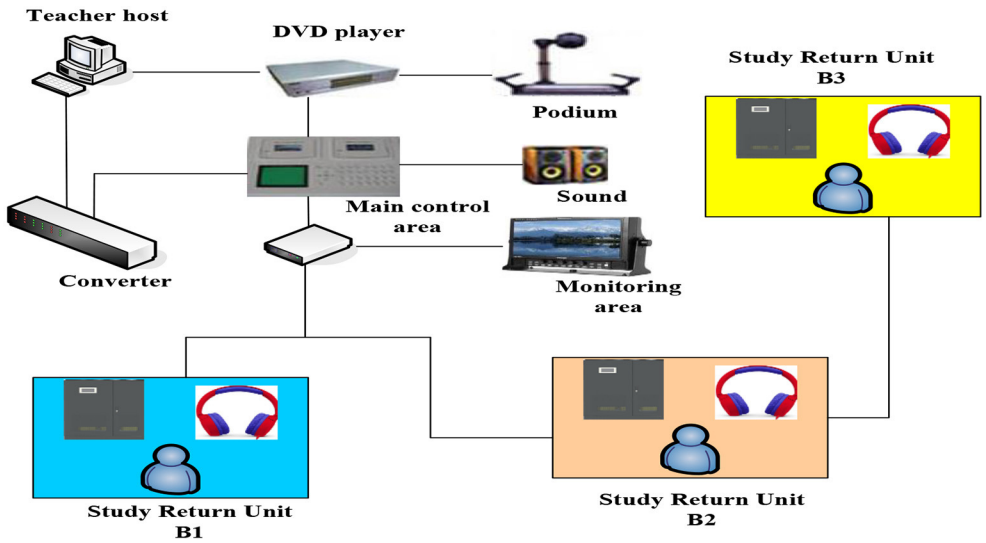


FIGURE 3 | The English competitive learning model.

The topic vector distribution of English learning participants is obtained through the topic model of English learning participants

$$Q = \{(a_1, b_1), (a_2, b_2), ..., (a_n, b_n)\} \tag{7}$$

$$b = \arg\max \sum_{a_n \in W_K(a)} |(b_n = c_n) \tag{8}$$

Web-based English learning participants' similarity calculation, model fusion calculation and parameter selection method calculation are shown in the following formula:

$$G(v,e)=[\zeta_1 c_1(t) + \zeta_2 c_2(k) + \zeta_3 c_3(k) + \zeta_4 c_4(k) + \zeta_5 c_5(k) + \zeta_6 w_{ik}] \tag{9}$$

$$c_1(t) \geq 0, c_2(k) \geq 0, c_3(k) \geq 0, c_4(k) \geq 0, c_5(k) \geq 0 \tag{10}$$

Given a network G (V, e), V is the vertex set, e is the edge set, | V | = n, | e | = m, let the weight value of Min be:

$$\min w_k(t) = [\omega_1(\frac{d_k}{V}) + \omega_2(\frac{d_k}{V}) + \omega_3(\frac{T_k}{ND_K}) + \omega_1(P_K T_K)] \tag{11}$$

$$W = \vartheta^* = \frac{2k}{k+1} + \frac{2c_1 + c_2 + 3et - 2et\zeta}{3} \tag{12}$$

The adjacency matrix of a graph is w. The degree of node I is Di, and the degree matrix of the graph is d. The classical network representation learning method is selected as the baseline method for embedded representation of English learning participants.

$$(In - \alpha W)y = (In - \alpha W)X\beta + \varepsilon \tag{13}$$

$$D_\kappa = \frac{2k}{k+1} + [\frac{1}{2} + \frac{1}{2k}][\frac{c_2 - c_1}{3}]^2 + \frac{2(c_2 - c_1)}{3} \tag{14}$$

$$\psi = \sum_{x=1}^{\theta} Vx = \sum_{x=1}^{\vartheta} (\frac{Wx}{\sum_{1}^{n} W_{\Im}} Sx) \tag{15}$$

The Stochastic Gradient descent (SGD) method is used to update the numerical value convergence

$$\Delta Q_L + \Delta Q_S + \Delta Q_R = \Delta Q \tag{16}$$

$$w_{ik} = \sum_{a}^{n} \tau_1 X_{ik} + \sum_{b}^{n} \tau_2 U(Y_{ik}) + B_{ik} \tag{17}$$

$$w_G^{SGD} = \max\left\{0, W_G \cdot \varepsilon(f_G^{A_i}, f_G^{A_j})\right\} \tag{18}$$

In deep walk model, a random walk is used to generate a vertex sequence. Each node sequence is similar to a sentence in the language, and each node pair is similar to a word. Skip

gram method is used for learning and training, and the vector representation of the node is obtained.

$$I = f(W^e D_1 + \delta^e) \tag{19}$$

$$D_a = g(W^d I + \delta^d) \tag{20}$$

The representation vectors of nodes with strong proximity are closer, and the objective of first-order optimization is

$$f_R{}^{A_i} = w_G{}^{A_i A_j} \cdot V \tag{21}$$

$$P\left(d_i, w_j\right) = P\left(d_i\right) P\left(w_j \left| d_i\right.\right); P\left(w_j \left| d_i\right.\right) = \sum_{k=1}^{K} P\left(w_j \left| z_k\right.\right) P\left(z_k \left| d_i\right.\right) \tag{22}$$

$$\lambda(Wi, Wj) = \left[\log\left(\frac{\left|x_{A_i} - a_{A_j}\right|}{w_{A_j}}\right), \log\left(\frac{\left|y_{A_i} - y_{A_j}\right|}{h_{A_j}}\right), \log\left(\frac{w_{A_i}}{w_{A_j}}\right), \log\left(\frac{h_{A_i}}{h_{A_j}}\right)\right] \tag{23}$$

The objective of second-order optimization is as follows

$$MES(y, y') = \frac{\sum_{i=1}^{m} \left(y_i - y_i^{'}\right)}{m} \tag{24}$$

Because the algorithm is simple, the result of line node prediction is not very stable and the dependence on the initial value is also serious.

CONSTRUCTION OF THE MODE OF COOPERATION AND COMPETITION IN ENGLISH LEARNING

Content

This article constructs a cooperative and competitive model of English learning based on multimodal information fusion. The main performance is that students form teams in groups, and then there is a competitive relationship between groups. The competition of students in learning is the common interest of the whole group, so that each student will be more competitive for the overall interest. This article refers to the subject association model in the literature and makes the following adjustments: English grammar, vocabulary, and language sense abilities learn together in team communication and improve students' abilities in many aspects. Finally, a questionnaire was designed to investigate the participants.

Methods and Steps

The students are divided into groups. The groups usually have four to six people. Keep the group level close. Each group member maintains a logical difference in many aspects, including personality, language, foundation, and hobbies. In particular, avoid the situation where the group members are all cross-graded and introverted students, so that the students in the group have their own characteristics, learn from each other's strengths, and help each other. At the same time, it solves the diversity and fairness of collaborative learning. This work can be done by the monitor and teacher together to promote mutual cooperation and interdependence between the team members. Then the teacher can choose the appropriate group leader for each group. Leaders need to have advantages in academic performance, interpersonal relationships, learning attitudes, organizational skills, etc. Team leaders need to have certain skills to organize English learning activities under the guidance of teachers. For example: Motivate team members to actively participate, supervise everyone to express in English, discuss topics, understand the progress of activities, etc. Create a good learning environment, stimulate students' learning motivation, and work hard in the process of collaborative learning. Team members work together for the same learning goal, communicate with each other, supervise each other, encourage each other, strengthen mutual emotions between members, and cultivate students' autonomy and learning abilities.

The traditional prediction methods based on network representation learning only consider the structural characteristics of the cooperative network of English learning participants, and do not consider the impact of the common research content between two English learning participants on scientific research cooperation. How to measure the similarity of the content between two English learning participants is also worth considering. The traditional content similarity calculation methods use one hot model, Term Frequency-Inverse Document Frequency (TF-IDF) representation, etc. When the number of words is large, it will bring the problem of dimension disaster. Based on the deductive and evaluative model of English competitive learning described in this article, this article proposes that we should pay attention to the influence of five factors on classroom communication: professional knowledge, language knowledge, strategic ability, language environment, and personal characteristics. Based on the above model, this article needs to start from these five aspects to redesign the strategies and methods used in teaching. This article also gives some specific suggestions. In recent years, the curriculum practice shows that this new teaching mode has achieved good results in activating the classroom atmosphere and promoting active learning, and has significantly improved the teaching quality and curriculum satisfaction. The cooperative learning model introduces a competitive mechanism. This type of group collaborative learning method can generate cohesion and centrifugal force between groups. Working together for the honor of the team has strengthened the spirit of unity and cooperation, and the resulting group consciousness can promote the development of usual collaborative learning activities. At the same time, this article believes that this concept and model

can also be used for reference by other non-English majors. At this stage, the teaching modes adopted in this article are: students preview textbooks, fill in the experimental report, watch the teaching video in class, teachers explain the experimental steps after class experimental data processing, and complete the experimental report. The teacher scored according to the experimental operation, preview, experimental report, and other aspects. At the same time, some experimental projects were also opened. Students learned by watching Massive Open Online Courses (MOOC) video and completed the experimental report. Some students like to study independently and think that collaborative learning takes a lot of time, while others feel that they have to take courses with weak foundations. All these require authors and colleagues to continuously explore the educational process to help students to improve their ability to learn English. However, it should also be noted that the shortcomings of the learning cooperative competition model are that it may cause some unnecessary conflicts between students. The form of cooperation may cause negative emotions in individual students and affect the efficiency of English learning.

Questionnaire Design

The sampling survey method adopts the method of random learning, and the questionnaire is distributed randomly to the survey subjects to ensure the randomness of the results. Two different scales were used in this test, one for pretest and one for the posttest, and one for the middle test. Use different perspectives to design the questionnaires for teachers and students, and consider the differences between teachers and students to ensure the wide applicability of the questionnaire. According to the content of the experiment, 15 single choice questions were designed for the pretest and posttest questions, with 1 point for each question, a total of 15 points. The midterm test is composed of 15 single choice questions and 12 subjective questions designed by the experimental content of the required stage. Because our students only need to complete four experiments at this stage, they only need to choose four of them to answer 1 point for each question, and the full score is 4 points. In this way, the measured results of students are more accurate to prevent students from being unfamiliar with other experiments, resulting in the final test data are not accurate enough. At the same time, this article also adds some subjective surveys in the second half of the questionnaire combined with cooperative learning and autonomous learning, and attaches the analytic hierarchy process scale. In particular, the main purpose of setting the midterm test questions in this project is to use the subjective test questions to screen out the students who have independently joined the appropriate cooperative learning or autonomous learning mode in the experimental process, to provide a certain range in the later data processing. The purposes of this test are: first, the type of the test is to provide a new idea for the future college English experimental test; second, to investigate the cognitive level of different levels of students in cooperative learning mode or autonomous learning mode under the framework of the new test.

CONSTRUCTION OF THE MODE OF COOPERATION AND COMPETITION IN ENGLISH LEARNING

Students' learning motivation is shown in **Table 1**. According to **Table 1**, both external motivation and internal motivation are below 3 points, and the average value of internal motivation is lower than the average value of external motivation 0.12. This shows that among the respondents, internal reading motivation still dominates. The average score for unmotivated is above 3 and is higher than the average score of 1.1 for internal motivation, indicating that students still have a certain learning motivation when learning English, and they know why they want to learn English.

The types of learning motivation and academic performance are shown in **Table 2**. The student's academic performance is closely related to the internal and external learning motivation. The internal motivation (2.52) of the students with good academic performance is significantly higher than the external motivation (3.02). There is no obvious difference between the internal motivation (2.78) and the external motivation (2.81) of the students of middle learning. Students with poor academic performance have significantly lower internal motivation (3.06) and external motivation (2.82). In general, the higher the students' internal motivation, the better their performance in learning, and *vice versa*. Internal motivation can provide students with long-term enthusiasm and motivation for learning. The English class itself is a kind of enjoyment for them, and English learning makes them feel happy.

As shown in **Figure 4**, for students with different learning abilities and grades, the same teaching method may have different effects, which is due to the differences between different individuals. With the development of science and technology, research in various fields has shown a trend of rapid differentiation and integration, and research on individual differences has also been minimized and integrated. Individuals are affected by the interaction of genetics and environment during their development, which leads to individual manifestations of various physical and psychological characteristics. Therefore, the influence of cooperative learning

TABLE 1 | Students' learning motivation.

Motivation type	Internal motivation	External motivation	No motivation to learn
Average value	2.76	2.88	3.86

TABLE 2 | The types of learning motivation and academic performance.

Academic performance	External motivation	Internal motivation
Good	3.02	2.52
Medium	2.81	2.78
Difference	2.82	3.06

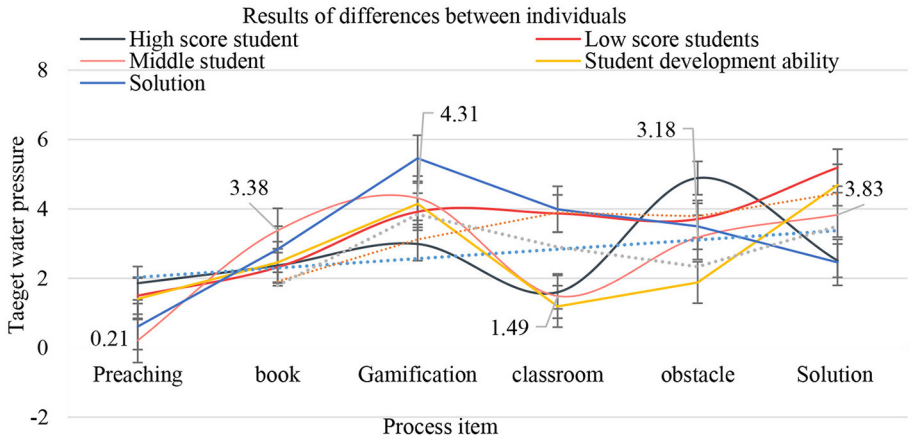


FIGURE 4 | Results of differences between individuals.

TABLE 3 | Small-scale teams are tacit knowledge learning.

Item	High-score student	Low-score students	Middle student	Student development ability	Solution
Preaching	1.86	1.5	0.21	1.41	0.61
Book	2.37	2.32	3.38	2.46	2.84
Gamification	2.99	3.92	4.31	4.15	5.46
Classroom	1.6	3.87	1.49	1.19	3.99
Obstacle	4.89	3.71	3.18	1.88	3.5
Solution	2.51	5.19	3.83	4.69	2.46

mode and autonomous learning mode on students' performance in different grades may be different. To explore the difference of this influence, this survey further refines the scale on the basis of the original scale of students' experimental ability, so that students only need to answer the relevant questions of their experiments at the end of the course, to improve the effectiveness of the questionnaire. At the same time, the students who have chosen the cooperative learning mode or autonomous learning mode in the experiment are selected, and they are divided into different levels according to the test scores, and then, whether cooperative learning or autonomous learning improves their scores significantly is studied.

As shown in **Table 3**, small-scale team is the best way to learn tacit knowledge and the most creative organizational unit. English team learning is easier to improve the overall team performance by stimulating the individual creativity of members. Moreover, in large-scale teams, the mediating path of team interaction and team cohesion has the least effect. The reason is that when students are in the English learning team, they have the least effect. When the number of members is large, the difficulty of interaction and cooperation among members increases, and the demand for team internal personnel management increases. Therefore, the team interaction process is easy to be blocked, and it is difficult to be transformed into team performance. In terms of the heterogeneity of team members' school composition, English team learning has a greater impact on team innovation performance.

As shown in **Figure 5**, whether the team members come from the same type of school or not, the direct effect of English team learning on team innovation performance is greater than the indirect effect. Comparing the contribution of different paths, it is found that the contribution rate of direct path and the mediating path through team interaction and cohesion in the heterogeneous team formed by schools is slightly less than that in the homogeneous team formed by schools; the contribution rate of mediating path through member creativity in the heterogeneous team formed by schools is greater than that in the homogeneous team formed by schools.

The results of the questionnaire are shown in **Table 4**. First, when the members come from different levels of colleges and universities, the interaction process and the formation of cohesion among team members are affected by background differences and lack of tacit understanding. Therefore, English team learning is more likely to affect team performance by stimulating the individual creativity of the members. Second, when the members come from the same level of colleges and universities, due to the lack of university training, it is easier to translate English team learning platform support into team performance or to influence team performance by improving team members' interaction and cohesion.

As shown in **Figure 6**, in terms of the heterogeneity of team members' subject composition, when the number of subjects in the team is large, the overall impact of English team learning on team innovation performance is greater, and multidisciplinary

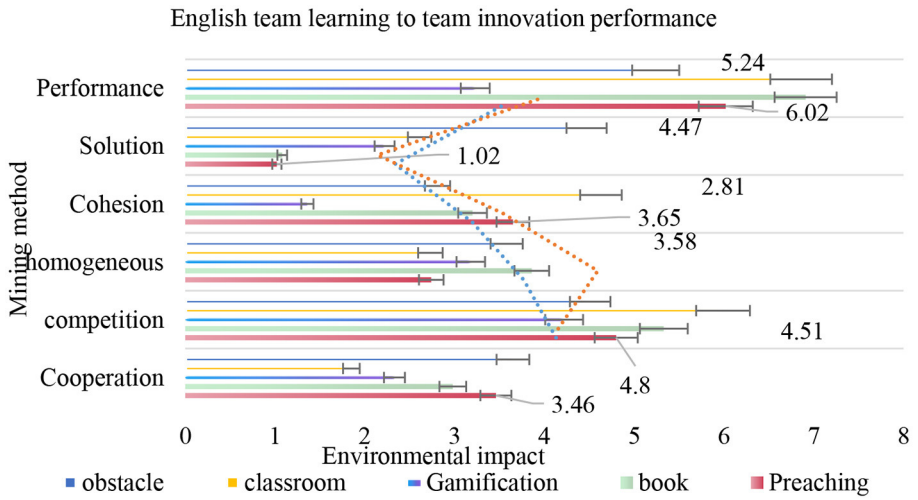


FIGURE 5 | English team learning to team innovation performance.

TABLE 4 | Affect the interaction process and cohesion between team members.

Item	Preaching	Book	Gamification	Classroom	Obstacle
Cooperation	3.46	2.98	2.33	1.85	3.65
Competition	4.8	5.33	4.22	5.99	4.51
Homogeneous	2.74	3.86	3.18	2.73	3.58
Cohesion	3.65	3.2	1.36	4.63	2.81
Solution	1.02	1.08	2.22	2.61	4.47
Performance	6.02	6.91	3.23	6.86	5.24

teams can promote the performance transformation of English learning space. By comparing the team path coefficients of different disciplines, it is found that the contribution rate of promoting performance transformation by improving member creativity is greater in teams with fewer disciplines than in teams with more disciplines; in teams with more disciplines, it is easier to improve team cohesion by promoting team interaction process and to promote team innovation performance transformation. The average value of the teaching method survey is shown in **Table 5**.

As shown in **Figure 7**, when the number of subjects in the English learning team is small, it is easy to produce adverse effects due to the lack of voice in different professional fields; however, due to the relatively concentrated disciplinary background of the members and the team members' in-depth discussion of professional knowledge, the team mainly relies on the individual creativity of the members to promote the team's innovation performance; second, the English learning team is built-in the greater the differences in the professional fields of the members, the more they can examine and solve problems from a diversified perspective, the richer the content of ideological exchange and collision among the members, in the end, a benign interaction will be formed in the creative process, so that the process of transforming the support of the English team learning platform into team innovation performance is smoother. As shown in **Table 6**, English team learning, as the main carrier of innovation and entrepreneurship activities, has an important impact on the innovation performance of students' English learning teams.

The measurement and heterogeneity test results of English team learning and team innovation performance are shown in **Figure 8**. At present, the students' English learning teams feel that the software and hardware support provided by English team learning is insufficient. In the small-scale team and the team with high homogeneity of members' school sources, the common goal and vision is to create a positive creative atmosphere; while in the large-scale team and the team with high heterogeneity of members' school sources, a reasonable division of tasks can make up for the lack of common vision and create a good English team learning atmosphere. In the team with larger scale and higher homogeneity of member schools, the contribution of team innovation achievement is lower; in the team with smaller scale and higher heterogeneity of member schools, the contribution of member innovation intention is lower. The team members' subject heterogeneity has no significant effect on English team learning platform and team innovation performance.

Based on the heterogeneity test results of students' English learning team characteristics, as shown in **Figure 9**, large-scale teams are more likely to directly transform the support provided by the English team learning platform into team performance; in small-scale teams, English team learning is more likely to improve the team's overall performance by stimulating individual creativity; while in large-scale teams, the team interaction process is difficult to transform into team performance. When the source of member universities is homogeneous, it is easier to promote team performance through the positive interaction among team members, while when the source of member universities is heterogeneous, it is mainly through individual creativity to achieve team performance. When the members come from the same discipline, they mainly rely on the individual creativity of the members to promote the team innovation performance; when the team members have different professional fields, it is easy to

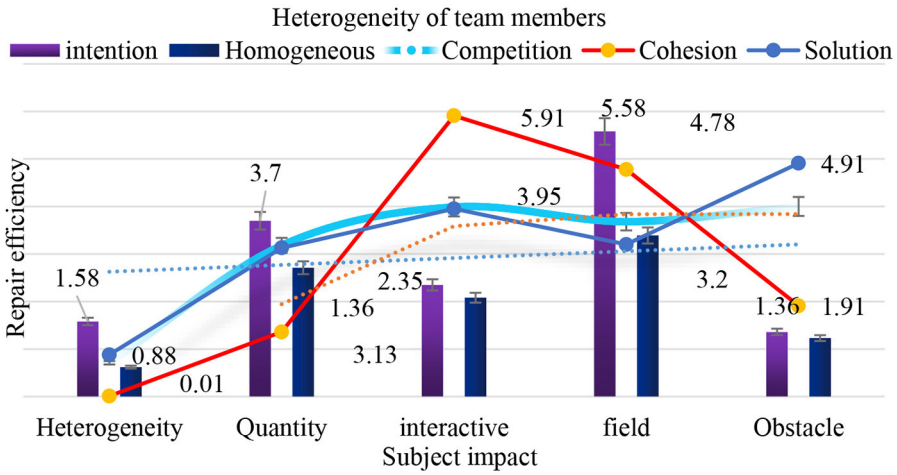


FIGURE 6 | Heterogeneity of team members.

TABLE 5 | The average value of the teaching method survey.

Project	1	2	3	4
Result	3.77	3.01	2.09	2.08

form the positive interaction in the process of creation, and then turn into the team innovation performance.

As shown in **Figure 10**, after changing the English team learning mode and optimizing the support system of English team learning for students' English learning team, the combination mode of English learning cooperation and competition proposed in this article can improve the learning effect by 55–60%. With the vigorous promotion of national policies, English team learning, as an important carrier of innovation and entrepreneurship, has entered a stage of rapid development, but the speed of optimizing the internal quality of English team learning cannot catch up with the growth rate of the total scale. In addition to providing hardware infrastructure to ensure the normal development of daily activities of English learning teams, team learning also needs to explore the establishment of a systematic internal management systems.

As shown in **Table 7**, we should strengthen the influence of innovation atmosphere and English learning culture, and create an innovation atmosphere of "free innovation" by carrying out a series of activities such as lectures, entrepreneurship guidance, and exchange meetings in English team learning. At the same time, English team learning should cooperate closely with colleges and universities and improve team innovation performance through scientific selection and scientific team formation.

As shown in **Table 8**, English team learning should be combined with the specific talent needs of team innovation and entrepreneurship projects, mining and gathering students with different advantages and talents, scientifically forming an English learning team, and scientifically determining the team size and member structure characteristics. For example, for high-quality small-scale innovation and entrepreneurship projects, we should fully explore the unique advantages of individual creativity of members, and at the same time, we should ensure the diversified member structure of the team to prevent adverse effects. After the team has entered a period of stable development, we should take advantage of the resource advantages of English learning space platform to employ well-known innovative and entrepreneurial talents, venture capitalists, and entrepreneurs to form an English team learning tutor library, to improve the team performance by improving the individual creativity of members.

As shown in **Figure 11**, we should give full play to the "aggregation" role of English team learning and enhance the contribution of the creative process to team innovation performance. First of all, English team learning, as a public space for communication and resource sharing in students' English learning, should create opportunities and build a platform for positive interaction and communication and cooperation among team members by means of various theme activities such as exchange salon and entrepreneurship competition, to enhance team cohesion and help the output of innovation and entrepreneurship achievements.

As shown in **Table 9**, with the rapid increase in the number of English learning teams and members, it is increasingly difficult to manage and coordinate within the team. Based on the characteristics of the team, it is necessary to further explore the scientific mode of information and resource sharing among members, the institutional mechanism of team cooperation and win-win results, and the effective path of innovation and entrepreneurship achievement transformation. Finally, according to the needs of English learning teams in different stages of innovation and entrepreneurship, combined with the field of entrepreneurial tutors, through scientific guidance and planning, we can improve the ability of students' English learning teams to reasonably allocate the tasks and the enthusiasm of team members, so as to enhance the team's "joint force."

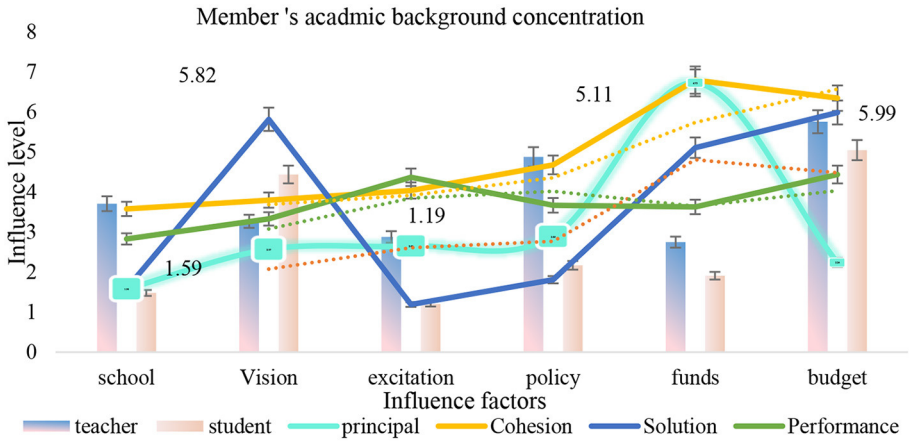


FIGURE 7 | Member's academic background concentration.

TABLE 6 | English team learning as an innovation and entrepreneurship activity.

Item	Intention	Competition	Homogeneous	Cohesion	Solution	Performance
Heterogeneity	1.58	0.71	0.62	0.01	0.88	1.71
Quantity	3.7	3.18	2.71	1.36	3.13	1.63
Interactive	2.35	3.99	2.08	5.91	3.95	3.03
Field	5.58	3.68	3.39	4.78	3.2	5.51
Obstacle	1.36	4	1.23	1.91	4.91	3.88

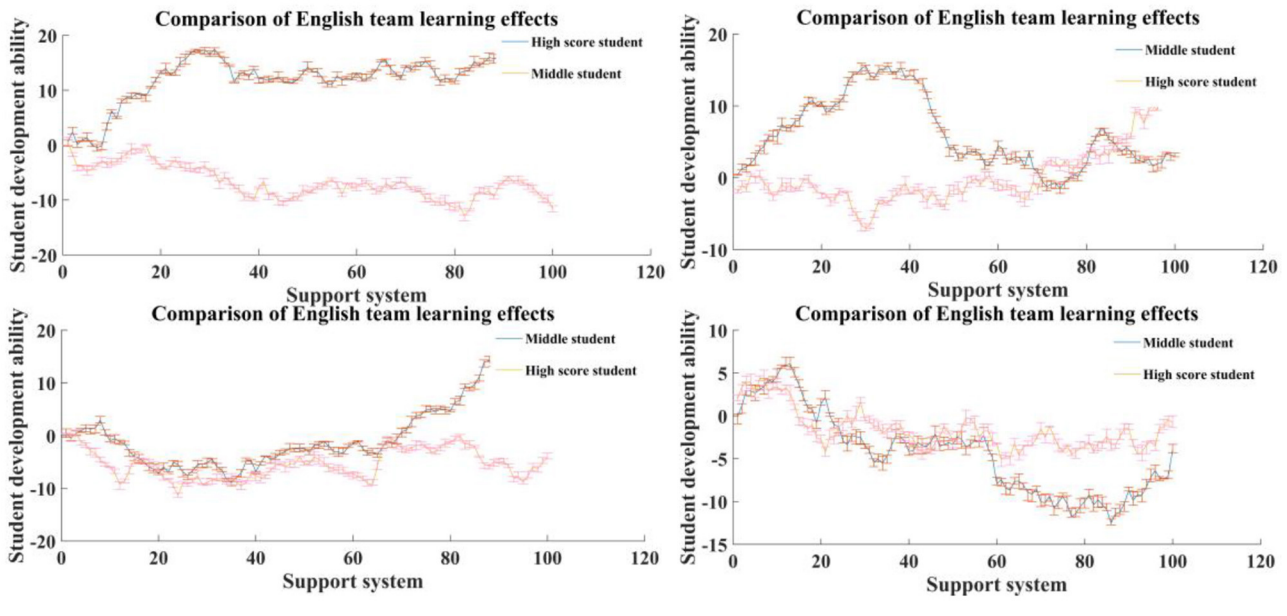


FIGURE 8 | Comparison of English team learning effects.

As shown in **Figure 12**, in view of some common problems in the current English teaching of English subjects, combined with the practical experience in the English teaching of relevant English courses in recent years and some reflections on applied linguistics, this article expounds some new opinions on the classroom communication and feedback link in the English teaching. In all English teaching, the two dimensions of professional knowledge and English ability training are not

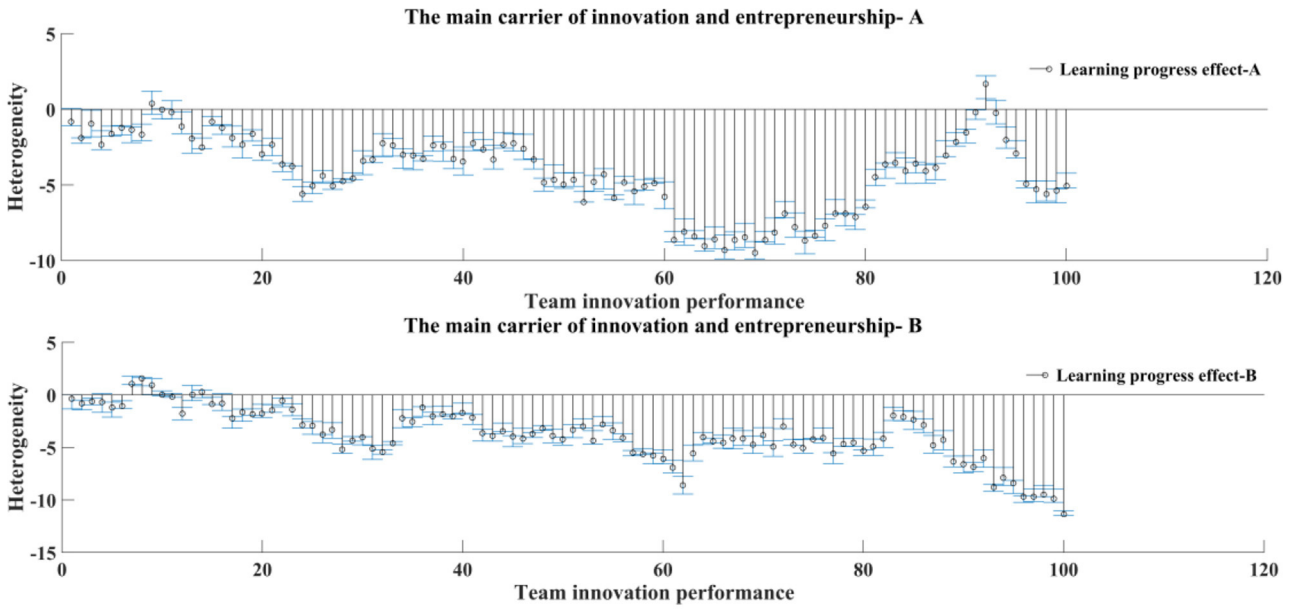


FIGURE 9 | The main carrier of innovation and entrepreneurship.

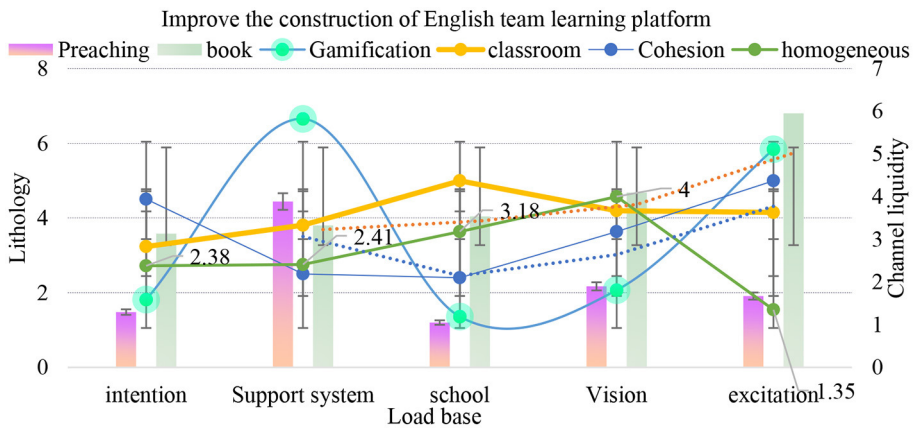


FIGURE 10 | Improve the construction of English team learning platform.

TABLE 7 | Innovative atmosphere and the influence of English learning culture.

Item	Support system	School	Vision	Excitation	Policy	Funds
Heterogeneity	2.71	1.36	3.13	1.63	1.55	1.8
Quantity	2.08	5.91	3.95	3.03	4.93	2.79
interactive	3.39	4.78	3.2	5.51	2.71	2.75
field	1.23	1.91	4.91	3.88	1.85	1.46
Intention	4.13	4.28	6.82	2.07	6.56	2.9

orthogonal and mutually exclusive, but should be mutually supportive and dependent. To form an effective teaching mode of "student-centered and teacher-led," active and rich communication and feedback in the classroom are the keys, which also help to form a gradual cycle of teaching and learning.

At the beginning of the semester, four monitors of the experimental class and the comparison class were called

TABLE 8 | English team learning, team innovation, and entrepreneurship projects.

Item	Heterogeneity	Quantity	Interactive	Field	Intention
Vision	1.36	3.13	1.63	1.55	1.8
Excitation	5.91	3.95	3.03	4.93	2.79
Dimension	4.78	3.2	5.51	2.71	2.75
Policy	1.91	4.91	3.88	1.85	1.46
Funds	4.28	6.82	2.07	6.56	2.9

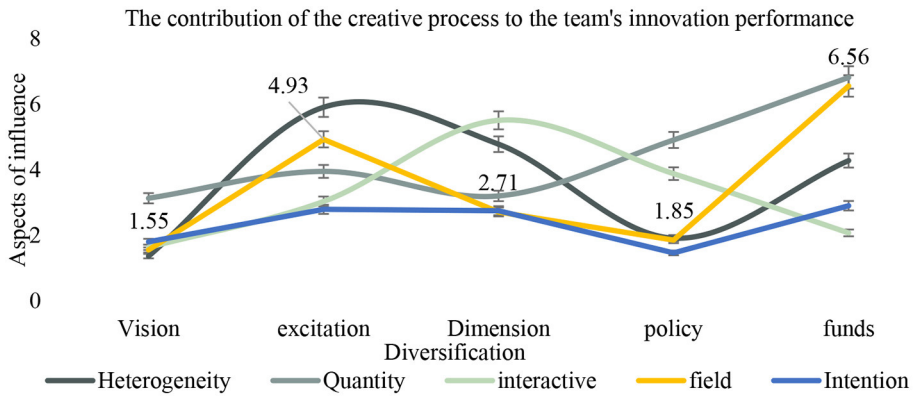


FIGURE 11 | The contribution of the creative process to the team's innovation performance.

TABLE 9 | The number of teams and members in English team learning.

Item	Heterogeneity	Quantity	interactive	field	Obstacle	Partial	Disorders
Intention	2.69	5.53	3.27	2.57	4.44	3.8	5.82
Competition	4.01	5.29	2.88	2.65	1.2	4.04	1.19
Homogeneous	3.74	1.16	4.88	2.89	2.17	4.68	1.81
Cohesion	6.57	6.15	2.75	6.73	1.91	6.8	5.11
Solution	4.66	3	5.76	2.24	5.05	6.35	5.99
Performance	1.46	4.49	4.3	5.68	4.36	3.57	6.03

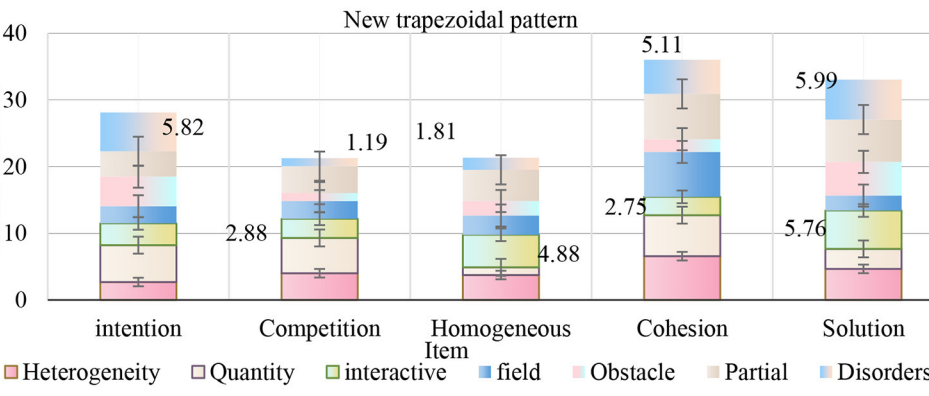


FIGURE 12 | New trapezoidal pattern.

TABLE 10 | Survey results.

Class	Evenly divided into classes	Average score of student motivation	Teachers' self-evaluation
1	75.4	79.5	90
2	75.1	85.0	90
3	52.7	65.5	70
4	47.6	60.9	75

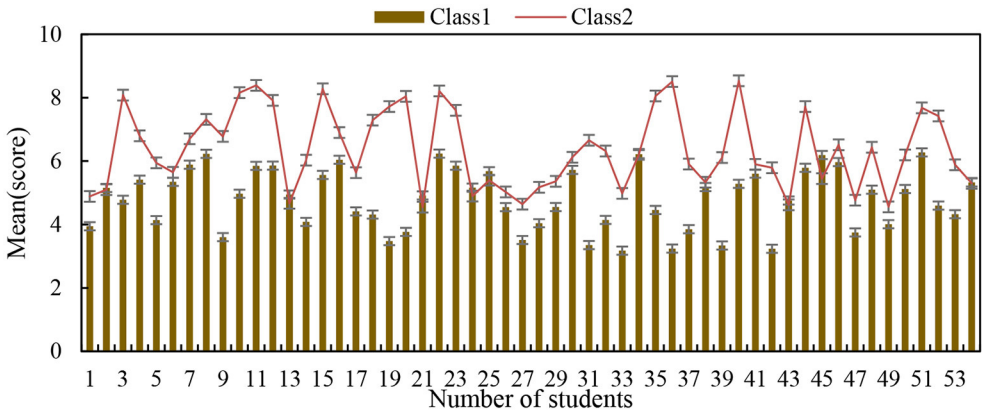


FIGURE 13 | Pretest results comparison between the experimental class and the control class.

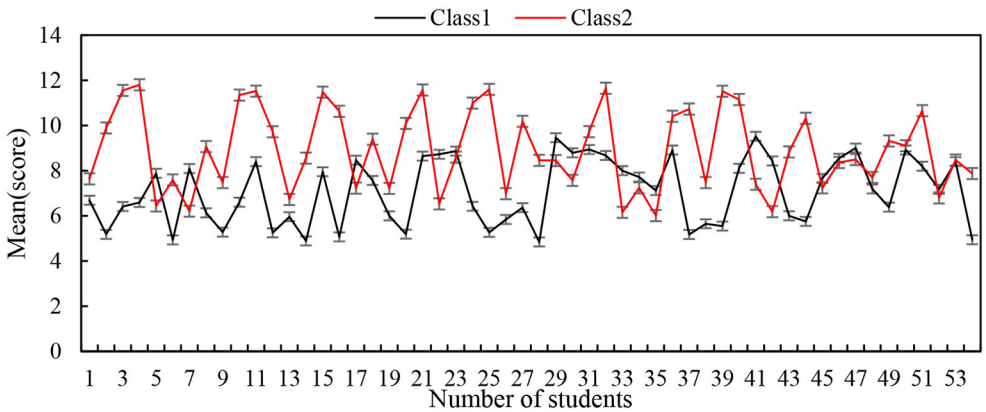


FIGURE 14 | The test results of the two classes after the experiment.

to distribute the questionnaire forms and explain related matters to obtain relevant data. The survey results are shown in **Table 10**. It can be seen from the table that the average score of each class varies greatly, and the average scores of student motivation and teacher self-evaluation are relatively low.

Figure 13 shows the pretest results comparison between the experimental class and the control class. The average score of the experimental class is 0.16 lower than that of the control class (the two classes are 8.59 and 8.75, respectively), and the T value is −0.164, which is for below 5% of the theoretical critical value of 1.69. Therefore, we can conclude that there is no significant difference between the two classes in the pretest results.

The test results of the two classes, after the experiment, are shown in **Figure 14**. The average number of the experimental class is 1.88 higher than that of the control class (the two classes are 14.25 and 12.37, respectively), and the T value is 2.1, which is significantly higher than 5% of the theoretical critical value of 1.69. From this, we can see that the cooperative learning mode in the experimental class played an exact role and achieved remarkable results, thus verifying that the cooperative learning model is significantly better than the traditional teaching method.

CONCLUSIONS

At present, students' English learning teams feel that the software and hardware support provided by English team learning is insufficient. In the small-scale team and the team with high homogeneity of members' school sources, the common goal and vision create a positive creative atmosphere; while in the large-scale team and the team with high heterogeneity of members' school sources, a reasonable division of tasks can make up for the lack of common vision and create a good English team learning atmosphere. In the team with larger scale and higher homogeneity of member schools, the contribution of team innovation achievement is lower; in the team with smaller scale and higher heterogeneity of member schools, the contribution of member innovation intention is lower. Team members' subject heterogeneity has no significant effect on English team learning platform and team innovation performance.

By changing the English team learning model and optimizing the English team learning support system of the students' English learning team, the English learning model based on multimode information fusion combining cooperation and competition proposed in this article can improve the learning effect by 55–60%. With the vigorous advancement of national policies, English team learning has entered a stage of rapid development as an important carrier of innovation and entrepreneurship, but the optimization speed of the internal quality of English team learning has not kept up with the growth rate of the overall scale. In addition to providing hardware infrastructure to ensure the normal development of the daily activities of the English learning team, team learning also needs to explore the establishment of a systematic internal management system.

In the process of learning English, the training process needs to spend more time and money, which greatly increases the difficulty and cost of coordination and management, thus, further reducing the willingness and efficiency of cooperative innovation. In the aspect of trust building, people will react more negatively to opinions and behaviors that do not conform to the existing norms. When the distance between the informal institutions is large, the innovative behaviors and opinions of the two sides often deviate from each other's informal institutional norms. Therefore, it is difficult to establish a high sense of trust and identity in learning English, which seriously hinders mutual knowledge sharing and learning, and is not conducive to the development of cooperative innovation. This article tests that team achievement goal orientation can effectively predict team learning creativity, which not only promotes the development of team achievement goal theory in the field of learning creativity, but also expands the research scope of antecedents of team learning creativity. The next research will focus on the important role of team situation factors in the formation of team learning creativity, and reveal the dynamic mechanism of team learning creativity from the perspective of positive psychology.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HT is responsible for conceptualization, methodology, software data curation, writing, contributed to manuscript revision, read, and approved the submitted version.

REFERENCES

Cano, W. N. S., and Villón, G. M. A. (2018). Cooperative learning as an effective tool in teaching EFL reading comprehension and in multi-level classes. *RECIMUNDO Rev. Científica Investig. Conocimiento* 2, 364–376. doi: 10.26820/recimundo/2.esp.2018.364-376

Cao, Q., Zhang, W., and Zhu, Y. (2021). Deep learning-based classification of the polar emotions of "Moe"-style cartoon pictures. *Tsinghua Sci. Technol.* 26, 27–38. doi: 10.26599/TST.2019.9010035

Carrillo, N. M. R., Labre, M. G. P., and Valle, V. V. Y. (2019). The effects of cooperative learning on reading comprehension. *Explorador Dig.* 3, 143–163. doi: 10.33262/exploradordigital.v3i3.1.875

Corell, A., Regueras, L. M., Verdú, E., Verdú, M. J., and de Castro, J. P. (2018). Effects of competitive learning tools on medical students: a case study. *PLoS ONE* 13, 96–104. doi: 10.1371/journal.pone.0194096

Eknc, M. (2020). Kahoot for good: firing competitive learning for EFL learners. *Int. J. Lang. Educ.* 4, 305–318. doi: 10.29228/ijlet.46623

Emerson, T. L. N., English, L. K., and Mcgoldrick, K. M. (2018). The high costs of large enrollment classes: can cooperative learning help?. *East. Econ. J.* 44, 1–20. doi: 10.1057/s41302-017-0093-2

Freedberg, M., Toader, A. C., Wassermann, E. M., and Voss, J. L. (2019). Competitive and cooperative interactions between medial temporal and striatal learning systems. *Neuropsychologia* 136, 100–101. doi: 10.1016/j.neuropsychologia.2019.107257

Ghodbane, T., and Achachi, H. H. E. (2019). Narrowing the achievement gap between EFL students in oral expression through cooperative learning. *Arab World English J.* 10, 365–378. doi: 10.24093/awej/vol10no2.28

Ghufron, M. A., and Ermawati, S. (2018). The strengths and weaknesses of cooperative learning and problem-based learning in EFL writing class: teachers and students' perspectives. *Int. J. Instruct.* 11, 657–672. doi: 10.12973/iji.2018.11441a

Hamzah, U. S., and Nasri, N. M. (2020). Students' perception in english subject using cooperative learning in primary school. *Int. J. Modern Educ.* 2, 128–140. doi: 10.35631/IJMOE.270010

Juliati, J. (2019). Improving students' writing skills using cooperative learning aT SMPN 2 SUKODONO. *J. English Acad. Spec. Purp.* 2, 36–37. doi: 10.18860/jeasp.v2i2.7853

Li, W., and Gu, Y. (2020). Improvement of stochastic competitive learning for social network. *comput. mater. Continua* 63, 755–768.

Liao, C. C., Chen, Z. H., Cheng, H. N., and Chan, T. W. (2018). Effects of friendship relations and gender differences on classroom-based surrogate competitive learning. *J. Comput. Educ.* 5, 1–19. doi: 10.1007/s40692-018-0108-y

Liao, H. C., Li, Y. C., and Wang, Y. H. (2019). Optimal cooperative learning grouping to improve medical university students' english competencies. *SAGE Open* 9, 215–219. doi: 10.1177/2158244019861454

Lobov, S. A., Chernyshov, A. V., Krilova, N. P., Shamshin, M. O., and Kazantsev, V. B. (2020). Competitive learning in a spiking neural network: towards an intelligent pattern classifier. *Sensors* 20, 545–549. doi: 10.3390/s20020500

Mohammed, H. J., Kasim, M. M., Hamadi, A. K., and Al-Dahneem, E. A. (2018). Evaluating of collaborative and competitive learning using MCDM technique. *J. Comput. Theor. Nanosci.* 24, 4084–4088. doi: 10.1166/asl.2018.11546

Mustafa, M. (2018). The use of picture-Kwl technique in improving students' reading comprehension: a cooperative learning approach. *ELS J. Interdiscip. Stud. Human*. 1, 530–536. doi: 10.34050/els-jish.v1i4.5765

Nelli, N., and Hartati, E. (2018). Improving students' reading comprehension through cooperative learning strategies using numbered heads together. *J. English Lang. Educ.* 4, 28–29. doi: 10.26486/jele.v4i1.317

Nievecela, L. C., and Ortega, D. P. (2019). Using cooperative learning strategies to develop rural primary students' english oral performance. *English Lang. Teach.* 12, 74–76. doi: 10.5539/elt.v12n11p74

Siddiqui, A., and Georgiadis, D. (2020). Global collaboration through local interaction in competitive learning. *Neural Netw.* 123, 393–400. doi: 10.1016/j.neunet.2019.12.018

Telaumbanua, Y., Yalmiadi, Y., Martini, M., and Hasbi, H. (2020). E-political caricatures-assisted cooperative learning: disengaging EFL learners from individualistic, competitive and racial discrimination's learning settings. *Stud. English Lang. Educ.* 7, 504–526. doi: 10.24815/siele.v7i2.16492

Vellayan, G., Singh, C. K. S., Tek, O. E., Yunus, M. M., Singh, T. S. M., and Mulyadi, D. (2020). A review of studies on cooperative learning strategy to improve ESL students' speaking skills. *Univ. J. Educ. Res.* 8, 7555–7560. doi: 10.13189/ujer.2020.082540

Wariyo, L. G. (2019). A Instructional goal structure, gender and second language motivation affecting English language achievement. *Glob. J. Foreign Lang. Teach.* 9, 69–81. doi: 10.18844/gjflt.v9i1.3137

Zidan, M., Abdel-Aty, A. H., El-shafei, M., Feraig, M., Al-Sbou, Y., Eleuch, H., et al. (2019). Quantum classification algorithm based on competitive learning neural network and entanglement measure. *Appl. Sci.* 9:1277. doi: 10.3390/app9071277

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Tu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 25 November 2021
doi: 10.3389/fpsyg.2021.759268

Evaluation of the Virtual Economic Effect of Tourism Product Emotional Marketing Based on Virtual Reality

Manati Talafubieke, Sheng Mai and Nuermila Xialifuhan*

School of Economics and Management, Shihezi University, Shihezi, China

Rapid development of Internet technology has led to great changes in all fields of life. Tourism industry has also gone to the Internet with the change of the times. Regarding the application of Internet technology in tourism, in fact, virtual tourism products are directly used to promote economic development. Online tourism stores have all kinds of tourism information in the database, realize network roaming, provide convenient information services for both the supply and demand of tourism, and can realize the rapid and stable sustainable development of tourism. This article mainly studies the evaluation of the virtual economic effect of emotional marketing of tourism products based on virtual reality (VR), the meaning algorithm of VR, etc., analyzes the virtual economy of tourism products based on VR, and helps tourism enterprises to grasp the needs of tourism consumers more accurately, so as to design products and services that are more in line with the demands of tourism consumers, the process of ultimately improving the user value. This study mainly uses the VR method, VR platform eye map processing, VR eye map regional detection and feature extraction method, and VR tourism product emotional marketing related experiments, to study the evaluation of the virtual economic effect of tourism product emotional marketing based on VR. The economic performance information of the virtual tourism industry is equivalent to the amount of these indicators, and the impact of the comprehensive evaluation on the economic performance of virtual tourism is quite equal. The results show that from 2017 to 2020, the virtual economic revenue of VR tourism product emotional marketing increased from 50 million to 1 billion. However, people give more importance to the tourism experience brought by virtual tourism products, and the reasonable consumption structure of virtual tourism hinders the improvement of tourism economic benefits. To promote the rationalization of the consumption structure of virtual tourism, the basic industry of tourism consumption is developed and the necessary design and processing is increased, in order to increase the supply of tourism products and to enrich tourism consumption.

Keywords: virtual reality, tourism products, emotional marketing, economic effect evaluation, consumption structure

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Ziyang Han,
Shenyang Jianzhu University, China
Xinxin Zhang,
Fujian Agriculture and Forestry University, China

****Correspondence:***
Sheng Mai
ma_shzu@126.com

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *16 August 2021*
Accepted: *22 October 2021*
Published: *25 November 2021*

Citation:
Talafubieke M, Mai S and Xialifuhan N (2021) Evaluation of the Virtual Economic Effect of Tourism Product Emotional Marketing Based on Virtual Reality. Front. Psychol. 12:759268. doi: 10.3389/fpsyg.2021.759268

INTRODUCTION

With the rapid development of computer technology and Internet technology, the network is indispensable in the national life. The progress of science and technology has greatly changed the daily life of people. Therefore, it is very important to explore the possibility of network development. Virtual reality (VR) is a technology that uses computers and other hardware devices to simulate the three-dimensional world. Ideal VR can mobilize five officials of human beings and immerse people in VR.

To develop the national economy and improve the living standard of people, people will pay more attention to the development of the spiritual world. Tourism is one of the important events in holiday life, and the deputy director of China National Administration Report Bureau and the vice president of China Report Association said that tourism has become the most reasonable and potential field to promote economic growth, which also provides great development potential for new forms of real tourism products. The current tourism products cannot provide real virtual displays and structured virtual services for tourists. With the continuous progress of VR technology and yearning for VR technology of tourists, the current tourism products cannot meet the needs of tourism consumers in terms of interaction, immersion, and authenticity. Thus, VR technology needs to be applied to the field of Internet tourism. We will continue to promote the healthy development of internet tourism.

Virtual reality technology will gradually become one of the important technologies to promote the digitalization of tourism information in the future. The VR technology is applied to digital tourism technology to realize the VR of tourism products. Bagus et al. believe that the concept of wireless augmented VR has recently swept the entire 5G ecosystem, causing unprecedented interest in academia, industry and other fields (Bagus et al., 2020). However, the success of the VR experience in the present situation depends on solving a large number of major challenges across multiple disciplines. At the same time, it emphasizes the importance of VR technology as a destructive use case of 5G, which takes advantage of the latest development in storage/memory, fog/edge computing, computer vision, artificial intelligence, and so on. In particular, the main requirements of wireless interconnection VR are described, and then, the key enabling technology is chosen; the research approaches and potential major challenges are introduced. In addition, we have studied three VR cases, given the numerical results under different storages, calculations, and network configurations, also revealed the limitations of the current network, and put forward more theories and innovative reasons, which laid the groundwork for the development of the mass VR but lacked specific data before the project (Bastug et al., 2017). Riva et al. believed that it is possible to make significant changes in body image, attitude, and behavior of obese and related diseases using VR-based psychological methods. The design used is a clinical intervention study, which is treated five times a week based on VR. This improvement is always related to a reduction in problematic diet and social behavior. In traditional weight loss programs, the possibility of inducing significant changes in body image and related behaviors by using short-term VR-based therapy can improve physical satisfaction, but there is no necessary experimental data (Riva et al., 2016). Lenoir J believed that the potential of virtual tourism in Africa is a way to promote economic growth and sustainability under the influence of the coronavirus disease 2019 (COVID-19) epidemic, archives is engaged in the research of archival methods or literature review, and virtual tourism may benefit millions of people. The document helps to innovate strategies for integrated and regional virtual or online tourist attractions such as game parks, protected areas, and other tourist attractions in Africa. African governments must seriously consider establishing partnerships with the private sector to build virtual technology experiments for virtual interaction and enterprise resilience but lack numerical analysis contents (Lenoir et al., 2017). Elbamby et al. believed that VR is expected to be one of the killer applications of 5G network. However, many technical bottlenecks and challenges need to be overcome to promote their wide adoption. Especially, the demand of VR in high throughput, low latency, and reliable communication requires innovative solutions and basic research across multiple disciplines. The challenge and implementation method of super reliable and low delay VR are needed. In addition, in a case study of an interactive VR game machine, the experiment shows an intelligent network design. With the use of millimeter-wave communication, edge computing, and active cache, the future vision of VR can be realized, and it can be used for reference but some of the theories are not accurate (Elbamby et al., 2018).

The innovation of this study is to study the evaluation of the virtual economic effect of tourism product emotional marketing based on VR, the meaning algorithm of VR, and analyze the virtual economy of tourism products based on VR, so as to meet the daily needs of tourists. In the virtual tourism experience, we should pay more attention to beautifying the public and scenic area ecological environment, increasing the attraction of the city, and realizing the concept of "everywhere scenery" of the whole tourism (Freeman et al., 2017).

VR METHOD

Virtual Reality

Virtual reality technology is to simulate a 3D fictional space for users by computer technology. It can simulate visual, tactile, auditory, and other sensory feelings and quickly observe real things. In the process of interaction, the computer uses its powerful computing power to transfer all kinds of information in the simulated three-dimensional virtual world to users, so as to provide the user with the experience of being in the real world (Freeman et al., 2016). With the rapid development of VR technology, more and more VR products are available in the market, but the natural interaction and dizziness in VR technology have not been solved well (Saposnik et al., 2016; Sharar et al., 2016). Therefore, the importance of eye-tracking technology in VR is self-evident. It can be used as a supplement to the human-computer interaction in VR, increase the immersion of users, and also provide immersion and enhanced experience. It has been widely used in various fields (Falconer et al., 2016).

FIGURE 1 | Virtual reality (VR) (http://alturl.com/boffn).

The application of VR in reality mainly includes aerospace and military, games and entertainment, and medical research. In recent years, technology has developed rapidly in entertainment games. VR technology can provide more real and rich perception for gamers, and let users experience 3D reality effects and the perception effect of the situation in movies and games as shown in **Figure 1**.

Eye Image Processing of VR Platform

Methods of Pupil Location

Accurate pupil location is the key to obtain effective eye movement data. The common methods of pupil detection include circle detection, ellipse fitting, and CAMSHIFT algorithm. In image processing, Hough transform is a common method to detect circle area (Berg and Vance, 2017). From a mathematical point of view, the algebraic equation is used to represent the contour of the measured area, and the Hough transform method is used to define the standard equation of the circle, as shown in the formula:

$$(x-m)^2+(y-n)^2=r^2 \tag{1}$$

In the above formula, the coordinates (m and n) are used to represent the coordinates of the center of the circle, and the radius of the circle is expressed by R. In general, the pupil area can be regarded as a circle, so Hough transform can be used to detect the circle to identify the pupil area. It is a common method to fit the edge of the target region with the least square ellipse. This method is mature and suitable for the situation that the edge of the target region is close to the ellipse (Donath et al., 2016; Munafo et al., 2017). In Cartesian coordinates, the general form of the elliptic equation is as follows:

$$mx^2+nxy+ay^2+bx+cy=0 \tag{2}$$

where x and y are the center coordinates of the ellipse, and m and N are the long and short half axes of the ellipse.

$$x=(nb-2ab)/(4ma-n^2) \tag{3}$$

Non-linear Transformation

The pixel data in the circular lens area of the eye movement image are processed, and the gray value of the data in the area is transformed non-linearly. Some data information may be lost through the non-linear transformation of image gray, but the information that the gray value of the original image does not change significantly can be enhanced (Munafo et al., 2017). We can change the mapping relation of non-linear transformation to get different effects and obtain the required image information. The non-linear gray transformation of the image can be divided into non-linear logarithmic transformation and non-linear exponential transformation to lay a good foundation for logarithmic relation:

$$h(x,y)=\lg[h(x,y)] \tag{4}$$

where (x, y) is the pixel coordinates in the image, $h(x, y)$ is the gray value corresponding to the pixel coordinates, and $h(x, y)$ is the gray value after logarithmic non-linear transformation of the gray value of the original coordinates. It is worth noting that one of the characteristics of the logarithmic function is that it can stretch the lower gray value of the image and compress the high gray area of the image (Vankipuram et al., 2017). Logarithmic change of image gray can expand the low gray area of the image, and the effect is better when the image is dark. Exponential non-linear transformation means that the gray value between the transformed image and the image before transformation satisfies the exponential relation:

$$h(x,y)=b^{h(x,y)} \tag{5}$$

where b is the base of the function, (x, y) is the coordinates of the pixels in the image, $h(x, y)$ is the coordinates of the pixels, the corresponding gray value, and $h(x, y)$ refers to the gray value of the original coordinate point after logarithmic non-linear transformation. It should be pointed out that the exponential transform is mainly used to expand the high gray area, which can expand the gray value of the brighter pixels in the original image, and is generally suitable for the over bright image (Meilin, 2017).

Near IR Light Source

The radiation intensity of the near-IR light source is 350 MW/Sr, it is assumed that the exposure time of human eyes to near-IR light is much longer than 1,000 s, the distance between human eyes and near-IR light source is ∼6 cm, therefore, the radiation intensity value at this time needs to be solved, and the calculation formula is as follows:

$$D=E/l^2 \tag{6}$$

FIGURE 2 | Application of VR technology (http://alturl.com/pxm5z).

where D refers to the illuminance of the radiation, E refers to the intensity of the radiation, and l refers to the distance between the near-IR light and the human eye.

$$h(x,y) = [h(x,y) * (500 - h(x,y))] \tag{7}$$

The ultimate goal of the clustering algorithm is to obtain close and independent classes through continuous iteration. k-means clustering aims to divide into k classes, where $k < n$, so that the original data information can be expressed as follows:

$$k = \{k_1, k_2, \cdots k_n\} \tag{8}$$

In other words, the goal is to find the minimum value of the mathematical model as follows:

$$\arg\min \sum_{i=1}^{n} \sum_{x \in k} \left\| x - k_i \right\|^2 \tag{9}$$

VR Technology

Virtual reality technology is referred to as "VR." VR technology is in the exploration period. At first, there were only some military and academic laboratory concepts. Until the beginning of this century, with the continuous development of computer and sensor technology, it gradually changed from the workplace to the practical stage and was widely used in different fields of human life, creating a new situation (Schmitz and Tsobgou, 2016). It realizes the natural communication between the user and the system, which not only improves the working efficiency of the system but also facilitates the operation of the user, as shown in **Figure 2**.

Virtual reality technology is to put the user into the virtual environment completely and realize the interaction with the electronic system using the natural perception of humans. Explicitly, with the own visual, auditory, tactile, and other senses of human beings, the user can get the same feeling of the visual, auditory, tactile, and other stimuli provided by the system in the system as in the real world (Bagus et al., 2020; Szczsna and Wesoowska, 2020).

Region Detection and Feature Extraction of VR Eye Map

The maximum stable extremal region detection is based on the idea of the watershed, which is described as continuous water injection into an area. With the increase of water, any gap in this area will narrow, but not all areas will rise underwater at the same time but by the way of water overflow. When the two regions are connected, the water surface of the two regions increases at the same time, but in this process, the increase of water surface height will not change significantly in some regions (Lan, 2020). The concept of drainage is used to analyze the basic image of eye movement. In this process, some areas of the image will not change significantly or even remain unchanged when the scaling threshold is increased. Then, these regions are regarded as the maximum stable regions, and their mathematical definitions are as follows:

$$F(i) = |F_i + \Delta - F_i| \,/\, |F_i| \tag{10}$$

where F represents one of the connected regions whose binarization threshold is I, represents the change of gray value, and obviously represents the change rate of connected region f(I) with I. For each pixel in the detected stable region, the calculation formula is as follows:

$$h = \sum I(x,y) \tag{11}$$

After calculating the center of gravity of the whole region, that is the center of gravity of the ellipse (Oladipo and Modupe, 2020), then, we need to solve the major axis and minor axis of the ellipse as follows:

$$m = \sum (x,y)^2 I(x,y) \tag{12}$$

The gray level histogram is a very useful tool in digital image processing. It mainly reflects the content of the image gray level. The gray level information is very important for any image. Gray level histogram statistics refers to the number of gray level pixels in an image or the accuracy of gray level pixels. The structure of the one-dimensional gray level histogram is as follows:

$$F(n) = [m(x_1), m(x_2), \cdots m(x_n)] \tag{13}$$

where $f(x)$ is the number of pixels, $m(x)$ is the total number of pixels, in the case of image gray histogram, and the abscissa is gray. The ordinate is the frequency of gray scale, and it can also be the number of pixels to calculate the gray scale value. Then,

we analyzed the gray scale histogram of the intercepted eye image as follows:

$$m(x_i) = \frac{f(x_i)}{\sum f(x_i)} \tag{14}$$

If the calculated fixation angular rate is greater than the given threshold, the eye motion mode of the frame is set to saccade, and there is no need to calculate the line of sight. Otherwise, it is fixation. To calculate the line of sight point, the variance of the angular velocity of fixation point needs to be calculated, as shown in the formula:

$$\sigma = \frac{1}{n}\sqrt{\sum_{i=0}^{n-1}({f_{n-i}}^2 - f^{-2})} \tag{15}$$

EXPERIMENTS ON EMOTIONAL MARKETING OF TOURISM PRODUCTS BASED ON VR

Function of Emotional Marketing

To meet the emotional needs of consumers is the core of emotional marketing. It is a process to stimulate the inner emotional needs of consumers by stimulating and meeting the emotional needs of consumers and to achieve marketing goals. Emotions of consumers can be divided into two different types, one is the result of daily accumulation and the other is a short-term attitude, whether positive or negative, calm or strong, can be classified into the emotional category. At present, society has huge material wealth, fierce competition, and increasingly personalized consumption. Emotional factors become important and special factors in the marketing process (Fabisiak, 2018). Enterprises have to create a positive emotional experience for consumers, for example, to make other consumers feel happy, happy and excited, so as to avoid their negative emotions. Emotion is the key factor of modern marketing. To obtain the recognition and loyalty of consumers, enterprises must connect consumers with emotion, that is, enterprises must pay attention to the emotional connection between consumers and brands, and emotional marketing emphasizes an enterprising spirit. To meet the emotional needs of consumers and make them feel psychological identity, consumers have a preference for corporate brands. Brand does not need the formation of consumer psychology (Khatwani and Srivastava, 2018).

Tourism Product Module Function of VR

Virtual reality tourism products adopt the VR technology relied on the combination of graphics and images and adopt different development technologies for different modules, which not only saves the development cost to the greatest extent but also enhances the performance of VR tourism products. In this system, visitors can not only watch and listen when walking but also show the real environment and facilities of the hotel in front of tourists. In addition, tourists can fully integrate into the VR world through games and immerse themselves in the world of VR tourism products. It provides the possibility to develop different types of VR tourism products, which can meet different types of tourist experiences as much as possible. The tourism product module of VR is shown in **Table 1**.

TABLE 1 | Tourism product function module of virtual reality (VR).

Serial number	Modular
1	Information inquiry and reservation
2	Online shopping
3	Virtual community
4	GIS
5	Virtual tourism

TABLE 2 | Development of virtual tourism.

Year	Total revenue of virtual tourism products	Total revenue growth rate	Total number of virtual tourism products	Total number growth rate
2016	26,456	15.7%	15.6	9.5%
2017	3,47,666	18.5%	14.5	8.4%
2018	4,21,562	20.6%	12.4	6.6%
2019	2,53,567	14.4%	11.4	5.6%
2020	1,45,678	5.6%	17.6	10.2%

Virtual tourism products not only improve the tourism experience of VR tourism products but also help to improve the tourism experience of real tourists. The main information source of this survey is a questionnaire survey, which is based on the use of tourism products and the consumption of tourism products by ordinary users. This study studies the impact of the use of tourism products on brand acceptance, cognition, attitude, and behavioral change of users. The purpose is to measure the effect of brand communication of tourism products from the perspective of users and investigate the development of virtual tourism products in recent years, as shown in **Table 2**.

Emotional VR Model of Virtual Tourism Products

The process of the research model is that based on the understanding of the connotation and dimension of the interaction between users in the tourism virtual community, we can understand the experience projects brought by the VR tourism products to the users. Then, according to the main elements of the theory of value co-creation and the context of its impact, this study constructs the research model and discusses and verifies the role of community users and online tourism enterprises in the interaction of virtual community users in tourism. Finally, based on the previous theories, this study verifies the influence of interaction and purchase decision, customer value and purchase decision, and

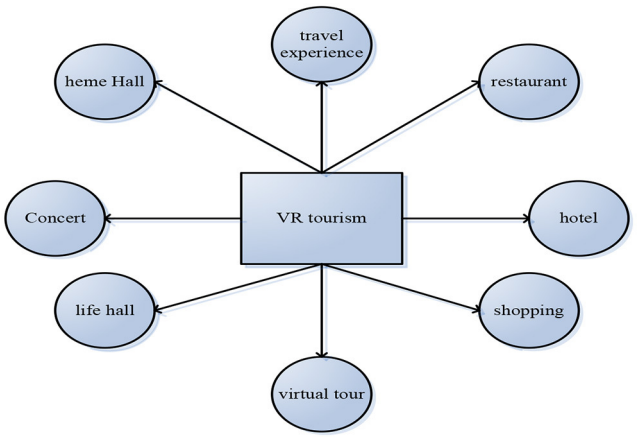


FIGURE 3 | Virtual tourism emotional VR model.

verifies the regulatory effect of enterprise control on the model (Peng et al., 2019). First, the practical application of virtual community in the tourism industry is tourism virtual community, its main characteristics are the same as a virtual community, different from that of participants in tourism virtual community, the common interest among participants is tourism products and information, and the formation of virtual tourism emotional VR model is shown in **Figure 3**.

Virtual Tourism

In recent years, during the small and long holidays, the news of overcrowding in popular tourist destinations at home and abroad occupies the headlines of the major media. In the case of such a large number of holiday tourists, it is particularly important to divert the number of tourists from popular tourist destinations by means of time shifting, traffic shifting, and choosing non-popular tourist routes. In addition to the above diversion methods, the virtual tourism developed by computer and Internet technology will also provide a new tourism mode for tourists to choose. Virtual tourism refers to the process that tourists participating in virtual tourism can personally carry out tourism activities by simulating or surpassing the real tourism landscape on the basis of the real tourism landscape, which can be divided into broad sense and narrow sense. In a broad sense, virtual tourism is closely related to traditional tourism, which mainly refers to the process of tourists obtaining relevant information of tourism destination before travel, such as through traditional and new media channels such as books, newspapers, television, and Internet. From the perspective of virtual tourism in a broad sense, virtual tourism and traditional tourism are simultaneous activities. The narrow sense of virtual tourism is more closely related to computer and Internet technology, mainly refers to the use of a variety of imaging technology VR tourism, and the process of making tourists get information about the tourist destination.

TABLE 3 | Six stages of tourism destination life cycle theory and its application.

Sequence number	Life cycle	
1	Exploration period	Natural tourism destination
2	Participation period	Basic tourism destination
3	Development period	Development tourism destination
4	Consolidation period	Popular tourism destination
5	Stagnation period	Popular tourism destination
6	Decline and recovery period	Decline and recovery period

EVALUATION OF VIRTUAL ECONOMY OF EMOTIONAL MARKETING OF TOURISM PRODUCTS

Virtual Tourism Plays Different Roles in the Six Types of Tourism Destination Development

The life cycle of a tourism destination is generally divided into six stages, namely, exploration period, participation period, development period, integration period, stagnation period, and recession period or recovery period. The development of the tourism industry in China is quite usual. According to Butler's tourism destination life cycle theory, the development of domestic tourism destinations can be divided into natural tourism destination, basic tourism destination, development tourism destination, popular tourism destination, tourism destination in decline period and tourism destination in recovery period as shown in **Table 3**.

According to **Table 3**, the main virtual tourism products can be used to try to produce leading or diversion effects in popular tourism destinations and also try to stimulate different effects in the six types of tourism destination development stages:

(1) Promotion function: in the natural tourist destination, the promotion form of virtual tourism products and is selected to increase the original tourism resources.

Industrial exhibition, to show the world the charm of natural tourism, plays a leading role in the real sense.

(2) Attraction: in the basic tourism destination, the advertising form of virtual tourism products is selected and put on the market to attract more tourists.

(3) Introduction function: in the development of tourism destinations, the introduction form of virtual tourism products and is selected to let tourists in many scenic spots.

Choose a closer destination in the future.

(4) Distributary function: in popular tourist destinations, we should pay attention to the sales and popularization of virtual tourism products and pay attention to saturated tourism in peak season.

The number of customers is divided.

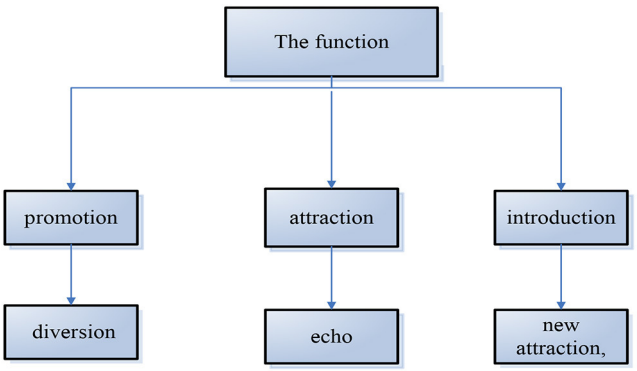


FIGURE 4 | The function of virtual tourism products.

(5) Echo function: in the mature tourist destination, choose the propaganda form of virtual tourism products, to the tourists who have visited the tourist destination.
(6) New attraction: in declining or reviving tourist destinations, virtual tourism should be selected as a new attraction to arouse reconsideration of tourists and understanding of tourist destinations. Of course, virtual tourism products should play a good role in the six types of tourist destinations at the development stage, provided that policy makers should correctly understand and pay attention to the promotion of virtual tourism products, as shown in **Figure 4**.

Virtual tourism is a more convenient and relaxed choice. In addition to the good relationship between traditional tourism and virtual tourism, virtual tourism can also play different roles in the development of tourism destination.

Advantages of Virtual Tourism Products

First, tourism products in tourism and emotional factors are culture, experience, and memory. In the construction of virtual tourism product scene, we can also put tourism and emotion-related factors into it for thinking and transform them into corresponding three designs of tourism emotion as follows: "cultural factors" form "cultural design," "experience factors" form "interactive design," and "memory factors" form "transformation design." Guided by the three-level theory of emotional design, combined with the three-level design of tourism emotion, the "cultural factor" is transformed. For the level of "cultural design" corresponding to emotional "instinct design," "experience factor" converted to "interactive design" corresponding to emotional "behavior design," "memory factor" converted to "transformation design" corresponding to emotional "reflection design," and the construction of virtual tourist attractions discussed in this study is shown in **Table 4**.

It can be seen from **Table 4** that the basic model framework of virtual tourism emotional design layer consists of three parts, namely, cultural design layer, emotional design layer, and transformation design layer, and "the cultural design" part is to use the cultural elements of virtual tourist attractions to awaken the recognition of tourists and appreciation of traditional tourist attractions. Considering the actual behavior needs of tourists in tourist attractions and the very important interactive part in virtual tourist attractions, the interactive function design reflects this, and the transformation design emphasizes the promotion of emotional sublimation of tourists in virtual tourism.

TABLE 4 | Emotional design three-level theory.

Serial number	Design	Factors
1	Instinct design	Cultural factors
2	Behavior design	Interaction factors
3	Reflective design	Memory factors

TABLE 5 | The economic effect of virtual tourism.

Year	Virtual tourism product volume	Virtual tourism store	Number of virtual tourism experience
2017	300	50	1,000
2018	500	80	5,000
2019	1,000	150	10,000
2020	5,000	400	70,000

Virtual Economic Effect of Virtual Tourism Products

With the rapid development of virtual tourism in recent years, the income of VR tourism products has been greatly improved. Its development should not be just a vague imitation, and specific development needs clear scientific planning and formulation including the comprehensive integration strategy of independent operation of provincial tourism bureaus and publicity of the overall image of national tourism bureaus. As a comprehensive and balanced tourism aggregate, China should formulate a comprehensive national tourism plan based on the specific development of local tourism characteristics, form differentiated tourism products, guide the complementary development of local tourism, and avoid fierce competition. Therefore, the quantitative analysis of China's VR tourism product income and the corresponding strategic suggestions are helpful to avoid the blindness and irrationality in the planning, form a scientific and reasonable industrial planning, reduce redundant construction and waste, strengthen the local cooperation and the overall collaborative development, improve the development efficiency, and strive to form a healthy and efficient industrial chain, According to the pattern of maximizing the internal and external economic and social benefits, the virtual tourism products, shops, and number of people experience in a county in recent 5 years are shown in **Table 5**.

It can be seen from the table that with the increase of years from 2017 to 2020, the volume of virtual tourism products has increased from 300 to 5,000, the number of virtual tourism shops has increased from 50 to 400, and the number of virtual tourism experience has increased from 1,000 to 70,000. A three-dimensional virtual world can make people feel more real

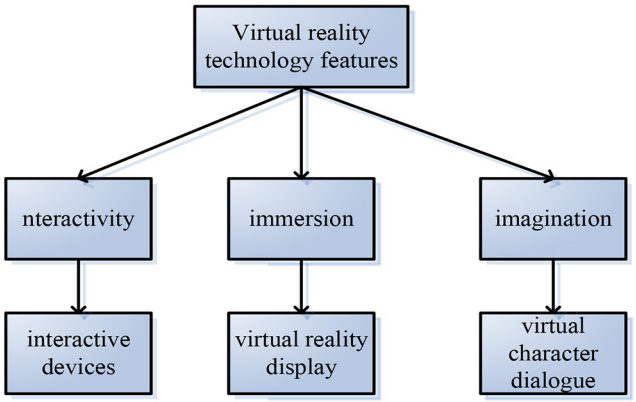


FIGURE 5 | Characteristics of VR technology.

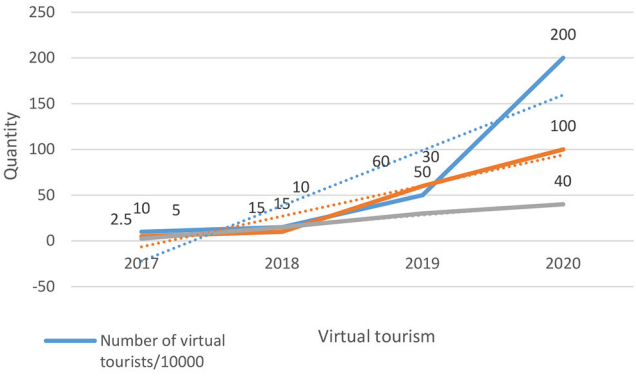


FIGURE 6 | Parameter selection for different applications.

"reality" experience. The goal of combining VR technology with computer graphics, computer simulation, artificial intelligence, multimedia, sensors, and other technologies is to make the perception experience of people in the virtual world more and more delicate and real. The technical advantages of VR technology can enrich the design possibilities of more functions, experiences, and interfaces that can be added in the construction of virtual tourist attractions. The interactivity, immersion, and imagination of VR technology will stimulate more virtual experiences in future, as shown in **Figure 5**.

Although VR technology is designed based on the real world, its virtual nature can make the virtual world more imaginative and redesigning space than the real world. Designers add more or more interesting functions to the virtual user experience so that users can immediately feel the services that are difficult to have in reality and stimulate more imagination and thinking after experiencing VR products. From this point of view, VR will change the cognitive and learning styles of people. The economic effect of the VR tourism products investigated above is shown in **Figure 6**.

It can be seen that from 2017 to 2020, the number of virtual tourists will increase from 1,00,000 to 2 million, the revenue of virtual products will increase from 50 million to 1 billion, and the growth rate will increase from 2.5 to 40%. In the VR tourism products, immersive VR strives to create an immersive environment for users, put the five senses of users in the VR space for natural interaction, use various output and input devices for interaction in line according to the law of movement, and pay attention to the experience of virtual simulation. This type of equipment has high requirements and strong experimental nature and is not easy to be popularized before the cost performance and portability of related equipment are improved. Augmented reality VR is a kind of VR based on reality. It solves different problems by superimposing virtual illusion on the real scene. In recent years, with the progress of software and hardware, the topic of VR and augmented reality is more and more discussed. People pay more attention to and look forward to how VR and augmented reality will change the way of human life in the future. With the development of science and technology, VR technology will have great development prospects, as shown in **Figure 7**.

In the stage of data collection of VR tourism products, local cultures need to be substituted into the construction of a virtual tourism scene. The image information collected by the designer as the best "texture" performance of the scenic spot will directly reflect the appearance image of the characteristic landscape, architectural culture, and other material culture of the scenic spot. Whether in the construction of virtual tourist attractions based on two-dimensional images or three-dimensional geometric models, it is necessary to collect the appearance data of tourist destinations in a certain standard and way, as a preliminary presentation of cultural factors. In the process of field data collection, the local intangible culture obtained will be collected as reference materials for future design, as shown in **Figure 8**.

The more simulation the scene of virtual tourism VR is designed, the easier it is for users to experience the feeling of being on the scene. Therefore, from the perspective of "people (users)," through the investigation and redesign of the real scene of the tourist destination, the "authenticity" of the scene in the virtual tourist attractions should be put in the first place of the construction principles of the virtual tourist attractions. Second, in terms of immersion in interaction, we need to design a natural interaction process to achieve a more "real" interaction behavior, as shown in **Table 6**.

In the stage of scene design in the construction of virtual tourist attractions, processing cultural factors are more important. With the careful embellishment and auxiliary design of cultural factors, the characteristic cultural atmosphere of the tourist destination can be created, so that users can feel the charm of local culture in the scene. In the gradation of scene design, it is difficult to add the idea of developers or designers to the two-dimensional virtual tourist attractions based on image construction design. The purpose of constructing virtual tourist attractions is to enable users to quickly reach the tourist destination in the virtual world for sightseeing activities because the plane images are based on the reality of the tourist destination. In two-dimensional virtual tourist attractions, the

FIGURE 7 | Technological innovation of information industry (http://alturl.com/knbeo).

FIGURE 8 | Virtual tourism experience on site (http://alturl.com/g2aeg).

appearance features of material culture are directly displayed through two-dimensional images, as shown in **Figure 9**.

CONCLUSION

This study studies the virtual economy effect evaluation of emotional marketing of tourism products based on VR and studies the meaning and algorithm of VR. The analysis of the virtual economy situation of tourism products based on VR shows that the existence and development of VR tourism products are inevitable and historical; however, they cannot completely replace real tourism and become the theme of future tourism mode. Only as an effective supplement to real tourism, they can improve the satisfaction of tourists with

TABLE 6 | Scene interaction process of virtual tourist attractions.

Serial number	Experience feeling
1	Authenticity
2	Interaction
3	Randomness
4	Fluency
5	Expansibility

FIGURE 9 | Virtual tourism products (http://alturl.com/aaa7o).

real tourism and the quality of the tourism experience. VR tourism products are the products of the combination of tourism and information technology and are the requirements of the development of modern society and the growing tourism demand of human beings. Therefore, VR tourism products have their supporting conditions for existence and development. The characteristics of VR tourism products determine that they will inevitably bring great influence to the current tourism system. It not only effectively solve the lack of transferability of real tourist attractions and the restrictions of real tourism on the leisure and income of tourists but also be affected by the current VR technology and the nature of VR tourism products. At present, VR tourism products can bring tourists only superficial and vulgar tourism experiences.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

MT: experimental designing and editing. SM: experimental process implementing and writing. NX: data collecting and data analysis. All authors contributed to the article and approved the submitted version.

REFERENCES

Bagus, C., Brahmasari, I. A., and Suryani, T. (2020). The effects of local culture on tourism products and domestic tourist satisfaction in tourism destinations. *Arch. Bus. Res.* 7, 233–243. doi: 10.14738/abr.712.7560

Bastug, E., Bennis, M., and Medard, M. (2017). Toward interconnected virtual reality: opportunities, challenges, and enablers. *IEEE Commun. Mag.* 55, 110–117. doi: 10.1109/MCOM.2017.1601089

Berg, L. P., and Vance, J. M. (2017). Industry use of virtual reality in product design and manufacturing: a survey. *Virtual Real.* 21, 1–17. doi: 10.1007/s10055-016-0293-9

Donath, L., Rössler, R., and Faude, O. (2016). Effects of virtual reality training (Exergaming) compared to alternative exercise training and passive control on standing balance and functional mobility in healthy community-dwelling seniors: a meta-analytical review. *Sports Med.* 46, 1293–1309. doi: 10.1007/s40279-016-0485-1

Elbamby, M. S., Perfecto, C., and Bennis, M. (2018). Towards low-latency and ultra-reliable virtual reality. *IEEE Netw.* 32, 78–84. doi: 10.1109/MNET.2018.1700268

Fabisiak, L. (2018). Web service usability analysis based on user preferences. *J. Organ. End User Comput.* 30, 1–13. doi: 10.4018/JOEUC.2018100101

Falconer, C. J., Rovira, A., and King, J. A. (2016). Embodying self-compassion within virtual reality and its effects on patients with depression. *Bjpsych Open* 2, 74–80. doi: 10.1192/bjpo.bp.115.002147

Freeman, D., Bradley, J., and Antley, A. (2016). Virtual reality in the treatment of persecutory delusions: randomised controlled experimental study testing how to reduce delusional conviction. *Br. J. Psychiatry* 209, 62–67. doi: 10.1192/bjp.bp.115.176438

Freeman, D., Reeve, S., and Robinson, A. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychol. Med.* 47, 1–8. doi: 10.1017/S003329171700040X

Khatwani, G., and Srivastava, P. R. (2018). Impact of information technology on information search channel selection for consumers. *J Organ. End User Comput* 30, 63–80. doi: 10.4018/JOEUC.2018070104

Lan, N. (2020). Designing tourism products for Dak Nong province in the perspectives of regional linkage. *Sci. Technol. Dev. J.* 3, 162–171. doi: 10.32508/stdjssh.v3i3.527

Lenoir, J., Cotin, S., and Duriez, C. (2017). Interactive physically-based simulation of catheter and guidewire. *J. Prev. Med. Inf.* 61, 2132–2141.

Meilin, W. (2017). Exploration strategy of sports tourism products viewed from experience marketing. *J. Sports Adult Educ.* 33, 30–33. doi: 10.16419/j.cnki.42-1684/g8.20170420.018

Munafo, J., Diedrick, M., and Stoffregen, T. A. (2017). The virtual reality head-mounted display Oculus Rift induces motion sickness and is sexist in its effects. *Exp. Brain Res.* 235, 889–901. doi: 10.1007/s00221-016-4846-7

Oladipo, O. S., and Modupe, O. F. (2020). Impact of tourism products development on osun osogbo sacred grove and badagry slave trade relics. *J. Tour Manag. Res.* 7, 170–185. doi: 10.18488/journal.31.2020.72.170.185

Peng, J. P., Quan, J., and Peng, L. (2019). It application maturity, management institutional capability and process management capability. *J. Organ. End User Comput.* 31, 61–85. doi: 10.4018/JOEUC.2019010104

Riva, G., Bacchetta, M., and Baruffi, M. (2016). Virtual reality environment for body image modification: a multidimensional therapy for the treatment of body image in obesity and related pathologies. *Cyberpsychol. Behav.* 3, 421–431. doi: 10.1089/10949310050078887

Saposnik, G., Cohen, L. G., and Mamdani, M. (2016). Efficacy and safety of non-immersive virtual reality exercising in stroke rehabilitation (EVREST): a randomised, multicentre, single-blind, controlled trial. *Lancet Neurol.* 15, 1019–1027. doi: 10.1016/S1474-4422(16)30121-1

Schmitz, S., and Tsobgou, D. L. (2016). Developing tourism products and new partnerships through participatory action research in rural Cameroon. *Geogr. Res.* 54, 143–152. doi: 10.1111/1745-5871.12174

Sharar, S. R., Alamdari, A., and Hoffer, C. (2016). Circumplex model of affect: a measure of pleasure and arousal during virtual reality distraction analgesia. *Games Health J.* 5, 197–202. doi: 10.1089/g4h.2015.0046

Szczsna, J., and Wesoowska, M. (2020). Assessment of the effect of new tourism products on the economic recovery of rural areas in the context of empirical research on selected communes of the lublin voivodeship. *Tourism* 30, 53–60. doi: 10.18778/0867-5856.30.1.17

Vankipuram, A., Khanal, P., and Ashby, A. (2017). Design and development of a virtual reality simulator for advanced cardiac life support training. *IEEE J. Biomed. Health Inform.* 18, 1478–1484. doi: 10.1109/JBHI.2013.2285102

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Talafubieke, Mai and Xialifuhan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 04 January 2022
doi: 10.3389/fpsyg.2021.799132

An EEG Neurofeedback Interactive Model for Emotional Classification of Electronic Music Compositions Considering Multi-Brain Synergistic Brain-Computer Interfaces

*Mingxing Liu**

Music School, Shanxi Normal University, Linfen, China

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Jin Yan,
Hunan Normal University, China
Kambiz Badie,
Iran Telecommunication Research Center, Iran
Yangxu Zhou,
Jilin University, China

***Correspondence:**
Mingxing Liu
liumingxingyy@sxnu.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *21 October 2021*
Accepted: *17 November 2021*
Published: *04 January 2022*

Citation:
Liu M (2022) An EEG Neurofeedback Interactive Model for Emotional Classification of Electronic Music Compositions Considering Multi-Brain Synergistic Brain-Computer Interfaces. Front. Psychol. 12:799132. doi: 10.3389/fpsyg.2021.799132

This paper presents an in-depth study and analysis of the emotional classification of EEG neurofeedback interactive electronic music compositions using a multi-brain collaborative brain-computer interface (BCI). Based on previous research, this paper explores the design and performance of sound visualization in an interactive format from the perspective of visual performance design and the psychology of participating users with the help of knowledge from various disciplines such as psychology, acoustics, aesthetics, neurophysiology, and computer science. This paper proposes a specific mapping model for the conversion of sound to visual expression based on people's perception and aesthetics of sound based on the phenomenon of audiovisual association, which provides a theoretical basis for the subsequent research. Based on the mapping transformation pattern between audio and visual, this paper investigates the realization path of interactive sound visualization, the visual expression form and its formal composition, and the aesthetic style, and forms a design expression method for the visualization of interactive sound, to benefit the practice of interactive sound visualization. In response to the problem of neglecting the real-time and dynamic nature of the brain in traditional brain network research, dynamic brain networks proposed for analyzing the EEG signals induced by long-time music appreciation. During prolonged music appreciation, the connectivity of the brain changes continuously. We used mutual information on different frequency bands of EEG signals to construct dynamic brain networks, observe changes in brain networks over time and use them for emotion recognition. We used the brain network for emotion classification and achieved an emotion recognition rate of 67.3% under four classifications, exceeding the highest recognition rate available.

Keywords: multi-brain collaborative brain-computer interface, EEG neural, interactive electronic music, emotion classification, brain-computer interfaces

INTRODUCTION

Brain-computer interface (BCI) is a communication system that does not rely on the brain's normal peripheral nerve and muscle output pathways. BCI enables direct communication between the human brain and the outside world or device control by creating a virtual channel between the human brain and an external device, which provides a possible path to recovery for people with paralysis or other motor dysfunction disorders (Rossetti et al., 2020). BCI technology involves neuroscience, computers, signal processing, pattern recognition, and other disciplines, and through the processing and classification of physiological signals collected from the human brain, it identifies the various intentions of the subject, and then converts the intentions into specific control commands, thus achieving direct control of external devices by the human brain. However, few BCIs systems can be put into large-scale applications. BCI technology involves neuroscience, computer, signal processing, pattern recognition and other disciplines. By processing and classifying the physiological signals collected from the human brain, it recognizes various intentions of the subjects, and then converts the intentions into specific controls. Command, so as to realize the direct control of the human brain to the external equipment. This is partly because real-time systems do not meet the requirements of large-scale applications in terms of performance, and partly because most of the current BCI systems are still stuck in the synchronization mode. Synchronous mode means that the system pre-defines the time window in which the user needs to operate, and when the system starts, the user always needs to pay attention to the system prompt and start or stop the brain control operation according to the time window prompted by the system, and the system will continuously classify and identify each time window, and perform the operation according to the classification result (Farrell, 2021). On the one hand, the user needs to focus on the system prompts all the time, which prevents the user from performing daily actions outside the system and increases the user's mental load, which can lead to fatigue. On the other hand, the user cannot start or stop the operation at any time according to his or her wish. When the user needs to operate, the system needs to be started first. When the user does not need to operate, the system cannot stop the operation at any time according to the user's will, but continues to continuously output commands and execute the operation until the system is shut down, which greatly destroys the autonomy of the BCI system.

In the field of art, visual art and auditory art are the most common forms of art, each of which can provide people with both physical and psychological aesthetic experiences (Dahan, 2020). In the traditional sense, auditory art can only be experienced through the auditory senses, and the form of experience is relatively single. With the development of multimedia technology, people have put forward higher requirements for the aesthetic experience of auditory art, seeking more diverse forms of artistic expression. The visualization of sound provides people with a new form of visual and auditory expression for the appreciation of auditory art, and the new visual interpretation of auditory art adds to the expressiveness of auditory art, providing people with the dual beauty of the combination of visual and auditory senses, and the visualization of sound is conducive to people's deep understanding of the formal structure and emotional content of artworks. For example, in the live performance of the Czechia Philharmonic Orchestra, the designer graphically designed the music to be played according to the rhythm, and displayed the 3D effect simultaneously with the performance, bringing a visual feast to the audience. With the advancement of technology and different application needs, forms of interaction between humans and machines and systems are used in various fields, and interactive sound visualization is developing from the purely artistic field to more practical directions, such as aiding applications in medical, entertainment, and teaching fields (Simurra and Borges, 2021). Interactive sound visualization can play the audience's specificity to produce different visual images, so it can be used in the diagnosis of certain articulation diseases and aided medical applications; in music teaching, interactive sound visualization can show students the visual structure of music, facilitating students to operate intuitively according to their learning situation; interactive sound visualization can play an active role for hearing-impaired people in speech recognition. Interactive sound visualization can play an active role in speech recognition, music performance, and other aspects.

Different cultures understand music differently, but they all have one thing in common and that is that through music one can experience the culture for a short period. Music is a way for composers to express their thoughts and feelings to listeners through specific sound structures and melodic variations; music portrays emotions and moods in different environments to listeners through sound (Turchet et al., 2021). The user needs to pay attention to the system always prompts, start, or stop the brain control operation according to the time window prompted by the system, the system will continuously classify and identify each time window, and perform operations according to the classification results. One of the basic cognitive functions of humans is the emotional response to music. Different people may have different emotional reactions to the same music because of their different life experiences, but music expresses the composer's thoughts and feelings and he will infect the listener's thoughts and feelings with a strong rendering, and from this perspective, the impact of music on different people's thoughts and feelings is again similar. Accurate recognition of musical emotions can achieve harmonious interaction between human and machine emotions, which has great application prospects in various fields. For example, in daily life, real-time emotion recognition can help people regulate their emotional state promptly and change the current emotional state of a person through music to reduce accidents caused by emotional factors.

CURRENT STATUS OF RESEARCH

The new research method of parsing musical elements and describing the changes in the corresponding quantitative values of different elements and emotions has pioneered the empirical study of musical emotions, allowing a close connection between

the two. The results available so far show that emotions related to music are divided into two main aspects: object and subject (Wolffenbüttel and Pellin, 2020). The object is the listener's perception of the emotion expressed in the musical work itself, i.e., the perception of the emotion the composer wants to express, which is universal and does not vary according to individual differences; the subject is the emotional response of the listener's nature that is evoked by the musical stimulus (Iverson, 2019). Instead, it continues to output instructions and perform operations until the system is shut down, which greatly undermines the autonomy of the BCI system. The main difference between object and subject is that the object is universal and the subject has individual differences. The Musical Cue Coherence Model posits that brain processing of musical events is the process by which the brain extracts different musical features and then achieves integration (Turchet et al., 2021). Emotion is the brain's representation of these different musical features, and the listener's emotional perception can be predicted by elements of music such as key, melody, and rhythm. Major keys in classical music correspond mostly to positive emotions, while minor keys correspond mostly to negative emotions (Biasutti and Concina, 2021). The synesthesia theory of musical emotion suggests that the process of the listener's enjoyment of music is the process of achieving interaction with the creator and that the occurrence of musical emotion is the process of achieving emotional empathy and sympathy between the listener and the composer (Malecki et al., 2020).

Synesthesia theory suggests that a person interacting with the external environment causes the person to align with the environment by adapting themselves to it (Romão, 2020). The listener's process of listening to music is the process of aligning himself or herself with the emotions of the music. This theory emphasizes the music itself at the expense of the listener's subjective feelings (Lakatos, 2020). The areas activated by dissonant music are more pronounced and include the hippocampus, Para hippocampal gyrus, and temporal pole which equate to areas highly relevant to emotional perception. At the same time, the arousal of the inferior frontal gyrus, and inferior Brodmann's area was stronger for harmonic music, and the activation of the relevant parts reflected, to some extent, the operational processes of musical structure, syntactic analysis, and memory. It is noted that the activation of brain regions increases with increasing stimulation time, indicating to some extent the dynamic effect of emotion (Redhead, 2021). In conjunction with the use of ER-related event analysis in his study, it was found that the pattern of neural activation differed between the two situations, as expected and differently expected and that the brain regions and brain activity activated during listening to favorite, familiar music, or singing belonged to positive emotions. This finding is similar when complexity changes are introduced, their activation patterns, and discordant intervals are present. It is easy to see that the expression of emotions by music involves extremely complex cognitive processes. The new visual interpretation of auditory art adds to the expressive power of auditory art, and provides people with a dual aesthetic that combines audiovisual senses.

The first thing that is evoked when listening to music is a memory mechanism, where the brain integrates self-knowledge and activates autobiographical nerves. So, there is a strong emotional trigger on familiar songs. Also because of the similarity of the left-skewed spiking tract that occurs when predicting the appearance of desired music to the characteristics of positive emotions (Sofer, 2020). So, it is said that familiarity with various types of music directly influences to some extent the emotional experience it evokes. When the musical experience is sufficiently rich for non-harmonic tunes to become part of its familiarity, the negative emotional experience is reduced to some extent. To meet the demand for continuous mining and exploration of the information carried by brain neural signals and comprehensive, flexible and in-depth exploration of peak potential signals, the thesis, based on the study of the principles of peak potential signal analysis and decoding algorithms, designs a componentized algorithm module, abstractly encapsulates the algorithm module for multiple operations, realizes flexible configuration of the analysis process, and provides two analysis modes, real-time and offline mode, which meets the experimental research requirements of peak potential signal analysis.

ANALYSIS OF EEG NEUROFEEDBACK INTERACTIVE ELECTRONIC MUSIC COMPOSITIONS WITH A MULTI-BRAIN SYNERGISTIC BRAIN-COMPUTER INTERFACE FOR EMOTIONAL CLASSIFICATION

Multi-Brain Synergistic Brain-Computer Interface EEG Neurofeedback Analysis

The brain has a powerful information processing capacity, and several functional brain regions are involved in the processing of visual and auditory information. To investigate the specific connection process and information interaction of neural connections within the brain under brain cognitive activities, it is necessary to understand the neurophysiological basis of brain structure and the basic methods of brain cognitive experiments (Görgün, 2020). The progress of neuroscience research cannot be achieved without a rigorous experimental design and a standardized experimental process to obtain more realistic and effective experimental data of EEG signals. In this experiment, we used long-time music to induce emotion in the subjects, and at the same time, we asked the subjects to fill in the emotion scale to record their emotional states and recorded the EEG signals of the subjects during the whole experiment. It is helpful for students to operate and learn intuitively according to their own learning situation. Interactive sound visualization can play an active role in speech recognition and appreciation of music performances for the hearing impaired. The corresponding cognitive patterns were derived by analyzing the EEG signals in different emotional states, exploring the relationship between music and emotion evocation, and exploring the differences in EEG characteristics in different emotional states. It is an improved adaptive algorithm using the

eigenmodal decomposition to solve the Hilbert transform in the frequency band requirement that cannot be too wide. Among the requirements for the eigenmodal decomposition function (IMF) are two. The number of zero-poles is equal or at most a little different throughout the data set. At any moment, the upper envelope consisting of the very large values and the lower envelope consisting of the very small values both have zero mean, or concerning the time axis, the upper and lower envelopes are symmetrical then the division is considered good and the next level of decomposition is performed.

First, we use two algorithms to model the offline data with different data lengths (0.5–1s with 0.1s steps) in the training phase and use the "leave-one-out" method for cross-validation: we divide the offline training data with 6 frequencies and 10 blocks into 10 groups, each containing 6 frequencies, and then use one of the groups as the test set and the other nine as the training set. We divide the offline training data with 6 frequencies and 10 blocks into 10 groups, each containing 6 frequencies, and then use one group as the test set and the other 9 groups as the training set. The template data is then used to classify and identify the data for each trial of the test set to obtain the correct classification rate for a set of test sets. After ten sets of data are tested for test set traversal, the final offline recognition correct rate is obtained by superimposing and averaging the classification correct rates of each set. When the online experiment is conducted, when the online system detects the time-synchronized signal sent from the stimulation program to the EEG amplifier, 1s of data from the sampling point where the time-synchronized signal is located is intercepted and matched with the template signal of six frequencies, respectively to obtain the final decision result and perform speech feedback.

$$p_{u,v} = \frac{a^T \sum_{12} b^2}{\sqrt{a^T \sum_{12} a^2}\sqrt{b^T \sum_{22} c^2}} \tag{1}$$

Explore the differences in EEG characteristics in different emotional states. The principle of linear discriminant analysis is very simple (Smith et al., 2021). It projects two classes of data onto a straight line through a projection matrix. It is required that the projection onto the straight line maximizes the difference between the two classes of data by making the mean values of the two classes the farthest apart while maximizing the degree of intra-class aggregation in each class. The test data will be projected onto the same straight line and the class it belongs to will be decided based on the projected position. Due to its simple principle and fast computation, it is often used in the classification of EEG data, as shown in **Figure 1**.

EEG signals obtained by experimental paradigm evocation are subject to various noise disturbances during acquisition, as well as electrical activity caused by neural responses unrelated to the inclusion of other cognitive psychological activities, while the potential changes that are truly relevant to brain cognition are only a few microvolts, which are usually much smaller than the potential changes caused by spontaneous activity of nerve cells, so the measured EEG signals need to be preprocessed before the study of EEG signal data (D'Agostino, 2020). The common preprocessing process of the EEG technique contains steps such as EEG preview, filtering, eye-electricity removal, EEG signal segmentation, baseline correction, and conversion reference. The acquired data contains many EEG artifacts due to the accompanying blinks or eye movements of the subject during the completion of the task. The traditional method of removing the oculomotor is by placing eye electrodes, acquiring the maximum value of the oculomotor, defining the oculomotor artifacts by the percentage of the maximum value, and finally removing the oculomotor point by point. But now we can separate the oculomotor components by independent component analysis (ICA) and reduce the signal of the removed oculomotor components to achieve the effect of removing the oculomotor.

$$\delta(t) = \sum_{k=2}^{K} IM_k(t^2) \tag{2}$$

$$\delta_p(t) = \sum_{k=1}^{K} IM_k(t^2) / \sum_{k=1}^{M} I^2 M_k(t) \tag{3}$$

Judgments are made after the criteria have been derived and those greater than the average contribution are recorded therein, while those less than are not calculated. After completion, the proportion of the overall that is accounted for is the one that has a particular contribution to that frequency band over a certain period and can also be considered as the dominant change. Therefore, it is further used as a parameter. The best or worst of these features are selected and given a score. After removing that feature continue the above operation with the remaining features. The result is a score of all features. It is a greedy algorithm that finds the optimal feature subspace by traversing it. One thing worth noting is that much of the stability of RFE comes from the choice of model. If a general regression algorithm is used then instability may arise, if a Ridge regression algorithm is used then Ridge itself is stable and its results are stable.

Self-encoder is an unsupervised learning neural network, whose core idea is to encode the data in the input layer (dimensionality reduction) and continuously improve the encoder and decoder by reducing the residuals between it and the original data. Ultimately, the purpose of dimensionality reduction representation is achieved. Similarly, for the input defined as a sample, supervised learning can be done by a combination of several layers of self-encoders. The method of removing electrooculogram is to obtain the maximum value of electrooculogram by placing the eye electrodes, define the electrooculogram artifact by the percentage of the maximum value, and finally remove the electrooculogram point-to-point. It is also possible to train to generate new data that is different from the samples and can be used as a generative model. Based on this the development of denoising self-encoders, stacked self-encoders, etc., is also based on the principle of reducing bias by its nature.

$$\phi_{ij}(x) = \frac{1}{\delta^d} \exp\left(\frac{\left(x - x_{ij}\right)\left(x + x_{ij}\right)^T}{\delta^2}\right) \tag{4}$$

$$Entropy = \sum_{i=1}^{n} p_i \cdot \ln(p_i) \tag{5}$$

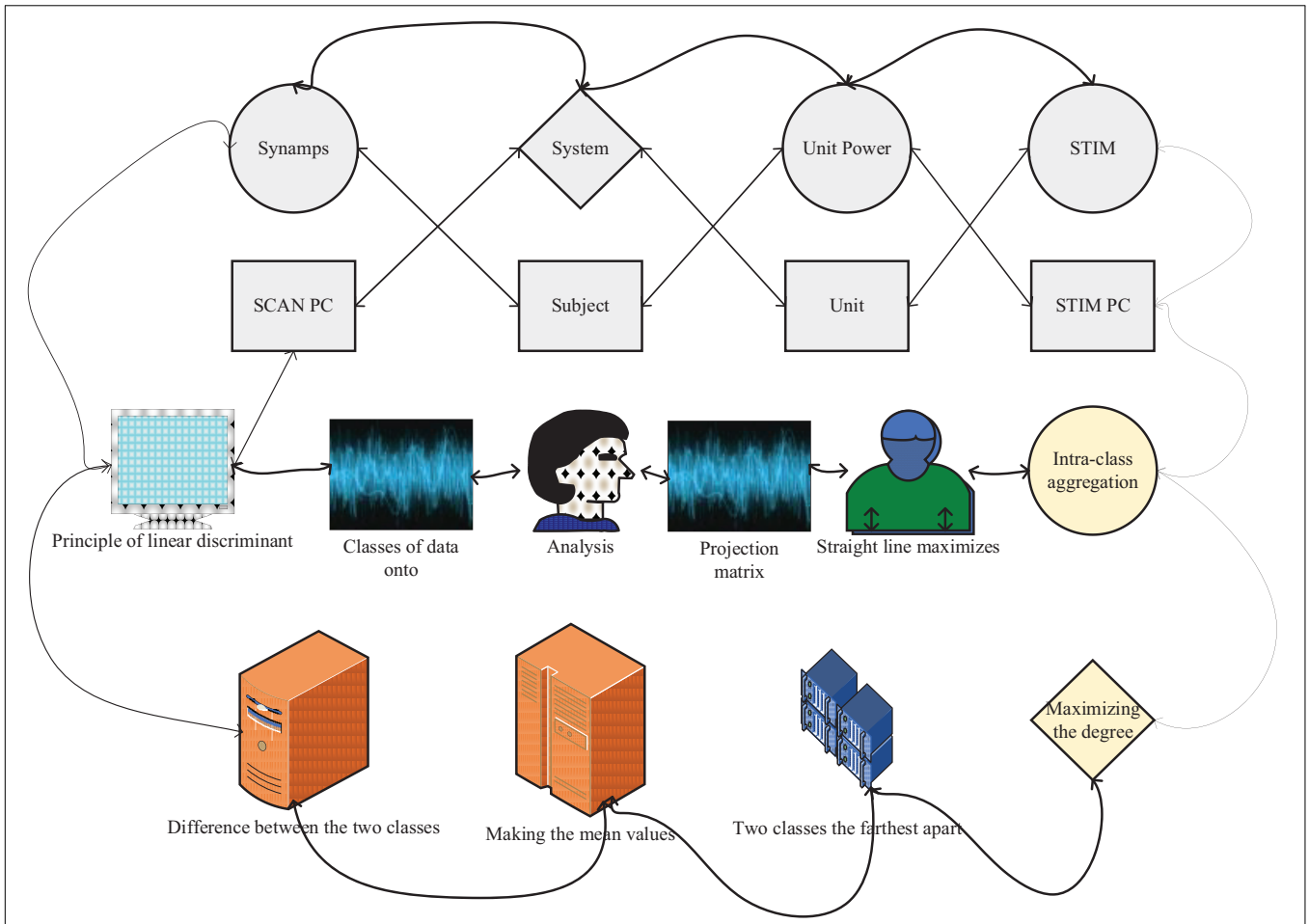


FIGURE 1 | Multi-brain collaborative brain-computer interface (BCI) EEG neurofeedback framework.

As with the Shannon first, the higher the Gini coefficient means the less pure the data is based on the information first improvement there is also the use of information gain rate to overcome the problem that the number of sample categories varies and the decision tree is biased to split the branch with the most samples, as well as splitting stop conditions, pruning and other conditions not detailed here (Kerékfy, 2020). However, through the decision tree, we can find that the decision tree is also divided based on the category samples. This means that different distribution of sample categories may produce different tree nodes. As an anisotropic classification, decision trees also meet the requirements of music emotion model learning well but face the same huge dimensionality problem. Decision trees can be combined to obtain an optimal feature subspace, as shown in **Figure 2**. The physiological characteristics of the EEG signal representations vary with frequency, so the EEG signals in different frequency domains differ in their discrimination between different categories of motor imagery. There are also differences in the function of different brain regions, so the degree to which different channels of EEG signals discriminate between different categories of motor imagery varies. The ERD/ERS phenomenon shows that the frequency domains associated with motor imagery mainly concentrated in the alpha band (8–13 Hz) and the bate band (13–30 Hz), and the channels mainly concentrated in the motor areas of the parietal lobe.

Although these frequency domains and channels are the main frequency domains and channels for performing motor imagery classification, the study shows that motor imagery has individual variability and there is still a degree of variation in the specific frequency domains and channels of different subjects; on the other hand, these frequency domains and channels are only for the energy features in the synchronous mode. There is no clearer frequency and channel selection tendency for the distinction between NC and IC states in asynchronous mode as well as PLV features. For these reasons, it is necessary to perform feature screening on a larger scale.

$$p_n = \sum_{i=1}^{n} a(m) p_i \cdot \ln(p_i) r_n^m \tag{6}$$

After music feature extraction for music, we successfully convert the music signal into a music feature signal, and the next step is to find the event points on the music feature signal. The music feature signal contains many mutations, but not all mutations

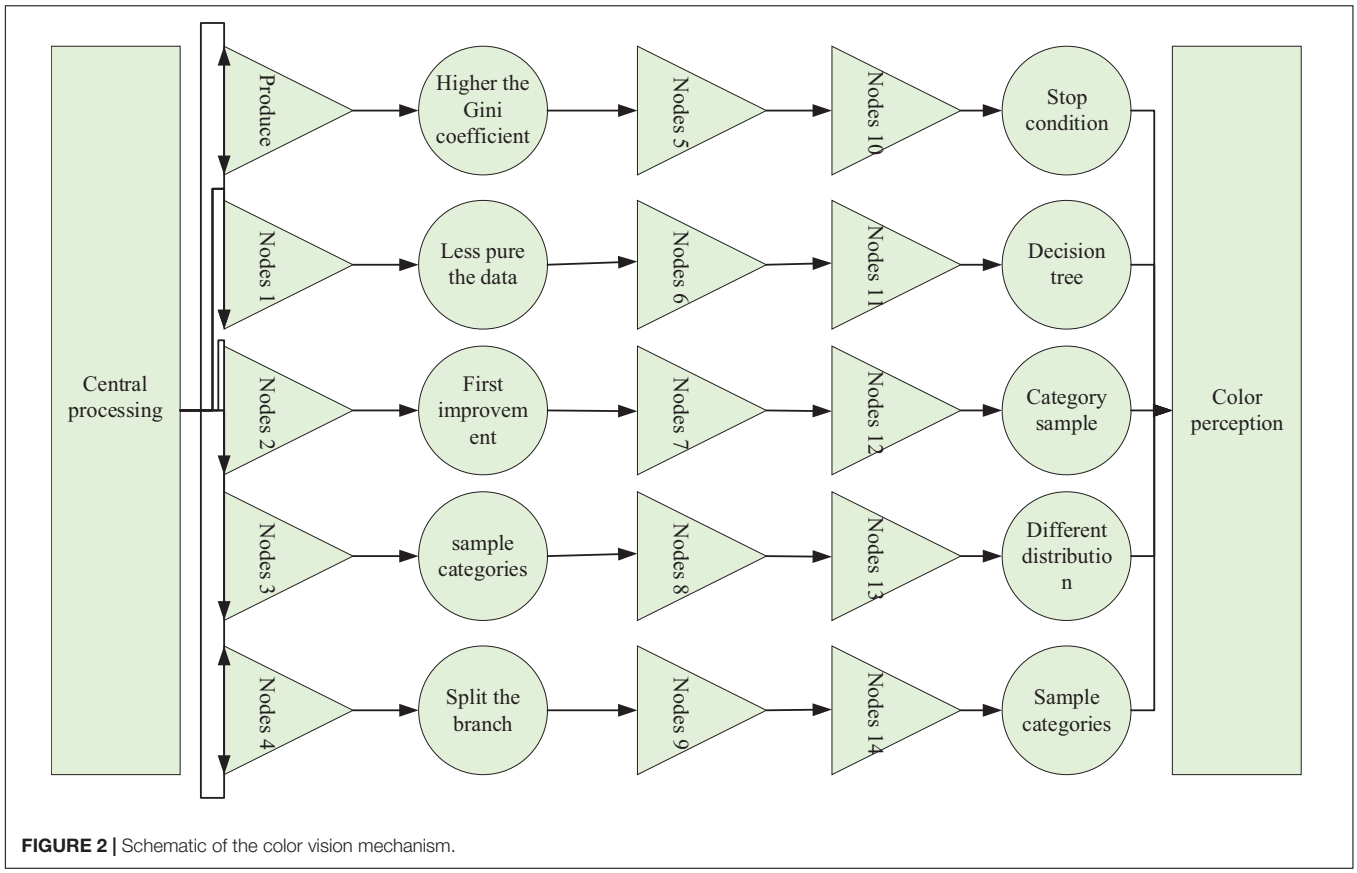


FIGURE 2 | Schematic of the color vision mechanism.

can be regarded as feature event points because we use the ERP technique, so the feature event points should satisfy the characteristics of both mutation points and ERP events, therefore, we define the feature event points simply.

Experimental Design of Interactive Electronic Music Piece Sentiment Classification

The energy signature reflects the physiological properties of motor imagery, but it is a single-channel signature that is more susceptible to the non-smoothness of the EEG signal. Our brain is a complex information processing center, and to process information more efficiently, the brain is divided into multiple functional areas, each of which is responsible for a different task; for example, image information processing is handled by the visual cortex located in the occipital lobe, and motor function is handled by the motor area located in the frontal lobe. The processing of complex tasks relies on the cooperation of multiple functional areas. The cooperation of multiple functional areas is carried out through information interaction, which is reflected in the scalp EEG as a dual-channel feature between channels (Mitrović, 2020). It has been shown that dual-channel features carry more information, reflect changes in brain state, and are more stable than single-channel features. This is advantageous for differentiating between controlled and uncontrolled states in asynchronous motor imagery. Common methods for analyzing dual-channel features include Granger causality, directed transition functions, Pearson correlation coefficients, phase synchronization, etc. Granger causality and directed transition functions reflect directed features, i.e., the features will differ with the order of the two channels. In the eyes of many people, empathy and compassion are similar concepts. They have the same roots and basically the same emotion generation and triggering. However, from a psychological point of view, the two are completely different. Compassion is like a bottle of fermented rice wine. Although it is also produced by the catalyzed fermentation of external interference factors, it does not take a long time. And empathy is like a bottle of high-year aging. It has a long time and a long and long process, and the temperament is completely different. The Pearson correlation coefficient and phase synchronization reflect undirected features, i.e., for a channel pair, the same result is obtained regardless of who comes first and who comes second in the calculation of the features.

Sound is produced and propagated by the vibration of a sound source causing the same rhythmic vibration of the propagating medium, and different frequencies of sound have different senses of vibration. The physical phenomenon of vibration does not occur only in sound phenomena; vibration is a universal phenomenon that occurs in the universe and occurs frequently in the visual matter, and the use of sound vibrations can trigger a description of the form of vibration produced by visible matter (Tahıroğlu et al., 2020). Sound

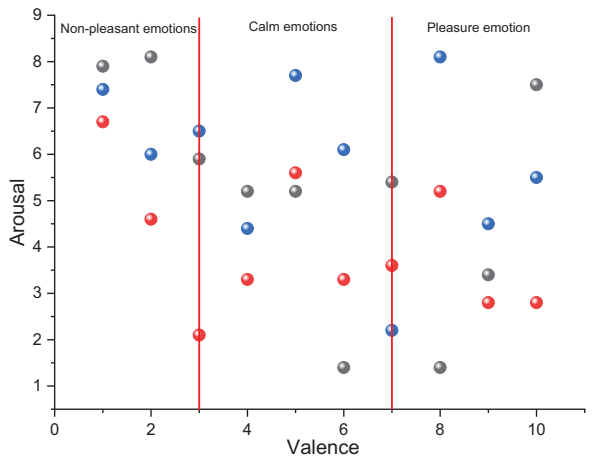


FIGURE 3 | Emotional division map.

waves can be represented numerically and precisely by physical quantities such as frequency and amplitude, and through the perception of the extremely physical properties of the physical sound generation process, in physics sound waves are usually described and calculated mathematically for the commonality of their generation principles, i.e., a curved sonogram describing the periodic changes of sound energy at different moments experiencing different displacements utilizing a mathematical number axis, as shown in **Figure 3**. The sonogram can accurately show the numerical variation of physical quantities of sound waves such as frequency and amplitude with time and is important for analytical work on sound waves.

In the same experiment, the default parameter records may not fully satisfy the user's needs, so the interface provides an interface to enable the user to store their own experiment information record file, which can be saved by clicking on it, or set to be saved automatically when the index is updated, and to complete the update of the index to that file path record at the same time. The stored data can also be analysis result data, each analysis result is marked with a timestamp, so the analysis result data can be directly displayed with the original data time match. The start and end times of data collection are recorded, and the data type of the file is recorded, with the data type taking on different values depending on the content of the experiment and the data stored. Offline analysis may be multiple times, by the index field to mark the number of times this processing results, to correspond to the offline processing of the analysis process records, to facilitate the comparison of experimental results.

According to the storage structure of index and data file separation and the mechanism of splitting the data storage by experiment and date to avoid oversized file storage, the approach used in this thesis achieves more efficient data retrieval. An example of the data retrieval process is shown in the figure below, first retrieve the experiment type and list all index file names of that type according to the experiment type. Then read the index file, the playback sub-interface displays the date with valid data recorded in the current index file, the user selects a date that needs to be played back, after selecting the interface will display the experimental records corresponding to that date, including the valid channel and the start and end time, according to the valid channel of the experimental records to select the channel number of the required playback, according to the listed start and end period to select the time, and finally locate the storage file. After locating the data file, read the data segment offset recorded in the corresponding index file, mark the valid period of the playback progress bar according to the start and end time of different data segments, and then read the data at the file offset corresponding to that time according to the progress bar position, as shown in **Figure 4**.

During the algorithm processing, the complexity of the set algorithm may be too high and the data in the buffer is not read in time. The mechanism of the buffer is that if the old data is not read in time for the same channel and the same frame number, it will be overwritten by the new data, and when the overwriting is done, the flag amount of data will be checked, and if it has not been read, it will be counted, and when this count exceeds a certain threshold (Calderón-Garrido et al., 2020). Music characteristic signal contains many mutations, but not all mutations can be regarded as characteristic event points. Because we adopt ERP technology, the characteristic event points should satisfy both the characteristics of mutation points and ERP events at the same time. The algorithm processing complexity is too high, resulting in data loss. The system parameter configuration includes local parameter configuration, acquisition system, and stimulation system parameter configuration. Since there is no hardware available for testing the stimulation system, only the first two are tested, which correspond to the local parameter column and acquisition system parameter column, respectively. After modifying the parameters in the corresponding parameter column, click the "Configure" button to issue the configuration command, the command successfully configured and prompted in the status bar; when the configuration command is issued in the case of network disconnection caused by disconnecting the network cable, the interface prompts "Configuration failed." The interface prompts the normal connection logic and verifies the correctness of the parameter configuration function.

ANALYSIS OF RESULTS

Multi-Brain Synergistic Brain-Computer Interface EEG Neurofeedback Results

A comparison of multiple feature combinations and learner combinations was attempted after completing the preprocessing and sentiment label modeling. Firstly, since the effect of the experimental subjects on the classification situation was not determined, the subjects with similar sentiment patterns were first determined by looking at the correlation coefficient matrix of the sentiment labels of the experimental subjects (32 people in total). As shown in **Figure 5**, a correlation coefficient matrix analysis was done for the 5 categorical labels of the 32 individuals in the Valence effective direction. Compared with the single-channel feature, the dual-channel feature carries more information, can better reflect the state of the brain,

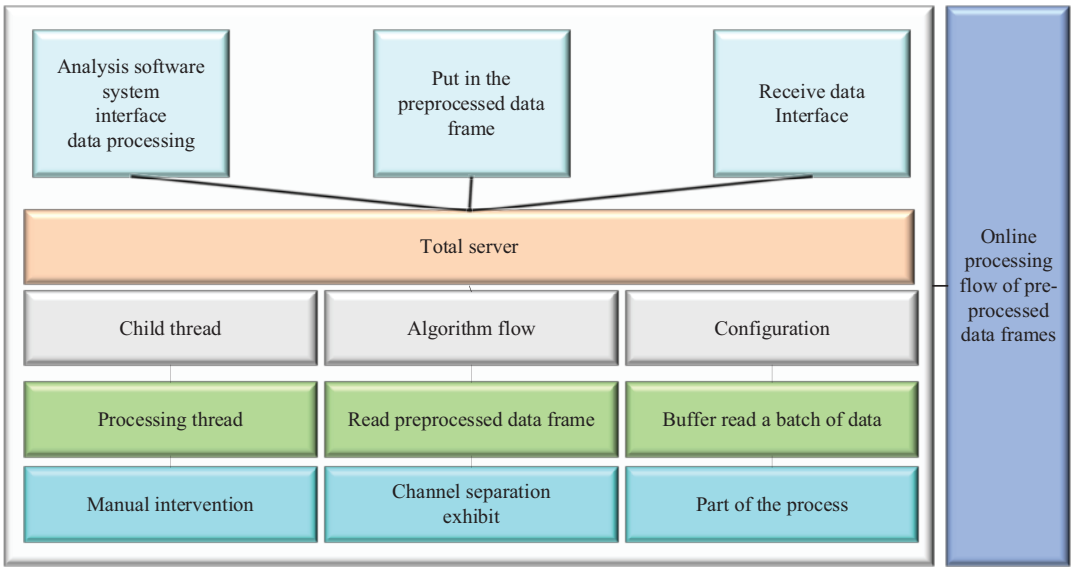


FIGURE 4 | Online processing flow of pre-processed data frames.

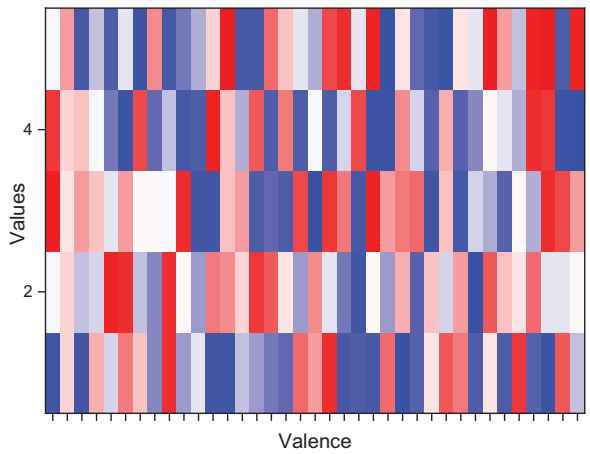


FIGURE 5 | Object valence sentiment correlation coefficient.

and has stronger stability. A dark red correlation coefficient of 1 represents a perfect correlation, while dark blue represents no linear relationship between the two. There are few red components in the whole graph (i.e., each person's emotional perception of the same music varies widely) in which the 18 subjects with the closest emotional experience were selected 2, 3, 4, 8, 9, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 24, 29, and 32 to verify by testing all subjects and selecting subjects for the validity of the EEG emotion classification.

Compared to the pre-trained model on the DEAP training set itself, there is a decrease in the self-provided dataset. Compared to the model under PNN simultaneous training, the model under classification with SVM has average decrease incorrectness of 29.29% with Liking, which has always been accurate in classification, decreasing significantly. The overall correctness rate in the self-collected dataset is relatively lower. Of note is that Dominance has a significantly lower correct rate than several other labeled classifications. This may be partly due to the lack of a more uniform understanding of Dominance. On the other hand, compared to DEAP model predictions, the model under training using PNN neural networks may be relatively more consistent with the data distribution under DEAP, with reduced generalizability and the possibility of overfitting. There may be multiple offline analysis, and the number of processing results is marked by the index field to record the analysis process corresponding to offline processing to facilitate the comparison of experimental results. Next is the specific data of the experimental results. While the sample size of the self-collected training set is smaller and less likely to converge, this part of the test may also have differences in EEG equipment and various processing details at the same time. Therefore, attention needs to be paid to the explanation and filling of experimental labels, the debugging and processing of experimental EEG equipment, and the effects on EEG and emotion in different cultural backgrounds, different experiences, etc., as shown in **Figure 6**.

From the experimental results, we can see that the difference between the recognition effect of PSD620 and F108 is small, while the recognition rate of PSD108 obtained by dimensionality reduction of PSD620 is significantly lower than the above two feature sets. This indicates that we have achieved the purpose of dimensionality reduction based on the cognitive laws while ensuring the recognition rate. Subsequently, we performed sentiment recognition on dataset 2 using F108. Dataset 2 had 32 subjects watching 40 music videos, so there were 1,280 samples, and the same treatment was done for Dataset 1. A total of 15 electrodes in the frontal and central regions were selected for the DEAP database, with 6 sets of electrodes being paired electrodes, so the feature dimension of the DEAP database was again 108-dimensional features.

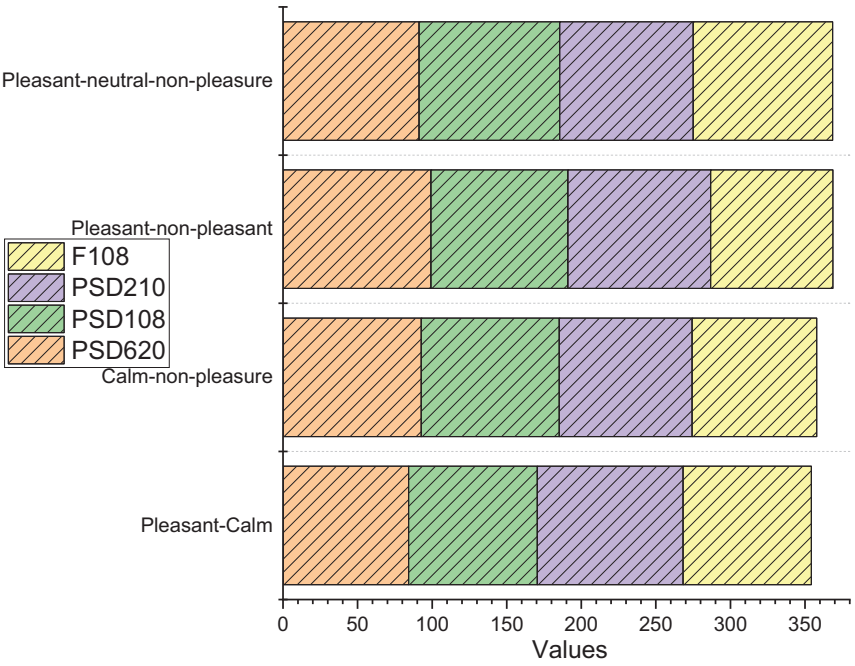


FIGURE 6 | Classification results.

The main reasons for the low recognition rate obtained in this experiment are: the features extracted through this experiment are based on music-evoked emotion, and the DEAP database not only has music-evoked emotion factors, but also video-evoked emotion factors, so there are some problems in using the method of this experiment directly; secondly, as discussed before, the DEAP database has two channels of emotional stimuli, and there is a partial problem of mismatching the emotional content of the music videos, thus making the emotions evoked by the subjects not unique and thus resulting in low recognition rates in this paper. Brain cognitive mechanisms were investigated and emotions were classified using information based on the time and frequency domains of EEG signals. Brain activity in different affective states was first analyzed by brain topography to summarize the main active brain regions and frequency bands of emotion. Subsequently, the discovered cognitive laws were used in practical emotion recognition, and feature extraction was filtered and optimized to achieve the purpose of reducing feature dimensionality and improving recognition rate.

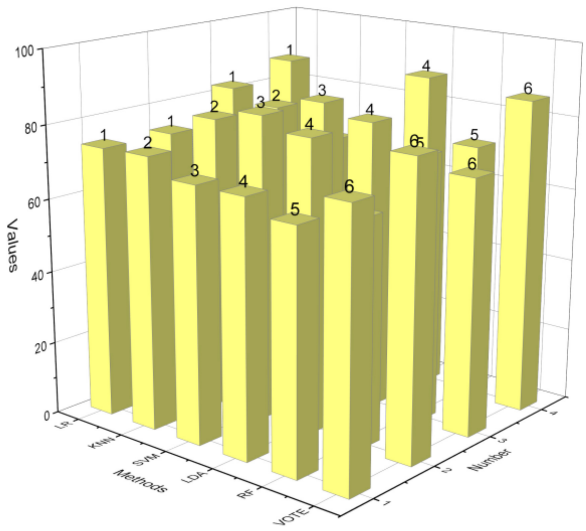


FIGURE 7 | Taxonomy results.

Experimental Results of Emotional Classification of Interactive Electronic Music Pieces

After analyzing the variability of brain load in multiple states in asynchronous motor imagery, we verified the applicability of the brain load metric MWL to asynchronous motor imagery. Next, we investigate the effectiveness of MWL for the classification of asynchronous motor imagery, and we add MWL metrics to energy and PLV features of motor areas to investigate its enhancement of the classification performance of conventional features. We added MWL features to energy features, PLV features, and composite features for classification, and integrated the energy features with MWL features added with PLV features with MWL for learning classification, and used the same two-step classification method and direct triple classification method as in Chapter 3, and the average classification results for the seven subjects are shown in **Figure 7**.

The accuracy rate has dropped by 29.29% on average, and Liking, which has always been classified accurately, has dropped

significantly. In the self-collected data set, the overall accuracy rate is relatively lower. The classification results from the two-step classifier show that for each of the motion area features, the addition of MWL features resulted in an improvement in accuracy on most of the classification algorithms. Among them, the one that improved on all six classifiers was the composite feature, with an improvement of 2.2–4.9%. The classification results from the direct tri-classification method show that for each of the motion area features, the addition of the MWL features resulted in an accuracy improvement on most of the classification algorithms. The highest average improvement for all 6 classifiers is the classification result for the integration learning of energy features with PLV features, with an average improvement of 3.35% and a maximum improvement of 7.3%. The classification result with the highest average improvement of 3.35% and the highest improvement of 7.3% for all 6 classifiers is the classification result of integrated learning of energy features with PLV features. This shows that the brain load metric has a boosting effect on the conventional features of an asynchronous motor imagery classification.

To verify the presence of elevated brain load from the uncontrolled to the controlled state, we investigated the variability of power spectral density in the theta and alpha bands in the controlled and uncontrolled states. From the brain topography, there was indeed a change in power spectral density consistent with elevated brain load. We then verified the significance of the MWL metric between the different states of asynchronous motor imagery, showing that for most subjects, there were significant differences both between the control and non-control states and between imagining the left hand and imagining the right hand and non-control states. Finally, we added MWL metrics to the regular energy and PLV features for classification, and the results showed boosts on most feature-classifier combinations, with the most significant boosts for composite features in the two-step classification method and integrated learning classification in the direct triple classification method. This suggests that brain load, a brain state feature, can improve feature stability in asynchronous motor imagery.

The difference in recognition correctness between the two conditions in the pre-test and post-test experiments shown in **Figure 8**. The difference in online recognition correctness for different data lengths in the pre-test and post-test experiments statistically analyzed using a paired *t*-test. The results show that there is a significant difference in the difference in recognition rate between the two conditions in the pre-test and post-test experiments at data lengths of 0.6, 0.9, and 1 s. From the statistical results, the correct recognition rate of c SSVEP-BCI gradually increases with the increase of data length, but the difference of recognition rate between the two conditions becomes larger with the increase of data length in the pre-test experiment, and the difference of recognition rate between the two conditions stabilizes in the post-test experiment.

The highest recognition rate of the dichotomous task using brain network features for emotion classification reached 73.8%, which is higher than the recognition of emotion recognition based on time-frequency features. This indicates that our proposed dynamic brain network can effectively distinguish

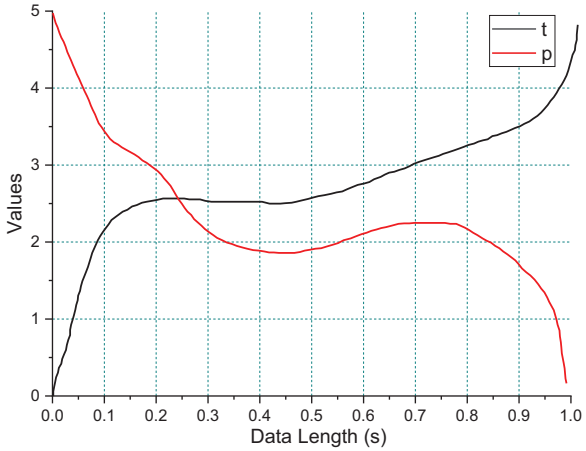


FIGURE 8 | Results of paired *t*-test for difference in online recognition correctness for different data lengths.

between different kinds of emotions Also the results of arousal-based emotion recognition are higher than those of pleasure-based emotion recognition are also consistent with our previous findings that brain network properties are more discriminative for differences in arousal. Among them, 6 groups of electrodes are pairs of electrodes, so the feature dimension of the DEAP database is also 108-dimensional features. From the classifiers, we can see that MLP and SVM perform the best, with near-identical recognition rates for both. This is exactly the opposite of the average clustering coefficient, which illustrates the dynamic balance of the brain. When the average clustering coefficient is elevated, it indicates that there is frequent information exchange between some brain regions and the surrounding brain regions, when decreasing the average path length between brain regions can improve the efficiency of information interaction between brain regions. This, in part, accounts for the self-regulatory nature of the brain. Similarly, LVHA emotions and LVLA emotions, the distinction between the two is more obvious. We perform feature extraction and optimization based on the previous findings; we optimize the traditional power spectral density-based feature extraction by cognitive laws to achieve feature dimensionality reduction while ensuring the recognition rate. We used brain networks as topological features of EEG signals for emotion classification and achieved an emotion recognition rate of 67.3% under four classifications, which exceeds the highest recognition rate available.

CONCLUSION

We analyze the time-frequency characteristics of EEG signals, we construct the corresponding brain topography map by power spectral density and analyze the brain topography map, we find that the main brain regions associated with emotion evocation are central, frontal, and occipital regions and the main emotion-related EEG bands are delta band, theta band, alpha band, and

gamma band. Ten subjects completed the game experiments, and the experimental results showed that all subjects were able to complete the game tasks better and achieve more efficient output. The experimental results showed that all subjects were able to complete the game task better and achieve more efficient output, proving the stability and reliability of the system and the effectiveness of the asynchronous classification algorithm. There seems to be a link between visual and auditory senses, which leads to a common audio-visual association. At this time, reducing the average path length of the brain section can improve the efficiency of the brain section information exchange. This explains the self-regulation of the brain to a certain extent. The phenomenon of audiovisual association provides the basis for the mapping transformation between sound and visual representation in this paper. Based on the analysis of the physiological and psychological causes of audiovisual association, various audiovisual mapping models are proposed in terms of people's perception and aesthetics of sound, respectively. In the same interactive sound visualization work, one or more audiovisual mapping modes can be chosen to be applied. Such audiovisual mapping transformation patterns are based on the audiences' audiovisual association phenomena, which are in line with the audiences' perceptual and aesthetic perceptions of sound, and provide a basis for the subsequent research on visual representation methods.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

ML: kualization and writing methodology, formal analysis, investigation resources, organized the database, and analyzed and interpreted the data.

FUNDING

This work was funded by the Reform of practical teaching system and evaluation mechanism of Music Major, Shanxi Provincial Education Reform Project (No: 2019GJ03).

REFERENCES

Biasutti, M., and Concina, E. (2021). Online composition: strategies and processes during collaborative electroacoustic composition. *Br. J. Music Educ.* 38, 58–73.

Calderón-Garrido, D., Gustems-Carnicer, J., and Carrera, X. (2020). Digital technologies in music subjects on primary teacher training degrees in Spain: teachers' habits and profiles. *Int. J. Music Educ.* 38, 613–624. doi: 10.1177/0255761420954303

D'Agostino, M. E. (2020). Reclaiming and preserving traditional music: aesthetics, ethics and technology. *Organ. Sound* 25, 106–115. doi: 10.1017/s1355771819000505

Dahan, K. (2020). A temporal framework for electroacoustic music exploration. *Organ. Sound* 25, 248–258. doi: 10.1017/s1355771820000151

Farrell, H. (2021). Considering dissemination. *Irish J. Technol. Enhanced Learn.* 6, 14–21.

Görgün, İ (2020). Exploring temporality in Horacio Vaggione's compositional thought. *Organ. Sound* 25, 168–178. doi: 10.1017/s1355771820000072

Iverson, J. (2019). *Electronic Inspirations: Technologies of the Cold War Musical Avant-Garde*. New York, NY: Oxford University Press.

Kerékfy, M. (2020). "Zero point" the beginnings of György Ligeti's Western career. *Stud. Musicol.* 60, 103–114. doi: 10.1556/6.2019.00005

Lakatos, M. (2020). Sights and sounds of big data: Ryoji Ikeda's immersive installations. *Acta Univ. Sapientiae Film Media Stud.* 18, 109–129. doi: 10.2478/ausfm-2020-0006

Malecki, P., Piotrowska, M., Sochaczewska, K., and Piotrowski, S. (2020). Electronic music production in ambisonics-case study. *J. Audio Eng. Soc.* 68, 87–94. doi: 10.17743/jaes.2019.0048

Mitrović, M. (2020). Transmedial (R) Evolution:{SinOsc. ar (400, 800, 0, 0.1) multisensory. experience}. play. *INSAM J. Contemp. Music Art Technol.* 1, 60–76.

Redhead, L. (2021). *Makis Solomos, From Music to Sound: The Emergence of Sound in 20th-and 21st-Century Music*. Abingdon: Routledge.

Romão, J. (2020). Volker Müller & Co.: electronic music and sound engineering at the WDR. *Contemp. Music Rev.* 39, 648–662.

Rossetti, D., Antunes, M., and Manzolli, J. (2020). Compositional procedures in electronic music and the emergence of time continuum. *Organ. Sound* 25, 156–167.

Simurra, I., and Borges, R. (2021). Analysis of Ligeti's atmosphères by means of computational and symbolic resources. *Rev. Música* 21, 369–394.

Smith, B. E., Pacheco, M. B., and Khorosheva, M. (2021). Emergent bilingual students and digital multimodal composition: a systematic review of research in secondary classrooms. *Read. Res. Q.* 56, 33–52. doi: 10.1002/rrq.298

Sofer, D. (2020). Categorising electronic music. *Contemp. Music Rev.* 39, 231–251.

Tahıroğlu, K., Magnusson, T., Parkinson, A., Garrelfs, I., and Tanaka, A. (2020). Digital musical instruments as probes: how computation changes the mode-of-being of musical instruments. *Organ. Sound* 25, 64–74. doi: 10.1017/s1355771819000475

Turchet, L., West, T., and Wanderley, M. M. (2021). Touching the audience: musical haptic wearables for augmented and participatory live music performances. *Pers. Ubiquitous Comput.* 25, 749–769.

Wolffenbüttel, D. C. R., and Pellin, H. (2020). Sound research and composition with everyday sounds. *Int. J. Dev. Res.* 10, 34464–34471.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 25 January 2022
doi: 10.3389/fpsyg.2021.799926

Position-Enhanced Multi-Head Self-Attention Based Bidirectional Gated Recurrent Unit for Aspect-Level Sentiment Classification

Xianyong Li[1]*, *Li Ding*[1], *Yajun Du*[1], *Yongquan Fan*[1] and *Fashan Shen*[2]

[1] School of Computer and Software Engineering, Xihua University, Chengdu, China, [2] Sichuan Suitang Science and Technology Co., Ltd., Chengdu, China

Aspect-level sentiment classification (ASC) is an interesting and challenging research task to identify the sentiment polarities of aspect words in sentences. Previous attention-based methods rarely consider the position information of aspect and contextual words. For an aspect word in a sentence, its adjacent words should be given more attention than the long distant words. Based on this consideration, this article designs a position influence vector to represent the position information between an aspect word and the context. By combining the position influence vector, multi-head self-attention mechanism and bidirectional gated recurrent unit (BiGRU), a position-enhanced multi-head self-attention network based BiGRU (PMHSAT-BiGRU) model is proposed. To verify the effectiveness of the proposed model, this article makes a large number of experiments on SemEval2014 restaurant, SemEval2014 laptop, SemEval2015 restaurant, and SemEval2016 restaurant data sets. The experiment results show that the performance of the proposed PMHSAT-BiGRU model is obviously better than the baselines. Specially, compared with the original LSTM model, the Accuracy values of the proposed PMHSAT-BiGRU model on the four data sets are improved by 5.72, 6.06, 4.52, and 3.15%, respectively.

Keywords: aspect-level sentiment classification, attention mechanism, position information, BiGRU, long short term memory networks, aspect terms

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
JiaQiang Wan,
Chongqing University of Technology, China
Xulong Zhang,
Shanxi Datong University, China
Chunming Zhang,
Guangdong Medical University, China

***Correspondence:**
Xianyong Li
lixy@mail.xhu.edu.cn;
xian-yong@163.com

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *22 October 2021*
Accepted: *10 November 2021*
Published: *25 January 2022*

Citation:
Li X, Ding L, Du Y, Fan Y and Shen F (2022) Position-Enhanced Multi-Head Self-Attention Based Bidirectional Gated Recurrent Unit for Aspect-Level Sentiment Classification. Front. Psychol. 12:799926. doi: 10.3389/fpsyg.2021.799926

1. INTRODUCTION

In natural language processing (NLP), the purpose of sentiment analysis (Pang and Lee, 2008) is to divide the texts into two or more sentiment categories (such as positive, neutral, and negative) based on the meaningful information from some texts. The aspect-level sentiment classification (ASC) is an important fine-grained sentiment classification. Its aim is to predict sentiment polarities of different aspect terms in a sentence (Thet et al., 2010). For example, in the sentence: "The environment of this restaurant is beautiful and the food is delicious, but the service is terrible," the sentiment polarities of the aspect terms "environment," "food," and "service" are positive, positive, and negative, respectively. Since the traditional sentiment analysis only consider the polarities of sentiment for sentences (Mullen and Collier, 2004), the ASC is more complicated than traditional sentiment classification.

In machine learning models, a series of features, e.g., a set of words and sentiment dictionaries (Jiang et al., 2011; Zhang and Lan, 2015), were set up to train classifiers, such as SVM and KNN. Their classification effect heavily depended on the features' quality. Another more important models are deep learning models (Zhang et al., 2018). Because they did not deliberately design feature engineering, they can be effectively applied to automatically achieve the task of the ASC (Tang et al., 2016b). In recent, the recurrent neural network (RNN) (Socher et al., 2011; Nguyen and Shirai, 2015; Liu et al., 2016) and its variant models have been widely used in ASC tasks. These models can capture the relationships between sequences. Lai et al. (2015) used a two-way loop structure to obtain text information. Compared with traditional window-based neural networks, their method reduced more noise. Their method also retained the word order in a large range when it learned text expressions. For targeted sentiment classification, Gan et al. (2020) put forward a sparse attention mechanism based on a separable dilated convolution network. Their method is superior to the existing methods. Tang et al. (2016a) proposed a target-dependent long-term short-term memory network (TD-LSTM). This network is modeled by the contexts before and after the target word. By combining the information of the two LSTM hidden layer states, they further achieved the ASC tasks. Compared with the RNN model, the performances of these RNN variant models have small improvements on the ASC task.

For specific aspect terms in a sentence, the RNN model paid little attention to its contextual information. Based on visual attention (Mnih et al., 2014), the attention mechanism is extensively borrowed in neural networks (Luong et al., 2015; Yin et al., 2015; Liu and Lane, 2016). A lot of attention-based neural network models (Yin et al., 2015; Wang et al., 2016; Ma et al., 2017; Zeng et al., 2019) are proposed to solve ASC tasks. For a sentence, the attention mechanism makes the neural network model pay more attention to the sentiment descriptions of specific aspects, i.e., the sentiment polarities of aspect words, while ignoring other noise words that are not related to the aspect words. Xu et al. (2020) proposed a multi-attention network. They used the global and local attention modules to obtain the interactive information of different granularities between aspect words and contexts. Chen et al. (2017a) proposed a recurrent attention network model on memory for sentiment classification. Their model is established on cognition grounded data. The proposed cognition-based attention mechanism can be applied in sentence-level and document-level sentiment analysis. Based on the attention mechanism and LSTM networks, Ma et al. (2017) proposed an interactive attention network (IAN) model. Their model obtained good performance on SemEval 2014. When the aspect terms contain more than one word, their method may lead to the loss of useful information. The self-attention mechanism (Letarte et al., 2018) could make sentiment analysis models pay more attention to the useful information of aspect terms in the context and the internal structure of sentences. It improved the performance of neural network models. Xiao et al. (2020) used multi-head self-attention to get the semantic and interactive information in sentences. They further proposed a multi-head self-attention based gated graph convolutional network model. Their model can effectively achieve aspect-based sentiment classification. Leng et al. (2021) modified the transformer encoder to propose the enhanced multi-head self-attention. Through this attention, the inter-sentence information can be encoded. Combining with the enhanced multi-head self-attention and BiLSTM or BiGRU, they proposed a sentiment analysis model which performed better than some baselines in some evaluation indices. Therefore, the attention mechanism is becoming more and more important in the ASC task.

In addition, the position information between the aspect terms and their contexts has been confirmed that it was capable of improving the accuracy of the ASC (Chen et al., 2017a; Gu et al., 2018). For the RNN model (Liu and Lane, 2016; Liu et al., 2016), the calculation at the current moment depends on the result at the previous moment. This will result in a lack of contextual semantic information for aspect words. Zhou et al. (2019) used R-Transformer to get this semantic information. They further combined the self-attention mechanism and position relationship to propose the position and self-attention mechanism-based R-Transformer network (PSRTN) model for the ASC. Their experiment results are better than some baseline models. It is, thus, clear that the position information needs to consider in the context attention calculation.

Based on the above observations, this article proposes a position-enhanced multi-head self-attention based BiGRU (PMHSAT-BiGRU) model which integrates the position influence vector, multi-head self-attention mechanism, and bidirectional gated recurrent unit (BiGRU). This model considers three influence factors for the ASC task: the keywords in aspect terms, the position relationship between aspect terms and context, and semantic information of the context. In order to avoid noise words and make better use of the keywords in the aspect, it uses a self-attention mechanism to calculate the attention scores of the aspect words and each word in the sentence. To better obtain the semantic information of the context, it also uses multi-head attention to learn the relevant information from different representation subspaces. Finally, the PMHSAT-BiGRU model will be evaluated on the SemEval2014 restaurant, SemEval2014 laptop, SemEval2015 restaurant, and SemEval2016 restaurant dataset. Abundant experiments will verify its effectiveness on the ASC task.

In general, the main contributions of this article are as follows:

(1) Based on the position information between the aspect terms and context, a positional information vector is designed. It uses the relative position method to participate in the calculation of the attention weight.
(2) To get a contextual representation of the specific aspect terms, a self-attention mechanism is used to calculate the words' weights in aspect terms. The multi-head attention mechanism is employed to represent the semantic information of the context in different representation subspaces.
(3) A PMHSAT-BiGRU model is proposed. Considering that three main factors, including the keywords in aspect terms, the position relationship between aspect terms and

context, and the semantic information of the context for a sentence, affect the ASC, the PMHSAT-BiGRU model integrates the position influence vector, multi-head self-attention mechanism, and BiGRU.

(4) Extensive experiments on four datasets including SemEval2014 restaurant, SemEval2014 laptop, SemEval2015 restaurant, and SemEval2016 restaurant data sets are conducted. The performance of the PMHSAT-BiGRU model is evaluated by using the Accuracy (Acc) and Macro-Average F1 (Macro-F1).

The rest of this article is organized as follows. Section 2 introduces the related work of the ASC. Section 3 elaborates the proposed PMHSAT-BiGRU model. In section 4, we carry out a large number of experiments to prove the validity of the proposed model. Finally, we make the summary and forecast to the full text in section 5.

2. RELATED WORK

The ASC focuses on the sentiment polarities of aspect terms in a sentence. Since neural network models (Santos and Gattit, 2014; Zhang et al., 2018; Chen and Huang, 2019) are superior to the machine learning methods (Mullen and Collier, 2004; Jiang et al., 2011; Zhang and Lan, 2015) in sentiment classification, many new research results are based on neural networks. On the basis of the RNN (Mikolov et al., 2010; Akhtar et al., 2020), Hochreiter et al. explored the long short-term memory network (LSTM) (Hochreiter and Schmidhuber, 1997) and the gated recurrent unit (GRU) (Dey and Salemt, 2017). These models could solve the gradient descent and explosion problems. Tang et al. (2016a) integrated the information of the target words and context words to establish the sentence semantically. They presented two improved LSTM models, i.e., the target-dependent LSTM and target-connection LSTM. These models are significantly superior to the original LSTM model. Jiang et al. (2011) took the content, sentiment lexicon and context into consideration to improve the target-dependent sentiment classification for Twitter. Tan et al. (2020) proposed an aligning aspect embedding method to train aspect embeddings for the ASC. The embeddings are applied to the gated convolutional neural networks (CNNs) and attention-based LSTM. Their experiment results showed that the model with the aspect embedding obtained better performance than other baseline models. Xue and Li (2018) proposed Gated Tanh-Rectified Linear Unit (ReLU) Units. They further built a new CNN model with this mechanism to predict the sentiment polarities of aspect terms. The training time of the model was faster than other baseline models.

The attention mechanism and position information are also considered in different neural network models for the ASC. Wang et al. (2016) designed a novel attention mechanism to capture the vital part of sentences with different aspect terms. Based on this mechanism, they presented an ATAE-LSTM model to effectively achieve the binary and 3-class prediction problems in the ASC. Considering the explicit memory, position, and context attentions, Tang et al. (2016b) designed deep memory networks. To a certain extent, their models achieved good performance on the ASC tasks. Liu et al. (2015); Chen et al. (2017b) introduced position information into attention mechanism to handle tasks of question answering and machine translation. The performance of the two tasks was obviously improved.

Although these models have provided a good performance on the ASC tasks, the neural network models with position relationships and multi-head self-attention mechanism have yet to be studied for the ASC.

3. PMHSAT-BiGRU FOR THE ASC

In this section, we will minutely describe the PMHSAT-BiGRU model (refer to **Figure 1**), including the task definition, position modeling, word representation, BiGRU, attention mechanism, sentiment classification, and model training.

3.1. Task Definition

For a sentence with N-words and M aspect terms, let $\langle aspect\ term\ i, sentence\rangle$ be the aspect-sentence pair for the aspect term $i, i = 1, 2, \cdots, M$. Then, using $\langle aspect\ term\ i,\ sentence\rangle$ as an input of the ASC, the sentiment category $\langle positive, neural, negative\rangle$ will be predicted for the aspect term i in the sentence. For example, the sentence "Great food but the service was dreadful!" involves two aspect terms, namely [food] and [service]. The sentence will generate two aspect-sentence pairs including $\langle food, sentence\rangle$ and $\langle service, sentence\rangle$ as the inputs of the ASC, then expectation outputs of the aspect terms [food] and [service] are positive and negative, respectively.

3.2. Position Modeling

In the ASC task, the sentiment polarity of a particular aspect will be severely affected by adjacent context words in a sentence. Inspired by Shaw et al. (2018), we employ relative position to model the position information of the aspect words in the corresponding sentence. For a sentence with aspect terms, the position indices of the words contained in an aspect term are marked as "0," and the position indices of other words will be expressed as the relative distances from the current aspect term. Therefore, the position index of a word for the sentence is the following:

$$p_i = \begin{cases} |i - a_{start}|, & i < a_{start}, \\ 0, & a_{start} \le i \le a_{end}, \\ |i - a_{end}|, & i < a_{end}, \end{cases} \tag{1}$$

where a_{start} and a_{end} respectively represent the start and end indices of the aspect term; and p_i represents the relative distance from the ith word to the aspect term in the sentence. According to these indices from the first word to the last word in the sentence, a position index sequence with the sentence length of N is $p = [p_1, p_2, \cdots, p_N]$ for an aspect term. For example, in the sentence "The seafood menu is interesting and quite reasonably priced.", there are two aspect terms "seafood menu" and "priced." Then the position index sequences of "seafood menu" and "priced" are expressed as p = [1,0,0,1,2,3,4,5,6] and

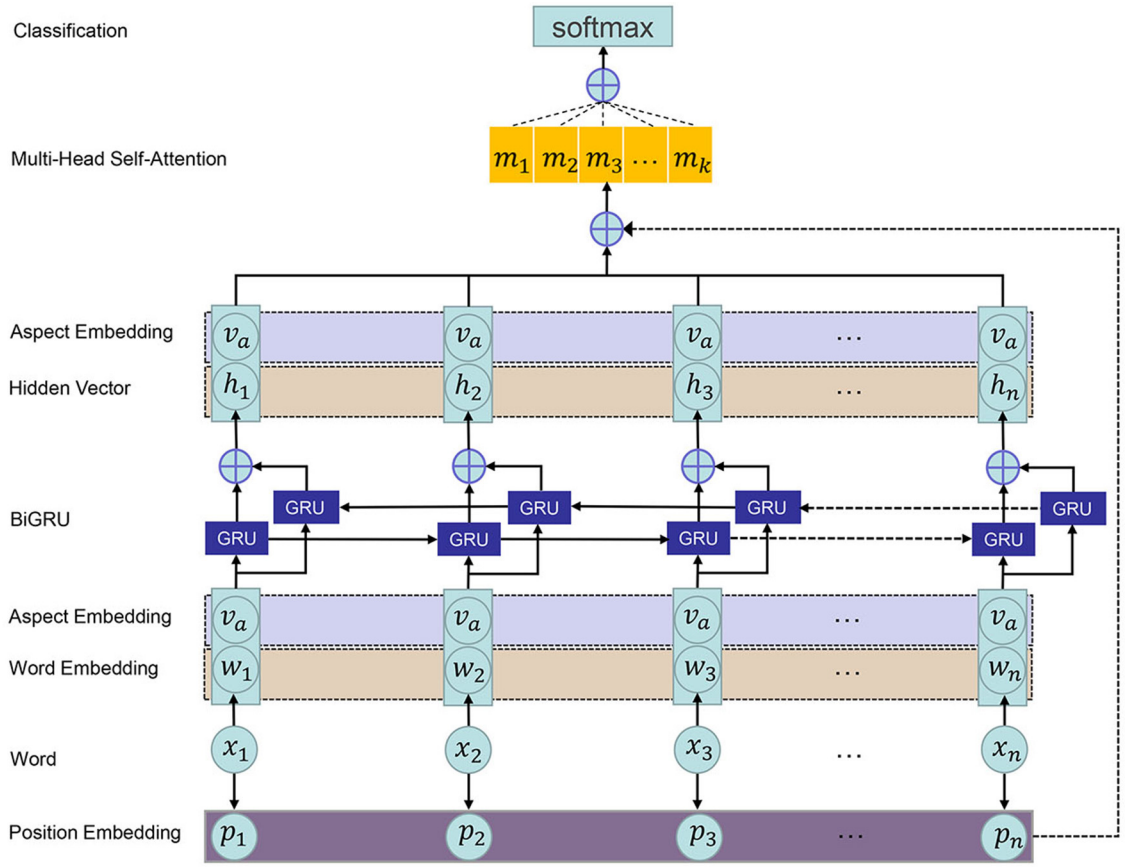


FIGURE 1 | The architecture of the PMHSAT-BiGRU for ASC.

p = [8,7,6,5,4,3,2,1,0], respectively. By looking up the position embedding matrix $P \in \mathbb{R}^{d_p \times N}$, the corresponding position embeddings are obtained, where d_p is the dimension of the position embedding, and N is the length of the sentence. Then, the position embeddings are randomly initialized and updated during the training process. After transforming the position indices into position embeddings, the embeddings can model the different weights of words with different distances. In the example above, the sentiment word "interesting" is more important than the words "quite reasonably" for the aspect term "seafood menu." It implies that when an aspect term needs to predict its sentiment polarity, the words with relatively small distances and sentiment polarities are more important than other words.

3.3. Word Representation

By word embedding technology, each word is embedded into a unique word vector with the information of the word itself in the vector space. To obtain the word embedding, we will apply Glove (Pennington et al., 2014) pre-trained at Stanford University. In the following, all word embeddings are denoted by $E \in \mathbb{R}^{d_w \times |V|}$, where d_w represents the dimension of the word embeddings and $|V|$ represents the size of the vocabulary. All aspect embeddings are expressed as $A \in \mathbb{R}^{d_a \times |L|}$, where d_a is the dimension of aspect embeddings, and $|L|$ is the size of aspect terms. For a sentence with N-words $[w_1, w_2, \cdots, w_N]$, if it contains an aspect term $[a_1, a_2, ..., a_M]$ with M words, then the sentence embedding and the aspect embedding will be obtained by finding the embedding matrix E and A, respectively.

3.4. Bidirectional Gated Recurrent Unit

Recurrent neural network has been successfully applied in the field of the NLP. However, the standard RNN often faces the problem of gradient disappearance or gradient explosion. As a special RNN, LSTM adjusts the cell state through three gated mechanisms at each time step, better solving the problem of the long dependence. Compared with the one-way LSTM, BiLSTM can learn more contextual information. It establishes the context dependence in the forward and reverse directions. Concretely, the forward LSTM processes sentences from the left to the right, and the reverse LSTM processes sentences from the right to the left. From this, it gets two hidden representations, and then connects the forward hidden state and backward hidden state of each word as the final representation.

In contrast with the LSTM, the GRU, which uses two gated mechanisms to adjust cell state and has fewer parameters

and lower computational complexity has relatively better performance than LSTM in the NLP. Specifically, at time t, we obtain the embedding vector $w_t \in \mathbb{R}^{d_w}$ of the current input word from E and the aspect embedding vector $v_a \in \mathbb{R}^{d_a}$ from A, then the current hidden layer vector h_t in GRU is updated by the following:

$$\begin{aligned} z_t &= \sigma\left(W_z h_{t-1} + U_z[w_t, v_a] + b_z\right), \\ r_t &= \sigma\left(W_r h_{t-1} + U_r[w_t, v_a] + b_r\right), \\ \widetilde{h_t} &= \tanh\left(W_h\left(h_{t-1} \odot r_t\right) + U_h[w_t, v_a] + b\right), \\ h_t &= h_{t-1} \odot (1 - z_t) + z_t \odot \widetilde{h_t}, \end{aligned} \tag{2}$$

where z and r are the update gate and reset gate, respectively; the sigmoid function $\sigma(\cdot)$ is used to control the retention of useful information and the discarding of the useless information; $W_z, W_r, W_h \in \mathbb{R}^{d_h \times d_h}, U_z, U_r, U_h \in \mathbb{R}^{d_h \times (d_w + d_a)}, b_z, b_r, b \in \mathbb{R}^{d_h}$ represent the weight matrices and biases learned in the GRU training process; $\odot$ denotes an element multiplication; and $[w_t, v_a]$ stands for the splicing vector of the word embedding w_t and the aspect embedding v_a. Then, the hidden vector $[h_1, h_2, ..., h_N]$ of the sentence with the length N is regarded as the final context word representation.

In the following, we will adopt the BiGRU to obtain the contextual representation of a sentence. Compared with the one-way GRU, BiGRU includes the forward hidden state $\overrightarrow{h_i^t} \in \mathbb{R}^{d_h}$ and the backward hidden state $\overleftarrow{h_i^t} \in \mathbb{R}^{d_h}$ at time t, where d_h represents the number of hidden layer units. Then, the forward hidden state $\overrightarrow{h_i^t}$ and the backward hidden state $\overleftarrow{h_i^t}$ are connected as the final context hidden representation $h_i^t = [\overrightarrow{h_i^t}; \overleftarrow{h_i^t}] \in \mathbb{R}^{2d_h}$.

3.5. Attention Mechanism

The attention mechanism can help the model focus on the important parts of a sentence in the ASC tasks. In particular, the multi-head attention mechanism allows the model to learn some relevant information in different representation subspaces. Furthermore, the self-attention mechanism can learn the word dependency relationships within the sentence and then capture the internal structure of the sentence. This mechanism can process in parallel, reducing the complexity of calculations. In view of these advantages, the overall semantics of a sentence can be represented by the multi-head self-attention mechanism (Zhou et al., 2019). Based on the last hidden layer state h_i^t output by BiGRU, the current context representation can be represented as $h_1^t, h_2^t, \cdots, h_N^t$. Then, feeding them into the multi-head self-attention, a new representation s_t for the sentence can be obtained by the following:

$$\begin{aligned} s_t &= MultiHeadAttention\left(h_1^t, h_2^t, \cdots, h_N^t\right) \\ &= Concat\left(head_1\left(h_N^t\right), head_2\left(h_N^t\right), \cdots, head_k\left(h_N^t\right)\right) W^o, \end{aligned} \tag{3}$$

where $head_i\left(h_N^t\right)$ denotes the value of the i-th attention head; and W^o stands for the linearization mapping matrix. For $head_i\left(h_N^t\right)(i = 1, 2, \cdots, N)$, it is calculated by the following formulas:

$$\begin{aligned} \alpha_1, \alpha_2, \cdots, \alpha_N &= softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V, \\ head_i\left(h_N^t\right) &= \sum_{j=1}^{N} \alpha_j v_j, \end{aligned} \tag{4}$$

where Q, K, and V represent the query, key, and value matrices, respectively. In these matrices, their vectors q, k_i and v_i are calculated as follows:

$$q, k_i, v_i = W^q h_N^t, W^k h_i^t, W^v h_i^t, \tag{5}$$

where W^q, W^k, and W^v are the weight matrices whose values are different in different attention heads.

3.6. Sentiment Classification

For the multi-head self-attention representation s_t, we map it to the target space with C sentiment polarities by a non-linear layer:

$$x = \tanh\left(W_r s_t + b_r\right), \tag{6}$$

where $x = (x_1, x_2, \cdots, x_C)$, W_r and b_r are the weight matrix and bias within the non-linear layer, respectively. Then, x is transformed into the conditional probability distribution through a Softmax layer. Therefore, the final distributions of the C sentiment polarities are obtained by the following:

$$y_c = \frac{exp\,(x_c)}{\sum_{c=1}^{C} exp\,(x_c)}. \tag{7}$$

From this result, the sentiment polarity corresponding to the maximum probability, i.e., $max_{c=1}^{C}\{y_c\}$, is chosen as the final sentiment classification.

3.7. Model Training

In the PMHSAT-BiGRU model, the cross entropy and L_2 regularization

$$L = -\sum_{d \in D}\sum_{c=1}^{C} y_c\left(d\right) log\left(g_c\left(d\right)\right) + \frac{1}{2}\lambda \parallel \theta \parallel^2, \tag{8}$$

will be regarded as the loss function, where D denotes the data set which consists of different sample d; $y_c\left(d\right) \in \mathbb{R}^C$ represents the real sentiment polarity distribution of sample d; $g_c\left(d\right) \in \mathbb{R}^C$ stands for the sentiment polarity vector of sample d; λ is the L_2 regularization coefficient; and θ includes all model parameters. For the sake of optimizing all model parameters, the loss function should be minimized as much as possible. By the back-propagation method, the parameters θ is updated by the following:

$$\theta = \theta - \lambda_l \frac{\partial L\left(\theta\right)}{\partial \theta}, \tag{9}$$

where λ_l is the learning rate. In order to prevent overfitting during training process, the dropout strategy is adopted as the method of discarding some learned features.

TABLE 1 | Samples of semeval 2014–2016 datasets.

Datasets		Positive	Negative	Neutral
R14	Train	2,164	807	637
	Test	728	196	196
L14	Train	994	870	464
	Test	341	128	169
R15	Train	1,178	382	50
	Test	439	328	35
R16	Train	1,620	709	88
	Test	597	190	38

TABLE 2 | The numbers of terms in statistical datasets.

Datasets		Len = 1	Len = 2	Len ≥ 3
R14	Train	2,720 (75.38%)	604 (16.74%)	284 (7.87%)
	Test	801 (71.52%)	215 (19.20%)	104(9.29%)
L14	Train	1,473 (63.27%)	649 (27.88%)	206 (8.85%)
	Test	351 (52.78%)	209 (31.43%)	78 (11.73%)
R15	Train	1,272 (79.00%)	216 (13.41%)	122 (9.59%)
	Test	638(79.55%)	94(11.72%)	70(8.73%)
R16	Train	1,941 (80.31%)	301 (12.45%)	175 (7.24%)
	Test	668 (80.97%)	101 (12.24%)	56 (6.79%)

4. EXPERIMENTS

In this section, we will make some experiments under the proposed PMHSAT-BiGRU model and several baseline models on several large data sets. By comparing the results of these experiments, the effectiveness of the proposed PMHSAT-BiGRU model will be verified. Then several ablation experiments are set to affirm the effectiveness of the modules in the proposed model. Finally, we visualize the dataset in the experiment based on the proposed PMHSAT-BiGRU model.

4.1. Experimental Setting

4.1.1. Dataset

The ASC benchmark data sets, officially published by SemEval including SemEval 2014 Task4[1], SemEval 2015 Task12[2], and SemEval 2016 Task5[3], will be adopted. In these datasets, the SemEval 2014 contains Restaurant14 (R14) and Laptop14 (L14) datasets; the SemEval 2015 uses Restaurant15 (R15) dataset; and the SemEval 2016 uses the Restaurant16 (R16) dataset. More specifically, each dataset contains a training set and a test set. In each dataset, every data is a single sentence, including the review text, aspect terms, sentiment labels corresponding to the aspect terms, and the starting position of the aspect terms. There are four aspect-level sentiment polarities, i.e., positive, negative, neutral, and conflict in these data sets. To facilitate subsequent experiments, we only use the positive, negative, and neutral aspect-level sentiment polarities and remove the conflict aspect-level sentiment polarity from these data sets, i.e, the number of the sentiment polarity categories $C = 3$. For all adopted datasets, their details of the training sets and test sets are shown in **Table 1**. In addition, we count the number of words in the aspect terms in **Table 2**. It easily finds that more than one-fourth of the datasets have the aspect terms with multiple words.

4.1.2. Parameters Setting

In our experiments, the Glove[4] (Pennington et al., 2014) is used to initialize the aspect and contextual word embedding. It sets the embedding dimension of each word to 300. The weight matrices are initialized from the uniform distribution $U(-\mu, \mu)$, where $\mu = 0.01$ and all offsets are set to 0. The aspect embedded dimension is also set to 300; the BiGRU hidden unit is set to 200; and the position embedded dimension is set to 100. The maximum length of a sentence is 80; the batch size is 4; and the number of multi-head self-attention heads is 8. In our PMHSAT-BiGRU model, the dropout rate is set as 0.5; the L_2 regularization coefficient is set as 1e-5; the Adam optimizer is used to optimize the training parameters; and the learning rate is set as 1e-4. To implement our PMHSAT-BiGRU model, we employ Pytorch[5] in the experiments.

4.1.3. Evaluation

In the experiments, we used two common evaluation indexes, i.e., Acc and Macro-F1 in classification tasks. In detail, the Acc represents the proportion of correctly classified samples to the total sample number, and its calculation is as follows:

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn}, \tag{10}$$

where tp denotes the number of the samples whose true labels and sentiment labels predicted by the model are both positive categories; and tn represents the number of the samples whose true labels are positive categories and sentiment labels predicted by the model are negative categories. Correspondingly, fp represents the number of the samples whose true labels are negative categories and sentiment labels predicted by the model are positive categories; and fn represents the number of the samples whose true labels and sentiment labels predicted by the model are both negative categories.

Next, the Recall, Precision, and F1-score (RPF value) are calculated by the following.

$$\begin{aligned} Recall &= \frac{tp}{tp + fn}, \\ Precision &= \frac{tp}{tp + fp}, \\ F1 - score &= \frac{2 \times Precision \times Recall}{Precision + Recall}. \end{aligned} \tag{11}$$

In the experiments, we will calculate the RPF values for the positive, negative, and neutral categories. Then, we obtain

[1]https://alt.qcri.org/semeval2014/task4/
[2]https://alt.qcri.org/semeval2015/task12/
[3]https://alt.qcri.org/semeval2016/task5/
[4]http://nlp.stanford.edu/projects/glove/
[5]https://pytorch.org/

the Macro-F1 values by averaging the F1-score values of the three categories.

4.2. Baselines

In order to verify the effectiveness of the PMHSAT-BiGRU model, the experiment results will compare with the following baseline models:

Context word vectors average (ContextAvg): It averages the word embedding and aspect vectors, and then input the result into the softmax classifier, which was cited as the baseline model in Tang et al. (2016b).

Long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997): For a sentence, the one-way LSTM network is used to model the sentence; the last hidden layer vector is regarded as the final representation of the sentence, and then sent to the Softmax classifier for the final classification.

Target-dependent long-term short-term memory (TD-LSTM) (Tang et al., 2016a): For a sentence with target words, the sentence is divided into two different parts based on a target word of the sentence, and then respectively uses two LSTMs to model the context on the left side of the target word and the right side of the target word. Finally, it connects the related representations of the two parts as the classifier input to predict the sentiment polarity of the target word.

Target-connection long-term short-term memory (TC-LSTM) (Tang et al., 2016a): This model is similar to TD-LSTM. However, the difference is that TC-LSTM has added the aspect word information at the input; and the word vector and the aspect vector are connected, obviously integrating the correlation information between the aspect word and the context word.

Attention-based long short-term memory (AE-LSTM) (Wang et al., 2016): Based on the standard LSTM, the aspect embeddings are designed to represent the aspect information; and the aspect embeddings are regarded as a part of the training parameters.

Attention-based long short-term memory with aspect embedding (ATAE-LSTM): On the basis of AE-LSTM, the aspect is embedded in each word embedding and hidden vector; and the attention mechanism is used to further strengthen the effect of the aspect embedding. This model was cited as the baseline model in Zhou et al. (2020).

Memory Network (MemNet): Using the deep memory network instead of the RNN-based method for sentence modeling, it repeatedly employs the attention mechanism to capture the connections between the context words and aspect words. This model was cited as the baseline model in Zhou et al. (2020).

Interactive attention network (IAN) (Ma et al., 2017): Two LSTMs are respectively used to model the aspect terms and context words. Through the interactive attentions from the sentences to their corresponding aspects and from the aspects to the sentences, the sentence representations and aspect representations are generated. Then, the two representations are connected to input the Softmax classifier for the classification.

Gated convolutional network with aspect embedding (GCAE) (Xue and Li, 2018): Many pairs of convolution kernels are used to extract local N-gram features, where each pair of convolution kernels contains one aspect-independent convolution kernel and one aspect-dependent convolution kernel. Then, the model respectively adopts tanh and ReLU gated units to output the sentiment features of a given aspect.

Attention-based long short-term memory with position context (PosATT-LSTM) (Zeng et al., 2019): On the basis of the one-way LSTM, the position relationships between the aspect words and the context are considered. The relationships are applied to the calculations of the attention weights.

4.3. Compared Methods

Based on the proposed PMHSAT-BiGRU model and baseline models, some experiments on R14, L14, R15, and R16 are made. These models' Acc and Macro-F1 values are shown in **Table 3**.

From **Table 3**, compared with other models, the performance of the ContextAvg model is the worst because it is directly classified by average word embedding and aspect embedding. Among the sequential models, the performance of the LSTM model is the worst because the model does not consider the attention mechanism and aspect word information but equally treats the aspect words and other words in the model. Compared with the LSTM model, the aspects are embedded into the LSTM model for training in the AE-LSTM model. So, the Accuracy values under the AE-LSTM model are respectively 1.32, 1.75, 0.79, and 2.17% better than the values under the LSTM model on R14, L14, R15, and R16. Compared with the TD-LSTM model, although the TC-LSTM model considers the aspect word information at the input end, its performance is worse than the TD-LSTM model.

Because the attention mechanism is used to model the relationships between the aspect words and context under the Memnet model, the performance of the Memnet model is better than the AE-LSTM model. Compared to the Memnet and ATAE-LSTM model, the Accuracy and Macro-F1 values on the Memnet model are slightly higher than the values on the ATAE-LSTM model under partial datasets. The performance of the IAN model is better than the ATAE-LSTM model because, in the IAN model, two LSTMs are respectively adopted to model the aspect terms and context, and an interactive attention mechanism is used to obtain context related to aspect terms. However, the importance of the position relationships between aspect words and context is not considered in the IAN model. Therefore, the performance of the IAN model is worse than the PosATT-LSTM model. Because the GCAE model uses the CNN and gated mechanism to realize parallel computation, making the model insensitive to position information, its performance on R16 has little promotion compared with the IAN model.

For the PMHSAT-BiGRU model, the aspect embedding information and the importance of the aspect words and context position information are applied in the calculations of the attention weights. Meanwhile, the multi-head attention mechanism is used to learn the dependent information in different contexts. The self-attention mechanism is also employed to capture the important words in the aspect terms. For the PosATT-LSTM model, only the semantic representation in a single context is captured; and each word in aspect terms

TABLE 3 | The results of the PMHSAT-BiGRU model and baseline models in Semeval2014–2016.

	R14		L14		R15		R16	
	Acc	**Macro-F1**	**Acc**	**Macro-F1**	**Acc**	**Macro-F1**	**Acc**	**Macro-F1**
ContextAvg	71.53	58.02	61.59	53.92	73.79	47.43	79.87	55.68
LSTM	74.80	59.08	67.08	60.53	75.15	51.27	80.09	55.09
TD-LSTM	77.30	63.33	68.10	62.02	77.28	59.04	82.56	56.15
TC-LSTM	76.60	61.52	68.30	62.26	76.44	57.65	81.90	55.01
AE-LSTM	76.12	61.08	68.83	62.08	75.94	50.11	82.26	56.96
ATAE-LSTM	78.12	68.40	69.44	62.45	78.34	58.47	83.24	61.99
MemNet	78.16	65.83	70.33	64.09	77.89	59.52	83.04	57.91
IAN	78.60	66.31	72.10	65.92	78.62	55.34	82.19	56.30
GCAE	77.41	65.06	69.12	62.17	78.25	54.31	82.35	56.49
PosATT-LSTM	79.40	-	72.80	-	-	-	-	-
PMHSAT-BiGRU	**80.27**	**69.25**	**73.14**	**68.27**	**79.67**	**61.89**	**83.24**	**62.39**

The bold value indicates that the effect of this method is the best compared with other baseline models.

TABLE 4 | Analysis of position-enhanced multi-head self-attention based BiGRU model (PMHSAT-BiGRU) model.

	R14	L14	R15	R16
BiGRU	77.14	69.44	76.92	81.55
MHSAT-BiGRU	78.31	70.06	77.28	81.96
PAT-BiGRU	79.38	71.32	77.75	82.75
PMHSAT-BiGRU	**80.27**	**73.14**	**79.67**	**83.24**

The bold value indicates that the effect of this method is the best compared with other baseline models.

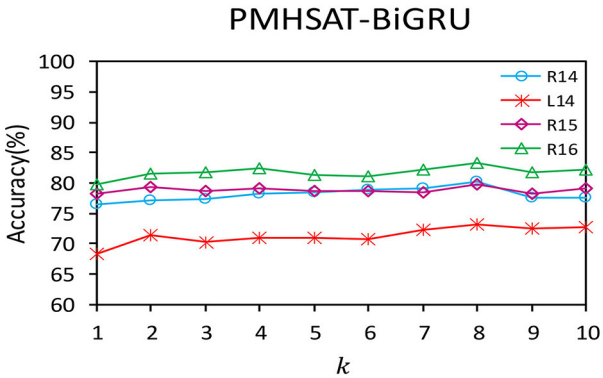


FIGURE 2 | The influence of the number of the heads k of the multi-head attention mechanism on the accuracy of the model.

is equally treated. Therefore, the performance of the PMHSAT-BiGRU model is obviously better than the PosATT-LSTM.

Overall, the performance of the PMHSAT-BiGRU model is superior to above baseline models. In particular, compared with the original LSTM model, the Accuracy values of the PMHSAT-BiGRU model on R14, L14, R15, and R16 are improved by 5.72, 6.06, 4.52, and 3.15%, respectively.

4.4. Model Analysis

In this section, a series of models will be designed to verify the effectiveness of the PMHSAT-BiGRU model. First of all, in order to verify the validity of the position information, the position information is removed from the PMHSAT-BiGRU model, denoted by the MHSAT-BiGRU model. In the MHSAT-BiGRU model, the representations of aspect words and sentences, and the multi-head self-attention mechanism is adopted to model the relationships between aspect words and sentences. Second, in order to verify the effectiveness of the multi-head self-attention mechanism for the PMHSAT-BiGRU model, the multi-head self-attention mechanism is replaced with a normal attention mechanism and the other parts are kept unchanged in PMHSAT-BiGRU, denoted by the PAT-BiGRU model. The structure of the PAT-BiGRU model is almost similar to the ATAE-LSTM model, except that the PAT-BiGRU model considers the position relationship between aspect words and context and uses the BiGRU structure instead of LSTM. Finally, we also use the BiGRU model to verify the effectiveness of our multi-head self-attention mechanism. The experimental results of these models are shown in **Table 4**.

From **Table 4**, the performance of the BiGRU model is the worst in all models. The reason is that this model equally treats every word in sentences. In contrast, the multi-head self-attention mechanism can learn contextual information related to terms from different contexts. So the MHSAT-BiGRU model gets better grades than the BiGRU model. Because the PAT-BiGRU model uses position embedding and aspect embedding to calculate the weight of attention, while the MHSAT-BiGRU model only adopts the aspect embedding, the PAT-BiGRU model performs better than the MHSAT-BiGRU model. Compared with the PMHSAT-BiGRU model, since the PAT-BiGRU model ignores the aspect words with different meanings in different contexts and the role of important words in aspect terms, the performance of the PAT-BiGRU model is lower than the PMHSAT-BiGRU model.

On the basis of the above analysis, the PMHSAT-BiGRU model performs the best in all models. The reason is that

Aspect term	Sentence	Polarity
wine list	The wine list was extensive - though the staff did not seem knowledgeable about wine pairings.	positive
staff	The wine list was extensive - though the staff did not seem knowledgeable about wine pairings.	negative

FIGURE 3 | Visualize the weight of attention for aspect terms and sentences by PMHSAT-BiGRU.

the model not only fully considers the position information of aspect terms in the corresponding sentences but also regards the relationships between aspect terms and sentences from multiple levels. Besides, the model pays more attention to the important words in aspect terms, which is mainly realized by the multi-head self-attention mechanism.

The multi-head self-attention mechanism is employed to learn the semantic information in different representation subspaces for the PMHSAT-BiGRU model, where the number of the subspaces is controlled by the number of the heads k in the multi-head attention mechanism. In the following, the influence of the parameter k on the Accuracy of the PMHSAT-BiGRU model is shown in **Figure 2**.

It can be observed that when k increases, the changing trends of the Accuracy values for the PMHSAT-BiGRU model on the four data sets are similar. Specifically, when $k = 1$, the multi-head self-attention mechanism is equivalent to an ordinary single-head self-attention mechanism. As the values of k increase, the performance of the model almost increases from 1 to 8, and then the performance of the model declines with the rise of k. The main reason is that when the value of k is more than 8, some heads will learn same attention weights, which bring noise for the sentiment classification of aspect terms. Evidently, when $k = 8$, the performance of the model on the four data sets is the best.

4.5. Case Study

In order to intuitively show the validity of the model, we will take a sentence with aspect terms as an example for predicting the aspect terms of sentences by the PMHSAT-BiGRU model. For example, the sentiment polarities of the sentence "The wine list was extensive-though the staff did not seem knowledgeable about wine pairings." will be predicted by the model. For the sentence, the attention weights of the aspect terms and the sentence are visualized in **Figure 3**, the darker the color of words is, the more the words are important for predicting the sentiment polarities of aspect terms. It easily finds that the model focuses on the words adjacent to the aspect terms. When the model predicts the sentiment polarity of the aspect term "wine list," the word "extensive" is closer to the position of the words "wine list," so the model pays more attention to "extensive" which plays an important role in calculating the sentiment polarity of the aspect term "wine list;" whereas, the words "not" and "knowledgeable" are farther away from the aspect term, and then they receive less attention. In the aspect term "wine list," the word "wine" gets more attention which is mainly realized by the self-attention mechanism. So the model can correctly predict the sentiment polarity of the aspect term "wine list" as positive. Similarly, when the model predicts the sentiment polarity of the aspect term "staff," the words: "knowledgeable" and "not" get more attention than other words in the model. Since the positive polarity of the word "knowledgeable" for the "staff" is eventually reversed by the word "not," the model can correctly predict the sentiment polarity of "staff" as negative. Therefore, the PHMSAT-BiGRU model accurately predicts the sentiment polarities of all aspect terms of the sentence. From this, even if a given sentence contains multiple aspect terms, the PMHSAT-BiGRU model can find the relevant sentiment descriptors of the given aspect terms, exactly predicting the sentiment polarities of its aspect terms.

5. CONCLUSION AND FUTURE STUDY

In this article, a PMHSAT-BiGRU based on the position influence vector, multi-head self-attention mechanism, and BiGRU is proposed for the ASC. The PMHSAT-BiGRU model considers the aspect terms contained in multi-words and the importance of each context word. The model also integrates the aspect word and its relative position information of the context into the semantic model. First, this model establishes position vectors based on the position information between the aspect words and its context. Then the position vectors and aspect embeddings are added to the hidden representations of BiGRU. Finally, the keywords in aspect terms and the sentiment features related to the aspect terms are captured by the multi-head self-attention mechanism. The experimental results on the SemEval 2014, 2015, and 2016 datasets show that the PMHSAT-BiGRU model can learn effective features and obtain better performance than the baseline model on the ASC tasks. In future study, the individual models and different fused approaches of the three important factors will be further improved.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

XL and YD contributed to conception and design of the study. LD and XL made all experiments and wrote sections of the manuscript. YF and FS organized the database. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This study was partially supported by the National Natural Science Foundation (Nos. 61802316, 61872298 and 61902324), Chunhui Plan Cooperation and Research Project, Ministry of Education of China (Nos. Z2015109 and Z2015100), Young Scholars Reserve Talents program of Xihua University, Scientific Research Fund of Sichuan Provincial Education Department (No. 15ZA0130), Science and Technology Department of Sichuan Province (Nos. 22ZDYF3157 and 2021YFQ0008), Key Scientific Research Fund of Xihua University (No. z1422615), Xihua University Innovation Fund (Nos. YCJJ2021031, YCJJ2021025, and YCJJ2021124), and Opening Project of Intelligent Policing Key Laboratory of Sichuan Province (No. ZNJW2022ZZZD003).

REFERENCES

Akhtar, M. S., Garg, T., and Ekbal, A. (2020). Multi-task learning for aspect term extraction and aspect sentiment classification. *Neurocomputing* 398, 247–256. doi: 10.1016/j.neucom.2020.02.093

Chen, F., and Huang, Y. (2019). Knowledge-enhanced neural networks for sentiment analysis of Chinese reviews. *Neurocomputing* 368, 51–58. doi: 10.1016/j.neucom.2019.08.054

Chen, P., Sun, Z., Bing, L., and Yang, W. (2017a). "Recurrent attention network on memory for aspect sentiment analysis," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Copenhagen), 452–461. doi: 10.18653/v1/D17-1047

Chen, Q., Hu, Q., Huang, J. X., He, L., and An, W. (2017b). "Enhancing recurrent neural networks with positional attention for question answering," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Tokyo), 993–996. doi: 10.1145/3077136.3080699

Dey, R., and Salemt, F. M. (2017). "Gate-variants of gated recurrent unit (GRU) neural networks," in *IEEE International Midwest Symposium on Circuits and Systems* (Boston, MA), 1597–1600. doi: 10.1109/MWSCAS.2017.8053243

Gan, C., Wang, L., Zhang, Z., and Wang, Z. (2020). Sparse attention based separable dilated convolutional neural network for targeted sentiment analysis. *Knowledge Based Syst.* 188, 1–10. doi: 10.1016/j.knosys.2019.06.035

Gu, S., Zhang, L., Hou, Y., and Song, Y. (2018). "A position aware bidirectional attention network for aspect level sentiment analysis," in *Proceedings of the 27th International Conference on Computational Linguistics (ICCL)* (Santa Fe, NM), 774–784.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Jiang, M., Yu, M., Ming, Z., Liu, X., and Zhao, T. (2011). "Target-dependent twitter sentiment classification," in *Proceedings of the 49th Conference on Association for Computational Linguistics (ACL)* (Portland, ON), 151–160.

Lai, S., Xu, L., Liu, K., and Zhao, J. (2015). "Recurrent convolutional neural networks for text classification," in *Proceedings of the Conference on Association for the Advancement of Artificial Intelligence (AAAI)* (Austin, TX), 2267–2273.

Leng, X.-L., Miao, X.-A., and Liu, T. (2021). Using recurrent neural network structure with enhanced multi-head self-attention for sentiment analysis. *Multimedia Tools Appl.* 80, 12581–12600. doi: 10.1007/s11042-020-10336-3

Letarte, G., Paradis, F., Giguére, P., and Laviolette, F. (2018). "Importance of self-attention for sentiment analysis," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Brussels), 267–275. doi: 10.18653/v1/W18-5429

Liu, B., An, X., and Huang, J. X. (2015). "Using term location information to enhance probabilistic information retrieval," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Santiago), 883–886. doi: 10.1145/2766462.2767827

Liu, B., and Lane, I. (2016). "Attention-based recurrent neural network models for joint intent detection and slot filling," in *Proceedings of the 17th Conference on International Speech Communication Association* (San Francisco, CA), 685–689. doi: 10.21437/Interspeech.2016-1352

Liu, P., Qiu, X., and Huang, X. (2016). "Recurrent neural network for text classification with multi-task learning," in *Proceedings of the 2016 Conference on International Joint Conference on Artificial Intelligence(IJCAI)* (New York, NY), 912–921.

Luong, M. T., Pham, H., and Manning, C. D. (2015). "Effective approaches to attention-based neural machine translation," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Lisbon), 1412–1421. doi: 10.18653/v1/D15-1166

Ma, D., Li, S., Zhang, X., and Wang, H. (2017). "Interactive attention networks for aspect-level sentiment classification," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)* (Melbourne, VIC), 4068–4074. doi: 10.24963/ijcai.2017/568

Mikolov, T., Karafiát, M., Burget, L., Cernocký J., and Khudanpur, S. (2010). "Recurrent neural network based language model," in *Proceedings of the 11th International Speech Communication Association* (Makuhari), 26–30. doi: 10.21437/Interspeech.2010-343

Mnih, V., Heess, N., Graves, A., and Kavukcuoglu, K. (2014). "Recurrent models of visual attention," in *Proceedings of the 28th Conference on Neural Information Processing Systems (NIPS)* (Montreal, VIC), 2204–2212.

Mullen, T., and Collier, N. (2004). "Sentiment analysis using support vector machines with diverse information sources," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Barcelona), 25–26.

Nguyen, T. H., and Shirai, K. (2015). "PhraseRNN: phrase recursive neural network for aspect-based sentiment analysis," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Lisbon), 2509–2514. doi: 10.18653/v1/D15-1298

Pang, B., and Lee, L. (2008). Opinion mining and sentiment analysis. *Found. Trends Inform. Retrieval* 2, 1–135. doi: 10.1561/1500000011

Pennington, J., Socher, R., and Manning, C. (2014). "Glove: global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Doha), 1532–1543. doi: 10.3115/v1/D14-1162

Santos, C. N. D., and Gattit, M. (2014). "Deep convolutional neural networks for sentiment analysis of short texts," in *Proceedings of the 25th International Conference on Computational Linguistics (ICCL)* (Dublin), 69–78.

Shaw, P., Uszkoreit, J., and Vaswani, A. (2018). "Self-attention with relative position representations," in *Proceedings of the 2018 Conference on Association for Computational Linguistics (ACL)* (New Orleans, LA), 464–468. doi: 10.18653/v1/N18-2074

Socher, R., Lin, C. Y., Ng, A. Y., and Manning, C. D. (2011). "Parsing natural scenes and natural language with recursive neural networks," in *Proceedings of the 28th International Conference on Machine Learning (ICML)* (Bellevue, WA), 129–136.

Tan, X. W., Cai, Y., Xu, J. Y., Leung, H. F., and Li, Q. (2020). Improving aspect-based sentiment analysis via aligning aspect embedding. *Neurocomputing* 383, 336–347. doi: 10.1016/j.neucom.2019.12.035

Tang, D., Qin, B., Feng, X., and Liu, T. (2016a). "Effective LSTMs for target-dependent sentiment classification," in *Proceedings of the 26th Conference*

on International Conference on Computational Linguistics (ICCL) (Osaka), 3298–3307.

Tang, D., Qin, B., and Liu, T. (2016b). "Aspect level sentiment classification with deep memory network," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Austin, TX), 214–224. doi: 10.18653/v1/D16-1021

Thet, T. T., Na, J. C., and Khoo, C. S. G. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. *J. Inform. Sci.* 36, 823–848. doi: 10.1177/0165551510388123

Wang, Y., Huang, M., Zhu, X., and Zhao, L. (2016). "Attention-based LSTM for aspect-level sentiment classification," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Austin, TX), 606–615. doi: 10.18653/v1/D16-1058

Xiao, L., Hu, X., Chen, Y., Xue, Y., Chen, B., Gu, D., et al. (2020). Multi-head self-attention based gated graph convolutional networks for aspect-based sentiment classification. *Multimedia Tools Appl.* 1-2, 1–20. doi: 10.1007/s11042-020-10107-0

Xu, Q., Zhu, L., Dai, T., and Yan, C. (2020). Aspect-based sentiment classification with multi-attention network. *Neurocomputing* 388, 135–143. doi: 10.1016/j.neucom.2020.01.024

Xue, W., and Li, T. (2018). "Aspect based sentiment analysis with gated convolutional networks," in *Proceedings of the 56th Association for Computational Linguistics (ACL)* (Melbourne, VIC), 2514–2523. doi: 10.18653/v1/P18-1234

Yin, W., Schütze, H., Xiang, B., and Zhou, B. (2015). ABCNN: Attention-based convolutional neural network for modeling sentence pairs. *Trans. Assoc. Comput. Linguist.* 4, 259–272. doi: 10.1162/tacl_a_00097

Zeng, J., Ma, X., and Zhou, K. (2019). Enhancing attention-based LSTM with position context for aspect-level sentiment classification. *IEEE Access* 7, 20462–20471. doi: 10.1109/ACCESS.2019.2893806

Zhang, L., Wang, S., and Liu, B. (2018). Deep learning for sentiment analysis: a survey. *Wiley Interdiscipl. Rev. Data Mining Knowledge Discov.* 8:e1253. doi: 10.1002/widm.1253

Zhang, Z., and Lan, M. (2015). "ECNU: extracting effective features from multiple sequential sentences for target dependent sentiment analysis in reviews," in *Proceedings of the 9th Conference on International Workshop on Semantic Evaluation (SemEval2015)* (Denver, CO), 736–741. doi: 10.18653/v1/S15-2125

Zhou, J., Huang, J. X., Hu, Q. V., and He, L. (2020). Is position important? Deep multi-task learning for aspect-based sentiment analysis. *Appl. Intell.* 50, 3367–3378. doi: 10.1007/s10489-020-01760-x

Zhou, Z., Liu, F. A., and Wang, Q. (2019). R-transformer network based on position and self-attention mechanism for aspect-level sentiment classification. *IEEE Access* 7, 127754–127764. doi: 10.1109/ACCESS.2019.2938854

Conflict of Interest: FS was employed by company Sichuan Suitang Science and Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Li, Ding, Du, Fan and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 07 March 2022
doi: 10.3389/fpsyg.2022.810507

Exploring the Domain of Emotional Intelligence in Organizations: Bibliometrics, Content Analyses, Framework Development, and Research Agenda

Baobao Dong[1], Xing Peng[1] and Na Jiang[2]*

[1] School of Business and Management, Jilin University, Changchun, China, [2] School of Music, Dance and Drama, Changchun Humanities and Sciences College, Changchun, China

Emotion is a kind of micro foundation that can affect human behaviors even in the digital era. Emotional intelligence (EI) is an important psychological factor that affects the growth and development of organizations from the view of emotion. Based on current bodies of literature, a comprehensive review of EI can contribute to its theory development in organization research and facilitate EI research burgeoning. We visualize the landscape of EI by analyzing 1,996 articles with CiteSpace their concepts, dimensions, and measurement. We propose two specific mechanisms, which clarify how individuals with high EI use emotional information to influence themselves and others. Following this, we develop a theoretical framework of EI at levels of individual, team, and organization. Finally, future directions and research agenda are addressed. This research contributes to the literature of EI and provides practical insight for practitioners.

Keywords: emotional intelligence, group emotional intelligence, framework development, bibliometric, research agenda

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Ye Wenping,
Jinan University, China
Zhongwei Cao,
Jilin University of Finance and Economics, China

****Correspondence:***
Baobao Dong
markruby@126.com

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *07 November 2021*
Accepted: *03 February 2022*
Published: *07 March 2022*

Citation:
Dong B, Peng X and Jiang N (2022) Exploring the Domain of Emotional Intelligence in Organizations: Bibliometrics, Content Analyses, Framework Development, and Research Agenda. Front. Psychol. 13:810507. doi: 10.3389/fpsyg.2022.810507

INTRODUCTION

Emotion is fundamental to human experiences influencing our daily activities such as cognition, communication, learning, and decision making. For centuries, psychologists have tried to understand and define emotions. Recently, emotional intelligence (herein referred to as EI), as a special unique resource within organizations, has gained attention from scholars and practitioners. Recent studies highlight the importance of EI as a predictor in important domains, such as psychology (e.g., job satisfaction, self-efficacy), behaviors (e.g., organizational citizenship behavior, workplace deviant behavior, ethical behavior) and work outcomes (e.g., job performance, leadership effectiveness, career success) (Wong and Law, 2002; Brackett et al., 2004; Landy, 2005; Momm et al., 2014; Tuncdogan et al., 2017). As such, we argue that EI is an important factor affecting organizational development and future growth.

Emotional intelligence, as an individual-level variable, means affective tendency to effectively use emotional information to achieve expected results (Bell, 2007). Members in an organization with high EI can successfully affect the social environment at work and achieve high performance by regulating their emotions (Momm et al., 2014), which is also considered as the main reason why

early studies on EI focused on the individual level. However, it should be noted that an organization is a social structure interwoven with relationships, and the flow of emotional information will not only affect individual behavior but also have cross-level effects. On the one hand, because decisions and behaviors of organizational members are always affected by emotional factors, EI of high-power members may have a significant impact on team or organizational effectiveness (Azouzi and Jarboui, 2014). On the other hand, in social communication, EI may affect emotional or behavioral responses of others. Individual EI can also be aggregated into a higher level of group EI. Such group norms that deal with emotions effectively and flexibly not only regulate the emotional state of individual members or teams inward but also affect the atmosphere of other teams or organizations outward, thus producing cross-level effects (Druskat and Wolff, 2001). As an antecedent of organizational performance, EI may be regarded as a hidden "driving force" affecting organizational growth (Momm et al., 2014).

However, since the concept of EI was proposed in the 1990s, more and more scholars began to pay attention to EI, but it cannot be ignored that the literature is still limited in the following two ways. First, the concept and measurement of EI are still controversial. There are various concepts of EI under different theoretical frameworks, and its measurements are not completely consistent, which leads to inconsistent results (Landy, 2005; Tuncdogan et al., 2017; MacCann et al., 2020). Second, the scope of research topics is a little narrow and concentrated. EI research mostly focuses on the workplace, and its influence at the level of team and organization is relatively single. As an individual-level variable in an organization, the influence of EI on all levels of an organization may be more complex, and future studies need to further explore the differentiated process and influences of EI at different levels.

Therefore, we believe that a review of EI should be conducted to understand the current research situation and trend of EI. First, we present the development status of EI in the field of organization by analyzing bodies of literature. We use big data analysis methods, such as CiteSpace, to conduct quantitative analysis and present the landscape and evolution of EI. Second, we review the definitions, dimensions, and measurements of EI. Third, we establish and analyze two emotional influence mechanisms of individuals with high EI, which illustrate how individuals with high EI use emotional information to influence themselves and others. Fourth, after reviewing the existing bodies of literature on EI, we construct a theoretical framework focusing on the impact of EI and its corresponding moderating effect. Finally, we propose research agenda for future research.

VISUALIZING THE LANDSCAPE OF EMOTIONAL INTELLIGENCE

Based on bibliometrics, CiteSpace is used to analyze the trend of EI research. The principle of collecting data is as follows. First, data are collected from the Web of Science database ranging from 1990 to 2020. The year 1990 is the base year, because the concept of EI was formally introduced by Salovey and Mayer in 1990. Second, we choose bodies of literature using terms such as "emotional intelligence" and "emotional ability" in the subject bar, set the operator to "OR," and retain empirical articles and reviews. In addition, by screening literature topics, we select relevant literature in the fields of economics, management, and sociology (e.g., "psychology social," "management," "psychology applied," "behavioral sciences," "business," "communication," and "economics"). Finally, a total of 1,996 articles are retrieved as research objects by manual inspection and exclusion of irrelevant ones.

In order to clearly and intuitively display the full picture of EI, we conducted an analysis on views of publication, journals, scholars, node literature, and cooperation network.

Analysis of Publications

Review on EI is of help to provide a visualized picture of the popularity and trend of EI research. **Figure 1** shows a total of 1,996 records that were published in the past 30 years *via* Web of Science. There is an increasing trend in publication, and it can be divided into three stages: infancy stage (1990–1997), development stage (1997–2008), and burst stage (2008–2020). During the infancy stage, it was controversial whether EI coincided with the concept of personal characteristics or cognitive intelligence in a wide range. Some scholars were skeptical about the research value of EI. Therefore, there were relatively few studies on EI in this stage. The field did not experience much growth until 1997. Scholars have started to pay their attention to EI since the ability model and mixed model of EI were first proposed in1997. The increasing bodies of literature provided evidence that EI had attracted an extraordinary attention from numerous scholars who focused on the concept and dimensions of EI and began to explore its effects. During its burst stage, the maturity of measurement was conducive to empirical research in this field, and publication showed a "blowout" growth.

Analysis of Journals and Scholars

In terms of citation or centrality, **Table 1** shows the top ten journals that publish EI research, which reflects that these journals show more in-depth research. As indicated in **Table 1**, the top ten journals include Journal of Organizational Behavior, Academy of Management Journal and other journals belonging to UTD24 and FT50, which shows that international journals are interested in EI. **Table 2** shows the top ten scholars based on citation or centrality, which can help scholars understand the development and direction of EI.

Analysis of Key Node Literature

Table 3 shows the top ten cited key node bodies of literature in EI field that explain the core constructs of EI and mainly cover conceptual dimensions, relationship among dimensions, and measurement scales. For instance, Mayer et al. (1999) and Petrides et al. (2007) provide insight into the concept and nature of EI. Joseph and Newman (2010) explored relationships among the dimensions of EI and proposed a cascading model. Schutte et al. (1998), Brackett and Mayer (2003), Law et al. (2004),

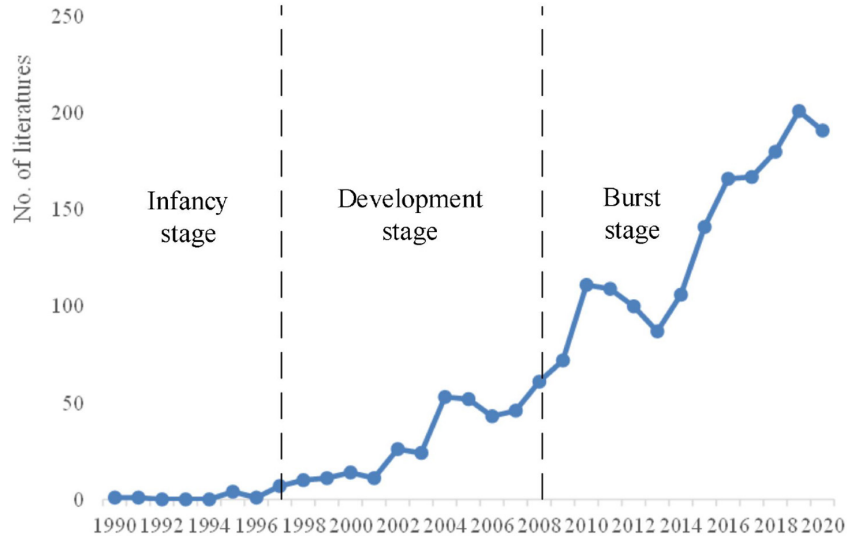


FIGURE 1 | Chart of output trends in emotional intelligence (EI).

TABLE 1 | Ranking of journals based on citation rate/centrality (top 10).

Ranking	Journal	Citation rate	Journal	Centrality
1	Personality and Individual Differences	942	Emotion	0.17
2	The Journal of Applied Psychology	583	Personality and Social Psychology Bulletin	0.13
3	Journal of Organizational Behavior	439	Journal of Personality Assessment	0.13
4	Journal of Personality and Social Psychology	399	Personality and Individual Differences	0.11
5	Emotion	378	The Journal of Applied Psychology	0.11
6	Leadership Quarterly	292	Journal of Organizational Behavior	0.10
7	Academy of Management Journal	266	Cognition and Emotion	0.10
8	Annual Review of Psychology	256	Psychological Bulletin	0.10
9	Journal of Vocational Behavior	242	Journal of Management	0.09
10	Cognition and Emotion	227	Leadership and Organization Development Journal	0.09

and Van Rooy and Viswesvaran (2004) conducted tests on measurement scales and content validity of EI. These node bodies of literature are conducive to subsequent theoretical development and empirical studies.

Analysis of Cooperation Network

Figure 2 shows a distribution of a country-based cooperation network. The top five countries ranked by output of EI literature are the United States (690), the United Kingdom (212), Australia (195), China (136), and Canada (130). In terms of publications, China has gradually begun to pay attention to EI, but the number is still quite different from that of the United States, the United Kingdom and other countries that lead the frontier of research on EI. In terms of centrality, China ranked 14th with 0.06, much lower than that of the United Kingdom (0.26) and the United States (0.21) which indicates that important future endeavors to address significant gaps in EI research between China and other countries is necessary.

TABLE 2 | Ranking of authors based on citation rate/centrality (top 10).

Ranking	Authors	Citation rate	Authors	Centrality
1	Mayer, John D.	407	Petrides, K. V.	0.21
2	Petrides, K. V.	337	Mayer, John D.	0.14
3	Schutte, Nicola S.	196	Cote, Stephane	0.14
4	Brackett, Marc A.	182	Salovey, Peter	0.13
5	Joseph, Dana L.	181	Schutte, Nicola S.	0.12
6	Zeidner, M.	172	Elfenbein, Hillary Anger	0.10
7	Goleman, Daniel	161	Brackett, Marc A.	0.09
8	Cote, Stephane	139	Austin, Elizabeth J.	0.08
9	Austin, Elizabeth J.	136	Ashkanasy, Neal M.	0.08
10	Bar-On, Roi	132	Judge, Timothy A.	0.07

DEFINITION, DIMENSION, AND MEASUREMENT OF EMOTIONAL INTELLIGENCE

Definition of Emotional Intelligence

Compared with intelligence or personality, EI is a relatively new construct. Based on social intelligence theory (Thorndike, 1920) and multiple intelligence theory Salovey and Mayer (1990), (Gardner, 1993), first proposed the concept of EI, which is "the ability to monitor one's own and others' feelings and emotions, to

TABLE 3 | Studies on emotional intelligence (EI) based on co-citation analysis from 1990 to 2020 (top 10).

References	Title	Journal
Joseph and Newman, 2010	Emotional intelligence: An integrative meta-analysis and cascading model	Journal of Applied Psychology
Mayer et al., 2008	Human abilities: Emotional intelligence	Annual Review of Psychology
O'Boyle et al., 2011	The relation between emotional intelligence and job performance: A meta-analysis	Journal of Organizational Behavior
Petrides et al., 2007	The location of trait emotional intelligence in personality factor space	British Journal of Psychology
Mayer et al., 1999	Emotional intelligence meets traditional standards for an intelligence	Intelligence
Brackett and Mayer, 2003	Convergent, discriminant, and incremental validity of competing measures of emotional intelligence	Personality and Social Psychology Bulletin
Martins et al., 2010	A comprehensive meta-analysis of the relationship between emotional intelligence and health	Personality and Individual Differences
Schutte et al., 1998	Development and validation of a measure of emotional intelligence	Personality and Individual Differences
Van Rooy and Viswesvaran, 2004	Emotional intelligence: A meta-analytic investigation of predictive validity and nomological net	Journal of Vocational Behavior
Law et al., 2004	The construct and criterion validity of emotional intelligence and its potential utility for management studies	Journal of Applied Psychology

Source: Collated according to relevant literature.

discriminate among them and to use this information to guide one's thinking and actions (p. 5)," and defined it as a subset of social intelligence. EI was first introduced into management filed by Goleman (1995), who believed that EI is the ability to maintain self-control, enthusiasm and perseverance, and self-motivation, and that it consists of five major parts: (a) being aware of one's emotions, (b) managing emotion, (c) motivating oneself, (d) identifying emotions in others, and (e) dealing with interpersonal relationships. The notion is quietly different from that of Salovey and Mayer (1990), who believed that EI is focused on the emotional ability to connect emotion with cognition.

There is a general agreement on the construction of models that divides EI into ability-based model and mixed model. The ability-based model of EI is focused on specific competence, and its core is emotion. The ability-based model was first proposed by Mayer and Salovey in 1997, which partially overlapped with cognitive ability. Mayer and Salovey (1997) defined EI as "the ability to perceive emotions, to access and generate emotions so as to help thought, to understand emotions and emotional knowledge, and to reflectively regulate emotions so as to promote emotional and intellectual growth (p. 5)." Based on the definition of EI by Mayer et al. (2008), MacCann et al. (2014) empirically verified EI as a second-stratum factor of intelligence and defined EI as the ability to process and reason emotional information, and that perception, understanding, and management of emotions are three dimensions of EI. Most scholars regard EI as a kind of ability, but Minbashian et al. (2017) conceptualized EI as a knowledge structure after analyzing measurement items of ability-based EI (Mayer and Salovey, 1997; Joseph and Newman, 2010). The knowledge structure specifically reflects an individual's declarative knowledge of emotions, including knowledge of motivation and cognition of affective states, how emotions swing, how to form a more complex affective state, and strategies for regulating one's emotions. In addition, EI is essential in social communication. Kidwell et al. (2011) introduced the concept of EI into marketing and defined EI as "the ability to acquire and apply knowledge from one's emotions and those of others to produce beneficial outcomes (p. 78)." In this regard, Petrides and Furnham (2003) defined EI as a behavioral orientation related to an individual's ability to recognize, process, and use emotional information, as well as self-cognition. **Table 4** shows definitions of EI.

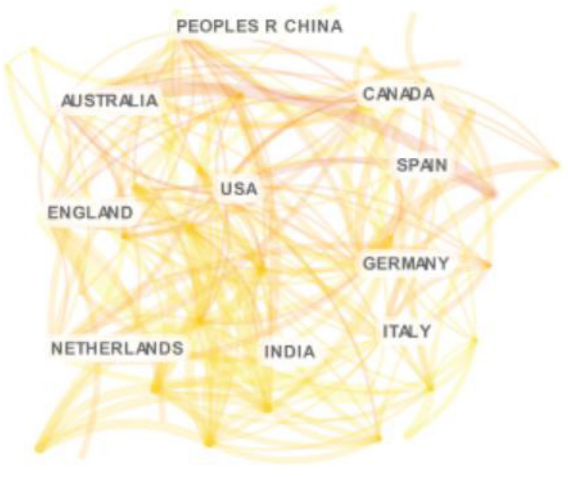


FIGURE 2 | Countries/regions of cooperative networks of emotional intelligence research in organization field from 1990 to 2020.

Dimensions and Measures of Emotional Intelligence

The dimension of EI is mainly based on two theoretical models: ability-based model and mixed model. Parallel to the theoretical model of EI is two measurement models, ability measurement and rating measurement. After distinguishing measurements based on the ability-based model and the mixed model, scholars have summarized three streams on measuring EI: (a) ability

TABLE 4 | Definitions of EI and its value.

References	Definition	Value
Salovey and Mayer, 1990	The ability to monitor one's own and others' feelings and emotions, to discriminate among them and to use this information to guide one's thinking and actions.	The concept of emotional intelligence was first proposed and defined.
Goleman, 1995	The ability to control impulses, delay gratification, regulate moods, keep distress from obstructing cognitive functioning, and empathize.	Emotional intelligence was first introduced into management field and widely discussed in various fields.
Mayer and Salovey, 1997	The ability to perceive emotions, to access and generate emotions so as to help thought, to understand emotions and emotional knowledge, and to reflectively regulate emotions so as to promote emotional and intellectual growth.	The ability model of emotional intelligence was proposed firstly.
Bar-On, 1997	A set of non-cognitive capabilities and skills that influence one's ability to succeed in coping with environmental demands and pressures.	A mixed model of emotional intelligence was first proposed.
Goleman, 1998	A synthesis of self-awareness, self-management, self-motivation, empathy and interpersonal skills.	The concept of emotional competence was first proposed and the emotional competence inventory (ECI) was developed.
Mayer et al., 2000	The ability to carry out accurate reasoning about emotions and the ability to use emotions and emotional knowledge to enhance thought.	A slight adjustment was made to the definition of Mayer and Salovey (1997) to develop the multifactorial emotional intelligence scale.
Petrides and Furnham, 2001, 2003	A constellation of emotion-related self-perceived abilities and dispositions, including individual differences in the ability to understand, process, and utilize affect-laden information.	The most comprehensive mixed model of EI.
Kidwell et al., 2011	The ability to acquire and apply knowledge from one's emotions and those of others to produce beneficial outcomes.	The concept of emotional intelligence was first introduced into marketing exchange.
MacCann et al., 2014	The ability to process and reason affective information.	In the ability-based framework, the essence of emotional intelligence was empirically verified as a part of intelligence.
Minbashian et al., 2017	Including knowledge of the motivational and cognitive effects of various affective states, how emotions transition over time, how they combine to form more complex affective states, and strategies that can be used to regulate one's affective states.	In the ability-based framework, the concept of emotional intelligence is extended to the concept of knowledge structure.

Source: Collated according to relevant literature.

scales, (b) ratings of ability (self-reported), and (c) ratings of mixed model (self-reported or peer-reported) (Sackett et al., 2017; MacCann et al., 2020).

Mainstream research supports the ability-based model proposed by Mayer and Salovey (1997). This model divides EI into four dimensions: (a) perceiving emotions (the ability to identify and accurately express emotion), (b) emotions to facilitate thought (including not only using existing emotions to promote goal achievement but also generating new emotions in a particular situation to accomplish tasks), (c) understanding emotions (the ability to process emotional information), and (d) managing emotions (the ability to strengthen or weaken emotions). The measurement was called the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT), which is the most commonly used ability-based measure.

In the earliest ability-based model using self-report, EI consisted of the following three dimensions: perception of emotions, management of emotions, and emotional facilitation of thinking (Schutte et al., 1998). The measurement was the Assessing Emotions Scale (AES). However, Davies et al. (1998) proposed four dimensions: (a) one's self emotional appraisal, (b) others' emotional appraisal, (c) regulation of emotions, and (d) use of emotions. Subsequently, aiming at the four dimensions of EI, Wong and Law (2002) developed the Wong and Law Scale (WLEIS). In addition, in the Self-Rated Emotional Intelligence Scale (SREIS) developed by Brackett et al. (2006), EI is divided into four dimensions: perceiving emotions, using emotions, understanding emotions, and managing emotions.

In the mixed model, trait features are emphasized. Based on this view, there are many types of dimensions of EI, and major measurement scales include Emotional Competence Inventory (ECI) (Goleman, 1998), Emotional Quotient Inventory (EQ-i) (Bar-On, 2006), and Trait Emotional Intelligence Questionnaire (TEIQue) (Petrides et al., 2007). Goleman (1998) outlined that EI consists of four major competencies: self-awareness, self-management, social awareness, and social skills. Since Bar-On (2006) regarded the relationship between emotion and social function as the main content of EI in his study, EI consisted of five major domains, intrapersonal competence, interpersonal competence, stress management, adaptability, and general mood. Petrides et al. (2007) pointed out four dimensions of EI, happiness, emotion regulation, emotions, and relationships. **Table 5** shows the dimensions and measurement of EI.

INFLUENCING MECHANISMS OF EMOTIONAL INTELLIGENCE

In proving the rationality of EI, Salovey and Mayer (1990) pointed out that EI needs to process specific emotional information individually. Mayer and Salovey (1997) believed that EI is a group of abilities of processing emotional information. In

fact, most studies generally believe that EI should be interpreted as "a specific ability to help individuals reason and use emotional information" (Fineman, 2006, p. 278; also see Mayer et al., 2000; Beck et al., 2001). This also means that the effectiveness of EI depends on the degree of effective recognition and utilization of emotional information.

After analyzing the bodies of literature, we propose two specific mechanisms of how individuals with high EI use emotional information to influence themselves and others. The first mechanism involves processes that influence the self. Individual with high EI is good at using emotional cues to change their emotions so as to achieve goals such as reshaping their mental state, reducing stress, and improving the quality of decision-making (Bar-On, 2000; Kidwell et al., 2008). The second one is the mechanism by which emotional information is consciously released in social interactions to drive and screen the reaction of others. In the process, EI can not only be displayed publicly but also arouse the emotions of others. The two mechanisms are described in detail below (see **Figure 3**).

The internal mechanism refers to the process by which EI influences the self, and its trigger is the emotional fluctuation caused by events. Afterward, based on the abilities of emotion perception, emotion understanding, emotional promotion thinking, and emotion management described by

TABLE 5 | Dimension and measurement of EI.

References	Model/Measures	Dimension	Content	Scale
Mayer et al., 2000, 2008	Ability model (ability scales)	Perceiving emotion	The ability to perceive of one's and others' emotion.	MSCEIT Scale
		Emotions to facilitate thought	The ability to using emotions to facilitate cognitive activities, such as thinking and problem solving.	
		Understanding emotion	The ability to understand verbal or non-verbal information.	
		Managing emotion	The ability to regulate emotions in oneself and others.	
Davies et al., 1998; Wong and Law, 2002	Ability model (self-report)	One's self emotional appraisal	The ability to understand their deep emotions and be able to express these emotions naturally.	WLEIS Scale
		Others' emotional appraisal	The ability to perceive and understand the emotions of those people around them.	
		Regulation of emotion	The ability to regulate their emotions and rapid recovery from psychological distress.	
		Use of emotion	The ability to use emotions toward constructive activities	
Schutte et al., 1998	Ability model (self-report)	Perception of emotion	Appraisal and expression of emotion in the self and appraisal of emotion in others.	AES Scale
		Management of emotion	Regulation of emotions in the self and regulation of emotions in others.	
		Emotional facilitation of thinking	Flexible planning, creative thinking, redirected attention and motivation.	
Jordan et al., 2002	Ability model (self-report)	Perceive own emotions	Recognize one's own emotions.	WEIP Scale
		Discuss own emotions	Understand and assimilate one's own emotions	
		Manage own emotions	Regulate and generate one's own emotions.	
		Perceive others' emotions	Recognize others' emotions.	
		Manage others' emotions	Empathize and manage others' emotions.	
Brackett et al., 2006	Ability model (self-report)	Perceiving emotion	The ability to identify emotions in oneself and others, as well as in other stimuli.	SREIS Scale
		Using emotion	The ability to harness feelings	
		Understanding emotion	The ability to analyze emotions.	
		Managing emotion	The ability to reduce, enhance, or modify an emotional response in oneself and others, as well as the ability to experience a range of emotions.	
Goleman, 1998	Mixed model	Self-awareness	Accurate in one's emotions.	ECI Scale
		Self-management	Control one's emotions and behaviors.	
		Social awareness	Showing empathy to others, and having a service orientation and organizational awareness.	
		Social skills	Manage interpersonal relationships	
Bar-On, 2006	Mixed model	Intrapersonal competence	The ability to deal with internal emotions.	EQ-I Scale
		Interpersonal competence	The ability to deal with interpersonal emotions	
		Adaptability	The ability to deal with change flexibly.	
		Stress management	The ability to manage external pressure.	
		General mood	Description of general mood	
Petrides et al., 2007; Petrides, 2009	Mixed model	Happiness	More adaptable in general.	TEIQue Scale
		Emotion regulation	More willpower	
		Emotion	Egotism	
		Relationships	Interpersonal skills	

Source: Collated according to relevant literature.

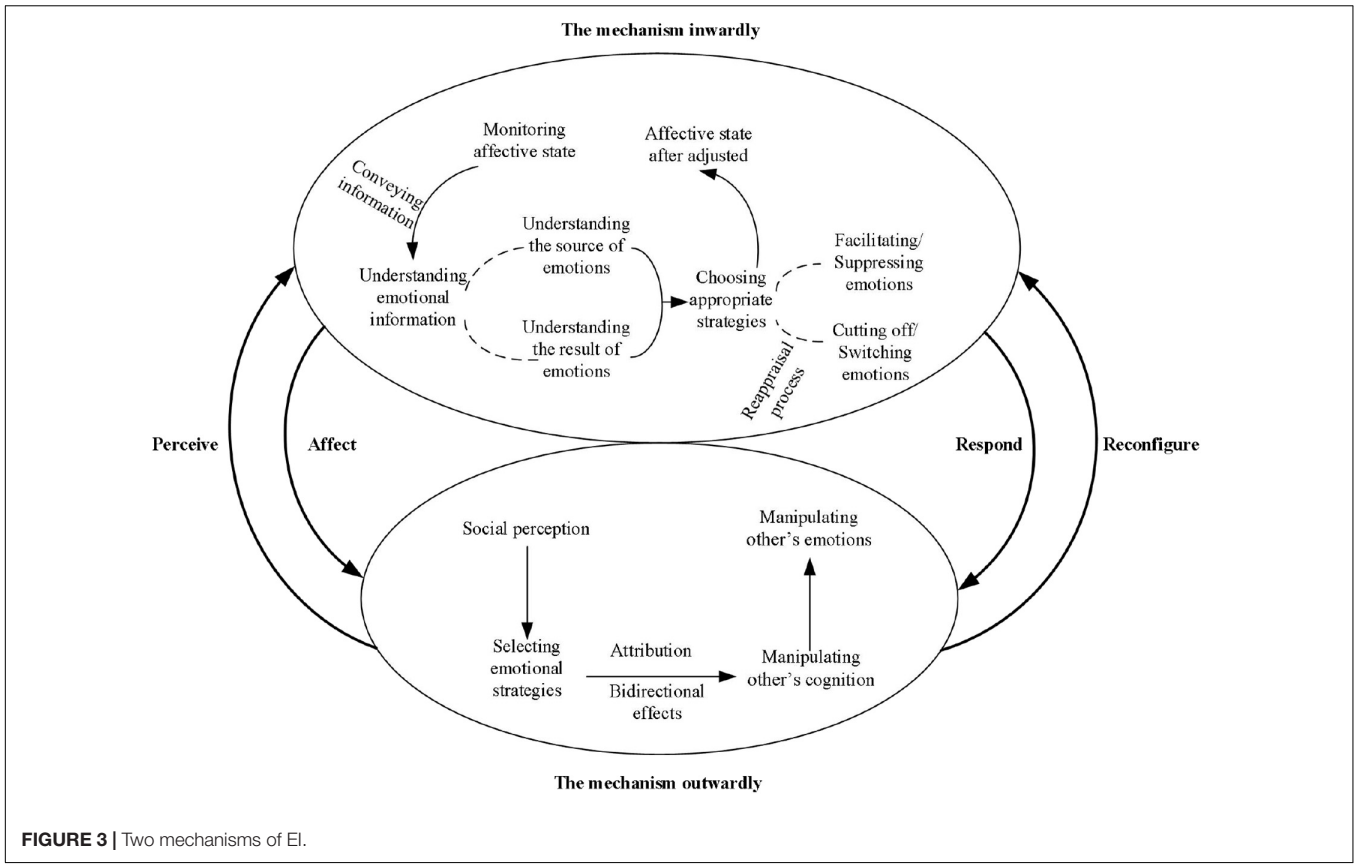


FIGURE 3 | Two mechanisms of EI.

Mayer et al. (2000), emotional information will undergo processing, which is the main content of this mechanism. Individuals with high EI tend to continuously monitor their emotional state (McCarthy et al., 2016). When emotions fluctuate, it is identified and transmitted as a set of unique emotional information (Izard, 1993). The understanding of this information not only lies in the emotion itself but also includes the source of emotion fluctuation and its consequences (MacCann et al., 2020). For example, when faced with unfair situations, job insecurity can trigger unpleasant emotions that stem from unfair behavior at work. In fact, people with low EI are more likely to engage in negative emotions than those with high EI, and it is difficult to find the root cause (Cheung et al., 2016; Dust et al., 2018). Both negative and positive emotions may result in potential threats or goal attainment (Lindebaum and Jordan, 2014; McFarland et al., 2016).

Based on sources and outcomes of emotions, individuals with high EI can use relevant knowledge to select the best strategy (Côté et al., 2010), such as promoting or suppressing emotions, cutting off or switching emotions after reappraisal to maintain a good emotional state. Specifically, if the desired outcome is of help to individual goals, emotional management capability will help individuals to maintain or reinforce existing emotional states, thereby inducing or enhancing individual behavior or motivation (Joseph and Newman, 2010). On the contrary, when existing emotions may lead to negative results, individuals with high EI may suppress current emotions or reappraise their emotional sources. Emotion suppression is a reasonable strategy when excessive emotion can lead to negative results. For example, individuals with high EI may suppress their emotions when they find that self-perceived emotions are too optimistic and may lead to wrong decisions (Lerner and Keltner, 2001; Côté et al., 2010). There are two steps to reappraise the source of emotion. First, individuals should think about whether the source of the emotion is worthy of attention, which ensures that reasonable concerns are proactively addressed and irrelevant distractions are ignored (Dust et al., 2018). Once the emotional source can affect task performance, individuals with high EI will no longer consider the information related to this source, which helps to cut off the existing adverse emotion to restore the desired emotional state. Second, when the emotional source needs to be paid attention, individuals with high EI will reappraise their cognition of the emotional source. For example, when there is anxiety, tension, and other negative emotions due to work pressure, employees with high EI will regard stressful tasks as challenges rather than threats, thus arousing work enthusiasm (Dong et al., 2014), which also means emotional transformation. In general, individuals with high EI can continuously monitor their emotional states, identify and analyze the sources and results of emotional cues, and then adopt beneficial strategies to achieve desired emotional states and expected goals.

Social perception influences the process by which individual EI is associated with responses of others (Chowdry and Newcomb, 1952; Côté et al., 2010). Individuals with high EI

can not only accurately identify others' emotions and their sources but also have a strong perception of subtle changes in emotional atmosphere as well as the direction or reason of such changes, which can help them acquire a lot of knowledge of emotional cues. This knowledge enables them to adopt more appropriate emotional strategies by identifying, understanding, and processing needs from the other person (George, 2000; Wolff et al., 2002). Then, individuals with high EI will select emotional strategies, such as promoting, suppressing, inducing, and faking emotions to display their emotional information to their counterparts by adopting reasonable emotional strategies.

In the context of interaction, attribution is an indispensable step in the process of emotional expression (Eberly and Fong, 2013). When others receive emotional information displayed by individuals, they will explain and analyze the emotion through the attribution process, including the source of the emotion and the motivation and intention of individuals with this emotional information (Dasborough and Ashkanasy, 2002). This means that individuals with high EI may be able to use appropriate strategies to only express emotions that others wish to perceive and interpret (Humphrey et al., 2008). This kind of emotional information can affect others' cognition of the expressor's personal characteristics or behavior. For example, a leaders' emotional display of enthusiasm and concern when communicating and motivating members also encourages members to attribute this to the leader's motivation to try to implement transformational leadership (Dasborough and Ashkanasy, 2002). In addition, people with high EI are able to manipulate other people's emotional response by directing their cognitive processes.

It is important that in social interactions, the effect of EI may not be one-way but bidirectional. Referring to the emotional and intentionality attribution model in leader-member relationship proposed by Dasborough and Ashkanasy (2002), EI can influence intentionality attribution in a social context. Specifically, for a person who exhibits emotional cues, the emotional strategies he uses and the extent to which he uses them may influence the attributions of another person. However, individuals with high EI can accurately perceive and interpret the intention of another party, thus producing the related emotional response. In addition, individuals with high EI may manipulate the cognitive attribution and emotion of another party to effectively promote the generation and reinforcement of the other party's motivation or behavior, thus affecting their performance (Humphrey et al., 2008).

FRAMEWORK DEVELOPMENT OF EMOTIONAL INTELLIGENCE RESEARCH

In order to comprehensively review the development of EI and provide insight for future research, a theoretical framework is developed (see **Figure 4**) in this study. This framework is based on the following parts: (a) notion of EI (individual and team EI); (b) outcomes of EI; (c) moderating factors that regulate the relationship between EI and its outcomes; (d) moderating effect of EI. As the concept of EI is controversial (Elfenbein, 2008; Joseph and Newman, 2010), the ability-based model is widely regarded as the most suitable model for EI (Jordan et al., 2003), so we reviewed EI from the perspective of the ability-based model.

Individual Emotional Intelligence

Mayer and Salovey (1997) proposed the ability-based model of EI, which covers a variety of abilities to understand and process emotional information, including emotional perception, emotional promotion, emotional understanding, and emotional regulation. This is a generally agreed theoretical model describing the composition of EI capabilities (Kidwell et al., 2008; MacCann et al., 2020). Since each can act independently or in combination with each other, Mayer and Salovey (1997), based on this model, argued that the above four aspects are progressive (Mayer and Salovey, 1997; Brackett et al., 2006; Joseph and Newman, 2010). Specifically, emotional perception as the ability to perceive emotions in oneself and others precedes other capabilities. Emotional promotion helps individuals promote cognitive activities using perceived emotions. Since emotion understanding requires language skills and logical thinking, the cognitive enhancement brought by emotion promotion can be directly reflected in emotion understanding. Emotion regulation relies on the processing and analysis of emotional information in emotion understanding (Mayer and Salovey, 1997).

Group Emotional Intelligence

When a high level of interpersonal interaction and emotional cues arise in teams, the influence of individual EI can be reflected at the team level through the activated interpersonal communication mechanism, and the aggregation of individual EI will form a team-level construct called group emotional intelligence (GEI, also known as collective emotional intelligence) (Farh et al., 2012; Troth et al., 2012; Wang, 2015). GEI is first proposed by Druskat and Wolff (2001) and is defined as "the ability of a group to develop a set of norms that manage emotional processes" (p. 132). These norms encourage the expression and regulation of emotional dynamics within and outside a group, thus helping group members to deal with emotional problems more effectively (Curseu et al., 2015). Extant studies on the formation of GEI mainly fall into two streams: one believes that GEI is the sum of individual EI resources (Jordan and Troth, 2004), and the other believes that GEI is the extent to which groups use EI when communicating with each other (Elfenbein, 2006). The former is focused on personal resources brought by members, while the latter is focused on interaction between members (Curseu et al., 2015). Since teams with high EI are better able to "interpret" emotional information and respond to different emotional situations, most studies believe that GEI includes group emotional awareness and group emotional regulation (Jordan and Lawrence, 2009; Troth et al., 2012).

Outcomes of Emotional Intelligence

Emotional intelligence can directly or indirectly affect results through mediating variables. Considering that EI as an ability to use emotions provides a potential emotional background for most of our behaviors and ongoing thought processes (Forgas and George, 2001), and its influence on behavior at different levels

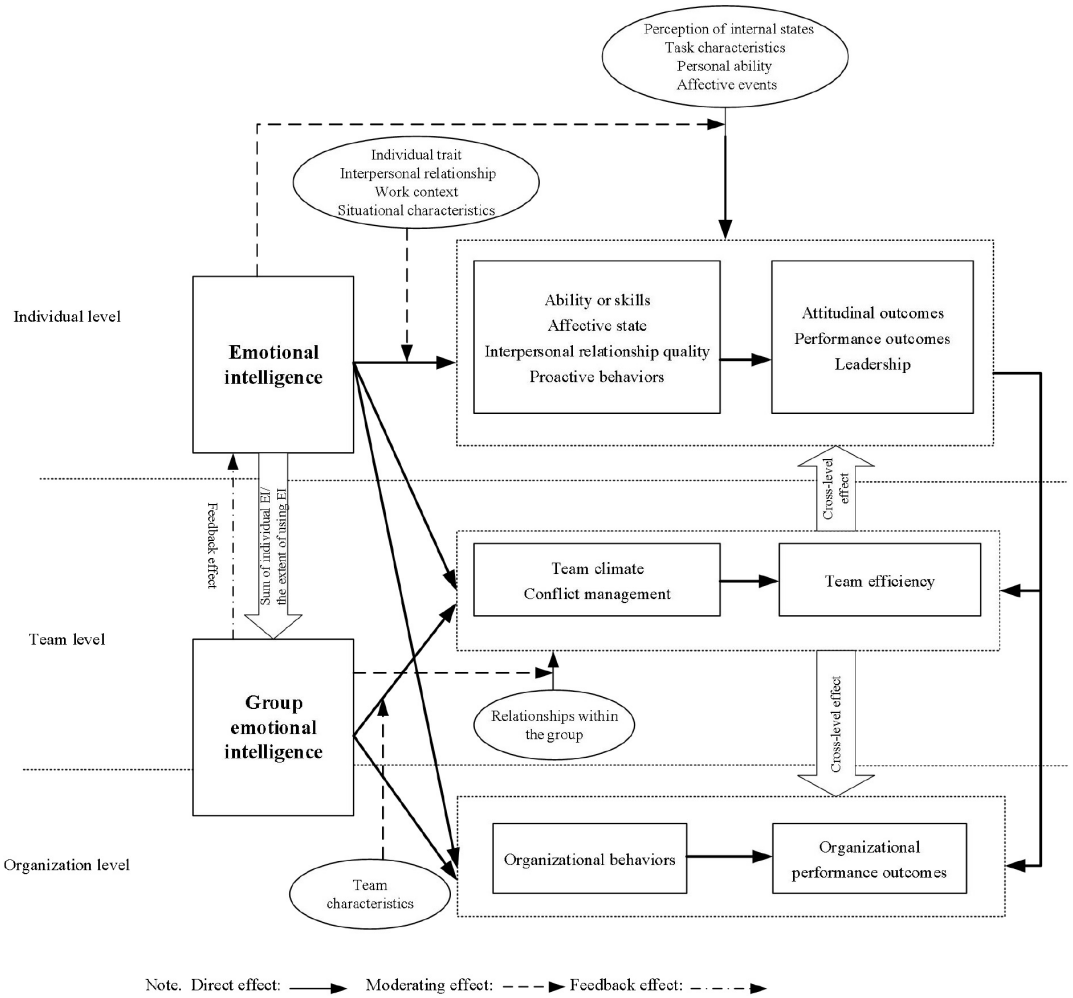


FIGURE 4 | Theoretical framework of EI.

is quite profound. Therefore, we reviewed the effect of EI on individual, team, and organizational outcomes.

Outcomes at Individual Level

As shown above, individual EI can directly affect abilities or skills, affective state, interpersonal relationship quality, and proactive behaviors at individual level.

Abilities or Skills

Individual EI can be of help to improve skills in interaction with others. The perception, understanding, and regulation of emotions are conducive to the ability to adapt to environment and stress in organizations. In addition, emotion promotion can help employees perceive problems from multiple perspectives, which may make them more willing to consider and even seek opinions from others, thus improving self-cognition and skills (Sheldon et al., 2014). Besides, EI enables individuals to have a keen understanding of interpersonal dynamics. On the one hand, it enables individuals to adjust their emotions to the environment more quickly, which helps to strengthen individual interpersonal skills and improve social and political skills (Zaccaro et al., 2018). On the other hand, it encourages individuals to more accurately identify the behavioral connotations of interactions. Dasborough and Ashkanasy (2002) believed that EI could enable individuals to more accurately perceive and interpret emotional cues, assess and classify others' motivations more accurately, and, thus, enhance their ability to perceive others' intentions.

Affective State

Individuals with high EI have a better understanding of how to regulate their emotions to achieve a favorable emotional state. Geddes and Callister (2007) pointed out after exploring emotional expression in the workplace that employees with high EI are more likely to deal with anger rationally. Emotional understanding and regulation can also speed up members' recovery from negative emotions (Shepherd, 2009) and induce employees to experience positive emotional states (Parke et al., 2015). In addition, objective appraisal and emotional knowledge encourage individuals to reappraise and regulate their emotions

in order to achieve and maintain a stable emotional state (Kidwell et al., 2008; Dust et al., 2018).

Interpersonal Relationship Quality

Most studies support the positive relationship between EI and interpersonal relationship quality (Kellett et al., 2006; Kotsou et al., 2011; Yela Aránega et al., 2020). An individual with high EI is very keen to non-verbal cues and can accurately catch the emotions of others and needs not clarified, and provide appropriate help to regulate the emotional responses of other parties, so as to achieve the purpose of strengthening relationship quality, such as establishing trust, or enhance interpersonal effectiveness (Chun et al., 2010; Farh et al., 2012; Momm et al., 2014; Vidyarthi et al., 2014).

Proactive Behaviors

Because of effective emotion regulation, members with high EI are more likely to exhibit positive behaviors than those with low EI (Kim et al., 2005). Studies argued that individuals with low EI have a weak ability in emotional awareness and emotion management, so they feel that it is more difficult to deal with the consequences of negative emotions compared with individuals with high EI (Jordan et al., 2002). Even when facing pressure, they tend to adopt more negative coping strategies (Kim et al., 2009). Conversely, individuals with high EI are able to interpret interpersonal behavior and subtle emotional cues more accurately, which means that the benefits of taking the initiative may far outweigh the cost of personal resources if they are able to solve their problems through social interaction. Therefore, motivated by resource conservation, they are more likely to engage in proactive behaviors, such as advising leaders, seeking feedback, or developing relationships with supervisors (Kim et al., 2009; Grant, 2013; Halbesleben et al., 2014). In addition, in order to develop social relationships, individuals with high EI may also adopt positive behaviors related to emotions, such as using humor to manage conflicts, adopting emotional labor strategies to achieve higher performance, and self-monitoring to establish relationships with others (Romero and Pescosolido, 2008; Caldwell, 2009; Cheung and Tang, 2009). Besides, as individuals with high EI tend to regard others' sadness and anxiety as a signal to seek help, they tend to carry out an organizational citizenship behavior (Kluemper et al., 2013).

In addition to the above direct influence of EI, individual EI can also indirectly affect subjective attitude, performance outcomes, and leadership.

Attitudinal Outcome

Studies have shown that individual EI is correlated with work attitude outcomes, including job satisfaction, organizational commitment, and personal well-being (Wong and Law, 2002; Pillai and Krishnakumar, 2019). These attitude outcomes may be self-relevant or related with others (Wong and Law, 2002; Chiva and Alegre, 2008). The organizational members' keen perception of emotional cues and reaction-centered emotional management ability will enable them to build stronger interpersonal relationships and maintain continuous positive emotional states, which will both help improve their job satisfaction and enhance their emotional commitment to an organization (Wong and Law, 2002; Dong et al., 2014). Chiva and Alegre (2008) believed that individual EI would have a positive impact on organizational learning ability, thus improving job satisfaction. Furthermore, emotional commitment to others is a necessary component of social interaction (Ashkanasy and Hooper, 1999). According to social exchange theory, members of an organization can respond to other's emotional states or behavioral information accordingly. In this sense, members with high EI can obtain desired attitudinal results by satisfying the psychological needs of other parties. For example, Wong and Law (2002) pointed out that leaders with high EI were more inclined to maintain employees' positive emotional states, which positively affected subordinates' happiness and job satisfaction.

Performance Outcomes

More and more studies have proved the positive relationship between EI and performance outcomes (Zhou and George, 2003; Vidyarthi et al., 2014; Deming, 2017), which include task performance, decision quality, creativity, and productivity. Specifically, members with high EI are good at regulating their emotions and are more likely to acquire a great deal of knowledge of how to use emotions to achieve their goals, such as participating in brainstorming in a passionate or excited mood (Gohm and Clore, 2002; Parke et al., 2015), which, in turn, is conducive to improve employees' creativity and job performance, and help members maintain a favorable emotional state. Even if they are having negative emotions, members with high EI will face the source of negative emotions and control their emotions within an appropriate threshold to cultivate more acute awareness and make wise decisions (Gohm et al., 2005). In addition, Kidwell et al. (2008) pointed out in their empirical study that individuals with high EI can maintain stable emotional states through emotional regulation, thus improving the quality of decision-making.

Because of understanding of the source of emotions, individuals with high EI are more likely to be employed in organizations that match their values (Tugade and Fredrickson, 2007; Dust et al., 2018), which means that they may be more proactive after being employed, such as taking a positive voice behavior (Grant, 2013), which helps improve their work performance. In a social context, the improvement of social skills brought by EI contributes to the exchange of heterogeneous resources and information during interaction, which will improve the efficiency of members (Kim et al., 2009; Deming, 2017). In the same context, individuals with high EI also influence others' emotional states or behavioral tendencies through exchange of social emotional resources, thus affecting others' job performance (Vidyarthi et al., 2014). For example, Zhou and George (2003) believed that leaders with high EI can accurately perceive employees' frustration at work and encourage employees to cultivate positive emotional states, thus promoting the generation of high-quality ideas and improving employees' work efficiency.

Leadership

Studies have confirmed the positive relationship between EI and leadership. Leadership refers to "a process of social interaction where leaders attempt to influence the behavior of their followers"

(Yukl et al., 2002, p. 615). When exploring the above, Walter et al. (2012) believed that individuals with high EI might be better at using emotional information to coordinate team tasks in persuasive ways, thus demonstrating their leadership. In general, current studies are mainly focused on the influence of EI on leadership effectiveness and different types of leadership, such as transformational leadership, charismatic leadership, and relationship-oriented leadership (Brown and Moshavi, 2005; Cavazotte et al., 2012).

To be specific, leaders with high EI are good at manipulating employees' perceptions with high-quality vision statements to encourage employees' commitment to a vision, thus improving the effectiveness of transformational leadership (Hur et al., 2011). Leaders with high EI can strengthen charismatic leadership by maintaining a positive attitude and effectively using impression management strategies (Walter and Bruch, 2009). Recognizing and regulating others' emotions can also promote the exchange of emotional resources between each other, which is conducive to the development of relational-oriented leadership (Kellett et al., 2006). In addition, George (2000) believed that leaders with high EI can improve their leadership efficiency by setting collective goals, instilling awareness in employees, maintaining enthusiasm, and strengthening trust. Rosette and Ciarrochi (2005) also empirically confirmed the positive correlation between leader's EI and leadership effectiveness.

Outcomes at Team Level

Both individual EI and GEI have an impact on team level outcomes. For the former, leaders' EI can influence internal process and team performance at team level. For the latter, GEI helps teams understand and respond to members' emotional responses through normative and shared behavioral patterns, and influences team effectiveness through mutual trust and group identification (Yang and Mossholder, 2004; Lee and Wong, 2017). In general, current studies show that team climate and conflict management are considered to be results of EI at the team level, and that EI may have an indirect impact on team efficiency.

Team Climate

Team climate is the emotional atmosphere that affects employees' emotional expression and experience. Emotions are often recognized as drivers of behavior and ultimately affect employee performance (Ashkanasy et al., 2017). Pirola-Merlo et al. (2002) believed that leader's EI has a strong influence on the formation of team climate. Leaders with high EI can better identify members' emotional needs and consciously manage their emotions (Eberly and Fong, 2013). Emotional expressions of a leader also affect subordinates. Therefore, leaders can create a positive emotional climate by setting good examples (Humphrey et al., 2015) or suppressing negative emotions. For example, leaders suppress self-doubt in the face of adversity in order to express positive emotions to team members (Hur et al., 2011). Wilderom et al. (2015) also believed that leaders with high EI can create or maintain a cohesive atmosphere in a team by stimulating positive group identity, establishing group norms, or encouraging team members to participate in emotional expression.

Conflict Management

It is obvious that conflict is a reflection of internal emotions in a team (Jordan and Troth, 2004). Individual or group EI may affect a conflict within a team. Leaders with high EI have a stronger perception of emotional information in a team and are able to catch conflict and tension in the team. Therefore, they may be better at correctly using the social influence brought about by power distance to effectively coordinate conflicts between members (Castro et al., 2012). In addition, because high-EI leaders can more accurately understand their inner emotions and needs, they can also set workplace norms accepted by a group, thus reducing the occurrence of team conflict and maintaining a harmonious atmosphere within a team (Wilderom et al., 2015).

Current studies show that teams with high GEI are better at managing conflict than teams with low GEI (Druskat and Wolff, 2001; Jordan et al., 2002; Jordan and Troth, 2004). Specifically, because individuals have different ideas about team tasks, team members may have perceived threats, which may lead to adverse emotional conflicts. Teams with high EI resolve these differences through open discussion and collaboration among members (Jordan and Troth, 2004). In addition, teams with high GEI can develop a set of norms to support emotional regulation to better identify conflicts in a timely manner and find different creative solutions to avoid escalation of conflicts (Yang and Mossholder, 2004; Curseu et al., 2015). For example, Lee and Wong (2017) suggested that GEI can manage task conflicts and relationship conflicts in a collaborative manner and prevent task conflicts from transforming into relationship conflicts (Curseu et al., 2015).

Team Efficiency

Druskat and Wolff (2001) proposed a team effectiveness model in the study of GEI. In this model, GEI contributes to better decisions, more creative solutions, and higher productivity. Indeed, in line with the suggestions put forward by Druskat and Wolff (2001), numerous studies have examined empirically the link between GEI and team effectiveness, and believed that there is a positive relationship between GEI and team efficiency (Hur et al., 2011; Curseu et al., 2015; Lee and Wong, 2017; Jamshed and Majeed, 2019). In order to effectively deal with emotional challenges, leader's or group's EI will assist teams to build a positive team climate and encourage proactive solving of intra-team problems; the latter is mainly manifested in the coordination of conflicts (Druskat and Wolff, 2001; Pirola-Merlo et al., 2002). Specifically, leaders or groups with high EI create an emotional atmosphere that enables members to perceive the information expected by an organization and generate corresponding emotions or motivations. For example, an open and cooperative atmosphere promotes the emergence and proliferation of new ideas. Under these circumstances, emotional contagion among members will be of help to improve team performance and enhance team creativity (Hur et al., 2011). In addition, teams and leaders with high EI are good at coordinating intra-team conflicts, reducing spread of adverse emotions, and keeping task conflicts within a favorable threshold. A moderate tension within a team is conducive to promoting team creativity and improving team effectiveness (Jamshed and Majeed, 2019).

Outcomes at Organization Level

The influence of EI on the organizational outcomes is mainly in three ways, namely, entrepreneur's EI, top management teams 'EI and high EI team affect organizational behavior respectively, thus affecting organizational performance.

Upper echelons theory holds that entrepreneurs and top management teams have a great effect on strategic behavior and performance (Hambrick and Mason, 1984). Facing high risks and uncertain environments, entrepreneurs tend to be emotional when making decisions (Cardon et al., 2012). Since entrepreneurs may have preferences or beliefs such as loss aversion and overconfidence, an entrepreneur's bias against alternative options may reduce the effectiveness of decision-making. Entrepreneurs with high EI have a broader vision and sharper insight, can objectively evaluate their behavior and market response, reduce bias, and coordinate different cognitive processes to improve decision-making quality and adaptability (Azouzi and Jarboui, 2014). Blume and Covin (2011) also pointed out that entrepreneurs with high EI have stronger intuitive judgment ability. This also means that entrepreneurs with high EI are more likely to develop effective strategies. In addition, entrepreneurs with high EI are able to more convincingly present their vision to employees and effectively manage social networks, which not only help them build and maintain trust with stakeholders but also gain access to information and resources (Azouzi and Jarboui, 2014; Naude et al., 2014).

Influencing factors of strategic decision quality of top management teams are mainly divided into two categories: one is the characteristics of senior management team, including GEI, and the other is the operation process of top management teams, including communication and coordination within a team (Hambrick and Mason, 1984). As one of the characteristics of top management teams, GEI can reduce the negative impact of emotional biases on decision-making by developing emotional norms and maintaining the dynamic stability of team emotions. Besides, GEI can also affect the operation process of top management teams, such as by enhancing intra-team communication, accurately interpreting intra-team information, and carrying out high-quality feedback loops to enhance decision-making quality (Wang, 2015).

Moderators in Emotional Intelligence Research

In recent years, more and more studies have revealed boundary conditions under which EI plays a role. It is found that moderators affect the relationship between EI and its outcomes at the individual and team levels. At the individual level, individual characteristics (including behavior tendency and trait motivation), interpersonal relationship, work background, and situational characteristics (including others' EI and GEI) may moderate relationships between individual EI and its outcomes (such as skills, emotional state, relationship quality, and proactive behavior). At the team level, team characteristics are considered to be the moderating variable of the relationship between team EI and team atmosphere and conflict (Shepherd, 2009; Wang, 2015).

Moderators at the Individual Level

Individual Trait

Individual differences may influence the relationship between EI and work outcomes (Côté and Miners, 2006; Rode et al., 2007). As the main source of individual differences, trait characteristics can affect individual behavioral tendencies or trait motivations (Wright et al., 1995; Petrides, 2009). For individuals with behavioral tendencies, high EI is more likely to lead to active behaviors. For example, Walter et al. (2012) believed that EI is more likely to trigger visible behaviors for individuals with high extroversion. Driven by a motivation, EI may have a great influence on expected results. For example, conscientiousness is seen as a motivation to achieve goals. Members with high conscientiousness are more likely to enhance the ability of EI to get rid of negative emotions, thus improving job performance (Rode et al., 2007).

Interpersonal Relationships

Empirical studies have shown that interpersonal factors can moderate the relationship between individual EI and others' behavior (Vidyarthi et al., 2014). This interpersonal relationship has been described as psychological distance from others; the stronger the relationship, the closer the psychological distance. In the interaction between leaders and employees, the high-power distance caused by difference in rank reflects the long psychological distance between leaders and employees. This psychological distance hinders the flow of social emotional resources in the interaction process, resulting in the inability to get corresponding feedback in the transmission of emotional information and limitation in the impact of leaders' emotional perception on employee behavior (Offermann and Hellmann, 1997; Vidyarthi et al., 2014; Pillai and Krishnakumar, 2019). Information theory holds that the understanding of emotional information is valuable. In highly intimate relationships, individuals with high EI perceive others' emotional information more quickly and accurately, and are more easily affected by others' emotional information. Pillai and Krishnakumar (2019) also pointed out that individuals with high EI are more likely to engage in positive social behaviors under the condition of high relationship intimacy, which will further improve their own happiness after receiving emotional feedback from another party.

Work Context

A large number of studies have found that work contexts, such as emotional labor and social needs, moderate the relationship between EI and work outcomes (Elfenbein and Ambady, 2002; Wong and Law, 2002; Vidyarthi et al., 2014). Empirical studies show that compared with low emotional labor, a work background with high emotional labor is more likely to contain emotional cues and social information, which will effectively activate individual EI (Ybarra et al., 2014). When individuals rely on others (colleagues or team leaders) and obtain support from others to complete their work with high task dependence, individual emotion recognition and emotion management will bring smoother interpersonal communication, thus affecting individual work results (Elfenbein and Ambady, 2002). When exploring the relationship between EI and job performance,

Farh et al. (2012) also pointed out that EI is more strongly correlated with job performance in a work context with high management requirements.

Situational Characteristics

Situational characteristics (which refer to the EI of another party in an interaction, or GEI) can influence the effect of EI. The interaction between individual EI and others' EI may produce complementary effects (Chun et al., 2012). The effect of EI is also affected by situational factors, and that is the EI of the other party in the interaction, or the EI of the group. Chun et al. (2010) argued that both parties with high EI have a better understanding of each other's emotions and needs, so as to timely regulate their emotional responses and promote the development of high-quality relationships such as trust. If one party has low EI, the other party with high EI can also enhance the quality of the relationship through a complementary effect. In addition, with high GEI, teams can handle negative emotions and find resources to deal with emotional recovery (Shepherd, 2009), which will strengthen the positive effect of EI of team members.

Moderators at the Team Level

Boundary conditions affecting the relationship between GEI and team-level outcomes are mainly related to team characteristics, such as number of individuals with high EI in teams and team informational diversity. Shepherd (2009) suggested in a multi- and meso-level model that the positive relationship between GEI and its speed to recover from adverse events is enhanced for teams with more emotionally intelligent members. Teams with high GEI store norms that direct members to fulfill different roles (Sameroff, 1994), while behaviors of members with high EI may change or enrich team norms and behavior patterns, which will enhance the influence of GEI (Schoka Traylor et al., 2003). In addition, these norms developed by GEI help teams to efficiently and intelligently process emotions, maintain emotionally stable communication (Druskat and Wolff, 2001); it also means that in a need of communication situations, for example, under the situation of diversity information being exchanged continually among members, there is a stronger link between EI and performance (Wang, 2015).

Moderating Effects of Emotional Intelligence

There have been studies that have used EI as a moderator in the organization field. Individual EI and GEI could moderate these relationships between predictors and work outcomes at the individual and team levels, respectively. What follows is a brief description of moderating effects of EI in different contexts.

The Moderating Effect of Emotional Intelligence at Individual Level

When EI operates as a moderator, it mainly moderates the relationship among internal state perception, personal ability, task characteristics, and emotional events on individual behaviors and work outcomes.

Internal State Perception

According to self-consistency theory (Korman, 1967; Dust et al., 2018), internal states, such as workplace anxiety, stress, and emotional state, will directly affect behaviors or work outcomes. Negative internal states can lead to negative outcomes, and EI, as the ability to monitor and manage emotional cues and facilitate emotional shift, is considered a promising source for regulating emotion (Di Fabio and Saklofske, 2021). Many scholars have proposed the importance of EI in coordinating internal states and emotional responses or behaviors (Cheng and McCarthy, 2018; Di Fabio and Kenny, 2019). Jordan et al. (2002) pointed out that members with high EI could not completely avoid the negative emotions associated with job insecurity, but that they could break the relationship between job insecurity and negative behaviors. Cheng and McCarthy (2018) also believed that members in an anxious state but with high EI are better able to understand the extent to which concerns affected their concentration and shifted their attention from the distraction to a current task. In addition, because high-EI members have more accurate emotional knowledge base and can better understand how their emotions affect their thought process and environmental perception, when they perceive negative situations such as threats or stress, they can also reduce negative effects through reappraisal (McFarland et al., 2016).

Personal Ability

Emotional intelligence can moderate the relationship between individual abilities and work outcomes. Obviously, EI and cognitive intelligence have an independent and complementary influence on job performance. When an individual's cognitive intelligence is low, EI can fill the gap by acquiring goal-related information, strengthening the quality of social relationships, and enhancing the quality of motivation and decision-making, thus positively affecting the relationship between cognitive ability and job performance. EI can also strengthen the influence of organization members on customers, because high-EI members can more accurately know customer needs and exert influence with the best solution (Kidwell et al., 2011). When EI is low, interpersonal skills do not work.

Task Characteristics

Transactional theory of stress and coping (TTSC) argues that stressful jobs are regarded as challenges or threats and lead to different emotional experiences that affect productivity (Lazarus and Folkman, 1984, 1987). This theory explains the moderating effect of EI on the relationship between task characteristics and performance. Members with high EI are more sensitive to emotional experiences at work (Clercq et al., 2014). For tasks with multiple attributes, members with high EI can effectively reappraise these tasks and convert them into motivation or opportunities, thus completing tasks more effectively (Day and Carroll, 2004). For instance, Clercq et al. (2014) proposed that employees with high EI are better able to recognize the value of goal congruence and use EI when exploring the relationship between goal congruence and organizational deviation. Such positive emotional information embedded in their relationship can reduce their possible organizational biases. Members with

high EI can also reevaluate development opportunities to eliminate negative impacts (Dong et al., 2014). Parke et al. (2015) also confirmed that EI can transform information and organizational requirements into internal incentives, thus improving employees' creativity (Thory, 2013).

Affective Events

Affective events theory proposes that accumulation of emotional events influences an individual's emotional state, which in turn influences the individual's attitude and behavior (Weiss and Cropanzano, 1996; Herman et al., 2018). As an effective theory to explain emotional fluctuations, this theory can explain the regulatory role of EI in emotional events. When an event is internally stress-induced, EI affects an individual's ability significantly to process the event (Dust et al., 2018). Lebel (2017) believes that EI can help individuals focus on future actions rather than events that trigger emotions. When facing fears, members can take advantage of e emotions, constructively guide fear emotions, and strive to seek feedback or take positive actions. When facing failure, managers with high EI may continuously face failure without hesitation, which will let them know their own capabilities (Sheldon et al., 2014). For managers with low EI, when they face negative feedback, the concern of self-protection may lead to the motivation of avoiding feedback, so they cannot timely respond fast.

The Moderating Effect of Emotional Intelligence at Team Level

Intra-group relationships not only reflect the quality of social interaction and emotional experiences in a team but also convey the social emotional information related to intra-group conflicts (Lawler, 2001; Yang and Mossholder, 2004). Since group norms developed by GEI can reflect the expected emotional tone of behavior during a conflict, GEI has been considered as a boundary condition to explain the relationship between intra-team conflicts and team effectiveness in many studies (e.g., Yang and Mossholder, 2004; Ambrosini et al., 2007; Lee and Wong, 2017). Yang and Mossholder (2004), in exploring the role of EI in the process of intra-team conflicts, indicated that teams with high GEI can keep task conflicts from spreading to an interpersonal relationship, that is, reduce links between task conflicts and relationship conflicts. In addition, GEI can effectively reduce the adverse effects of relationship conflicts and weaken the negative relationship between relationship conflicts and team performance (Koman and Wolff, 2008). Speaking of negative effects of conflicts, Ayoko et al. (2008) also showed empirically that GEI can moderate the relationship between conflicts and destructive reactions to a conflict, that, when GEI is high, the negative effects weaken. Relationships within a group are reflected not only in intra-team conflicts, but also in group affective tones. Collins et al. (2015) provided some evidence to show that a positive emotional tone positively predicts team performance when GEI is high; otherwise, it's a negative predictor.

Based on the above analysis, we propose a theoretical framework of EI in an organization, as shown in **Figure 4**. In this model, the following three aspects need to be noted: first, not all facets of EI exert an equal and consistent influence on outcomes, and the sub-dimension of EI might exert an influence independently as a single factor. Second, EI in organizations can have a cross-level effect. For example, when organizational members have a high degree of control in a team or an organization, they can influence the quality of decision-making by improving their EI, thus indirectly affecting team efficacy and organizational effectiveness. The enhancement of team effectiveness driven by GEI can also affect organizational performance. For example, a top management team will be directly affected in an organization. The improvement of the effectiveness of multiple teams with high GEI can promote organizational efficiency. In addition, a team climate created by GEI can also drive individual behaviors or emotional states across levels. Third, the feedback effect emerges in the developing process of EI. The construction of GEI relies on the aggregation of individual EI in the interaction, in turn, EI of members in emotionally intelligent teams and can be developed by emotional training or cultivated through organizational culture (Kaplan et al., 2014; Kidwell et al., 2015).

DISCUSSION AND CONCLUSION

On the basis of visualized analysis and mechanism analysis, we mainly draw the following conclusions: (1) the results of visualization show that EI has been gradually paid attention by many scholars, and that the number of articles published has been increasing; (2) the influencing mechanism of EI mainly includes self-effect mechanism and influencing others mechanism. The influence chain of "internal state monitoring – emotional information understanding – optimal strategy selection – emotional state" is adopted in the self-effect mechanism. The influencing others mechanism means that individuals manipulate others' cognition and emotion through an attribution process after external perception and application of emotional strategies; (3) EI can be divided into two categories, individual EI and group EI, and its effects are mainly at the level of individuals, teams, and organizations. Compared with the multiple effects of EI on the individual level, the team level mainly focuses on the influence of EI on the internal atmosphere or a team conflict, while the organization level mostly explores the influence of EI under the upper echelon theory. It is obvious that there are few perspectives and narrow research scope at the team and organization levels; (4) EI is mainly regarded as an independent variable or a moderating variable, and hardly as a mediator variable. It goes without saying that EI has attracted more attention from scholars, but that there are still many valuable issues that need to be paid attention in the future. Therefore, we propose possible future research directions.

First, future research can validate a more appropriate EI measurement scale. Based on the ability-based model and mixed model, current studies have proposed complex measures of EI. However, some scholars still hold critical opinions on the measurement scale of EI. Future research can supplement or revise the EI scale from the following aspects. First of all, because the concepts of EI overlap with traditional personality factors (O'Connor and Little, 2003), scholars believe that it lacks convergent and discriminant validity (Mayer et al., 2008; Joseph and Newman, 2010). The problem of validity can be solved in

the future. Besides, self-reported surveys are susceptible to social expectations, and respondents may falsely report the results. In this regard, scholars can improve the language of measurement items to reduce emotional bias or adopt group measurement and use reverse questions in the items to avoid self-reported concerns. Finally, the validity of measures of EI is called into question, because the same level of EI may trigger different behaviors in different cultures. Therefore, cultural factors can be considered in the measurement scale of EI in future studies.

Second, EI, as an important factor at multiple levels in an organization, is focused on the individual level from the perspective of social capital or social network (Momm et al., 2014; Naude et al., 2014). Seldom attention is given to the team and organizational levels. Team level studies mainly emphasize changes in the internal state of teams from the perspective of communication, while those at the organizational level mostly analyze it from the perspective of managers' emotions (Curseu et al., 2015). These above perspectives limit potential EI research. Therefore, future research should focus on antecedents of EI and check under what circumstances individual or group EI can be activated to trigger or promote corresponding behaviors. Furthermore, it is essential to track the development of EI from a dynamic perspective. Since EI can be improved through training, dynamic research on the relationship between EI and its results is helpful to characterize the specific influence of EI and explore the long-term influence mechanism of EI on organizational outcomes, which means that cross-level research on EI is of importance in the near future, and that more attention should be paid to group EI's influence on the organizational level.

Third, future research should deepen the comparative research on EI and its different dimensions. Researchers found that the effect of EI with different constructs may be quite different (Côté and Miners, 2006; O'Boyle et al., 2011) because of a discordant concept of EI. In addition, many scholars no longer explore EI as a whole variable but to study different influences of a single dimension of EI, for example, emotion regulation (Walter et al., 2012; Momm et al., 2014). However, the dimensions of EI are not completely independent of each other. Future research should explore and compare the effects of different types of EI on multiple levels. In addition, it is also meaningful to compare the influence of each dimension under different backgrounds.

Fourth, future research should focus on bidirectional effects between EI and its outcomes. Most studies argue that EI is a born factor or the result of professional training (Kaplan et al., 2014; Kidwell et al., 2015). However, EI is rooted in a special context, and the bidirectional effects of EI and its outcomes on each other are very complicated. When EI influences subsequent outcomes, it may also affect EI through feedback and recursive process. The influencing process should be longitudinal, in which we can not only address the influence of EI on its outcomes but also resolve the concerns of the effects of these outcomes on EI. The feedback path can give explanation to the bidirectional effects. Current studies on the measurement of EI use cross-section data (Jordan and Troth, 2004; Kim et al., 2009) and cannot show the longitudinal feature of EI. Future studies should focus on longitudinal designs to observe the fluctuations of EI, which will be of help to test the dynamic influence of and on EI.

Finally, future research should focus more on negative effects of EI, which may be interesting. Most studies believed that EI has a positive impact on all levels of an organization. However, other research suggests that EI may not be generally good. Higher EI may drive people to know too much about situations and respond to interfering emotional cues (Elfenbein and Ambady, 2002; Farh et al., 2012). Although scholars have pointed out possible adverse effects of EI, there is still lack of empirical studies on it. We suggest that future research should put more emphasis on the negative role of EI; one is to explore the negative relationship between EI and its outcomes, and, for example, when employees are full of emotional information, while its working performance is less related to social communication, so that employees with high EI may bring negative performance outcomes. Besides, considering that there may be multiple thresholds between EI and its outcomes, whether there is a non-linear effect is also worth further exploring, for example, during an interaction with leaders, appropriate EI displayed by employees can lead to positive feedback from leaders; whereas when leaders find that employees' EI is so high that it can even threaten their power or position, they may not give positive feedback to employees and may even show a negative response as EI increases.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

BD contributed to the mechanism, modeling, and conclusion. XP contributed to the other parts of the manuscript. NJ works with translation and revision. All authors contributed equally to the manuscript.

FUNDING

This study was financially supported by the National Natural Science Foundation of China under the Grant No. 72072068.

REFERENCES

Ambrosini, V., Bowman, C., and Burton-Taylor, S. (2007). Inter-team coordination activities as a source of customer satisfaction. *Hum. Relat.* 60, 59–98. doi: 10.1177/0018726707075283

Ashkanasy, N. M., and Hooper, G. S. (1999). Perceiving and managing emotion in the workplace: a research agenda based on neurophysiology. *Aust. J. Psychol.* 51:51.

Ashkanasy, N. M., Humphrey, R. H., and Huy, Q. N. (2017). Integrating emotions and affect in theories of management. *Acad. Manage. Rev.* 42, 175–189. doi: 10.5465/amr.2016.0474

Ayoko, O. B., Callan, V. J., and H€artel, C. E. (2008). The influence of team emotional intelligence climate on conflict and team members' reactions to conflict. *Small GR Res.* 39, 121–149. doi: 10.1177/1046496407304921

Azouzi, M. A., and Jarboui, A. (2014). CEO emotional intelligence and firms' financial policies. bayesian network method. *Contemp. Econ.* 8, 5–24.

Bar-On, R. (1997). *The Emotional Intelligence Inventory (EQ-I): Technical Manual.* Toronto, ON: Multi-Health Systems. Inc.

Bar-On, R. (2000). "Emotional and social intelligence: Insights form the Emotional Quotient Inventory (EQ-i)," in *Handbook of emotional intelligence*, eds R. Bar-On and J. D. A. Parker (San Francisco: Jossey-Bass), 363–388.

Bar-On, R. (2006). The Bar-On model of emotional-social intelligence (ESI). *Psicothema* 18, 13–25.

Beck, J. H., Ciarrochi, J., Forgas, J. P., and Mayer, J. D. (2001). *Emotional intelligence in everyday life: A scientific inquiry*. Philadelphia, PA: Psychology Press.

Bell, S. T. (2007). Deep-level composition variables as predictors of team performance: a meta-analysis. *J. Appl. Psychol.* 92, 595–615. doi: 10.1037/0021-9010.92.3.595

Blume, B. D., and Covin, J. G. (2011). Attributions to intuition in the venture founding process: do entrepreneurs actually use intuition or just say that they do? *J. Bus. Ventur.* 26, 137–151. doi: 10.1016/j.jbusvent.2009.04.002

Brackett, M. A., and Mayer, J. D. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Pers. Soc. Psychol. Bull.* 29, 1147–1158. doi: 10.1177/0146167203254596

Brackett, M. A., Mayer, J. D., and Warner, R. M. (2004). Emotional intelligence and its relation to everyday behaviour. *Pers. Individ. Differ.* 36, 1387–1402. doi: 10.1016/S0191-8869(03)00236-8

Brackett, M. A., Rivers, S. E., Shiffman, S., Lerner, N., and Salovey, P. (2006). Relating emotional abilities to social functioning: A comparison of self-report and performance measures of emotional intelligence. *J. Pers. Soc. Psychol.* 91, 780–795. doi: 10.1037/0022-3514.91.4.780

Brown, F. W., and Moshavi, D. (2005). The incubator - transformational leadership and emotional intelligence: a potential pathway for an increased understanding of interpersonal influence. *J. Organ. Behav.* 26, 867–871. doi: 10.1002/job.334

Caldwell, C. (2009). Identity, self-awareness, and self-deception: Ethical implications for leaders and organizations. *J. Bus. Ethics.* 90, 393–406. doi: 10.1007/s10551-010-0424-2

Cardon, M. S., Foo, M., Shepherd, D., and Wiklund, J. (2012). Exploring the heart: Entrepreneurial emotion is a hot topic. *Entrep. Theory Pract.* 36, 1–10. doi: 10.1111/j.1540-6520.2011.00501.x

Castro, F., Gomes, J., and de Sousa, F. C. (2012). Do intelligent leaders make a difference? the effect of a leader's emotional intelligence on followers' creativity: dointelligent leaders make a difference? *Creat. Innov. Manag.* 21, 171–182. doi: 10.1111/j.1467-8691.2012.00636.x

Cavazotte, F., Moreno, V., and Hickmann, M. (2012). Effects of leader intelligence, personality and emotional intelligence on transformational leadership and managerial performance. *Leadersh. Q.* 23, 443–455. doi: 10.1016/j.leaqua.2011.10.003

Cheng, B. H., and McCarthy, J. M. (2018). Understanding the dark and bright sides of anxiety: a theory of workplace anxiety. *J. Appl. Psychol.* 103, 537–560. doi: 10.1037/apl0000266

Cheung, F. Y., and Tang, C. S. (2009). The influence of emotional intelligence and affectivity on emotional labor strategies at work. *J. Individ. Differ.* 30, 75–86. doi: 10.1027/1614-0001.30.2.75

Cheung, S. Y., Gong, Y., and Huang, J. C. (2016). Emotional intelligence, job insecurity, and psychological strain among real estate agents: a test of mediation and moderation models. *Int. J. Hum. Resour. Manag.* 27, 2673–2694. doi: 10.1080/09585192.2015.1091369

Chiva, R., and Alegre, J. (2008). Emotional intelligence and job satisfaction: The role of organizational learning capability. *Pers. Rev.* 37, 680–701. doi: 10.1108/00483480810906900

Chowdry, K., and Newcomb, T. M. (1952). The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. Abnorm. Psychol.* 47, 57–61. doi: 10.1037/h0056411

Chun, J. U., Litzky, B. E., Sosik, J. J., Bechtold, D. C., and Godshalk, V. M. (2010). Emotional intelligence and trust in formal mentoring programs. *Group Organ. Manage.* 35, 421–455. doi: 10.1177/1059601110378293

Chun, J. U., Sosik, J. J., and Yun, N. Y. (2012). A longitudinal study of mentor and protégé outcomes in formal mentoring relationships. *J. Organ. Behav.* 33, 1071–1094. doi: 10.1002/job.1781

Clercq, D. D., Bouckenooghe, D., Raja, U., and Matsyborska, G. (2014). Unpacking the goal Congruence-Organizational deviance relationship: The roles of work engagement and emotional intelligence. *J. Bus. Ethics.* 124, 695–711. doi: 10.1007/s10551-013-1902-0

Collins, A. L., Jordan, P. J., Lawrence, S. A., and Troth, A. C. (2015). Positive affective tone and team performance: The moderating role of collective emotional skills. *Cogn. Emot.* 30, 167–182. doi: 10.1080/02699931.2015.1043857

Côté, S., Lopes, P. N., Salovey, P., and Miners, C. T. H. (2010). Emotional intelligence and leadership emergence in small groups. *Leadersh. Q.* 21, 496–508. doi: 10.1016/j.leaqua.2010.03.012

Côté, S., and Miners, C. T. H. (2006). Emotional intelligence, cognitive intelligence, and job performance. *Adm. Sci. Q.* 51, 1–28. doi: 10.2189/asqu.51.1.1

Curseu, P. L., Pluut, H., Boros, S., and Meslec, N. (2015). The magic of collective emotional intelligence in learning groups: No guys needed for the spell. *Br. J. Psychol.* 106, 217–234. doi: 10.1111/bjop.12075

Dasborough, M. T., and Ashkanasy, N. M. (2002). Emotion and attribution of intentionality in leader–member relationships. *Leadersh. Q.* 13, 615–634. doi: 10.1016/S1048-9843(02)00147-9

Davies, M., Stankov, L., and Roberts, R. D. (1998). Emotional intelligence: In search of an elusive construct. *J. Pers. Soc. Psychol.* 75, 989–1015. doi: 10.1037/0022-3514.75.4.989

Day, A. L., and Carroll, S. A. (2004). Using an ability based measure of emotional intelligence to predict individual performance, group performance and group citizenship behavior. *Pers. Individ. Differ.* 36, 1443–1458. doi: 10.1016/S0191-8869(03)00240-X

Deming, D. J. (2017). The growing importance of social skills in the labor market. *Q. J. Econ.* 132, 1593–1640. doi: 10.1093/qje/qjx022

Di Fabio, A., and Kenny, M. E. (2019). Resources for enhancing employee and organizational well-being beyond personality traits: the promise of emotional intelligence and positive relational management. *Pers. Individ. Differ.* 151:22. doi: 10.1016/j.paid.2019.02.022

Di Fabio, A., and Saklofske, D. H. (2021). The relationship of compassion and self-compassion with personality and emotional intelligence. *Pers. Individ. Differ.* 169:110109. doi: 10.1016/j.paid.2020.110109

Dong, Y., Seo, M., and Bartol, K. M. (2014). No pain, no gain: an affect-based model of developmental job experience and the buffering effects of emotional intelligence. *Acad. Manage. J.* 57, 1056–1077. doi: 10.5465/amj.2011.0687

Druskat, V. U., and Wolff, S. B. (2001). Building the emotional intelligence of groups. *Harv. Bus. Rev.* 79, 80–91. doi: 10.1177/1470595809335725

Dust, S. B., Rode, J. C., Arthaud-Day, M. L., Howes, S. S., and Ramaswami, A. (2018). Managing the self-esteem, employment gaps, and employment quality process: The role of facilitation and understanding based emotional intelligence. *J. Organ. Behav.* 39, 680–693. doi: 10.1002/job.2265

Eberly, M. B., and Fong, C. T. (2013). Leading via the heart and mind: The roles of leader and follower emotions, attributions and interdependence. *Leadersh. Q.* 24, 696–711. doi: 10.1016/j.leaqua.2013.05.003

Elfenbein, H. A. (2006). "Team emotional intelligence: What it can mean and how it can affect performance," in *Linking emotional intelligence and performance at work: Current research evidence with individuals and groups*, eds V. U. Druskat, F. Sala, and G. Mount (Mahwah, NJ: Lawrence Erlbaum), 165–184.

Elfenbein, H. A. (2008). Emotion in organizations: a review and theoretical integration. *Acad. Manag. Ann.* 1, 315–386. doi: 10.5465/078559812

Elfenbein, H. A., and Ambady, N. (2002). Predicting workplace outcomes from the ability to eavesdrop on feelings. *J. Appl. Psychol.* 87, 963–971. doi: 10.1037/0021-9010.87.5.963

Farh, Crystal, I. C., Chien, Seo, M., and Tesluk, P. E. (2012). Emotional intelligence, teamwork effectiveness, and job performance: The moderating role of job context. *J. Appl. Psychol.* 97, 890–900. doi: 10.1037/a0027377

Fineman, S. (2006). On being positive: Concerns and counterpoints. *Acad. Manage. Rev.* 31, 270–291. doi: 10.5465/AMR.2006.20208680

Forgas, J. P., and George, J. M. (2001). Affective influences on judgments and behavior in organizations: an information processing perspective. *Organ. Behav. Hum. Decis. Process.* 86, 3–34. doi: 10.1006/obhd.2001.2971

Gardner, H. (1993). *Multiple intelligences: The theory in practice*. New York, NY: Basic Books.

Geddes, D., and Callister, R. R. (2007). Crossing the line(s): A dual threshold model of anger in organizations. *Acad. Manage. Rev.* 32, 721–746. doi: 10.5465/AMR.2007.25275495

George, J. M. (2000). Emotions and leadership: The role of emotional intelligence. *Hum. Relat.* 53, 1027–1055. doi: 10.1177/0018726700538001

Gohm, C. L., and Clore, G. L. (2002). "Affect as information: An individual differences approach," in *the wisdom in feeling: Psychological processes in emotional intelligence*, eds L. F. Barrett and P. Salovey (New York, NY: Guilford Press), 89–113.

Gohm, C. L., Corser, G. C., and Dalsky, D. J. (2005). Emotional intelligence under stress: Useful, unnecessary, or irrelevant? *Pers. Individ. Differ.* 39, 1017–1028. doi: 10.1016/j.paid.2005.03.018

Goleman, D. (1995). *Emotional intelligence: Why it can matter more than IQ*. London: Bloomsbury.

Goleman, D. (1998). *Working with emotional intelligence*. London: Bloomsbury.

Grant, A. M. (2013). Rocking the boat but keeping it steady: The role of emotion regulation in employee voice. *Acad. Manage. J.* 56, 1703–1723. doi: 10.5465/amj.2011.0035

Halbesleben, J. R. B., Neveu, J., Paustian-Underdahl, S. C., and Westman, M. (2014). Getting to the COR: Understanding the role of resources in conservation of resources theory. *J. Manag.* 40, 1334–1364. doi: 10.1177/0149206314527130

Hambrick, D. C., and Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *Acad. Manage. Rev.* 9, 193–206. doi: 10.2307/258434

Herman, H. M., Troth, A. C., Ashkanasy, N. M., and Collins, A. L. (2018). Affect and leader-member exchange in the new millennium: a state-of-art review and guiding framework. *Leadersh. Q.* 29, 135–149. doi: 10.1016/j.leaqua.2017.10.002

Humphrey, R. H., Ashforth, B. E., and Diefendorff, J. M. (2015). The bright side of emotional labor. *J. Organ. Behav.* 36, 749–769. doi: 10.1002/job.2019

Humphrey, R. H., Pollack, J. M., and Hawver, T. (2008). Leading with emotional labor. *J. Manage. Psychol.* 23, 151–168. doi: 10.1108/02683940810850790

Hur, Y., van den Berg, Peter, T., and Wilderom, C. P. M. (2011). Transformational leadership as a mediator between emotional intelligence and team outcomes. *Leadersh. Q.* 22, 591–603. doi: 10.1016/j.leaqua.2011.05.002

Izard, C. E. (1993). Four systems for emotion activation: cognitive and noncognitive processes. *Psychol. Rev.* 100, 68–90. doi: 10.1037//0033-295X.100.1.68

Jamshed, S., and Majeed, N. (2019). Relationship between team culture and team performance through lens of knowledge sharing and team emotional intelligence. *J. Knowl. Manag.* 23, 90–109. doi: 10.1108/JKM-04-2018-0265

Jordan, P. J., Ashkanasy, N. M., and Hartel, C. E. (2002). Emotional intelligence as a moderator of emotional and behavioral reactions to job insecurity. *Acad. Manage. Rev.* 27, 361–372. doi: 10.5465/amr.2002.7389905

Jordan, P. J., Ashkanasy, N. M., and Härtel, C. E. (2003). The case for emotional intelligence in organizational research. *Acad. Manage. Rev.* 28, 195–197. doi: 10.2307/30040707

Jordan, P. J., and Lawrence, S. A. (2009). Emotional intelligence in teams: development and initial validation of the short version of the Workgroup Emotional Intelligence Profile (WEIP-S). *J. Manag. Organ.* 15, 452–469. doi: 10.1017/S1833367200002546

Jordan, P. J., and Troth, A. C. (2004). Managing emotions during team problem solving: emotional intelligence and conflict resolution. *Hum. Perform.* 17, 195–218. doi: 10.1207/s15327043hup1702_4

Joseph, D. L., and Newman, D. A. (2010). Emotional intelligence: an integrative meta-analysis and cascading model. *J. Appl. Psychol.* 95, 54–78. doi: 10.1037/a0017286

Kaplan, S., Cortina, J., Ruark, G., LaPort, K., and Nicolaides, V. (2014). The role of organizational leaders in employee emotion management: A theoretical model. *Leadersh. Q.* 25, 563–580. doi: 10.1016/j.leaqua.2013.11.015

Kellett, J. B., Humphrey, R. H., and Sleeth, R. G. (2006). Empathy and the emergence of task and relations leaders. *Leadersh. Q.* 17, 146–162. doi: 10.1016/j.leaqua.2005.12.003

Kidwell, B., Hardesty, D. M., and Childers, T. L. (2008). Emotional calibration effects on consumer choice. *J. Consum. Res.* 35, 611–621. doi: 10.1086/591107

Kidwell, B., Hardesty, D. M., Murtha, B. R., and Sheng, S. (2011). Emotional intelligence in marketing exchanges. *J. Mark.* 75, 78–95. doi: 10.1509/jmkg.75.1.78

Kidwell, B., Hasford, J., and Hardesty, D. M. (2015). Emotional ability training and mindful eating. *J. Mark. Res.* 52, 105–119. doi: 10.1509/jmr.13.0188

Kim, T., Cable, D. M., Kim, S., and Wang, J. (2009). Emotional competence and work performance: The mediating effect of proactivity and the moderating effect of job autonomy: emotional competence and work performance. *J. Organ. Behav.* 30, 983–1000. doi: 10.1002/job.610

Kim, T.-Y., Cable, D. M., and Kim, S.-P. (2005). Socialization tactics, employee proactivity, and person-organization fit. *J. Appl. Psychol.* 90, 232–241. doi: 10.1037/0021-9010.90.2.232

Kluemper, D. H., DeGroot, T., and Choi, S. (2013). Emotion management ability: predicting task performance, citizenship, and deviance. *J. Manag.* 39, 878–905.

Koman, E. S., and Wolff, S. B. (2008). Emotional intelligence competencies in the team and team leader: a multi-level examination of the impact of emotional intelligence on team performance. *J. Manag. Dev.* 27, 55–75. doi: 10.1108/02621710810840767

Korman, A. K. (1967). Self-esteem as a moderator of the relationship between self-perceived abilities and vocational choice. *J. Appl. Psychol.* 51, 65–67. doi: 10.1037/h0024237

Kotsou, I., Nelis, D., Grégoire, J., and Mikolajczak, M. (2011). Emotional plasticity: conditions and effects of improving emotional competence in adulthood. *J. Appl. Psychol.* 96, 827–839. doi: 10.1037/a0023047

Landy, F. J. (2005). Some historical and scientific issues related to research on emotional intelligence. *J. Organ. Behav.* 26, 411–424. doi: 10.1002/job.317

Law, K. S., Wong, C. S., and Song, L. J. (2004). The construct and criterion validity of emotional intelligence and its potential utility for management studies. *J. Appl. Psychol.* 89, 483. doi: 10.1037/0021-9010.89.3.483

Lawler, E. J. (2001). An affect theory of social exchange. *Am. J. Sociol.* 107, 321–352.

Lazarus, R. S., and Folkman, S. (1984). *Stress, appraisal and coping*. New York, NY: Springer Publishing.

Lazarus, R. S., and Folkman, S. (1987). Transactional theory and research on emotions and coping. *Eur. J. Personal.* 1, 141–169. doi: 10.1002/per.2410010304

Lebel, R. D. (2017). Moving beyond fight and flight: A contingent model of how the emotional regulation of anger and fear sparks proactivity. *Acad. Manage. Rev.* 42, 190–206. doi: 10.5465/amr.2014.0368

Lee, C., and Wong, C. (2017). The effect of team emotional intelligence on team process and effectiveness. *J. Manag. Organ.* 25, 1–16. doi: 10.1017/jmo.2017.43

Lerner, J. S., and Keltner, D. (2001). Fear, anger, and risk. *J. Pers. Soc. Psychol.* 81, 146–159. doi: 10.1037/0022-3514.81.1.146

Lindebaum, D., and Jordan, P. J. (2014). When it can be good to feel bad and bad to feel good: exploring asymmetries in workplace emotional outcomes. *Hum. Relat.* 67, 1037–1050. doi: 10.1177/0018726714535824

MacCann, C., Jiang, Y., Brown, L. E. R., Double, K. S., Bucich, M., and Minbashian, A. (2020). Emotional intelligence predicts academic performance: a meta-analysis. *Psychol. Bull.* 146, 150–186. doi: 10.1037/bul0000219

MacCann, C., Joseph, D. L., Newman, D. A., and Roberts, R. D. (2014). Emotional intelligence is a second-stratum factor of intelligence: evidence from hierarchical and bifactor models. *Emotion* 14, 358–374. doi: 10.1037/a0034755

Martins, A., Ramalho, N., and Morin, E. (2010). A comprehensive meta-analysis of the relationship between emotional intelligence and health. *Pers. Individ. Differ.* 49, 554–564. doi: 10.1016/j.paid.2010.05.029

Mayer, J. D., Caruso, D. R., and Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence* 27, 267–298. doi: 10.1016/S0160-2896(99)00016-1

Mayer, J. D., Roberts, R. D., and Barsade, S. G. (2008). Human abilities: emotional intelligence. *Annu. Rev. Psychol.* 59, 507–536. doi: 10.1146/annurev.psych.59.103006.093646

Mayer, J. D., and Salovey, P. (1997). "What is emotional intelligence?," in *Emotional development and emotional intelligence: Educational implications*, eds P. Salovey and D. Sluyter (New York, NY: Basic Books), 3–31.

Mayer, J. D., Salovey, P., and Caruso, D. R. (2000). "Models of emotional intelligence," in *Handbook of intelligence*, ed. R. J. Sternberg (Cambridge, MA: Cambridge University Press), 392–420.

McCarthy, J. M., Trougakos, J. P., and Cheng, B. H. (2016). Are anxious workers less productive workers? It depends on the quality of social exchange. *J. Appl. Psychol.* 101, 279–291. doi: 10.1037/apl0000044

McFarland, R. G., Rode, J. C., and Shervani, T. A. (2016). A contingency model of emotional intelligence in professional selling. *J. Acad. Mark. Sci.* 44, 108–118. doi: 10.1007/s11747-015-0435-8

Minbashian, A., Beckmann, N., and Wood, R. E. (2017). Emotional intelligence and individual differences in affective processes underlying task-contingent conscientiousness. *J. Organ. Behav.* 39, 1182–1196. doi: 10.1002/job.2233

Momm, T., Blickle, G., Liu, Y., Wihler, A., Kholin, M., and Menges, J. I. (2014). It pays to have an eye for emotions: Emotion recognition ability indirectly predicts annual income. *J. Organ. Behav.* 36, 147–163. doi: 10.1002/job.1975

Naude, P., Zaefarian, G., Tavani, Z. N., Neghabi, S., and Zaefarian, R. (2014). The influence of network effects on SME performance. *Ind. Mark. Manage.* 43, 630–641. doi: 10.1016/j.indmarman.2014.02.004

O'Boyle, E. H. Jr., Humphrey, R. H., Pollack, J. M., Hawver, T. H., and Story, P. A. (2011). The relation between emotional intelligence and job performance: a meta-analysis. *J. Organ. Behav.* 32, 788–818. doi: 10.1002/job.714

O'Connor, R. M. Jr., and Little, I. S. (2003). Revisiting the predictive validity of emotional intelligence: Self-report versus ability-based measures. *Pers. Individ. Differ.* 35, 1893–1902. doi: 10.1016/S0191-8869(03)00038-2

Offermann, L. R., and Hellmann, P. S. (1997). Culture's consequences for leadership behavior: national values in action. *J. Cross-Cult. Psychol.* 28, 342–351. doi: 10.1177/0022022197283008

Parke, M. R., Seo, M., and Sherf, E. N. (2015). Regulating and facilitating: the role of emotional intelligence in maintaining and using positive affect for creativity. *J. Appl. Psychol.* 100, 917–934. doi: 10.1037/a0038452

Petrides, K. V. (2009). "Psychometric properties of the Trait Emotional Intelligence Questionnaire (TEIQue)," in *Assessing emotional intelligence: Theory, research, and applications*, eds C. Stough, D. H. Saklofske, and J. D. A. Parker (New York, NY: Springer Science? Business Media), 85–101.

Petrides, K. V., and Furnham, A. (2001). Trait emotional intelligence: psychometric investigation with reference to established trait taxonomies. *Eur. J. Personal.* 15, 425–448. doi: 10.1002/per.416

Petrides, K. V., and Furnham, A. (2003). Trait emotional intelligence and happiness. *Soc. Behav. Pers.* 31, 815–823. doi: 10.2224/sbp.2003.31.8.815

Petrides, K. V., Pita, R., and Kokkinaki, F. (2007). The location of trait emotional intelligence in personality factor space. *Br. J. Psychol.* 98, 273–289. doi: 10.1348/000712606X120618

Pillai, R. G., and Krishnakumar, S. (2019). Elucidating the emotional and relational aspects of gift giving. *J. Bus. Res.* 101, 194–202. doi: 10.1016/j.jbusres.2019.03.037

Pirola-Merlo, A., Hartel, C., Mann, L., and Hirst, G. (2002). How leaders influence the impact of affective events on team climate and performance in R&D teams. *Leadersh. Q.* 13, 561–581. doi: 10.1016/S1048-9843(02)00144-3

Rode, J. C., Mooney, C. H., Arthaud-Day, M. L., Near, J. P., Baldwin, T. T., Rubin, R. S., et al. (2007). Emotional intelligence and individual performance: Evidence of direct and moderated effects. *J. Organ. Behav.* 28, 399–421. doi: 10.1002/job.429

Romero, E., and Pescosolido, A. (2008). Humor and group effectiveness. *Hum. Relat.* 61, 395–417. doi: 10.1177/0018726708088999

Rosette, D., and Ciarrochi, J. V. (2005). Emotional intelligence and its relationship to workplace performance outcomes of leadership effectiveness. *Leadersh. Org. Dev. J.* 26, 388–399. doi: 10.1108/01437730510607871

Sackett, P. R., Lievens, F., Van Iddekinge, C. H., and Kuncel, N. R. (2017). Individual differences and their measurement: A review of 100 years of research. *J. Appl. Psychol.* 102, 254–273. doi: 10.1037/apl0000151

Salovey, P., and Mayer, J. D. (1990). Emotional intelligence.*Imag. Cogn. Pers.* 9, 185–211. doi: 10.2190/DUGG-P24E-52WK-6CDG

Sameroff, A. (1994). Developmental systems and family functioning. *Exploring. Family. Rel.* 8, 199–214.

Schoka Traylor, E., Hayslip, B., Kaminski, P. L., and York, C. (2003). Relationships between grief and family system characteristics: a cross lagged longitudinal analysis. *Death Stud.* 27, 575–601. doi: 10.1080/07481180302897

Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., et al. (1998). Development and validation of a measure of emotional intelligence. *Pers. Individ. Differ.* 25, 167–177. doi: 10.1016/S0191-8869(98)00001-4

Sheldon, O. J., Dunning, D., and Ames, D. R. (2014). Emotionally unskilled, unaware, and uninterested in learning more: reactions to feedback about deficits in emotional intelligence. *J. Appl. Psychol.* 99, 125–137. doi: 10.1037/a0034138

Shepherd, D. A. (2009). Grief recovery from the loss of a family business: a multi- and meso-level theory. *J. Bus. Ventur.* 24, 81–97. doi: 10.1016/j.jbusvent.2007.09.003

Thorndike, E. L. (1920). Intelligence and its uses. *Harpers. Magazine.* 140, 227–235. doi: 10.1234/12345678

Thory, K. (2013). Teaching managers to regulate their emotions better: insights from emotional intelligence training and work-based application. *Hum. Resour. Dev. Int.* 16, 4–21. doi: 10.1080/13678868.2012.738473

Troth, A. C., Jordan, P. J., Lawrence, S. A., and Herman, H. M. T. S. E. (2012). A multilevel model of emotional skills, communication performance, and task performance in teams. *J. Organ. Behav.* 33, 700–722. doi: 10.1002/job.785

Tugade, M. M., and Fredrickson, B. L. (2007). Regulation of positive emotions: Emotion regulation strategies that promote resilience. *J. Happiness Stud.* 8, 311–333. doi: 10.1007/s10902-006-9015-4

Tuncdogan, A., Acar, O. A., and Stam, D. (2017). Individual differences as antecedents of leader behavior: Towards an understanding of multi-level outcomes. *Leadersh. Q.* 28, 40–64. doi: 10.1016/j.leaqua.2016.10.011

Van Rooy, D. L., and Viswesvaran, C. (2004). Emotional intelligence: a meta-analytic investigation of predictive validity and nomological net. *J. Vocat. Behav.* 65, 71–95. doi: 10.1016/S0001-8791(03)00076-9

Vidyarthi, P. R., Anand, S., and Liden, R. C. (2014). Do emotionally perceptive leaders motivate higher employee performance? the moderating role of task interdependence and power distance. *Leadersh. Q.* 25, 232–244. doi: 10.1016/j.leaqua.2013.08.003

Walter, F., and Bruch, H. (2009). An affective events model of charismatic leadership behavior: a review, theoretical integration, and research agenda. *J. Manag.* 35, 1428–1452. doi: 10.1177/0149206309342468

Walter, F., Cole, M. S., Van Der Vegt, G. S., Rubin, R. S., and Bommer, W. H. (2012). Emotion recognition and emergent leadership: unraveling mediating mechanisms and boundary conditions. *Leadersh. Q.* 23, 977–991. doi: 10.1016/j.leaqua.2012.06.007

Wang, S. (2015). Emotional intelligence, information elaboration, and performance: The moderating role of informational diversity. *Small Group Res.* 46, 324–351. doi: 10.1177/1046496415578010

Weiss, H. M., and Cropanzano, R. (1996). Affective events theory: a theoretical discussion. *Res. Organ. Beh.* 18, 17–19.

Wilderom, C. P. M., Hur, Y., Wiersma, U. J., Peter, T., and Lee, J. (2015). From manager's emotional intelligence to objective store performance: through store cohesiveness and sales-directed employee behavior. *J. Organ. Behav.* 36, 825–844. doi: 10.1002/job.2006

Wolff, S. B., Pescosolido, A. T., and Druskat, V. U. (2002). Emotional intelligence as the basis of leadership emergence in self-managing teams. *Leadersh. Q.* 13, 505–522. doi: 10.1016/S1048-9843(02)00141-8

Wong, C., and Law, K. S. (2002). The effects of leader and follower emotional intelligence on performance and attitude: an exploratory study. *Leadersh. Q.* 13, 243–274. doi: 10.1016/S1048-9843(02)00099-1

Wright, P. M., Kacmar, K. M., McMahan, G. C., and Deleeuw, K. (1995). P= f(MXA): cognitive ability as a moderator of the relationship between personality and job performance. *J. Manag.* 21, 1129–1139. doi: 10.1016/0149-2063(95)90025-X

Yang, J., and Mossholder, K. W. (2004). Decoupling task and relationship conflict: The role of intragroup emotional processing. *J. Organ. Behav.* 25, 589–605. doi: 10.1002/job.258

Ybarra, O., Kross, E., and Sanchez-burks, J. (2014). The "big idea" that is yet to be: toward a more motivated, contextual, and dynamic model of emotional intelligence. *Acad. Manag. Perspect.* 28, 93–107. doi: 10.5465/amp.2012.0106

Yela Aránega, A., Del Val Núñez, M. T., and Castaño Sánchez, R. (2020). Mindfulness as an intrapreneurship tool for improving the working environment and self-awareness. *J. Bus. Res.* 115, 186–193. doi: 10.1016/j.jbusres.2020.04.022

Yukl, G., Gordon, A., and Taber, T. (2002). A hierarchical taxonomy of leadership behavior: integrating a half century of behavior research. *J. Leadersh. Organ. Stud.* 9, 15–32. doi: 10.1177/107179190200900102

Zaccaro, S. J., Green, J. P., Dubrow, S., and Kolze, M. (2018). Leader individual differences, situational parameters, and leadership outcomes: a comprehensive review and integration. *Leadersh. Q.* 29, 2–43. doi: 10.1016/j.leaqua.2017.10.003

Zhou, J., and George, J. M. (2003). Awakening employee creativity: the role of leader emotional intelligence. *Leadersh. Q.* 14, 545–568. doi: 10.1016/S1048-9843(03)00051-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Dong, Peng and Jiang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 08 March 2022
doi: 10.3389/fpsyg.2022.830342

A Study on the Cognition and Emotion Identification of Participative Budgeting Based on Artificial Intelligence

*Yuan Zhou[1], Tianjiao Zhang[1], Lan Zhang[1], Zhaoxin Xue[1], Mingxu Bao[2] and Lingbing Liu[3]**

[1] College of Accounting, Jilin University of Finance and Economics, Changchun, China, [2] School of Innovation and Entrepreneurship, Changchun University of Chinese Medicine, Changchun, China, [3] School of Accounting, Dongbei University of Finance and Economics, Dalian, China

Cognition and emotion exert a powerful influence on human behavior. Based on cognitive psychology and organizational behavior theory, this paper examines the role of cognition and emotion in participative budgeting and corporate performance using a questionnaire survey. The questionnaires were sent to 345 listed companies in China. The results support the hypothesis that human cognition and emotion have a positive moderating effect on the relationship between participative budgeting and corporate performance. Cognition and emotion can promote the effect of participative budgeting on corporate performance. Furthermore, according to the theory of artificial intelligence (AI), this paper designs an AI-based cognition and emotion identification system. This system can help managers identify the budget participants' cognitive and emotional states and undertake the interventions necessary to improving corporate performance.

Keywords: cognition, emotion, artificial intelligence, participative budgeting, performance

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Yongjun Sun,
Heilongjiang University, China
Xiuwei Chu,
Southwest University, China

****Correspondence:***
Lingbing Liu
liulingbing@dufe.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *07 December 2021*
Accepted: *16 February 2022*
Published: *08 March 2022*

Citation:
Zhou Y, Zhang T, Zhang L, Xue Z, Bao M and Liu L (2022) A Study on the Cognition and Emotion Identification of Participative Budgeting Based on Artificial Intelligence. Front. Psychol. 13:830342. doi: 10.3389/fpsyg.2022.830342

INTRODUCTION

In the postepidemic era, enterprise risk management has become key to achieving sustainable development (Yu et al., 2020). Budgeting is an important tool for enterprise risk control. Participative budgeting prefers that employees at all levels of the organization participate in the preparation and evaluation of the budgeting. As a multiparty budgeting system, the results are often multifaceted. Most studies on the relationship between participative budgeting and corporate performance focus on economics and management, and many scholars believe that participative budgeting can promote corporate performance (Peter and Morris, 1986). However, some scholars put forward a different view, that participative budgeting is negatively correlated or not significantly correlated with corporate performance (Cherrington and Cherrington, 1973). This indicates that there may be a complex interactive mechanism between participative budgeting and corporate performance.

Cognition and emotion may play an important role in the relationship between participative budgeting and corporate performance. As a management tool, participative budgeting is closely related to the participants (Milani, 1975), and the participants' psychological characteristics affect the result of participative budgeting (Shields and Shields, 1998). Cognition and emotion are

important psychological links between organizations and employees. People with high cognitive and emotional abilities are more likely to take positive actions. Employee cognition can affect the identification and transmission of information in budget participation and then influence the quality of decision-making. Employees with great emotional ability usually participate actively in the formulation and implementation of budgeting. Therefore, cognition and emotion may play an important role in the relationship between participative budgeting and corporate performance. Exploring how of cognition and emotion affect participative budgeting and corporate performance will play a positive role in improving corporate performance.

With the development of science, it has become possible to identify human cognition and emotion through artificial intelligence (AI). Connectionism holds that cognition arises from the human nervous system so cognitive processes can be simulated by using mathematical models of the nervous system. At present, researchers have constructed a variety of cognitively and emotionally intelligent identification models, such as the CogAff model (Solman, 2004), EI model (Ryan et al., 2018), MEM model (Yoichiro et al., 2020) among others. These models can imitate cognition and emotion to some extent, and provide a reference for research on the identification of cognition and emotion. Therefore, this paper attempts to construct an AI-based cognition and emotion identification system. Meanwhile, the cognition and emotion of participants in budgeting can be reasonably interfered with through this system to promote corporate performance.

In summary, our study has two contributions. First, we divide cognition and emotion into three important parts, cognitive ability, emotional intelligence and self-efficacy, and verify their positive moderating effects on corporate performance in participative budgeting by empirical analysis. Our results can extend the research scope of cognitive psychology and organizational behavior. Second, an AI-based cognition and emotion identification system is constructed to identify the cognitive and emotional processes of participants in budgeting. Corresponding management suggestions are put forward by this system. Our aim is to help managers identify their employees' cognitive and emotional problems and implement more valid solutions. The main characteristics of the system are as follows:

(1) A questionnaire data processing system (QDP) is used for matching questionnaire data and cognitive and emotional labels to identify the cognitive and emotional processes of participants in budgeting.
(2) An artificial neural network (ANN) is used for extracting the features of cognition and emotion of participants in budgeting, and a production expert system (PES) is used for identifying cognitive and emotional problems and providing solutions.
(3) The management application innovation system (MAI) combines the solutions provided by machine analysis with the manager's own wisdom.

THEORETICAL ANALYSIS OF THE ROLE OF COGNITION AND EMOTION IN ORGANIZATIONAL BEHAVIOR AND PERFORMANCE

Cognition, Emotion and Organizational Behavior

In the continuous development of corporate management practice, managers have been aware that it is impossible to arouse the enthusiasm of workers simply by adopting classical management theory. Exploring the emotional basis in management, that is, seeking the relationship between emotion and behavior in organizations, should be of great importance.

The change in behavior is a process that includes three stages: "knowing" (cognition), "feeling" (emotion) and "action" (behavior) (Westbrook and Oliver, 1991). Cognition is the process of acquiring knowledge and information processing, which is the most basic psychological process of human beings. The human brain accepts the information input from the outside world. Then, the information is processed by the brain and converted into mental activities, thus dominating human behavior. It is the process of information processing, also known as cognitive process. Emotion is a complex and stable physiological evaluation and experience of attitude dominated by cognition. When individuals' desires and needs are met, it causes positive emotions and thus brings changes to their behaviors. The process of emotion formation occurs through the processes of cognition, restraint and motivation. Restraint is the expression of emotional intelligence. Emotion can motivate individuals by guiding and promoting them to achieve predetermined goals, and the strength of this effect is closely related to self-efficacy. A survey of 77 senior managers found that the more intense the competition in the market is, the more positively correlated the degree of budget participation with management performance and job satisfaction (Vincent et al., 2005). Budget participation enhances communication among group members. Participation enables members to fully understand and share the goals and values of the organization, arouses the sense of responsibility of employees, enhances the cohesion of the team, and thus enhances their commitment to the organization (DeCotiis and Summers, 1987). Therefore, it can be inferred that cognition and emotion can affect individual behavior in the organization.

Organizational Behavior and Organizational Performance

An organization is a systematic arrangement of people to accomplish a specific mission. With increasingly advanced technical means of interpersonal communication, the spatial barriers between people are getting smaller, which significantly improves the effectiveness of contact between people. Along with the increasing socialization of services, the form and degree of interactions between individuals and organizations are complicated and diversified (Zhang, 2011). Modern management has developed from "material-oriented" management to a "people-oriented" one, from a "discipline" study to a

"human behavior" one and from "supervision" management to "incentive" management.

Based on modern organizational behavior theory, participative budgeting is a decentralized and bidirectional management mode in which all the staff actively participate in the company budget. By participating in the budgeting process, employees can integrate their own goals into organizational goals. This fully stimulates the creativity and work enthusiasm of employees and enhances their sense of mission within the organization to improve corporate performance. However, in practice, the effect of participative budgeting on corporate performance is complicated. Human behavior is the manifestation of human psychology, which is dominated by cognition and emotion. People with different cognitive and emotional states tend to behave differently within organizations. Correspondingly, behavioral consequences—the impact on corporate performance— varies. Cognitive and emotional factors are not only the results of the influence of organization on individuals but also the internal source of the effect of individual behavior on organizational performance. Therefore, to influence the behavior of employees, the cognitive and emotional factors of participants need to be considered. Sensible psychological intervention can contribute to taking full advantage of the positive role of participative budgeting.

EMPIRICAL ANALYSIS AND TEST RESULTS OF COGNITIVE AND EMOTIONAL INFLUENCES ON PARTICIPATIVE BUDGETING

Research Hypotheses

Participative Budgeting and Corporate Performance

The discussion on participative budgeting can be traced back to the 1950s (Argyris, 1952). Most scholars have focused on the relationship between participative budgeting and corporate performance. The study found that employees could have more choices in participative budgeting and feel that they can create a working atmosphere of joint efforts. According to self-determination theory, this voluntary participation can motivate employees to contribute to the budget. In this process, employees become more willing to think deeply and formulate strategic budget goals that are conducive to the long-term development of the organization. As a result, it is likely to reinforce job satisfaction and enhance corporate performance (Chenhall, 1986; Wu et al., 2019; Li et al., 2021). Some scholars believe that the effects of participative budgeting on corporate performance are probably influenced by mediating variables or moderating variables (Li et al., 2021). A study of manufacturing in Hong Kong found that budget participation is positively correlated with management performance under a high level of decentralization (Gul et al., 1995), and this positive correlation is also moderated by budget incentives (Wu et al., 2011). As employees participate in budgets, the restraining effect of inconsistencies in budget participation on performance becomes weaker; that is, corporate performance improves with the decrease in budget participation inconsistency (Clinton, 1999). Other studies have found that involving employees in budgeting can alleviate organizational dysfunction caused by budget compulsion and thus indirectly improve corporate performance (Li and Miao, 2013). Thus, building on the above arguments, we propose that:

> H1: Participative budgeting has a significant positive effect on corporate performance.

The Moderating Effect of Cognition and Emotion

Processes of cognition and emotion include cognition, constraint and motivation. People's different emotional capacities may play different roles in the influence of participative budgeting and corporate performance. Therefore, this paper divides cognition and emotion into three aspects, namely, cognitive ability, emotional intelligence and self-efficacy.

As a psychological factor with individual differences, cognitive ability is the ability to process, store and extract information. Employee cognitive ability refers to the ability of employees to accurately identify useful information at work and apply it to decision-making. In people-oriented management, the level of cognitive ability could exert a far-reaching impact on the perception of others' needs, the maintenance of interpersonal relationships and even the enhancement of team cohesion. Rich cognitive perspectives and information sources can significantly improve employees' ability to creatively propose solutions to tackle problems. Cognition consists of a variety of elements, such as thought, need, attitude and belief. When these elements conflict, cognitive dissonance will occur. Cognitive dissonance makes employees feel strong work pressure, resulting in low work efficiency and a high turnover rate (Festinger, 1957). The study found that the participation of team members in the discussion is conducive to enhancing the overall complexity of cognition. This enables them to think more deeply from multiple perspectives before making final decisions, thereby improving cognitive consistency and decision-making efficiency (Gruenfeld and Hollingshead, 1993; Liu et al., 2020). In general, people with higher cognitive abilities more precisely perceive changes in interpersonal relationships, which is beneficial for further coordination with others. Creating a good interpersonal atmosphere in budget participation is conducive to realizing team cooperation and improving corporate performance (Che and Qin, 2009). Employees with great cognitive ability can better grasp the nature of the budget goals. They can combine the current situation and their own capabilities to make the budget goals more objective and feasible (Li et al., 2020). The participation of employees with great cognitive ability in budgeting can promote a team's cognitive awareness of the organization's goals. In this way, the efficiency of resource allocation and budget transparency can be improved, which reduces unnecessary conflicts (Yang and He, 2020). According to these arguments, we propose the following:

> H2a: Cognitive ability positively moderates participative budgeting and corporate performance.

Another important process in cognition and emotion is emotional restraint. Emotional self-discipline refers to the

effective control of emotions, which is shown as emotional intelligence. Research has shown that emotional intelligence, based on individuals or the entire team, plays an important role in achieving strategic goals (Zhang, 2012). Moderate emotional excitement can put people in an optimal state, both physically and mentally, and its physiological activation function can lead people to achieve the highest level of performance (Hebb, 1955). As an important psychological feature, emotional intelligence helps individuals maintain good emotional states owing to its effect on emotional regulation. Emotional intelligence is key for individuals to fully utilize their abilities. Emotion can induce behavior through the arousal function. A moderate level of arousal can motivate individuals to achieve set goals and improve their work performance (Yerkes and Dodson, 1908). The level of employees' emotional intelligence determines their work performance to a certain extent. Employees with high emotional intelligence are capable of accurately detecting, integrating and evaluating information, so they can deeply perceive the budget information required by managers. This makes it easy for them to deliver the results the organization expects (Wong and Law, 2002; Wei et al., 2015). It is found that employees with low emotional intelligence have poorer information absorption and summary abilities. As a result, they have difficulties in effectively exchanging information in the budgetary participation process, which hinders the implementation effect of participative budgeting. Emotional intelligence involves the ability of emotions to guide individuals to achieve predetermined goals. Employees with high emotional intelligence can mobilize and direct their emotions to make them subordinate to a given goal. Therefore, they are not easily affected by negative emotions but actively look for ways to solve problems. In this way, they can better complete the budget targets and promote corporate performance (Goleman et al., 2002). Emotional intelligence develops gradually with the growth of people and can be improved through training and intervention (Bar-On and Parker, 2000). Employees use emotional intelligence, the bond between the emotional system and cognitive system, to manage their emotions, which could exert a certain influence on behavior in the workplace (Liu, 2018). Thus, we propose the following:

> H2b: Emotional intelligence positively moderates participative budgeting and corporate performance.

With an important role moderating the effect of cognition and emotion on behavior, emotion-based motivation affects adherence to one's ideal goals and helps overcome negative emotions. This is called self-efficacy. Self-efficacy is "an individual's definite belief in his or her ability to successfully accomplish a specific task in a specific situation." It can mobilize motivation, cognitive resources and a set of actions (Fred, 2003). The study found that employees with high self-efficacy are more actively involved in their work compared to employees with low self-efficacy. They tend to set challenging goals and maintain high levels of commitment while persevering despite difficulties, thus showing better performance (Bandura, 1986; Zhao et al., 2018). A high level of self-efficacy means that employees are confident in their ability to overcome difficulties and have a strong willingness to participate in new things (Lu et al., 2001). Based on this, high self-efficacy employees demonstrate high loyalty and are willing to share new knowledge and experience. When employees with a strong willingness to share and high working ability participate in budget preparation, the exchange, sharing and integration of budget information will be more effective. This makes budget planning more objective and feasible and thus promotes corporate performance. At the same time, some studies have shown that employees with high self-efficacy are often less susceptible to external influences and are more willing to set challenging goals and later put them into practice. They are usually able to withstand challenging workloads and better adapt to organizational innovations, thus better achieving budget goals (Seijts and Latham, 2011; Li and Guo, 2021). Employees with high self-efficacy tend to be passionate about their jobs, which strengthens goal commitments and eventually corporate performance. Accordingly, we propose the following:

> H2c: Self-efficacy positively moderates participative budgeting and corporate performance.

Measurement of Variables

There are five latent variables in this paper, namely, participative budgeting, cognitive ability, emotional intelligence, self-efficacy and corporate performance. Maturity scales are used.

1. Participative budgeting (PB). Participative budgeting was measured by a six-item scale adapted from Maiga (2005), as shown in **Table 1**.
2. Cognition and emotion mechanism. The cognition and emotion mechanism consists of cognitive ability, emotional intelligence and self-efficacy.

(1) Cognitive ability (CA)

According to Spreitzer, cognitive ability was divided into three dimensions with nine items (Spreitzer, 1995), as shown in **Table 2**.

(2) Emotional Intelligence (EI)

Based on mixed model theory, this paper adopts the scale of Law, K. S, to divide emotional intelligence into four dimensions with 16 items (Law et al., 2004). The specific items are shown in **Table 3**.

(3) Self-efficacy (SE)

According to the definition of self-efficacy (Yao, 2008), this paper adopts the scale of Yao, K, which contains 11 items in three dimensions. The specific items are shown in **Table 4**.

3. Corporate performance (CP). According to Green and Inman (2005), corporate performance is divided into two dimensions, marketing performance and financial performance, with 7 items. The specific items are shown in **Table 5**.

Model Construction

Based on the above theoretical analysis and deduction, this paper constructs the following theory model. As shown in **Figure 1**.

TABLE 1 | Item of participative budgeting.

Variable	Items
PB (Maiga, 2005)	I am involved in setting all of my budget.
	My superior clearly explains budget revisions.
	I have frequent budget-relate discussions with my superior.
	I have a great deal of influence on my final budget.
	My contribution to the budget is very important.
	My superior initiates frequent budget discussion when the budget is being prepared.

TABLE 2 | Item of cognitive ability.

Variable		Items
CA (Spreitzer, 1995)	Meaning	The work I do is very important to me.
		My job activities are personally meaningful to me.
		The work I do is meaningful to me.
	Competence	I am confident about my ability to do my job.
		I am self-assured about my capabilities to perform my work activities.
		I have mastered the skills necessary for my job.
	Self-Determination	I have significant autonomy in determining how I do my job.
		I can decide on my own how to go about doing my work.
		I have considerable opportunity for independence and freedom in how I do job.

TABLE 3 | Item of emotional intelligence.

Variable		Items
EI (Law et al., 2004)	Self-Emotions Appraisal (SEA)	I have a good sense of why I have certain feelings most of the time.
		I have good understanding of my own emotions.
		I really understand what I feel.
		I always know whether or not I am happy.
	Regulation of Emotion (ROE)	I am able to control my temper so that I can handle difficulties rationally.
		I am quite capable of controlling my own emotions.
		I can always calm down quickly when I am very angry.
		I have good control of my own emotions.
	Use of Emotion (UOE)	I always set goals for myself and then try my best to achieve them.
		I always tell myself I am a competent person.
		I am a self-motivating person.
		I would always encourage myself to try my best.
	Others-Emotions Appraisal (OEA)	I always know my friends' emotions from their behavior.
		I am a good observer of others' emotions.
		I am sensitive to the feelings and emotions of others.
		I have good understanding of the emotions of people around me.

Questionnaire Design and Data Collection

This paper adopts the questionnaire survey method to collect data, distributed to listed companies in China. We use a five-level scale to score and measure the items. To ensure the accuracy of the questionnaire items, during the formulation of the questionnaire we first invited experts to make preliminary modifications to the questionnaires. Then, we interviewed employees in the enterprise for a second revision. Finally, a small number of MBA students and senior managers of enterprises were tested to form the final draft. A total of 345 questionnaires were sent out, and 320 responses were received. There were 305 valid questionnaires for an effective recovery rate of 88.4%.

Data Quality Analysis

Homologous Deviation Analysis

Harman's single-factor test was used for homologous bias analysis. Unrelated exploratory factor analysis was performed on the measurement items of all variables by SPSS17.0. The results show that there are 13 factors with eigenvalues greater than 1, and the total variance explained by the first factor was 14.418%, less than the critical value of 40%. Therefore, it can be shown that the homology deviation of the survey data does not affect the final data analysis results.

TABLE 4 | Item of self-efficacy.

Variable		Items
SE (Yao, 2008)	Work Cognition and Motivation Efficacy	I feel that I was right to be in this profession at the beginning.
		Facing new tasks, I am always confident.
		I like to use new methods or techniques to complete work.
	Interpersonal Communication and Recovery Efficacy	Even in a new or unfamiliar working environment, I don't feel very anxious.
		I can take the initiative to communicate with people around me, whether I am familiar with him or not.
		I am in good health and can reasonably regulate the emotional problems that arise at work.
		Even if I encounter difficulties in interpersonal communication, I have the confidence to solve it through my own efforts.
	Stress and Innovation Ability Efficacy	When faced with a problem, I can usually find a new and creative solution to the problem.
		At work, I can often put forward reasonable and innovative ideas for the project.
		With my wit, I believe I can deal with unexpected situations.
		I can turn innovative ideas into practical applications.

TABLE 5 | Item of corporate performance.

Variable		Items
CP (Green and Inman, 2005)	Financial Performance	Average return on investment over the past 3 years.
		Average profit over the past 3 years.
		Profit growth over the past 3 years.
		Average return on sales over the past 3 years.
	Marketing Performance	Average market share growth over the past 3 years.
		Average sales volume growth over the past 3 years.
		Average sales (in dollars) growth over the past 3 years.

Reliability, Validity and Correlation Test

Table 6 shows descriptive statistics and correlation analysis among variables, namely, the mean value, standard deviation and correlation coefficient. The results showed that participative budgeting (PB), cognitive ability (CA), emotional intelligence (EI), self-efficacy (SE) and corporate performance (CP) were positively correlated. The results provide support for related hypothesis research.

To test the reliability and validity of all variables, SPSS 22.0 and AMOS 21.0 software were used to test the data. First, the reliability of the data was tested. The results in **Table 7** show that the Cronbach's α value of each variable was greater than 0.7, and the combined reliability CR value was also greater than 0.7, indicating that each variable scale was highly reliable and valid. Second, the validity of the data was tested. In this study, the data were tested by concorporateatory factor analysis (CFA). The results showed that the factor loading value ranged from 0.552 to 0.895, and the average variance extraction value (AVE value) was greater than 0.5, indicating that the scale has good validity. The square root of each variable is greater than the correlation coefficient, indicating that the validity of the scale is high.

Moderating Effect Test

In this paper, model 1, 2, 3 and 4 were constructed according to the following formulae to test the moderating effect.

$$CP_i = \beta_{11}+\beta_{12}Controls+\varepsilon_i \quad (1)$$

$$CP_i = \beta_{21}+\beta_{22}PD_i+\beta_{23}Controls+\varepsilon_i \quad (2)$$

$$CP_i = \beta_{31}+\beta_{32}PD_i+\beta_{33}CA_i+\beta_{34}EI_i+\beta_{35}SE_i+\beta_{36}Controls+\varepsilon_i \quad (3)$$

$$CP_i = \beta_{41}+\beta_{42}PD_i+\beta_{43}CA_i+\beta_{44}EI_i+\beta_{45}SE_i+\beta_{46}PD_i \times CA_i+\beta_{47}PD_i \times EI_i+\beta_{48}PD_i \times SE_i+\beta_{49}Controls+\varepsilon_i \quad (4)$$

In this study, multiple linear regression was used to test the moderating effects of cognitive ability (CA), emotional intelligence (EI), and self-efficacy (SE) on participative budgeting (PB) and corporate performance (CP) (Models 2, 3, and 4). Model 1 shows the relationships between the control variables and dependent variables. According to **Table 8**, the effect of participative budgeting on corporate performance (Model 2, $\beta = 0.248$, $p < 0.001$) had a significant positive promoting effect. H1 was verified. Model 3 ($\beta = 0.116$, $\beta = 0.140$, $p < 0.05$ and $\beta = 0.212$, $p < 0.001$) and Model 4 ($\beta = 0.122$, $\beta = 0.116$ and $\beta = 0.120$, $p < 0.05$) demonstrated the moderating effects of cognitive ability (CA), emotional intelligence (EI) and self-efficacy (SE), which played a positive moderating role between

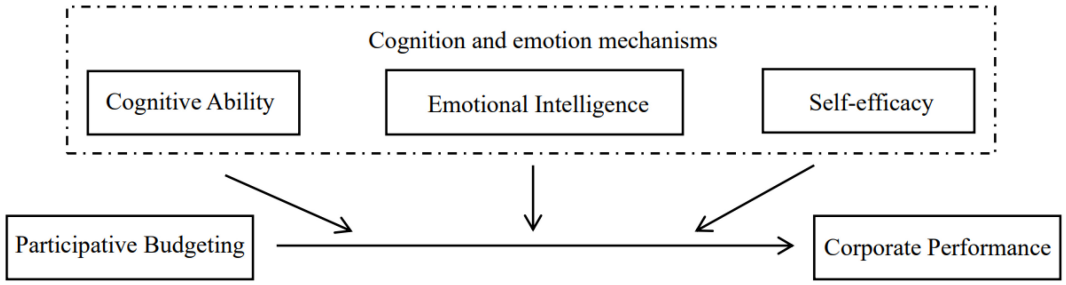


FIGURE 1 | Theory model of cognition and emotion.

TABLE 6 | Mean, SD, and correlation coefficient of variables.

Variables	Mean	SD	1	2	3	4	5	6	7	8
1 Corporate scale	3.40	0.84	1							
2 Position of Respondents	4.11	0.84	−0.188**	1						
3 Office term	2.16	1.01	−0.040	−0.086	1					
4 PB	5.08	0.83	−0.062	0.070	−0.001	1				
5 CA	5.11	0.62	−0.030	0.125*	0.039	0.053	1			
6 EI	5.22	0.67	−0.095	0.173**	0.107	0.148**	0.295**	1		
7 SE	5.25	0.59	0.005	0.026	0.017	0.314**	−0.051	−0.003	1	
8 CP	5.36	0.71	−0.192**	−0.011	0.139*	0.255**	0.152**	0.209**	0.254**	1

*$p < 0.05$ and **$p < 0.01$.

participative budgeting (PB) and corporate performance (CP) and thus verified H2a, H2b and H2c, as shown in **Table 8**.

The Empirical Results

Hypothesis 1, Hypothesis 2a, Hypothesis 2b and Hypothesis 2c in this study are all valid, which proves that cognitive ability, emotional intelligence and self-efficacy have a positive moderating effect on the relationship between participative budgeting and corporate performance. That is, cognitive and emotional factors can promote the positive effect of participative budgeting on corporate performance. It provides a theoretical basis for constructing an artificial intelligence identification system of cognition and emotion.

ARTIFICIAL INTELLIGENCE IDENTIFICATION SYSTEM OF COGNITION AND EMOTION BASED ON PARTICIPATIVE BUDGETING

Overview of the Development of Artificial Intelligence

In recent decades, artificial intelligence technology has made great progress. Artificial intelligence (AI) began as a simple computer program in the 1950s and gradually evolved to be capable of automatic recovery and self-analysis. By the 1980s, AI gained learning and cognitive abilities (Yao, 2018). Since the beginning of the 21st century, the development of society has shown an accelerating trend of increased connection between sectors due to the widespread use of the internet. The exponential rise of the scale of data creation in the new century has shown a new era of big data thanks to the rapid development of AI. As a tool to explore, develop and extend human intelligence, AI aims to unravel the mystery hidden behind intelligence and develop machines with intelligent response systems similar to human behavior, such as intelligent robots, image recognition technology and more. The emergence and development of AI is changing or even subverting the human way of life and thereby leading society into a new era of connected intelligence.

Cognition and Emotion Identification Method of Artificial Intelligence

Scholars have carried out many studies about the comprehensive artificial machine simulation of human cognitive and emotional activities over nearly a century, using the knowledge of psychology, neuroscience, neurophysiology, computer science and so on. From the perspective of information science, identifying, understanding and simulating psychological states can be achieved by quantifying psychological information. Furthermore, AI can recognize and process information based on numerical values by constructing a cognitive and emotional information processing model.

With the development of analog neuron technology, artificial neural networks have gradually been used to identify human cognition and emotion. Artificial neural networks (ANNs), information processing systems based on the structure and function of neural networks of the human brain, perform the functions of distributed storage, associative memory and self-learning. Human psychology is a non-linear complex system

TABLE 7 | Reliability and validity of all variables (*N* = 305).

Variables	Indicators	Factor loading	Cronbach's α	AVE	CR
PB	PB1	0.749	0.849	0.546	0.878
	PB2	0.774			
	PB3	0.792			
	PB4	0.631			
	PB5	0.751			
	PB6	0.727			
CA	CA1	0.870	0.733	0.678	0.949
	CA2	0.802			
	CA3	0.725			
	CA4	0.737			
	CA5	0.921			
	CA6	0.803			
	CA7	0.755			
	CA8	0.883			
	CA9	0.888			
EI	EI1	0.848	0.852	0.666	0.969
	EI2	0.847			
	EI3	0.862			
	EI4	0.806			
	EI5	0.731			
	EI6	0.875			
	EI7	0.849			
	EI8	0.703			
	EI9	0.796			
	EI10	0.789			
	EI11	0.856			
	EI12	0.780			
	EI13	0.774			
	EI14	0.806			
	EI15	0.843			
	EI16	0.873			
SE	SE1	0.836	0.716	0.689	0.960
	SE2	0.842			
	SE3	0.895			
	SE4	0.856			
	SE5	0.857			
	SE6	0.679			
	SE7	0.726			
	SE8	0.850			
	SE9	0.885			
	SE10	0.858			
	SE11	0.819			
CP	CP1	0.561	0.754	0.549	0.893
	CP2	0.552			
	CP3	0.785			
	CP4	0.738			
	CP5	0.792			
	CP6	0.861			
	CP7	0.834			

TABLE 8 | Regression analysis results (*N* = 305).

Variables	CP			
	Model 1	Model 2	Model 3	Model 4
Corporate scale	−0.194**	−0.181**	−0.180**	−0.189***
Position of Respondents	−0.036	−0.051	−0.091	−0.111*
Office term	0.128*	0.128*	0.102	0.088
PB		0.248***	0.157**	0.144*
CA			0.116*	0.109*
EI			0.140*	0.125*
SE			0.212***	0.205***
PB × CA				0.122*
PB × EI				0.116*
PB × SE				0.120*
R2	0.055	0.116	0.190	0.239
Adjusted R2	0.046	0.105	0.171	0.213
F	5.895**	9.872***	9.978***	9.227***

$^*p < 0.05$; $^{**}p < 0.01$; and $^{***}p < 0.001$.

with uncertainties and non-linearity, while ANNs can activate the processing units to perform non-linear operations. Thus, the study of human psychology by adopting the technology of ANNs is worth exploring. From the perspective of artificial intelligence psychology, it is possible to identify mental activities more accurately by exploring a system (artificial brain) that controls all kinds of behaviors (Newell, 1992). Neural network simulation refers to the design of electronic circuits to simulate neural networks in the human brain. This method can realize part of the brain's thinking function. By simulating the neuronal path of processing uncertain information, it is found that the decision-making process of humans tends to follow the rule of optimal Bayesian computation, which lays a foundation for the simulation of brain information processing. Some studies build a hierarchical generative model to simulate the neural efficiency of the human cortex. The model can form top-down perceptual data flow and thus simulate the processes that occur when the brain faces uncertainty (Clark, 2013). Some scholars have established an artificial psychological model by applying RBF neural networks, which are composed of three layers of neurons: an input layer, a hidden layer and an output layer. This model can extract a variety of information features and perform non-linear transformation. Then, users' psychological values can be used to realize the recognition and prediction of their psychological states (Wang, 2017). As a function of the brain, psychology is the reflection of objective reality in the brain. Thus, analog neuron technology is of great significance to psychological identification. At present, some models have imitated mental activities to a limited extent. From the perspective of engineering science, a perfect description and analysis of all mental activities are difficult to realize. Recognizing and simulating limited mental activities are also valuable and meaningful.

Construction of the Artificial Intelligence-Based Cognition and Emotion Identification System

Through the above analysis, employees' cognitive ability, emotional intelligence and self-efficacy play an important moderating role in the relationship between participative

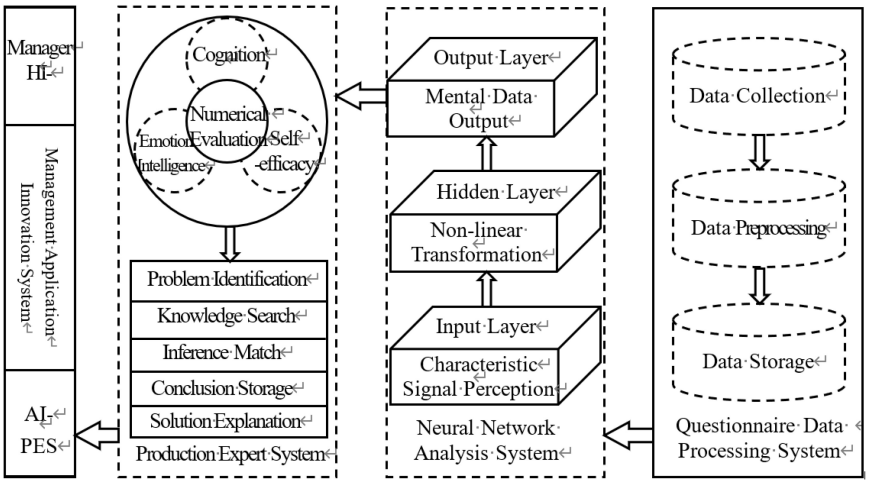


FIGURE 2 | AI-based cognition and emotion identification system.

budgeting and corporate performance. The cognitive and emotional states of employees in these three aspects affect the level of corporate performance. Therefore, it is necessary to build an AI-based cognition and emotion identification system that makes full use of employees' cognitive and emotional information to realize the automatic identification and intervention of employees' cognitive and emotional states.

According to artificial intelligence psychology, psychology is the natural reflection of the brain. AI can recognize the processes of mental activity by simulating the information processing of brain activity. It is known that the perception, analysis, and evaluation of information by the brain relies on neural networks to function. Intelligent models based on the nervous system, such as artificial neural networks, can understand overall brain activity through the connection of each processing unit, thereby mapping the psychological characteristics. In this paper, by combining data processing technology, radial basis function neural network technology and expert system tools, we built an idealized system to simulate the process of cognitive and emotional information acquisition, recognition, analysis, expression and application based on affective computing research. We try to simulate the processes intelligently through four subsystems: a questionnaire data processing system, neural network analysis system, production expert system and management innovation application system. We expect to realize the intelligent identification of employees' cognitive and emotional states based on the simulation of brain functions and provide some inspiration for management practices. Based on this, we constructed an AI-based cognition and emotion identification system, as shown in **Figure 2**.

Questionnaire Data Processing System

The questionnaire data processing system consists of a data collection module, a data preprocessing module and a data storage module. Among them, the data collection module collects and records data about the employees' cognitive and emotional characteristics by designing questionnaires. As a subjective

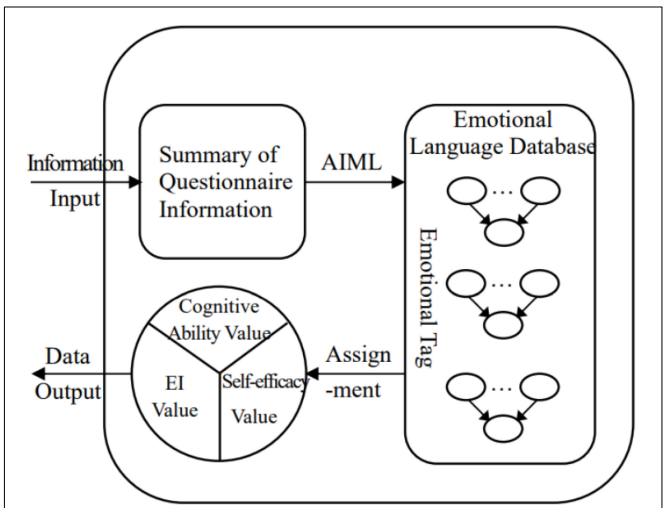


FIGURE 3 | Schematic diagram of data preprocessing module.

experience measurement method, the questionnaire can directly obtain the cognitive and emotional states of employees and avoid the measurement errors caused by inferring mental activities through behaviors in intelligent recognition (Picard, 2003). The data preprocessing module matches and classifies the collected cognitive and emotional information. The module is embedded with a perceptual language database, which mainly collects and summarizes adjectives that reflect the cognitive and emotional characteristics of employees in budgeting. Relying on artificial intelligence markup language (AIML), different cognitive and emotional characteristics are designated with emotional tags, such as "confident" to describe self-efficacy. By matching the questionnaire data with emotional tags, the topic information is classified and summarized, and the characteristics of employees' cognitive ability, emotional intelligence and self-efficacy are formed and assigned. Finally, valid data are input into the data

storage module to provide support for later retrieval and analysis, as shown in **Figure 3**.

Neural Network Analysis System

Adopting ANNs—the adaptive dynamic system composed of non-linear processing units—for information processing and pattern recognition can move technology closer to human thinking. Based on a radial basis function neural network, this system is a feedforward neural network with multiple inputs and a single output, which uses multidimensional data feature values as network inputs and employees' mental values as outputs. The neural network has a strong non-linear fitting ability and self-learning ability and can better simulate the complex non-linear system of human psychology (Wang, 2017). As a three-layer feedforward neural network, it is composed of an input layer, a hidden layer and an output layer, as shown in **Figure 4**. The input layer consists of signal source nodes, in which the neurons are responsible for sensing and transmitting the inputs of employees' cognitive and emotional characteristics. After initial identification of the information, it is transmitted to the hidden layer, which extracts the cognitive and emotional feature values and uses the Gaussian function to perform non-linear conversion, as shown in Formula 2-1. In the formula, $X_{(p)} = (x_{p1}, x_{p2}..., x_{pm})$ is the *p-th* input sample in the cognitive and emotional feature sample set, and $C_i = (c_1, c_2..., c_n)$ is the center of the *i-th* hidden layer neuron. σ^2 is the normalized parameter of the *i-th* hidden layer node. The output layer linearly combines the values of non-linear cognitive and emotional characteristics to output the data on the cognitive and emotional values of employees, such as cognitive ability value, emotional intelligence value or self-efficacy value, as shown in Formula 2-2. In the formula, M is the total number of hidden units, $p = 1, 2..., N$ (N is the total number of samples in the cognitive and emotional feature sample set), and w_i is the connection weighted value between the *i-th* basis function and the output node (Chen et al., 1991). Finally, the calculated cognitive and emotional values are imported into the production expert system (PES) for processing to identify employees' cognitive and emotional problems and provide corresponding solutions.

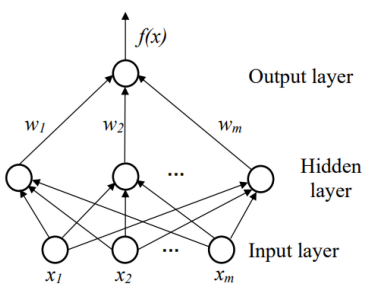
Production Expert System

The Production Expert System is an intelligent computer program that contains expert-level knowledge and experience in a specific field, which can simulate the reasoning and decision-making process of human experts to solve a range of sophisticated problems. This system expresses expertise based on rules, conditions, and actions, which includes a knowledge base, fact database and inference machine. It can be used to deal with diagnostic and prescriptive problems (Haag and Cummings, 2009). The Production Expert System is usually used for problem identification and analysis. The system first calls on the expert knowledge stored in the knowledge base. Then, the inference machine is used to compare and evaluate the cognitive and emotional values of employees participating in budgeting. Finally, the psychological states and problems of employees can be diagnosed. Under the premise of clarifying the characteristics of the employees' psychological states, the system answers the mandatory questions, that is, analyzing, judging and reasoning about existing problems, drawing conclusions and storing them in the fact database. The system derives solutions through analysis and uses an inference machine to explain the analysis process and deliver final conclusions. It can provide managers with professional solutions and help them effectively guide employees' cognitive and emotional activities.

Management Application Innovation System

As an innovative system based on human–machine coordination, the management application innovation system is jointly affected by human intelligence and artificial intelligence. It eliminates the traditional method of relying on managers' personal knowledge and experience to make program decisions and creatively introduces artificial intelligence into management decision-making. By combining the solutions provided by PES with the manager's own wisdom, we can design more valuable plans to solve employees' cognitive and emotional problems. Furthermore, we are able to manage and control employees' behaviors by guiding their cognitive and emotional activities. From the perspective of intelligent management, this system realizes man-machine interaction and has the compound effect of $1+1 > 2$. The sum of man-machine intelligence produced by it will synchronously promote human intelligence and artificial intelligence.

The AI-based cognition and emotion identification system is an identification system composed of a number of independent subsystems that are interconnected and restricted to each other, forming a complex non-linear system. The system



Hidden Layer Nonlinear Function

$$G(\|X_{(p)} - C_i\|) = exp\left|-\frac{1}{2}\left|\frac{\|X_{(p)} - C_i\|}{\sigma_i}\right|^2\right| \quad (2\text{-}1)$$

Neural Network Output Layer Formula

$$F(X_{(p)}, \omega_i, C_i) = \sum_{i=1}^{m} \omega_i G(\|X_{(P)} - C_i\|) \quad (2\text{-}2)$$

FIGURE 4 | Radial basis function neural network structure diagram.

has learnability, flexibility and subjective initiative. These characteristics give the system strong adaptability in changing environments. The questionnaire data processing system mainly sets up items to objectively and accurately evaluate problems, classifies and sorts the data, and later transmits the effective data to the input layer of the neural network analysis system. As the core component of the non-linear system, the neural network analysis system is a three-layer feedforward neural network composed of an input layer, a hidden layer and an output layer. It can effectively simulate the human cognitive and emotional process and output the employees' cognition and emotion values to the production expert system. The production expert system can simulate the thinking and reasoning processes of the brain, thereby clarifying the cognitive and emotional states of employees. It can assist managers in effectively intervening in employees' cognitive and emotional activities. By analyzing and controlling their possible cognitive and emotional problems, corporate performance could be improved.

RESEARCH CONCLUSION AND PROSPECTS

From the psychological and behavioral perspectives, this paper empirically examines the role of cognition and emotion in the relationship between participative budgeting and corporate performance. It is found that cognition and emotion can exert a positive effect from participative budgeting on corporate performance. On this basis, the AI-based cognition and emotion identification system is designed according to the theory of artificial intelligence. Through the cognition and emotion identification system, managers can identify the cognitive and emotional characteristics of budget participants and take necessary measures to intervene to improve corporate performance.

With the development of artificial intelligence, it is possible for people to evaluate cognitive and emotional characteristics through artificial intelligence systems and intervene with human psychology through management. With the continuous development of AI, its application in psychology and management is sure to expand. Additionally, the theoretical construction and practice of psychology and management will evolve.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YZ: designing and writing. LL: designing. TZ, LZ, and ZX: writing. MB: processing data. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Social Science Foundation of China (21AGL010), the Jilin Provincial Education Department (JJKH20190735SK), the Scientific Research Project of Jilin University of Finance and Economics (2017P37), and the Project of Jilin Provincial Corporate Finance and Accounting Research Center (JCK2020015). This work was also supported by China Scholarship Council.

REFERENCES

Argyris, C. (1952). *The Impact of Budgets on People*. Ph.D. Dissertation. Ithaca: Cornell University.

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social-Cognitive View*. Englewood Cliffs, NJ: Prentice-Hall, 23.

Bar-On, R., and Parker, J. D. (2000). *Handbook of Emotional Intelligence: Theory, Development, Assessment, and Application at Home, School and in the Workplace*. San Francisco, CA: Jossey-Bass, 43–50.

Che, L. P., and Qin, Q. W. (2009). *Management Psychology*. Wuhan: Wuhan University Press, 218–220.

Chen, S., Cowan, C. F. N., and Grant, P. M. (1991). Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Trans. Neural Netw.* 2, 302–309.

Chenhall, R. H. (1986). Authoritarianism and participative budgeting: a dyadic analysis. *Account. Rev.* 61, 263–272.

Cherrington, D. J., and Cherrington, J. O. (1973). Appropriate reinforcement contingencies in the budgeting process. *J. Account. Res.* 11, 225–253.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204.

Clinton, B. D. (1999). Antecedents of budgetary participation: the effects of organizational, situational, and individual factors. *Adv. Manage. Account.* 8, 45–70.

DeCotiis, T. A., and Summers, T. P. (1987). A path analysis of a model of the antecedents and consequences of organizational commitment. *Hum. Relat.* 37, 445–470.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Palo Alto: Stanford University Press, 32–40.

Fred, R. (2003). *Organizational Behavior*. Beijing: Posts and Telecommunications Press, 218.

Goleman, D., Boyatzis, R., and McKee, A. (2002). *Primal Leadership: Realizing the Power of Emotional Intelligence*. Boston: Harvard Business School Press, 25–34.

Green, K. W. Jr., and Inman, R. A. (2005). Using a just-in-time selling strategy to strengthen supply chain linkages. *Int. J. Prod. Res.* 43, 3437–3453.

Gruenfeld, D. H., and Hollingshead, A. B. (1993). Sociocognition in work groups: the evolution of group integrative complexity and its relation to task performance. *Small Group Res.* 24, 383–405. doi: 10.1177/1046496493243006

Gul, F. A., Tsui, J. S. L., and Fong, S. C. C. (1995). Decentralization as a moderating factor in the budgetary participation-performance relationship: some Hong Kong evidence. *Account. Bus. Res.* 25, 107–113. doi: 10.1080/00014788.1995.9729933

Haag, S., and Cummings, M. (2009). *Haag Management Information System*. Beijing: China Renmin University Press, 99–101.

Hebb, D. O. (1955). Drives and the C. N. S. (conceptual nervous system). *Psychol. Rev.* 62, 243–254. doi: 10.1037/h0041823

Law, K. S., Wong, C. S., and Song, L. J. (2004). The construct and criterion validity of emotional intelligence and its potential utility for management studies. *J. Appl. Psychol.* 89, 483–496. doi: 10.1037/0021-9010.89.3.483

Li, B. Z., Yin, S., Zhang, X. Y., and Luo, X. F. (2020). Cognitive ability and innovation performance of innovative employees from the perspective of sleep quality. *J. Harbin Eng. Univ.* 41, 937–944. doi: 10.1186/s12913-016-1423-5

Li, H. Z., Anna, and Liu, J. Y. (2021). The impact of strategic planning process and performance indicator causal chain on budget goal setting: an experimental study. *J. Central Univ. Financ. Econ.* 10, 52–64.

Li, Z. B., and Miao, Y. J. (2013). Can budget participation improve corporate performance. *J. Xinjiang Univ.* 41, 14–19.

Li, Z. D., and Guo, S. S. (2021). How does work stress affect employee turnover intention. *J. East Chin. Univ. Technol.* 36, 69–85.

Liu, W. (2018). The impact of transformational leadership on employee learning: the mediating role of employee emotion and the moderating role of emotional intelligence. *Manage. Rev.* 30, 128–140.

Liu, X. S., Chen, T., Zhang, P., and Peng, K. P. (2020). Cognitive complexity: an assessment method and related research. *Psychol. Res.* 13, 8–15.

Lu, C. Q., Ling, W. Q., and Fang, L. L. (2001). A review of managerial self-efficacy. *Adv. Psychol. Sci.* 9, 179–185.

Maiga, A. S. (2005). The effect of manager's moral equity on the relationship between budget participation and propensity to create slack: a research note. *Adv. Account. Behav. Res.* 8, 139–165.

Milani, K. (1975). The relationship of participation in budget-setting to industrial supervisor performance and attitudes: a field study. *Account. Rev.* 50, 274–284.

Newell, A. (1992). Précis of unified theories of cognition. *Behav. Brain Sci.* 15, 425–492.

Peter, B., and Morris, M. (1986). Budgetary participation, motivation, and managerial performance. *Account. Rev.* 61, 587–600.

Picard, R. W. (2003). Affective computing: challenges. *Int. J. Hum. Comput. Stud.* 59, 105–114.

Ryan, S., William, D. S., Killgorea, A., Alkozeia, R., and Lanea, D. (2018). A neuro-cognitive process model of emotional intelligence. *Biol. Psychol.* 139, 131–151. doi: 10.1016/j.biopsycho.2018.10.012

Seijts, G. H., and Latham, G. P. (2011). The effect of commitment to a learning goal, self-efficacy, and the interaction between learning goal difficulty and commitment on performance in a business simulation. *Hum. Perform.* 24, 189–204. doi: 10.1080/08959285.2011.580807

Shields, J. F., and Shields, M. D. (1998). Antecedents of participative budgeting. *Account. Org. Soc.* 23, 49–76. doi: 10.1111/jonm.12622

Solman, A. (2004). "What is the emotion theories about? Architectures for modeling emotion," in *Proceedings of the Cross Disciplinary Foundations American Association for Artificial Intelligence 2004 Spring Symposium* (Edinburgh: The University of Edinburgh), 2–8.

Spreitzer, G. M. (1995). Psychological empowerment in the workplace: dimensions, measurement, and validation. *Acad. Manag. J.* 38, 1442–1465. doi: 10.1111/jonm.12045

Vincent, K. E., Ian, R. C., and Michele, K. C. (2005). The impact of market competition and budgetary participation on performance and job satisfaction: a research note. *Br. Account. Rev.* 23, 115–133. doi: 10.1016/j.bar.2004.06.007

Wang, Z. L. (2017). *Machine Intelligence: Artificial Psychology*. Beijing: China Machine Press, 16–23.

Wei, X. H., Liu, Y. M., and Che, X. L. (2015). Relationship conflict management: the moderating role of team efficacy and team emotional intelligence. *J. Syst. Manag.* 24, 138–145.

Westbrook, R. A., and Oliver, R. L. (1991). The dimensionality of consumption emotion patterns and consumer satisfaction. *J. Consum. Res.* 18, 84–91. doi: 10.1086/209243

Wong, C. S., and Law, K. S. (2002). The effects of leader and follower emotional intelligence on performance and attitude: an exploratory study. *Leadersh. Q.* 13, 243–274. doi: 10.1016/s1048-9843(02)00099-1

Wu, G., Wang, Y., and Zhou, Y. F. (2019). The impact of budget participation on corporate management performance. *Financ. Res.* 3, 30–41.

Wu, L., Wang, F. F., and Yuan, Z. (2011). An empirical study on the impact of managers' budget participation on management performance. *J. Northeast. Univ.* 32, 149–152.

Yang, S. C., and He, M. (2020). Participatory budgeting: value review and feasibility boundary. *Gansu Soc. Sci.* 5, 185–192.

Yao, K. (2008). A review of self-efficacy: a new trend of organizational behavior. *J. Org. Behav.* 3, 463–468.

Yao, W. (2018). The wide application of artificial intelligence and the innovation of income distribution system. *Academia* 4, 66–74.

Yerkes, R. M., and Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *J. Comp. Neurol. Psychol.* 18, 459–482. doi: 10.1002/cne.920180503

Yoichiro, M., Tensei, S., Katsuari, K., and Cooper, E. W. (2020). "Human-robot interaction based on facial expression recognition using deep learning," in *Proceedings of the 2020 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems (SCIS-ISIS)* (Tokyo: Hachijo island), 1–6.

Yu, X., Li, Y., Su, Z., Tao, Y., Nguyen, B., and Xia, F. (2020). Entrepreneurial bricolage and its effects on new venture growth and adaptiveness in an emerging economy. *Asia Pac. J. Manage.* 37, 1141–1163.

Zhang, D. (2011). *Organizational Behavior*. Beijing: Higher Education Press, 9–15.

Zhang, H. H. (2012). Team emotional intelligence from the perspective of social network. *Adv. Psychol. Sci.* 29, 1381–1395.

Zhao, J. J., Shao, D. F., Yi, M., and Peng, J. Q. (2018). Coaching leadership, self-efficacy and employee performance. *J. South. Econ.* 3, 99–112.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhou, Zhang, Zhang, Xue, Bao and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 09 March 2022
doi: 10.3389/fpsyg.2022.842076

Research on Digital Business Model Innovation Based on Emotion Regulation Lens

*Shan Lu and Haijing Yu**

School of Business and Management, Jilin University, Changchun, China

Digital technologies, such as artificial intelligence, brain-computer interfaces technology and big data, enable many firms to innovate their business model. It is clearly an emotional process due to its complex and uncertain nature, and involves individuals' emotion regulation, yet the current research lacks an effective conversion path from emotion to digital business model innovation (BMI). Drawing on theories and research on emotion regulation and business model innovation, we investigate how emotion regulation of entrepreneurs (i.e., cognitive reappraisal and expressive suppression) influence digital BMI. Data from 126 new ventures show that entrepreneurs' reappraisal positively affects digital BMI, while entrepreneurs' suppression exerts opposite effects on digital BMI. Moreover, we find that environmental dynamism moderates this relationship. The findings explain the emotional complexity in digital technology empowerment, which has implications for the development and design of brain computer interface applications and the literature on emotions and business model innovation.

Keywords: emotion regulation, digital business model innovation, entrepreneurship, environmental dynamism, digital technology

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Qihua Cai,
Zhengzhou University, China
Jing Liu,
Inner Mongolia University, China

***Correspondence:**
Haijing Yu
yuhj17@mails.jlu.edu.cn

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *23 December 2021*
Accepted: *11 February 2022*
Published: *09 March 2022*

Citation:
Lu S and Yu H (2022) Research on Digital Business Model Innovation Based on Emotion Regulation Lens. Front. Psychol. 13:842076. doi: 10.3389/fpsyg.2022.842076

INTRODUCTION

Digital technologies, such as artificial intelligence, brain-computer interfaces technology and big data, are changing entrepreneurial behaviors and outcomes fundamentally, including recognize and regulate emotion using advanced artificial intelligence brain-computer interface to avoid the bias caused by subjective emotion, and creating a new business model (BMI) through the combination of BMI with digital technology, which is labeled as *digital BMI* (e.g., Fichman et al., 2014; Remane et al., 2017). We follow Soluk et al. (2021) in defining it as "a significantly new way of creating and capturing the business value that is embodied in or enabled by digital technologies." Practice shows that more and more new ventures have achieved rapid growth through digital BMI. Examples of digital business model innovators include Tik Tok, which develops a unique business proposition through building social networking and video-sharing platforms to deliver a new way to create value for consumers (Ma and Hu, 2021). Similarly, some theoretical studies have also revealed the importance of digital BMI to long-term business success (Remane et al., 2017; Sorescu, 2017; Teece, 2018). However, as yet, we have a limited understanding on how emotion influence digital BMI emerge (Soluk et al., 2021). To address this gap, we explore the factors that influence digital BMI from the individual-level emotion lens.

Prior research has indicated the importance of individual entrepreneurs for BMI, focusing on how BMI is triggered by individual-level cognition, including creativity, mindfulness, or analogical

reasoning (Snihur and Zott, 2020). However, few studies have explored the emergence of BMI from an emotional lens. This lack of research is especially surprising, since innovation, including BMI, is ambiguous, uncontrollable event (Huy, 2005) laden with uncertainty and potentially affects, positively or negatively, which can trigger emotional ambivalence of entrepreneurs and their stakeholders (Ashforth et al., 2014). Therefore, entrepreneurs need to recognize and regulate their own and stakeholders' emotions to achieve emotional resonance for innovation adoption. In particular, we propose that entrepreneurs need more emotional regulation in the context of the digital BMI than conventional BMI. The reason is that the constantly changing attribute of digital technology embedded in digital BMI triggers continuous experiment and implementation, which makes the innovation process and results more uncertain (Nambisan et al., 2017). Such higher complexity makes digital BMI trigger more intense emotional experiences of entrepreneurs and stakeholders.

In this regard, we explore how different emotion regulation strategies influence digital BMI. More specially, we focus on two of the most well-established emotion regulation strategies, namely cognitive reappraisal, an antecedent-focused approach, which shape emotion-inner experience by changing the way we see things, and expressive suppression, a response-focused approach relating to inhibiting emotion-external expression (Webb et al., 2012; Gross, 2013). We propose that entrepreneurs' cognitive reappraisal positively affects digital BMI, while expressive suppression strategies negatively affect digital BMI. We further posit that that such effect of emotion regulation strategies on digital BMI can be shifted by environmental dynamism. We believe that this factor may exert important effects because past research suggests that the effects of emotion regulation is largely context dependent (De Cock et al., 2020). Environmental dynamism refers to the rate of unpredicted change occurring within a given industry (Duncan, 1972; Dess and Beard, 1984). In particularly, such rapid and unpredictable change is more obvious in the digital context (Nambisan, 2017; Lu et al., 2021a; Zhao and Zhou, 2022). Research has shown that entrepreneurs in dynamic environments often experience high levels of stress and anxiety (Hmieleski and Baron, 2008; Shan and Lu, 2020). On the basis of these considerations, we reason that the environmental dynamism, and the emotions that environmental dynamism may evoke, will influence the effectiveness and usefulness of emotion regulation.

We contribute to the present status of research in three main ways. First, we contribute to emotion research in entrepreneurship by investigating the link between emotion regulation and digital BMI. We address the calls for more attention to be paid to the role of emotion regulation in the entrepreneurial process by previous research (e.g., Burch et al., 2013; De Cock et al., 2020; Sirén et al., 2020). While the widespread agreement that the entrepreneurial process is seen as an emotional rollercoaster, entrepreneurs' emotion regulation has not received much attention. Second, we contribute to research on digital BMI by extending the emotion lens on digital BMI by showing that emotion regulation can be a nuanced explanation of the mechanisms leading to digital BMI. Third, we contribute to the emotion regulation literature by investigating when and how it influence digital BMI in the specific context of entrepreneurship. We find the dynamics of the entrepreneurial environment as boundary conditions for the relationship between emotion regulation and digital BMI. Our theoretical model integrates individual factors and environment factors to explain how certain aspects of emotion regulation become more or less beneficial given different dynamic levels of the entrepreneurial environment.

THEORETICAL BACKGROUND AND HYPOTHESES

Taking an Emotive View on Digital Business Model Innovation

Business model innovation is defined as "designed, novel, non-trivial changes to the key elements of a firm's business model and/or the architecture linking these elements" (Foss and Saebi, 2018, p. 201). In recent years, with technological change and global competition, many firms have reshaped the industry pattern and created growth myths through BMI, becoming an essential driving force of innovation-driven development (Johnson, 2010; McGrath, 2010; Afuah, 2014; Hartmann and Vanpoucke, 2017; Narayan et al., 2021). Given its economic importance, BMI has gained increasing attention from academics and practitioners (Foss and Saebi, 2017; Snihur and Zott, 2020; Bhatti et al., 2021; Zhang et al., 2021). In their comprehensive review on BMI, Foss and Saebi (2017) indicated that BMI studies have been growing rapidly since it was first explicitly discussed in Mitchell and Coles (2003). However, with the exception of a few studies, little is known about the possible antecedents of BMI (Foss and Saebi, 2017; Bhatti et al., 2021; Zhang et al., 2021). For example, in their meta-analytic review on BMI, Zhang et al. (2021) suggested that the external factors include market opportunity, situational factors, value network and technology innovation can impact BMI. In addition, individual-level cognition, internal resources and capabilities, as well as organization characteristics as internal factors can foster BMI. In recent studies, there is some new evidence that entrepreneurs influence BMI, including CEO leadership (Colovic, 2021), founder imprinting (Snihur and Zott, 2020), top management mindfulness (Bhatti et al., 2021). Despite the role of entrepreneurs is increasingly studied, there is still a lack of both theoretical and empirical research on BMI from individual entrepreneur's emotive lens. To address this gap, we draw on the recently developed emotive approach to digital BMI, which suggests that entrepreneurs need to regulate their emotions in the process of developing digital BMI.

Past research has shown that entrepreneurs experience more extreme emotions in the innovation process and that using appropriate emotion regulation strategies to deal with these extreme fluctuating emotions can facilitate innovation adoption and implementation (Huy, 2005; Raffaelli et al., 2019; Shan et al., 2021). More specially, innovation involves mobilizing sufficient resources and challenges members' basic assumptions about the organization, which define intersubjective reality

and provide a way of dealing with ambiguous, uncontrollable events (Schein, 1992; Yu et al., 2020). Organization members are emotionally invested in these non-negotiable assumptions, which challengers their cognitive and emotional stability for core identity, thus, triggers strong defense mechanisms, such as anxiety and defensiveness (Schein, 1992; Huy, 1999). Moreover, emotion theory also predicts a similar emotional response to innovation, which suggests emotions arise from the appraisal of events that are concerning personal goals and important on some valued dimension (Pirola-Merlo et al., 2002; Morris et al., 2012; Davis et al., 2017). If people appraise events as leading to potential benefits, positive emotions are aroused, whereas negative emotions are aroused if they appraise the events as potentially harmful (Huy, 2005; Davis et al., 2017). In this regard, innovation, as a major event of organizational change, is related to entrepreneurs' goals and organization value creation, which can trigger entrepreneurs' assessment of whether it's an opportunity or a threat (Saebi et al., 2017), thus, produce positive or negative emotions. In particular, the digital BMI process experiences more uncertainty, discontinuity and complexity than the traditional innovation process (Soluk et al., 2021). Conventional BMI usually remain in place for several years, while digital BMI involves repeated cycles of experimentation and implementation as the result of the constantly changing nature of digital technologies inherent in digital BMI (Nambisan et al., 2017). Moreover, compared with conventional BMI, digital BMI is also associated with more rapidly changing customer needs and market developments (Sorescu, 2017). Therefore, entrepreneurs are expected to experience greater emotional experience in the digital BMI process than in the traditional innovation process. Thus, we argue that how entrepreneurs regulate their emotions will have an impact on whether their ventures will adapt digital BMI, or conversely, abandon.

Emotion Regulation and Entrepreneurship

Emotion regulation refers to "attempts to influence which emotions one has, when one has them, and how one experiences or expresses these emotions (Gross, 2015)." According to the time series of emotion generation process, Gross and John (2003) proposed that people could choose five emotion regulation strategies sequentially. These strategies are (1) situation selection, (2) situation modification, (3) attention deployment, (4) cognitive change (cognitive reappraisal), and (5) response modulation (e.g., expressive suppression). The first four of these strategies belong to antecedent-focused emotion regulation strategies, which refers to changing individual's emotional experience before the emotion is fully evoked. The last one is a response-focused emotion regulation strategy, which refers to altering the individual's emotional response once it is fully underway. We focus on two different types of strategies—cognitive reappraisal and expressive suppression, as the most commonly used emotion regulation strategies in daily life, which have been most established, intensively studied in psychological research (Gross and John, 2003; Webb et al., 2012; Gross, 2013). More specially, cognitive reappraisal refers to shape emotional experience by altering one's own perceptions of a situation before emotions are generated. Expressive suppression refers to inhibiting emotion-expressive behavior after emotions are generated (Gross and John, 2003). We expect that these two different types of strategies will have different impacts on digital business model innovation.

Emotion regulation was first proposed in developmental psychology studies in the 1980s (Gross, 1999). In recent years, with the increasing of emotion research in entrepreneurship, some studies began to focus on emotional regulation in entrepreneurial activities. For example, in their case study, Huy and Zott (2019) investigated the effect of emotion regulation on resource mobilization. They found that managers' emotion regulation of the self (defined as the regulation of one's own emotion) helps them mobilize human capital resources by creating psychic benefits, whereas their emotion regulation of others (defined as emotion regulation of other stakeholders such as investors and employees) helps mobilize social capital by facilitating legitimacy judgments. In their empirical study. De Cock et al. (2020) investigated how entrepreneurs' two well-established types of emotion regulation-cognitive reappraisal and expressive suppression influence the likelihood of their venture surviving. They found that cognitive reappraisal lowers their survival odds when ventures are low-performing. In contrast, expressive suppression is generally associated with lower survival chances, except when the venture's performance is very low, in which case suppression actually significantly increases the survival odds. Sirén et al. (2020) hypothesized and found that in a nascent venture team, an individual with a stronger tendency to reappraise emotions is more likely to emerge as a leader, whereas an individual with stronger tendency to suppress emotions is less likely to emerge as a leader. Moreover, positive team emotions negatively moderate the relationship between reappraisal and leader emergence, whereas a team's negative emotions magnify the positive relationship between reappraisal and leader emergence.

In sum, as yet, the empirical studies of emotion regulation in an entrepreneurial context are scarce (Huy and Zott, 2019; De Cock et al., 2020). Moreover, the previous scant research on emotion regulation in entrepreneurship has highlighted and empirically demonstrated the importance of this concept for several domains of entrepreneurial outcomes (e.g., resource mobilization, venture survival and leader emergence). However, as yet there is no empirical evidence on the role that emotion regulation play in new venture digital BMI. In the following, we will derive our hypotheses regarding this relationship based on previous evidence from entrepreneurship and the emotion research in psychology.

Emotion Regulation and Digital Business Model Innovation

First, we expect cognitive reappraisal of entrepreneurs is positively related is digital BMI. As a kind of antecedent-focused emotion regulation strategy, cognitive reappraisal can

change an individual's internal emotional experience by changing how we interpret emotional events (Gross and John, 2003; De Cock et al., 2020; Sirén et al., 2020). For example, an entrepreneur may initially interpret project failures as the end of an entrepreneurial career, and then reappraise it as an opportunity to learn and grow. According to hedonic models of wellbeing, individuals prefer to use cognitive reappraisal strategy to experience more positive emotions and reduce negative emotions (Gross, 2015).

Previous studies have shown that individuals who use cognitive reappraisal are more likely to reframe threats as opportunities in ambiguous contexts (De Cock et al., 2020). The reprogrammable and re-combinable nature of digital technologies inherent in digital BMI enables repeated cycles of experimentation and implementation, making it unclear as to when a particular phase starts and/or ends (Nambisan et al., 2017; Soluk et al., 2021). Such uncertainty induces digital BMI might be perceived as both a threat and an opportunity to the entrepreneur (Saebi et al., 2017). Prior research has shown that the perceived opportunity of managing is related to BMI adaption, whereas perception of threat is related to resisting BMI (Dewald and Bowen, 2010). Thus, we expect that cognitive reappraisal is positively associated with BMI because reappraisers tend to reframe digital BMI as an opportunity rather than a threat. Moreover, many experimental studies on emotion regulation have shown that individuals using cognitive reappraisal strategies experienced more positive emotions and fewer negative emotions (e.g., Gross, 1998; Ray et al., 2010; Lieberman et al., 2011; Szasz et al., 2011; Wolgast et al., 2011; Feinberg et al., 2012). Such positive effect that is often interpreted as a sign that "all is going well" encourages the entrepreneurs to expand their repertoires of useful skills and the range of their social networks, which improves the dynamic capacity of entrepreneurs to respond effectively to environmental change (Baron, 2008; De Cock et al., 2020). Such dynamic capacity has been demonstrated to be crucial for digital BMI (Soluk et al., 2021).

In contrast, we expect expressive suppression of entrepreneurs is negatively related is digital BMI. Expressive suppression, as a kind of response-focused emotion regulation strategy, refers that a person decreasing emotion-expressive behavior but not the internal experience of the emotion while emotionally aroused (Gross, 2014). For example, entrepreneurs may hide their feelings of sadness when faced with a negative event, such as a failed investment. Thus, although expressive suppression maintains the stability of emotions and avoids the spread of emotions among stakeholders, it leads to more negative internal experience (De Cock et al., 2020). Experimental evidence suggested that the person who uses suppression experience less positive emotion and more negative emotion, including painful feelings of inauthenticity as well as depressive symptoms (Gross and John, 2003; Moore et al., 2008; Nezlek and Kuppens, 2008). Such negative emotions are associated with less risk-taking behavior, lower levels of alertness and less engagement in search activities, all of which are important antecedents of innovation (Baron, 2008; Tang et al., 2012; Foo et al., 2015; De Cock et al., 2020). Moreover, prior research has shown that suppression leads to the unreal perception of the entrepreneur by stakeholders, which in turn leads to less social interaction and hinders the establishment of positive relationships with stakeholders (Butler et al., 2003; Gross, 2014; De Cock et al., 2020). However, the development of digital BMI depends on the participation of distributed actors, including external stakeholders (Yoo et al., 2010; Faraj et al., 2011; George et al., 2021). Therefore, we expect that suppression hinders the development of digital BMI by inhibiting the social relationships with stakeholders.

Thus, we hypothesize:

Hypothesis 1a. There is a positive relationship between the cognitive reappraisal of a new venture's entrepreneur and digital BMI.

Hypothesis 1b. There is a negative relationship between the expressive suppression of a new venture's entrepreneur and digital BMI.

Moderating Effects on Environmental Dynamism

Dynamic environments are characterized by unpredictable and rapid change, which increases uncertainty for entrepreneurs and firms implementing entrepreneurial activities within them (Duncan, 1972; Dess and Beard, 1984; Ensley et al., 2006; Hmieleski and Baron, 2008, Hmieleski and Baron, 2009). Because of such uncertainty, entrepreneurs working in dynamic environments often suffer from heavy information processing burdens. As a result, they may experience high levels of distress and anxiety (Markman et al., 2005; Ensley et al., 2006; Hmieleski and Baron, 2008, Hmieleski and Baron, 2009). Thus, when entrepreneurs develop digital BMI in a highly dynamic environment, they seek approaches to protect themselves against the negative emotions. Emotion regulation are often seen to afford such protection (Gross and John, 2003). Therefore, in a highly dynamic environment, entrepreneurs are encouraged to use emotion regulation strategies to reduce stress and anxiety. Drawing on this line of thinking, we suggest that high environmental dynamism encourages entrepreneurs to use emotion regulation that might help them deal with uncertain times. Thus, we expect that the effect of both emotion regulation strategies on digital BMI is strengthened by the dynamic environment.

Moreover, as the high-level environmental dynamism makes it difficult for organizations to assimilate and anticipate environmental conditions and increase the difficulty of developing digital BMI, firms need to develop multiple solutions to deal with the uncertainty in this process. Firms thus need to encourage their employees to express their ideas and free communication to inspire more ideas and solutions (Ashforth and Humphrey, 1995; Akgün et al., 2008). Previous studies have shown that how entrepreneurs regulate their emotions affects how employees regulation their emotions (De Cock et al., 2020). Entrepreneurs who use cognitive reappraisal create more positive emotions, encouraging communication and expression in ventures. However, new and potentially better opportunities are less likely to be discussed openly within the venture under suppressing emotions (De Cock et al., 2020).

Thus, we hypothesize:

Hypothesis 2a. The positive relationship between cognitive reappraisal and digital BMI is strengthened by environmental dynamism.

Hypothesis 2b. The negative relationship between expressive suppression and digital BMI is strengthened by environmental dynamism.

MATERIALS AND METHODS

Sample and Data Collection

To test our hypotheses, we used a questionnaire survey from China firms. We chose China as our research context given that digital technology is driving the development of Chinese firms. Moreover, the innovation-driven development, international competition, and digital transformation challenges of key industries in China make this country an ideal context to study digital BMI and the role of environmental dynamism.

We used a sample of Chinese new ventures within 10 years (Milanov and Fernhaber, 2009). To implement the survey, we used the online survey platform of China (i.e., Questionnaire Star). Before sending the survey link to the potential participants, we piloted and pretested the survey through two steps: (1) face-to-face interviews with three entrepreneurial scholars and three entrepreneurs who put forward opinions on possible problems about the questionnaire. (2) sending the web survey to another 25 entrepreneurs to assess the survey structure, wording, and overall length, thus following previous survey-based research (e.g., Obal, 2017). The two pretesting phases led to a final questionnaire.

We collected a total of 200 questionnaires. Further, we excluded those questionnaires that filled out by firms over 10 years and with incomplete information, and finally there were 126 valid questionnaires left, reflecting an effective response rate of 63%. 87 samples (69.05%) are from traditional manufacturing, 13 samples are from IT industry which account for 10.32%. The remaining 26 samples are from other industries, accounting for 20.63%. In terms of the number of employees, 90 samples (71.43%) have less than 200 employees and 36 samples (28.57%) have more than 200 employees.

Variables

Digital Business Model Innovation

We measured digital BMI using nine items derived from Soluk et al. (2021). Participants answered each question on a 7-point Likert-type scale ranging from 1 (strongly disagree) to 7 (strongly agree) (Cronbach's alpha = 0.90).

Emotion Regulation

We measured cognitive reappraisal and expressive suppression using Gross and John's (2003) 10-items scale. Cognitive reappraisal sub-scale includes six items such as "I control my emotions by changing the way I think about the situation" (Cronbach's alpha = 0.90). Expressive suppression sub-scale includes four items, such as "I control my emotions by not expressing them" (Cronbach's alpha = 0.87). Respondents rated their level of agreement with each item using a 7-point Likert-type scale ranging from 1 "strongly disagree" to 7 "strongly agree."

Environmental Dynamism

We measured environmental dynamism using the scale developed by Jansen et al. (2006). The scale is comprised of four items in total. All the items were measured using a 7-point Likert-type scale, with response categories ranging from "strongly disagree" to "strongly agree."

Control Variables

We chose control variables that were also controlled in prior studies on (digital) BMI, including firm's financial performance (respondents were asked to indicate how successful in terms of ROA their company on a five-point scale: 1 = very unsuccessful; 5 = very successful) (Soluk et al., 2021). We also controlled firm size through the number of employees (1–20, 21–50, 51–200, 201–500, and more than 500) and firm age using the number of years the firm has existed (Guo et al., 2016; Soluk et al., 2021). Finally, we controlled industry by setting three dummy variables (traditional manufacturing, IT, and other industries). These data were collected as demographic items at the end of the administered survey.

Common Method Bias

This study employed two statistical methods to check the potential of common method bias. First, we took a Harman's one-factor test and the result showed that the first principal component encompassed only 29.785% of the total variation (less than 40%), thus excluding the existence of common homology bias (Podsakoff et al., 2003). Second, we conducted a confirmatory factor analysis (CFA) to compare a model with all items loading on one latent factor (one-factor model) with our four-factor model). The results shown that our model (χ^2/df = 1.393; comparative fit index [CFI] = 0.944; root mean square error of approximation [RMSEA] = 0.056) indicates superior fit over the one-factor model (χ^2/df = 4.936; CFI = 0.428; RMSEA = 0.177) as indicated by the fit indices (Hu and Bentler, 1999).

Reliability and Validity

We tested the reliability and validity of the scales (**Table 1**). The results showed that the Cronbach's alpha coefficient of each variable was over 0.7, indicating that the scales had good reliability. Furthermore, we tested the composite reliability (CR) and convergent validity of the sample using a CFA. The results show that factor loading for the variables is more than 0.6. In addition, the CR of each variable is higher than 0.8, and the average variances extracted (AVE) for each construct is higher than 0.5 (Fornell and Larcker, 1981). Thus, CR and convergent validity are demonstrated by the CFA results as well. Finally, the correlation coefficients between variables are all less than the square root of AVE, indicating good discriminant validity.

Descriptive Analyses and Correlations

As shown in **Table 2**, there is a significant correlation between independent variables and dependent variable. First, cognitive

TABLE 1 | Reliability and validity.

Variables	Items	Factor loading	Cronbach's alpha	CR	AVE
Cognitive reappraisal (CR)	CR1	0.827	0.896	0.897	0.592
	CR2	0.775			
	CR3	0.757			
	CR4	0.738			
	CR5	0.720			
	CR6	0.795			
Expressive suppression (ES)	ES1	0.804	0.872	0.875	0.636
	ES2	0.834			
	ES3	0.756			
	ES4	0.794			
Digital BMI (DBMI)	DBMI1	0.680	0.902	0.902	0.507
	DBMI2	0.682			
	DBMI3	0.715			
	DBMI4	0.689			
	DBMI5	0.731			
	DBMI6	0.758			
	DBMI7	0.669			
	DBMI8	0.793			
	DBMI9	0.684			
Environmental dynamism (ED)	ED1	0.695	0.868	0.869	0.626
	ED2	0.806			
	ED3	0.809			
	ED4	0.846			

reappraisal is positively associated with digital BMI ($R = 0.32$, $p < 0.01$). Second, expressive suppression is negatively associated with digital BMI ($R = -0.39$, $p < 0.01$). Meanwhile, to eliminate the multi-collinearity between variables, we calculate the variance inflation factor (VIF) of each model. The results show that all of the values are less than 2 and thus multi-collinearity is not a serious problem.

Hypotheses Testing

We model a hierarchical regression to test our hypotheses (in **Table 3**).

Hypotheses 1a and 1b predicted that cognitive reappraisal and expressive suppression, respectively, have relationships with digital BMI. As shown in Model 2 of **Table 3**, there were significant effect of cognitive reappraisal on digital BMI ($\beta = 0.544$, $p < 0.001$), thus Hypothesis 1a was supported. Meanwhile, in Model 2, expressive suppression had a significantly negative effect on digital BMI ($\beta = -0.584$, $p < 0.001$). Thus, Hypothesis 1b was supported.

Next, we tested the moderating effect of environmental dynamism on the relationship between emotion regulation and digital BMI (Hypotheses 2a and 2b). As presented in **Table 3**, we entered the moderator environmental dynamism and the interaction of emotion regulation (cognitive reappraisal and expressive suppression) and environmental dynamism in Model 4 and Model 5. The results in Model 4 shown that the interaction items of cognitive reappraisal and environmental dynamism ($\beta = 0.123$, $p < 0.05$) was significant. To facilitate the interpretation of this interaction effect, we visualize the interaction in **Figure 1A**. As shown in the figure, when environmental dynamism was high, the positive effect of cognitive reappraisal on digital BMI was significant, hypothesis 2a thus was supported. The results in Model 5 shown that the interaction items of expressive suppression and environmental dynamism ($\beta = -0.128$, $p < 0.001$) was significant. Meanwhile, the graph of interaction of expressive suppression and environmental dynamism revealed (see **Figure 1B**) that as environmental dynamism go from low to high, the negative relationship between expressive suppression and digital BMI is amplified, which supported Hypothesis 2b.

DISCUSSION

Discussion and Implications for Research

The widespread agreement that entrepreneurial process is seen as an emotional roller coaster. However, studies of emotion regulation in entrepreneurial contexts are scarce. Our study theorized and tested the role of entrepreneurs' emotion regulation in new ventures' digital BMI, which offers two novel findings on how and when the emergence of BMI from emotion lens. First, we find that entrepreneurs' cognitive reappraisal is positively related to digital BMI, whereas entrepreneurs' expressive suppression has a negative relationship with digital BMI. Second, we predicted and found that the positive impact of cognitive reappraisal on digital BMI is strengthened by environmental dynamism. Also, the negative impact of expressive suppression on digital BMI is strengthened by environmental dynamism.

TABLE 2 | Description statistics and correlation coefficients.

	M	*SD*	1	2	3	4	5	6
Firm age	6.300	2.489	1					
Firm size	3.170	0.901	−0.026	1				
Financial performance	2.890	1.045	0.044	−0.006	1			
Cognitive reappraisal	4.370	1.241	−0.084	0.108	0.058	1		
Expressive suppression	3.919	1.173	0.071	0.140	0.182*	0.341**	1	
Environmental dynamism	4.786	1.053	0.017	−0.070	0.184*	0.137	0.149	1
Digital BMI	4.662	1.232	−0.032	−0.080	0.016	0.320**	−0.394**	0.427**

*$p < 0.05$, **$p < 0.01$. Dummy variables of the industry are not shown in the table.*

TABLE 3 | Regression models of emotion regulation and digital BMI.

	Digital BMI					
	Model1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Control variables						
Firm age	−0.034	0.049	0.045	0.042	0.027	0.012
Firm size	−0.076	−0.059	−0.016	−0.023	−0.024	−0.039
Industry	−0.059	0.056	0.054	0.073	0.060	0.094
Financial performance	0.021	0.085	0.013	0.008	−0.001	−0.018
IVs						
Cognitive reappraisal		0.544***	0.496***	0.520***	0.478***	0.506***
Expressive suppression		−0.584***	−0.627***	−0.614***	−0.637***	−0.620***
Moderating variable						
Environmental dynamism			0.450***	0.458***	0.460***	0.478***
Interaction						
Cognitive reappraisal × Environmental dynamism				0.123*		0.200**
Expressive suppression × Environmental dynamism					−0.128*	−0.204**
R^2	0.011	0.405	0.595	0.609	0.610	0.642
Adj-R^2	−0.021	0.375	0.571	0.582	0.583	0.614
F	0.344	13.513***	24.758***	22.790***	22.887***	23.108***

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Our findings contribute to the existing literature in the following ways. First, we investigate the link between emotion regulation and digital BMI in the entrepreneurial context in response to the call of scholars (cf. Burch et al., 2013; De Cock et al., 2020; Sirén et al., 2020) to focus on the role of emotional regulation in entrepreneurship. While the widespread agreement that the entrepreneurial process is seen as an emotional roller coaster, entrepreneurs' emotion regulation has not received much attention.

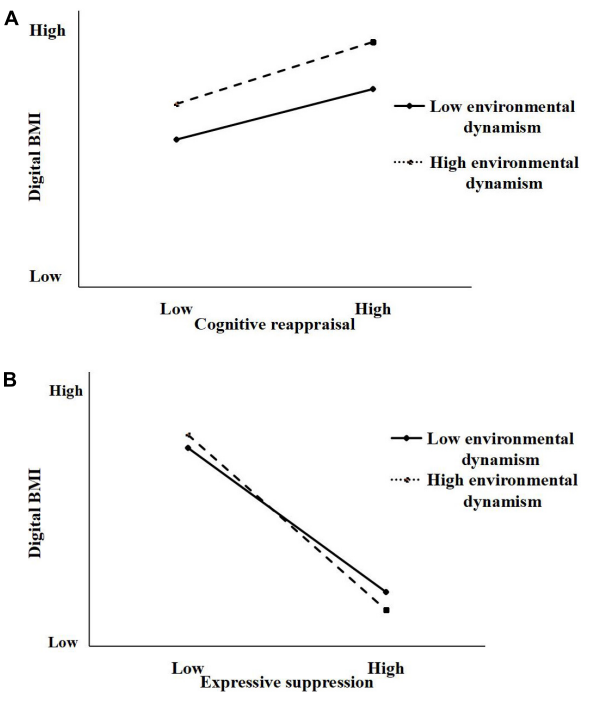


FIGURE 1 | (A) Cognitive reappraisal × environmental dynamism interaction (H2a). **(B)** Expressive suppression × environmental dynamism interaction (H2b).

Second, we contribute to research on digital BMI by extending the emotion lens on digital BMI by showing that emotion regulation can be a nuanced explanation of the mechanisms leading to digital BMI. Although digital BMI has attracted more and more attention from scholars, there are still few existing studies on the antecedents of business model innovation. This paper explores digital BMI from the perspective of emotion, providing new insights for business model innovation research.

Third, by investigating the moderating role of environmental dynamism in the relationship between emotion regulation and digital BMI, we contribute to the emotion regulation literature by investigating when and how it influence digital BMI in the specific context of entrepreneurship. In particular, we find the dynamics of the entrepreneurial environment as boundary conditions for the relationship between emotion regulation and digital BMI. Our theoretical model integrates individual factors and environment factor to explain how certain aspects of emotion regulation become more or less beneficial given different dynamic levels of entrepreneurial environment, offering a new perspective for further exploring the role of emotion regulation in the entrepreneurial process.

Implications for Practice

Digital technologies have broken down the structural boundaries of the product and service (Nambisan, 2017; Lu et al., 2021b). In such context, firms that only operate technological innovation or product innovation cannot meet the ever-changing needs of customers,they also need to provide new solutions and create new values for customers through digital BMI (Teece, 2010, 2018). However, many firms are facing the challenge of digital

innovation. Through empirical research, we find that the digital BMI of new ventures is influenced by entrepreneurs' emotional regulation, where cognitive reappraisal emotion promotes digital BMI, while expressive suppression emotion inhibits it. Therefore, entrepreneurs can use cognitive reappraisal emotion consciously to gain the support of stakeholders, thus promoting the adoption and implementation of new business models. In addition to provide useful knowledge for using digital BMI, understanding how entrepreneurs regulate their emotions is also important for employees and team members, which enables them to better understand the decisions of entrepreneurs, thus reducing team conflict and increasing communication efficiency.

Limitations and Future Research

Several limitations should be taken into consideration. First, we only focus on two types of emotion regulation (i.e., cognitive reappraisal and expressive suppression), other types of emotion regulation (e.g., attention deployment) may also play a role in digital BMI. Further research may build on our study by investigating the relationship between other important emotion and business model innovation. In addition, studies have shown that individuals can use a variety of emotional regulation at the same time (Gross, 2015). Therefore, future research can explore the impact of combination and sequence of different emotion regulation strategies on the entrepreneurial process. Third, our findings have shown that emotion regulation has a critical effect on digital BMI, which reminds us entrepreneurs or firms should pay attention to the emotional aspects in innovation process. Since advanced digital technologies can help entrepreneur to recognize and regulate own and others' emotion, so it will be an interesting question that how the digital technologies affect the relationship between individual emotion and innovation. Finally, the data in this study are cross-sectional, making it difficult to assert causality. Therefore, we call for longitudinal studies that resolve this problem.

CONCLUSION

In conclusion, our study offers valuable insights on the emergence of digital BMI from emotion lens. We show that cognitive reappraisal and expressive suppression-two different emotion regulation—influence digital BMI, and that this relationship is moderated by environmental dynamism. We believe that our findings promise important insights into our understanding of digital BMI and further explores the role of other emotional mechanism in various entrepreneurial activities.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Department of Technology Economy and Management, Jilin University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

Both authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

FUNDING

This research was supported by the major project of the National Natural Science Foundation of China (NSFC) "Research on Major Theoretical and Practice Issues in Innovation-Driven Entrepreneurship" (Grant No. 72091310), Project 1 "Developing Theory on Innovation-Driven Entrepreneurship in the Digital Economy" (Grant No. 72091315).

REFERENCES

Afuah, A. (2014). *Business Model Innovation: Concepts, Analysis, and Cases*. (New York, NY: Routledge).

Akgün, A. E., Keskin, H., and Byrne, J. (2008). The moderating role of environmental dynamism between firm emotional capability and performance. *J. Organ. Chang. Manage.* 21, 230–252.

Ashforth, B. E., and Humphrey, R. H. (1995). Emotion in the workplace: a reappraisal. *Hum. Relat.* 48, 97–125.

Ashforth, B. E., Rogers, K. M., Pratt, M. G., and Pradies, C. (2014). Ambivalence in organizations: a multilevel approach. *Organ. Sci.* 25, 1453–1478. doi: 10.1287/orsc.2014.0909

Baron, R. A. (2008). The role of affect in the entrepreneurial process. *Acad. Manage. Rev.* 33, 328–340.

Bhatti, S. H., Santoro, G., Khan, J., and Rizzato, F. (2021). Antecedents and consequences of business model innovation in the IT industry. *J. Bus. Res.* 123, 389–400.

Burch, G. F., Batchelor, J. H., and Humphrey, R. H. (2013). Emotional labor for entrepreneurs: a natural and necessary extension. *Entrep. Res. J.* 3, 331–366.

Butler, E. A., Egloff, B., Wilhelm, F. W., Smith, N. C., Erickson, E. A., and Gross, J. J. (2003). The social consequences of expressive suppression. *Emotion* 3, 48–67. doi: 10.1037/1528-3542.3.1.48

Colovic, A. (2021). Leadership and business model innovation in late internationalizing SMEs. *Long Range Plan.* 55:102083.

Davis, B. C., Hmieleski, K. M., Webb, J. W., and Coombs, J. E. (2017). Funders' positive affective reactions to entrepreneurs' crowdfunding pitches: the influence of perceived product creativity and entrepreneurial passion. *J. Bus. Ventur.* 32, 90–106.

De Cock, R., Denoo, L., and Clarysse, B. (2020). Surviving the emotional rollercoaster called entrepreneurship: the role of emotion regulation. *J. Bus. Ventur.* 35:105936. doi: 10.1016/j.jbusvent.2019.04.004

Dess, G. G., and Beard, D. W. (1984). Dimensions of organizational task environments. *Adm. Sci. Q.* 29, 52–73.

Dewald, J., and Bowen, F. (2010). Storm clouds and silver linings: responding to disruptive innovations through cognitive resilience. *Entrep. Theory Pract.* 34, 197–218.

Duncan, R. B. (1972). Characteristics of organizational environments and perceived environmental uncertainty. *Adm. Sci. Q.* 17, 313–327.

Ensley, M. D., Pearce, C. L., and Hmieleski, K. M. (2006). The moderating effect of environmental dynamism on the relationship between entrepreneur leadership behavior and new venture performance. *J. Bus. Ventur.* 21, 243–263.

Faraj, S., Jarvenpaa, S. L., and Majchrzak, A. (2011). Knowledge collaboration in online communities. *Organ. Sci.* 22, 1224–1239. doi: 10.1287/orsc.1100.0614

Feinberg, M., Willer, R., Antonenko, O., and John, O. P. (2012). Liberating reason from the passions. *Psychol. Sci.* 23, 788–795. doi: 10.1177/0956797611434747

Fichman, R. G., Dos Santos, B. L., and Zheng, Z. (2014). Digital innovation as a fundamental and powerful concept in the information systems curriculum. *MIS Q.* 38, 329–354. doi: 10.25300/misq/2014/38.2.01

Foo, M. D., Uy, M. A., and Murnieks, C. (2015). Beyond a?ective valence: untangling valence and activation in?uences on opportunity identifcation. *Enterp. Theory Pract.* 39, 407–431. doi: 10.1111/etap.12045

Fornell, C., and Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: algebra and statistics. *J. Mark. Res.* 18, 382–388.

Foss, N. J., and Saebi, T. (2017). Fifteen years of research on business model innovation: how far have we come, and where should we go? *J. Manage.* 43, 200–227.

Foss, N. J., and Saebi, T. (2018). Business models and business model innovation: between wicked and paradigmatic problems. *Long Range Plann.* 51, 9–21. doi: 10.1016/j.lrp.2017.07.006

George, G., Merrill, R. K., and Schillebeeckx, S. J. (2021). Digital sustainability and entrepreneurship: how digital innovations are helping tackle climate change and sustainable development. *Entrep. Theory Pract.* 45, 999–1027. doi: 10.1177/1042258719899425

Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: divergent consequences for experience, expression, and physiology. *J. Pers. Soc. Psychol.* 74, 224–237. doi: 10.1037//0022-3514.74.1.224

Gross, J. J. (1999). Emotion regulation: past, present, future. *Cogn. Emot.* 13, 551–573. doi: 10.1080/026999399379186

Gross, J. J. (2013). Emotion regulation: taking stock and moving forward. *Emotion* 13:359. doi: 10.1037/a0032135

Gross, J. J. (2014). "Emotion regulation: conceptual and empirical foundations," in *Handbook of Emotion Regulation* 2nd Edn, ed. J. J. Gross (New York, NY: The Guilford Press). 1–20. doi: 10.4337/9781783476855.00008

Gross, J. J. (2015). Emotion regulation: current status and future prospects. *Psychol. Inquiry* 26, 1–26. doi: 10.1254/fpj.151.21

Gross, J. J., and John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *J. Pers. Soc. Psychol.* 85:348. doi: 10.1037/0022-3514.85.2.348

Guo, H., Su, Z., and Ahlstrom, D. (2016). Business model innovation: the effects of exploratory orientation, opportunity recognition, and entrepreneurial bricolage in an emerging economy. *Asia Pac. J. Manage.* 33, 533–549. doi: 10.1007/s10490-015-9428-x

Hartmann, T., and Vanpoucke, E. (2017). User acceptance of technologies in their infancy: the case of 3D printing business models. *J. Organ. End User Comput.* 29, 1–24. doi: 10.4018/joeuc.2017040101

Hmieleski, K. M., and Baron, R. A. (2008). When does entrepreneurial self-efficacy enhance versus reduce firm performance? *Strateg. Entrep. J.* 2, 57–72.

Hmieleski, K. M., and Baron, R. A. (2009). Entrepreneurs' optimism and new venture performance: a social cognitive perspective. *Acad. Manage. J.* 52, 473–488. doi: 10.5465/amj.2009.41330755

Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Modeling* 6, 1–55.

Huy, Q. (1999). Emotional capability, emotional intelligence, and radical change. *Aca. Manage. Rev.* 24, 325–345. doi: 10.2307/259085

Huy, Q., and Zott, C. (2019). Exploring the affective underpinnings of dynamic manageerial capabilities: how manageers' emotion regulation behaviors mobilize resources for their firms. *Strateg. Manage. J.* 40, 28–54.

Huy, Q. N. (2005). "An emotion-based view of strategic renewal," in *Strategy Process (Advances in Strategic Management*, 22, eds G. Szulanski, J. Porac, and Y. Doz (Bingley: Emerald Group Publishing Limited). 3–37. doi: 10.1016/s0742-3322(05)22001-6

Jansen, J. J., Van Den Bosch, F. A., and Volberda, H. W. (2006). Exploratory innovation, exploitative innovation, and performance: effects of organizational antecedents and environmental moderators. *Manage. Sci.* 52, 1661–1674. doi: 10.1287/mnsc.1060.0576

Johnson, M. W. (2010). *Seizing the White Space: Business Model Innovation for Growth and Renewal*. (Boston, MA: Harvard Business Press).

Lieberman, M. D., Inagaki, T. K., Tabibnia, G., and Crockett, M. J. (2011). Subjective responses to emotional stimuli during labeling, reappraisal, and distraction. *Emotion* 11, 468–480. doi: 10.1037/a0023503

Lu, S., Shan, B., and Fei, Y. (2021a). Exploring an inverted U-shaped relationship between entrepreneurial experience and Chinese new venture performance: the moderating role of environmental uncertainty. *Asia Pac. Bus. Rev.* 1–18. doi: 10.1080/13602381.2021.1899464

Lu, X., Shan, B., and Pu, Y. (2021b). Sustainability in the entrepreneurial ecosystem: operating mechanisms and enterprise growth. *Asia Pac. Bus. Rev.* 27, 620–626. doi: 10.1080/13602381.2021.1928941

Ma, Y., and Hu, Y. (2021). Business model innovation and experimentation in transforming economies: bytedance and tiktok. *Manage. Organ. Rev.* 17, 382–388. doi: 10.1017/mor.2020.69

Markman, G. D., Baron, R. A., and Balkin, D. B. (2005). Are perseverance and self-efficacy costless? Assessing entrepreneurs' regretful thinking. *J. Organ. Behav.* 26, 1–19. doi: 10.1002/job.305

McGrath, R. G. (2010). 'Business models: a discovery driven approach'. *Long Range Plann.* 43, 247–261. doi: 10.1016/j.lrp.2009.07.005

Milanov, H., and Fernhaber, S. A. (2009). The impact of early imprinting on the evolution of new venture networks. *J. Bus. Ventur.* 24, 46–61. doi: 10.1016/j.jbusvent.2007.11.001

Mitchell, D., and Coles, C. (2003). The ultimate competitive advantage of continuing business model innovation. *J. Bus. Strategy* 24, 15–22.

Moore, S. A., Zoellner, L. A., and Mollenholt, N. (2008). Are expressive suppression and cognitive reappraisal associated with stress-related symptoms? *Behav. Res. Ther.* 46, 993–1000. doi: 10.1016/j.brat.2008.05.001

Morris, M. H., Kuratko, D. F., Schindehutte, M., and Spivack, A. J. (2012). Framing the entrepreneurial experience. *Entrep. Theory Pract.* 36, 11–40. doi: 10.1111/j.1540-6520.2011.00471.x

Nambisan, S. (2017). Digital entrepreneurship: toward a digital technology perspective of entrepreneurship. *Entrep. Theory Pract.* 41, 1029–1055. doi: 10.1111/etap.12254

Nambisan, S., Lyytinen, K., Majchrzak, A., and Song, M. (2017). Digital innovation manageement: reinventing innovation manageement research in a digital world. *MIS Q.* 41, 223–238. doi: 10.25300/MISQ/2017/41:1.03

Narayan, S., Sidhu, J. S., and Volberda, H. W. (2021). From attention to action: the influence of cognitive and ideological diversity in top manageement teams on business model innovation. *J. Manage. Stud.* 58, 2082–2110.

Nezlek, J. B., and Kuppens, P. (2008). Regulating positive and negative emotions in daily life. *J. Pers.* 76, 561–579. doi: 10.1111/j.1467-6494.2008.00496.x

Obal, M. (2017). What drives post-adoption usage? Investigating the negative and positive antecedents of disruptive technology continuous adoption intentions. *Ind. Mark. Manage.* 63, 42–52. doi: 10.1016/j.indmarman.2017.01.003

Pirola-Merlo, A., Härtel, C., Mann, L., and Hirst, G. (2002). How leaders influence the impact of affective events on team climate and performance in RandD teams. *Leadersh. Q.* 13, 561–581. doi: 10.1016/s1048-9843(02)00144-3

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., and Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J. Appl. Psychol.* 88:879. doi: 10.1037/0021-9010.88.5.879

Raffaelli, R., Glynn, M. A., and Tushman, M. (2019). Frame flexibility: the role of cognitive and emotional framing in innovation adoption by incumbent firms. *Strateg. Manage. J.* 40, 1013–1039. doi: 10.1002/smj.3011

Ray, R. D., McRae, K., Ochsner, K. N., and Gross, J. J. (2010). Cognitive reappraisal of negative affect: converging evidence from EMG and self-report. *Emotion* 10, 587–592. doi: 10.1037/a0019015

Remane, G., Hanelt, A., Nickerson, R. C., and Kolbe, L. M. (2017). Discovering digital business models in traditional industries. *J. Bus. Strat.* 38, 41–51. doi: 10.1108/JBS-10-2016-0127

Saebi, T., Lien, L., and Foss, N. J. (2017). What drives business model adaptation? The impact of opportunities, threats and strategic orientation. *Long Range Plann.* 50, 567–581. doi: 10.1016/j.lrp.2016.06.006

Schein, E. H. (1992). *Organizational Culture and Leadership*. 2nd Edn, (San Francisco, CA: Jossey-Bass).

Shan, B., and Lu, X. (2020). Founder's social ties, learning and entrepreneurial knowledge acquisition in China. *Asia Pac. Bus. Rev.* 26, 209–229.

Shan, B., Pu, Y., Chen, B., and Lu, S. (2021). New technologies' commercialization: the roles of the leader's emotion and incubation support. *Front. Psychol.* 12:710122. doi: 10.3389/fpsyg.2021.710122

Sirén, C., He, V. F., Wesemann, H., Jonassen, Z., Grichnik, D., and von Krogh, G. (2020). Leader emergence in nascent venture teams: the critical roles of individual emotion regulation and team emotions. *J. Manage. Stud.* 57, 931–961.

Snihur, Y., and Zott, C. (2020). The genesis and metamorphosis of novelty imprints: how business model innovation emerges in young ventures. *Acad. Manage. J.* 63, 554–583.

Soluk, J., Miroshnychenko, I., Kammerlander, N., and De Massis, A. (2021). Family influence and digital business model innovation: the enabling role of dynamic capabilities. *Entrep. Theory Pract.* 45, 867–905.

Sorescu, A. (2017). Data-driven business model innovation. *J. Prod. Innov. Manage.* 34, 691–696. doi: 10.1111/jpim.12398

Szasz, P. L., Szentagotai, A., and Hofmann, S. G. (2011). The effect of emotion regulation strategies on anger. *Behav. Res. Ther.* 49, 114–119. doi: 10.1016/j.brat.2010.11.011

Tang, J., Kacmar, K. M., and Busenitz, L. W. (2012). Entrepreneurial alertness in the pursuit of new opportunities. *J. Bus. Ventur.* 27, 77–94. doi: 10.1016/J.JBUSVENT.2010.07.001

Teece, D. J. (2010). Business models. business stmtegy and innovation. *Long Range Plann.* 43, 172–194.

Teece, D. J. (2018). Business models and dynamic capabilities. *Long Range Plann.* 51, 40–49. doi: 10.1016/j.lrp.2017.06.007

Webb, T. L., Miles, E., and Sheeran, P. (2012). Dealing with feeling: a meta-analysis of the e?ectiveness of strategies derived from the process model of emotion regulation. *Psychol. Bull.* 138:775. doi: 10.1037/a0027600

Wolgast, M., Lundh, L.-G., and Viborg, G. (2011). Cognitive reappraisal and acceptance: an experimental comparison of two emotion regulation strategies. *Behav. Res. Ther.* 49, 858–866. doi: 10.1016/j.brat.2011.09.011

Yoo, Y., Henfridsson, O., and Lyytinen, K. (2010). Research commentary: the new organizing logic of digital innovation: an agenda for information systems research. *Info. Sys. Res.* 21, 724–735. doi: 10.1287/isre.1100.0322

Yu, X., Li, Y., Su, Z., Tao, Y., Nguyen, B., and Xia, F. (2020). Entrepreneurial bricolage and its effects on new venture growth and adaptiveness in an emerging economy. *Asia Pac. J. Manage.* 37, 1141–1163. doi: 10.1007/s10490-019-09657-1

Zhang, H., Xiao, H., Wang, Y., Shareef, M. A., Akram, M. S., and Goraya, M. A. S. (2021). An integration of antecedents and outcomes of business model innovation: a meta-analytic review. *J. Bus. Res.* 131, 803–814.

Zhao, Y., and Zhou, Y. (2022). Measurement method and application of a deep learning digital economy scale based on a big data cloud platform. *J. Organ. End User Comput.* 34, 1–17.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Lu and Yu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 07 April 2022
doi: 10.3389/fpsyg.2022.833007

Cross-Modal Transfer Learning From EEG to Functional Near-Infrared Spectroscopy for Classification Task in Brain-Computer Interface System

Yuqing Wang[1], Zhiqiang Yang[1], Hongfei Ji[1], Jie Li[1], Lingyu Liu[2]* and Jie Zhuang[3]**

[1] Shanghai Yangzhi Rehabilitation Hospital (Shanghai Sunshine Rehabilitation Center), Collage of Electronic and Information Engineering, Tongji University, Shanghai, China, [2] Department of Neurorehabilitation, Shanghai Yangzhi Rehabilitation Hospital (Shanghai Sunshine Rehabilitation Center), School of Medicine, Tongji University, Shanghai, China, [3] School of Psychology, Shanghai University of Sport, Shanghai, China

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Lihui Wang,
Shanghai Jiao Tong University, China
Lanlan Chen,
East China University of Science and Technology, China
Isao Nambu,
Nagaoka University of Technology, Japan

***Correspondence:**
Jie Li
nijanice@163.com
Lingyu Liu
Happyneurologist@163.com
Jie Zhuang
jzhuang255@163.com

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *10 December 2021*
Accepted: *07 March 2022*
Published: *07 April 2022*

Citation:
Wang Y, Yang Z, Ji H, Li J, Liu L and Zhuang J (2022) Cross-Modal Transfer Learning From EEG to Functional Near-Infrared Spectroscopy for Classification Task in Brain-Computer Interface System. Front. Psychol. 13:833007. doi: 10.3389/fpsyg.2022.833007

The brain-computer interface (BCI) based on functional near-infrared spectroscopy (fNIRS) has received more and more attention due to its vast application potential in emotion recognition. However, the relatively insufficient investigation of the feature extraction algorithms limits its use in practice. In this article, to improve the performance of fNIRS-based BCI, we proposed a method named R-CSP-E, which introduces EEG signals when computing fNIRS signals' features based on transfer learning and ensemble learning theory. In detail, we used the Independent Component Analysis (ICA) algorithm for the correspondence between the sources of the two signals. We then introduced the EEG signals when computing the spatial filter based on a modified Common Spatial Pattern (CSP) algorithm. Experimental results on public datasets show that the proposed method in this paper outperforms traditional methods without transfer. In general, the mean classification accuracy can be increased by up to 5%. To our knowledge, it is an innovation that we tried to apply transfer learning between EEG and fNIRS. Our study's findings not only prove the potential of the transfer learning algorithm in cross-model brain-computer interface, but also offer a new and innovative perspective to research the hybrid brain-computer interface.

Keywords: fNIRS signals, transfer learning, brain computer interface, ICA, RCSP

INTRODUCTION

The brain-computer interface (BCI) is a communication way between the brain of users and the outside world, which can be used to exchange information between the brain and the equipment (Khosrowabadi et al., 2011). It has been employed in many fields, such as neuronal rehabilitation (Daly and Wolpaw, 2008), military (Rebsamen et al., 2007), traffic (Yue and Wang, 2019), and entertainment (Royer et al., 2010). In BCI, motor imagery is the most investigated paradigm. Many previous studies have demonstrated that the motor imagery brain-computer interface system (MI-BCI) has a broad prospect in rehabilitation training and neural recovery (Pichiorri et al., 2015; Bundy et al., 2017).

Among most studies about the MI-BCI system, researchers mainly focus on EEG signals, which record neural activity in the brain (Wolpaw et al., 2006). Considering its non-invasiveness and high

temporal resolution (milliseconds), EEG is the most popular brain signal and provides encouraging results. fNIRS is a functional brain imaging tool, which reflects the activation degree of the cerebral cortex by measuring the level of oxygen in the blood (Ferrari and Quaresima, 2012). In contrast to EEG, fNIRS provides better spatial resolution and shows better tolerance to motion artifacts. The fNIRS-based BCI was mentioned firstly by Coyle et al. (2004). With the development of brain science and breakthroughs in cognitive neuroscience technology, fNIRS has developed rapidly as a tool to study cognitive neuroscience in the past decades.

However, as we know, due to the lower time resolution and lack of feature extraction algorithms, the fNIRS-based BCI system usually has poor performance. Common Space Pattern (CSP) is an effective method to extract features in the EEG-based BCI (Spuler et al., 2014; Wu et al., 2017). However, fNIRS signal is different from EEG. For EEG, it is a signal reflecting brain electrophysiology with rich high-frequency fluctuations while fNIRS is a slowly changing signal detecting cerebral hemodynamic response. Therefore, applying the CSP algorithm to fNIRS signals directly usually leads to poor effect. In article (Zhang et al., 2017), CSP algorithm was utilized to integrate multi-channel fNIRS signals into a single temporal sequence with the usage of an optimal spatial filter and subsequently, variance was extracted as part of the feature vector for brain state classification. They reported this method can improve the classification accuracy more than 9% in average. In our research, to improve the CSP algorithm in fNIRS, we introduced EEG signals based on the theory of transfer learning.

In addition, because fNIRS relies on cerebrovascular dynamics, the response is not as sensitive as EEG. Taking motor imagery as an example, the change of hemodynamic concentration is delayed relative to the start and end of the movement lasting about 5 s (Lachert et al., 2017). This results in the fact that it takes longer than EEG to collect a certain amount of data during the experimental session. Considering that the experiment time is too long, the subjects will be tired, which also limits the classification accuracy and application of fNIRS-BCI. Previous studies often used deep learning technology to improve classification accuracy. In the article (Trakoolwilaiwan et al., 2017), a convolutional neural network (CNN) was utilized to automatically extract features and classify in left or right-hand motor execution task. Results showed that the CNN-based methods improve the classification accuracy compared to traditional methods. Another popular approach is integrating the fNIRS signals with the EEG signals. EEG is one of the electrophysiological signals, and fNIRS measures the level of oxygen in the blood. Hence, combining the two signals could offer us different types of data that involved the same brain activity. Experimental results revealed that the fusion method could indeed enhance the performance of BCI (Fazli et al., 2012; Lee et al., 2019).

Recently, transfer learning (TL) has been widely discussed in the BCI system and has gradually become a new research hotspot (Jayaram et al., 2015). The existing methods are mainly applied in the EEG brain-computer interface system, which promotes learning a new subject/session/equipment/task by utilizing the data or knowledge from similar or relevant subjects/sessions/equipment/tasks (Wu et al., 2022). It dramatically solves the dilemma of small training data, significant individual differences, and few executable commands. Transfer learning is currently the most crucial method to reduce calibration in BCI. In this article (Khazem et al., 2021), a transfer learning approach called MDWM was proposed, which aims to reduce calibration time by transferring knowledge from other subjects to the target user. Similarly, this paper described a multi-source fusion transfer learning algorithm (Liang and Ma, 2020). However, there is little research in the multimodal brain-computer interface that involved the thought of transfer learning.

In this paper, to achieve TL between EEG and fNIRS, we proposed a method called R-CSP-E based on the Regularized Common Spatial Pattern (RCSP) theory, which states that the weighted covariance matrix can be transferred from EEG to fNIRS. RCSP, as a variant of the CSP algorithm, is used for cross-subject adaption (Xu et al., 2019). EEG signals and fNIRS signals were recorded simultaneously under the same thinking task activity. We believe that there is a certain amount of information between the two signals that can be transferred and utilized. Besides, our method also embodies the idea of ensemble learning. The results of multi-experiments showed that our method could recognize the characteristic patterns of the brain better and improve the classification accuracy of motor imagination tasks.

In summary, the contribution of our paper lies in the following three aspects. Firstly, we use the ICA algorithm to trace the source and find the region with the highest activation level to establish the correspondence between channels. Secondly, inspired by the thought of TL, we consider the features of the EEG signals to improve the performance of the brain-computer interface system based on fNIRS. Finally, we provide a new perspective to research the hybrid brain-computer interface.

The remainder of the sections of our paper is organized as follows: section "Materials and Methods" describes the dataset and methodology used in the experiment. Section "Experiments" is devoted to showing the experimental design and results, and sections "Results" and section "Discussion" are the discussion and conclusion of this article.

MATERIALS AND METHODS

Dataset

An open-access dataset for EEG-fNIRS hybrid brain-computer interface with two brain activities, motor imagery (MI) and mental arithmetic (MA), was involved in our research (Shin et al., 2017). In our work, we mainly made full use of the MI dataset. For the EEG, the signals were recorded at 1000 Hz. As shown in **Figure 1**, thirty EEG active electrodes were placed according to the international 10-5 standard (AFp1, AFp2, AFF1h, AFF2h, F3, F4, F7, F8, FCz, FCC3h, FCC4h, FCC5h, FCC6h, T7, T8, Cz, CCP3h, CCP4h, CCP5h, CCP6h, CPz, Pz, P3, P4, P7, P8, PPO1h, PPO2h, POO1, POO2, and Fz for ground electrode) (Oostenveld and Praamstra, 2001). For fNIRS, fourteen sources and sixteen detectors represented thirty-six channels with a 12.5 Hz sampling

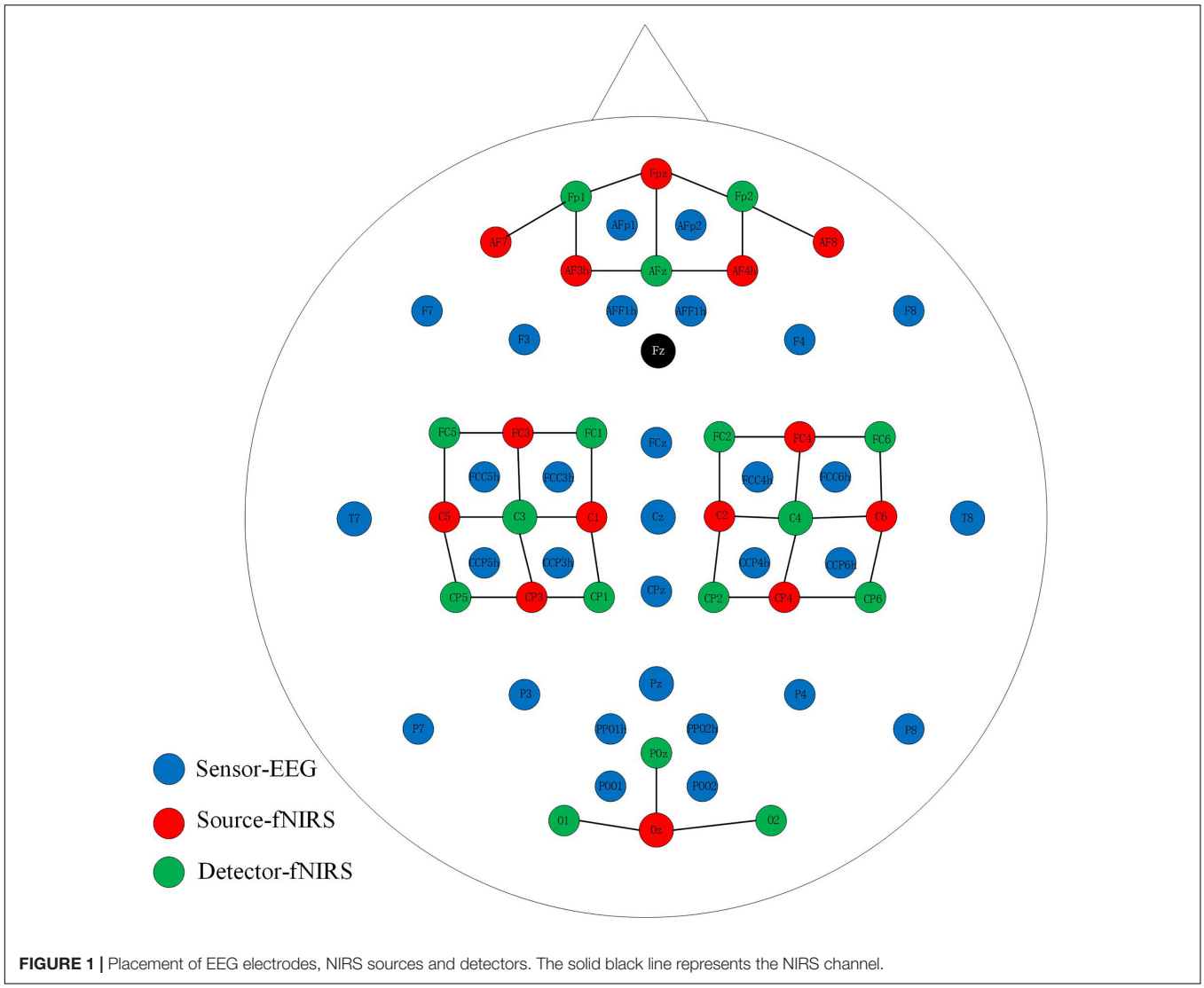


FIGURE 1 | Placement of EEG electrodes, NIRS sources and detectors. The solid black line represents the NIRS channel.

rate. The detail and specific placement of electrodes can refer to this article (Shin et al., 2017). In general, placed channels near the frontal, motor, and visual areas. In order to realize the synchronous acquisition of signals, the fNIRS optodes and EEG electrodes were fixed on the same fabric cap.

All subjects sat on a chair 50 inches from the screen and asked to remain still and relaxed. Participants were instructed to execute left or right-hand squeezing imagery (i.e., to imagine the finger's movement up and down in the brain) by visual stimulation. Each subject performed three sessions (each session contains twenty trials) of left or right-hand MI tasks. As is illustrated in **Figure 2**, a single trial consisted of instruction (2s), task (10 s), and rest (15 task (periods. The left or right arrow was displayed on the computer screen during the instruction.

Pre-processing

All data processing was completed by MATLAB R2019b, and the BBCI toolbox was also applied for further data processing (Benjamin et al., 2010). The original EEG signals were re-referenced firstly and filtered with a passband of 0.5–50 Hz. When we collect EEG signals, there will be some unavoidable noises, such as electrooculogram (EOG), electromyogram (EMG). These signals will cause interference to our brain electricity. Hence, in data processing, these interference signals need to be eliminated. Then we performed the independent component analysis (ICA)-based artifact rejection to remove the ocular artifacts. This method assumes that the observed signals are composed of clean EEG data and independent noise components.

Down-sampled the EEG signals to 200 Hz and filtered between 8 and 30 Hz. Then normalized by Z-score and segmented into epochs ranging from −3 to 7 s. After that, subtract the mean value of the period ranging from −3 to 0s to correct the baseline. Electrodes attached to the motor cortex (FCC5h, FCC6h, FCC3h, FCC4h, CCP5h, CCP6h, CCP3h, CCP4h) were selected for further analysis.

Raw fNIRS signal is filtered at 0.01–0.1 Hz to remove internal noise generated by physiological activities such as respiration

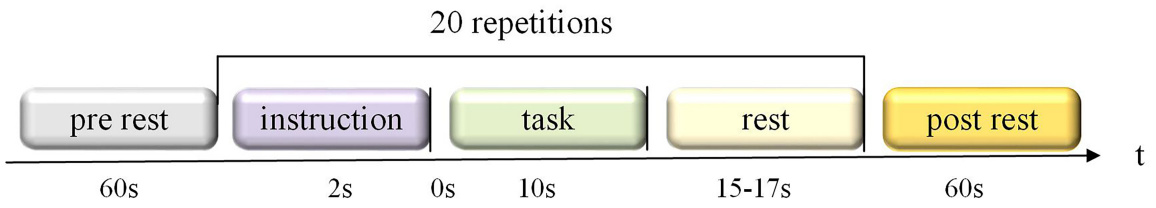


FIGURE 2 | Task sequence of experimental paradigm.

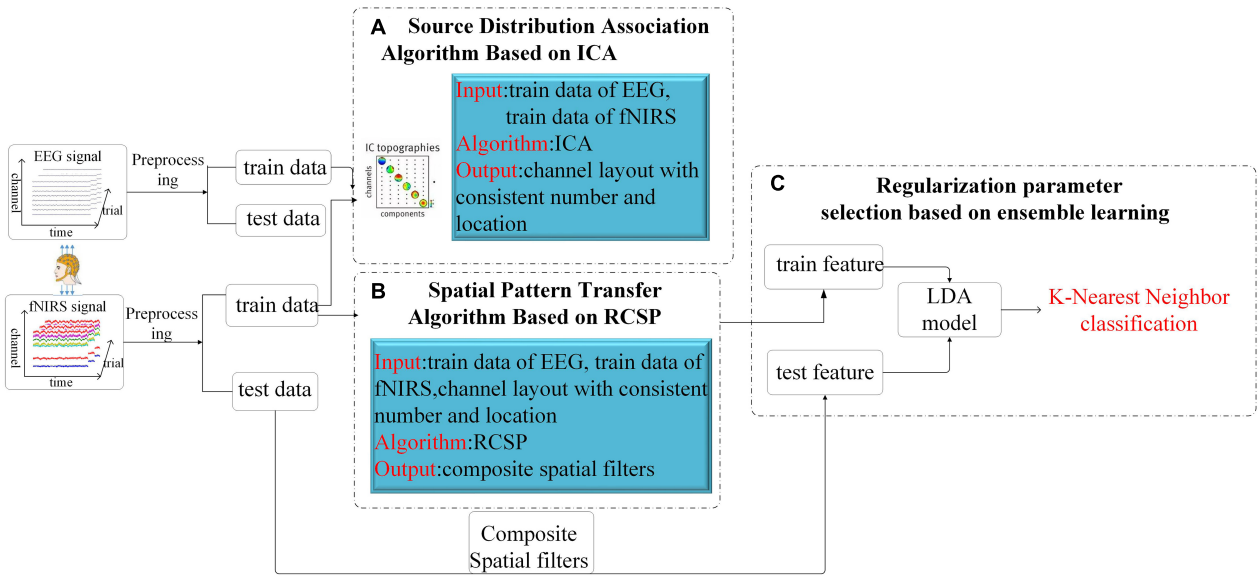


FIGURE 3 | The total structure of the proposed Algorithm. After the EEG signal and the fNIRS signal are preprocessed respectively, they are divided into a training set and a test set. The training set data is adjusted by the ICA algorithm for channel order **(A)**, then the RCSP framework is used for feature generation **(B)**, and finally combined with LDA and KNN for classification **(C)**.

and heartbeat. After that, the filtered data are converted to a concentration of deoxy- and oxy-hemoglobin (HbR and HbO) according to the modified Beer-Lambert law (Kocsis et al., 2006). Then, considering that the hemodynamic response of fNIRS is relatively slow, the HbO and HbR signals are divided into periods of 0–10 s. Finally, we conducted the baseline correction by diminishing the mean value from 0 to 3 s. Similar to EEG, twenty-four channels near the motor cortex were used for further NIRS data processing.

Methods

This part will introduce the overall framework of our algorithm and then describe the three essential components in detail. The whole scheme of our algorithm is shown in **Figure 3**, which consists of two parts. The first part is the model training. After preprocessing, the synchronously collected multimodal data are divided into the training and test datasets. After adjusting the channel through the proposed ICA-based source distribution association algorithm, the EEG training dataset is introduced into the calculation of the spatial filter and then constructed the composite spatial filters from the fNIRS data based on the RCSP framework. In RCSP, regularization parameter selection is crucial. In our study, this problem is solved by the idea of ensemble learning. Finally, the features extracted by the filter are input into Linear Discriminant Analysis (LDA) for dimensionality reduction, and K-nearest neighbors (KNN) is used for classification. Another part is the model testing. Test data is fed to the model to calculate the accuracy.

Source Distribution Association Algorithm Based on Independent Component Analysis

To realize the transfer between EEG and fNIRS signals, we first need to adjust the positions and numbers of the two signals' channels to be consistent. Although EEG and fNIRS channels show one-to-many correspondence, which means that each EEG channel is associated with multiple fNIRS channels. But for the convenience of calculation, we need to make a one-to-one correlation. Therefore, this paper proposes channel matching from the perspective of source distribution. This section will introduce the channel adjustment strategy in detail. ICA is widely used as a popular technique to remove artifacts in the BCI field. In addition, for the motor imagery task, because the interstitial

activity is significant and its source can usually be considered specific, the ICA algorithm is also suitable for extracting the exciting regions. ICA is a generative model, which describes how to generate observation data by mixing independent components (Hyvarinen and Oja, 2000). Considering the calculation speed, the ICA algorithm used in this article is FastICA (Dominic et al., 2010).In our study, we first obtain the source signal through the ICA algorithm. We believe that both EEG and fNIRS are based on motor imagery tasks, and the brain regions where the source signals generated have some certain consistency. We then matched channels by correlation of source signals.

As described in **Algorithm 1**, we first used the ICA algorithm on the EEG signals to get eight independent sources (the number of channels is eight), which were marked as IC_1∼IC_8 in turn. Then, draw the power spectral density of these independent components. As shown in **Figure 4**, the power spectra of different independent components vary greatly. Calculate the distance between different types of power spectral density (PSD) curves on the power spectrum (the distance is calculated as the mean distance across all the frequencies). Then select the most obvious components that distinguish the left and right hands, and treat them as the most relevant source of classification, which is marked as S_{eeg1}, S_{eeg2}.

Algorithm 1 | Source Distribution Association Algorithm Based on ICA.

Input: The multi-channel EEG signals, E_n; The multi-channel fNIRS signals, F_n;

Output: Channel layout with consistent number and location, *ch*;

1: Extract the set of independent sources of E_n based on ICA algorithm, named IC_i;

2: Draw power spectral density on these independent components.

3: Select the two most obvious components that distinguish the left and right hands, which are marked as S_{eeg1}, S_{eeg2}

4: Extract the set of independent sources of F_n;

5: Calculated the change of the source signals' oxyhemoglobin concentration between the task period and rest period;

6: Pick up the two components with the largest variation. Denoted these two components as S_{fnirs1}, S_{fnirs2};

7: Observe the columns in the mixing matrix belong to the four source signals, and sort the channels according to their weights;

8: **return** *ch*;

As shown in **Figure 5**, in the fNIRS data, using the left and right-handed tasks did not work because the response of the fNIRS signal was slower than that of the EEG signal. For the fNIRS signals, ICA decomposition was performed first and then calculated the change of the source signal's oxyhemoglobin

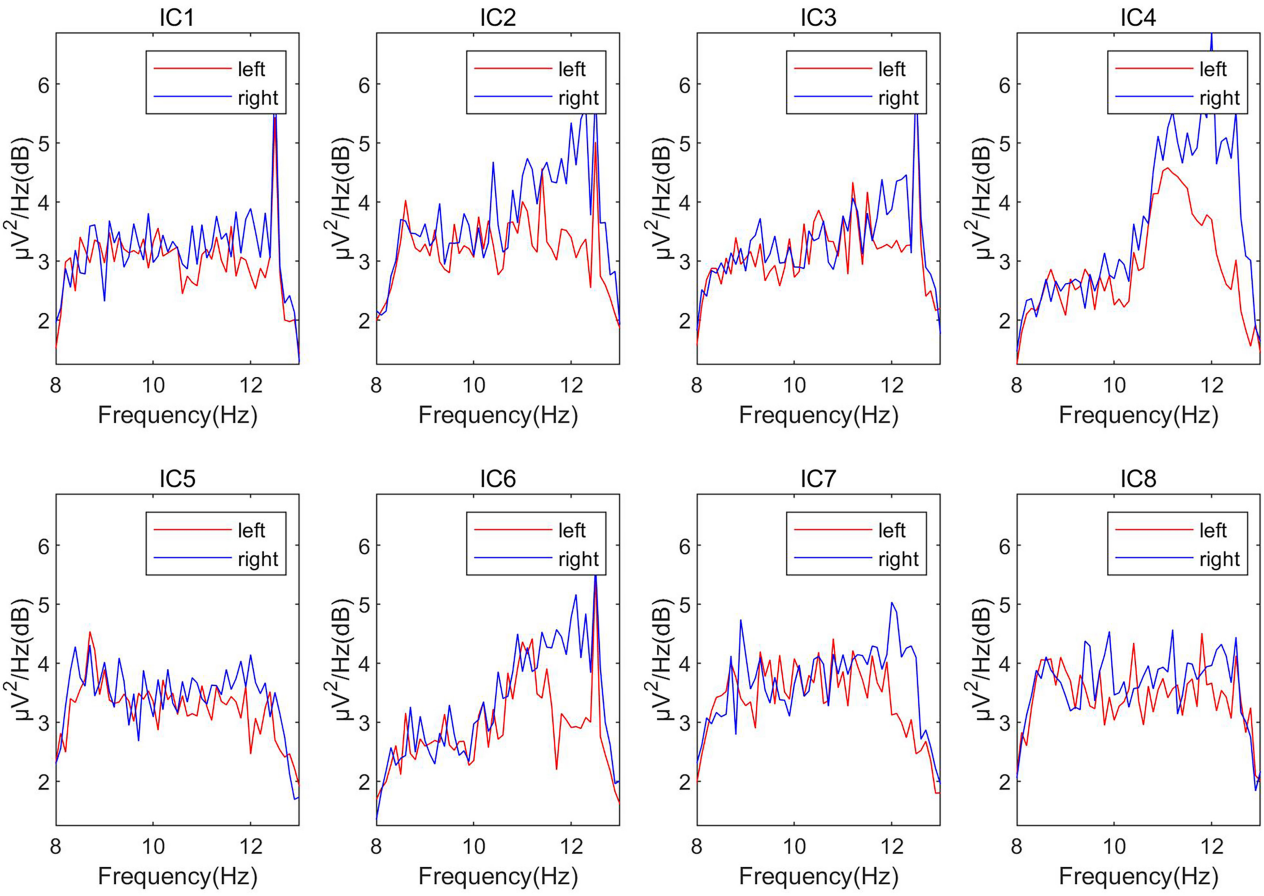


FIGURE 4 | The power spectra of the independent components.

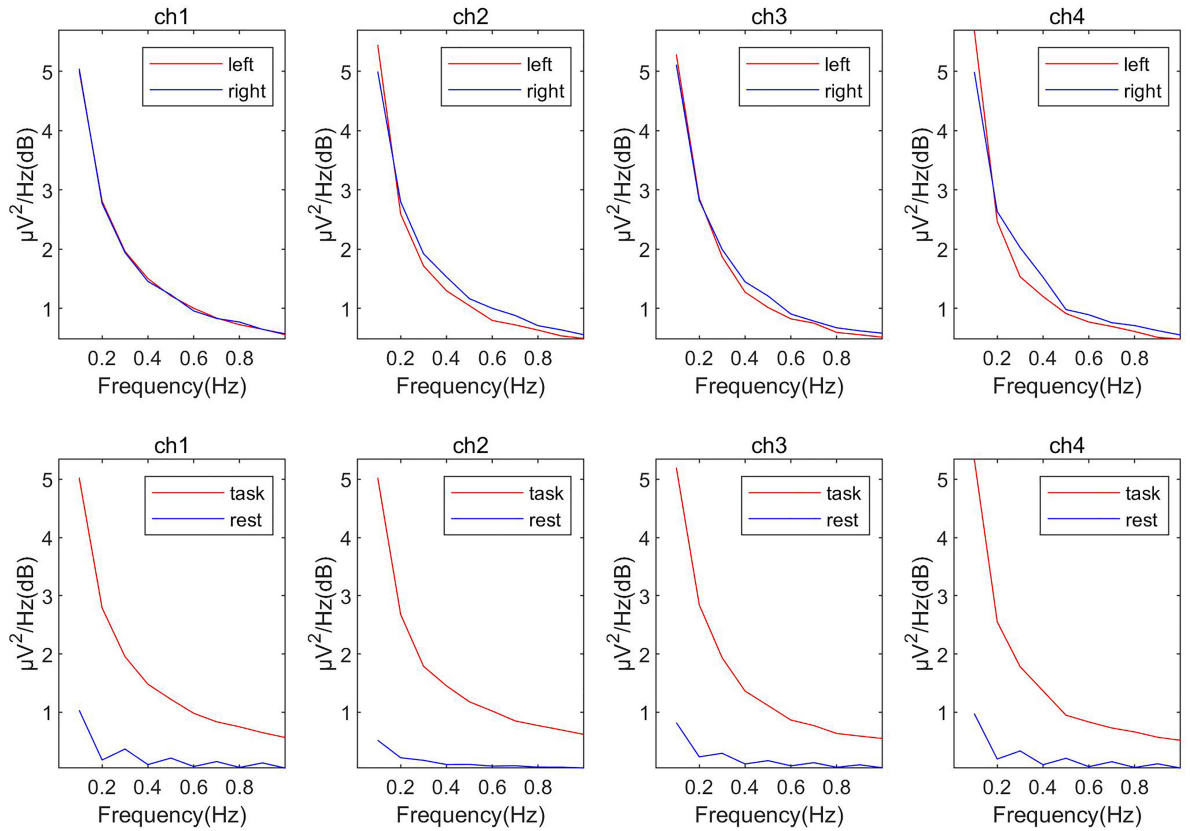


FIGURE 5 | The power spectrum comparison of left and right handed and task-resting in fNIRS data.

concentration between the task period and rest period. Like EEG, the two components with the largest variation were picked up as the most relevant sources of classification. Denoted these two components as S_{fnirs1}, S_{fnirs2}.

Usually, when we compare the columns of the mixing matrix A in ICA, we can analyze how much the source data contributes to each channel of the observation data. Therefore, the columns of the mixing matrix can be used to calculate and visualize the topography of components. Motivated by the theory above, we observed the columns in the mixing matrix belong to the four source signals and sorted the channels according to their weights. For example, the pattern diagram of S_{eeg1} is shown in **Figure 6**. We sorted the channels according to the weight, the order of the channels is "FCC5h," "FCC6h," "CCP4h," "CCP5h," "FCC3h," "FCC4h," "CCP6h," "CCP3h." We deduced the channel sequence corresponding to the other three source signals by analogy. By combining them, the purpose of matching the EEG optode and the fNIRS channel was achieved.

Spatial Pattern Transfer Algorithm Based on Regularized Common Spatial Pattern

Common spatial patterns (CSP), which is regarded as the most effective feature extraction algorithm in the MI task, aims to extract the spatial distribution components of each class. The core of CSP is designing a set of spatial filters, thus the variance of one type of signal is maximized, and the other type is minimized, to obtain the eigenvectors with higher discrimination (Lu et al., 2009b).

Suppose X_1 and X_2 are the multi-channel signals during the motor imagery task, and their dimensions are both N × T (N is the number of channels, and T is the number of samples within each channel). The average spatial covariance matrix of each class can be expressed as Equation (1):

$$R_1 = \frac{1}{k} * \sum_i^k \frac{X_1^i (X_1^i)^T}{trace(X_1^i (X_1^i)^T)},\ R_2 = \frac{1}{k} * \sum_i^k \frac{X_2^i (X_2^i)^T}{trace(X_2^i (X_2^i)^T)} \quad (1)$$

k represents the number of trials, X^T means the transposition of X, and trace (X) is the sum of the elements on the diagonal of the matrix. By calculating the average covariance matrix of each class separately, the composite spatial covariance is obtained.

$$R = R_1 + R_2 \quad (2)$$

Then, according to Equation (3), the eigenvalue decomposition is performed on the composite spatial covariance matrix R.

$$R = U \lambda U^T \quad (3)$$

Among them, U is the eigenvector matrix, and λ is the diagonal matrix composed of the corresponding eigenvalues.

Rearrange the eigenvalues in descending order, and then obtained the whitening transformation.

$$P = \sqrt{\lambda^{-1}} U^T \tag{4}$$

After the whitening transformation, R1 and R2 can be written as S1and S2, respectively.

$$S_1 = PR_1P^T,\ S_2 = PR_2P^T \tag{5}$$

Then, S1 and S2 can be factorized according to Equation (6), where λ1 and λ2 are the diagonal matrices composed of eigenvalues.

$$S_1 = B_1\lambda_1 B_1^T,\ S_2 = B_2\lambda_2 B_2 \tag{6}$$

Through the above formula, it can be proved that the eigenvectors of matrix S_1 and S_2 are equal, namely:

$$B_1 = B_2 = V \tag{7}$$

Meanwhile, the sum of λ_1 and λ_2 is the identity matrix. This attribute makes the eigenvectors B useful to distinguish the different classes.

$$\lambda_1 + \lambda_2 = I \tag{8}$$

$\tilde{B}$ is concatenated by the first m and last m feature vectors in B (m is equal to 5). Next, calculate the projection matrix according to Equation (9). It contains 2m spatial filters.

$$W = \tilde{B}^T P \tag{9}$$

The feature vectors of the i_{th} trial as follows, where $var\,(Z_i)$ is the variance.

$$Z_i = W * X_i,\ f_i = log\left(\frac{var(Z_i)}{sum(var(Z_i))}\right) \tag{10}$$

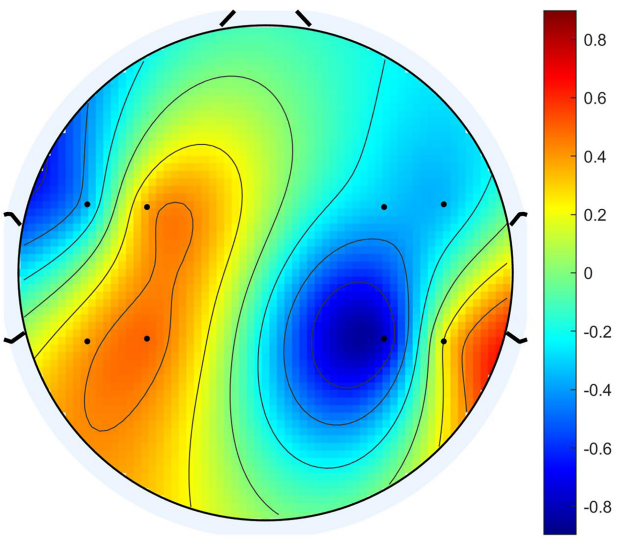


FIGURE 6 | The pattern diagram of Seeg1.

Based on the theory of CSP and matrix regularization, researchers proposed the framework of RCSP to overcome the shortcomings of the traditional CSP model (Lu et al., 2009b). In our work, the regularized average spatial covariance matrix of each class is given as Equation (11).

$$\widehat{\Sigma}_c(\beta,\gamma) = (1-\gamma)\,\widehat{\Omega}_c(\beta) + \frac{\gamma}{N} tr\left[\widehat{\Omega}_c(\beta)\right] \cdot I \tag{11}$$

where β and γ are both regularization parameters ($0 \leq \beta \leq 1$, $0 \leq \gamma \leq 1$), and I is N × N identity matrix. $\widehat{\Omega}_c(\beta)$concludes the covariance matrices from the fNIRS signals as well as EEG signals. It is defined as Equation (12).

$$\widehat{\Omega}_c(\beta) = \frac{(1-\beta)R_c + \beta\hat{R}_c}{(1-\beta)M_c + \beta\hat{M}_c} \tag{12}$$

where R_c is the sum of covariance matrices for all M_c training trials of the fNIRS signals in class c, and $\hat{R}_c$ is the sum of the covariance matrices for $\hat{M}_c$ generic training trials from EEG signals in class c. The composite spatial covariance in R-CSP is formed and factorized as Equation (13).

$$R = \widehat{\Sigma}_1(\beta,\gamma) + \widehat{\Sigma}_2(\beta,\gamma) = U\lambda U^T \tag{13}$$

TABLE 1 | Classification accuracy (%) of all subjects of different algorithms in experiment 1.

Subject	Algorithm		
	CSP	R-CSP-CV	R-CSP-E
1	83.3	78.3	85.0
2	65.0	71.7	78.3
3	75.0	73.3	83.3
4	71.7	75.0	76.7
5	66.7	65.0	75.0
6	63.3	66.7	70.0
7	71.7	75.0	81.7
8	68.3	68.3	75.0
9	75.0	71.7	78.3
10	80.0	78.3	83.3
11	71.3	75.0	78.3
12	76.7	78.3	83.3
13	63.3	65.0	70.0
14	71.7	76.7	81.7
15	68.3	73.3	75.0
16	71.7	78.3	85.0
17	55.0	61.3	73.3
18	81.3	78.3	86.7
19	68.3	70.0	78.3
20	66.7	71.7	68.3
21	71.7	78.3	83.3
22	58.3	61.3	65.0
23	73.3	65.0	75.0
24	65.0	76.7	81.7
25	75.0	71.7	80.0
26	81.3	85.0	91.7
27	71.7	73.3	80.0
28	56.7	55.0	63.3
29	68.3	75.0	81.7

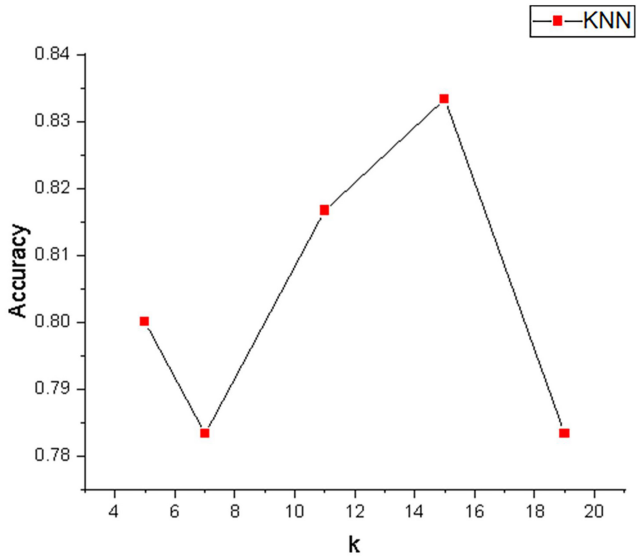


FIGURE 7 | The average accuracy of classification on different *k*-value.

In this study, we tried to apply this approach in transfer between EEG and fNIRS. the EEG signal is introduced according to Equation (12) when computing the spatial filter for fNIRS in the RCSP algorithm. The spatial pattern features of EEG are incorporated into fNIRS to realize the transfer of feature patterns.

Regularization Parameter Selection Based on Ensemble Learning

The critical problem of the RCSP algorithm proposed above is the determination of the regularization parameters. This question is crucial because it is challenging to know the optimal values of the parameters in advance. In earlier work, this problem is often solved through cross-validation. Our paper adopts the approach of ensemble learning for regularization parameter determination, named R-CSP-E. Ensemble learning refers to training multiple individual learners for training data, and finally we can form a powerful learner based on a certain combination strategy (Zhang and Ma, 2012). The commonly used combination strategy in regression tasks is the averaging method; the outputs of several weak learners are averaged to obtain the final predicted result. As for classification tasks, we usually use voting methods. The simplest voting method is the relative majority voting method, which is what we often say that the minority obeys the majority.

In R-CSP-E, a set of regularization parameter pairs from the interval [0,1] were utilized instead of applying fixed parameters. We recorded the results from different regularization parameters and combined them to form an aggregation solution, which reveals the theory of ensemble learning. Different regularization parameter pairs will produce different discriminative characteristics. Such diversity is conducive to training the classifier and improving the performance of the BCI system, based on the principle of boosting. The combination scheme originated from ensemble learning, which has been mainly developed in the following ways: feature, matching score, and ensemble-based learning, such as boosting (Lu et al., 2009a). In our study, the KNN classification algorithm was employed. This method embodies the idea of ensemble learning, but it is not ensemble learning in the strict sense. The principle of KNN is that if most of its K nearest neighbors in the training set belong to a specific category, the sample also belongs to this category. K is usually an odd number not greater than 20. In our algorithm, KNN is used as a classifier, and the feature dimension of training data (test data) is all 30 (the number of combinations of regularization parameters). For the test data, the distances to all dimensions of the training data are calculated and summed, and finally the labels of the k nearest training data to the test data are counted to complete the classification.

EXPERIMENTS

In this section, we carried out a set of experiments to investigate how the accuracy of fNIRS signals classification is affected by the EEG signals. As mentioned, the experiments were carried out on an open-access dataset. We only used the MI data, and the motor cortex was regarded as the region we were interested. Considering the differences between trials, we applied the five-fold cross-validation technique, which means that the training set of each iteration contains 48 trials, and the test set contains 12 trials. Besides, based on the study in Xu et al. (2019),we selected six values for β and five values for γ,which forms $A = 6 \times 5 = 30$ different parameter combinations. This setting for R-CSP-E was used in all experiments in the following.

$$\beta \in \{0, 0.01, 0.1, 0.2, 0.4, 0.6\}, \gamma \in \{0, 0.001, 0.01, 0.1, 0.2\} \quad (14)$$

Two experiments are executed as detailed in the following.

(1) To evaluate the effect of the R-CSP-E, we applied this method on EEG signals. R-CSP-E is a combination

of cross-modal transfer learning and ensemble learning. In experiment I, only EEG data was involved, which means both R_c and $\hat{R}_c$ are EEG data in Equation (12). Hence, the results for experiment 1 showed the effectiveness of ensemble learning. We compared the accuracy of R-CSP-E with the conventional CSP algorithm as well as other classification algorithms, such as R-CSP-CV. In R-CSP-CV, the training set is divided into n parts, each part is used as a validation set in turn, and the remaining n-1 parts are used as a training set. In this way, for n times of training, the average value of the errors obtained in n times is taken as the final error of the model, so as to select the optimal model parameters. To investigate R-CSP-E, we also compared the impact of different values of k on the classification results.

(2) To study the transfer learning in EEG and fNIRS, we compared the correct classification rate of fNIRS before and after adding EEG signals. In detail, we calculated the mean value of the HbO concentration in each channel. Then, we extracted the features of channel-wise, conventional CSP algorithm and R-CSP-E separately.

TABLE 2 | Classification accuracy (%) of all subjects of different algorithms in experiment 2.

Subject	Algorithm			
	HbO	CSP	R-CSP-E	R-CSP-CV
1	45.0	51.7	58.3	50.0
2	71.7	63.3	66.7	65.0
3	63.3	55.0	61.7	53.3
4	60.0	61.7	68.3	63.3
5	61.7	65.0	71.7	58.3
6	48.3	50.0	58.3	51.7
7	61.7	63.3	68.3	65.0
8	66.7	68.3	71.7	63.3
9	55.0	61.3	66.7	58.3
10	60.0	55.0	63.3	56.7
11	55.0	61.3	68.3	63.3
12	58.3	63.3	65.0	55.0
13	61.7	60.0	68.3	58.3
14	55.0	68.3	71.7	65.0
15	61.7	63.3	70.0	65.0
16	60.0	66.7	68.3	61.7
17	48.3	50.0	58.3	53.3
18	53.3	63.3	73.3	60.0
19	55.0	58.3	61.7	55.0
20	61.7	65.0	68.3	58.3
21	63.3	70.0	65.0	61.7
22	51.7	55.0	58.3	51.3
23	68.3	65.0	71.7	65.0
24	51.7	53.3	63.3	55.0
25	60.0	51.7	65.0	48.3
26	61.7	65.0	73.3	65.0
27	58.3	63.3	68.3	61.7
28	68.3	66.7	71.7	65.0
29	66.7	63.3	73.3	58.3

In the above experiments, in R-CSP-E, we chose the LDA algorithm for dimensionality reduction. After LDA, the feature dimension is [number of trials] by [number of regularization parameters combinations]. KNN is used as the classifier. Finally, the average accuracy was shown, because we applied five-fold cross-validation. In other methods, such as CSP, KNN is also used as the classifier to make sure the classifiers are consistent. In R-CSP E, the input dimension for classification module is 60×30 (60 is the number of trials, 30 is the number of regularization parameter combinations). In R-CSP-CV and CSP, the input dimension for classification module is 60×10 (60 is the number of trials, 10 is 2 m). In HbO, the input dimension for classification module is 60×24 (60 is the number of trials, 24 is the number of channels).

In order to prove whether the effect of the R-CSP-E algorithm is significantly different, this paper uses the *t*-test (the full name is independent sample *t*-test) to study the difference in the accuracy of different algorithms for the results of experiment 2. The independent samples *t*-test is used to compare the means of two independent samples, which assumes that the difference between the two-sample means is equal to 0. Calculates the difference between two means and the confidence interval (CI) for that difference. Next follow the test statistic t, degrees of freedom (DF), and two-tailed probability P. When the *p*-value is less than the usual 0.05, the null hypothesis is rejected and it is concluded that the two means are indeed significantly different.

TABLE 3 | Independent *t*-test between HbO and R-CSP-E.

	Algorithm (Mean ± SD)		*t*	*p*
	HbO (*n* = 29)	R-CSP-E (*n* = 29)		
Accuracy	59.08 ± 6.45	66.83 ± 4.81	−5.183	0.000**

*$p < 0.05$, **$p < 0.01$.

TABLE 4 | Independent *t*-test between CSP and R-CSP-E.

	Algorithm (Mean ± *SD*)		*t*	*p*
	CSP (*n* = 29)	R-CSP-E (*n* = 29)		
Accuracy	60.94 ± 5.81	66.83 ± 4.81	−4.203	0.000**

*$p < 0.05$, ** $p < 0.01$.

TABLE 5 | Independent *t*-test between R-CSP-CV and R-CSP-E.

	Algorithm (Mean ± SD)		*t*	*p*
	R-CSP-CV (*n* = 29)	R-CSP-E (*n* = 29)		
Accuracy	59.00 ± 5.21	66.83 ± 4.81	−5.946	0.000**

*$p < 0.05$, **$p < 0.01$.

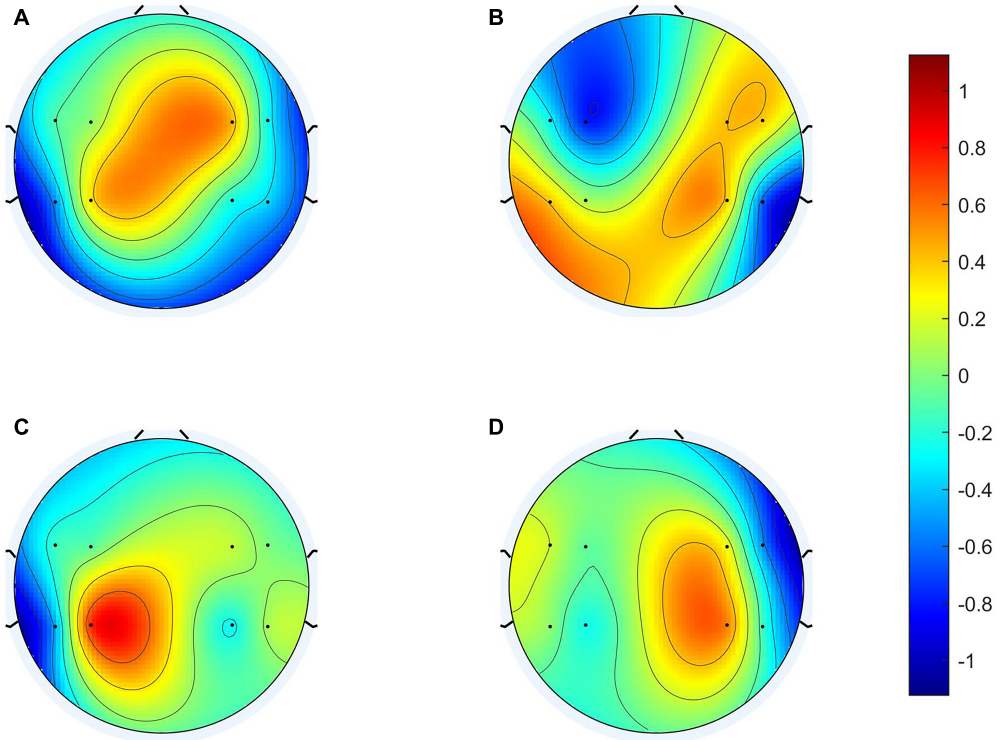


FIGURE 8 | 18th subject's spatial patterns. **(A,C)** Are spatial patterns of LH, **(B,D)** are spatial patterns of RH. The spatial patterns in the first row are generated from Conventional CSP, and the spatial patterns in the second row are generated from our algorithm.

RESULTS

Table 1 depicts the EEG classification performance of experiment I. The overall average classification accuracy for CSP, R-CSP-CV, and R-CSP-E are 70.2, 72.2, and 78.2%, respectively. As shown in **Table 1**, in all participants, the R-CSP-E algorithm outperforms the traditional CSP algorithm, with an average improvement of 7% in the correct classification rate with respect to the conventional CSP method. This result demonstrates the effectiveness of the CSP regularization scheme and illustrates the advantages of the proposed ensemble learning method over traditional cross-validation. In addition, as shown in **Figure 7**, in the R-CSP-E algorithm, different values of k will lead to different results. When $k = 15$, the average performance is the highest. Therefore, in subsequent experiments, we set $k = 15$.

The complete results for all subjects on the testing data set for experiment II are summarized in **Table 2**. In general, the classification rate of R-CSP-E showed superiority. Compared with the channel-wise feature extraction method and the conventional CSP algorithm, R-CSP-E can improve the classification accuracy up at least 6% on average. In order to clarify that the improvement of accuracy is derived from the effects of transfer learning or ensemble learning. The R-CSP-CV, which only includes effects of transfer learning, is tested for the fNIRS data. As depicted in **Table 2**, the result of R-CSP-E is better in the majority of subjects. This not only shows that the proposed R-CSP-E method is effective, but also concludes that the combination of the transfer learning and ensemble learning is crucial to improve accuracy. Besides, an independent t-test was performed to validate the outperformance of R-CSP-E over HbO, CSP and R-CSP-CV, separately.

As shown in **Tables 3–5**, for mean value of R-CSP-E algorithm compared with that of HbO, it shows a higher accuracy with $p = 0.00001 < 0.01$, for mean value of R-CSP-E algorithm against that of CSP, higher accuracy is observed with $p = 0.001 < 0.01$, and for mean value of R-CSP-E algorithm against that of R-CSP-CV, higher accuracy is observed with $p = 0.00001 < 0.01$.

To clarify the characteristics of the transfer algorithm, we visualized and compared the spatial patterns generated by the traditional CSP algorithm and our R-CSP-E method. **Figures 8**, **9** are the most convincing of them, extracted from the 18th and 8th subjects, respectively. In both figures, the first row is CSP, and the second row is our algorithm. Applying traditional CSP and our algorithm to the 8th subject, the accuracy rates are 68.3 and 71.7%, respectively, which is relatively high. Hence, the optimization of our proposed algorithm has almost no improvement. As for the 18th subject, the accuracy is 63.3 and 73.3%, respectively. The CSP algorithm has a low accuracy rate, while the transfer algorithm achieves a significant increase in the classification rate, which can be explained by the spatial patterns. In **Figure 7**, the result reveals that our algorithm has a higher discrimination weight on the CP3, FC4, and CP6 channels.

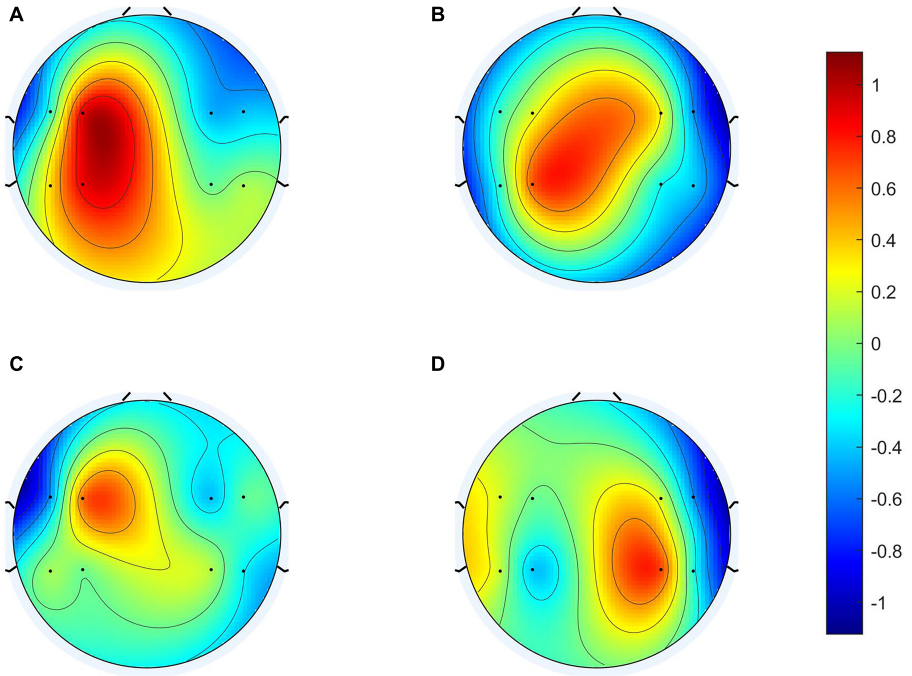


FIGURE 9 | 8th subject's spatial patterns. **(A,C)** Are spatial patterns of LH, **(B,D)** are spatial patterns of RH. The spatial patterns in the first row are generated from Conventional CSP, and the spatial patterns in the second row are generated from our algorithm.

The R-CSP-E filter is more discriminative than the CSP filter by introducing the EEG signal. More interestingly, it is expected from the neurophysiology literature. For example, for subject 18, the CSP filter appears to be rough and noisy, with greater weights at potentially different electrode positions, while the corresponding R-CSP-E filter shows greater weight in a few areas located on the motor cortex of the brain.

DISCUSSION

The main strength of our method is that the feature space of EEG signals is considered when computing the features of fNIRS signals. However, there are some limitations in this study that could be addressed in future research:

(1) At present, our transfer learning framework is based on the RCSP framework, which a simple but effective. In recent years, many more advanced feature extraction methods have been proposed successively, which can improve the classification performance by exploiting more spatiotemporal information (Lachert et al., 2017; Qi et al., 2020). Therefore, in future work, we can try to utilize these methods to further explore the application of transfer learning in cross-modal BCI.
(2) In our experiments, the regularization parameters are selected through the idea of ensemble learning. In the future, we will try more parameter selection methods.
(3) Currently, our cross-modal transfer learning is from EEG to fNIRS. However, in theory, transfer is mutual. In the future, we will further explore reverse transfer, the transfer of feature space from fNIRS to EEG.

CONCLUSION

In summary, we proposed an algorithm named R-CSP-E based on the RCSP framework and transfer learning theory. Instead of simply fusing them, we innovatively applied transfer learning to the EEG-fNIRS multimodal brain-computer interface. Our outcomes showed that the R-CSP-E algorithm significantly improved the classification accuracy compared with conventional CSP and channel-wise methods in fNIRS.

As mentioned, MI-BCI has a broad prospect in rehabilitation training and neural recovery. fNIRS has also shown its potential in the investigation of functional brain activation patterns and neurorehabilitation. For example, a study evaluated the changes in cerebral cortex activation in stroke patients 2 months before and after rehabilitation (Arun et al., 2020). This study proved the potential of fNIRS in detecting changes in brain activation related to exercise recovery, which is reflected in the hemispheres affected after rehabilitation that increased activation of the premotor cortex. In this way, fNIRS has shown great potential as a neurorehabilitation tool to monitor the patient's movement and cognitive improvement over time. In addition, fNIRS can also be used in the BCI system to treat movement disorders. Therefore, it is necessary to enhance the performance of fNIRS BCI, which

will promote the development of cognitive neuroscience and neural plasticity.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://doc.ml.tu-berlin.de/hBCI.

AUTHOR CONTRIBUTIONS

YW carried out experiment and writing. JL and HJ designed the overall framework. LL, JZ, and ZY contributed to methodological guidance and formal analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Shanghai Municipal Science and Technology Major Project (2021SHZDZX0100), the Fundamental Research Funds for the Central Universities, the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (No. TP2018056), and the Science and Technology Innovation Action Plan of Shanghai Science and Technology Commission (19441908000).

REFERENCES

Arun, K. M., Smitha, K. A., Sylaja, P. N., and Kesavadas, C. (2020). Identifying resting-state functional connectivity changes in the motor cortex using fnirs during recovery from stroke. *Brain Topogr.* 33, 710-719. doi: 10.1007/s10548-020-00785-2

Benjamin, B., Michael, T., Carmen, V., Siamac, F., Claudia, S., Stefan, H., et al. (2010). The berlin brain–computer interface: non-medical uses of BCI technology. *Front. Neurosci.* 4:198. doi: 10.3389/fnins.2010.00198

Bundy, D. T., Souders, L., Baranyai, K., Leonard, L., Schalk, G., Coker, R., et al. (2017). Contralesional brain–computer interface control of a powered exoskeleton for motor recovery in chronic stroke survivors. *Stroke* 48, 1908–1915. doi: 10.1161/STROKEAHA.116.016304

Coyle, S., Ward, T., Markham, C., and McDarby, G. (2004). On the suitability of near-infrared (NIR) systems for next-generation brain–computer interfaces. *Physiol. Meas.* 25, 815–815. doi: 10.1088/0967-3334/25/4/003

Daly, J. J., and Wolpaw, J. R. (2008). Brain-computer interfaces in neurological rehabilitation. *Lancet Neurol.* 7, 1032–1043. doi: 10.1016/S1474-4422(08)70223-0

Dominic, L., Sylvain, C., and Dominique, G. (2010). An introduction to independent component analysis: infomax and fastica algorithms. *Tutor. Quant. Methods Psychol.* 6, 31–38. doi: 10.20982/tqmp.06.1.p031

Fazli, S., Mehnert, J., Steinbrink, J., Curio, G., Villringer, A., Müller, K., et al. (2012). Enhanced performance by a hybrid NIRS–EEG brain computer interface. *NeuroImage* 59, 519–529. doi: 10.1016/j.neuroimage.2011.07.084

Ferrari, M., and Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage* 63, 921–935. doi: 10.1016/j.neuroimage.2012.03.049

Hyvarinen, A., and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Netw.* 13, 411–430. doi: 10.1016/s0893-6080(00)00026-5

Jayaram, V., Alamgir, M., Altun, Y., Schlkopf, B., and Grosse-Wentrup, M. (2015). Transfer learning in brain-computer interfaces. *IEEE Comput. Intell. Mag.* 11, 20–31. doi: 10.1109/mci.2015.2501545

Khazem, S., Chevallier, S., Barthélemy, Q., Haroun, K., and Nos, C. (2021). "Minimizing subject-dependent calibration for BCI with riemannian transfer learning," in *Procceedings of the 10th International IEEE/EMBS Conference on Neural Engineering (NER)* (Rome: *IEEE*).

Khosrowabadi, R., Chai, Q., Kai, K. A., Tung, S. W., and Heijnen, M. (2011). "A brain-computer interface for classifying EEG correlates of chronic mental stress," in *International Joint Conference on Neural Networks*) (San Jose, CA: IEEE), 757–762.

Kocsis, L., Herman, P., and Eke, A. (2006). The modified beer-lambert law revisited. *Phys. Med. Biol.* 51, N91-8. doi: 10.1088/0031-9155/51/5/N02

Lachert, P., Janusek, D., Pulawski, P., Liebert, A., Milej, D., and Blinowska, K. J. (2017). Coupling of oxy- and deoxyhemoglobin concentrations with EEG rhythms during motor task. *Sci. Rep.* 7:15414. doi: 10.1038/s41598-017-15770-2

Lee, S., Shin, Y., Kumar, A., Kim, M., and Lee, H. N. (2019). Dry electrode-based fully isolated EEG/fNIRS hybrid brain-monitoring system. *IEEE Trans. Biomed. Eng.* 66, 1055–1068. doi: 10.1109/TBME.2018.2866550

Liang, Y., and Ma, Y. (2020). Calibrating EEG features in motor imagery classification tasks with a small amount of current data using multisource fusion transfer learning. *Biomed. Signal Process. Control* 62:102101. doi: 10.1016/j.bspc.2020.102101

Lu, H., Plataniotis, K. N., and Venetsanopoulos, A. N. (2009a). "Boosting discriminant learners for gait recognition using MPCA features," in *EURASIP Journal on Image & Video Processing* 2009(video) (New York, NY: SpringerOpen), 1–11. doi: 10.1155/2009/713183

Lu, H., Plataniotis, K. N., and Venetsanopoulos, A. N. (2009b). "Regularized common spatial patterns with generic learning for EEG signal classification," in *Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Piscataway, NJ: IEEE), 6599–6602.

Oostenveld, R., and Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* 112, 713–719. doi: 10.1016/s1388-2457(00)00527-7

Pichiorri, F., Morone, G., Petti, M., Toppi, J., Pisotta, I., Molinari, M., et al. (2015). Brain–computer interface boosts motor imagery practice during stroke recovery. *Ann. Neurol.* 77, 851–865. doi: 10.1002/ana.24390

Qi, F., Wu, W., Yu, Z. L., Gu, Z., Wen, Z., Yu, T., et al. (2020). Spatiotemporal-filtering-based channel selection for single-trial EEG classification. *IEEE Trans. Cybern.* 51, 558–567. doi: 10.1109/TCYB.2019.2963709

Rebsamen, B., Burdet, E., Guan, C., Zhang, H., Teo, C. L., Zeng, Q., et al. (2007). Controlling a wheelchair indoors using thought. *IEEE Intell. Syst.* 22, 18–24. doi: 10.1109/IEMBS.2008.4650144

Royer, A. S., Doud, A. J., Rose, M. L., and He, B. (2010). EEG control of a virtual helicopter in 3-dimensional space using intelligent control strategies. *IEEE Trans. Neural Syst. Rehabil. Eng.* 18, 581–589. doi: 10.1109/TNSRE.2010.2077654

Shin, J., Lühmann, A., Blankertz, B., Kim, D. W., Jeong, J., Hwang, H. J., et al. (2017). Open access dataset for EEG+NIRS single-trial classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1735–1745. doi: 10.1109/tnsre.2016.2628057

Spuler, M., Walter, A., Rosenstiel, W., and Bogdan, M. (2014). Spatial filtering based on canonical correlation analysis for classification of evoked or event-related potentials in EEG Data. *IEEE Trans. Neural. Syst. Rehabil. Eng.* 22, 1097–1103. doi: 10.1109/TNSRE.2013.2290870

Trakoolwilaiwan, T., Behboodi, B., Lee, J., Kim, K., and Choi, J. W. (2017). Convolutional neural network for high-accuracy functional near-infrared spectroscopy in a brain–computer interface: three-class classification of rest, right-, and left-hand motor execution. *Neurophotonics* 5:011008. doi: 10.1117/1.NPh.5.1.011008

Wolpaw, J. R., Loeb, G. E., Allison, B. Z., Donchin, E., Do, O. N., Heetderks, W. J., et al. (2006). BCI meeting 2005–workshop on signals and recording methods. *IEEE Trans. Neural Syst. Rehabil. Eng.* 14:138. doi: 10.1109/TNSRE.2006.875583

Wu, D., King, J. T., Chuang, C. H., Lin, C. T., and Jung, T. P. (2017). Spatial filtering for EEG-based regression problems in brain–computer interface (BCI). *IEEE Trans. Fuzzy Syst.* 26, 771–781. doi: 10.1109/tfuzz.2017.2688423

Wu, D., Xu, Y., and Lu, B.-L. (2022). Transfer learning for EEG-based brain-computer interfaces: a review of progress made since 2016. *IEEE Trans. Cogn. Dev. Syst.* 14, 4–19. doi: 10.1109/TCDS.2020.3007453

Xu, Y., Wei, Q., Hua, Z., Hu, R., Liu, J., Jing, H., et al. (2019). Transfer learning based on regularized common spatial patterns using cosine similarities of spatial filters for motor-imagery BCI. *J. Circuits Syst. Comput.* 28:1950123. doi: 10.1142/s0218126619501238

Yue, K., and Wang, D. (2019). EEG-based 3D visual fatigue evaluation using CNN. *Electronics* 8:1208. doi: 10.3390/electronics8111208

Zhang, C., and Ma, Y. (2012). in *Ensemble Machine Learning Methods and Applications*, eds C. Zhang and Y. Ma (New York, NY : Springer), 1–34.

Zhang, S., Zheng, Y., Wang, D., Wang, L., Ma, J., Zhang, J., et al. (2017). Application of a common spatial pattern-based algorithm for an fNIRS-based motor imagery brain-computer interface. *Neurosci. Lett.* 655, 35–40. doi: 10.1016/j.neulet.2017.06.044

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang, Yang, Ji, Li, Liu and Zhuang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 13 April 2022
doi: 10.3389/fpsyg.2022.843531

Managing Students' Creativity in Music Education – The Mediating Role of Frustration Tolerance and Moderating Role of Emotion Regulation

Lei Wang[1†] *and Na Jiang*[2*†]

[1] School of Music, Liaoning Normal University, Dalian, China, [2] Changchun Humanities and Sciences College, Changchun, China

Artificial intelligence (AI) era challenges the use and functions of emotion in college students and the students' college life is often experienced as an emotional rollercoaster, negative and positive emotion can affect the emotional outcomes, but we know very little about how students can ride it most effectively to increase their creativity. We introduce frustration tolerance as a mediator and emotion regulation as a moderator to investigate the mechanism of creativity improvement under negative emotion. Drawing on a sample of 283 students from professional music colleges or music major in normal universities, we find that negative emotion are generally associated with a lower creativity, while frustration tolerance can mediate the relationship between negative emotion and creativity, but these effects depend on the emotion regulation. Cognitive reappraisal exerts a negative effect on the relationship between negative emotion and creativity, while expressive suppression has the opposite effect. Our study contributes to the literatures on student's emotions and creativity in music education and to the emotion regulation literature.

Keywords: creativity, emotion regulation, negative emotion, frustration tolerance, artificial intelligence

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Hongxia Sun,
Suzhou University of Science and Technology, China
Hong Yang,
Jilin University, China

***Correspondence:**
Na Jiang
jiangna0122@126.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *26 December 2021*
Accepted: *07 March 2022*
Published: *13 April 2022*

Citation:
Wang L and Jiang N (2022) Managing Students' Creativity in Music Education – The Mediating Role of Frustration Tolerance and Moderating Role of Emotion Regulation. Front. Psychol. 13:843531. doi: 10.3389/fpsyg.2022.843531

INTRODUCTION

In the new era of artificial intelligence (AI), big data and cloud computing, more and more attention has been deployed to emotion, because emotion is fundamental to human experiences influencing our daily activities including cognition, communication, learning, and decision making and it has been a very crucial factors influencing students' creativity in college to facilitate the multiple creation (Brodin, 2018), for example, music improvisation. In reality, college music education called for the training of students' creativity, so did the society and parents, because music needs more creativity in creation. When students have more creativity, they can create wonderful work, which can bring more for music students for their development and growth. With the reform of the education, the students' creativity has not been effectively improved. Faced with the pressure in life and study, there are a lot of negative emotions from student, which will inhibit students' creativity (Chen et al., 2019) in music education. As we all know, for centuries, psychologists have tried to understand and define emotions because of its importance in affecting creativity. The most

widely recognized basic emotions proposed by Ekman were happiness, sadness, fear, anger, surprise, and disgust. Humans express and recognize emotions in multiple ways in which the expression of emotions is one of the heavily studied areas. So it is necessary to study this topic in the view of emotion and find ways to regulate the influence on students' creativity in college life.

However, the current literatures elaborated the influence of students' creativity in the four aspects: *first*, the relationship between student and their parents can bring negative mood for their behavior, such as family background, economic situation, which have an awful influence on student's mood (Rossetto and Chapple, 2019; Cheong et al., 2021; Colson et al., 2021). *Second*, the relationship between student and teachers exerts special influence on the creativity in the view of imbalance, for example, their conflict can affect the mood and behaviors (Chang et al., 2016; Soh, 2017). *Third*, the internal control and coordination can influence student's creativity in learning process. *Fourth*, the boundary conditions need more explanation among the negative emotion and creativity and the current moderators focus on internal and external conditions, such as emotion experience, emotion labor and mental elasticity (Stephenson et al., 2018). *Fifth*, the black box between negative emotion and creativity is not open because the extant studies paid more attention to the negative effect of negative emotion on creativity, but seldom focus on how to change negative emotion into creativity and the path received little attention.

In order to answer the above concerns, we introduce frustration tolerance as an essential path between emotion and creativity. But the effects of frustration tolerance on students' creativity and the moderating role of emotion regulation have received little attention (Runco et al., 2017; Gál et al., 2021). Frustration tolerance is an individual's ability to tolerate adversity and unmet needs in order to maintain normal lives. In college, the level of individual frustration tolerance is affected by many subjective and objective factors, and is also related to the individual's past experience, especially in music education (Wu et al., 2014). If the degree of frustration exceeds the range of self-tolerance, it is bound to destroy the normal performance of self-function, and then hinder their life adaptation. Therefore, it is indeed necessary to study the relationship between frustration tolerance and creativity.

Based on cognitive appraisal theory and using data of 283 students from professional music colleges or music major in normal universities, we contribute to emotion research in creativity literature by showing the important impact of negative emotion on student's creativity. Our study provides insights into how students' emotion regulation enables themselves to bounce back from bad mood in music education. Students who tend to suppress their emotions will be more unable to improve creativity. In contrast, students who tend to reappraise situations may be able to perceive negative emotions accurately, thereby activating problem-focused coping mechanisms and make creative behaviors. Knowing students can use frustration tolerance to bounce back from negative emotion is not only vital for student's creativity, but also for future development of students by using emotion regulation.

LITERATURE AND HYPOTHESES

Taking an Emotive View on Creativity

There is a mount of research focusing on the determinant factors influencing students' creativity (Glover, 1980; Ashby et al., 1999). The creativity literature highlights that factors contributing to the improvement of students' creativity remain poorly understood, especially under the circumstances of negative emotion, and that there is still a significant gap in our understanding of how to improve students' creativity (Furnham et al., 2011; Zhao et al., 2019). We do not focus on traditional views which suggest that students' creativity is brought about by external factors, such as college's rules, student-teacher relationship or even social atmosphere that students have little or no control, or that creativity depends mainly on student's ability to create something innovatively. Instead, we draw on the recently developed emotive approach to student's creativity, which suggests that the student's creativity is not simply a matter of internal abilities (Kolyvas and Nikiforos, 2021), but that more subjective and psychological reasons contribute to creativity (Oliveira et al., 2021). By focusing on psychological attributes, the emotive view provides a powerful perspective on student's creativity, as it explains why some students with negative emotion have high creativity, while others with positive have low one (Isen et al., 1987). Therefore, we introduce cognitive appraisal theory, which believes that emotion is a reaction process in which individuals perceive environmental events as beneficial or harmful and students can react to the events to regulate their emotion and find psychological ways to deal with these matters.

Creativity

Creativity is the ability to combine or connect elements to form a new relationship (Rahimi and Shute, 2021). Individuals with creativity have psychological traits such as curiosity, adventure, challenge and imagination in their emotions. Creativity explains the psychological process of transcending existing experiences, breaking through habitual limitations and forming new ideas in new situations (Bereczki and Kárpáti, 2018). It is not restricted by conventional rules and can flexibly apply experience to solve problems. Students' creativity is a kind of ability with which the individual can function flexible, unique, progressive characteristics in the supportive environment through the process of thinking to produce divergent views, endow things unique, novel meaning, and gain the results not only make themselves, but also make others satisfied. For students majoring in music, their creativity is of great value to their own creation, especially improvisation (Falavarjani and Yeh, 2018). And individual creativity is diverse and variable, and should not be limited to the innovation idea itself, but should also include the generation of innovative ideas, content, implementation, promotion and development to ensure that innovative ideas can be implemented effectively in the music creation (Torrance and Myers, 1970; Zhou and George, 2001; Rossetto and Chapple, 2019). So we conceptualize the creativity as the ability of individuals to produce innovative and feasible ideas for solving problems. Torrance and Myers (1970) divided creativity into four

dimension, and they are originality, which means the degree of innovation in a method of solving a problem, constructive diversity, which means the number and diversity of methods proposed to solve problems, analytical power, which means the degree of details of the method of solving problem and feasibility, which refers to the degree to which the method proposed to solve the problem can be applied in real life.

Frustration Tolerance

In Zhang's Dictionary of Psychology, frustration tolerance is defined as an adaptive ability of individuals to avoid abnormal behavior and continue efforts when confronted with blows and setbacks, which causes individuals to withstand difficulties, setbacks and failures when encountering setbacks. It is an individual's ability to adapt to frustration and setbacks. Frustration tolerance reflects an individual's ability to endure possible failures and continue to participate in challenging work, which can affect an individual's willingness to take risks and his response to failure (Clifford, 1991). Frustration tolerance ensures that students are able to cope effectively with stress, adapt to daily challenges, recover from disappointment, adversity and trauma, and develop a sense of purpose. This plays an important role in solving problems and getting along with others for college students (Posner et al., 2005). Clifford (1984) proposed *the Theory of Constructive Failure*, which clarifies and reevaluates the functions of failures and setbacks. If the importance of success is blindly emphasized, it will deprive individuals of the opportunity to feel failure and frustration, which individuals cannot learn from Roth et al. (2021). Frustration tolerance in this study is defined as an individual's ability to endure failure and frustration in learning, and to persevere in the face of difficulties.

Negative Emotion

Emotional theory divided emotion into emotional tones, such as happiness, anger, sadness and emotional dynamics, such as emotional intensity, duration, and recovery time (Russell, 1980). Early studies focused on the tone of emotions and paid more attention to the nature of emotions (Chen et al., 2019), but recent studies tend to focus on the kinetic energy of emotions (Liu, 2022), such as the stimulation of emotions and the process of emotional calm. Yela et al. (2019) emphasized the importance of individuals' active construction and evaluation of situations. He believed that individuals undergo a cognitive evaluation process between the occurrence of emotional stimulation and emotional response, which is why the same event can produce different emotional responses for different individuals. Watson and Tellegen (1985) took the degree of individual subjective feeling of pleasure as the benchmark for differentiation and divided emotions into positive and negative ones. If the feeling is favorable to oneself, it will produce positive emotions, if the feeling is unfavorable to oneself, it will produce negative emotions. Negative emotion is a kind of external behavior of escape, resistance and other physiological unpleasant.

Emotion Regulation

Emotion regulation is defined as a kind of process that influences the emotional intensity over time and influences whether, when, and how we experience and express positive or negative emotions (Fredrickson, 1998; Gross, 1998, 2013; Gross and Thompson, 2007). Gross (1998) proposed a process model of emotion regulation, which has five sequential strategies, and they are situation selection, situation modification, attentional deployment, cognitive change (cognitive reappraisal) and response modulation (expressive suppression). Individuals may regulate their emotions by using different strategies (Gross and John, 2003). In our study, we follow Gross and John (2003) in defining, analyzing, and measuring emotion regulation as the tendency to habitually use a specific emotion regulation strategy.

We choose to focus on cognitive reappraisal and expressive suppression, two of the most-established and important emotion regulation strategies (Gross and John, 2003; Webb et al., 2012; Gross, 2013). Gross and John (2003) proposed the two emotion regulation strategies, and studied their different influences on emotion, cognition and social behavior. Lazarus and Alfert (1964) pointed out that cognitive reappraisal is a type of cognitive change, which involves explaining a potential emotional withdrawal situation in a non-emotional way, thus changing the impact of the current emotion. On the contrary, Gross (2013) pointed out that expressive suppression is a type of response modulation, and takes place later in the process to inhibit ongoing emotional expressions, which requires not only cognitive strategies, but also long-term use of these strategies to monitor and inhibit the expression of emotions.

Focusing on these two types allows us to contrast between an antecedent- and a response-focused emotion regulation strategy. We believe that cognitive reappraisal and expressive suppression are two crucial emotion regulation strategies that are particularly relevant in student's college learning (Roth et al., 2021) as they, respectively, relate to emotional experience and expression and can affect student's behaviors (Sun et al., 2022). We expect that how they deal with the student's negative emotions will be different to some extent.

Hypotheses Development

It is obvious that individual students with positive emotions can improve their positive effects in emotion, cognition and action (Watson, 2018). However, negative emotions restrict student's ability to think and act, and even show a state of self-protection, and inhibit the expression of creativity. Negative emotions such as depression and fear can lead students to have no confidence, disappointment, self-doubt, and make individual lose the courage to try, or even hinder creative ideas, which may inhibit creativity (Yeh et al., 2019). The higher the degree of positive emotion, the better individual creativity. Conversely, the higher the degree of negative emotion, the worse individual creativity. Huang (2019) discussed creativity from the perspective of neurophysiological responses and argued that negative emotions would reduce the stimulation of knowledge and interfere with information integration. Negative emotions tend to be conservative in the way individuals deal with problems and reduce creativity (Averill, 2000). Negative emotion can limit individual's thinking and reduce the ability of association and analysis, as well as lack of motivation, which makes it more difficult for individuals to exert their creativity when they have negative emotion. All of these

can stop music students to create new productions. Watson and Hubbard (1996) pointed out that when individuals are in negative emotions, they would try to isolate themselves from stressors, and even deny themselves. Based on this, the following hypotheses are proposed in this study:

Hypothesis 1. There is a negative relationship between the negative emotion and creativity.

The theory of constructive failure affirms the positive influence of setbacks and frustration. Tolerance is produced by the combination of an individual's attitude toward failure, external goals and intrinsic motivation (Pi et al., 2019; Shirish et al., 2021). Frustration tolerance is acquired through learning and cultivation, and it is a sign of good adaptation and mental health. In addition, Oliveira et al. (2021) also believed that there are obvious individual differences in frustration tolerance, even if the same person, when facing different frustration events, their physical and mental reaction and frustration tolerance are not the same.

When individuals face the impact of external events or its internal needs are not satisfied, they can tolerate and adjust the negative emotions brought by setbacks, solve the dilemma with a positive attitude and promote the improvement of creativity (Liu et al., 2020). Researchers believe that frustration tolerance should be a necessary ability for individuals, and it can be more perfect after acquired training and learning. In the face of negative emotions, frustration tolerance can help individuals to identify the problem, seek the root of the problem, promote individuals to be positive, and then enhance their creativity (Pi et al., 2019). In college education, it is normal for student to encounter setbacks and difficulties, and their negative emotions often exist. Only students with strong tolerance of setbacks and frustration will not be discouraged when they occasionally meet adversity (Shi et al., 2021). They can improve their own emotions and mentality by facing up to setbacks and enhance their creativity. Therefore, we proposed the following hypotheses:

Hypothesis 2. There is a positive relationship between frustration tolerance and creativity.
Hypothesis 3. Frustration tolerance mediate the relationship between negative emotion and creativity.

As shown above, cognitive appraisal theory believes that emotion is a reaction process in which individuals perceive environmental events as beneficial or harmful and this emotion can facilitate individuals to cope with the events by taking emotive measures (Lazarus, 1991). Emotion regulation is the ability to detect their own and other people's emotions, and then appropriately deal with emotions and adjust their own thinking and action (Snyder et al., 2019). Cognitive reappraisal is "a form of cognitive change that involves construing a potentially emotion-eliciting situation in a way that changes its emotional impact" (Gross and John, 2003: 349). As an antecedent-focused type of emotion regulation, it can help individual to deal with upcoming emotional events. Previous research shows that cognitive reappraisal is of help to student and assist them to reframe setbacks or obstacles as opportunities, because cognitive reappraisal is associated with better psychological health, optimistic views, and wellbeing (John and Gross, 2004; Gross, 2013), all of which may prevent students from suffering negative emotions and encourage them to recognize themselves and discover the source of negative emotions (Morina et al., 2018). As we all know, negative emotions can weaken the ability of cognitive improvement, thus receding the creativity.

In other words, individuals with high cognitive reappraisal can properly express their own emotional responses, and adopt flexible strategies to adjust their own emotions and deal with the emotional responses of others. Individuals can properly use the influence of emotions to strengthen their tolerance for setbacks. It's good for creative thinking, motivation and even distraction. Catanzaro (1990) pointed out that when individual is in an emotional state generated by stimulus, he can self-perceive the emotion, and try to adjust the emotion by behavioral and cognitive changes to adapt to the environment or events. Gross (2013) believed that individuals who often use cognitive reappraisal strategy can improve their emotional and interpersonal functions and enhance their creativity.

Expressive suppression is "a form of response modulation that involves inhibiting ongoing emotion-expressive behavior" (Gross and John, 2003: 349). It is a response-focused strategy, meaning that this strategy can change the public display but not the internal experience of the emotion. According to cognitive appraisal theory, Individuals using suppression frequently leads to high emotional exhaustion, which in turn strengthens the bad effect of negative emotions. Furthermore, Schroder (2021) found that expressive suppression decreases student's satisfaction on daily work, which in turn strengthens intentions to give up the current work, thus reducing the deeper thinking in production and students' giving up halfway (D'Mello and Duckworth, 2019; Pujol, 2019). In addition, students who suppress their emotions are also seen as desolate and lack of confidence for some matters, which can attenuate the creativity as a result of negative emotion (Shi et al., 2021). Although the individuals who often used "expressive suppression" strategy could effectively adjust their emotions, they could not reduce the influence of negative emotions, especially on music creativity.

"Emotional Alchemy" proposed by Cooper and Sawaf (1997) believes that individuals with high cognitive reappraisal can quickly deal with their negative emotions and focus on solving problems. Therefore, when individuals encounter negative emotions, they can adjust their emotions and transform the power originally used to fight against negative emotions into energy to stimulate creativity. In challenging situations, the individual expressive suppression would affect their curiosity and problem-solving ability, thus affecting creativity. This results in the following hypotheses:

Hypothesis 4. Cognitive reappraisal negatively moderates the relationship between negative emotion and creativity.

Hypothesis 5. Expressive suppression positively moderates the relationship between negative emotion and creativity.

Figure 1 depicts the conceptual model of the research.

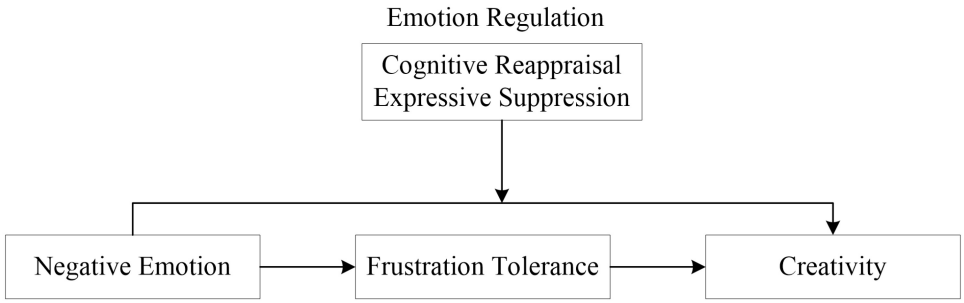


FIGURE 1 | The conceptual model.

MATERIALS AND METHODS

Sample and Data Collection

We used a sample of students from professional music colleges or music major in normal universities in Northeast of China. We contacted the 13 teachers in 7 colleges or universities and asked them to help distribute the survey questionnaire, through which we collected data on emotion regulation, negative emotions, frustration tolerance, and creativity and student demographics. Although emotion regulation, creativity and frustration tolerance may vary across an individual's university grade – differences are found in freshman, sophomore, junior and senior stages (Gross, 2013). But we believed it remain stable over this survey period.

In total, we contacted 1,360 students who agreed to participate in this survey. We delivered 1,360 questionnaires and collected 513 samples. After deleting the questionnaires with missing key items we obtained 283 complete, usable questionnaires, giving us a response rate of 20.8%. We evaluated non-response bias (283 vs. 230) by testing for differences in means for each variable ($t > 0.05$) and compared the former and latter questionnaires (177 vs. 106) and found that non-response bias ($t > 0.05$) doesn't exists.

Variable Measurement

Creativity

Recently, divergent thinking has become an important index to measure music student's creativity. TTCT (Torrance Tests of Creative Thinking) proposed by Torrance and Myers (1970) is widely used to measure creativity. TTCT divides creativity into language and graphics, and these division can give special explanation for music improvisation. In music education, language and graphics can teach students special knowledge and give them extra hints in their created work (Watson, 2018; Colson et al., 2021), so TTCT (Torrance Tests of Creative Thinking) is often used in testing students' creativity in education, especially in music, arts, and other humanities majors. Therefore, TTCT is very suitable and right for measure creativity for music education. TTCT uses four characteristics to measure creativity (see **Supplementary Appendix Table 1**). They are (1) originality (the degree of innovation in a method of solving a problem), (2) constructive diversity (the number and diversity of methods proposed to solve problems), (3) analytical power (the degree of details of the method of solving problem), and (4) feasibility (the degree to which the method proposed to solve the problem can be applied in real life). The Cronbach's alpha for this measure was 0.84.

Frustration Tolerance

We borrowed and revised 18 measurement items from School Failure Tolerance Scale (SFT) developed by Clifford (1984) to measure frustration tolerance. After factor analysis, we deleted three items with low factor loading (see **Supplementary Appendix Table 2**). The Cronbach's alpha for this measure was 0.91.

Negative Emotion

We measured negative emotion using Fisher's (1997) eight-item scale (see **Supplementary Appendix Table 3**). The Cronbach's alpha for this measure was 0.86.

Emotion Regulation

We measured cognitive reappraisal using Gross and John's (2003) six-item scale (see **Supplementary Appendix Table 4**). The Cronbach's alpha for this measure was 0.89. Expressive suppression was measured using Gross and John's (2003) four-item scale. The Cronbach's alpha was 0.80.

Control Variables

We chose student's age, gender as control variables because they are regarded as influencing factors for creativity (Lai et al., 2019; Kolyvas and Nikiforos, 2021) and emotion regulation (Liu, 2021). We also controlled for the Big Five factors (i.e., neuroticism, agreeableness, extraversion, conscientiousness, and openness to experience) and measured these using John et al.'s (1991) Big Five Inventory (BFI) (see **Supplementary Appendix Table 5**). The Cronbach reliability values were 0.92 for neuroticism, 0.94 for agreeableness, 0.93 for extraversion, 0.90 for conscientiousness, and 0.92 for openness to experience and the total Cronbach reliability values of Big Five is 0.96 after deleting the low factor loading items. All variables are valued by Likert-5 level.

TABLE 1 | Correlation matrix.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. age	1												
2. gender	0.00	1											
3. neuro	−0.09	0.05	1										
4. agree	0.11	0.12*	0.23**	1									
5. extra	−0.03	−0.03	0.25**	0.32**	1								
6. consc	0.09	0.12*	0.34**	0.19**	0.33**	1							
7. open	0.11	0.11	0.24**	0.29**	0.40**	0.28**	1						
8. creat	0.02	0.04	0.15**	0.21**	0.14*	0.23**	0.22**	1					
9. ft	0.01	0.11	0.29**	0.38**	0.37**	34**	0.68**	0.22**	1				
10. ne	−0.09	0.05	0.23*	0.21**	−0.31**	−0.23**	−0.25**	−0.14*	0.28**	1			
11. cr	0.06	−0.01	0.12*	0.20**	0.27**	0.20**	0.22**	−0.16**	0.21**	0.25**	1		
12. es	0.08	−0.02	0.25**	−0.26**	0.23**	0.28**	0.28**	−0.20**	−0.33**	0.20**	−0.19**	1	
13. grade	0.02	0.03	−0.03	0.00	0.03	−0.01	0.01	−0.01	−0.02	0.04	−0.018	0.05	1
14. mv	−0.07	−0.03	0.06	0.07	0.09	0.04	0.17*	0.11	0.24*	0.05	0.10	0.06	0.09
Mean	20.84	0.69	3.59	3.67	3.64	3.67	3.67	3.68	3.63	3.77	3.66	3.56	2.59
S.D	1.53	0.22	0.96	0.91	0.88	0.77	0.80	0.96	0.57	0.87	0.95	0.89	0.93

*N = 283; *p < 0.05, **p < 0.01, neuro stands for neuroticism, agree stands for agreeableness, extra stands for extraversion, consc stands for conscientiousness, open stands for openness to experience, creat stands for creativity, ft stands for frustration tolerance, ne stands for negative emotion, cr stands for cognitive reappraisal, es stands for expressive suppression. mv stands for method variable. The same below.*

TABLE 2 | The results of hierarchical linear regression (*n* = 283).

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
age	0.02 (0.06)	0.02 (0.06)	0.02 (0.06)	−0.01 (0.06)	−0.01 (0.06)	−0.01 (0.06)
gender	0.01 (0.06)	0.01 (0.06)	0.01 (0.06)	0.02 (0.6 = 06)	0.02 (0.06)	0.02 (0.06)
grade	−0.01 (0.06)	−0.01 (0.06)	−0.01 (0.06)	−0.01 (0.06)	−0.01 (0.06)	−0.02 (0.06)
neuro	−0.06 (0.09)	−0.05 (0.09)	−0.07 (0.09)	−0.07 (0.09)	−0.08 (0.09)	−0.07 (0.09)
agree	0.12 (0.46)	0.11 (0.45)	0.09 (0.46)	0.10 (0.45)	0.13 (0.46)	0.10 (0.44)
extra	0.08 (0.53)	0.07 (0.54)	0.10 (0.52)	0.08 (0.51)	0.10 (0.54)	0.11 (0.56)
consc	0.11 (0.74)	0.12 (0.78)	0.10 (0.72)	0.11 (0.74)	0.12 (0.76)	0.13 (0.74)
open	−0.11 (0.30)	−0.12 (0.34)	−0.13 (0.33)	−0.14 (0.37)	−0.11 (0.36)	−0.10 (0.35)
ne		−0.24*** (0.08)	−0.17* (0.08)	−0.19* (0.08)	−0.20* (0.09)	−0.19* (0.08)
ft			0.21*** (0.06)	0.24*** (0.06)	0.20*** (0.06)	0.22*** (0.06)
cr				0.17* (0.08)	0.20* (0.09)	0.21* (0.09)
es				−0.22* (0.10)	−0.18* (0.09)	−0.19* (0.09)
Ne × cr					−0.31*** (0.08)	
Ne × es						0.20* (0.09)
constant	4.03	4.40	4.39	5.07	5.20	4.67
R^2	0.21	0.23	0.25	0.26	0.28	0.30
Adjusted R^2	0.19	0.20	0.23	0.23	0.25	0.27
ΔR^2	–	0.02	0.02	0.01	0.02	0.02
F-value	27.12***	34.39***	35.13***	39.61***	40.77***	41.35***

****p ≤ 0.001, **p ≤ 0.01, *p ≤ 0.05. Two-tailed reported. The values in brackets are standard errors. The results are unstandardized.*

Common Method Bias

Common method bias may appear when both independent and dependent variables are derived from the same source at the same time (Podsakoff et al., 2003). This is the case in this study. Nevertheless, we tested for this in two ways. First, we used Harmon one-factor analysis without rotation and found that the first factor can only explain 37.71% of variance and there was no single factor, but five ones. Second, we used a latent method variable in our model to examine what influence common method variance might have on our results. The loadings of this latent construct on the variables in our model were relatively low (between 0.07 and 0.39), and the differences in estimates for our main variables between the models with and without the latent variable were small (average deviation = 0.03, maximum deviation = 0.11), and the correlation coefficients with other variables (see **Table 1**) range from −0.04 to 0.13, suggesting that common method bias was not a big concern.

ANALYSES AND RESULTS

We used a hierarchical linear regression model to test the hypotheses. **Table 1** presents the correlation matrix of the variables.

Table 2 presents the results of our hypotheses testing. Hypothesis 1 predicted that negative emotion would have a

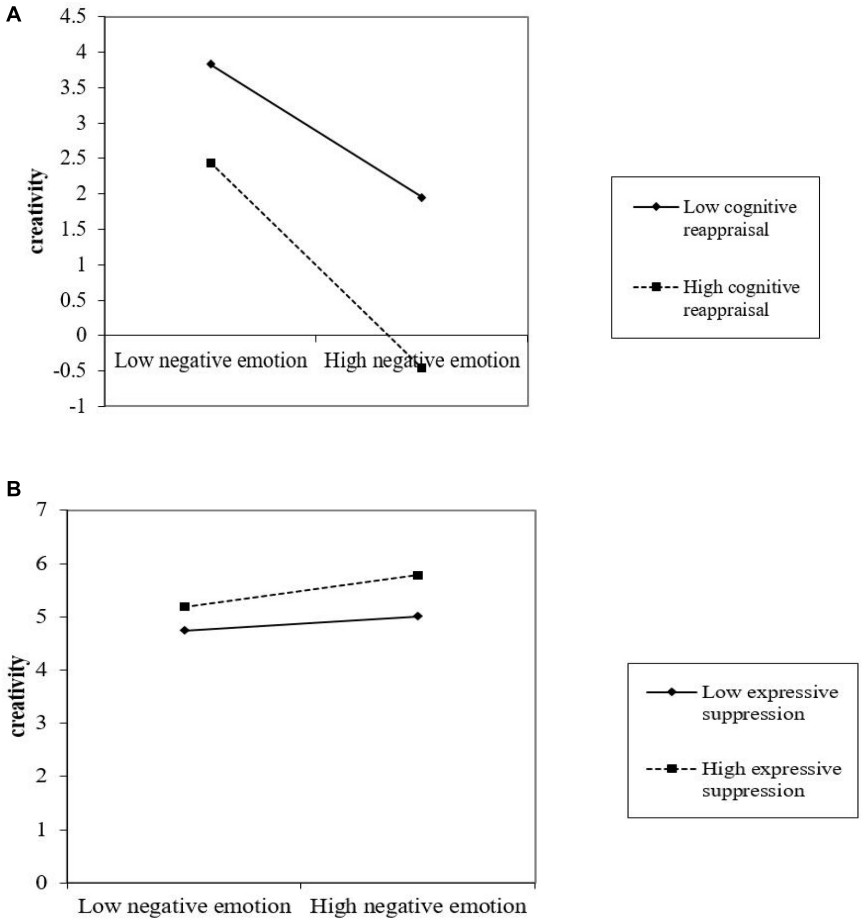


FIGURE 2 | (A) Negative emotion × cognitive reappraisal. **(B)** Negative emotion × expressive suppression.

negative effect on creativity. Model 2 in **Table 2** shows that the coefficient of negative emotion is significant ($\beta = -0.24$, $p < 0.001$), so hypothesis 1 is supported. We also found support for Hypothesis 2, which hypothesized a positive effect of students' frustration tolerance on their creativity ($\beta = 0.21$, $p < 0.001$). According to the research of Baron and Kenny (1986) for testing mediating effect, we found in model 1 and 2 that frustration tolerance can partially mediate the relationship between negative emotion and creativity, thus supporting Hypothesis 3. Models 5 includes the interaction effects between negative emotion and cognitive reappraisal, and the results show that cognitive reappraisal negatively moderates the relationship between negative emotion and creativity ($\beta = -0.31$, $p < 0.001$), as proposed by Hypothesis 4. Similarly, Models 6 includes the interaction effects between negative emotion and expressive suppression, and the results show that expressive suppression positively moderates the relationship between negative emotion and creativity ($\beta = 0.20$, $p < 0.05$), as proposed by Hypothesis 5. Therefore, Hypothesis 4 and 5 are all supported.

In order to gain a better understanding of their effects, we graphed these interaction effects in **Figure 2**. **Figure 2A** shows that students with high cognitive reappraisal had a negative impact on the relationship between negative emotion and creativity, that is to say, the higher the cognitive reappraisal, the weaker the negative impact of negative emotion on creativity. While **Figure 2B** shows that students with high expressive suppression had a positive impact on the relationship between negative emotion and creativity, that is to say, the higher the expressive suppression, the stronger the negative impact of negative emotion on creativity. **Figure 2** provides further evidence for Hypotheses 4 and 5.

DISCUSSION AND IMPLICATIONS

Discussion and Contributions

In line with the emotive view and cognitive appraisal theory, we integrated the concept of negative emotion, frustration tolerance, and creativity and emotion regulation into a research framework to provide an in-depth understanding of how emotions of student influence creativity in music education and can help student cope with the emotional swings in their college life. Specifically, we found that negative emotion

can decrease the creativity while frustration tolerance can facilitate the student's creativity. Besides, frustration tolerance can mediate the relationship between negative emotion and creativity. Considering the boundary conditions, students' cognitive reappraisal and expressive suppression, two strategies that represent antecedent- and response-focused ways of regulating emotions (Gross and John, 2003) have significant influence on creativity improvement.

Speaking of the negative effect of negative emotion on creativity, negative emotions can affect an individual's understanding and integration of information. In terms of mental model, negative emotions can also lead an individual to be more conservative, thus limiting the use of information and reducing the generation of individual creativity (Hershkovitz et al., 2019). In addition, negative emotions can reduce the stimulation of knowledge and make individuals tend to deal with problems in a conservative way, thus reducing the improvement of creativity.

Frustration tolerance, as we all know, is a kind of ability to defeat setbacks and obstacles. Individual with high frustration tolerance is not affected greatly by negative emotions of setbacks, can take the initiative to face setbacks, make good use of their own resources, find solutions and take effective actions. Most of the individuals with high frustration tolerance have strong willpower, believe that they have the ability to overcome difficulties, and will be braver when encountering failure and setbacks. For college students, frustration tolerance can solve the anxiety and confusion caused by negative emotions, so that they can fight back in adversity and improve their creativity.

Because cognitive reappraisal typically preempts extreme emotions (De Cock et al., 2019), and stimulates students to generate and discuss options for improving behaviors, we expected a positive moderating impact of cognitive reappraisal on the relationship between negative emotion and creativity, which can make it easy for students to use an antecedent-focused emotion regulation strategy effectively to anticipate what may happen in future and adjust their negative emotions by judging the emotional impact. It is simple for students to create a stable experience using cognitive reappraisal, which can keep them out of bad situations and perceive negative events accurately and activate problem-focused coping mechanisms. The negative effect of negative emotion on creativity is reversed principally by cognitive reappraisal.

For expressive suppression, we hypothesized and found a positive moderating effect on the relationship between negative emotion and creativity. As expressive suppression leads to emotional dissonance and lower satisfaction, which can cause students to give up the current work, lower the tendency for improvisation and quit discussion and communication. Students with negative emotions may suffer from expressive suppression, which may undermine student's chances of showing amiableness and talkativeness when facing frustration. Then students cannot improve creativity under the interaction of negative emotions and expressive suppression.

This study has several theoretical contributions. First, the main contribution is to the literature on creativity, where an amount of research argued that creativity is influenced by personality, Big Five factors, inspiration, training and learning, collaboration, information and communication technologies and psychology (Stolaki and Economides, 2018; Huang, 2019; Lai et al., 2019; Liu et al., 2020; Hsia et al., 2021; Rahimi and Shute, 2021; Roth et al., 2021), but little attention is paid to the impact of emotions (Falavarjani and Yeh, 2018). This study is one of the cutting-edge research to investigate how students in music education can deal with creativity to influence their performance. Studies of college student's creativity in emotive context are scarce, tend to overlook some important influential factors and outcomes (Dumas et al., 2021). Second, researchers found that negative emotion can cause unpleasant or bad outcomes, thus affecting the creative ideas or thinking (Guan et al., 2021). There is little research on how to change the adverse effects of negative emotions and the mechanism has not been effectively explored. We introduced frustration tolerance as a mediator to guide the positive transformation of negative emotions for college student, especially in music education. This study opens the black box between negative emotions and creativity and elaborate the path for students' creativity improvement under negative emotions. Last but not the least, our study contribute to the literature of emotion regulation. Previous studies focus on the direct influence of emotion regulation on negative emotions and creativity (in the perspective of psychology), its contingency effect lack enough attention. Our study focuses on the under-studied contingency effect of emotion regulation from two different views, one is antecedent-focused way (cognitive reappraisal) and the other is response-focused way (expressive suppression). By doing so, we address recent calls of researchers to investigate how students can influence their creativity to the benefit of their emotion regulation (e.g., Hershkovitz et al., 2019; Holdhus, 2019). Most remarkably, we also make contributions for cognitive appraisal theory in explaining how to cope with emotive events by find useful strategies to regulate negative emotions in music education.

Our study also has several practical implications. The first is for students, who should be aware that their emotions, and more importantly how they habitually regulate them, may have a significant impact on their daily behaviors and their creativity. Students should strengthen their frustration tolerance when facing setbacks or in bad mood, especially in music creation. In addition, considering the negative effect of negative emotion, colleges and teachers should pay close attention to the students' psychological state and encourage them to establish a correct outlook on life, be persistent when confronted with problems, and train them to be positive, innovative and creative. Most importantly, students should cultivate their tolerance to face difficulties and frustration, and creativity for production when facing negative mood. To some extent, cultivation of tolerance and creativity and control of emotion can be parallel.

Limitations and Future Research Directions

Our results suggest that antecedent- and response-focused emotion regulation has different impact on the relationship between negative emotion and creativity. But we know that cognitive reappraisal and expressive suppression are only two strategies out of five. Future research might investigate whether other types of emotion regulation strategies have similar effects

on the above relationship. For example, future studies might examine emotion regulation actions at the stages of situation selection, situation modification, and attentional deployment to determine whether these forms of antecedent-focused emotion regulation relate to creativity similarly as cognitive reappraisal and expressive suppression do.

Another limitation of our study is that all data are collected in a certain period, which incurred cross sectional problem. Because we use cross-sectional data to make the analyses, the direction of causality cannot be fully verified. Emotion regulation changes over time, so do frustration tolerance and students' creativity (Bardeen et al., 2017). Although we argue that frustration tolerance is conducive to students' creativity, it is possible to produce alternative results. For example, a student's high level of frustration tolerance might result in bad creativity because of the intensity of tolerance, which can cause more negative blowout of emotion and bring negative results. Future research should focus on the longitudinal view and study the above relationships with longitudinal data to investigate the possible law.

Finally, although the results provide support for our hypotheses, we acknowledge that our focus on data from Northeast of China, a relatively under-developed area in China, raises questions about the generalizability of our study beyond regions. Northeast of China is full of several special characteristics, including laggard economic development, solidified thought and weak creativity, which may exert influences on students' thinking and behavior. Future research should broaden the areas for data collection to evaluate the external validity of our model by testing it in different areas and compare the differences among different areas to get a general view.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

LW designed the model and collected the data. NJ analyzed the data. LW wrote the parts 1, 2, and 5, NJ finished the other parts. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported under the programs of 2019 Liaoning Humanities and Social Science Project "Research on all-round Collaborative Training Mechanism of Excellent Music Teachers" (WJ2019018), and 2019 Liaoning Normal University Doctoral funding project "Research on Professional Characteristics of Excellent Music Teachers based on MPCKS" (203070091922).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.843531/full#supplementary-material

REFERENCES

Ashby, F. G., Isen, A. M., and Turken, U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychol. Rev.* 106, 529–550. doi: 10.1037/0033-295X.106.3.529

Averill, J. R. (2000). "Intelligence, emotion, and creativity: from trichotomy to trinity," in *The Handbook of Emotional Intelligence*, eds B. O. Reuven and J. D. A. Parker (San Francisco, CA: Jossey-Bass Publishers), 277–298.

Bardeen, J. R., Fergus, T. A., and Orcutt, H. K. (2017). Examining the specific dimensions of distress tolerance that prospectively predict perceived stress. *Cogn. Behav. Ther.* 46, 211–223. doi: 10.1080/16506073.2016.1233454

Baron, R., and Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical consideration. *J. Pers. Soc. Psychol.* 51, 1173–1182. doi: 10.1037/0022-3514.51.6.1173

Bereczki, E. O., and Kárpáti, A. (2018). Teachers' beliefs about creativity and its nurture: a systematic review of the recent research literature. *Educ. Res. Rev.* 23, 25–56. doi: 10.1016/j.edurev.2017.10.003

Brodin, E. M. (2018). The stifling silence around scholarly creativity in doctoral education: experiences of students and supervisors in four disciplines. *High. Educ.* 75, 655–673. doi: 10.1007/s10734-017-0168-3

Catanzaro, S. J. (1990). Measuring generalized expectancies for negative mood regulation:initial scale development and implications. *J. Pers. Assess.* 54, 546–563. doi: 10.1080/00223891.1990.9674019

Chang, S., Wang, C., and Lee, J. (2016). Do award-winning experiences benefit students' creative self-efficacy and creativity? The moderated mediation effects of perceived school support for creativity. *Learn. Individ. Differ.* 51, 291–298. doi: 10.1016/j.lindif.2016.09.011

Chen, X., He, J., and Fan, X. (2019). Relationships between openness to experience, cognitive flexibility, self-esteem, and creativity among bilingual college students in the U.S. *Int. J. Biling. Educ. Biling.* 25, 342–354. doi: 10.1080/13670050.2019.1688247

Cheong, T. C., Alty, S. R., Yardim, A., and Howard, D. M. (2021). Creativity first, science follows: lessons in digital signal processing education. *IEEE Signal Process. Mag.* 38, 51–61. doi: 10.1109/MSP.2021.3058459

Clifford, M. M. (1984). Thoughts on a theory of constructive failure. *Educ. Psychol.* 19, 108–120. doi: 10.1080/00461528409529286

Clifford, M. M. (1991). Risk taking: theoretical, empirical, and educational considerations. *Educ. Psychol.* 26, 263–297. doi: 10.1016/j.addbeh.2021.107122

Colson, N., Shuker, M., and Maddock, L. (2021). Switching on the creativity gene: a co-creation assessment initiative in a large first year genetics course. *Assess. Eval. High. Educ.* doi: 10.1080/02602938.2021.2011133

Cooper, R. K., and Sawaf, A. (1997). *Executive EQ*. New York, NY: Penguin Putnam.

De Cock, R., Denoo, L., and Clarysse, B. (2019). Surviving the emotional rollercoaster called entrepreneurship: the role of emotion regulation. *J. Bus. Ventur.* 35:105936. doi: 10.1016/j.jbusvent.2019.04.004

D'Mello, S. K., and Duckworth, A. L. (2019). A brief behavioral measure of frustration tolerance predicts academic achievement immediately and two years later. *Emotion* 19, 1081–1092. doi: 10.1037/emo0000492

Dumas, D. G., Dong, Y., and Leveling, M. (2021). The zone of proximal creativity: what dynamic assessment of divergent thinking reveals about students' latent class membership. *Contemp. Educ. Psychol.* 67:102013. doi: 10.1016/j.cedpsych.2021.102013

Falavarjani, M. F., and Yeh, C. J. (2018). The impact of acculturation identification and acculturative stress on creativity among Iranian immigrants living in

Malaysia. *J. Ethn. Migr. Stud.* 44, 2219–2239. doi: 10.1080/1369183X.2017.1366301

Fisher, C. D. (1997). *Emotions at Work: What do People Feel and how should we Measure it? School of Business Discussion Papers.* Available Online at: http://epublications.bond.edu.au/discussion_papers/63 (accessed February 1, 1997).

Fredrickson, B. L. (1998). What good are positive emotions?. *Rev. Gen. Psychol.* 2, 300–319. doi: 10.1037/1089-2680.2.3.300

Furnham, A., Batey, M., Booth, T., Patel, V., and Lozinskaya, D. (2011). Individual difference predictors of creativity in art and science students. *Think. Skills Creat.* 6, 114–121. doi: 10.1016/j.tsc.2011.01.006

Gál, É., Tóth-Király, I., Szamosközi, I., and Orosz, G. (2021). Self-downing mediates the association between fixed intelligence mindset and different indicators of students' mental health. *Curr. Psychol.* doi: 10.1007/s12144-021-01457-5

Glover, J. A. (1980). *Becoming a more Creative Person*. Englewood Cliffs, NJ: Prentice Hall.

Gross, J. J. (1998). The emerging field of emotion regulation: an integrative review. *Rev. Gen. Psychol.* 2, 271–299. doi: 10.1037/1089-2680.2.3.271

Gross, J. J. (2013). Emotion regulation: taking stock and moving forward. *Emotion* 13, 359–365. doi: 10.1037/a0032135

Gross, J. J., and John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *J. Pers. Soc. Psychol.* 85, 348–362. doi: 10.1037/0022-3514.85.2.348

Gross, J. J., and Thompson, R. A. (2007). "Emotion regulation: conceptual foundations," in *Handbook of Emotion Regulation*, ed. J. J. Gross (New York, NY: Guilford Press), 3–24.

Guan, J., Wang, L., Chen, Q., Jin, K., and Hwang, G. (2021). Effects of a virtual reality-based pottery making approach on junior high school students' creativity and learning engagement. *Interact. Learn. Environ.* doi: 10.1080/10494820.2021.1871631

Hershkovitz, A., Sitman, R., Israel-Fishelson, R., Eguíluz, A., Garaizar, P., and Guenaga, M. (2019). Creativity in the acquisition of computational thinking. *Interact. Learn. Environ.* 27, 628–644. doi: 10.1080/10494820.2019.1610451

Holdhus, K. (2019). When students teach creativities: exploring student reports on creative teaching. *Qual. Inq.* 25, 690–699. doi: 10.1177/1077800418801377

Hsia, L., Lin, Y., and Hwang, G. (2021). A creative problem solving-based flipped learning strategy for promoting students' performing creativity, skills and tendencies of creative thinking and collaboration. *Br. J. Educ. Technol.* 52, 1771–1787. doi: 10.1111/bjet.13073

Huang, T. (2019). Do different learning styles make a difference when it comes to creativity? An empirical study. *Comput. Hum. Behav.* 100, 252–257. doi: 10.1016/j.chb.2018.10.003

Isen, A. M., Daubman, K. A., and Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *J. Pers. Soc. Psychol.* 52, 1122–1131. doi: 10.1037//0022-3514.52.6.1122

John, O. P., Donahue, E. M., and Kentle, R. L. (1991). *The Big Five Inventory: Versions 4a and 54, Institute of Personality and Social Research*. Berkeley, CA: University of California.

John, O. P., and Gross, J. J. (2004). Healthy and unhealthy emotion regulation: personality processes, individual differences, and life span development. *J. Pers.* 72, 1301–1334. doi: 10.1111/j.1467-6494.2004.00298.x

Kolyvas, S., and Nikiforos, S. (2021). Technology and creativity on early adolescence: a case study during COVID-19 pandemic. *Curr. Psychol.* [Online ahead of print]. doi: 10.1007/s12144-021-02349-4

Lai, C., Lai, Y., Hwang, R., and Huang, T. (2019). Physiological signals anticipatory computing for individual emotional state and creativity thinking. *Comput. Hum. Behav.* 101, 450–456. doi: 10.1016/j.chb.2018.05.015

Lazarus, R. S. (1991). *Emotion and Adaptation*. New York, NY: Oxford University Press.

Lazarus, R. S., and Alfert, E. (1964). Short-circuiting of threat by experimentally altering cognitive appraisal. *J. Abnorm. Soc. Psychol.* 69, 195–205. doi: 10.1037/h0044635

Liu, H. (2021). Moderating effects of task interdependence on interaction behaviours and creativity for nursing students on interdisciplinary teams. *J. Adv. Nurs.* 78, 131–141. doi: 10.1111/jan.14961

Liu, H. (2022). Promoting creativity of nursing students in different teaching and learning settings: a quasi-experimental study. *Nurse Educ. Today* 108:105216. doi: 10.1016/j.nedt.2021.105216

Liu, X., Zou, Y., Ma, Y., and Gao, W. (2020). What affects PhD student creativity in china? A case study from the joint training pilot project. *High. Educ.* 80, 37–56. doi: 10.1007/s10734-019-00463-8

Morina, N., Bryant, R. A., Doolan, E. L., Martin-Sölch, C., Plichta, M. M., Pfaltz, M. C., et al. (2018). The impact of enhancing perceived self-efficacy in torture survivors. *Depress. Anxiety* 35, 58–64. doi: 10.1002/da.22684

Oliveira, A. W., Brown, A. O., Zhang, W. S., LeBrun, P., Eaton, L., and Yemen, S. (2021). Fostering creativity in science learning: the potential of open-ended student drawing. *Teach. Teach. Educ.* 105:103416. doi: 10.1016/j.tate.2021.103416

Pi, Z., Yang, J., Hu, W., and Hong, J. (2019). The relation between openness and creativity is moderated by attention to peers' ideas in electronic brainstorming. *Interact. Learn. Environ.* 30, 344–352. doi: 10.1080/10494820.2019.1655458

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., and Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J. Appl. Psychol.* 88, 879–903. doi: 10.1037/0021-9010.88.5.879

Posner, J., Russell, J. A., and Peterson, B. S. (2005). The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* 17, 715–734. doi: 10.1017/S0954579405050340

Pujol, D. (2019). A reduction in pain intensity is more strongly associated with improved physical functioning in frustration tolerant individuals: a longitudinal moderation study in chronic pain patients. *Front. Psychol.* 10:907. doi: 10.3389/fpsyg.2019.00907

Rahimi, S., and Shute, V. J. (2021). First inspire, then instruct to improve students' creativity. *Comput. Educ.* 174:104312. doi: 10.1016/j.compedu.2021.104312

Rossetto, C., and Chapple, S. (2019). Creative accounting? The critical and creative voice of students. *Assess. Eval. High. Educ.* 44, 216–232. doi: 10.1080/02602938.2018.1492700

Roth, T., Conradty, C., and Bogner, F. X. (2021). Testing creativity and personality to explore creative potentials in the science classroom. *Res. Sci. Educ.* doi: 10.1007/s11165-021-10005-x

Runco, M., Acar, S., and Cayirdag, N. (2017). A closer look at the creativity gap and why students are less creative at school than outside of school. *Think. Skills Creat.* 24, 242–249. doi: 10.1016/j.tsc.2017.04.003

Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714

Schroder, H. S. (2021). Mindsets in the clinic: applying mindset theory to clinical psychology. *Clin. Psychol. Rev.* 83:101957. doi: 10.1016/j.cpr.2020.101957

Shi, S., Zhang, Z., Wang, Y., Yue, H., Wang, Z., and Qian, S. (2021). The relationship between college teachers' frustration tolerance and academic performance. *Front. Psychol.* 12:564484. doi: 10.3389/fpsyg.2021.564484

Shirish, A., Chandra, S., and Srivastava, S. C. (2021). Switching to online learning during COVID-19: theorizing the role of IT mindfulness and techno eustress for facilitating productivity and creativity in student learning. *Int. J. Inf. Manag.* 61:102394. doi: 10.1016/j.ijinfomgt.2021.102394

Snyder, H. T., Hammond, J. A., Grohman, M. G., and Katz-Buonincontro, J. (2019). Creativity measurement in undergraduate students from 1984-2013: a systematic review. *Psychol. Aesthet. Creat. Arts* 13, 133–143. doi: 10.1037/aca0000228

Soh, K. (2017). Fostering student creativity through teacher behaviors. *Think. Skills Creat.* 23, 58–66. doi: 10.1016/j.tsc.2016.11.002

Stephenson, E., Watson, P. J., Chen, Z. J., and Morris, R. J. (2018). Self-compassion, self-esteem, and irrational beliefs. *Curr. Psychol.* 37, 809–815. doi: 10.1007/s12144-017-9563-2

Stolaki, A., and Economides, A. A. (2018). The creativity challenge game: an educational intervention for creativity enhancement with the integration of information and communication technologies (ICTs). *Comput. Educ.* 123, 195–211. doi: 10.1016/j.compedu.2018.05.009

Sun, M., Wang, M., Wegerif, R., and Peng, J. (2022). How do students generate ideas together in scientific creativity tasks through computer-based mind mapping? *Comput. Educ.* 176:104359. doi: 10.1016/j.compedu.2021.104359

Torrance, E. P., and Myers, R. (1970). *Creative Learning and Teaching*. New York, NY: Dodd, Mead, and Company: Inc.

Watson, D., and Hubbard, B. (1996). Adaptational style and dispositional structure: coping in the context of the five-factor model. *J. Pers.* 64, 737–774. doi: 10.1111/j.1467-6494.1996.tb00943.x

Watson, D., and Tellegen, A. (1985). Toward a consensual structure of mood. *Psychol. Bull.* 98, 219–235. doi: 10.1037/0033-2909.98.2.219

Watson, J. (2018). Deferred creativity: exploring the impact of an undergraduate learning experience on professional practice. *Teach. Teach. Educ.* 71, 206–213. doi: 10.1016/j.tate.2017.12.018

Webb, T. L., Miles, E., and Sheeran, P. (2012). Dealing with feeling: a meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychol. Bull.* 138, 775–808. doi: 10.1037/a0027600

Wu, H., Wu, H., Chen, I., and Chen, H. (2014). Exploring the critical influential factors of creativity for college students: a multiple criteria decision-making approach. *Think. Skills Creat.* 11, 1–21. doi: 10.1016/j.tsc.2013.09.004

Yeh, Y., Rega, E. M., and Chen, S. (2019). Enhancing creativity through aesthetics-integrated computer-based training: the effectiveness of a FACE approach and exploration of moderators. *Comput. Educ.* 139, 48–64. doi: 10.1016/j.compedu.2019.05.007

Yela, A. A., Castano, S. R., and Garcia, P. C. (2019). Mindfulness' effects on undergraduates' perception of self-knowledge and stress levels. *J. Bus. Res.* 101, 441–446. doi: 10.1016/j.jbusres.2019.01.026

Zhao, D., Wu, J., and Gu, J. (2019). Can high leader–member exchange spark low creativity among graduate students? The role of stress and personal initiative. *Curr. Psychol.* 40, 4488–4499. doi: 10.1007/s12144-019-00389-5

Zhou, J., and George, J. M. (2001). When job dissatisfaction leads to creativity: encouraging the expression of voice. *Acad. Manag. J.* 44, 682–696. doi: 10.2307/3069410

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang and Jiang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 28 June 2022
doi: 10.3389/fpsyg.2022.864047

Using Facial Micro-Expressions in Combination With EEG and Physiological Signals for Emotion Recognition

Nastaran Saffaryazdi[1], Syed Talal Wasim[1], Kuldeep Dileep[1], Alireza Farrokhi Nia[1], Suranga Nanayakkara[2], Elizabeth Broadbent[3] and Mark Billinghurst[1]*

[1] *Empathic Computing Laboratory, Auckland Bioengineering Institute, The University of Auckland, Auckland, New Zealand,*
[2] *Augmented Human Laboratory, Auckland Bioengineering Institute, The University of Auckland, Auckland, New Zealand,*
[3] *Department of Psychological Medicine, The University of Auckland, Auckland, New Zealand*

Emotions are multimodal processes that play a crucial role in our everyday lives. Recognizing emotions is becoming more critical in a wide range of application domains such as healthcare, education, human-computer interaction, Virtual Reality, intelligent agents, entertainment, and more. Facial macro-expressions or intense facial expressions are the most common modalities in recognizing emotional states. However, since facial expressions can be voluntarily controlled, they may not accurately represent emotional states. Earlier studies have shown that facial micro-expressions are more reliable than facial macro-expressions for revealing emotions. They are subtle, involuntary movements responding to external stimuli that cannot be controlled. This paper proposes using facial micro-expressions combined with brain and physiological signals to more reliably detect underlying emotions. We describe our models for measuring arousal and valence levels from a combination of facial micro-expressions, Electroencephalography (EEG) signals, galvanic skin responses (GSR), and Photoplethysmography (PPG) signals. We then evaluate our model using the DEAP dataset and our own dataset based on a subject-independent approach. Lastly, we discuss our results, the limitations of our work, and how these limitations could be overcome. We also discuss future directions for using facial micro-expressions and physiological signals in emotion recognition.

Keywords: emotion recognition, electroencephalography (EEG), facial micro-expressions, physiological signals, neural networks, decision fusion, OpenBCI

OPEN ACCESS

Edited by:
Maozhen Li,
Brunel University London,
United Kingdom

Reviewed by:
Zhaoqiang Xia,
Northwestern Polytechnical University,
China
Chang Li,
Hefei University of Technology, China

****Correspondence:***
Nastaran Saffaryazdi
zsaf419@aucklanduni.ac.nz

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *28 January 2022*
Accepted: *30 May 2022*
Published: *28 June 2022*

Citation:
Saffaryazdi N, Wasim ST, Dileep K, Nia AF, Nanayakkara S, Broadbent E and Billinghurst M (2022) Using Facial Micro-Expressions in Combination With EEG and Physiological Signals for Emotion Recognition. Front. Psychol. 13:864047. doi: 10.3389/fpsyg.2022.864047

1. INTRODUCTION

Human emotions involve numerous external and internal activities and play an essential role in our daily life. Facial expressions, speech, and body gestures are some of the external activities affected by emotional situations. Changes in brain activity, heart rate, blood pressure, respiration rate, body temperature, and skin conductance are examples of internal emotional effects (Verma and Tiwary, 2014). Nowadays, we are surrounded by digital characters, intelligent devices, and computers in the modern world. There is a need for better interaction with these systems, and it is becoming increasingly important to recognize emotions in many human-human and human-computer interactions (Zheng et al., 2018). The effectiveness of our remote interactions,

therapy, consultations, or training sessions could be improved if they were equipped with emotion recognition systems. For example, recognizing emotion in remote e-learning (Khalfallah and Slama, 2015) could enhance the performance of learning. Similarly, in Empathic Computing applications, the goal is to measure the emotions of people teleconferencing together and use the result to improve remote communications (Piumsomboon et al., 2017).

Finally, creating intelligent agents with emotion recognition capabilities could be helpful in health care, education, entertainment, crime investigation, and other domains (Huang et al., 2016). It could be beneficial for intelligent assistants (Marcos-Pablos et al., 2016) or humanoid robots (Bartlett et al., 2003) to be able to measure the emotions of their users. Zepf et al. (2020) discuss the importance of emotion-aware systems in cars. Similarly, Hu et al. (2021) presented a conversational agent that recognizes emotions based on the acoustic features of speech. According to Chin et al. (2020) empathy between conversational agents and people can improve aggressive behavior. Schachner et al. (2020) discussed developing intelligent conversational agents for health care, especially for chronic diseases. Similarly, Aranha et al. (2019) reviewed software with smart user interfaces capable of recognizing emotions in various fields, including health, education, security, and art. According to their review, emotion recognition has often been used for adjusting sounds, user interfaces, graphics, and content based on user emotion.

Facial expressions are one of the most commonly used input modalities analyzed to identify emotional state (Sun et al., 2020). They are used in many HCI applications (Samadiani et al., 2019). Although studies have shown significant results in recognizing emotion from facial expressions (Li and Deng, 2020), using these methods in daily life faces some challenges because they can be controlled or faked by humans (Hossain and Gedeon, 2019). Many methods for recognizing emotions from facial expressions are based on datasets with non-spontaneous facial expressions or exaggerated facial expressions which do not correctly reflect genuine emotions (Weber et al., 2018; Li and Deng, 2020). In the real world, people usually show subtle involuntary expressions (Zeng et al., 2008) or expressions with lower intensity according to the type of stimuli. These studies show the importance of developing and improving robust methods for recognizing spontaneous emotions.

1.1. Recognizing Spontaneous Emotions

Three main approaches have been proposed in the literature for recognizing subtle, spontaneous emotions in the real world, which are listed as follows:

- Extracting involuntary expressions from faces.
- Using physiological signals that cannot be faked.
- Using a combination of various input modalities.

1.1.1. Extracting Facial Micro-Expressions From Faces

In this approach, the focus is on extracting facial micro-expressions instead of facial macro-expressions. Facial macro-expressions or intense facial expressions are voluntary muscle movements in the face that are distinguishable, cover a large area of the face, and their duration is between 0.5 and 4 s (Ekman and Rosenberg, 1997). In contrast, facial micro-expressions refer to brief and involuntary facial changes like the upturn of the inner eyebrows or wrinkling of the nose that happen spontaneously in response to external stimuli, typically over a short time frame of between 65 and 500 ms (Yan et al., 2013). Facial micro-expressions are difficult to fake and can be used to detect genuine emotions (Takalkar et al., 2018). The short duration of these expressions and their subtle movements make it difficult for humans to identify them (Qu et al., 2016); **Figure 1** shows some examples of facial micro-expressions compared to facial macro-expressions.

1.1.2. Using Physiological Signals That Cannot Be Faked

This approach relies on physiological responses that are difficult to fake and provide a better understanding of underlying emotions. These responses come from the central (brain and spinal cord) and autonomic nervous systems (regulating body functions like heart rate) (Kreibig, 2010). Electroencephalography (EEG) is one of the methods for measuring brain activity that is commonly used in emotional studies (Alarcao and Fonseca, 2017). Galvanic Skin Response (GSR) and Heart Rate Variability (HRV) can also be used to reliably measure emotional state and have been used widely in emotion recognition studies (Perez-Rosero et al., 2017; Setyohadi et al., 2018; Shu et al., 2018). Although EEG and physiological signals are more reliable and can not be controlled or faked by humans (Wioleta, 2013). These signals can be very weak and easily contaminated by noise (Jiang et al., 2020). So, recognizing emotions using only physiological signals can be pretty challenging.

1.1.3. Using a Combination of Various Input Modalities

In this approach, various modalities are combined to overcome the weaknesses of each individual modality. Combining different physiological signals for emotion recognition (Yazdani et al., 2012; Shu et al., 2018) or fusing only behavioral modalities have been widely explored (Busso et al., 2008; McKeown et al., 2011). Recently some studies tried to improve emotion recognition methods by exploiting both physiological and behavioral techniques (Zheng et al., 2018; Huang et al., 2019; Zhu et al., 2020). Many studies used a combination of facial expressions and EEG signals to achieve this improvement (Koelstra and Patras, 2013; Huang et al., 2017; Zhu et al., 2020). Usually, these researchers work on data that has been collected from subjects while they are watching videos or looking at still images (Koelstra et al., 2011; Soleymani et al., 2011). However, people often do not show many facial expressions in these tasks. Therefore, regular facial expression strategies may not be able to accurately recognize emotions. A limited number of studies used facial micro-expressions instead of facial macro-expressions (Huang et al., 2016), but this area still needs more research and exploration.

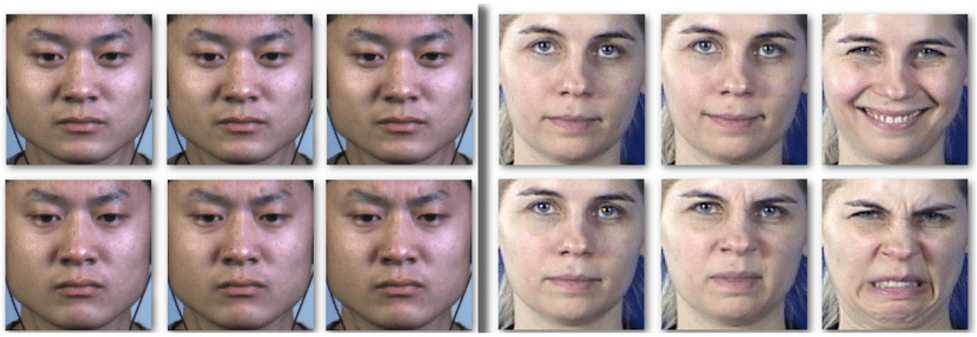


FIGURE 1 | Facial micro-expressions compared to facial macro-expressions. Facial micro-expressions **(left)** and macro expressions **(right)** for happiness (line1) and disgust (line2), from CASME II (Pantic et al., 2005) and MMI (Yan et al., 2014) datasets (Allaert et al., 2018).

Moreover, based on the research of Doma and Pirouz (2020), it is not clear when genuine emotion starts. They hypothesized that participants might still be in their previous emotional state during the first seconds of watching video stimuli. While in the last seconds, they may be more immersed in the video and feel genuine emotion. This is because they better understand the video in the final seconds. They found that the last seconds of EEG data were more informative and showed better emotion prediction results. We believe that the peak time of feeling emotions with the most intensity is affected by many factors such as the stimuli flow, participant personality, or previous experiences.

1.2. Goals, Overview, and Contributions

Our hypothesis is that by identifying and analyzing the most emotional part of each stimulus or the time of emerging emotions, we can better understand the body's reaction to emotions and create more robust models for identifying emotions. A primary objective of our research is to improve emotion recognition by combining facial micro-expression strategies with EEG and physiological signals.

In this paper, firstly, each facial video is scanned for micro-expressions that roughly indicate the emergence of emotional stimulation. The micro-expression window is used to approximately determine the time of arising emotions. Then we analyze the EEG and physiological data around the emergence of micro-expressions in each trial in comparison to the analysis of the entire trial. Finally, we compare these two strategies and evaluate our methods based on a subject-independent approach. In the end, we present the results, limitations, and future works. We also use the DEAP dataset as a benchmark to evaluate our method. Additionally, we conduct a user study to collect facial video, EEG, PPG, and GSR data while watching a video task similar to the DEAP dataset but with different sensors.

The main contributions of this research are as follows:

- Fusing facial micro-expressions with EEG and physiological signals to recognize emotions.
- Utilizing facial micro-expressions to identify the emotional stimulation or more informative period of data to improve recognition accuracy.
- Creating a new multimodal dataset for emotion recognition using a low-cost and open-source EEG headset.

2. PRELIMINARIES

2.1. Emotion Models

Some researchers believe that a few universal emotions exist that apply to all ages and cultures (Maria et al., 2019). A deeper understanding of emotion modeling is necessary to avoid making mistakes in emotion recognition and design a reliable system. Researchers have represented Emotions in two ways. The first perspective is the well-known discrete emotion model introduced by Ekman and Friesen (1971) which categorized emotions into six basic types; happiness, sadness, surprise, anger, disgust, and fear. In contrast, the second perspective considers emotions as a combination of three psychological dimensions: arousal and valence and one of dominance or intensity. Earlier research has demonstrated that two dimensions of arousal and valence are sufficient to explain the underlying emotions, which are primarily driven by neurophysiological factors (Eerola and Vuoskoski, 2011). The most common dimensional model used in the literature is Russel's Circumplex Model (Posner et al., 2005), which only uses valence and arousal for representing emotions, where valence represents a range of negative to positive emotions. In contrast, arousal represents a passive to active emotion.

Based on Russel's Circumplex Model, it is incorrect to categorize emotional states into discrete emotions because the human emotional state is always a mixture of several emotions. So, when people report fear as their emotion, it may be a mixture of excitement, joy, and fear or a combination of negative feelings and fear. So, in positive and negative scary situations, the pattern of the brain and physiological signals are not the same, and categorizing them in a single class leads to incorrect recognition. Additionally, the perception of emotions varies widely based on experience, culture, age, and many other factors, which makes evaluation difficult (Maria et al., 2019). Lichtenstein et al. (2008) showed that the dimensional approach is more accurate for self-assessments. Similarly, Eerola and Vuoskoski (2011) found that the discrete emotion model is less reliable than the dimensional model in rating complex emotional stimuli.

They also observed a high correspondence between the discrete and dimensional models.

Facial macro-expressions and facial micro-expressions are usually expressed with discrete emotions, and previous studies used the discrete emotion model to evaluate their strategy. However, most research on neurophysiological emotion recognition and the benchmark dataset that we used, used the Circumplex Model to assess their methods. Since the focus of our study is on revealing underlying emotions and used three neurophysiological cues besides facial micro-expressions, we used the two-dimensional Circumplex Model to evaluate our methodology on the benchmark dataset and our dataset.

2.2. Emotion Stimulation Methods

There are different ways of inducing emotions. However, the effect of all emotion induction methods is not the same. Siedlecka and Denson (2019) have classified emotional stimuli into five strategies; (1) watching visual stimuli like images and videos, (2) listening to music, (3) recalling personal emotional memories, (4) accomplishing psychological procedures, and (5) imagining emotional scenes. They showed how different types of stimuli could affect various physiological variables differently. Based on their research, visual stimuli are the most effective induction methods used more frequently in the literature. Quigley et al. (2014) have added Words, body movements, physiological manipulators like caffeine, and Virtual Reality (VR). Roberts et al. (2007) also found that dyadic interactions can be considered as an emotion eliciting method.

2.3. Facial Micro Expressions

Facial micro-expressions are brief facial movements in response to emotional stimuli which reveal hidden emotions (Ekman, 2003). Micro-expressions have been used in lie detection, security systems, and clinical and psychological fields to reveal underlying emotions (Yan et al., 2013). Lesser movements and shorter duration times are the main characteristics of facial micro-expressions in comparison to macro-expressions (Liong et al., 2015). Yan et al. (2013) studied the duration of micro-expressions and showed that their duration varies between 65 and 500 ms. Since video episodes are dynamic, long-lasting emotional stimuli, they have been used in micro-expression studies and creating most of the micro-expressions datasets (Li et al., 2013; Yan et al., 2014). To prevent facial macro-expression contamination in micro-expression recording, in many studies, participants are asked to inhibit any facial movements and keep a poker face when watching video (Li et al., 2013; Yan et al., 2013, 2014). However, suppression is brutal to achieve in response to emotional video stimuli (Yan et al., 2013).

A micro-expression has three phases; the onset, apex, and offset phases. In response to emotional stimuli, rapid muscle movements happen in the onset phase, which is involuntary and shows genuine emotional leakage. Sometimes these responses last for a moment as the apex phase. Finally, the emotional reactions disappear in the offset phase, and the face returns to a relaxed state. Returning to a relaxed state may take longer for some people because of natural skin tension or may not happen because of merging with the subsequent emotional stimuli (Yan et al., 2013). The first frame of the onset phase indicates the onset frame in a recorded video, while the frame with the most expressive emotion is the apex frame. The offset frame is when the expression disappears (Goh et al., 2020).

Recognizing emotions using facial micro-expressions has two main steps. The first step is spotting or locating the frame or frames with micro-expressions in a video sequence. The second step is recognizing the micro-expression emotional state (Oh et al., 2018; Tran et al., 2020). Several works have used hand-crafted strategies like Local Binary Pattern with Three Orthogonal Planes (LBP-TOP) (Pfister et al., 2011) or Histogram of Oriented Gradients (HOG) (Davison et al., 2015) to extract features from frames for spotting and recognizing emotions. (Guermazi et al., 2021) proposed an LBP-based micro-expression recognition method to create a low-dimensional high correlated representation of the facial video and used a Random Forest classifier to classify micro-expressions.

Recently, deep learning techniques have been used to extract deep features and classify emotions using facial micro-expressions (Van Quang et al., 2019; Tran et al., 2020). Hashmi et al. (2021) proposed a lossless attention residual network (LARNet) for encoding the spatial and temporal features of the face in specific crucial locations and classifying facial micro-expressions. Although they achieved a promising recognition of emotions in real-time, their model was efficient only when the frame rate was more than 200 fps. Xia et al. (2019) proposed a recurrent convolutional neural network (RCN) to extract spatiotemporal deformation of facial micro-expressions. They used an appearance-based and a geometric-based method to transform facial sequence into a matrix and extract the geometric features of facial movements. They evaluated their strategy based on both leave-one-video-out (LOVO) and leave-one-subject-out (LOSO) approaches and achieved satisfactory results. Similarly, Xia et al. (2020) proposed an RCN network to recognize micro-expressions across multiple datasets. They also discussed the effect of input and model complexity on the performance of deep learning models. They showed that lower-resolution input data and shallower models are beneficial when running models on a combination of datasets.

Ben et al. (2021) reviewed available datasets of facial micro-expressions and discussed different feature extraction methods for recognizing facial micro-expressions. In this research, they introduced a new dataset of micro-expressions and discussed the future directions for micro-expressions research. Similarly, Pan et al. (2021) summarized and compared the available spotting and micro-expression strategies and discussed the limitations and challenges in this area. Detecting facial micro-expressions has received growing attention. Many datasets have been created, and spotting and recognition methods have developed significantly. However, recognizing facial micro-expressions still faces many challenges (Weber et al., 2018; Zhao and Li, 2019; Tran et al., 2020). Oh et al. (2018) discussed various challenges in the dataset, spotting, and recognition areas. They showed that handling facial macro movements, developing more robust spotting strategies, and ignoring irrelevant facial information like head movements and cross-dataset evaluations still needs more attention and research.

2.4. Electroencephalography (EEG) Signals

Recently, many neuropsychological studies have investigated the correlations between emotions and brain signals. Electroencephalography (EEG) is one of the neuro-imaging techniques that reads brain electrical activities through electrodes mounted on the scalp. EEG devices differ based on the type and number of electrodes, the position of electrodes (flexible or fixed position), connection type (wireless or wired), type of amplifier and filtering steps, the setup, and wearability (Teplan et al., 2002). EEG devices with higher data quality like g.tec[1] or Biosemi[2] or EGI[3] are usually expensive and bulky and require a time-consuming setup. Alternatively, there are some EEG devices with lower data quality, like the Emotiv Epoc[4] or MindWave[5]. These EEG devices are affordable and are wireless devices that require less setup time (Alarcao and Fonseca, 2017). OpenBCI[6] provides a lightweight and open-source (hardware and software) EEG headset, which is positioned in between these two product categories. It captures high-quality data while it is low-cost and easy to set up. Nowadays, because of the improved wearability and lower price of EEG devices, recognizing emotions using EEG signals has attracted many researchers (Alarcao and Fonseca, 2017).

EEG-based emotion recognition is an exciting and rapidly growing research area. However, due to the weak amplitude of EEG signals, it is challenging to recognize emotion using EEG (Islam et al., 2021). Some research has focused on extracting hand-crafted features and using shallow machine learning methods to classify emotions in different application areas like health-care (Aydın et al., 2018; Bazgir et al., 2018; Pandey and Seeja, 2019a; Huang et al., 2021). Several review studies have discussed the effect of various hand-crafted features like brain band powers as well as using various classifiers like Support Vector Machine (SVM) or Random Forest (RF) for recognizing emotions. For instance, Alarcao and Fonseca (2017) reviewed EEG emotion recognition studies. They discussed the most common data cleaning and feature extraction that have been used in the literature for emotion recognition. Based on their review, brain band powers, including alpha, beta, theta, gamma, and delta bands, are effective features for emotion recognition. Similarly, Wagh and Vasanth (2019) provided a detailed survey on various techniques involved in the analysis of human emotions based on brain-computer interface and machine learning algorithms.

Recently many researchers have used raw EEG signals and applied deep learning methods to extract deep features and recognize emotions (Keelawat et al., 2019; Aydın, 2020). Sharma et al. (2020) used an LSTM-based deep learning method to classify emotional states based on EEG signals. Topic and Russo (2021) used deep learning to extract the topographic and holographic representations of EEG signals and classify emotional states. EEG-based emotion recognition methods have been comprehensively reviewed by Islam et al. (2021). They discussed various feature extraction methods and shallow and deep learning methods for recognizing emotions.

Researchers have focused on more advanced network architectures to increase performance in recent years. Li et al. (2021) proposed a neural architecture search (NAS) framework based on reinforcement learning (RL). They trained a Recurrent Neural Network (RNN) controller with an RL to maximize the generated model performance on the validation set. They achieved a high average accuracy of around 98% for arousal and valence on the DEAP dataset in a subject-dependent approach. In another research (Li et al., 2022), they proposed a multi-task learning mechanism to do the learning step for arousal, valence, and dominance simultaneously. They also used a capsule network to find the relationship between channels. Finally, They used the attention mechanism to find the optimal weight of channels for extracting the most important information from data. They reached the average accuracy of 97.25, 97.41% for arousal and valence in the subject-dependent approach. Similarly, Deng et al. (2021) used the attention mechanism to assign weights to channels and then capsule network and LSTM to extract spatial and temporal features. They achieved the average accuracy of 97.17, 97.34% for arousal and valence levels subject-dependently.

2.5. Galvanic Skin Responses (GSR) Signals

Previous studies have shown a connection between the nervous system and sweat glands on human skin. Changes in the level of sweat secretion because of emotional arousal lead to changes in skin resistance (Tarnowski et al., 2018; Kołodziej et al., 2019), which is known as the Electrodermal Activity (EDA) or Galvanic Skin Responses (GSR).

When the skin receives the brain's exciting signals caused by emotional arousal, sweating in the human body changes, and GSR signals rise. Kreibig (2010) showed that although EDA signals show changes in emotional arousal, more research is needed to identify the type of emotion using EDA signals. Tarnowski et al. (2018) used GSR local minimum as an indicator for emotional epochs of EEG. They showed that GSR is a good indicator of emotional arousal. In many studies, the GSR signal's statistical features have been used as the features for emotion classification (Udovičić et al., 2017; Yang et al., 2018). Kołodziej et al. (2019) calculated some statistics of peaks (local maxima) and raw GSR signal to use as the feature of signals. They used different classifiers and showed that SVM works better than other classifiers for identifying emotional arousal using these statistical features.

Some studies have used the time series or an averaging signal as the feature vector. Setyohadi et al. (2018) collected the average signal in each second and applied feature scaling. They used this data to classify positive, neutral, and negative emotional states. They used different classifiers, and SVM with Radial Based Kernel showed the best accuracy. Kanjo et al. (2019) used GSR time series and deep learning analysis to understand the valence level during walking in the middle of the city. Ganapathy et al. (2021) showed that Multiscale Convolutional Neural Networks

[1]https://www.gtec.at/
[2]https://www.biosemi.com
[3]https://www.egi.com/
[4]https://www.emotiv.com/epoc/
[5]https://store.neurosky.com/pages/mindwave
[6]https://openbci.com

(MSCNN) are effective in extracting deep features of GSR signals and classifying emotions.

In many studies, GSR signals have been used independently for recognizing emotion. But, they are mainly used as a supplementary signal or combined with other physiological signals for recognizing emotion (Das et al., 2016; Udovičić et al., 2017; Wei et al., 2018; Yang et al., 2018; Maia and Furtado, 2019).

2.6. Photoplethysmography (PPG) Signals

Photoplethysmography (PPG) is a novel method for measuring Blood Volume Pulse (BVP) using infrared light (Elgendi, 2012). It has been shown that PPG can measure heart rate variability (HRV). HRV is a measure of temporal changes in the heart rate to reveal medical or mental states (Maria et al., 2019). Due to the advent of wearable devices like smartwatches that transmit PPG signals, studies that utilize PPG signals have received more attention. Kreibig (2010) have shown changes in HRV and HR in a different emotional state. Recently a limited number of studies used deep learning strategies to extract deep features of PPG signals. Lee et al. (2019) used a one-dimensional convolutional neural network (1D CNN) to extract deep features of PPG signals and classify emotional states. Similar to GSR signals, PPG data is usually used with other physiological signals to recognize the emotional state.

3. RELATED WORKS

3.1. Multimodal Datasets for Emotion Recognition

Multimodal emotion recognition has attracted the attention of many researchers. A limited number of multimodal datasets with facial video, EEG, and physiological signals for emotion recognition are available for download. The DEAP dataset (Koelstra et al., 2011) and MAHNOB-HCI dataset (Soleymani et al., 2011) are the most popular datasets in multimodal emotion recognition, which include all these modalities. Since EEG signals are sensitive to muscle artifacts (Jiang et al., 2019), these kinds of datasets used passive tasks like watching videos or listening to music to minimize the subject movements.

3.1.1. DEAP Dataset

The DEAP dataset contains EEG data, facial video, GSR, blood volume pressure (BVP), temperature, and respiration data of 32 participants. It used 40 music videos for stimulating emotions, while EEG data were collected using the Biosemi ActiveTwo EEG headset[7], which has 32 channels. Participants reported their arousal, valence, dominance, and liking level using the self-assessment manikins (SAM) questionnaire (Bradley and Lang, 1994). However, in this dataset, only 22 participants have video data, and for 4 of them, some trials have been missed. The illumination in the facial video is low, and some sensors on the face cover part of the facial expressions.

3.1.2. MAHNOB-HCI Dataset

In the MAHNOB-HCI dataset, eye movements, sound, EEG data, and respiration patterns have been collected for image and video content tagging. After watching video clips, the participants reported their emotional state using the valence-arousal model. Thirty participants were recruited to create this dataset. The Biosemi active II EEG headset[8] with 32 channels was used for collecting the EEG data.

3.2. Exploring the Relationship Between Modalities

Some studies focused on the relationship between behavioral responses and physiological changes in multimodal emotion recognition. For example, Benlamine et al. (2016) and Raheel et al. (2019) used EEG signals to recognize facial micro-expressions. Hassouneh et al. (2020) used single-modality strategies for recognizing emotion in physically disabled people or people with autism using EEG and facial data. Although they did not use multimodal strategies, they showed that emotion could be recognized successfully using each facial expression or EEG signal. They achieved an accuracy of 87.3% for EEG and 99.8% for facial micro-expression from their experimental dataset.

Sun et al. (2020) investigated a strong correlation in emotional valence between spontaneous facial expression and brain activities measured by EEG and near-infrared spectroscopy (fNIRS). However, Soleymani et al. (2015) argued that although EEG signals have some complementary information for facial expression-based emotion recognition, they cannot improve the accuracy of the facial expression system. However, later studies showed improvement by combining EEG and facial expressions. The following section describes these studies.

3.3. Fusing Behavioral and Physiological Modalities

In many studies, researchers have shown the impact of emotional stimuli on physiological changes like heart rate, body temperature, skin conductance, respiration pattern, etc. However, they could not identify which emotions had been aroused. Some studies showed that combining physiological emotion recognition and behavioral modalities improves recognition outcomes. Combining facial expressions with physiological modalities attracted the focus of some researchers in this area. Most of these studies focused on traditional facial expression methods and used all recorded video frames to recognize emotions. For example, Koelstra and Patras (2013) used a combination of EEG and facial expressions to generate affective tags for videos. They extracted the power spectral density of power bands and the lateralization for 14 left-right pairs and extracted 230 features of EEG data. They tried to recognize the activation of action units frame-by-frame and finally extracted three features from them for each video. They used feature-level and decision-level fusion strategies. Based on their results, fusion strategies improved tagging performance compared to a single modality. By fusing EEG and face data,

[7]https://www.biosemi.com/

[8]https://www.biosemi.com/

TABLE 1 | Comparison of recent related works that used the DEAP or MAHNOB-HCI datasets.

References	Modalities	Dataset	Classification method	Evaluation	Arousal	Valence
Zhu et al. (2020)	EEG, FE, Physiological	DEAP	DNN	Cross-subject	72.20	78.47
Huang et al. (2019)	EEG, FE	DEAP	EEG SVM FE: CNN	Dependent	74.00	80.00
Pandey and Seeja (2019b)	EEG	DEAP	DNN	Independent	61.25	62.50
Li et al. (2018)	EEG	DEAP	SVM	Independent	–	59.60
Lan et al. (2018)	EEG	DEAP	DA	Independent	–	48.93 for 3 levels
Kwon et al. (2018)	EEG,GSR	DEAP	CNN	Dependent	76.56	80.46
Rayatdoost and Soleymani (2018)	EEG	DEAP MAHNOB	NN	Independent	55.70 61.46	59.22 71.25
Huang et al. (2016)	EEG, FE	MAHNOB	SVM/KNN	Independent	63.22	66.28
Koelstra and Patras (2013)	EEG, FE	MAHNOB	GNB	Dependent	73.00	70.90

Three evaluation methods have been considered including subject-dependent (dependent), subject-independent (independent), and cross-subject (when there are some trials of all subjects in both the training and test sets). FE, facial expression; SVM, support vector machine; GNB, Gaussian Naive Bayes; KNN, K-nearest neighbor; DNN, deep neural network; CNN, convolutional neural network; DA, domain adaptation. The arousal and valence scores are the percentage accuracy of recognition.

arousal accuracy was improved to 70.9% from 64.7% for EEG and 63.8% for the face. This improvement was from 70.9% for EEG and 62.8% for face to 73% by fusion for valence values.

Huang et al. (2017) investigated fusing facial macro-expressions and EEG signals for emotion recognition at the decision level. They used a feed-forward network to classify basic emotions in the extracted face of each video frame. They used their experimental data in this study and achieved 82.8% accuracy in a subject-dependent strategy when fusing EEG and facial expressions. Later, they extended their work by improving facial expression recognition using a CNN model (Huang et al., 2019). They pre-trained a model using the FER2013 dataset (Goodfellow et al., 2013) and used wavelets for extracting power bands and SVM classification for the EEG data. They achieved 80% accuracy for valence and 74% for arousal on the DEAP dataset using a subject-dependent strategy in a multimodal approach. In a similar study, Zhu et al. (2020) used a weighted decision level fusion strategy for combining EEG, peripheral physiological signals, and facial expressions to recognize the arousal-valence state. They used a 3D convolutional neural network (CNN) to extract facial features and classify them, and they also used a 1D CNN to extract EEG features and classify them. They achieved higher accuracy when combining facial expressions with EEG and physiological signals. Chaparro et al. (2018) also presented a feature-level fusion strategy for combining EEG and facial features (using 70 landmark coordinates) to improve recognition results.

In most multimodal emotional datasets' recorded video, no expressions could be observed in many frames. These datasets use passive tasks like video watching to stimulate emotions, so emotional faces can be seen in only a small portion of frames. So, considering all frames in the data analysis or using a majority vote among frames without considering this issue cannot produce a good emotion recognition result. However, many micro-expression can be observed in response to these passive tasks. To the best of our knowledge, only Huang et al. (2016) considered the presence of neutral faces and subtle expressions. They extracted Spatio-temporal features of all frames based on Local Binary Patterns (LBP) strategies. They then trained a linear kernel SVM using these features to calculate expression percentage features and used this feature vector for emotion classification. They extracted all frequencies and frequency bands and then used the ANOVA test to select a subset of these features for EEG. For facial classification, they used the K-Nearest-Neighbor (KNN) classifier for EEG and Support Vector Machine (SVM). They showed that a decision-level fusion strategy works better than a single modality or feature fusion. They achieved an accuracy of 62.1 and 61.8% for valence and arousal, respectively.

Table 1 summarizes the most recent related works. As can be seen, a limited number of studies combined facial expressions with EEG data. Most previous works evaluate their methods subject dependently or cross-subject when there are some trials of all participants in the training and test sets. Although designing general models that identify emotions in unseen participants is extremely useful in our daily lives, only a few studies have used a subject-independent approach to design and evaluate their methods. The accuracy of subject-independent methods is low compared to subject-dependent and cross-subject evaluations and needs more research and exploration. Also, although some research has focused on combining facial expressions with physiological signals, most of them have been designed and trained based on intense facial expressions. While in most used multimodal datasets, people are not allowed to show intense expressions.

This research addresses this gap by investigating the best ways to use facial micro-expression strategies combined with EEG and physiological signals for multimodal emotion recognition. We also explore how facial micro-expressions can be used to identify the most emotional part of the facial video, EEG, and physiological data. Furthermore, we create a new dataset

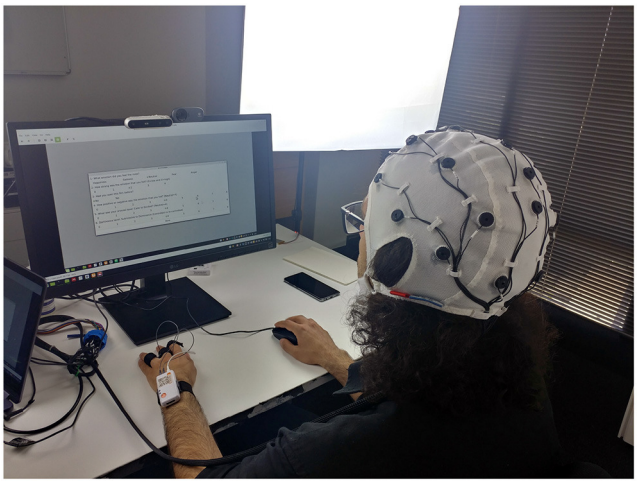


FIGURE 2 | The experiment setup.

of facial video, physiological signals, and EEG signals, which helps develop robust models for emotion recognition. We also explore the performance and quality of collected data from the OpenBCI EEG headset, a low-cost EEG headset for recognizing emotions. Moreover, we propose our strategy of using multimodal data for emotion recognition and finally evaluate it. Overall, the main novelty of this research is fusing facial micro-expressions recognition with the EEG and physiological signals. Another significant contribution of this work is using facial micro-expression to identify a neutral state vs. an emotional state to improve emotion recognition using the EEG and physiological signals.

4. EXPERIMENTAL SETUP

We created a new multimodal dataset for emotion recognition using lightweight wearable devices and a webcam. We recruited 23 volunteers (12 female and 11 male) aged between 21 and 44 years old ($\mu = 30, \sigma = 6$) from university students and staff. We only targeted four of the six basic emotions, including happiness, sadness, anger, and fear, plus a neutral state. We collected facial video, EEG, PPG, and GSR signals in a watching video task. We used the arousal-valence model for measuring emotions, and self-report data was also used as the ground truth.

4.1. Study Design

The data collection was performed in a room with a controlled temperature. We turned off the room lights, closed the door and curtains, and used two soft-boxes lightings facing the participant to control illumination. One Intel Realsense camera with a frame rate of 30 Hz was used to record facial expressions. Participants wore the OpenBCI EEG soft headset[9] with the cyton-daisy board to record EEG signals. A Shimmer3 sensor[10] was used to record PPG and GSR data. **Figure 2** shows the experiment setup. Participants wore the Shimmer sensor as a wristband, with PPG and GSR sensors attached to their three middle fingers. We used an Asus laptop (TP410U) to run the experiment scenario and record data. We designed the Octopus-Sensing library, a multi-platform, open-source python library[11], to create the scenario and simultaneously record data and send synchronization markers to the devices.

4.2. Stimuli Set

We considered happiness, sadness, anger, fear, and neutral emotions and used two video clips to stimulate each emotion. Ten video clips with the same length of 80 s were shown in a random order for emotion stimulation. We tried to choose videos with strong emotional scenes and subjects. Most of these videos have been used in previous emotion studies. **Table 2** shows the list of movies, their references, and their details.

4.3. Scenario

Each session started with introducing the devices, questionnaire, the purpose of the experiment, the meaning of arousal and valence levels, and the overall experiment process for the participant. Then, the EEG headset was placed on the participant's head, a shimmer3 wristband was worn on the participant's non-dominant hand, and the Shimmer's PPG and GSR sensors were attached to their three middle fingers. While watching the videos, participants were asked not to move their heads or bodies and put their hands with the Shimmer3 sensor on a table or on their legs.

The 10 videos were shown to the participants in random order. The experiment started with showing a gray screen for 5 s, then a fixation cross for 3 s, and then the video was displayed for 80 s. After each video, participants reported their emotional state by filling out a questionnaire similar to the SAM questionnaire (**Figure 3**) and then moving to the next video by pressing a button. Since the emotional effect of some video clips may have remained for a while, we asked participants to move to the next video clip after resting for a while and when they felt that they were in a neutral state. We left the participant alone in the room during the experiment to prevent any distraction or psychological effects of a stranger's presence.

5. METHODOLOGY

5.1. Ground Truth Labeling

We used self-report data from the SAM questionnaire for ground truth labeling. We only used the reported arousal and valences for the DEAP and our datasets. To classify arousal and valence levels, although there are nine levels for arousal and valence in the SAM questionnaire, similar to previous studies, we used binary classification. We considered five as the threshold for creating binary labels, corresponding to high and low arousal and valence values.

Table 3 shows the average of self-report ratings for arousal and values when rating values were between 1 and 9. This table also

[9]https://openbci.com/
[10]http://www.shimmersensing.com/
[11]https://octopus-sensing.nastaran-saffar.me

TABLE 2 | The video stimuli set for inducing emotion.

Emotion	Movie	Scene	References
Happiness	Pursuit of happiness	Offering job	–
Happiness	Benny and Joone	Benny (Johnny Depp) plays the fool in a coffee shop	Schaefer et al. (2010)
Sadness	The Champ	A kid cries at father's death	Gross and Levenson (1995)
Sadness	E.T.	Saying goodbye	Uhrig et al. (2016)
Fear	Silence of the Lambs	Darkness, chasing	Gross and Levenson (1995), Schaefer et al. (2010)
Fear	Chucky 2	Chucky beats Andy's teacher with a ruler	Schaefer et al. (2010)
Anger	My Bodyguard	Bullying scenes	Gross and Levenson (1995)
Anger	Cry freedom	Police abuse protesters	Gross and Levenson (1995)
Neutral	Weather news	News	Droit-Volet et al. (2011)
Neutral	Documentary	Documentary about soil	–

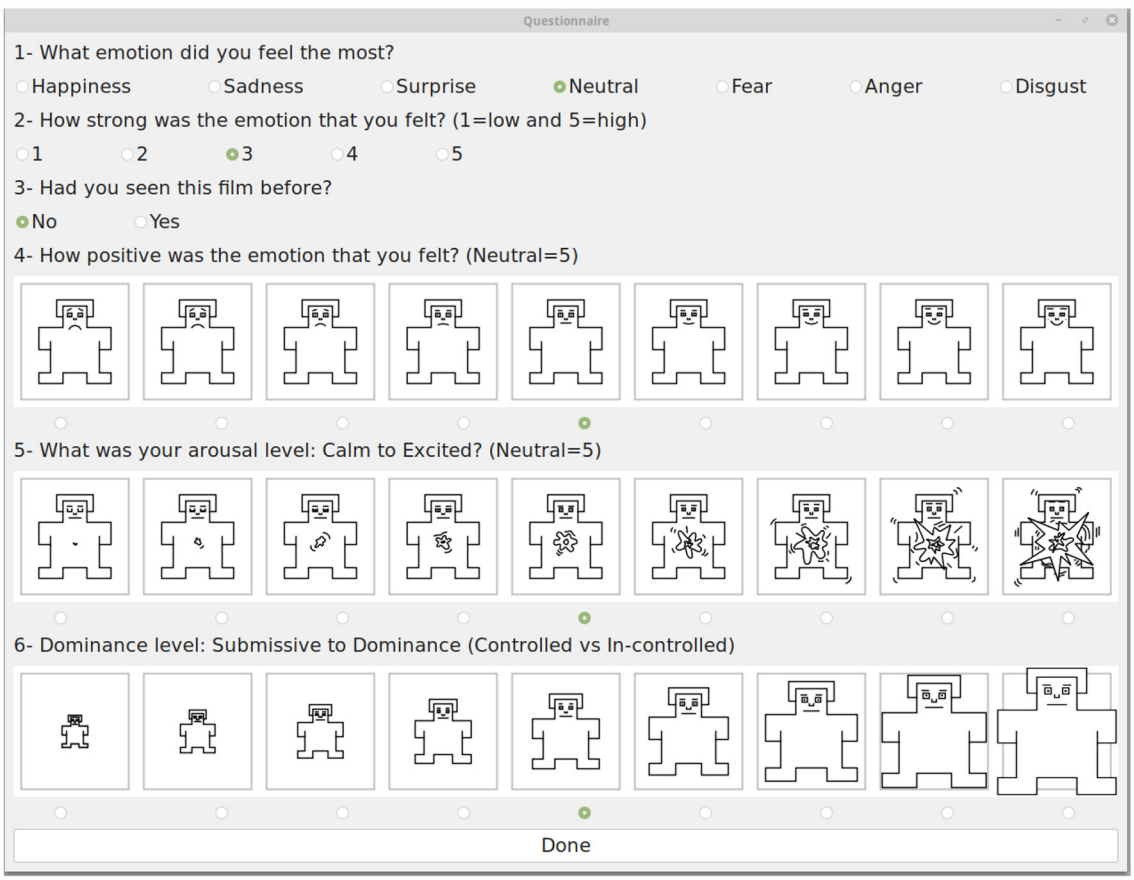


FIGURE 3 | Experimental questionnaire (Three last questions are from SAM questionnaire).

shows the percentage of participants who reported each emotion for each video clip. For example, 78.9% of participants reported happiness for the Pursuit of Happiness video clip, and only 4.3% reported fear, 7.8% reported neutral, and 8.7% reported sadness for this video clip. As can be seen, most of the participants reported the target emotion for all stimuli. Although we included all basic emotions in the self-report questionnaire, none of the participants reported other emotions except those in our target emotion list. So, we did not include other emotions in this table and in our evaluation results.

5.2. Imbalanced Data

In the DEAP dataset, the total number of low and high classes for all participants' trials for valence were 339 and 381, and for arousal, 279 and 444. These values for our dataset for valence classes were 100 and 130, and for arousal were 94 and 136 for

TABLE 3 | The Mean arousal and valence rating values and the percentage of participants who reported each emotion for each video-clip in our dataset.

	Target emotion	Valence	Arousal	Anger	Fear	Happiness	Neutral	Sadness
Pursuit of happiness	Happiness	6.83	5.48	0.00	4.30	**78.30**	8.70	8.70
Benny and Joone	Happiness	6.57	6.22	0.00	0.00	**87.00**	13.00	0.00
The champ	Sadness	3.35	5.17	4.30	0.00	0.00	0.00	**95.70**
E.T.	Sadness	5.61	4.57	0.00	4.30	26.10	13.00	**56.50**
Silence of the lambs	Fear	4.35	2.82	0.00	**69.60**	0.00	30.40	0.00
Chucky 2	Fear	3.83	7.22	0.00	**91.30**	8.70	0.00	0.00
My bodyguard	Anger	3.83	5.61	**73.90**	4.30	0.00	8.70	13.00
Cry freedom	Anger	3.26	6.35	**60.90**	0.00	0.00	0.00	39.10
News	Neutral	4.74	4.35	8.70	0.00	4.30	**78.30**	87.00
Documentary	Neutral	5.65	4.52	0.00	0.00	30.40	**69.60**	0.00

The bolded values prove that for each stimulus, most people reported the target emotion of stimuli.

low and high classes, respectively. As can be seen, both datasets were not balanced among classes. Also, we used a leave-some-subject-out strategy for splitting the training and test data. Hence, the imbalance state among the training and test sets for each set depended on the participants' rating. We used cost-sensitive learning (Ling and Sheng, 2008) to handle the imbalanced data. Cost-sensitive learning used the costs of prediction errors during the model training. It employed a penalized learning algorithm, which raised the cost of classification errors in the minority class. We used the Scikit-learn library[12] to measure class weights and used the estimated weights while training the models. We also used the cost-sensitive SVM and RF to handle imbalanced data.

5.3. Video Emotion Recognition

In the DEAP dataset and our dataset, we asked participants to keep a poker face while watching videos because of the sensitivity of EEG signals to muscle artifacts. This condition is entirely the same as micro-expression datasets. In micro-expression datasets, participants were asked to inhibit their expressions and keep a poker face while watching the videos to prevent macro-expression contamination (Goh et al., 2020). This condition leads to neutral faces in almost all frames, and only genuine emotions will leak as micro-expressions. **Figure 4** shows some frames of a trial from the DEAP dataset, our dataset, the SMIC dataset (Li et al., 2013) and some images from FER2013 dataset (Goodfellow et al., 2013). The SMIC dataset has been specifically collected for facial micro-expression emotion recognition studies. As seen in all of these datasets, emotions can hardly be noticed, and we mostly saw a neutral face. In contrast, in the facial macro-expressions datasets like FER2013 (Goodfellow et al., 2013) and CK+ (Lucey et al., 2010), there are sets of faces with intense expressions (**Figure 4**).

We trained a deep convolutional neural network using the FER2013 dataset, tested it on all trials' frames, and mainly got neutral emotions from facial expression recognition. The model had five blocks of convolutional and pooling layers, and its structure was similar to the VGG-16 (Simonyan and Zisserman, 2014) with some extra layers in each block. **Figure 5** shows the structure of model. FER2013 is a large-scale dataset automatically collected by the Google image search API and has been widely used in facial emotion recognition studies. It contains 28,709 training images, 3,589 validation images, and 3,589 test images with seven expression labels: anger, disgust, fear, happiness, sadness, surprise, and neutral. We preprocessed the data by converting the images to grayscale images, extracting the face area using the face detection module from the Dlib library, normalizing and resizing them, and finally, feeding them to the deep convolutional network. We removed the non-detected faces

FIGURE 4 | Six frames of a trial of three different datasets from a watching video task. **(A)** Our dataset, micro-expressions in eyes and lips, **(B)** DEAP, micro-expressions around lips, **(C)** SMIC, micro-expressions on the forehead around the eyebrows, **(D)** FER2013, macro-expressions.

[12]https://scikit-learn.org/stable/index.html

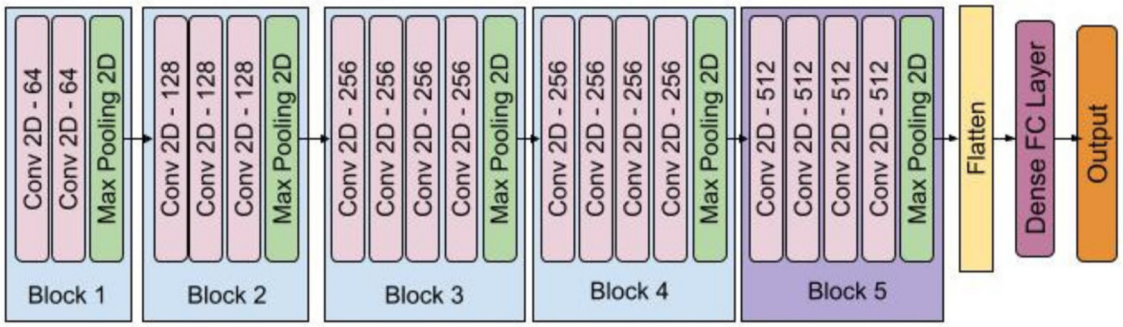


FIGURE 5 | Facial macro-expression model.

TABLE 4 | The result of using facial macro-expression model for detecting emotions from all frames.

	Percentage of trials with all neutral frames	Percentage of trials with majority vote neutral	Percentage of frames with detected emotion						
			Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
DEAP dataset	80.3	100	0.0	0.0	0.0	0.0	98.7	1.3	0.0
Our dataset	6.9	89.1	0.0	0.0	0.0	0.0	87.5	12.5	0.0

from the training and test set and achieved 85% accuracy on the FER2013 test set data. We used the trained model for detecting emotions from each recorded video frame in the DEAP dataset and our dataset. Using the trained model, we applied the same preprocessing steps and predicted each frame's emotion.

Table 4 shows the result of prediction for the DEAP and our datasets. As can be seen, based on the majority vote strategy of all frames' emotions, the detected emotion for 100% of DEAP's trials and 89.1% of the experimental's trials is neutral. For a limited number of participants, the neutral faces were mistakenly predicted as sadness emotion in all trials.

This result shows that neutral faces or faces with subtle or micro-expressions cannot be easily identified with facial macro-expressions methods. Since the condition of recorded video in the DEAP and our datasets is the same as the micro-expression datasets, we used micro-expression methods to detect facial video expressions in these two datasets and investigated their performance. So we considered the facial data in the DEAP and our dataset as facial data with micro-expressions and used a facial micro-expression strategy for video-emotion recognition.

We used a two-steps facial micro-expression recognition strategy. Firstly, we used an automatic spotting strategy to automatically find the apex frame based on maximum facial components' movements compared to the first and last frame of the trial. Then we extracted a set of frames around the apex frame and considered these frames instead of the overall video for classification. Finally, we fed the extracted sequence to a 3D convolutional neural network.

To prepare frames for spotting micro-expressions, first of all, we employed a pre-trained YOLO v3 network (Redmon and Farhadi, 2018) on the WIDER FACE dataset (Yang et al., 2016) for face detection. We chose the WIDER FACE dataset because it contains images with varying degrees of scale, occlusion, and poses, enhancing the feature space for the model to learn better and giving better real-time performance under any condition. Then we followed the spotting method introduced in Van Quang et al. (2019) to identify the apex frame (frame with micro-expression) in each video. In this spotting method, firstly, we extracted ten regions of the face around facial components where muscle movements occur very frequently. For the next step, we considered the first frame of the video sequence as the onset frame and the last frame as the offset frame and calculated the absolute pixel differences between each frame and the onset and offset frames in the ten regions. Finally, we calculated the per-pixel average value for each frame. We considered the frame with the higher intensity differences as the apex frame. We considered a window of frames around the apex frame as the region of interest (ROI) and only used these frames in the classification step (**Figure 6A**).

Although the recorded videos in our dataset and the DEAP dataset are longer and may contain more neutral frames than facial micro-expressions datasets, facial micro-expression spotting methods could still find the apex frame. So we still have the onset, apex, and offset frames. There may be several neutral frames before the actual onset frame and after the offset frame, but all of them are the same and will not affect the result of the spotting algorithm. This is because the actual onset frame and the first frame or the real offset frame and the last frame are almost the same. Hence the measured absolute pixel differences between the apex frame and actual onset or offset frame will be practically the same as those between the apex frame and the first or last frame. Although there may be more facial expressions and apex frames in the video, the actual offset frame, and the last

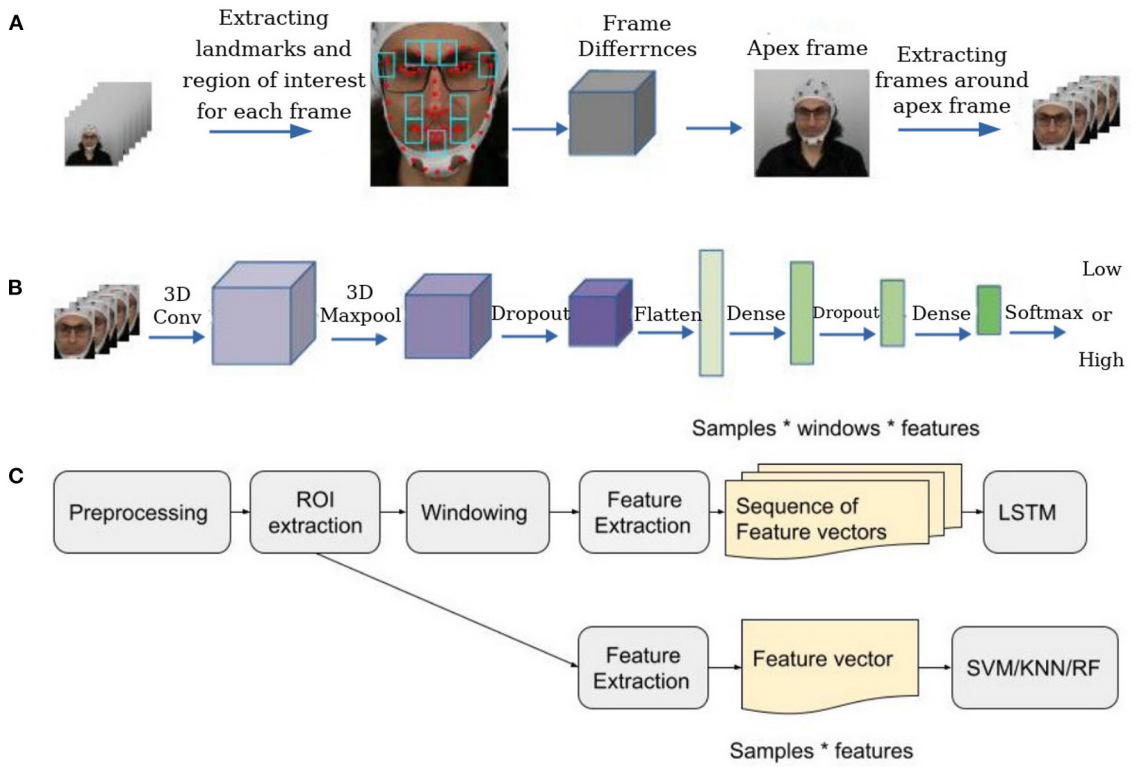


FIGURE 6 | Emotion recognition strategy. **(A)** Locating the apex frame in the sequence of all frames in each trial. **(B)** The architecture of the network for detecting arousal or valence states in facial videos based on micro-expressions. **(C)** The overall structure of EEG and physiological classification.

frame are nearly identical since both frames depict the face in a neutral state.

We considered different window sizes for the ROI and discussed it in the result section. **Figure 4** illustrates six frames of extracted sequences around the apex frame for DEAP and our dataset, in addition to a sequence of the SMIC dataset. We used a 3D Convolutional Neural Network (3D CNN) to classify micro-expression sequences. It is one of the state-of-the-art models in micro-expression emotion recognition (Reddy et al., 2019) which achieved good performance on two popular micro-expressions dataset CASME II (Yan et al., 2014), and SMIC (Li et al., 2013). This method achieved 87.8% accuracy on the CASME II dataset and 68.75% accuracy on the SMIC dataset. We used this model to extract deep features and classify micro-expressions in the DEAP and our datasets. Since the ground-truth labeling in both datasets is based on arousal and valence levels, instead of classifying micro-expressions based on basic emotions, we classified micro-expressions based on arousal and valence levels. To classify emotional states based on arousal or valence, we applied the model two times to the data, once for classifying arousal levels and once for classifying valence levels. In this model, instead of using six as the output shape in the last dense layer, we used 2 to classify micro-expressions based on low and high arousal or valence.

At first, we used the YOLO face detection algorithm to detect the face in each frame in the ROI, then converted it to the grayscale image, normalized it, and resized it. Finally, we fed the preprocessed sequences into two 3D CNN models introduced in Reddy et al. (2019) for classifying arousal and valence separately. **Figure 6B**) illustrates the structure of the 3D CNN model.

5.4. EEG and Physiological Emotion Recognition

We considered micro-expressions as an indicator for identifying the most emotional time of each trial. Then, we used an ROI-based strategy for recognizing arousal and valence using EEG and physiological data. We considered the time of the apex frame as the most emotional time of each trial. Then, we located corresponding samples in the EEG and physiological data at this time. Due to the difference in sampling rates between EEG, physiological data, and video frames, we multiplied the sampling rate of each signal at this time to determine the ROI. Finally, we extracted a couple of seconds of data around it, considered the extracted part as the ROI, and analyzed only the extracted data. We regarded different window sizes for extracting ROI and discussed it in the result section.

To analyze EEG and physiological data, we followed the main steps of emotion recognition: preprocessing, feature extraction, and classification. Firstly, we cleaned data and then extracted ROI sections and only used ROI data as the input of the feature extraction step. To classify data, we used two methods for classifying EEG and physiological data. In the first method, we

extracted some features—described in the following sections—from the whole data or ROI section. We used these features as the input of Support Vector Machine (SVM), K-Nearest Neighbor (KNN) (Bressan and Vitria, 2003) and Random Forest (RF) (Criminisi et al., 2011) classifiers. In the second method, firstly, we partitioned each trial into non-overlapping windows. Then extracted features the same as the previous method from each window and made a sequence of consequences feature vectors. We used these sequences as the input of a stacked Long-Short-Term-Memory (LSTM) network (Staudemeyer and Morris, 2019) with two layers of LSTM to extract temporal features. Finally, we used a Dense layer with Adam optimizer (Kingma and Ba, 2014) to separately classify the data for arousal and valence labels. **Figure 6C** shows the overall structure of EEG and physiological data analysis.

5.5. Data Cleaning

5.5.1. EEG

We used the preprocessed EEG data in the DEAP dataset, removed the first 8 s of data, including 3 s of baseline, and considered 5 s as the engagement time and finally normalized data. The engagement time was chosen by observation and was the average time participants were immersed in the video. For our dataset, we applied bandpass filters and extracted frequencies between 1 and 45 Hz which are the frequency range of brain waves (Huang et al., 2016). Then a common average reference was applied, and finally, we normalized the data. **Figures 7A,B** show the frequencies of EEG channels before and after data cleaning.

5.5.2. PPG and GSR

A bandpass filter with a low-cut frequency of 0.7 Hz and a high-cut frequency of 2.5 Hz has been used to remove noise from the PPG signals. Similarly, a low-cut frequency of 0.1 and a high-cut frequency of 15 Hz were used to clean the GSR signals. We also used a median filter to remove rapid transient artifacts from the GSR signal. Finally, we normalized these GSR and PPG signals. **Figures 7C–F** shows the amplitude of one sample of GSR and PPG signals before and after data cleaning.

5.6. Feature Extraction

5.6.1. EEG

To extract EEG features, we applied a Fast Fourier Transform (FFT) on each window of data to extract EEG band powers. We made a feature vector of five features by extracting EEG power bands from each window and considered the average of each as one feature. We extracted Delta (1–4 HZ), Theta (4–8 HZ), Alpha (8–12 HZ), Beta (12–30 HZ), and Gamma (30–45) bands. These features have commonly been used in previous studies (Wagh and Vasanth, 2019).

5.6.2. PPG and GSR

We calculated some statistical features for both GSR and PPG signals. The average and standard deviation of the GSR signal and the first and second-order discrete differences of the GSR signal made up the GSR feature vector. To build the PPG feature vector, we considered the average and standard deviation of the PPG signal. The PPG and GSR feature vectors have similar characteristics, so we concatenated the two feature vectors and referred to them as physiological data.

5.7. Fusion Strategy

There are several methods for fusing data from various sources. Fusing data can be done mainly in two major ways, (1) feature-level or early fusion and (2) decision-level fusion or late fusion (Shu et al., 2018). We fused the PPG and GSR signals at the feature level, addressed the created features as physiological features, and classified them. We used two different strategies for fusing facial micro-expressions, EEG and physiological classification results in the decision level. The first strategy was based on majority voting, where we selected the prediction that had the most votes among EEG, facial and physiological predictions as the final prediction. In the second strategy, we used the weighted sum of all probabilities as the decision level fusion strategy (Koelstra and Patras, 2013; Huang et al., 2017). We gave various weights in the range [0, 1] with 0.01 steps to these three classifiers, measured the best weights on the training data, and used these weights in the fusion step. The Equation (1), $p^x_{Modality}$ shows the probability of each class using a specific modality, and a, b, and c are weights.

$$p^x_o = a \times p^x_{Video} + b \times p^x_{EEG} + c \times p^x_{Physiological} \tag{1}$$

$$x \in [0, 1]$$

$$a + b + c = 1$$

6. RESULT AND DISCUSSION

6.1. Evaluation Strategy

We used a subject-independent strategy to evaluate our methods and find a general model. We used the leave-some-subject-out strategy cross-validation. Since our models were not complex and the size of datasets was not significant, we did not use a GPU for training models. All models were trained on a computer with Gnu-Linux Ubuntu 18.04, Intel(R) Core(TM) i7-8700K CPU (3.70 GHz) with six cores. We randomly shuffled participants into six-folds and trained models for all folds in parallel. For the DEAP dataset, 3 participants were considered in the test set in each fold. In our dataset, four participants were considered in the test set. The reported result is the average of all folds results.

The four main metrics in evaluating models are accuracy, precision, recall, and F-Score or F1. They are measured using the Equation (2) for binary classification. In this section, all of the results are based on F-Score. In these equations, TP is True Positive which means the number of correctly positive class predictions. The True Negative (TN) measures how many correctly negative predictions were made. False Positives are the number of incorrectly predicted positive classes. FN stands for False Negative, the number of incorrectly negative class predictions. We used binary classification for classifying arousal and valence separately and chose the F-Score for evaluating

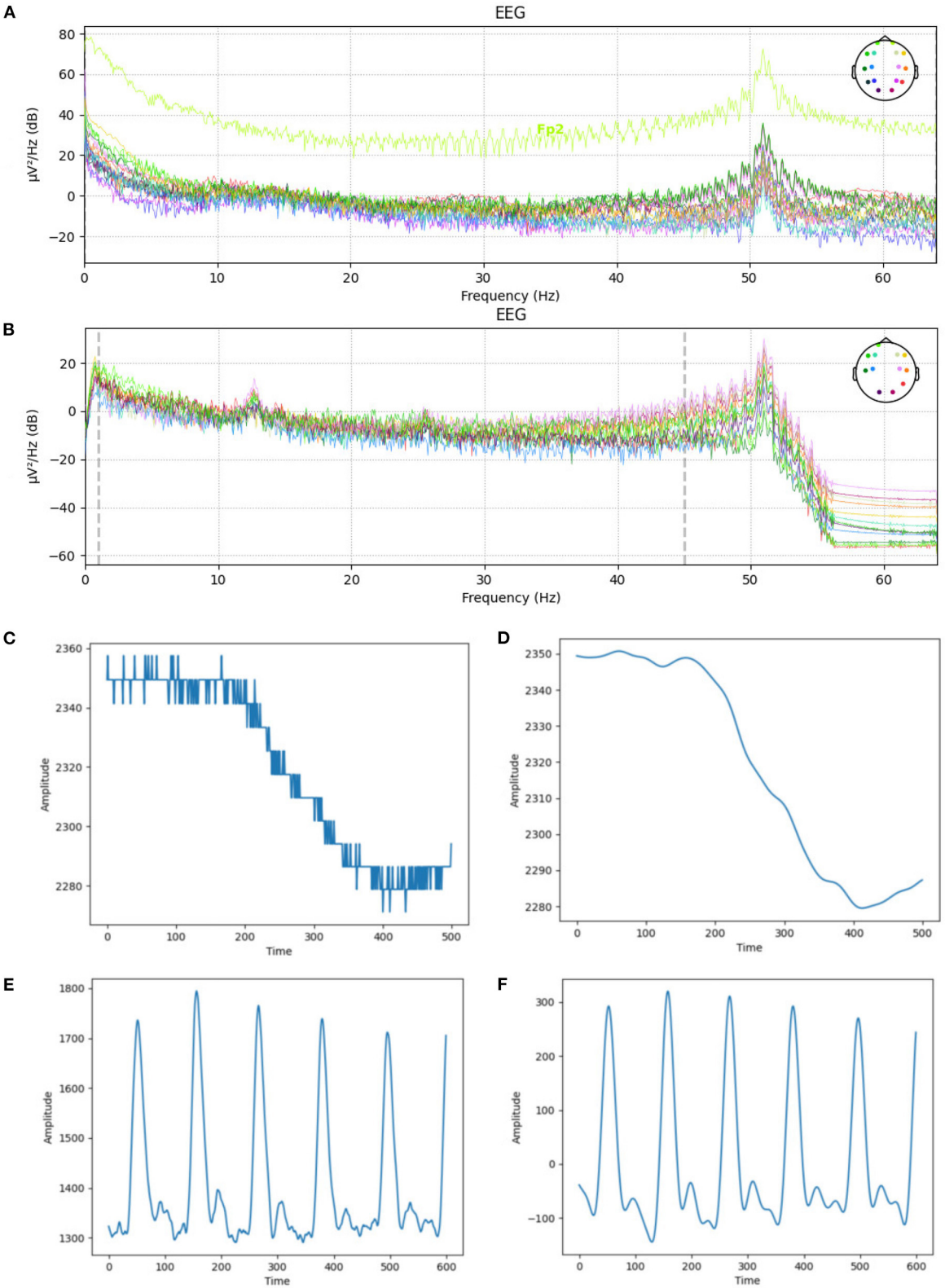


FIGURE 7 | The EEG, GSR, and PPG signals before and after applying preprocessing steps and noise removal. **(A)** EEG channels' frequencies before preprocessing. **(B)** EEG channels' frequencies after preprocessing. **(C)** GSR signal's amplitude before preprocessing. **(D)** GSR signal's amplitude after preprocessing. **(E)** PPG signal's amplitude before preprocessing. **(F)** PPG signal's amplitude after preprocessing.

TABLE 5 | The F-Score of facial micro-expression recognition when the window size is 20 frames around the apex frame or 60 frames.

	Arousal		Valence	
Window size	20	60	20	60
DEAP	55.0	**59.0**	55.7	**56.8**
Experimental	61.0	**62.2**	57.0	**61.1**

The bolded values show the highest F-scores for each dataset and emotional level according to the window size. Increasing the window size to 60 frames increase the F-score.

our methods which are appropriate for imbalanced data. (Sun et al., 2009).

$$Accuracy = (TP + TN)/(TP + FP + FN + TN) \quad (2)$$

$$Precision = TP/(TP + FP)$$

$$Recall = TP/(TP + FN)$$

$$F1 = 2 \times (Recall \times Precision)/(Recall + Precision)$$

6.2. Hyper-Parameter Tuning

To measure the best hyper-parameters for SVM, RF, and KNN, we used grid-search cross-validation parameter tuning (Claesen and De Moor, 2015). Hyper-parameters were tuned using six-fold cross-validation when split data based on the leave-some-subject-out strategy. We got the best result when considering Radial Basis Function (RBF) kernel with 200 as the regulation parameter for SVM, five neighbors for KNN, and 500 estimators for RF. For the 3D convolutional model for micro facial expression, we used the same parameters as the source study (Reddy et al., 2019). We only set the number of epochs to 50. We also empirically found that using two stacked LSTM generates better results when the first LSTM has 80 neurons and the second has 30 neurons. We considered 128, 32, and 64 as the batch size in LSTM model training for EEG, GSR, and PPG classifiers and set the number of epochs to 100 for them. We did not tune the learning rates. Instead, we used a reduced learning rate in the range of 0.001–0.0001, which decreases with a rate of 0.5 when validation loss is not changing.

6.3. Identifying ROI Size

Micro-expression duration varies between 65 and 500 ms. This time may increase when the emotion lasts for a while or may merge with the next micro-expression that is the response of the subsequent emotional stimulus (Yan et al., 2013). The DEAP dataset recorded facial data at 50 frames per second. This means that if we consider the length of a micro-expression as half a second, a micro-expression appears in 25 frames when the frame rate is 50 Hz. In our dataset, the frame rate is 30 frames per second, so the length of a micro-expression is 15 frames. We considered two different window sizes, including 20 and 60 frames, around the apex frames to cover short micro-expressions or long-lasting micro-expressions. We considered a bigger window size to cover micro-expressions that remain longer or overlap with the next micro-expression. **Table 5** compares the effect of these two window sizes on the prediction result when we want to classify emotions according to arousal and valence levels. As can be seen, the result of 60 frames is better for both datasets. Since increasing the window size increases the probability of including other head movements, adding non-informative data to the sequence, and increasing the computation cost, we did not consider a bigger window size. **Table 5** shows the f-score of 3D CNN models from the DEAP and our datasets for these two different window sizes. We used the prediction result of facial micro-expression classification combined with the other modalities at the decision level to classify arousal and valence levels.

We considered various sizes for extracting the ROI from EEG and physiological data and compared the effect of ROI size on the classification result. **Figure 8** shows the impact of various ROI sizes on the classification result when we used the LSTM method. The reported values are the average of F-Score values for all folds. As shown in **Figure 8**, for both datasets, the window size of 15 created almost the highest F-Score when using majority fusion. For the DEAP dataset, weighted fusion created the best result for predicting arousal when we considered all of the data. Despite not seeing any consistent pattern in the two different datasets shown here, assuming a small portion of data in the most emotional part can yield a similar or better result than using all the data. This indicates that if we accurately identify the most emotional part of data, we can accurately study brain and body responses to emotional stimuli.

We also used SVM, KNN, and RF classifiers to classify the ROI section when the window size is 15 and when the whole data was considered. We compared the F-Score of these classifiers with the LSTM method when all data or only the ROI section has been considered for classification in **Table 6**. The F-Score of facial micro-expression with window size 60 reported in **Table 5** has been considered in the fusion strategies. We fused the prediction result of the facial micro-expression method with all classifiers that we used. As can be seen in these tables, the LSTM method achieved the best result in both datasets for arousal and valence. This shows that exploiting both temporal and spatial features could help detect emotions. Also, the result of fusion strategies is considerably better than single modalities. The majority vote fusion in our dataset for arousal and valence, and only valence in DEAP outperforms weighted fusion. Combining PPG and GSR only improves the performance of the LSTM method in classifying valence levels when applied to ROI data. Also, the F-Score of the ROI-based LSTM is relatively close to or sometimes better than using LSTM on the whole of the data. This shows that using a small portion of data can be informative as using all of the data.

Although other classifiers generated a good result for some modalities for arousal or valence in one of these datasets, their predictions did not improve after fusing with other modalities compared to the LSTM method. When the prediction accuracy is below or around random prediction or 50% for binary classification, it could not find any particular pattern in the data. Thus, the mismatching between single modalities prediction

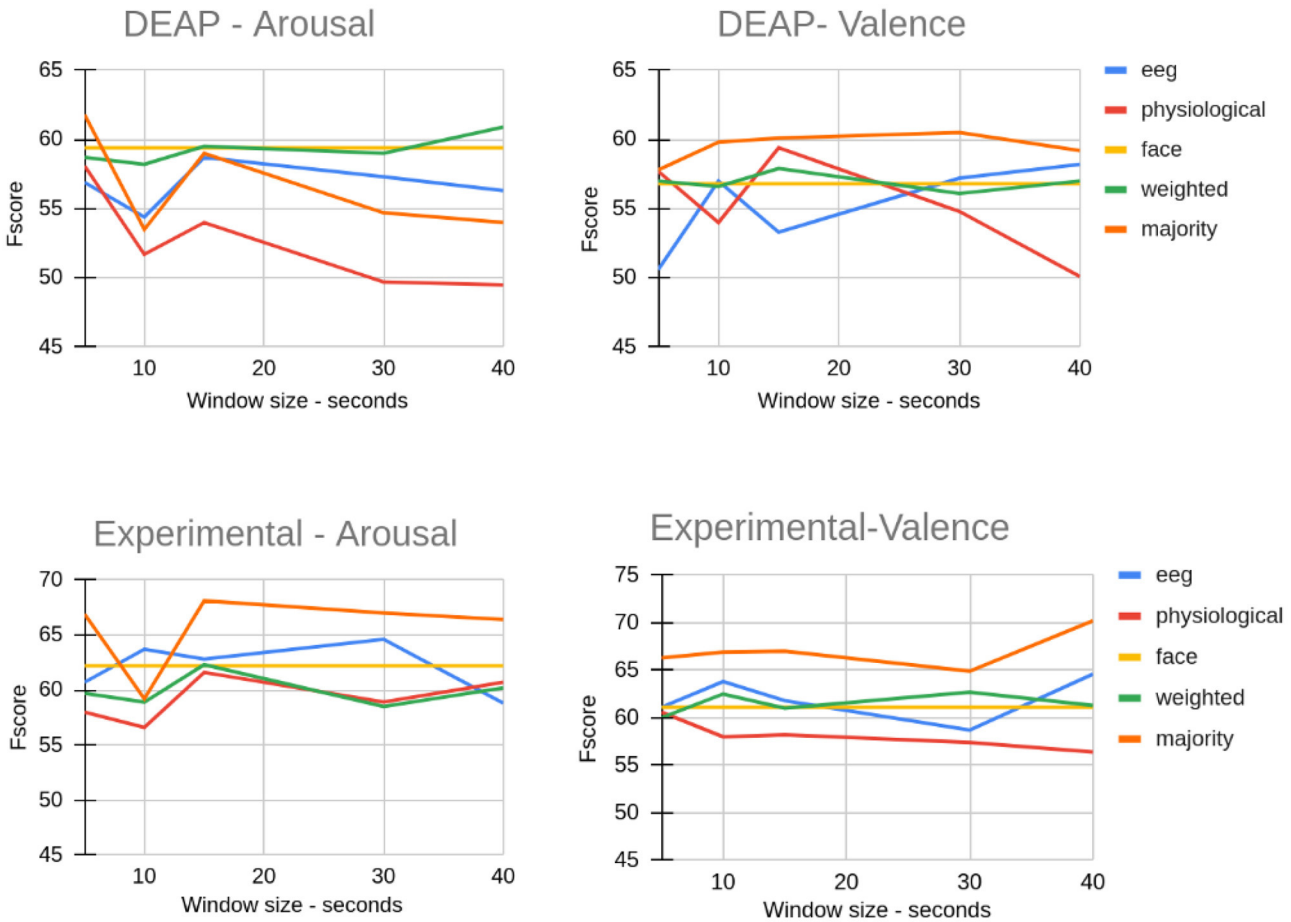


FIGURE 8 | The effect of ROI size on EEG, physiological, and fusion classification on the DEAP dataset and our experimental dataset (Window size 40 means all frames).

increases, leading to a degrading f-score in the fusion strategy. According to **Table 6**, the f-scores of SVM, KNN, or RFC for some modalities were below or around 50%. Therefore, this leads to ineffective fusion.

6.4. Computation Cost

Instead of using all frames as the input of the 3D convolutional model, only 60 frames of each video were employed as the model's input. The DEAP dataset has 3,000 frames in each video, and our dataset has 2,400 frames in each video. By extracting micro-expression ROI, we decrease the input size for DEAP with the rate (60/3,000) and our dataset with (60/2,400). This drop-off in input size leads to a considerable decline in computation cost. Our dataset has 230 (23 * 10) trials for all participants, while the DEAP dataset has 720 (18 * 40) trials for all participants. The face model's input for both datasets is (60 * 64 * 64), where 60 is the number of frames in each trial and 64 * 64 is the dimension of the frame in grayscale in the face area. Training six-folds of face models for the DEAP dataset in parallel took 1 h and 37 min (235 s for each epoch). This time was 33 min for our dataset (79 s for each epoch) because of each participant's lower number of trials.

Moreover, despite previous studies which used the LSTM network for classifying EEG signals and feeding raw signals as the input of network (Ma et al., 2019), we extracted a limited number of features from each second of data to decrease the input size. We created a new sequence of data that is considerably smaller than raw data while still being informative. For example, for the EEG data, the size of each trial was (duration in seconds * sampling rate * channels). We decreased this size to (duration in seconds * five power bands). This decrease is the same for physiological data. Training the LSTM models for EEG, PPG, and GSR were done in parallel for six-folds. It took 12 min and 14 s to train all these models for the DEAP dataset, while each epoch took around 1–3 s to run. The training time for our dataset took 5 min, with each epoch taking between 25 and 100 ms to run.

6.5. Final Result

Table 7 shows the final results of classifying the ROI section for a single modality or fusion strategy when the ROI window size is 15 s. As can be seen, fusing micro expressions with EEG and physiological signals leads to higher accuracy and F-Score than using a single modality in both datasets. We achieved similar or

TABLE 6 | Comparison of F-Score value of LSTM, RF, SVM, and KNN methods when ROI of 15 s or all of data was used.

				Modality					
Dataset	**Emotional state**	**Method**	**ROI length**	**EEG**	**PPG**	**GSR**	**PPG+GSR**	**Weighted fusion**	**Majority fusion**
Our dataset	**Arousal**	**SVM**	**15 s interval**	51.7	43.7	44.6	43.0	55.0	50.9
			All	50.5	44.3	44.8	43.7	57.3	51.1
		RF	15 s interval	57.9	50.9	52.9	51.6	60.3	61.3
			All	56.8	46.0	52.6	49.6	56.5	55.8
		KNN	15 s interval	62.5	50.0	52.9	47.7	53.6	61.0
			All	62.1	44.5	49.6	47.9	59.3	60.4
		LSTM	15 s interval	62.8	51.9	65.6	61.6	62.3	**68.1**
			All	58.8	54.9	60.7	60.7	60.2	66.4
	Emotional state	**Method**	**ROI length**	**EEG**	**PPG**	**GSR**	**PPG+GSR**	**Weighted fusion**	**Majority fusion**
	Valence	**SVM**	15 s interval	45	40.7	41.3	41.8	58.3	41.5
			All	44.3	49.5	47.0	39.4	57.3	43.0
		RF	15 s interval	52.2	49.4	45.5	38.3	50.9	54.3
			All	53.1	49.2	43.9	47.7	55.4	56.0
		KNN	15 s interval	54.5	43.9	48.9	43.5	57.1	57.1
			All	53.7	46.8	51.1	46.7	52.7	56.8
		LSTM	15 s interval	61.8	56.7	56.7	58.2	61.0	67.0
			All	64.6	59.5	61.0	56.4	61.3	**70.2**
Dataset	**Emotional state**	**Method**	**ROI length**	**EEG**	**PPG**	**GSR**	**PPG+GSR**	**Weighted fusion**	**Majority fusion**
DEAP	**Arousal**	**SVM**	15 s interval	48.2	47.2	47.2	47.1	58.9	48.1
			All	50.0	47.2	47.0	47.2	58.1	48.3
		RF	f15 s interval	57.2	56.4	51.6	51.5	52.7	56.0
			fAll	53.2	52.8	52.3	53.5	55.1	54.7
		KNN	15 s interval	55.4	55.5	49.1	49.0	55.4	56.1
			All	53.3	53.8	51.5	51.2	53.9	57.4
		LSTM	15 s interval	58.7	49.4	55.7	54.0	59.5	59.0
			All	56.3	47.2	50.1	49.5	**60.9**	54.0
	Emotional state	**Method**	**ROI length**	**EEG**	**PPG**	**GSR**	**PPG+GSR**	**Weighted fusion**	**Majority fusion**
	Valence	**SVM**	15 s interval	44.9	39.7	37.2	49.9	51.0	46.6
			All	47.2	40.5	38.3	42.0	54.7	51.8
		RF	15 s interval	50.8	48.7	50.4	47.4	52.4	54.8
			All	51.4	50.7	49.3	51.9	53.0	58.0
		KNN	15 s interval	49.5	51.1	48.9	52.0	52.9	53.6
			All	52.4	47.7	53.1	49.5	51.4	53.0
		LSTM	15 s interval	53.3	59.4	51.2	54.4	57.9	**60.1**
			All	58.2	50.1	47.0	50.8	57.0	59.2

The bolded values show the highest F-scores for each dataset and emotional level. The LSTM method shows the best result.

better accuracy in recognizing arousal and valence levels than related works that used subject-independent strategies.

There is not any standard benchmark for evaluating various emotion recognition studies. There are multiple datasets with different scenarios in data collection that record emotional data using various modalities and sensors. The variety in the datasets, emotion models, the way of splitting data, evaluation strategies, and evaluation metrics affect the final emotion recognition results. For this reason, we should consider all these factors for comparing various studies. Compared to

TABLE 7 | Accuracy, F-score, precision, and recall for arousal and valence.

Arousal

Accuracy	EEG	Face	PPG	GSR	Physio	Majority
Our dataset	66.0	61.4	62.6	69.6	64.6	**70.8**
DEAP	62.9	60.4	62.4	63.3	63.2	**65.1**
Precision	**EEG**	**Face**	**PPG**	**GSR**	**Physio**	**Majority**
Our dataset	64.4	64.0	48.3	68.0	62.0	**70.9**
DEAP	59.0	60.8	59.0	59.9	55.0	**69.0**

F-Score	EEG	Face	PPG	GSR	Physio	Majority
Our dataset	62.8	62.2	51.9	65.6	61.6	**68.1**
DEAP	58.7	59.4	49.4	55.7	54.0	**59.0**
Recall	**EEG**	**Face**	**PPG**	**GSR**	**Physio**	**Majority**
Our dataset	66.0	61.4	62.6	69.6	64.6	**70.8**
DEAP	62.9	60.4	62.4	63.3	63.2	**70.0**

Valence

Accuracy	EEG	Face	PPG	GSR	Physio	Majority
Our dataset	64.0	61.0	62.6	60.7	63.2	**69.2**
DEAP	58.1	57.9	58.5	59.0	60.3	**62.4**
Precision	**EEG**	**Face**	**PPG**	**GSR**	**Physio**	**Majority**
Our dataset	66.4	62.9	67.0	63.6	62.9	**69.2**
DEAP	56.9	58.9	61.1	55.5	60.5	**64.1**

F-Score	EEG	Face	PPG	GSR	Physio	Majority
Our dataset	61.8	61.1	56.7	56.7	58.2	**67.0**
DEAP	53.3	56.8	51.2	54.4	59.4	**60.1**
Recall	**EEG**	**Face**	**PPG**	**GSR**	**Physio**	**Majority**
Our dataset	64.0	61.0	62.6	60.7	63.2	**69.2**
DEAP	58.1	57.9	58.5	59.0	60.3	**62.4**

The best values for each measurement are bolded in each row for our dataset and DEAP dataset.

the previous work reported in **Table 1**, the accuracy of the proposed methods is considerably high while considering the subject-independent approach, which is the most challenging evaluation condition.

Although detecting facial micro-expressions is still a big challenge in the literature and needs more exploration, we have shown that it could considerably decrease computational costs for video emotion recognition. There are some challenges for detecting micro-expressions that affect emotion recognition performance, including contamination with other facial movements, pose changes, poor illumination, and the possibility of faked or posed micro-expressions (Zhao and Li, 2019). With the DEAP dataset and our dataset, the chance of faked micro-expressions is low due to the poker face condition. There are, however, some unwanted movements that can affect the results of detecting micro-expressions and identifying regions of interest.

We also found that the low-cost OpenBCI EEG cap could achieve similar performance to the Biosemi Active II cap used in the DEAP dataset. Our result shows that although this tool is low-cost, it can be used as a reliable tool for collecting brain signals for emotion recognition.

Similar to previous studies, our result shows that combining various modalities leads to a better recognition result with a 3–8% improvement after fusion. We achieved 65.1% accuracy for arousal and 69.2% accuracy for valence in the DEAP dataset, which is better than single modalities. These corresponding values are 70.8 and 69.2% for arousal and valence in our dataset. **Table 7** shows these improvements. Although there are some disadvantages to employing multimodal data, such as increased computing cost and data analysis complexity, the benefits of enhancing prediction performance outweigh them. Nowadays, most processing systems have multiple cores, making parallel processing easy. We can perform multimodal data analysis using parallel processing at almost the same time as a single modality analysis.

7. LIMITATIONS

Despite showing that facial micro-expressions can effectively identify emotions, we face some challenges that should be addressed in the future. Due to involuntary facial movements, such as eye blinking, head movements, or regular facial expressions, micro-expressions can be mistakenly detected (Tran et al., 2020). These movements result in incorrect detection of the apex frame. In the future, we could significantly improve the spotting strategy result by introducing new facial micro-expression datasets and using deep learning methods. In this paper, we used a simple traditional micro-expression spotting strategy to detect the apex frame. We showed that facial micro-expressions could be combined with other modalities in emotion recognition. In the future, we want to use more robust spotting strategies to improve recognition quality.

Furthermore, facial micro-expression methods face challenges similar to facial macro-expressions, including illumination conditions, cultural diversity, gender, and age. These limitations can be overcome by using new datasets and more robust deep learning methods. Also, combining facial micro-expressions with physiological signals will improve the recognition result. EEG headsets and physiological sensors are not as accessible as cameras for most people. We are now closer than ever to developing robust models for emotion recognition because more and more affordable and wearable devices like

smartwatches, activity trackers, and VR headsets are equipped with physiological sensors. We can achieve this objective by introducing more accurate, wearable, and affordable EEG sensors and developing a more robust algorithm for physiological emotion recognition.

We used a spotting strategy to detect apex frames in our study. Since spotting methods still need more exploration and are an open challenge (Oh et al., 2018), we could improve our results in the future by manually annotating the DEAP and our dataset. Manually annotating these datasets is a labor-intensive and time-consuming activity. Still, because they were collected under similar conditions as micro-expression datasets, we can use them as micro-expression datasets for making more robust micro-expression models.

8. CONCLUSIONS

It can be essential to accurately recognize emotions for human-human and human-machine applications. The previous techniques relied heavily on facial macro-expressions. This paper demonstrated our strategy for how facial micro-expressions can be used effectively with EEG and physiological signals to recognize emotional states.

In this paper, we used facial micro-expressions emotion recognition instead of facial macro-expressions emotion recognition combined with physiological modalities, which is more reliable in identifying genuine emotions. Also, we used a facial micro-expressions spotting strategy to roughly determine the most emotional and informative part of the data. We identified each trial's region of interest (ROI) using a landmark-based spotting strategy for detecting micro-expressions. Several frames around the micro-expression were extracted and fed to a 3D convolutional network. In addition, we extracted a sequence of feature vectors from EEG and physiological data in the ROI when the data was partitioned into 1 s windows. To extract temporal features from physiological signals and EEG signals, we employed LSTM. We evaluated ROI classification with LSTM, SVM, KNN, and RF classifiers compared to classifying all data. Our methods were evaluated based on a subject-independent approach. According to our results, we could obtain a similar or even better accuracy by using a small portion of data compared to all the data. According to our findings, facial micro-expressions could identify the more emotional part of data with sufficient information and low noise.

Moreover, we used a low-cost, open-source EEG headset to collect multimodal emotional data. We evaluated our method based on the DEAP dataset and our own data. Lastly, we combined multiple modalities and found that fusing their outputs improved emotion recognition. In addition, we found that facial micro-expressions were more effective at detecting genuine emotions than facial macro-expressions methods.

Due to the high data quality and ease of use of the OpenBCI hardware, we want to follow up our study with more data collection with various settings with OpenBCI. The collected data will be used to pre-train the upcoming models to create a robust model for recognizing emotions in EEG data. In the future, after getting ethics approval for publishing the dataset, we want to make the EEG and physiological data publicly available. This will help researchers to train more robust models for emotion recognition.

We would like to examine more features in the future and see if changing the feature set or using more complex features will improve the LSTM method performance. We would also like to use more complex fusion strategies to exploit multimodal sensors effectively.

In addition, another future direction is to explore how facial micro-expressions can be extracted from more natural head movements (for example, not requiring people to maintain a poker face). Moreover, it would be interesting to identify facial micro-expressions in the presence of regular facial expressions and explore how a combination of both could be used to recognize emotions.

Finally, we are interested in incorporating this emotion recognition approach into applications such as healthcare with remote therapy sessions, identifying emotional disorders in patients, or creating intelligent assistants to help patients or the elderly. In addition, this emotion model can be used in our daily interactions with humans, such as enhancing teleconferencing and making remote interactions more immersive. Furthermore, it will improve our interactions with virtual agents and other interactive devices we regularly use by giving them the ability to recognize and respond to our emotions.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not publicly available due to privacy concerns. Requests to access the datasets should be directed to zsaf419@aucklanduni.ac.nz.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Auckland Human Participants Ethics Committee. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

NS: the main author, conducting the research, data collection, data analysis, and writing. SW: facial micro-expression analysis. KD: facial detection model. AN: technical and writing feedback. SN: co-supervisor and technical feedback. EB: co-supervisor, psychological support, and feedback. MB: main supervisor and lab coordinator, technical, and writing feedback. All authors contributed to the article and approved the submitted version.

REFERENCES

Alarcao, S. M., and Fonseca, M. J. (2017). Emotions recognition using EEG signals: a survey. *IEEE Trans. Affect. Comput*. 10, 374–393. doi: 10.1109/TAFFC.2017.2714671

Allaert, B., Bilasco, I. M., and Djeraba, C. (2018). Advanced local motion patterns for macro and micro facial expression recognition. *arXiv preprint* arXiv:1805.01951. doi: 10.48550/arXiv.1805.01951

Aranha, R. V., Corrêa, C. G., and Nunes, F. L. (2019). Adapting software with affective computing: a systematic review. *IEEE Trans. Affect. Comput*. 12, 883–899. doi: 10.1109/TAFFC.2019.2902379

Aydın, S. (2020). Deep learning classification of neuro-emotional phase domain complexity levels induced by affective video film clips. *IEEE J. Biomed. Health Inform*. 24, 1695–1702. doi: 10.1109/JBHI.2019.2959843

Aydın, S., Demirtaş, S., and Yetkin, S. (2018). Cortical correlations in wavelet domain for estimation of emotional dysfunctions. *Neural Comput. Appl*. 30, 1085–1094. doi: 10.1007/s00521-016-2731-8

Bartlett, M. S., Littlewort, G., Fasel, I., and Movellan, J. R. (2003). "Real time face detection and facial expression recognition: Development and applications to human computer interaction," in *2003 Conference on Computer Vision and Pattern Recognition Workshop* (Madison, WI: IEEE).

Bazgir, O., Mohammadi, Z., and Habibi, S. A. H. (2018). "Emotion recognition with machine learning using EEG signals," in *2018 25th National and 3rd International Iranian Conference on Biomedical Engineering (ICBME)* (Qom: IEEE), 1–5. doi: 10.1109/ICBME.2018.8703559

Ben, X., Ren, Y., Zhang, J., Wang, S.-J., Kpalma, K., Meng, W., et al. (2021). Video-based facial micro-expression analysis: a survey of datasets, features and algorithms. *IEEE Trans. Pattern Anal. Mach. Intell*. doi: 10.1109/TPAMI.2021.3067464

Benlamine, M. S., Chaouachi, M., Frasson, C., and Dufresne, A. (2016). "Physiology-based recognition of facial micro-expressions using EEG and identification of the relevant sensors by emotion," in *International Conference on Physiological Computing Systems - PhyCS, Vol. 2* (Lisbon: SciTePress), 130–137. doi: 10.5220/0006002701300137

Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9

Bressan, M., and Vitria, J. (2003). Nonparametric discriminant analysis and nearest neighbor classification. *Pattern Recogn. Lett*. 24, 2743–2749. doi: 10.1016/S0167-8655(03)00117-X

Busso, C., Bulut, M., Lee, C.-C., Kazemzadeh, A., Mower, E., Kim, S., et al. (2008). Iemocap: interactive emotional dyadic motion capture database. *Lang. Resour. Eval*. 42:335. doi: 10.1007/s10579-008-9076-6

Chaparro, V., Gomez, A., Salgado, A., Quintero, O. L., Lopez, N., and Villa, L. F. (2018). "Emotion recognition from EEG and facial expressions: a multimodal approach," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Honolulu, HI: IEEE), 530–533. doi: 10.1109/EMBC.2018.8512407

Chin, H., Molefi, L. W., and Yi, M. Y. (2020). "Empathy is all you need: how a conversational agent should respond to verbal abuse," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (New York, NY: Association for Computing Machinery), 1–13. doi: 10.1145/3313831.3376461

Claesen, M., and De Moor, B. (2015). Hyperparameter search in machine learning. *arXiv preprint* arXiv:1502.02127. doi: 10.48550/arXiv.1502.02127

Criminisi, A., Shotton, J., and Konukoglu, E. (2011). *Decision Forests for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning*. Technical Report, Microsoft Research Cambridge, MSRTR-2011-114, 5, 12. doi: 10.1561/9781601985415

Das, P., Khasnobish, A., and Tibarewala, D. (2016). "Emotion recognition employing ECG and GSR signals as markers of ANS," in *2016 Conference on Advances in Signal Processing (CASP)* (Pune: IEEE), 37–42. doi: 10.1109/CASP.2016.7746134

Davison, A. K., Yap, M. H., and Lansley, C. (2015). "Micro-facial movement detection using individualised baselines and histogram-based descriptors," in *2015 IEEE International Conference on Systems, Man, and Cybernetics* (Hong Kong: IEEE), 1864–1869. doi: 10.1109/SMC.2015.326

Deng, L., Wang, X., Jiang, F., and Doss, R. (2021). Eeg-based emotion recognition *via* capsule network with channel-wise attention and LSTM models. *CCF Trans. Pervas. Comput. Interact*. 3, 425–435. doi: 10.1007/s42486-021-00078-y

Doma, V., and Pirouz, M. (2020). A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals. *J. Big Data* 7, 1–21. doi: 10.1186/s40537-020-00289-7

Droit-Volet, S., Fayolle, S. L., and Gil, S. (2011). Emotion and time perception: effects of film-induced mood. *Front. Integr. Neurosci*. 5:33. doi: 10.3389/fnint.2011.00033

Eerola, T., and Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychol. Mus*. 39, 18–49. doi: 10.1177/0305735610362821

Ekman, P. (2003). Darwin, deception, and facial expression. *Ann. N.Y. Acad. Sci*. 1000, 205–221. doi: 10.1196/annals.1280.010

Ekman, P., and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *J. Pers. Soc. psychol*. 17:124. doi: 10.1037/h0030377

Ekman, P., and Rosenberg, E. L. (1997). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford: Oxford University Press.

Elgendi, M. (2012). On the analysis of fingertip photoplethysmogram signals. *Curr. Cardiol. Rev*. 8, 14–25. doi: 10.2174/157340312801215782

Ganapathy, N., Veeranki, Y. R., Kumar, H., and Swaminathan, R. (2021). Emotion recognition using electrodermal activity signals and multiscale deep convolutional neural network. *J. Med. Syst*. 45, 1–10. doi: 10.1007/s10916-020-01676-6

Goh, K. M., Ng, C. H., Lim, L. L., and Sheikh, U. U. (2020). Micro-expression recognition: an updated review of current trends, challenges and solutions. *Vis. Comput*. 36, 445–468. doi: 10.1007/s00371-018-1607-6

Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., et al. (2013). "Challenges in representation learning: a report on three machine learning contests," in *International Conference on Neural Information Processing* (Berlin; Heidelberg: Springer), 117–124. doi: 10.1007/978-3-642-42051-116

Gross, J. J., and Levenson, R. W. (1995). Emotion elicitation using films. *Cogn. Emot*. 9, 87–108. doi: 10.1080/02699939508408966

Guermazi, R., Ben Abdallah, T., and Hammami, M. (2021). Facial micro-expression recognition based on accordion spatio-temporal representation and random forests. *J. Vis. Commun. Image Represent*. 79:103183. doi: 10.1016/j.jvcir.2021.103183

Hashmi, M. F., Ashish, B. K. K., Sharma, V., Keskar, A. G., Bokde, N. D., Yoon, J. H., et al. (2021). LARNet: Real-time detection of facial micro expression using lossless attention residual network. *Sensors* 21:1098. doi: 10.3390/s21041098

Hassouneh, A., Mutawa, A., and Murugappan, M. (2020). Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods. *Inform. Med. Unlock*. 20:100372. doi: 10.1016/j.imu.2020.100372

Hossain, M. Z., and Gedeon, T. (2019). Observers' physiological measures in response to videos can be used to detect genuine smiles. *Int. J. Hum. Comput. Stud*. 122, 232–241. doi: 10.1016/j.ijhcs.2018.10.003

Hu, J., Huang, Y., Hu, X., and Xu, Y. (2021). "Enhancing the perceived emotional intelligence of conver-sational agents through acoustic cues," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–7. doi: 10.1145/3411763.3451660

Huang, H., Xie, Q., Pan, J., He, Y., Wen, Z., Yu, R., et al. (2021). An eeg-based brain computer interface for emotion recognition and its application in patients with disorder of consciousness. *IEEE Trans. Affect. Comput*. 12, 832–842. doi: 10.1109/TAFFC.2019.2901456

Huang, X., Kortelainen, J., Zhao, G., Li, X., Moilanen, A., Seppänen, T., et al. (2016). Multi-modal emotion analysis from facial expressions and electroencephalogram. *Comput. Vis. Image Understand*. 147, 114–124. doi: 10.1016/j.cviu.2015.09.015

Huang, Y., Yang, J., Liao, P., and Pan, J. (2017). Fusion of facial expressions and EEG for multimodal emotion recognition. *Comput. Intell. Neurosci*. 2017:2107451. doi: 10.1155/2017/2107451

Huang, Y., Yang, J., Liu, S., and Pan, J. (2019). Combining facial expressions and electroencephalography to enhance emotion recognition. *Fut. Intern*. 11:105. doi: 10.3390/fi11050105

Islam, M. R., Moni, M. A., Islam, M. M., Rashed-Al-Mahfuz, M., Islam, M. S., Hasan, M. K., et al. (2021). Emotion recognition from EEG signal focusing on deep learning and shallow learning techniques. *IEEE Access* 9, 94601–94624. doi: 10.1109/ACCESS.2021.3091487

Jiang, X., Bian, G.-B., and Tian, Z. (2019). Removal of artifacts from EEG signals: a review. *Sensors* 19:987. doi: 10.3390/s19050987

Jiang, Y., Li, W., Hossain, M. S., Chen, M., Alelaiwi, A., and Al-Hammadi, M. (2020). A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition. *Inform. Fus.* 53, 209–221. doi: 10.1016/j.inffus.2019.06.019

Kanjo, E., Younis, E. M., and Ang, C. S. (2019). Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection. *Inform. Fus*. 49, 46–56. doi: 10.1016/j.inffus.2018.09.001

Keelawat, P., Thammasan, N., Kijsirikul, B., and Numao, M. (2019). "Subject-independent emotion recognition during music listening based on EEG using deep convolutional neural networks," in *2019 IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA)* (Penang: IEEE), 21–26. doi: 10.1109/CSPA.2019.8696054

Khalfallah, J., and Slama, J. B. H. (2015). Facial expression recognition for intelligent tutoring systems in remote laboratories platform. *Proc. Comput. Sci.* 73, 274–281. doi: 10.1016/j.procs.2015.12.030

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv preprint* arXiv:1412.6980. doi: 10.48550/arXiv.1412.6980

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., et al. (2011). Deap: a database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput*. 3, 18–31. doi: 10.1109/T-AFFC.2011.15

Koelstra, S., and Patras, I. (2013). Fusion of facial expressions and EEG for implicit affective tagging. *Image Vis. Comput*. 31, 164–174. doi: 10.1016/j.imavis.2012.10.002

Kołodziej, M., Tarnowski, P., Majkowski, A., and Rak, R. (2019). Electrodermal activity measurements for detection of emotional arousal. *Bull. Pol. Acad. Sci.*. 67:4. doi: 10.24425/bpasts.2019.130190

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: a review. *Biol. Psychol*. 84, 394–421. doi: 10.1016/j.biopsycho.2010.03.010

Kwon, Y.-H., Shin, S.-B., and Kim, S.-D. (2018). Electroencephalography based fusion two-dimensional (2d)-convolution neural networks (CNN) model for emotion recognition system. *Sensors* 18:1383. doi: 10.3390/s18051383

Lan, Z., Sourina, O., Wang, L., Scherer, R., and Müller-Putz, G. R. (2018). Domain adaptation techniques for EEG-based emotion recognition: a comparative study on two public datasets. *IEEE Trans. Cogn. Dev. Syst*. 11, 85–94. doi: 10.1109/TCDS.2018.2826840

Lee, M. S., Lee, Y. K., Pae, D. S., Lim, M. T., Kim, D. W., and Kang, T. K. (2019). Fast emotion recognition based on single pulse PPG signal with convolutional neural network. *Appl. Sci*. 9:3355. doi: 10.3390/app9163355

Li, C., Wang, B., Zhang, S., Liu, Y., Song, R., Cheng, J., and Chen, X. (2022). Emotion recognition from EEG based on multi-task learning with capsule network and attention mechanism. *Comput. Biol. Med*. 143:105303. doi: 10.1016/j.compbiomed.2022.105303

Li, C., Zhang, Z., Song, R., Cheng, J., Liu, Y., and Chen, X. (2021). EEG-based emotion recognition *via* neural architecture search. *IEEE Trans. Affect. Comput*. doi: 10.1109/TAFFC.2021.3130387

Li, S., and Deng, W. (2020). Deep facial expression recognition: a survey. *IEEE Trans. Affect. Comput*. doi: 10.1109/TAFFC.2020.2981446

Li, X., Pfister, T., Huang, X., Zhao, G., and Pietikäinen, M. (2013). "A spontaneous micro-expression database: inducement, collection and baseline," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (Shanghai: IEEE), 1–6. doi: 10.1109/FG.2013.6553717

Li, X., Song, D., Zhang, P., Zhang, Y., Hou, Y., and Hu, B. (2018). Exploring EEG features in cross-subject emotion recognition. *Front. Neurosci*. 12:162. doi: 10.3389/fnins.2018.00162

Lichtenstein, A., Oehme, A., Kupschick, S., and Jürgensohn, T. (2008). "Comparing two emotion models for deriving affective states from physiological data," in *Affect and Emotion in Human-Computer Interaction*, eds P. Christian and B. Russell (Berlin; Heidelberg: Springer), 35–50. doi: 10.1007/978-3-540-85099-1_4

Ling, C. X., and Sheng, V. S. (2008). Cost-sensitive learning and the class imbalance problem. *Encyclop. Mach. Learn*. 2011, 231–235.

Liong, S.-T., See, J., Wong, K., Le Ngo, A. C., Oh, Y.-H., and Phan, R. (2015). "Automatic apex frame spotting in micro-expression database," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)* (Kuala Lumpur: IEEE), 665–669. doi: 10.1109/ACPR.2015.7486586

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). "The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops* (San Francisco, CA: IEEE), 94–101. doi: 10.1109/CVPRW.2010.5543262

Ma, J., Tang, H., Zheng, W.-L., and Lu, B.-L. (2019). "Emotion recognition using multimodal residual LSTM network," in *Proceedings of the 27th ACM International Conference on Multimedia* (Nice), 176–183. doi: 10.1145/3343031.3350871

Maia, C. L. B., and Furtado, E. S. (2019). An approach to analyze user's emotion in HCI experiments using psychophysiological measures. *IEEE Access* 7, 36471–36480. doi: 10.1109/ACCESS.2019.2904977

Marcos-Pablos, S., Gonzàlez-Pablos, E., Martín-Lorenzo, C., Flores, L. A., Gòmez-García-Bermejo, J., and Zalama, E. (2016). Corrigendum: virtual avatar for emotion recognition in patients with schizophrenia: a pilot study. *Front. Hum. Neurosci*. 10:554. doi: 10.3389/fnhum.2016.00554

Maria, E., Matthias, L., and Sten, H. (2019). Emotion recognition from physiological signal analysis: a review. *Electron. Notes Theor. Comput. Sci*. 343, 35–55. doi: 10.1016/j.entcs.2019.04.009

McKeown, G., Valstar, M., Cowie, R., Pantic, M., and Schroder, M. (2011). The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Trans. Affect. Comput*. 3, 5–17. doi: 10.1109/T-AFFC.2011.20

Oh, Y.-H., See, J., Le Ngo, A. C., Phan, R. C.-W., and Baskaran, V. M. (2018). A survey of automatic facial micro-expression analysis: databases, methods, and challenges. *Front. Psychol*. 9:1128. doi: 10.3389/fpsyg.2018.01128

Pan, H., Xie, L., Wang, Z., Liu, B., Yang, M., and Tao, J. (2021). Review of micro-expression spotting and recognition in video sequences. *Virt. Real. Intell. Hardw*. 3, 1–17. doi: 10.1016/j.vrih.2020.10.003

Pandey, P., and Seeja, K. (2019a). "Emotional state recognition with EEG signals using subject independent approach," in *Data Science and Big Data Analytics*, eds M. Durgesh Kumar, Y. Xin-She, and U. Aynur (Singapore: Springer), 117–124. doi: 10.1007/978-981-10-7641-1_10

Pandey, P., and Seeja, K. (2019b). Subject independent emotion recognition from EEG using VMD and deep learning. *J. King Saud Univ. Comput. Inform. Sci.* 34, 1730–1738. doi: 10.1016/j.jksuci.2019.11.003

Pantic, M., Valstar, M., Rademaker, R., and Maat, L. (2005). "Web-based database for facial expression analysis," in *2005 IEEE International Conference on Multimedia and Expo* (Amsterdam: IEEE), 5. doi: 10.1109/ICME.2005.1521424

Perez-Rosero, M. S., Rezaei, B., Akcakaya, M., and Ostadabbas, S. (2017). "Decoding emotional experiences through physiological signal processing," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 881–885. doi: 10.1109/ICASSP.2017.7952282

Pfister, T., Li, X., Zhao, G., and Pietikäinen, M. (2011). "Recognising spontaneous facial micro-expressions," in *2011 International Conference on Computer Vision* (Barcelona: IEEE), 1449–1456. doi: 10.1109/ICCV.2011.6126401

Piumsomboon, T., Lee, Y., Lee, G. A., Dey, A., and Billinghurst, M. (2017). "Empathic mixed reality: sharing what you feel and interacting with what you see," in *2017 International Symposium on Ubiquitous Virtual Reality (ISUVR)*, 38–41. doi: 10.1109/ISUVR.2017.20

Posner, J., Russell, J. A., and Peterson, B. S. (2005). The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol*. 17, 715–734. doi: 10.1017/S0954579405050340

Qu, F., Wang, S.-J., Yan, W.-J., and Fu, X. (2016). "CAS (me) 2: a database of spontaneous macro-expressions and micro-expressions," in *International Conference on Human-Computer Interaction* (Toronto, ON: Springer), 48–59. doi: 10.1007/978-3-319-39513-5_5

Quigley, K. S., Lindquist, K. A., and Barrett, L. F. (2014). "Inducing and measuring emotion and affect: Tips, tricks, and secrets," in *Handbook of Research Methods in Social and Personality Psychology*, eds H. T. Reis and C. M. Judd (Cambridge University Press), 220–252.

Raheel, A., Majid, M., and Anwar, S. M. (2019). "Facial expression recognition based on electroence-phalography," in *2019 2nd International Conference on*

Computing, Mathematics and Engineering Technologies (iCoMET) (Sukkur: IEEE), 1–5. doi: 10.1109/ICOMET.2019.8673408

Rayatdoost, S., and Soleymani, M. (2018). "Cross-corpus EEG-based emotion recognition," in *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)* (Aalborg: IEEE), 1–6. doi: 10.1109/MLSP.2018.8517037

Reddy, S. P. T., Karri, S. T., Dubey, S. R., and Mukherjee, S. (2019). "Spontaneous facial micro-expression recognition using 3d spatiotemporal convolutional neural networks," in *2019 International Joint Conference on Neural Networks (IJCNN)*, 1–8. doi: 10.1109/IJCNN.2019.8852419

Redmon, J., and Farhadi, A. (2018). Yolov3: an incremental improvement. *arXiv preprint* arXiv:1804.02767. doi: 10.48550/arXiv.1804.02767

Roberts, N. A., Tsai, J. L., and Coan, J. A. (2007). "Emotion elicitation using dyadic interaction tasks," in Handbook of Emotion Elicitation and Assessment, eds J. A. Coan and J. B. Allen (New York, NY: Oxford University Press), 106–123.

Samadiani, N., Huang, G., Cai, B., Luo, W., Chi, C.-H., Xiang, Y., et al. (2019). A review on automatic facial expression recognition systems assisted by multimodal sensor data. *Sensors* 19:1863. doi: 10.3390/s19081863

Schachner, T., Keller, R., and Von Wangenheim, F. (2020). Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. *J. Med. Intern. Res.* 22:e20701. doi: 10.2196/20701

Schaefer, A., Nils, F., Sanchez, X., and Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: a new tool for emotion researchers. *Cogn. Emot.* 24, 1153–1172. doi: 10.1080/02699930903274322

Setyohadi, D. B., Kusrohmaniah, S., Gunawan, S. B., Pranowo, P., and Prabuwono, A. S. (2018). Galvanic skin response data classification for emotion detection. *Int. J. Electr. Comput. Eng.* 8:4004. doi: 10.11591/ijece.v8i5.pp4004-4014

Sharma, R., Pachori, R. B., and Sircar, P. (2020). Automated emotion recognition based on higher order statistics and deep learning algorithm. *Biomed. Signal Process. Control* 58:101867. doi: 10.1016/j.bspc.2020.101867

Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., et al. (2018). A review of emotion recognition using physiological signals. *Sensors* 18:2074. doi: 10.3390/s18072074

Siedlecka, E., and Denson, T. F. (2019). Experimental methods for inducing basic emotions: a qualitative review. *Emot. Rev.* 11, 87–97. doi: 10.1177/1754073917749016

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint* arXiv:1409.1556. doi: 10.48550/arXiv.1409.1556

Soleymani, M., Asghari-Esfeden, S., Fu, Y., and Pantic, M. (2015). Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Trans. Affect. Comput.* 7, 17–28. doi: 10.1109/TAFFC.2015.2436926

Soleymani, M., Lichtenauer, J., Pun, T., and Pantic, M. (2011). A multimodal database for affect recognition and implicit tagging. *IEEE Trans. Affect. Comput.* 3, 42–55. doi: 10.1109/T-AFFC.2011.25

Staudemeyer, R. C., and Morris, E. R. (2019). Understanding LSTM-a tutorial into long short-term memory recurrent neural networks. *arXiv preprint* arXiv:1909.09586. doi: 10.48550/arXiv.1909.09586

Sun, Y., Ayaz, H., and Akansu, A. N. (2020). Multimodal affective state assessment using fNIRS+ EEG and spontaneous facial expression. *Brain Sci.* 10:85. doi: 10.3390/brainsci10020085

Sun, Y., Wong, A. K., and Kamel, M. S. (2009). Classification of imbalanced data: a review. *Int. J. Pattern Recogn. Artif. Intell.* 23, 687–719. doi: 10.1142/S0218001409007326

Takalkar, M., Xu, M., Wu, Q., and Chaczko, Z. (2018). A survey: facial micro-expression recognition. *Multim. Tools Appl.* 77, 19301–19325. doi: 10.1007/s11042-017-5317-2

Tarnowski, P., Kołodziej, M., Majkowski, A., and Rak, R. J. (2018). "Combined analysis of GSR and EEG signals for emotion recognition," in *2018 International Interdisciplinary PhD Workshop (IIPhDW)* (Świnoujście: IEEE), 137–141. doi: 10.1109/IIPHDW.2018.8388342

Teplan, M., et al. (2002). Fundamentals of EEG measurement. *Measure. Sci. Rev.* 2, 1–11.

Topic, A., and Russo, M. (2021). Emotion recognition based on EEG feature maps through deep learning network. *Eng. Sci. Technol. Int. J.* 24, 1442–1454. doi: 10.1016/j.jestch.2021.03.012

Tran, T.-K., Vo, Q.-N., Hong, X., Li, X., and Zhao, G. (2020). Micro-expression spotting: a new benchmark. *arXiv preprint arXiv:2007.12421*. doi: 10.1016/j.neucom.2021.02.022

Udovičić, G., Derek, J., Russo, M., and Sikora, M. (2017). "Wearable emotion recognition system based on GSR and PPG signals," in *Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care* (Mountain View, CA), 53–59. doi: 10.1145/3132635.3132641

Uhrig, M. K., Trautmann, N., Baumgärtner, U., Treede, R.-D., Henrich, F., Hiller, W., et al. (2016). Emotion elicitation: a comparison of pictures and films. *Front. Psychol.* 7:180. doi: 10.3389/fpsyg.2016.00180

Van Quang, N., Chun, J., and Tokuyama, T. (2019). "Capsulenet for micro-expression recognition," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, 1–7. doi: 10.1109/FG.2019.8756544

Verma, G. K., and Tiwary, U. S. (2014). Multimodal fusion framework: a multiresolution approach for emotion classification and recognition from physiological signals. *NeuroImage* 102, 162–172. doi: 10.1016/j.neuroimage.2013.11.007

Wagh, K. P., and Vasanth, K. (2019). "Electroencephalograph (EEG) based emotion recognition system: a review," in *Innovations in Electronics and Communication Engineering* (Singapore: Springer), 37–59. doi: 10.1007/978-981-10-8204-7_5

Weber, R., Li, J., Soladie, C., and Seguier, R. (2018). "A survey on databases of facial macro-expression and micro-expression," in *International Joint Conference on Computer Vision, Imaging and Computer Graphics* (Funchal-Madeira: Springer), 298–325.

Wei, W., Jia, Q., Feng, Y., and Chen, G. (2018). Emotion recognition based on weighted fusion strategy of multichannel physiological signals. *Comput. Intell. Neurosci.* 2018:5296523. doi: 10.1155/2018/5296523

Wioleta, S. (2013). "Using physiological signals for emotion recognition," in *2013 6th International Conference on Human System Interactions (HSI)* (Sopot: IEEE), 556–561. doi: 10.1109/HSI.2013.6577880

Xia, Z., Hong, X., Gao, X., Feng, X., and Zhao, G. (2019). Spatiotemporal recurrent convolutional networks for recognizing spontaneous micro-expressions. *IEEE Trans. Multim.* 22, 626–640. doi: 10.1109/TMM.2019.2931351

Xia, Z., Peng, W., Khor, H.-Q., Feng, X., and Zhao, G. (2020). Revealing the invisible with model and data shrinking for composite-database micro-expression recognition. *IEEE Trans. Image Process.* 29, 8590–8605. doi: 10.1109/TIP.2020.3018222

Yan, W.-J., Li, X., Wang, S.-J., Zhao, G., Liu, Y.-J., Chen, Y.-H., et al. (2014). CASME II: an improved spontaneous micro-expression database and the baseline evaluation. *PLoS ONE* 9:e86041. doi: 10.1371/journal.pone.0086041

Yan, W.-J., Wu, Q., Liang, J., Chen, Y.-H., and Fu, X. (2013). How fast are the leaked facial expressions: the duration of micro-expressions. *J. Nonverb. Behav.* 37, 217–230. doi: 10.1007/s10919-013-0159-8

Yang, S., Luo, P., Loy, C.-C., and Tang, X. (2016). "Wider face: a face detection benchmark," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV: IEEE), 5525–5533. doi: 10.1109/CVPR.2016.596

Yang, W., Rifqi, M., Marsala, C., and Pinna, A. (2018). "Physiological-based emotion detection and recognition in a video game context," in *2018 International Joint Conference on Neural Networks (IJCNN)* (Rio de Janeiro: IEEE), 1–8. doi: 10.1109/IJCNN.2018.8489125

Yazdani, A., Lee, J.-S., Vesin, J.-M., and Ebrahimi, T. (2012). Affect recognition based on physiological changes during the watching of music videos. *ACM Trans. Interact. Intell. Syst.* 2:7. doi: 10.1145/2133366.2133373

Zeng, Z., Pantic, M., Roisman, G. I., and Huang, T. S. (2008). A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 39–58. doi: 10.1109/TPAMI.2008.52

Zepf, S., Hernandez, J., Schmitt, A., Minker, W., and Picard, R. W. (2020). Driver emotion recognition for intelligent vehicles: a survey. *ACM Comput. Surveys* 53, 1–30. doi: 10.1145/3388790

Zhao, G., and Li, X. (2019). Automatic micro-expression analysis: open challenges. *Front. Psychol.* 2019:1833. doi: 10.3389/fpsyg.2019.01833

Zheng, W.-L., Liu, W., Lu, Y., Lu, B.-L., and Cichocki, A. (2018). Emotionmeter: a multimodal framework for recognizing human emotions. *IEEE Trans. Cybern.* 49, 1110–1122. doi: 10.1109/TCYB.2018.2797176

Zhu, Q., Lu, G., and Yan, J. (2020). "Valence-arousal model based emotion recognition using EEG, peri-pheral physiological signals and facial expression," in *Proceedings of the 4th International Conference on Machine Learning and Soft Computing* (Haiphong City: ACM), 81–85. doi: 10.1145/3380688.3380694

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Saffaryazdi, Wasim, Dileep, Nia, Nanayakkara, Broadbent and Billinghurst. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 26 July 2022
doi: 10.3389/fpsyg.2022.819855

An Emotional Analysis Method for the Analysis of Cognitive and Psychological Factors in the Change of Second Language Learning Model of Chinese Mainland Students in the Post-epidemic Era

Gang Xie[1]* *and Xiaona Wang*[2]*

[1] School of Foreign Language Education, Jilin University, Changchun, China, [2] School of Arts, Agricultural Technology of Jilin, Jilin City, China

Since the sudden outbreak of the coronavirus disease 2019 (COVID-19) epidemic in 2020, the second language learning patterns of students in mainland China have encountered new challenges that have had a psychological impact on mainland Chinese students. The epidemic has not only inconvenienced students' normal second language learning but also greatly affected the second language learning patterns of mainland Chinese students. In the post-epidemic era, more and more students are becoming accustomed to studying and learning a second language online. The level of informatization of second language learning patterns of students in mainland China has increased significantly. This study first analyses the mechanisms of change in second language learning patterns and further analyses the influence of knowledge background on the perception of second language learning patterns on this basis. To design the influencing factors of second language learning patterns, a questionnaire was used to investigate the influence of knowledge background on the perception of second language learning patterns. The survey was conducted on students who were learning a second language in mainland China. Then, the survey data were statistically analyzed. In analyzing the influence of effect on second language learning behaviors of students in mainland China, observed variables were designed, including observed variables of affective factors and learning behaviors. After that, the findings of the experiment were summarized based on the results of the questionnaire survey, and the positive influence of emotional factors on second language learning behaviors of mainland Chinese students in the post-development era was concluded.

Keywords: post-epidemic era, second language learning model, change of learning model, cognitive and psychological factors, an emotional analysis method, statistical analysis

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Runmin Liu,
Wuhan Sports University, China
Tongguang Ni,
Changzhou University, China

***Correspondence:**
Gang Xie
xie_gang@jlu.edu.cn
Xiaona Wang
iamnanaanna@sina.com

Specialty section:
This article was submitted to Emotion Science, a section of the journal Frontiers in Psychology

Received: *22 November 2021*
Accepted: *19 April 2022*
Published: *26 July 2022*

Citation:
Xie G and Wang X (2022) An Emotional Analysis Method for the Analysis of Cognitive and Psychological Factors in the Change of Second Language Learning Model of Chinese Mainland Students in the Post-epidemic Era. Front. Psychol. 13:819855. doi: 10.3389/fpsyg.2022.819855

INTRODUCTION

While the new crown virus epidemic has caused many inconveniences to school teaching, it has also brought profound changes and impacts to the second language learning model of mainland Chinese students (Xiuqin and Yingli, 2020; Mohsen and Mahdi, 2021). Under the influence of the epidemic, the ecology of the second language learning model for mainland Chinese students has undergone

tremendous changes: students have developed the habit of learning a second language online, teachers' traditional teaching thinking has also changed, and the development of informatization of the second language learning model for students in mainland China has also taken a big step forward. First, there have been profound changes in students' habits of learning a second language in mainland China. Almost all students have formed the habit of learning a second language online as they are forced to leave school to continue their studies at home due to the epidemic prevention and control (Calafato, 2021). This habit is reflected in the following aspects: the goal of second language learning can be achieved without relying on the carrier of the classroom, good learning results can be achieved without relying on traditional learning methods, and the goal of second language learning can be achieved through autonomous learning without relying on the supervision of teachers to a certain extent (Mohsen and Mahdi, 2021). Therefore, judging from students' performance in learning a second language, the epidemic has had a profound impact on the teaching ecology of second language learning for students in mainland China. Second, teachers' teaching thinking has undergone profound changes. During the epidemic, all learning models were basically converted to online, and all teaching activities were carried out based on modern information technology. The teaching thinking of teachers was very different from the thinking of traditional teaching methods (Barrios and Hayes-Harb, 2020; Bigelow et al., 2020). This difference is mainly manifested in the following: teaching content materials must conform to the characteristics of online teaching methods, the way of interaction between teachers and students in the process of distance teaching has changed, and the way of feedback on teaching effects has also changed (Wahbeh et al., 2021). Therefore, the teaching thinking of teachers has also changed with the changes in the teaching ecology and has shown the characteristics of the Internet. Finally, the degree of informatization of the second language learning model for mainland Chinese students has been greatly improved. Driven by the epidemic, the online learning model of the second language for students on the Chinese mainland has been popularized, greatly improving the informatization level of the second language learning. On the one hand, the informationization of the second language learning model of students in mainland China breaks the limitation of physical space teaching and provides rich learning resources and teaching methods for second language learning (Martin, 2020). On the other hand, the informatization of students' second language learning model in mainland China has greatly improved the informatization ability and professional quality of second language teachers (Hsieh, 2020; Havron, 2021). Therefore, affected by the epidemic, the degree of informatization of the second language learning model for mainland Chinese students has greatly increased, which in turn improves the overall teaching level of the second language learning for mainland Chinese students.

This article consists of the following sections: the first section is the introduction, which introduces the significance and importance of this study; the second section examines the mechanism of change in the second language learning model; the third section analyses the influence of knowledge background on the perception of the second language learning model; the fourth section analyses the influence of emotion on learning behavior, and the fifth section is the conclusion.

MECHANISM OF CHANGE IN THE SECOND LANGUAGE LEARNING MODEL

Current studies generally simplify the generation of behavior into a driving model of "need-motive-attitude-intention-behavior" (Cong and Chen, 2021; Thomas et al., 2021). Need is the reflection in the brain of an individual's objective needs of the internal and external environment, including material, psychological, and spiritual needs. Motivation is the internal power that impels an individual to engage in some kind of activity (Hou and Aryadoust, 2021), the intrinsic condition that causes motive is need, and the extrinsic condition is inducement. Attitude is people's evaluation of things based on their own morality and values. Intention is the behavioral tendency of the individual toward the object of the attitude, that is, the state of preparation for the behavior; behavior refers to the external activities of a person under the influence of subjective and objective factors and is an overall process of action (Tabata-Sandom et al., 2020).

A need arises when an individual feels a lack of something materially, psychologically, or spiritually, and a need creates motivation to satisfy it (Gong et al., 2021). However, individuals do not immediately carry out the behavior but evaluate the behavior to be carried out and form an attitude toward the behavior (Gessinger et al., 2021). Iglesias et al. (2021) believed that adding intention as an intermediary variable determines whether a behavior occurs.

When the individual realizes the need and converts the need into the motivation for learning under the stimulation of external incentives, the individual, therefore, becomes the subject of the second language learning model. Next, the changed subjects will conduct cognition and understanding of the second language learning model, evaluate the usefulness and ease of use of the second language learning model, and form the attitude of the second language learning model after obtaining the results. At the same time, through the cognition and experience of the second language learning model, the changed subjects can generate certain emotions, which are positive and negative. When the positive emotions rise to a certain height, the person being changed will have learning intentions, and the intentions will ultimately determine the occurrence of the changed behavior (Loingsigh and Mozzon-Mcpherson, 2020; Lou and Noels, 2020). After the behavior occurs, the subject will evaluate the degree of pleasure in the second language learning model, thus producing the satisfaction of the subject (Martín-Monje and Borthwick, 2021). However, the changed subject is not a completely rational decision-maker. The behavior of the changed subject is affected by their own emotions to some extent, and they have preference for the second language learning model. Only when the changed subject has preference for the second language learning model emotionally, the behavior can be maintained. The formation

mechanism of learning behavior of those whose second language learning model is changed is shown in **Figure 1**.

According to the above theoretical analysis, the formation mechanism of the learning behavior of the person whose second language learning model is changed is obtained (**Figure 1**). This article proposes the hypothesis that the emotions and preferences of the changed subjects have a positive effect on the learning behavior of the changed subjects. However, for the second language learning model, the changed subject's cognition of the second language learning model will affect the changed subject's preference and emotions for the second language learning model, thereby affecting the learning behavior of the person being changed; hence, it is necessary to understand what factors affect the changed subject's perception of the second language learning model.

The following research will first analyze the influence of the changed subject's own factors on the cognition of the second language learning model, and then construct the emotional factors and the changed subject's learning structure model and the personal preference and the changed subject's learning structure model on this basis, and then conduct a questionnaire to investigate, and finally verify the hypothesis and make recommendations.

AN ANALYSIS OF THE INFLUENCE OF KNOWLEDGE BACKGROUND ON SECOND LANGUAGE LEARNING MODEL COGNITION

Due to the different factors of the altered subjects, the altered subjects' perceptions of the second language learning model also differ, which affects the altered subjects' preferences and emotions toward the second language learning model. This study focuses on analyzing the influence of subjects with different knowledge and experience backgrounds on the perception of second language learning models, i.e., observing and analyzing the differences in the perception of the content of second language learning models by subjects with different knowledge backgrounds.

Design of Influencing Factors

In this study, the knowledge background of the change in the second language learning model of Chinese mainland students in the post-epidemic era is divided into four dimensions: professional background, academic education, retrieval knowledge, and second language learning experience (Xu, 2020); according to this, the changed subjects were grouped to analyze the differences between groups. Specific user groups were divided as follows:

Professional background is divided into science and engineering, humanities, and other three groups.
Academic education is divided into undergraduate (junior and senior students), master's students, and doctoral students.
Retrieval knowledge is divided into proficient, skilled, and general groups, according to the degree of mastery of the changed subjects.
The second language learning experience is based on the frequency of second language learning, which is divided into four groups, namely, at least once a week, once a half month, once a month, and at most once a month.

As for the determination of website cognitive content, this study divides it into three parts: pattern cognition, content cognition, and efficiency cognition. The specific contents of each part are as follows: (1) Pattern cognition includes the cognition of the function, content, and process of the second language learning model of Chinese mainland students in the post-epidemic era. (2) Content cognition includes the cognition of the information comprehensiveness, content types, updating of teaching methods, and teaching quality of second language

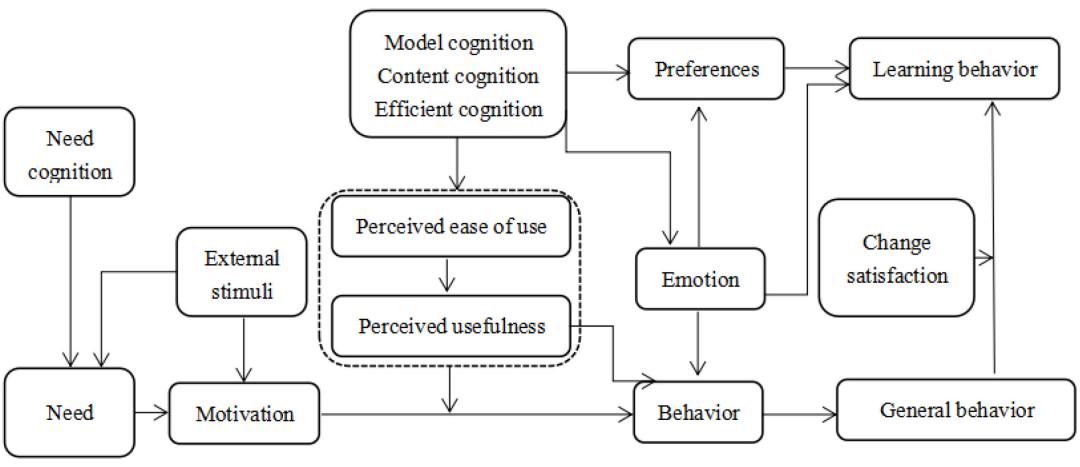


FIGURE 1 | The formation mechanism of learning behavior of those whose second language learning model is changed.

learning patterns of Chinese mainland students in the post-epidemic era. 3. The cognition of efficiency includes the cognition of the efficiency of changing the second language learning model of Chinese mainland students in the post-epidemic era.

The Questionnaire Survey

The subjects of this questionnaire survey are all mainland Chinese students who are learning a second language, mainly college students. A total of 200 questionnaires were distributed and 192 valid questionnaires were recovered. The effective questionnaire response rate was 96.00%. A summary of the basic information of the survey objects in the valid questionnaire is shown in **Table 1**.

Statistical Analysis of Data

The cognition experiment of the changed subjects mainly analyses the difference in the cognition of the change of the second language learning model of Chinese students in the post-epidemic era by the changed subjects with different second language learning knowledge experience. In this study, SPSS20.0 software (IBM, Chicago, IL, United States) is used to collect statistics on the data, and a one-way ANOVA method is used. It is generally believed that the *P*-value of the changed person variable is less than 0.1 after the cognitive univariate analysis of the change in the second language learning model of mainland Chinese students in the post-epidemic era. It shows that this variable has a significant impact on the change of cognition of the second language learning model of Chinese students in the post-epidemic era.

First, the impact of the second language learning knowledge experience of the changed subject on the cognition of the second language learning model of Chinese students in the post-epidemic era is analyzed. The analysis results are shown in **Table 2**.

According to **Table 2**, it can be seen that (1) the changed subjects of different professional backgrounds have significant differences in the basic cognition of the second language learning model and the cognitive evaluation of the role of the second language learning model and (2) there are significant differences in the cognitive evaluation of the effect of different levels of retrieval knowledge on the second language learning model.

Second, the influence of the knowledge and experience of the changed participants on the content cognition of the second language learning model is analyzed. The results are shown in **Table 3**.

According to **Table 3**, it can be seen that (1) changed subjects with different professional backgrounds have significant differences in the timeliness of the information update of the second language learning model and the cognitive evaluation of the comparison with the offline second language learning model (Meskill et al., 2020), (2) there are significant differences in the cognitive evaluation of the teaching quality of the second language learning model by the changed subjects with different second language learning experiences, and (3) there are significant differences in the cognitive evaluation of the effect of the second language learning model among those who have different levels of retrieval knowledge.

Finally, the impact of the knowledge and experience of the changed subjects on the cognitive effect of the change of the second language learning model of Chinese students in the post-epidemic era is analyzed. The results are shown in **Table 4**.

According to **Table 4**, it can be seen that (1) there are significant differences in the cognitive evaluation of the efficiency of the change in second language learning model of mainland Chinese students in the post-epidemic era among the subjects with different professional backgrounds and different second language learning experiences; (2) the subjects with different levels of retrieval knowledge had significant differences in their cognitive evaluation of the personalized learning efficiency of the change of second language learning model in the post-epidemic era; and (3) there are significant differences in cognitive

TABLE 1 | Summary of the basic information of the survey objects in the valid questionnaire.

Survey item	Description	Number of samples	Proportion (%)
Professional background	Science and engineering	102	53.13
	Humanities	56	29.17
	Other	34	17.70
Academic education	Undergraduate (junior and senior students)	130	67.71
	Master students	47	24.48
	Doctoral students	15	7.81
Retrieval knowledge	Proficient	43	22.40
	Skilled	69	35.94
	General	80	41.66
The second language learning experience	At least once a week	68	35.42
	Once a half month	62	32.29
	Once a month	48	25.00
	At most once a month	14	7.29

TABLE 2 | Variance test of the impact of second language learning knowledge and experience on the cognitive impact of the second language learning model of Chinese students in the post-epidemic era.

Individual characteristic variables	Basic understanding of the second language learning model	The role of second language learning model for cognition	Perception of acceptance of second language learning model
Professional background	0.002	0.004	0.413
Academic education	0.318	0.523	0.732
Retrieval knowledge	0.926	0.027	0.527
The second language learning experience	0.957	0.186	0.134

TABLE 3 | Test of variance of score of influence of knowledge experience on content cognition of second language learning model.

Individual characteristic variables	Comprehensiveness of content	Second language learning quality cognition	Real-time cognition of content update	Contrast cognition with offline second language learning models
Professional background	0.185	0.005	0.165	0.115
Academic education	0.563	0.415	0.530	0.862
Retrieval knowledge	0.421	0.352	0.481	0.034
The second language learning experience	0.436	0.647	0.836	0.047

evaluation of teaching effect among the subjects with different second language learning experiences (Hsieh, 2020).

ANALYSIS OF THE INFLUENCE OF EMOTION ON LEARNING BEHAVIOR

Based on the above theoretical research, the author puts forward a mechanism model of the effect of affective factors on the learning behavior of the changed subjects and puts forward hypothesis H1: Affective analysis has a positive effect on the learning behavior of the changed subjects (Warren, 2020). The main purposes of emotion analysis are as follows: (1) Analyze the hypothesis of the influence of sentimental factors on the learning behavior of the changed subject and examine the effect of emotion analysis on the learning behavior of the changed person in general (Ebadi et al., 2021); (2) Comparative analysis of emotion analysis and the weights of the observed variables of the learning behavior of the changed person.

Emotional factor is the inner psychological reflection of the tendency of people and objective things to be good or bad in the activities of people. Between people, a good emotional relationship is established, and it can make people feel affectionate. With a sense of cordiality, the mutual attraction is great, and the influence of each other is also great. Conversely, without the establishment of a good emotional relationship, it will cause a certain psychological distance between the two sides, and the psychological distance is a psychological repulsive force; confrontation will produce a negative influence. So, students must inject their own situational factors when learning a second language, but in the post-epidemic era, students have more time for online classes; hence, it is more important to adjust their emotional perceptions in time and try not to influence their learning behavior.

Design of Observed Variables

Design of Observed Variables for Emotion Factors

This study divides the emotion factors into the mood, satisfaction, and surprise of the changed subject after learning. In fact, the design and content of the second language learning model have an impact on the above-mentioned emotions of users. When the changed subject is learning a second language, the optimized second language learning model after emotion analysis can make the changed subject feel relaxed and happy in the second language learning process and enhance the second language learning experience (Gong et al., 2021). If the content and teaching effect provided by the second language learning exceed the expectation of the subject, the subject will feel surprised and be more patient in the process of second language learning, so as to ensure the effect of second language learning. Affected by the epidemic, students' demand for the informationization degree of the second language learning model has increased, and the second language learning model supported by informationization technology can break the limitation of physical space teaching, and improve the changed students' trust in the second language learning model, and significantly affect their emotions (Zheng et al., 2020).

Design of Observed Variables for Learning Behavior

In this study, the measurement items of second language learning behavior change of students in mainland China are divided into learning behavior and learning attitude. The observed variables of learning behavior were divided into percentage of second language learning and percentage of second language application. The observational variables of learning attitude were divided into learning intention, preferred future learning intention, and fault tolerance. The observed variables of emotion factors and learning behaviors are shown in **Table 5**.

TABLE 4 | Variance test of the cognitive impact of the knowledge and experience of the changed subjects on the efficiency of the change of the second language learning model of Chinese students in the post-epidemic era.

Individual characteristic variables	Perception of total efficiency	Recognition of personalized learning efficiency	Cognition of teaching effect
Professional background	0.075	0.218	0.117
Academic education	0.263	0.327	0.962
Retrieval knowledge	0.561	0.036	0.132
The second language learning experience	0.051	0.401	0.034

TABLE 5 | The observed variables of emotion factors and learning behaviors.

Structure variables	Observation variable
Emotion factors	Change experience mood
	Change satisfaction
	Change the degree of surprise
The changed subject's learning behavior	The number of sessions in a second language learning class
	Percentage of study time
	The ability to learn
	Study notes
	Learning effect

The conceptual model of affective factors influencing the learning behavior of the changed students is finally formed, as shown in **Figure 2**.

The Questionnaire Survey

This questionnaire is divided into three parts. The first part is the basic information of the interviewees. The second part is an investigation of the learning behavior of the people whose learning model is changed, which includes learning behavior and learning attitude. Learning behavior is measured by the frequency of learning, percentage of learning time, and learning ability in second language learning classes. Learning attitude is measured by learning experience, learning effect, and other indicators (Ez-Zaouia et al., 2020; Zheng and Tian, 2020). The third part is the measurement of emotional factors, including the changed mood of the subject, the degree of satisfaction with the change of the second language learning model of mainland Chinese students in the post-epidemic era, and the degree of surprise to the change of the second language learning model of mainland Chinese students in the post-epidemic era.

The respondents of this questionnaire are all Chinese mainland students, mainly college students, who are learning a second language. A total of 150 questionnaires were distributed, and 144 valid questionnaires were recovered.

TABLE 6 | The reliability and validity test results of structural variables.

Structure variables	Number of observed variables	AVE	Cronbach's α
Emotion factors	3	0.812	0.925
The changed subject's learning behavior	5	0.827	0.913

Statistical Analysis of Data

SPSS20.0 statistical software is used to complete the statistics of basic data. The reliability and validity test results of structural variables are shown in **Table 6**.

Table 6 shows that Cronbach's α values reach 0.925 and 0.913, which means that the questions show good internal consistency. In addition, the AVE value is significantly higher than the threshold value of 0.7. It shows that latent variables have strong convergent validity. It can be seen that the measure items of the questionnaire have high reliability. The weight summary table of observed variables is shown in **Table 7**.

From **Table 7**, it can be known that the weights of the observed variables of the affective factors are not much different, and they are all around 0.320. Among them, the weight of the emotion of the changed subject for the second language learning experience is relatively high, which shows that the emotion of the changed subject for the second language learning model change experience has a considerable influence on the cognitive effect of the second language learning model change. The weight of learning ability is more important than the weight of learning experience. In the attitude part, the weight of learning ability is the highest, while the weight of learning experience is the lowest.

Experimental Conclusion

Based on the above data analysis, the following conclusions can be drawn from the affective analysis. The hypothesis that affective factors have a positive influence on the second language

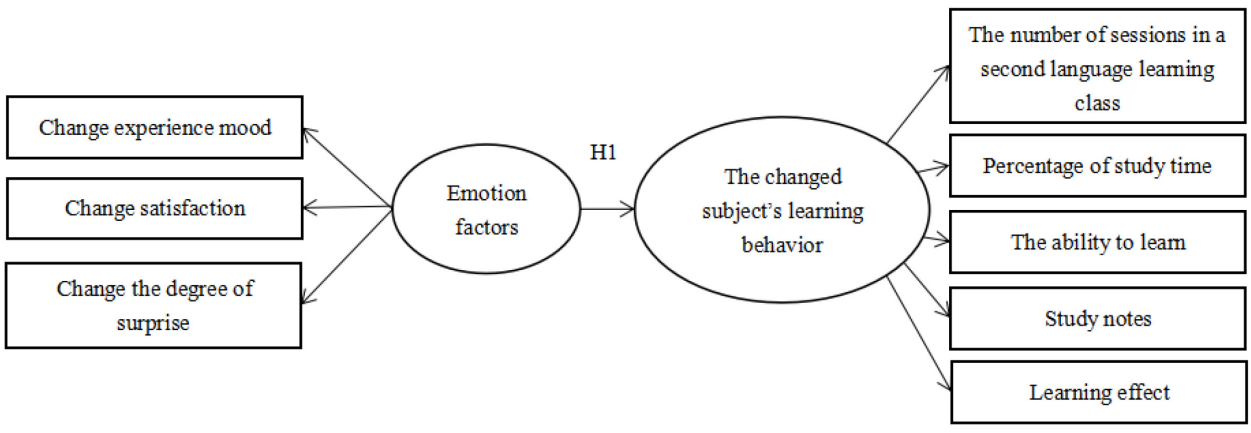


FIGURE 2 | The conceptual model of affective factors influencing the learning behavior of the changed students.

TABLE 7 | The weight summary table of observed variables.

Structure variables	Observation variable	Interpretation of observed variables	Weight of observed variable
Emotion factors	Change experience mood	The mood of the subject after experiencing the change of the second language learning model	0.326
	Change satisfaction	The satisfaction of the changed subjects after experiencing the change of second language learning model	0.354
	Change the degree of surprise	How surprised they are after experiencing the change in their second language learning model	0.320
The changed subject's learning behavior	The number of sessions in a second language learning class	The number of times the subject studied in a second language learning class	0.218
	Percentage of study time	Percentage of second language learning in total learning time	0.179
	The ability to learn	Changes in second language learning ability	0.282
	Study notes	Changes in second language learning experience	0.114
	Learning effect	Changes in the effect of second language learning	0.207

learning behavior of students in mainland China in the post-emergency era is verified, and affective factors have a significant positive influence on the learning effect; among the affective factors, experiencing mood has a greater influence on the effect of the changed students, which indicates that the change in the second language learning model of students in mainland China in the post-emergency era has a great influence on the mood of the changed students in the process of learning the second language. Therefore, our suggestion is to let students have a correct view of learning, whether in online classes or in classrooms. First, we teachers should observe students' emotional changes more before the class and actively make good emotional guidance, and second, if students are at home by themselves, then teachers should communicate more with students to eliminate students' loneliness. The last thing is to make students' homework not too much.

CONCLUSION

Understanding the cognitive and psychological processes of second language learners is helpful for teachers to adjust foreign language teaching methods and improve the efficiency of foreign language learning (Silvia, 2020). In the post-epidemic era, computer technology provides a good means for second language teachers to develop more attractive software and apply it in teaching (Lochlainn et al., 2021). The online learning model created by online education platforms and software in the post-epidemic era has exerted a subtle influence on students. Online learning is gradually attracting mainland Chinese students with its strong atmosphere of the times, convenient implementation methods, and massive knowledge treasures (Zeng et al., 2020). This second language learning model may play a greater role in the future learning of mainland Chinese students. In order to help students achieve better online learning of the second language learning, teachers still need to continue to study online teaching method to improve student's comprehensive ability and strive to increase online interaction and teaching platform for the school, to help students form a good habit of learning a second language, and at the same time, the students also need to continuously strengthen their self-cultivation to improve the consciousness of learning (Obermeier and Elgort, 2020).

In the post-epidemic era, the knowledge of mainland Chinese students has a significant impact on the second language learning model's model cognition and content cognition, while the knowledge and experience of the changed person have a significant impact on the second language learning model's service cognition. Therefore, strengthening the cognition of the changed subject's second language learning model can promote the changed subject's learning behavior. Emotion factors have a significant impact on the learning behavior of the changed second language learning model, among which the mood of the changed second language learning experience has a greater impact on loyalty. When the changed subject has positive feelings toward the second language learning model, the willingness of the changed person to choose the second language learning model next time can be improved. Personal preference has a significant impact on the learning behavior of the changed learners. The convenience of the interface design of the changed learners and the variety of the content of the changed learners have a greater impact on their learning behavior. Based on the above research conclusions, the second language learning model for students in mainland China in the post-epidemic era can improve the learning behavior of the changed learners from the following aspects.

In the future, we should focus more on students' behavioral cognition and affective cognition because we know from the above studies that these two cognitions can influence students' learning behaviors, and we can also add more factors in future designs so that we can evaluate students' learning behaviors more completely, regardless of future epidemics.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Jilin University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

GX collected and analyzed data. XW contributed to research and edited the article. Both authors contributed to the article and approved the submitted version.

REFERENCES

Barrios, S. L., and Hayes-Harb, R. (2020). Second language learning of phonological alternations with and without orthographic input: evidence from the acquisition of a German-like voicing alternation. *Appl. Psychol.* 41, 1–29.

Bigelow, K. M., Walker, D., Jia, F., Irvin, D., and Turcotte, A. (2020). Text messaging as an enhancement to home visiting: building parents' capacity to improve child language-learning environments. *Early Child. Res. Q.* 51, 416–429. doi: 10.1016/j.ecresq.2019.12.010

Calafato, R. (2021). "I'm a salesman and my client is China": language learning motivation, multicultural attitudes, and multilingualism among university students in Kazakhstan and Uzbekistan. *System* 103:102645–.

Cong, F., and Chen, B. (2021). The processing mechanism of morphologically complex words in second language learners. *Adv. Psychol. Sci.* 29, 12–13.

Ebadi, A., Xi, P., Tremblay, S., Spencer, B., Pall, R., Wong, A., et al. (2021). Understanding the temporal evolution of COVID-19 research through machine learning and natural language processing. *Scientometrics* 126, 725-739. doi: 10.1007/s11192-020-03744-7

Ez-Zaouia, M., Tabard, A., and Lavoué, E. (2020). Emodash:a dashboard supporting retrospective awareness of emotions in online learning. *Int. J. Hum. Comput. Stud.* 139, 102411–102412. doi: 10.1016/j.ijhcs.2020.102411

Gessinger, I., Mbius, B., Maguer, S. L., Raveh, E., and Steiner, I. (2021). Phonetic accommodation in interaction with a virtual language learning tutor: a wizard-of-Oz study. *J. Phon.* 86, 101029–101030. doi: 10.1016/j.wocn.2021.101029

Gong, Y., Guo, Q., Li, M., Lai, C., and Wang, C. (2021). Developing literacy or focusing on interaction: New Zealand students' strategic efforts related to Chinese language learning during study abroad in China. *System* 98:102462. doi: 10.1016/j.system.2021.102462

Havron, N. I. (2021). Starting Big: the effect of unit size on language learning in children and adults. *J. Child Lang.* 48, 244-260 doi: 10.1017/S0305000920000264

Hou, Z., and Aryadoust, V. (2021). A review of the methodological quality of quantitative mobile-assisted language learning research. *System* 2021:102568–102569.

Hsieh, Y. C. (2020). Learner interactions in face-to-face collaborative writing with the support of online resources. *ReCall* 32, 85–105. doi: 10.1017/s0958344019000120

Iglesias, A., Luo, R., Pace, A., Golinkoff, R. M., Levine, D. F., Wilson, M. S., et al. (2021) Home literacy environment and existing knowledge mediate the link between socioeconomic status and language learning in dual language learners. *Early Child. Res. Q.* 55, 1-14.

Lochlainn, C. M., Ming, M and Beirne, E. (2021). Clicking, but connecting. L2 learning engagement on an ab initio Irish language lmooc. *Recall* 33, 111–127.*ed doi: 10.1017/s0958344021000100

Loingsigh, D. N., and Mozzon-Mcpherson, M. (2020). Advising in language learning in a new speaker context: facilitating linguistic shifts. *System* 95, 102363–102364. doi: 10.1016/j.system.2020.102363

Lou, N. M., and Noels, K. A. (2020). Mindsets about language learning and support for immigrants' integration. *Int. J. Int. Relat.* 79, 46-57.

Martin, I. A. (2020). Pronunciation development and instruction in distance language learning. *Lang. Learn. Technol.* 24, 86-106.

Martín-Monje, E., and Borthwick, K. (2021). Researching massive open online courses for language teaching and learning. *ReCALL* 33, 107–110. doi: 10.1017/s0958344021000094

Meskill, C., Anthony, N., and Sadykova, G. (2020). Teaching languages online: professional vision in the making. *Lang. Learn. Technol.* 24, 160–175.

Mohsen, M. A., and Mahdi, H. S. (2021). Partial versus full captioning mode to improve L2 vocabulary acquisition in a mobile-assisted language learning setting: words pronunciation domain. *J. Comput. High. Educ.* 33, 524–543.

Obermeier, A., and Elgort, I. (2020). Deliberate and contextual learning of L2 idioms: the effect of learning conditions on online processing. *System* 97:102428.

Silvia, B. (2020). Extending digital literacies: proposing an agentive literacy to tackle the problems of distractive technologies in language learning. *ReCALL* 32, 250-271. doi: 10.1016/j.system.2020.102428

Tabata-Sandom, M., Nishikawa, Y., and Ishii, D. (2020). Metaphorical conceptualizations of language learning by post-tertiary learners of Japanese. *System* 94, 102335–102336. doi: 10.1016/j.system.2020.102335

Thomas, N., Bowen, N., and Rose, H. (2021). A diachronic analysis of explicit definitions and implicit conceptualizations of language learning strategies. *System* 103, 102619–102620. doi: 10.1016/j.system.2021.102619

Wahbeh, D. G., Najjar, E. A., and Sartawi, A. F. Abuzant, M., Daher, W. (2021). The role of project-based language learning in developing students'. *Life Skills. Sustain.* 13:6518. doi: 10.3390/su13126518

Warren, A. N. (2020). Language teachers' narratives of professional experience in online class discussions. *Text Talk* 40, 399–419. doi: 10.1515/text-2020-2063

Xiuqin, Z., and Yingli, W. (2020). The effect of language level on learners' output and synergistic effect in the reading follow-up speaking task. *J. PLA Foreign Lang. Inst.* 43, 9–10. doi: 10.37546/jalttlt41.2-2

Xu, P. (2020). Reform of online english teaching model from the perspective of second language acquisition theory. *Eng. Square* 36, 3–4.

Zeng, G., Wang, Y., and Tan, X. (2020). The influence of two peer feedback modes on oral english output in mobile assisted language learning environment. *Foreign Lang. Foreign Lang. Teach.* 10, 14–15.

Zheng, B., Lin, C. H., and Kwon, J. B. (2020). The impact of learner-, instructor-, and course-level factors on online learning. *Comput. Educ.* 150:103851 doi: 10.1016/j.compedu.2020.103851

Zheng, D., and Tian, J. (2020). Interactions between ecology, conversational and distributed horizons and virtual environments and second language learning. *Foreign Lang. China* 24, 9–10.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Xie and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

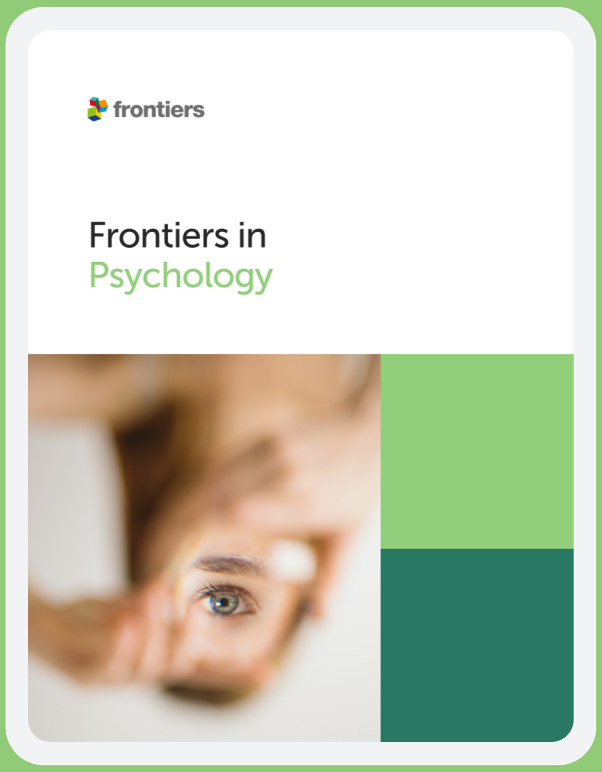
Frontiers in Psychology

Paving the way for a greater understanding of human behavior

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact