

## OPEN ACCESS

## EDITED BY

Umar Khokhar,  
Georgia Gwinnett College, United States

## REVIEWED BY

Shahid Latif,  
University of the West of England,  
United Kingdom  
Muhammad A. Khan,  
Korea University, Republic of Korea

## \*CORRESPONDENCE

Poh Ping Em  
✉ ppm@mmu.edu.my

RECEIVED 13 October 2025

REVISED 01 January 2026

ACCEPTED 12 January 2026

PUBLISHED 30 January 2026

## CITATION

Ahmad K, Em PP and Ab Aziz NA (2026) Security of drivers in intelligent transportation systems: privacy-preserving federated transfer learning for driver drowsiness detection. *Front. Comput. Sci.* 8:1723711. doi: 10.3389/fcomp.2026.1723711

## COPYRIGHT

© 2026 Ahmad, Em and Ab Aziz. 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.

# Security of drivers in intelligent transportation systems: privacy-preserving federated transfer learning for driver drowsiness detection

Khubab Ahmad, Poh Ping Em\* and Nor Azlina Ab Aziz

Faculty of Engineering and Technology, Multimedia University Malacca, Melaka (Malacca), Malaysia

Driver drowsiness is a serious concern for road safety within intelligent transportation systems, and it can undermine the safety and dependability of critical transport infrastructure. As modern vehicles become more connected and data-focused, centralized learning systems that share driver and vehicle information can expose private details and raise privacy and security concerns. This study presents a privacy-preserving framework that enables secure learning among multiple vehicles without sharing raw data. It uses the On-Board Diagnostic-II sensor data, combined with transfer learning, to detect driver drowsiness in real time within a federated learning framework. Signals such as speed, engine revolutions, throttle position, and steering torque are extracted from cars and then converted into image representations using Mel-Frequency Cepstral Coefficients so the model can identify changes in driving behavior. These image features are used to train a pretrained ResNet50 network; this trained model can classify driver states as drowsy or normal. Each vehicle trains on its own data while the central server updates the shared model weights through a client-weighted averaging strategy that keeps learning balanced for all clients. This process keeps data private while the model trained on different driving pattern. Using client weights DrowsyXnet achieved 98.29% accuracy, which is nearly matched the centralized baseline of 98.67%. The latent feature graph showed a clear separation between drowsy and normal states, indicating that the model learns the underlying signals rather than merely incidental correlations. The proposed framework improves intelligent transportation systems while preventing leakage of private data. The use of driver drowsiness detection system into vehicles can prevent drowsiness related accidents and enhance overall road safety.

## KEYWORDS

driver drowsiness, federated learning, intelligent transportation, Mel-Frequency Cepstral Coefficients, On-Board Diagnostic-II, privacy preservation

## 1 Introduction

Transportation systems are a major part of modern infrastructure to keep societies connected by the movement of people and goods. These transportation systems still encounter challenges related to human reliability and road safety. Despite many other causes, driver drowsiness is the most common and dangerous cause of accidents. It threatens both lives and the stability of intelligent transport networks that require drivers to be focused. Every year, 1.35 million people die in road crashes, averaging 3,700 daily fatalities. Beyond the human life loss these incidents causes massive economic costs to families and nations. In 2018, the

Malaysian government figured that every road death caused a loss of around 3.12 million, according to the Value of Statistical Life (Ministry of Transport Malaysia, 2020). Figure 1 shows a graph illustrating the number of road accidents and deaths in Malaysia from 2010 to 2024. Road accidents generally increased over the years and reached a top in 2019 but decreased during the pandemic in 2020 and 2021. In 2023, both accidents and fatalities increased and deaths jumped to over 12,000. Accidents decreased in 2024 but the death rate was still alarmingly high. The sources for this data include the Ministry of Transport Malaysia, as well as The Star and Paultan.org (Ministry of Transport Malaysia, 2021; Paultan.org, 2022; The Star, 2023, 2024).

Drowsiness causes serious risks for driver response time and lapses in awareness, which increases accidents globally. This affects thousands of drivers daily including long-haul truckers so there is need for solution. Artificial intelligence offers a solution to prevent accidents by detecting drowsiness in drivers. Recognizing that sleep issues and impaired driving performance emphasize the urgency of addressing drowsiness for road safety (Khan et al., 2022). According to Sharma et al. (2021), machine learning is a subset of artificial intelligence (AI) that learns from features through algorithms. It uses statistical learning to improve detection. Conversely, deep learning relies on extensive data for learning and involves multiple layers in the neural network. Artificial intelligence aims for human-like results and uses machine learning and deep learning algorithms because it requires a lot of data to train and boost performance. Many researchers are using machine learning and deep learning to solve complex problems in various fields (Umair et al., 2021; Ahmad et al., 2023; Umair et al., 2024). Using deep learning and machine learning techniques, researchers have provided solutions for detecting and preventing accidents caused by driver drowsiness. Advancements in understanding and identifying driver drowsiness are facing ongoing challenges, despite efforts to develop more effective measurement methods. However, Lenné and Jacobs (2016) reviewed research

methods for predicting drowsiness-related driving events and discussed future opportunities for enhancing detection techniques. The authors identify the challenges of driver drowsiness detection, which hinders the development of more effective measurement methods. A survey by Arceda et al. (2020) stated that most drowsiness detection methods still have to be tested in real driving conditions. Most testing is done in simulated environments that do not reflect real road conditions. De Naurois et al. (2019) investigate the challenge of incorporating contextual information into drowsiness detection. Factors such as traffic flow and time of day strongly influence how drivers respond. Most driver drowsiness detection systems have not utilized these elements due to data limitations, privacy concerns, and the unpredictable nature of real driving. Recent artificial intelligence methods show promising results in detecting driver drowsiness through physiological and behavioral signals, and machine learning models are increasingly used in vehicle safety systems to identify early signs of fatigue from sensor data. Traditional centralized methods still face serious limitations, as sending driving data to external servers increases privacy and security risks in connected vehicles. Federated learning offers a possible solution by training models locally on each vehicle or monitoring unit without exposing personal information.

This study introduces a privacy-preserving framework to detect driver drowsiness using On-Board Diagnostics II (OBD-II) data and transfer learning within a Class-Weighted Federated Averaging (CW-FedAvg) setup. OBD-II sensor data, such as speed, engine revolutions per minute (RPM), throttle position, and steering torque, are converted into Mel-Frequency Cepstral Coefficient (MFCC) images. This transformation allows the model to predict temporal changes and visual patterns. A pretrained ResNet50 network is fine-tuned to classify driver states on each client, and the server applies a CW-FedAvg strategy to keep learning balanced across different client cars. Model behavior is further analyzed through t-Distributed Stochastic Neighbor Embedding (t-SNE), Uniform Manifold

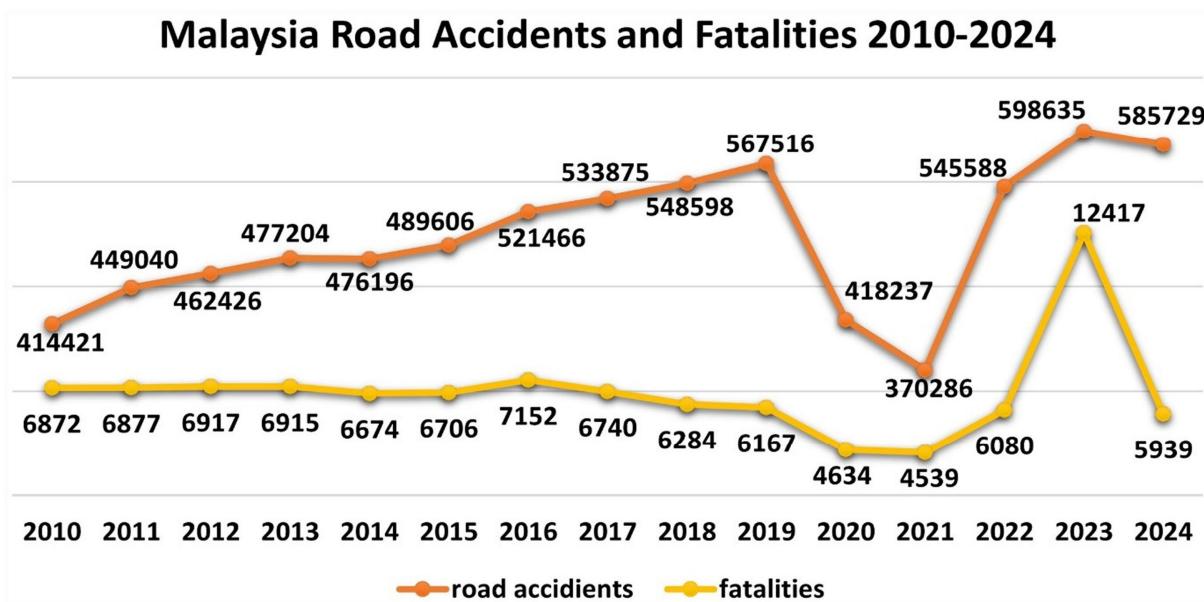


FIGURE 1  
Road traffic accidents and fatalities in Malaysia (2010–2024).

Approximation Projection (UMAP) and SHapley Additive exPlanations (SHAP). These techniques are used to evaluate the model and its ability to learned correctly. Statistical metrics, such as the Matthews Correlation Coefficient (MCC), Cohen's kappa ( $\kappa$ ), and the 95% Confidence Interval (95% CI), are also used to analyze the stability and reliability of the results. This framework improves detection accuracy and privacy within intelligent transportation infrastructure. The main contributions of this study are as follows:

- Use of OBD-II sensor data combined with MFCC for converting time-series signals into image representations.
- Integration of a pretrained ResNet50 model within a federated learning framework for improved accuracy and privacy preservation.
- A CW-FedAvg scheme mitigates class imbalance and stabilizes convergence under non-independent and identically distributed (non-IID) data.
- Comparative experiments with FedAvg, FedProx, and lightweight CNN backbones confirm near-centralized accuracy while maintaining privacy.
- Interpretability analyses (t-SNE, UMAP, SHAP) and statistical validation (MCC,  $\kappa$ , 95% CI) confirm model reliability.

Driver monitoring within intelligent transportation plays an important role in improving the safety and resilience of mobility systems. The proposed framework strengthens reliability and privacy in assessing driver alertness while protecting road transportation as a vital part of modern infrastructure. Altogether, this study supports the development of safer and more secure transport networks through practical means. Section 2 provides an overview of existing research on driver drowsiness detection methods through the literature review. Section 3, Materials and Methods, describes the workflow in detail, including the collection of OBD-II sensor data, preprocessing of time-series signals, conversion into image representations using MFCCs, and training of the pretrained ResNet50 model within a federated transfer learning framework. In Section 4, the study's results are presented, followed by a discussion of the results and limitations in Section 5. Section 6 concludes the paper and outlines future directions.

## 2 Literature review

In recent years, researchers have made steady progress in creating and improving methods that can detect driver's drowsiness with greater accuracy and reliability. Different detection methods for driver drowsiness have been developed in recent studies, including image-based, biological-based, vehicle-based, and hybrid. [Reddy et al. \(2017\)](#) divide driver monitoring methods into three groups, with the first analyzing vehicle behavior, such as acceleration, braking, and steering. The second method uses physiological signal inputs like heart rate and brain signals to check for drowsiness. The third method covers computer vision to study facial expressions and eye movements in real-time. In a recent study, [Dua et al. \(2021\)](#) suggest that combining vehicle data with physiological and behavioral features can improve the accuracy of drowsiness detection in intelligent transportation. Behavioral features include head face and eye movements while physiological measures include Electrooculography (EOG), Electroencephalograms (EEG), Electrocardiograms (ECG) and heart

rate. Vehicle sensors data collected through OBD-II port include speed, RPM, throttle position and steering behavior which can provide detection data for driver drowsiness. In a recent systematic review [Shaik \(2023\)](#) studied methods for detecting and predicting driver drowsiness using machine learning, computer vision and physiological data. These hybrid detection techniques are grouped using physiological signals, behavioral features, and vehicle sensors to highlight research trends and ongoing challenges in driver drowsiness detection.

In recent studies, researchers have utilized different approaches to detect driver drowsiness using machine learning and deep learning with data from driver behavior, physiological signals, and vehicle sensors. But in the domain of vehicle behavior, mainly OBD-II sensors, such as steering angle, throttle position, and speed sensors, are key to early detection of driver drowsiness. Other sources of vehicle behavior, such as the global positioning system (GPS), gyroscopes, lane position, and engine RPM, play an important role in identifying the driving patterns that lead to driver drowsiness. Studies by various researchers show that continuous vehicle data can be used to train models to prevent risky driving events. Physiological indicators, including EEG, ECG, EOG, EMG, and wearable sensor data, have also proven useful in classifying drowsy states ([Harkous and Artail, 2019](#); [Malik and Nandal, 2023](#); [Arefnezhad et al., 2019](#)). Also, recent studies by [Kundinger et al. \(2020\)](#) and [Nasri et al. \(2022\)](#) showed that models built on physiological and behavioral signals can reach high accuracy in detecting drowsiness. [Martins et al. \(2021\)](#) examined wearable systems for fatigue monitoring and noted their promise for real-time use although data stability and model generalization remain weak points, also visual indicators such as facial expressions and eye movement have also become a focus since they tend to shift noticeably as fatigue develops. Many studies have shown that visual patterns provide reliable indicators of driver drowsiness, while newer hybrid methods that combine vehicle data with physiological and behavioral cues are improving detection accuracy and adaptability ([Ed-Doughmi et al., 2020](#); [Vu et al., 2019](#); [Zhao et al., 2020](#)). Studies such as [Omerustaoglu et al. \(2020\)](#) and [Gwak et al. \(2020\)](#) show that blending multiple data sources leads to better accuracy and safer driving outcomes.

[Ping and Shie \(2022\)](#) investigated how a hybrid strategy could identify driver drowsiness in Malaysia. Their method integrates vehicle diagnostics, physiological signs, and remote sensing data to collect data by driving a specially outfitted car along the North–South Motorway to compare various detection systems. [Albadawi et al. \(2022\)](#) examined a wide range of drowsiness detection methods, focusing on physiological and sensory cues. The authors discussed how machine learning appears to shape the next phase of progress in this field. The paper highlights both the potential and the remaining gaps in current systems and suggests that a stronger link between technology and real-world use may be needed. [Ahmad et al. \(2023\)](#) conducted a systematic review of recent machine learning and deep learning techniques for detecting drowsiness using several data sources. Their findings show that machine learning may continue to improve road safety and reduce fatigue-related accidents. Even so, vehicle-based behavioral and physiological systems face challenges such as privacy issues, setup difficulty, and limited data availability. The evidence so far suggests that vehicle-based systems that rely on OBD-II data may offer a practical and effective way to identify driver fatigue while improving overall safety and efficiency.

Recent work in time series analysis and feature extraction shows how effective Mel spectrograms and MFCCs can be across many applications. Gupta et al. (2013) used MFCCs for hand gesture recognition, whereas Alves et al. (2021) applied high-dimensional arrays of features such as MFCCs and Tempograms to capture the structure of sound data. In existing research, Mel spectrograms combined with convolutional neural networks are used for many fields, including sound and time-domain signals. Another study developed a respiratory condition identification system that processed time signals using CNN models. These results show that converting time series into Mel spectrograms improves feature extraction and classification accuracy (Stankovic et al., 2024; Purkovic et al., 2024). Using deep learning and MFCC, Mohammed et al. (2023) transformed radio frequency signals into Mel spectrograms and used a pre-trained YAMNet model for drone classification. In the recent work, Bacanin et al. (2024) applied CNNs and optimization techniques to respiratory sounds. The results indicate that Mel spectrograms improve detection accuracy and improve strength of CNN model's feature extraction. These studies suggest that Mel spectrograms and MFCCs are effective techniques for feature extraction and temporal pattern analysis. Similarly, transforming OBD-II time-series data into MFCC images is a strong baseline for CNNs to use visual patterns in detecting driver drowsiness. These techniques improve the accuracy and reliability of driver drowsiness detection by allowing models to process temporal signals as images.

In recent studies, researchers have explored the use of vehicle sensors and OBD-II data in decentralized learning approaches to detect driver drowsiness. These sensors measure RPM vehicle speed throttle position and steering torque and these signals fluctuate according to fatigue or drowsiness during driving. Michailidis et al. (2025) work on these signals, which capture behavioral patterns without relying on video-based or physiological measurements that can be intrusive. Converting OBD-II readings into higher-level representations offers a lightweight and privacy-friendly input for deep learning models. The studies by Albadawi et al. (2023) and Safarov et al. (2023) showed promising results in detecting driver drowsiness using visual features. The use of centralized data collection raises privacy risks and affects the utilization of these systems in real cars on the road today. A federated transfer learning framework using OBD-II data provides a practical solution to this problem. Each vehicle can contribute to global model updates while keeping its local data safe and private. This introduced privacy and effective learning across distributed vehicle sources. Using client-weighted federated averaging with OBD-II signals offers a comparison between accuracy and data protection by equal bias of each client with protection of raw data (McMahan et al., 2017; Hong et al., 2022; Zeng et al., 2023; Michailidis et al., 2025). In model evaluation, t-SNE is used to visualize latent features with accurate prediction and interpretability. In their research Xu et al. (2020) used t-SNE that can find clustering patterns in microbiome data which helps to understand complex relationships and better evaluation of classification performance. Furthermore, SHAP values of each feature are important for individual predictions. This visualization explains how inputs drive the final model output results (Zhang et al., 2023).

### 3 Materials and methods

The proposed methodology utilized a structured approach to detect driver drowsiness by combining OBD-II sensor data with

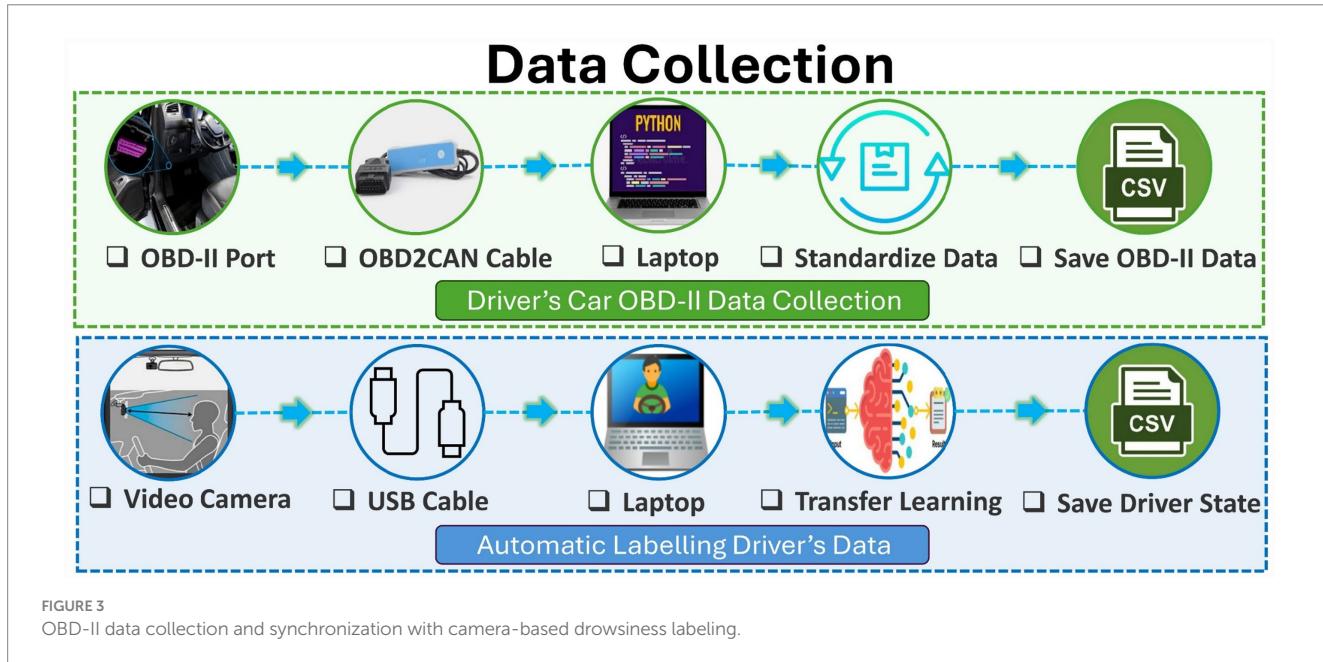
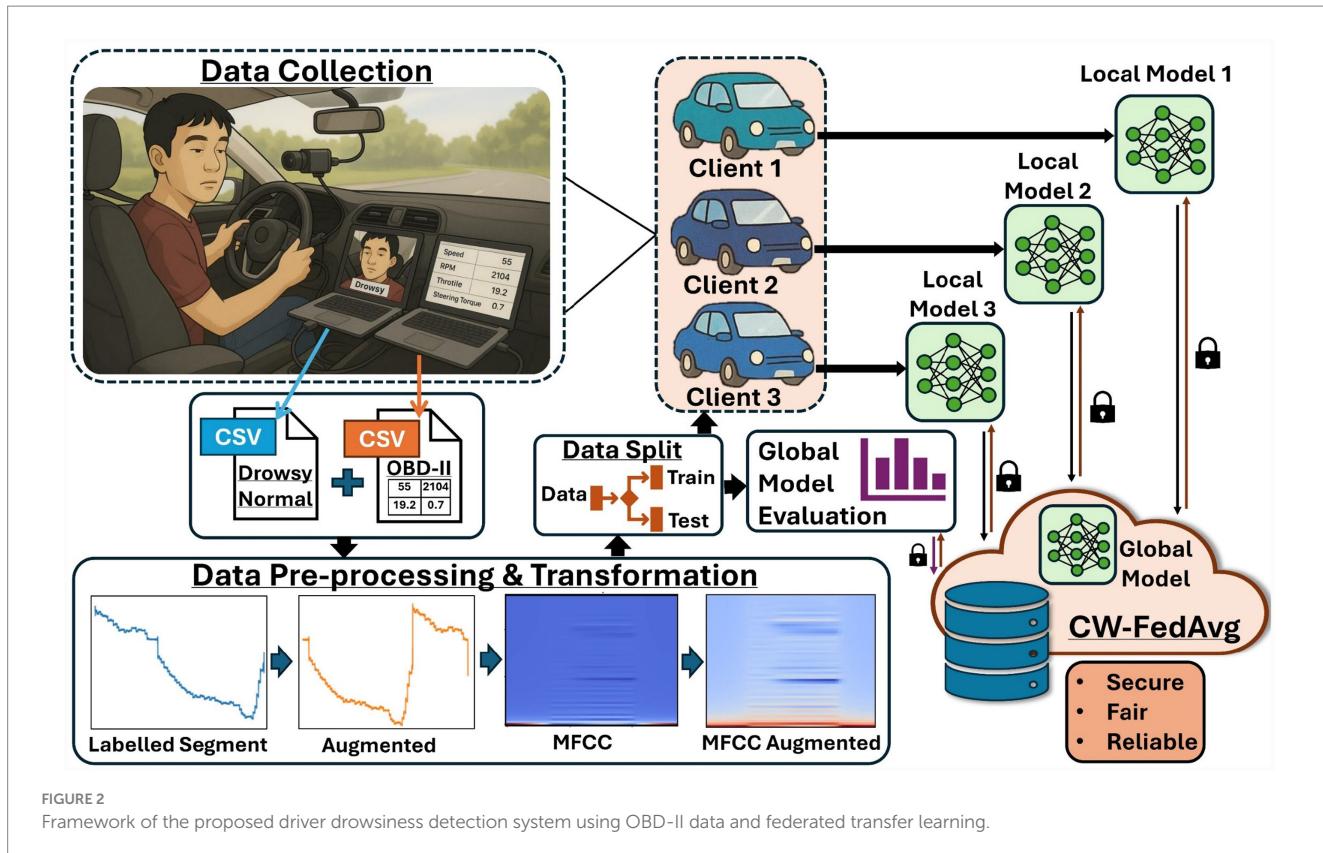
camera-based labeling. Vehicle parameters such as speed, RPM, throttle position, and steering torque were collected through the OBD-II port. Also, a camera detected facial features to determine the driver's state. Python scripts synchronized the two data sources using timestamps, ensuring each sensor reading corresponded to the correct facial label. This automatic labeling groups the final data as drowsy or normal producing a dataset for supervised training of model. OBD-II data is gathered as time-series signals and later converted into two-dimensional images using MFCC. The images are split into training and testing sets for model development, for training a pretrained ResNet50 model is using transfer learning setup and is fine-tuned to classify drowsiness based on the MFCC images. Each client vehicle trains the model locally within a federated learning framework and then global model is updated through CW-FedAvg which preserves data privacy while balancing updates from all clients. The final model known as DrowsyXnet learns patterns linked to driver drowsiness without any direct sharing of vehicle data. Model performance is evaluated on the test set and the results are saved with best model for evaluation and comparison. This approach presents a privacy-aware and efficient method for detecting driver drowsiness in real time using OBD-II data, as illustrated in Figure 2.

#### 3.1 Data collection

The data collection process was developed to record vehicle behavior and driver drowsiness in a synchronized and privacy-aware way. Data was collected from several drivers using two linked sources: OBD-II telemetry from the vehicle's diagnostic port and facial video recordings used only for labeling. The OBD-II stream provided real-time measurements such as speed, engine RPM, throttle position, and steering torque. These readings were retrieved through an OBD2CAN interface connected to a laptop running a custom Python script. The raw hexadecimal outputs were translated into numerical values using predefined formulas and stored as timestamped CSV files so each entry reflected the vehicle's exact operating condition. To obtain reliable ground truth for drowsiness, a camera placed in front of the driver continuously recorded facial video during data collection. A pre-trained transfer learning model for facial drowsiness detection analyzed the frames and classified the driver's state as Drowsy or Normal. The predictions with their timestamps were aligned with the OBD-II data so that every sequence of telemetry corresponded to the correct driver condition at that moment. Visual data were not used for model training but only for labeling the OBD-II data, which became the sole input for later stages. This setup maintains driver privacy while still producing accurate labels from visual evidence. The resulting dataset containing time-aligned OBD-II readings and drowsiness labels forms a strong base for feature extraction augmentation and federated transfer learning, as shown in Figure 3.

#### 3.2 Data preprocessing and transformation

The preprocessing stage played a key role in preparing the collected OBD-II data for analyzing driver drowsiness. The continuous stream of time-series signals from the vehicle sensors were cleaned and segmented to a structured format. The continuous stream data were then split into segments of 3,500 samples, with each segment



representing about 3 s of driving. Segments that are shorter than the target length were padded with zeros to keep a consistent structure across all samples. This step ensured that every input segment had the same dimensions. Each segment including speed, engine RPM, throttle position and steering torque was aligned with its corresponding drowsiness label. Synchronization involved matching each sensor segment with the ground truth labels so both referred to

the same three-second time window. This process ensures that each data segment reflects the driver's state. Each driver's dataset was balanced to address the imbalance between normal and drowsy samples. The drowsiness class had fewer entries so for balancing samples were recreated using time-stretching and expansion techniques. This process not only recreated new but realistic variations in the signals also preserved the original temporal relationships. This

dataset is class-balanced and more reflective of real driving conditions. An illustration of the up-sampling signal using data augmentation is shown in [Figure 4](#). After balancing, a two-stage augmentation strategy was utilized to increase the dataset while preserving stability across both classes during training. This approach introduced realistic variations to signals and MFCCs while preserving the critical patterns in the dataset. The two stages are as follows:

- Signal-level transformations included scaling, permutation, flipping, noise addition, magnitude warping, and slicing. These steps increased natural variation while keeping the temporal feature of the signal.
- Spectro-temporal transformations on the MFCC images used time-shift and spectral scaling to handle small alignment slips and sensor noise that often appear during driving.

Recent studies on sensor and audio-based data show that multi-stage augmentation strategies can significantly improve model performance. [Liang et al. \(2023\)](#) and [Yu et al. \(2023\)](#) stated that such augmentation enhances generalization across datasets and also stabilizes model performance. Signal-level augmentation was applied, including amplitude scaling, random permutation, signal flipping, Gaussian noise injection, magnitude warping, and window slicing. These steps are illustrated in [Figure 5](#).

After completing time-domain augmentation, each OBD-II signal was transformed into a spectral representation using MFCCs. This step captured frequency features that represent driver drowsiness behavior patterns during driving. MFCCs were computed with 55 coefficients, FFT window size of 2048, a hop length of 64, and a sampling rate of

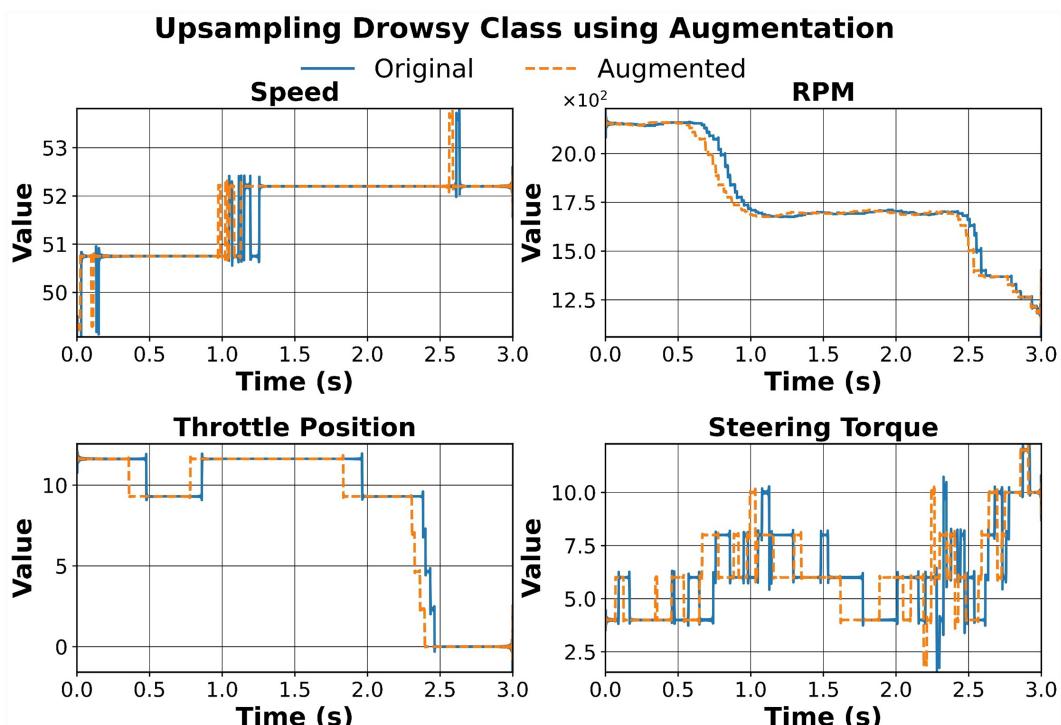
16 kHz as summarized in [Table 1](#). Each OBD-II parameter was processed individually, and the resulting coefficients were stacked to create four-channel MFCC images. The overall process of converting time-series signals into MFCC-based images is illustrated in [Figure 6](#).

A final augmentation step was applied to the MFCC representations to further enhance variety of features in dataset. This step involved small temporal shifts, spatial scaling and time-stretching techniques which generated variations in spectral patterns. The process expanded each dataset by a factor of 4, producing final sizes of (23,184, 55, 55, 4) for Driver 1, (22,400, 55, 55, 4) for Driver 2, and (12,936, 55, 55, 4) for Driver 3. [Figure 7](#) illustrates the MFCC augmentation process for each sensor's spectral image. This multistage preprocessing and transformation created a detailed dataset that captured both temporal and spectral aspects of driver behavior. This provided a strong foundation for federated learning experiments for detecting driver drowsiness.

The number of samples for the data balancing and augmentation process is summarized in [Table 2](#). The dataset was first balanced by up-sampling the drowsy class to match the number of normal samples for each driver. Then, signal-level augmentation expanded the dataset about sevenfold, yielding 5,796, 5,600, and 3,234 samples for drivers 1, 2, and 3, respectively. Following this, MFCC-level augmentation added temporal and spectral variations, further expanding the dataset by about fourfold to 23,184, 22,400, and 12,936 samples for the same drivers.

### 3.3 Data split

A stratified data split was applied to the MFCC-augmented OBD-II data to ensure balanced and reliable evaluation of the model



**FIGURE 4**  
Up-sampling the drowsy class using signal-level augmentation to balance the dataset.

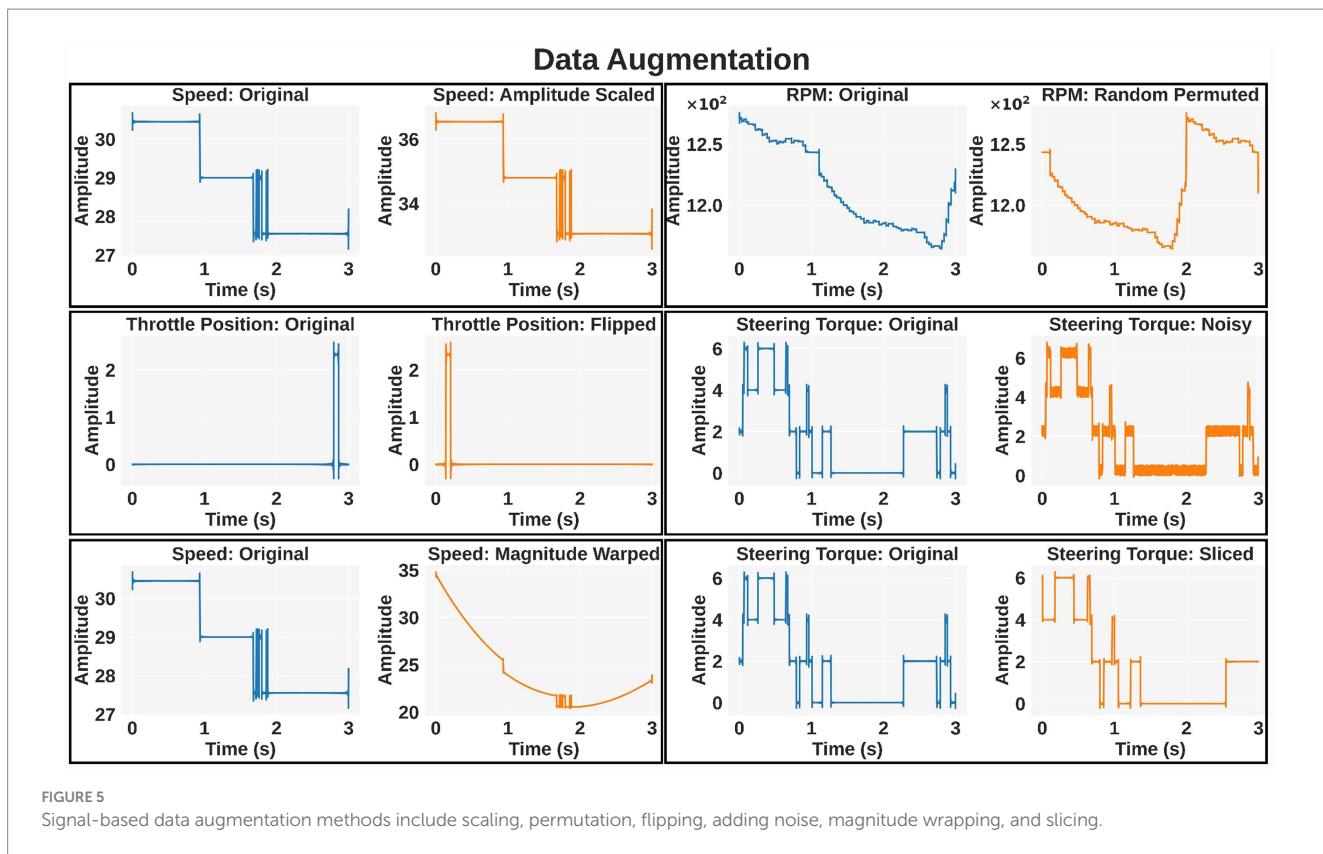


TABLE 1 Key parameters used for MFCC feature extraction.

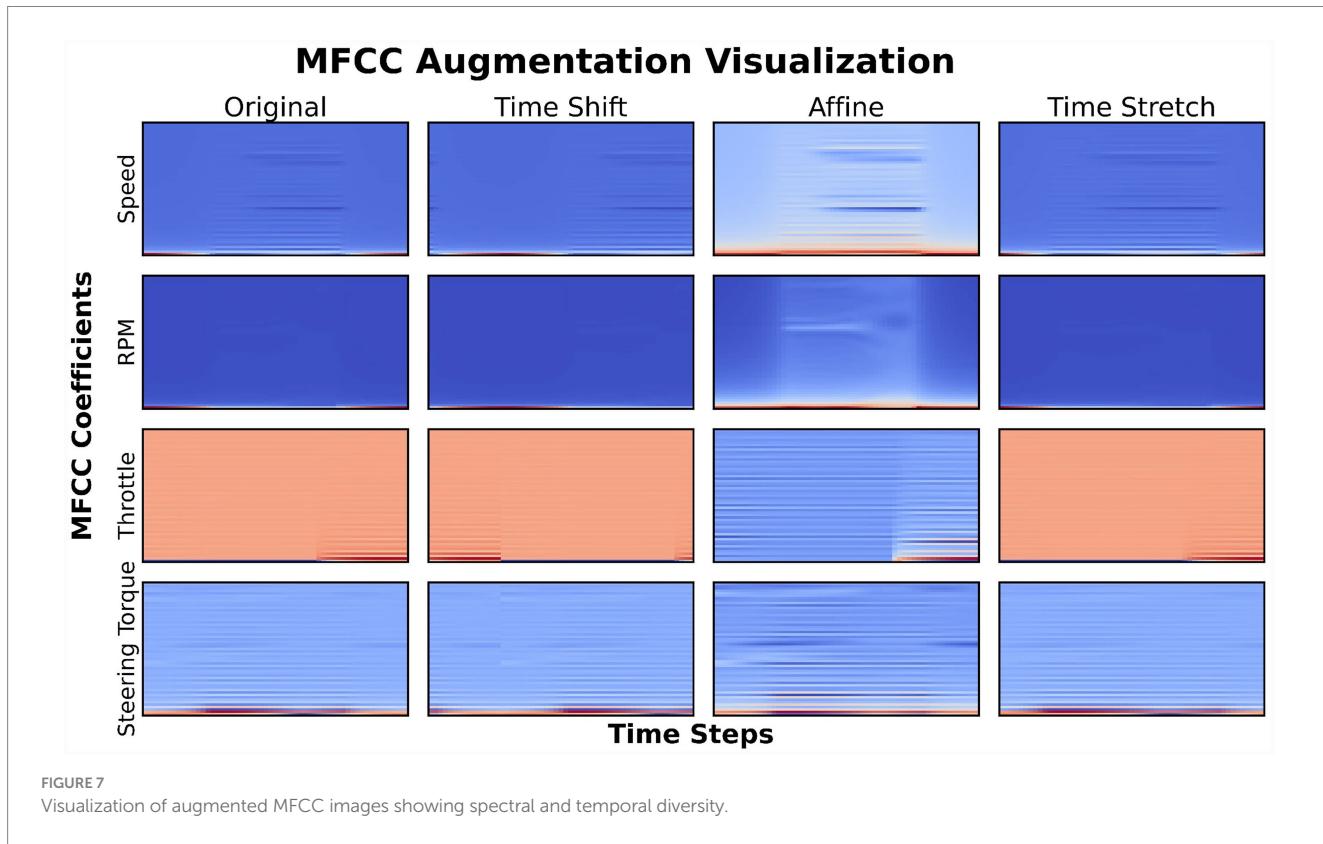
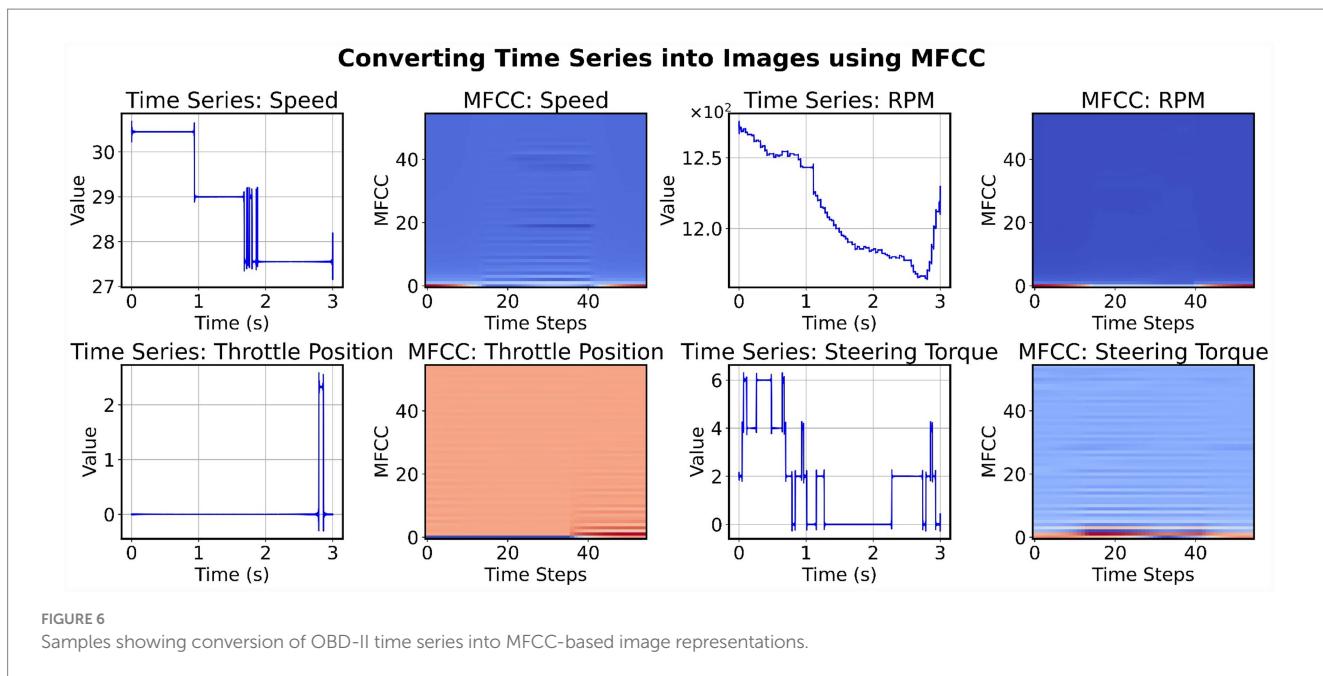
Parameter	Value	Purpose
MFCC coefficients	55	Rich spectral representation
FFT window size	2048	Balance time/frequency resolution
Hop length	64 samples	Smooth temporal continuity
Sampling rate	16 kHz	Preserve signal detail
Channels stacked	4 (Speed, RPM, Throttle, Steering Torque)	Multi-sensor fusion
MFCC image size	$55 \times 55 \times 4$	CNN input dimension

for driver drowsiness detection. Stratified sampling maintained proportional representation for normal and drowsy classes within all data subsets. Each driver's dataset was split using 80:20 ratio with 80% of the samples used for training and the remaining 20% for testing. This approach provided the model with enough data to learn each driver's behavioral patterns while reserving a portion for testing performance on unseen data. During federated learning, each client used its own training data to update its local model weights based on the driver's data distribution. These locally updated weights were then communicated to the central server and aggregated to update the global model using CW-FedAvg. Instead of testing models separately, the evaluation phase grouped all tests. Therefore, the testing subsets from all drivers were combined into a single global test set. This

combined test set was then used to evaluate both the aggregated global model and each client's locally trained model based on their respective weight updates. This approach provided a uniform and consistent evaluation framework by testing every model on the same global test dataset. That allowed direct performance comparison for clients and global model accuracy using identical evaluation conditions. This approach minimized the impact of local data biases by evaluating the aggregated model on a combined global test set. In addition, the figure shows that the global model can handle the diversity of driving behaviors across different drivers. The overall data split and the composition of the global test set are summarized in Table 3.

### 3.4 Federated transfer learning

The federated learning framework worked as the primary training strategy for developing the driver drowsiness detection system. This framework contains the DrowsyXnet model, which was proposed to capture both temporal and spectral patterns from MFCC-transformed OBD-II data. Training was done locally on each client's car using its own dataset. Then, the updated weights were aggregated through the secure CW-FedAvg method and updated a global model. The DrowsyXnet global model protects privacy while still performing reliably across all drivers. To speed up learning and trained on relevant features DrowsyXnet used transfer learning with a pretrained ResNet-50 backbone. As the MFCC inputs had four channels, an extra convolutional layer with three filters and a  $3 \times 3$  kernel was added to convert them into a three-channel format compatible with ResNet-50. Using this layer kept the main MFCC structure. The pretrained ResNet-50 initially trained on ImageNet, which served as a feature



extractor identifying key features in the signals. Transfer learning adapted these general visual patterns for the specific task of detecting driver drowsiness. Feature extraction outputs were passed through a Global Average Pooling layer to condense activation maps and preserve significant spatial information for training. A dense layer with 256 neurons and ReLU activation followed, introducing nonlinearity to learn complex feature interactions. To reduce overfitting, Dropout (rate 0.2) and L2 regularization were applied. The

final dense layer, equipped with a sigmoid activation function, performed binary classification between Drowsy and Normal states. The model was trained using the Adam optimizer with a learning rate of 0.0001 and binary cross-entropy loss. The architecture of DrowsyNet is illustrated in Figure 8.

The federated learning process followed multiple communication rounds using the CW-FedAvg approach. The server began by sending the global model to all clients, each holding the same model

TABLE 2 Summary of data balancing and augmentation effects on OBD-II data for driver drowsiness detection.

Driver ID	Class balancing (samples per class)			Dataset size after balancing	Dataset size after signal-level augmentation	Dataset size after MFCC-level augmentation
	Class	Before	After			
Driver 1	Normal	414	414	828 samples	5,796 samples (7 × increase from balanced set → $828 \times 7 = 5,796$ )	23,184 samples (4 × increase from signal-level → $5,796 \times 4 = 23,184$ )
	Drowsy	111	414		5,600 samples (7 × increase from balanced set → $800 \times 7 = 5,600$ )	22,400 samples (4 × increase from signal-level → $5,600 \times 4 = 22,400$ )
Driver 2	Normal	400	400	800 samples	3,234 samples (7 × increase from balanced set → $462 \times 7 = 3,234$ )	12,936 samples (4 × increase from signal-level → $3,234 \times 4 = 12,936$ )
	Drowsy	87	400		2,588 samples (7 × increase from balanced set → $365 \times 7 = 2,588$ )	10,340 samples (4 × increase from signal-level → $2,588 \times 4 = 10,340$ )
Driver 3	Normal	231	231	462 samples	1,848 samples (7 × increase from balanced set → $260 \times 7 = 1,848$ )	7,392 samples (4 × increase from signal-level → $1,848 \times 4 = 7,392$ )
	Drowsy	66	231		1,105 samples (7 × increase from balanced set → $158 \times 7 = 1,105$ )	4,420 samples (4 × increase from signal-level → $1,105 \times 4 = 4,420$ )

TABLE 3 Data split distribution for the dataset.

Driver ID	Total samples (after MFCC augmentation)	Training set (80%)	Testing set (20%)
Driver 1	23,184	18,547	4,637
Driver 2	22,400	17,920	4,480
Driver 3	12,936	10,348	2,588
Global test set	–	–	<b>11,705</b> (Combined)

architecture locally. Every client trained its model on its own OBD-II and MFCC data for one epoch in each round and kept the best weights based on test accuracy. After training, the clients sent only their updated weights, sample counts, and class distribution information to the server. No raw data or model structure was shared, which maintained data privacy. The server merged the received weights using CW-FedAvg, adjusting each client's contribution based on its dataset size and class balance to keep the aggregation fair. The global model was evaluated after each training round, and the best-performing version was saved on the server as the final model after 60 rounds. This process allowed the global model to improve continuously while preserving complete data privacy for all client cars.

### 3.5 Global model evaluation

The evaluation of the global model provides insight into the understanding and effectiveness of the federated transfer learning framework for driver drowsiness detection. Over 60 communication rounds, performance metrics were monitored for both the global and local models. After every round, client cars delivered updated weights to the central server where the testing of global model has to be done. Each local model was evaluated on the combined global test set using its own weights, and the results were saved for comparison. The server also evaluated the aggregated global model and saved a checkpoint of the best-performing model weights based on test accuracy. Figure 9a shows that global accuracy increased across rounds, and Figure 9b shows a steady decline in loss. This trend suggests effective learning and coordination among clients. The final global model achieved a test accuracy of 98.29% on the global test set, which included data from all drivers. Conversely, the local model showed some fluctuation due

to differences in individual datasets. This pattern proves that CW-FedAvg can increase consistency by weighting client contributions according to data size and class balance metrics. Global model checkpointing ensured the best configuration was saved, reducing the risk of overfitting during training and making it ready for deployment.

Latent features were extracted from the test dataset by passing inputs through all layers except the final output layer to analyze how the global model represents and classifies driver states. These features offered a compressed view of the patterns in the data, which were analyzed using dimensionality reduction techniques like t-SNE and UMAP. Both methods reduced the complex latent space to two dimensions, making it easier to observe differences between Drowsy and Normal states. In the t-SNE visualization, each point was colored according to its true label—the warm reds representing drowsy states and cooler blues representing normal ones. As shown in Figure 10, the DrowsyXnet model appears to separate the two groups reasonably well and form clusters that show distinct driving patterns. UMAP provided a second view and showed a similar cluster pattern, supporting the model's ability to successfully identify differences between drowsy and normal driving as presented in Figure 11. The separation is generally clear with few overlapping points due to natural variability in driver behavior, underscoring that the model performs perfectly.

The overall evaluation suggested that the federated learning approach provides stable global convergence and performs better than models trained singly on local data sets. Observing performance across multiple communication rounds provided clear insight into how learning evolved at both local and global levels effectively. The DrowsyXnet global model achieved high accuracy and showed resilience across different drivers. Also, keeping raw data private and allowing effective collaboration among client teams.

## 4 Results and discussion

The performance of the proposed DrowsyXnet global model is evaluated using classification metrics derived from the confusion matrix and classification report. These metrics assess the model's ability to distinguish between Drowsy and Normal driver states on the global test set. The confusion matrix consists of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), which are used to calculate precision, recall, and F1-score.

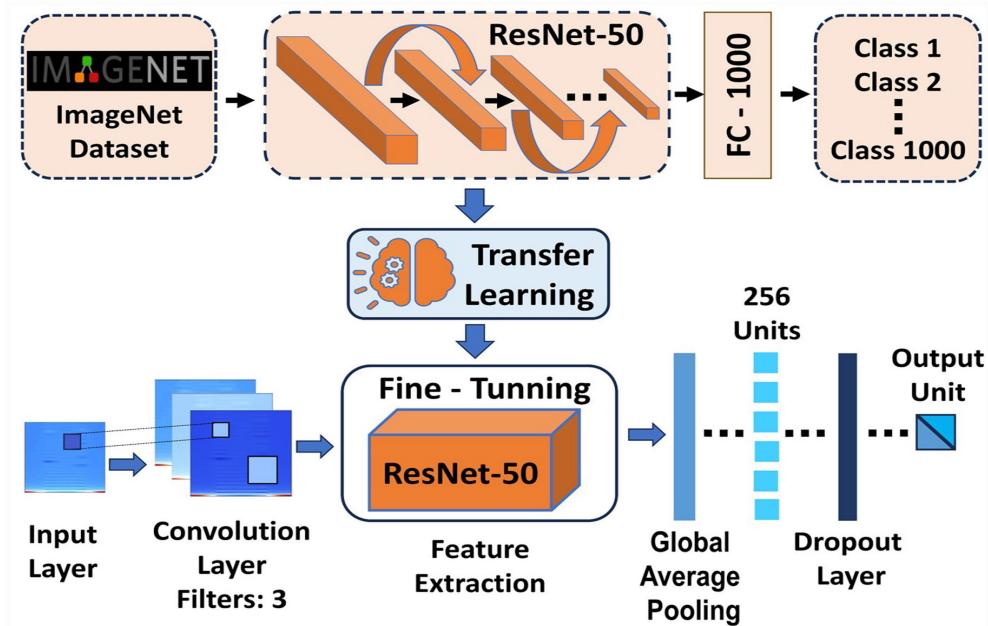


FIGURE 8  
Architecture of DrowsyXnet model based on transfer learning with ResNet-50.

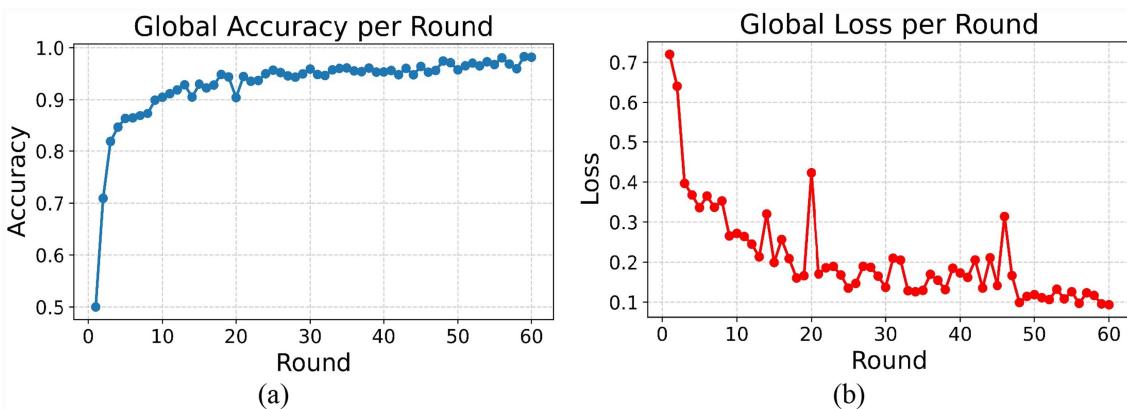


FIGURE 9  
Global model evaluation on global test set over 60 federated communication rounds: (a) global model accuracy and (b) global model loss.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Recall = \frac{TP}{TP + FN}$$

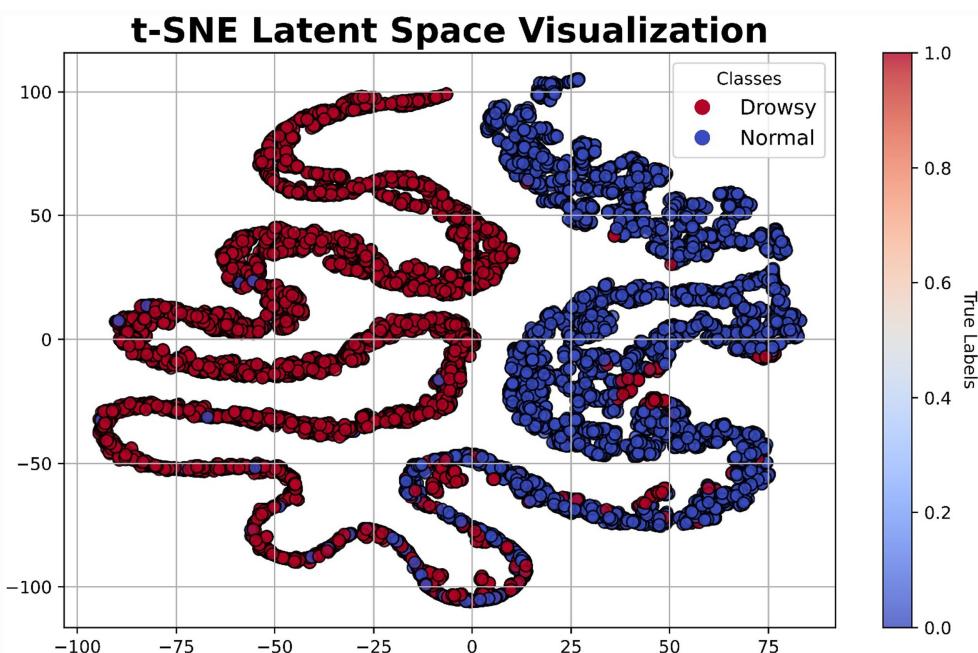
$$Precision = \frac{TP}{TP + FP}$$

$$F1-Score = 2 \times \frac{Recall \times Precision}{Recall + Precision}$$

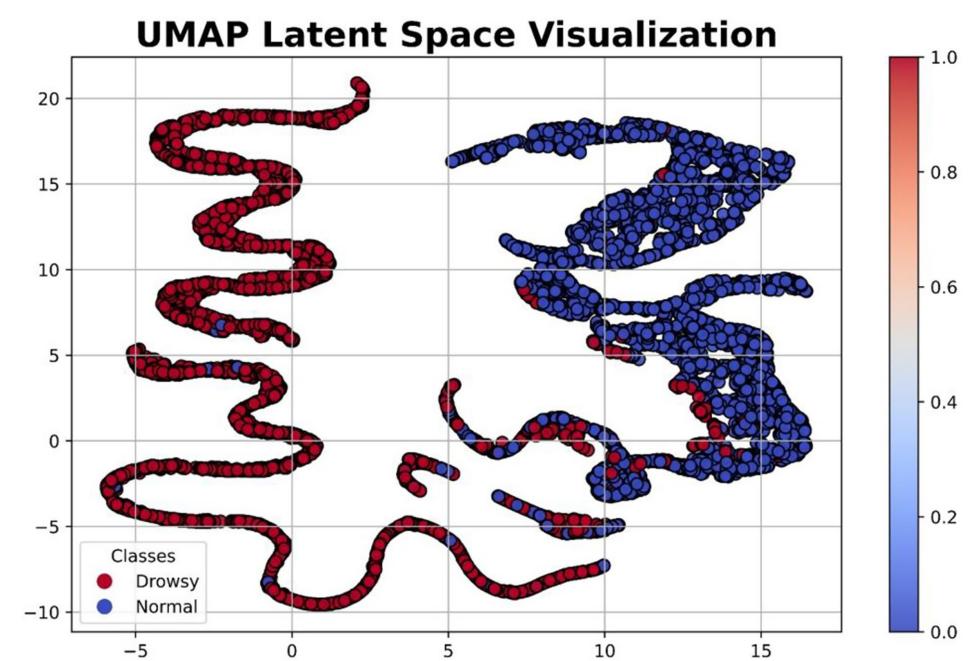
The values used in the above equations, i.e., TP, FP, TN, and FN values, were obtained from the confusion matrices of the proposed DrowsyXnet model. Figure 12 presents the confusion matrix for the

model DrowsyXnet global model in this proposed federated transfer learning framework. The confusion matrix shows the true and false predictions across both classes. Out of a total of 11,705 test samples, only 113 false negatives (drowsy samples incorrectly classified as normal) and 87 false positives (normal samples misclassified as drowsy) were observed. This low misclassification rate highlights the model's strong discriminative ability and generalization across varying driver behaviors and sensor readings. Furthermore, the parameters extracted from this confusion matrix are shown in Table 4.

Table 4 presents a detailed breakdown of performance metrics for the proposed CW-FedAvg-based DrowsyXnet model, classifying samples into “Drowsy” and “Normal” driver states. Each row of the table corresponds to a class, while the columns show precision, recall, F1-score, and support (the number of instances per class). For the Drowsy class, the model achieved a precision of 98.51% conforming



**FIGURE 10**  
t-SNE latent space visualization for DrowsyXnet model evaluation.



**FIGURE 11**  
UMAP latent space visualization for DrowsyXnet model evaluation.

that predictions of drowsiness are usually correct. The recall of 98.07% indicates that most actual drowsy cases are detected, though a few samples are misclassified. Achieving an F1 score of 98.29 percent proves a close balance between precision and recall, so the model manages detection very well. These results show that the model is identifying patterns in OBD-II signals such as RPM, vehicle speed,

throttle position, and steering torque that detect early fatigue or lapses in attention.

For the Normal class, precision reached 98.08%, showing the model correctly identifies typical driving states with few false positives. Recall was slightly higher at 98.51%; nearly all normal instances were captured. The F1-score was again 98.29%, reflecting consistent

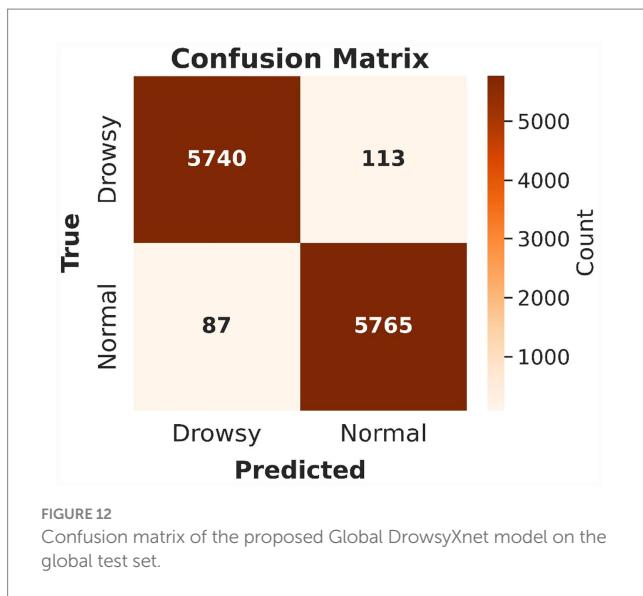


TABLE 4 Class-wise performance metrics of the proposed CW-FedAvg DrowsyXnet model.

Class	Precision	Recall	F1-Score	Support
Drowsy	0.9851	0.9807	0.9829	5,853
Normal	0.9808	0.9851	0.9829	5,852
Macro Avg	<b>0.9829</b>	<b>0.9829</b>	<b>0.9829</b>	<b>11,705</b>
Weighted Avg	<b>0.9829</b>	<b>0.9829</b>	<b>0.9829</b>	<b>11,705</b>

performance across metrics. Both the macro and weighted averages were 98.29%, indicating that the model maintains fairly uniform accuracy across classes. Overall, these results stated that the CW-FedAvg-based DrowsyXnet model performs reliably and consistently. It offered a balance between accuracy and privacy and provided a deployable solution for real-world driving scenarios.

#### 4.1 Client and global model comparison

The comparison of client-level and aggregated global model accuracies over 60 communication rounds using the global test set, as shown in Figure 13. Each client represents a different driver dataset collected under varying conditions. During federated training, all clients performed local updates, which were then aggregated on the central server using the CW-FedAvg strategy. The figure suggests accuracy improved continuously across all clients, which appears to reflect stable convergence over time. Client car 2 reached high accuracy faster, likely because its dataset was more balanced. Client car 1 followed a similar trend, while client car 3 showed slightly more fluctuation due to less data as compared to other client cars. The aggregated global model tended to outperform individual clients and showed smoother convergence across rounds. Accuracy increased during early rounds and appeared to level off near the fortieth round. By the end of training, the global model reached 98.29% on the global test set, showing strong generalization. Overall, the rising trend in the

global curve indicates that federated learning successfully integrated knowledge from all clients into a global model. The weighting strategy in CW-FedAvg can balance contributions from each client, preventing any single dataset from dominating the learning process. The global test results further confirmed that the aggregated model generalized well to unseen data from all clients. Evaluating each local model on the server-side global test set showed that local accuracies changed across rounds, and the global model consistently performed better. Aggregating class-weighted updates appears to improve learning stability and reduce the impact of non-identical data distributions across clients.

#### 4.2 Comparative model performance

To benchmark the proposed framework, its performance was compared with several previous machine learning and deep learning models trained under similar conditions. Conventional classifiers such as Logistic Regression (LR) with accuracy 60.90% and Support Vector Machines (SVM) achieved 72.18% accuracy which conclude that these models appeared to struggle. Time-series-oriented models such as long short-term memory (LSTM) achieved approximately 81.95% accuracy, and the One-Dimensional Convolutional Neural Network (1D-CNN) reached 88.35%, which improved performance compared to conventional classifiers. But these models cannot fully predict the complex spectral-temporal patterns present in MFCC representations. Modern deep architectures that leverage visual representations showed much stronger results. EfficientNetB0 reached 97.63%, DenseNet201 achieved 97.45%, and ConvNeXtTiny scored 96.81% accuracy. Overall, the comparison demonstrated that using MFCC-based visual representations in combination with deep learning yields an advantage over traditional and simpler time-series models. The DDD-GC-ViT model also achieved 97.90%, confirming the advantage of attention-based mechanisms. However, the DrowsyXnet model using CW-FedAvg achieved the highest accuracy of 98.29% while preserving data privacy, as shown in Table 5.

To assess real-time feasibility, two lightweight models, EfficientNet-B0 and MobileNet V3-Small, were compared with the main backbone (ResNet-50). Results were generated using the same MFCC inputs and training configuration. To ensure reproducibility, all inference time benchmarks were conducted on a Google Colab environment equipped with an NVIDIA T4 Tensor Core (16GB GDDR6) GPU. The results show that ResNet-50 achieves the highest accuracy but lightweight models offer 2 times faster inference with minimal performance loss as stated in Table 6.

A comparison of the training setups in Table 7. The centralized baseline marks the upper limit because it trains on all data at once reaching 98.67% accuracy, an F1-score of 98.65%, an MCC of 0.968 and a Cohen's  $\kappa$  of 0.967. FedAvg drops slightly on every metric settling at 97.89% accuracy, 97.74% F1, 0.955 MCC and 0.953  $\kappa$  while FedProx achieved 98.05% accuracy, 98.00% F1, 0.960 MCC and 0.959  $\kappa$ . The proposed CW-FedAvg model achieved the best accuracy 98.29, 98.29% F1, an MCC of 0.966, and a  $\kappa$  of 0.965 and narrowing the gap with the centralized setup still preserving data privacy.

These trends show that CW-FedAvg provides more stable performance across all evaluation metrics by giving balanced influence on clients with small or uneven datasets. The close match between the centralized and federated results confirms that high-quality

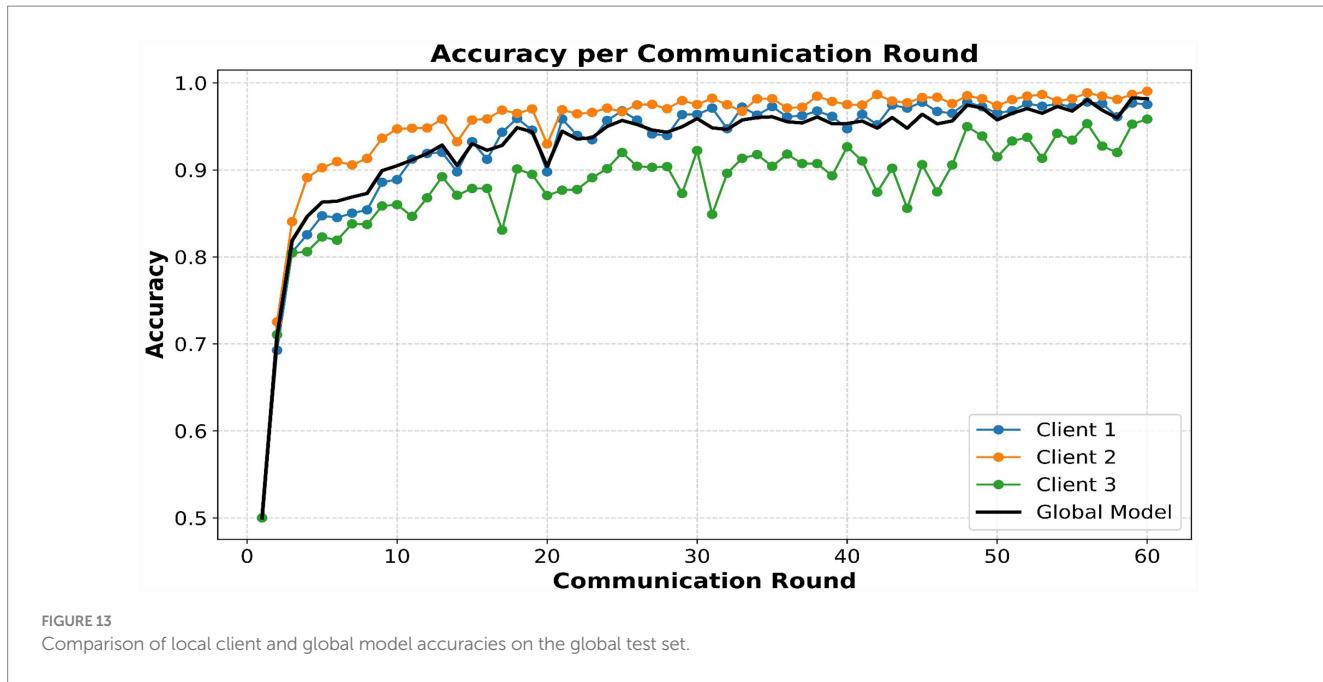


TABLE 5 Comparative model performance on driver drowsiness detection.

Baseline approaches	Model	Accuracy	Precision	Recall	F1-score
Baseline model	LR	60.90%	60.95%	60.90%	60.88%
Baseline model	SVM	72.18%	73.06%	72.25%	71.95%
Ahmad et al. (2024a)	LSTM	81.95%	82.21%	81.98%	81.92%
Ahmad et al. (2024b)	1D-CNN	88.35%	88.35%	88.35%	88.35%
Ahmad et al. (2025b)	EfficientNetB0	97.63%	97.64%	97.63%	97.63%
	DenseNet201	97.45%	97.45%	97.45%	97.45%
	ConvNeXtTiny	96.81%	96.82%	96.81%	96.81%
Ahmad et al. (2025a)	DDD-GC-ViT	97.90%	97.91%	97.90%	97.90%
Proposed study	DrowsyXnet (CW-FedAvg)	98.29%	98.29%	98.29%	98.29%

performance can be achieved without data sharing. This provides privacy-preserving deployment in real-world settings.

### 4.3 Federated learning under heterogeneous client conditions

The dataset was used to run three non-IID scenarios to evaluate the federated framework's handling of inter-client differences. The first introduced a strong class imbalance by limiting one client to a 1:20 drowsy-normal ratio. The second mimicked missing information by randomly masking either the RPM or throttle channel for a client. The third combined both imbalance and channel masking to create a more challenging scenario. The tests examined whether FedAvg, FedProx, and the proposed CW-FedAvg can deal with uneven and incomplete client data. Across all scenarios CW-FedAvg appeared more resilient keeping accuracy within about 0.6% of the balanced baseline. FedAvg showed a larger drop under strong class skew and feature removal while FedProx remained more stable as shown in Table 8. These results recommend that weighting clients according to data volume and class

distribution helps reduce bias and keeps the model fairly reliable. Even though individual clients differ significantly in data quality or class availability. The patterns also suggest CW-FedAvg fits real-world deployments where client datasets are rarely uniform.

### 4.4 Robustness and statistical significance

All experiments were repeated across three random seeds to check consistency and reduce the effect of random initialization. The model's performance stayed stable across the runs. Metric variance remained below 0.3%, which shows consistent convergence and effect from seed choice. McNemar tests were applied to determine the performance differences and statistical evaluation. Comparison between the centralized model and FedAvg provide a *p*-value of 0.018 which showing a statistically significant difference. The gap between the centralized model and CW-FedAvg was not significant, with a *p*-value of 0.12. CW-FedAvg remained close to the centralized model performance while addressing privacy concerns. The comparison between FedAvg and CW-FedAvg yielded a *p*-value of 0.037,

TABLE 6 Lightweight backbone model comparison.

Model	Params (M)	Inference time (ms/sample)	Accuracy (%)	Suitability
ResNet-50	23.5	4.9	98.29	Best accuracy
EfficientNet-B0	5.3	3.1	97.9	Balanced
MobileNet V3-Small	2.9	2.4	96.4	Edge devices

TABLE 7 Performance of DrowsyXnet under different training approaches.

Method	Test accuracy (%)	F1-score (%)	MCC	Cohen's $\kappa$
Centralized (pooled)	98.67	98.65	0.968	0.967
FedAvg	97.89	97.74	0.955	0.953
FedProx ( $\mu = 0.1$ )	98.05	98.00	0.960	0.959
CW-FedAvg (proposed)	<b>98.29</b>	<b>98.29</b>	<b>0.966</b>	<b>0.965</b>

TABLE 8 Heterogeneous client simulation results.

Condition	FedAvg (%)	FedProx (%)	CW-FedAvg (%)	Observation
Balanced clients	98.0	98.1	98.3	Reference
Skewed 1:20 class ratio	96.8	97.4	97.8	CW-FedAvg mitigates imbalance
Masked channels	96.9	97.5	97.9	Robust with missing features
Skew + mask	96.4	97.2	97.7	Most challenging case

TABLE 9 Confidence intervals and significance tests.

Comparison	Accuracy (%)	95% CI	p-value	Interpretation
Centralized vs. FedAvg	98.67/97.80	[98.4–98.9] / [97.5–98.1]	0.018	Statistically significant
Centralized vs. CW-FedAvg	98.67/98.29	[98.4–98.9] / [98.0–98.5]	0.12	Not significant
FedAvg vs. CW-FedAvg	97.80/98.29	[97.5–98.1] / [98.0–98.5]	0.037	Statistically significant

indicating that CW-FedAvg is better than standard FedAvg. Overall, the findings indicate that the proposed CW-FedAvg approach provides performance that is stable and reproducible across different runs and experiments. At the same time, it appears effective at reducing the performance gap between centralized training and federated training and maintaining results close to the upper bound even under varied client conditions, as demonstrated in Table 9. Minor fluctuations observed across different random seeds highlight the inherent variability in model behavior. The results confirm that CW-FedAvg provides a significant improvement over FedAvg, and the performance difference compared to the centralized model is not statistically significant.

#### 4.5 Interpretation of MFCC features

The evaluation shows that DrowsyXnet reliably predicts both drowsy and normal driver states. SHAP (SHapley Additive exPlanations) was used to make the model more interpretable and to illustrate the contribution of individual features to each prediction. Each input including speed, RPM, throttle position and steering torque represented as MFCC-based images and given a value showing its influence on the final output. The SHAP summary and dependence plots in Figure 14 show patterns of feature importance and

interactions, providing insight into the model's sensitivity. Using SHAP makes the model more transparent and builds trust while showing which behavioral indicators are most important for detecting driver drowsiness in safety-critical situations.

#### 4.6 Discussion and limitations

The results suggest that the DrowsyXnet model trained with the CW-FedAvg strategy performs reliably even under strict data privacy constraints. The global accuracy of 98.29% comes close to the 98.67% achieved with centralized training. This shows that federated learning can capture most of the predictive power of pooled-data systems without sharing raw data. Comparing local and global models shows the effect of federated collaboration. Clients with limited or imbalanced data gain from the knowledge shared across all clients. Client 3 had fewer data points and exhibited lower accuracy. The data distribution was skewed, and performance variance was higher. The local models gradually aligned with the global trend as communication rounds progressed. Overall, the results indicated that CW-FedAvg strategy preserve privacy and offer a consistent and performance across diverse client datasets. FedAvg showed a noticeable drop in accuracy, F1-score, MCC, and Cohen's  $\kappa$ . FedProx reduced some of the negative effects caused by non-IID data. The proposed CW-FedAvg

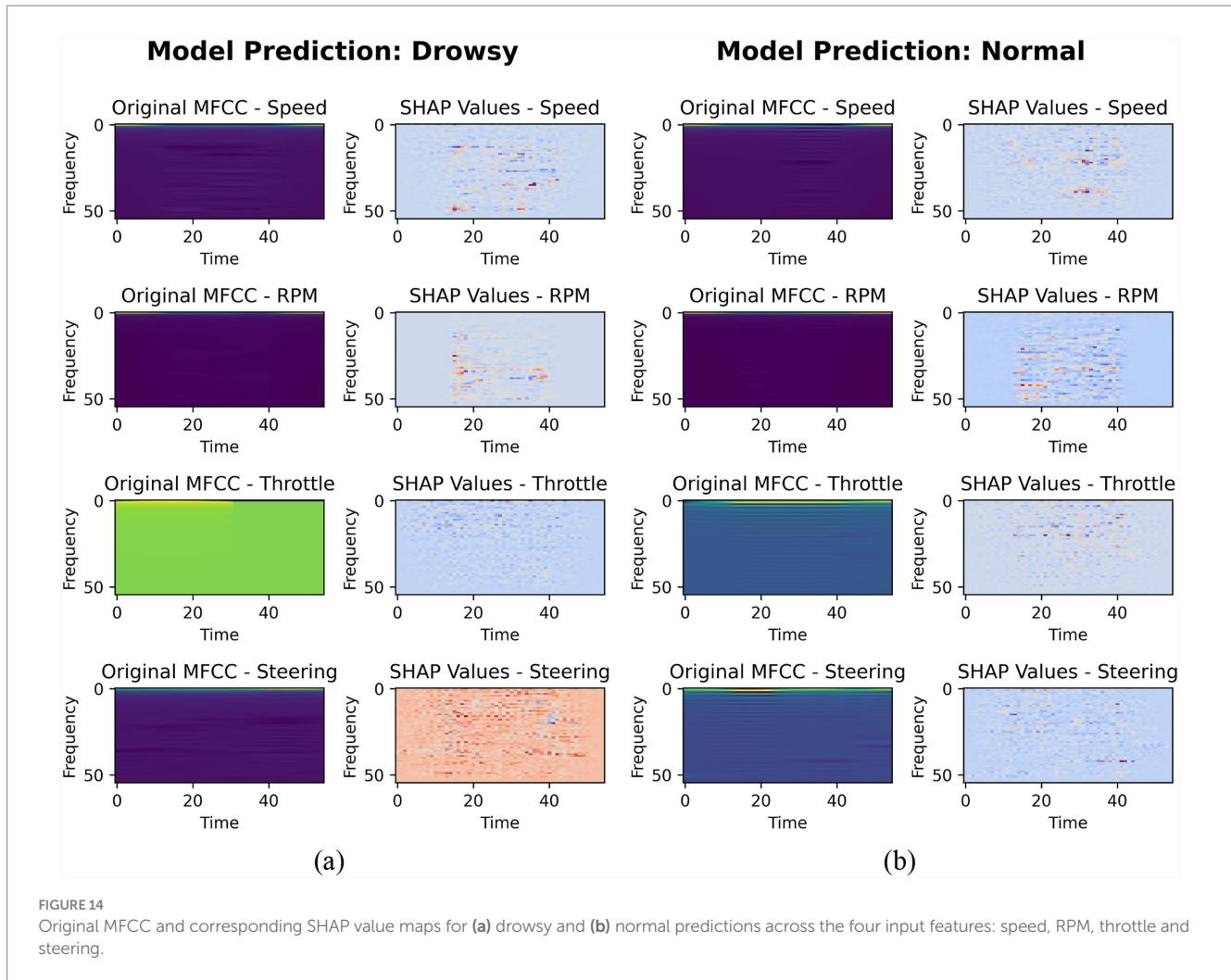


FIGURE 14

Original MFCC and corresponding SHAP value maps for (a) drowsy and (b) normal predictions across the four input features: speed, RPM, throttle and steering.

achieved the best balance across all metrics. Weighting updates according to class distribution prevents clients with dominant data from having excessive influence on global learning. Testing with clients who have different data conditions gives more insight into the model's performance. With class imbalance or missing sensor channels and combined non-IID distortions, CW-FedAvg maintained accuracy within about 0.6% of the balanced baseline. In comparison, FedAvg dropped more sharply, and FedProx provided only intermediate resilience. SHAP-based interpretability analysis offers additional confidence in the framework's reliability in real-world applications. The visualizations highlighted meaningful spectral patterns in the MFCC representations of speed, RPM, throttle, and steering position.

Comparing lightweight backbones provides useful information for real-time deployment, and ResNet-50 remains the most accurate architecture. EfficientNet-B0 and MobileNet V3-Small performed competitively while running about twice as fast and using far fewer parameters. The framework could be adapted for embedded automotive hardware. It allows a balance between accuracy and latency depending on the application. Statistical checks add further context to these findings. McNemar tests show that CW-FedAvg performs better than FedAvg while showing no significant difference from the centralized upper bound. This provides confidence for practical deployment in real-world driving scenarios. Despite the

strengths, several limitations are worth noting. The dataset includes a relatively small group of drivers and mostly controlled driving conditions. The findings are limited in their application to more diverse real-world scenarios such as night driving, heavy traffic, or adverse weather. Collecting data from different regions, vehicle types, and driver populations could improve robustness. MFCC-based feature extraction improves performance but adds computational overhead. This can be challenging for ultra-low-power edge devices, especially with deep convolutional backbones. Future work could explore more efficient spectral encoders or hardware-aware neural architectures. These approaches would help reduce the computational load. Adding temporal attention mechanisms or sequence-level federated models could help address this limitation.

Finally, federated learning improves privacy and practical deployment still faces hurdles such as device availability, communication delays and client dropout and limited bandwidth. Addressing these system-level challenges is important for smooth operation in commercial intelligent transportation systems. Future studies will need to carefully consider these real-world factors. Overall, the findings show that combining federated learning with data-aware aggregation and spectral representations and explainable AI provides a practical and privacy-conscious way to detect driver drowsiness. The CW-FedAvg framework provides high accuracy and handles non-IID conditions well. It also makes the decision process

more transparent. These qualities make it a useful foundation for real-world applications. Safety, scalability, and privacy are all critical considerations.

## 5 Conclusion

The study proposed a framework for detecting driver drowsiness using OBD-II sensor data and MFCC transformations with federated deep learning. The DrowsyXnet model used a pretrained ResNet-50 backbone and fine-tuned for MFCC image of OBD-II data. The DrowsyXnet model predicts drowsy and normal driver states effectively. In federated deep learning model is trained within a CW-FedAvg framework, it achieved an overall accuracy of 98.29% on the global test set. The achieved performance is consistent and promising across multiple drivers while keeping raw data private. In the centralized approaches model, a single client car data, the federated setup trains models locally and aggregates updates securely on the server. This reduces the impact of non-IID data and class imbalance in different client car data. The model maintained high precision, recall, and F1-scores for both classes. This global model can detect inconsistencies in driver behavior and vehicle dynamics that indicate drowsiness. Compared with other models, traditional classifiers such as Logistic Regression and SVM struggled with OBD-II signals, achieving accuracies of 60.90 and 72.18%. Time-series favorable models such as LSTM and 1D-CNN performed better but had difficulty capturing useful features for drowsiness detection. Modern deep learning architectures, including EfficientNetB0, DenseNet201, ConvNeXtTiny, and DDD-GC-ViT, achieved accuracies exceeding 96%, underscoring the advantage of deeper networks for the task. Overall, transforming time-series OBD-II signals into MFCC images and using transfer learning improves feature extraction. Furthermore, combining privacy-preserving federated learning with explainable AI provides the best solution for real-world driver safety and reliability for drowsiness detection in vehicles. Future work will focus on expanding the dataset to cover a wider range of drivers, road types, and environmental conditions. It will also investigate lighter transformer-based or hybrid architectures to enhance inference efficiency and reduce computational load in safer and more reliable vehicles. This framework provides a solution for secure driving and reducing drowsiness-related accidents. It can contribute to improved road safety and privacy-based intelligent transportation systems.

## 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 humans were approved by the Research Ethics Committee (REC) and the Technology Transfer Office (TTO) at Multimedia University, Malaysia, Ethics Approval Reference: TTO/REC/EA/047/2021, Approval Number: EA0472022; Approval Date: 21 July 2022. The studies were conducted in accordance with the local legislation and institutional requirements. The 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

KA: Software, Writing – review & editing, Conceptualization, Writing – original draft, Methodology, Data curation. PE: Writing – review & editing, Project administration, Funding acquisition, Conceptualization, Supervision. NA: Validation, Investigation, Project administration, Writing – review & editing, Supervision, Visualization.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. The research in this paper was supported by the Malaysian Ministry of Higher Education (MOHE) through the Fundamental Research Grant Scheme (FRGS/1/2022/TK0/MMU/02/13).

## Acknowledgments

The authors acknowledge the contributions of Telekom Malaysia (TM) and Multimedia University (MMU) for their valuable support and collaboration throughout the research project. The assistance of these authorities was crucial in achieving the objectives of this study.

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## 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.

## References

Ahmad, K., Em, P. P., and Ab Aziz, N. A. (2024a). Leveraging OBD II time series data for driver drowsiness detection: a recurrent neural networks approach. In *2024 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)* (518–523). IEEE.

Ahmad, K., Em, P. P., and Ab Aziz, N. A. (2024b). “Utilizing OBD II time series data for driver drowsiness detection: a one-dimensional CNN approach” in *2024 multimedia university engineering conference (MECON)* (Cyberjaya, Malaysia: IEEE), 1–6.

Ahmad, K., Em, P. P., and Aziz, N. A. A. (2023). Machine learning approaches for detecting driver drowsiness: a critical review. *Int. J. Membr. Sci. Technol.* 10, 329–346. doi: 10.15379/ijmst.v10i1.1815

Ahmad, K., Ping, E.P., and Ab Aziz, N.A. 2025a. Enhancing driver drowsiness detection using OBD-II and MFCC features: a transformer-based approach. In: *Proceedings of the 2025 Multimedia University Engineering Conference (MECON)*, Cyberjaya, Malaysia, 1–6.

Ahmad, K., Ping, E.P., and Ab Aziz, N.A. 2025b. Harnessing transfer learning for multimodal driver drowsiness detection: CNN-based pretrained models. In: *Proceedings of the 2025 Multimedia University Engineering Conference (MECON)*, Cyberjaya, Malaysia, 1–6.

Albadawi, Y., AlRedhaei, A., and Takruri, M. (2023). Real-time machine learning-based driver drowsiness detection using visual features. *J. Imaging* 9:91. doi: 10.3390/jimaging9050091

Albadawi, Y., Takruri, M., and Awad, M. (2022). A review of recent developments in driver drowsiness detection systems. *Sensors* 22:2069. doi: 10.3390/s22052069

Alves, A. A. C., Andrietta, L. T., Lopes, R. Z., Bussiman, F. O., Silva, F. F. E., Carvalheiro, R., et al. (2021). Integrating audio signal processing and deep learning algorithms for gait pattern classification in Brazilian gaited horses. *Front. Anim. Sci.* 2:681557. doi: 10.3389/fanim.2021.681557

Arceda, V.M., Nina, J.C., and Fabian, K.F. (2020). A survey on drowsiness detection techniques. In *Iberoamerican Conference of Computer Human Interaction, Arequipa, Perú*. 15,

Arefnezhad, S., Samiee, S., Eichberger, A., and Nahvi, A. (2019). Driver drowsiness detection based on steering wheel data applying adaptive neuro-fuzzy feature selection. *Sensors* 19:943. doi: 10.3390/s19040943

Arif, S., Munawar, S., and Ali, H. (2023). Driving drowsiness detection using spectral signatures of EEG-based neurophysiology. *Front. Physiol.* 14:1153268. doi: 10.3389/fphys.2023.1153268

Bacanin, N., Jovanovic, L., Stoean, R., Stoean, C., Zivkovic, M., Antonijevic, M., et al. (2024). Respiratory condition detection using audio analysis and convolutional neural networks optimized by modified metaheuristics. *Axioms* 13:335. doi: 10.3390/axioms13050335

de Naurois, C. J., Bourdin, C., Stratulat, A., Diaz, E., and Vercher, J. L. (2019). Detection and prediction of driver drowsiness using artificial neural network models. *Accid. Anal. Prev.* 126, 95–104. doi: 10.1016/j.aap.2017.11.038

Doudou, M., Bouabdallah, A., and Berge-Cherfaoui, V. (2020). Driver drowsiness measurement technologies: current research, market solutions, and challenges. *Int. J. Intell. Transp. Syst. Res.* 18, 297–319.

Dua, M., Shakshi, Singla, R., Raj, S., and Jangra, A. (2021). Deep CNN models-based ensemble approach to driver drowsiness detection. *Neural Comput. Appl.* 33, 3155–3168.

Ed-Doughmi, Y., Idrissi, N., and Hbali, Y. (2020). Real-time system for driver fatigue detection based on a recurrent neuronal network. *J. Imaging* 6:8. doi: 10.3390/jimaging6030008

Gupta, S., Jaafar, J., Ahmad, W. W., and Bansal, A. (2013). Feature extraction using MFCC. *Signal Image Proc. Int. J.* 4, 101–108.

Gwak, J., Hirao, A., and Shino, M. (2020). An investigation of early detection of driver drowsiness using ensemble machine learning based on hybrid sensing. *Appl. Sci.* 10:2890. doi: 10.3390/app10082890

Harkous, H., and Artail, H. (2019). A two-stage machine learning method for highly-accurate drunk driving detection. In *2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)* (1–6). IEEE.

Hong, M., Kang, S. K., and Lee, J. H. (2022). Weighted averaging federated learning based on example forgetting events in label imbalanced non-iid. *Appl. Sci.* 12:5806. doi: 10.3390/app12125806

Khan, M.A.A., Alsawaf, M., Arab, B., AlHashim, M., Almashharawi, F., Hakami, O., et al. (2022) Road damages detection and classification using deep learning and UAVs. In *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)* (1–6). IEEE.

Kundinger, T., Sofra, N., and Riener, A. (2020). Assessment of the potential of wrist-worn wearable sensors for driver drowsiness detection. *Sensors* 20:1029. doi: 10.3390/s20041029

Lenné, M. G., and Jacobs, E. E. (2016). Predicting drowsiness-related driving events: a review of recent research methods and future opportunities. *Theor. Issues Ergon. Sci.* 17, 533–553. doi: 10.1080/1463922X.2016.1155239

Liang, W., Liang, Y., and Jia, J. (2023). MiAMix: enhancing image classification through a multi-stage augmented mixed sample data augmentation method. *PRO* 11:3284. doi: 10.3390/pr11123284

Ma, Y., Chen, B., Li, R., Wang, C., Wang, J., She, Q., et al. (2019). Driving fatigue detection from EEG using a modified PCANet method. *Comput. Intell. Neurosci.* 2019, 1–9. doi: 10.1155/2019/4721863

Malik, M., and Nandal, R. (2023). A framework on driving behavior and pattern using on-board diagnostics (OBD-II) tool. *Mater Today Proc* 80, 3762–3768. doi: 10.1016/j.matpr.2021.07.376

Martins, N. R. A., Annaheim, S., Spengler, C. M., and Rossi, R. M. (2021). Fatigue monitoring through wearables: a state-of-the-art review. *Front. Physiol.* 12:790292. doi: 10.3389/fphys.2021.790292

McMahan, B., Moore, E., Ramage, D., Hampson, S. Y., and Arcas, B. A. (2017). “Communication-efficient learning of deep networks from decentralized data” in *Artificial intelligence and statistics* (Fort Lauderdale, Florida, USA: PMLR), 1273–1282.

Meireles, T., and Dantas, F. (2019). A low-cost prototype for driver fatigue detection. *Multimodal Technol. Interact.* 3:5. doi: 10.3390/mti3010005

Michailidis, E. T., Panagiotopoulou, A., and Papadakis, A. (2025). A review of OBD-II-based machine learning applications for sustainable, efficient, secure, and safe vehicle driving. *Sensors* 25:4057. doi: 10.3390/s25134057

Ministry of Transport Malaysia (2020) Ministry of Transport Malaysia official portal Malaysia road fatalities index. Available online at: <https://www.mot.gov.my/en/land-safety/malaysia-road-fatalities-index> (Accessed May 21, 2025).

Ministry of Transport Malaysia (2021) Road accident and fatalities. Available online at: <https://www.mot.gov.my/en/land/safety/road-accident-and-facilities> (Accessed May 21, 2025).

Mohammed, K. K., Abd El-Latif, E. I., El-Sayad, N. E., Darwish, A., and Hassanieh, A. E. (2023). Radio frequency fingerprint-based drone identification and classification using mel spectrograms and pre-trained YAMNet neural. *Internet Things* 23:100879. doi: 10.1016/j.iot.2023.100879

Nasri, I., Karrouchi, M., Kassmi, K., and Messaoudi, A., 2022. A review of driver drowsiness detection systems: techniques, advantages and limitations. [Epubh ahead of preprint]. doi: 10.45500/arXiv.2206.07489

Omerustaoglu, F., Sakar, C. O., and Kar, G. (2020). Distracted driver detection by combining in-vehicle and image data using deep learning. *Appl. Soft Comput.* 96:106657. doi: 10.1016/j.asoc.2020.106657

Paultan.org (2022) 255,532 road accidents reported in 2021, 3,302 deaths. Available online at: <https://paultan.org/2022/01/27/255532-road-accidents-jan-sep-2021-3302-deaths> (Accessed May 21, 2025).

Peppes, N., Alexakis, T., Adamopoulou, E., and Demestichas, K. (2021). Driving behaviour analysis using machine and deep learning methods for continuous streams of vehicular data. *Sensors* 21:4704. doi: 10.3390/s21144704

Ping, E.P., and Shie, T.T. 2022. Driver drowsiness detection system using hybrid features among Malaysian drivers: a concept. *Paper Presented at the Multimedia University Engineering Conference (MECON) 2022*.

Purkovic, S., Jovanovic, L., Zivkovic, M., Antonijevic, M., Dolicanin, E., Tuba, E., et al. (2024). Audio analysis with convolutional neural networks and boosting algorithms tuned by metaheuristics for respiratory condition classification. *J. King Saud Univ.* 36:102261. doi: 10.1016/j.jksuci.2024.102261

Reddy, B., Kim, Y.H., Yun, S., Seo, C., and Jang, J., 2017. Real-time driver drowsiness detection for embedded system using model compression of deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (121–128).

Safarov, F., Akhmedov, F., Abdusalomov, A. B., Nasimov, R., and Cho, Y. I. (2023). Real-time deep learning-based drowsiness detection: leveraging computer-vision and eye-blink analyses for enhanced road safety. *Sensors* 23:6459. doi: 10.3390/s23146459

Shaik, M. E. (2023). A systematic review on detection and prediction of driver drowsiness. *Transp. Res. Interdiscip. Perspect.* 21:100864. doi: 10.1016/j.trip.2023.100864

Sharma, N., Sharma, R., and Jindal, N. (2021). Machine learning and deep learning applications-a vision. *Glob. Transit. Proc.* 2, 24–28. doi: 10.1016/j.gltip.2021.01.004

Stankovic, M., Jovanovic, L., Bozovic, A., Budimirovic, N., Zivkovic, M., and Bacanin, N. (2024). Exploring the potential of combining Mel spectrograms with neural networks optimized by the modified crayfish optimization algorithm for acoustic speed violation identification. *Int. J. Hybrid Intell. Syst.* 20, 119–143. doi: 10.3233/his-240006

The Star (2023) Road accident and fatalities statistics. Available online at: <https://www.thestar.com.my/news/nation/2023/06/13/915874-road-accidents-recorded-in-2021-and-2022-says-transport-ministry> (Accessed May 21, 2025).

The Star (2024) Road accident deaths to be published daily Available online at: <https://www.thestar.com.my/news/nation/2024/03/07/road-accident-deaths-to-be-published-daily>

Umair, M., Khan, M. S., Ahmed, E., Baothman, F., Alqahtani, E., Alian, M., et al. (2021). Detection of COVID-19 using transfer learning and grad-CAM visualization on indigenously collected X-ray dataset. *Sensors* 21:5813.

Umair, M., Tan, W.H., and Foo, Y.L., (2024). Optimized 1D convolutional neural network for efficient intrusion detection in IoT networks. In *2024 IEEE 8th International Conference on Signal and Image Processing Applications (ICSIPA)* (1–6). IEEE.

Ustubioglu, A., Ustubioglu, B., and Ulutas, G. (2023). Mel spectrogram-based audio forgery detection using CNN. *Signal Image Video Proc.* 17, 2211–2219.

Vu, T. H., Dang, A., and Wang, J. C. (2019). A deep neural network for real-time driver drowsiness detection. *IEICE Trans. Inf. Syst.* 102, 2637–2641.

Xiao, Y., and bin Abas, A. (2021). A review on fatigue driving detection. *ASP Trans. Internet Things* 1, 1–14.

Xu, X., Xie, Z., Yang, Z., Li, D., and Xu, X. (2020). A t-SNE based classification approach to compositional microbiome data. *Front. Genet.* 11:620143. doi: 10.3389/fgene.2020.620143

Yu, H., Sun, H., Tao, J., Qin, C., Xiao, D., Jin, Y., et al. (2023). A multi-stage data augmentation and AD-ResNet-based method for EPB utilization factor prediction. *Autom. Constr.* 147:104734. doi: 10.1016/j.autcon.2022.104734

Zeng, D., Xu, Z., Liu, S. H. I. Y. U., Pan, Y., Wang, Q., and Tang, X., 2023. On the power of adaptive weighted aggregation in heterogeneous federated learning and beyond. In *The 28th International Conference on Artificial Intelligence and Statistics*.

Zhang, J., Ma, X., Zhang, J., Sun, D., Zhou, X., Mi, C., et al. (2023). Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. *J. Environ. Manag.* 332:117357. doi: 10.1016/j.jenvman.2023.117357

Zhao, Z., Zhou, N., Zhang, L., Yan, H., Xu, Y., and Zhang, Z. (2020). Driver fatigue detection based on convolutional neural networks using em-CNN. *Concurr. Comput.* 2020:e5927. doi: 10.1155/2020/7251280