



## OPEN ACCESS

## EDITED BY

Xiaofeng Li,  
Heilongjiang International University, China

## REVIEWED BY

Ahmed Ali Dawud,  
Jimma University, Ethiopia  
Hendra Kusuma,  
Institut Teknologi Sepuluh Nopember,  
Indonesia

## \*CORRESPONDENCE

Mohamed Atef  
✉ moh\_atef@uaeu.ac.ae

RECEIVED 19 June 2025

ACCEPTED 31 October 2025

PUBLISHED 27 November 2025

## CITATION

Jaya Prakash A, Nasreddine Belkacem A,  
Elfadel IM, Jelinek HF and Atef M (2025)  
Advances in machine and deep learning for  
ECG beat classification: a systematic review.  
Front. Digit. Health 7:1649923.  
doi: 10.3389/fdgth.2025.1649923

## COPYRIGHT

© 2025 Jaya Prakash, Nasreddine Belkacem,  
Elfadel, Jelinek and Atef. 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.

# Advances in machine and deep learning for ECG beat classification: a systematic review

Allam Jaya Prakash<sup>1</sup>, Abdelkader Nasreddine Belkacem<sup>2</sup>,  
Ibrahim M. Elfadel<sup>3</sup>, Herbert F. Jelinek<sup>4</sup> and Mohamed Atef<sup>1\*</sup>

<sup>1</sup>Electrical and Communication Engineering Department, College of Engineering, United Arab Emirates University, Abu Dhabi, United Arab Emirates, <sup>2</sup>Department of Computer and Network Engineering, College of IT, United Arab Emirates University, Abu Dhabi, United Arab Emirates, <sup>3</sup>Center for Cyber-Physical Systems and Department of Computer and Information Engineering, Khalifa University, Abu Dhabi, United Arab Emirates, <sup>4</sup>Department of Medical Sciences, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

The electrocardiogram (ECG) is an important tool for exploring the structure and function of the heart due to its low cost, ease of use, efficiency, and non-invasive nature. With the rapid development of artificial intelligence (AI) in the medical field, ECG beat classification has emerged as a key area of research for performing accurate, automated, and interpretable cardiac analysis. According to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses criteria, we examined a total of 106 relevant articles published between 2014 and 2024. This study investigates ECG signal analysis to identify and categorize various beats with better accuracy and efficiency, by emphasizing and applying vital pre-processing techniques for denoising the raw data. Particular attention is given to the evolution from traditional feature-engineering methods toward advanced architectures with automated feature extraction and classification, such as convolutional neural networks, recurrent neural networks, and hybrid frameworks with attention mechanisms. In addition, this review article investigates the common challenges observed in the existing studies, including data imbalance, inter-patient variability, and the absence of unified evaluation metrics, which restrict fair comparison and clinical translation. To address these gaps, future research directions are proposed, focusing on the development of standardized multi-center datasets, cross-modal fusion of physiological signals, and interpretable AI models to facilitate real-world deployment in healthcare systems. This systematic review provides a structured overview of the current state and emerging trends in ECG beat classification, offering clear insights for researchers and clinicians to guide future advancements in intelligent cardiac diagnostics.

## KEYWORDS

arrhythmia, classification, deep learning, electrocardiogram, feature extraction, machine learning

## 1 Introduction

The electrocardiogram (ECG) signal is a crucial non-invasive tool for diagnosing and monitoring cardiac disorders (1). Its quick and accurate results make it valuable in various clinical settings (1, 2), allowing healthcare providers to assess heart rate (HR), rhythm, and conduction mechanisms (2, 3). An ECG is commonly used to screen patients with risk factors such as hypertension, diabetes, or a family history of heart disease, as minor irregularities may signify a higher risk (4). It also reveals heart size, thickness, and blood supply, helping to detect conditions such as heart failure or

cardiomyopathy. In addition, an ECG is utilized during procedures or serious illnesses to monitor heart function and detect abnormal rhythms, allowing for prompt intervention (2). According to the World Health Organization (WHO), approximately 17.90 million deaths worldwide are caused by cardiovascular diseases (CVDs) each year (5). The most common CVDs (6) are arrhythmia, myocardial infarction (MI), congestive heart failure, rheumatic heart disease, cardiomyopathy, ischaemia, and heart stroke. ECGs play a vital role in diagnosing and monitoring CVDs. Therefore, a timely diagnosis and accurate ECG beat detection in cardiac patients are crucial. Identifying the morphological similarities among the many ECG beats from different classes is difficult when using the naked eye. Therefore, an automated diagnostic tool for ECG beat classification is required (7).

Figure 1 illustrates the primary ECG signal, composed of the following characteristic waves: the P-wave, QRS complex, T-wave, and U-wave (1). These characteristic waves are crucial for identifying the state of the heart's functioning. As per the Physionet arrhythmia database, there are 17 different types of ECG beats. These categories cover a wide range of arrhythmias, aiding in the comprehensive analysis and classification of ECG beats (8). These 17 different types of ECG beats are further sub-categorized into five classes, i.e., non-ectopic beat, supraventricular ectopic beat (S), ventricular ectopic beat (V), fusion beat (F), and unknown beat (Q), as per the American Association for the Medical Instrumentation (AAMI) (8). Among these ECG beats, S and V are clinically crucial, as these are sources of sudden heart attacks (9).

## 1.1 Automated ECG beat classification

ECG recordings are classified into two types based on the recording duration, i.e., resting and ambulatory ECG (10).

A resting ECG contains only 5 to 10 min of heart function data recording, whereas an ambulatory ECG records 24–48 h of information (11). Detecting abnormal episodes from this enormous quantity of data is very difficult. Therefore, effective automated diagnosis tools are required to detect important episodes in cardiac patients. Initially, we used template-based and rule-based techniques that are usually utilized to detect the type of ECG beat (12–17). Rule-based approaches rely on predefined rules and thresholds to classify ECG beats. These rules are often based on expert knowledge or heuristics. However, these rules may not be able to handle the wide range of variations and complexities observed in real-world ECG signals. As a result, rule-based approaches may struggle to adapt to different types of beats or handle new patterns that were not considered during rule creation (18). Developing accurate and comprehensive rules for ECG beat classification can be challenging. It requires a deep understanding of ECG signal characteristics and considerable domain expertise. Designing rules that cover all possible scenarios and variations is complex and time-consuming. Rule-based approaches are typically designed to classify beats based on specific features or patterns (19). They may struggle to generalize well to new or unseen data that do not conform to the predefined rules. The rule-based classifier can produce incorrect or inconsistent results if the ECG data deviates from the expected patterns. A set of templates representing different types of beats is required in template-based approaches (13). Choosing appropriate templates that accurately represent the various beat morphologies in ECG signals can be challenging. There is a need to consider inter-subject and intra-subject variability and variations due to different conditions and diseases (12). The classification in template-based approaches relies on comparing the input ECG beat with a set of templates to find the best match. However, template matching can be sensitive to noise, baseline wander (BW), and other artifacts present in the signal (20). These issues

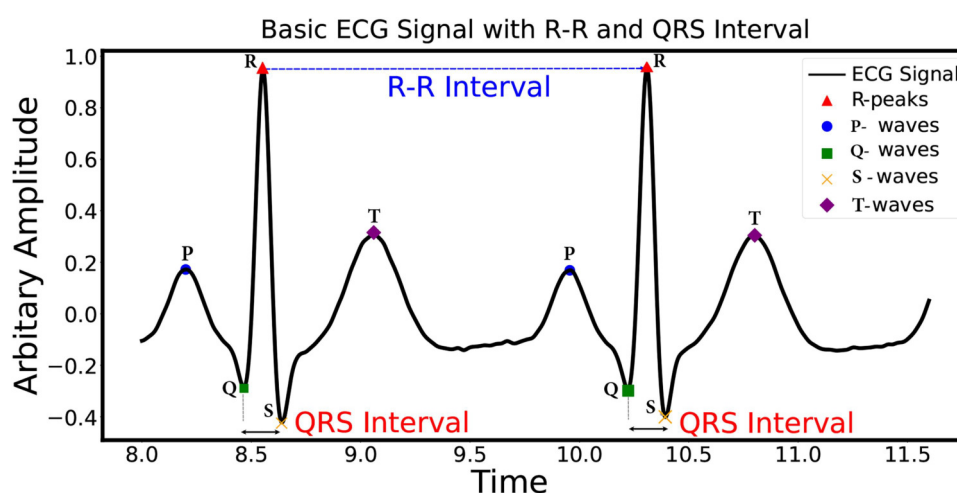


FIGURE 1  
Basic ECG signal and characteristic wave representation.

can affect the accuracy of the match and lead to misclassification. Template-based approaches often struggle with scalability when dealing with large datasets or real-time applications (17). Comparing each beat with a set of templates can be computationally expensive, especially if the number of templates is high. The time complexity increases as the number of beats and templates grows, making it impractical for large-scale applications (21).

## 1.2 Machine and deep learning for ECG beat classification

Machine learning (ML) algorithms are becoming popular for effective classification of ECG beats (18, 19, 22–28). Hierarchical representations of the latest ECG beat classification techniques based on machine and deep learning (DL) techniques have been reported in the literature, as shown in Figure 2. Handcrafted features are required to detect the type of ECG beat when using machine learning techniques (22–24). Most handcrafted features are extracted based on time, frequency, and time-frequency domains (29, 30). Time-domain features alone are not sufficient for effective ECG beat classification; together with this, the frequency-domain features increase beat detection performance (11, 29). In addition, features based on time and frequency are more effective than individual features in the time and frequency domain (31). Handcrafted approaches rely on manual feature engineering, where domain knowledge is used to design and extract features from ECG signals. This process can be time-consuming, involving the design of algorithms and signal processing techniques to extract meaningful features (5). Handcrafted approaches require selecting relevant features that capture the discriminative information from the ECG signals. Identifying the most informative and robust features is non-trivial and often requires domain expertise (32). Choosing inappropriate features or excluding important ones can lead to suboptimal classifier performance. Handcrafted feature sets may not generalize well to new or unseen data that significantly

differ from the training data. ECG signals vary considerably due to age, sex, underlying conditions, and noise (33). The classification accuracy may be compromised if the handcrafted features do not represent the new data. ECG beat classification involves capturing complex relationships and patterns within signals. Handcrafted approaches may struggle to capture these intricate relationships, as they typically rely on pre-defined algorithms and feature engineering techniques. These problems have motivated researchers to develop automatic feature extraction approaches.

Deep learning models can automatically learn relevant features directly from the raw ECG signals, eliminating the need for manual feature engineering. This feature learning process enables the model to capture intricate patterns and complex relationships that may be difficult to capture using handcrafted features (18, 19, 25–28). Deep learning models can learn hierarchical representations of ECG signals, allowing them to extract meaningful features at different levels of abstraction. The different deep learning architectures and their advantages in the extraction of meaningful features are presented in Table 1. Deep learning models are known for generalizing unseen data well. They can learn from large amounts of labeled ECG data and capture the underlying patterns that are characteristic of different beat types. Deep learning models can handle large-scale ECG datasets efficiently (20). Once trained, the models can process ECG beats quickly, making them suitable for real-time applications. In addition, deep learning models can be deployed on parallel computing architectures, such as graphical processing units (GPUs), to improve computational performance, enabling rapid and scalable ECG beat classification (9, 18). Deep learning models can be updated and fine-tuned with new data to improve performance and adapt to changes in ECG signals (25). This ability for continual learning allows the model to incorporate new knowledge and adjust its classification capabilities as new data become available. It enables the model to stay up-to-date with the emerging beat types or changes in the data distribution (28). Deep learning models can adapt and generalize well to new beat types and variations not encountered

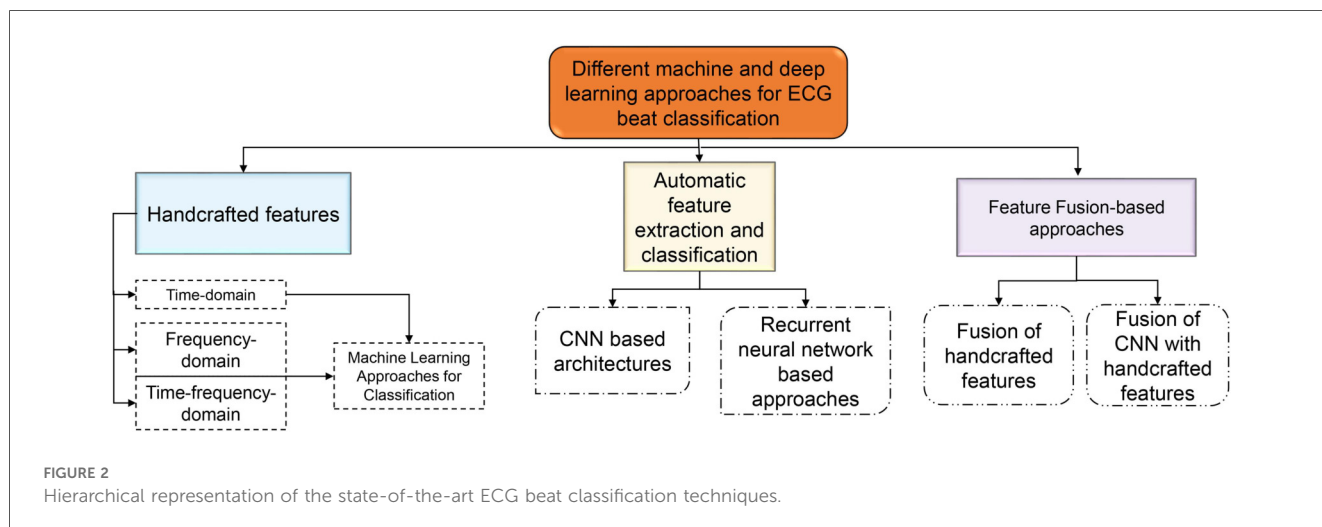


TABLE 1 Hierarchy of different deep learning architectures for ECG beat classification.

| Architecture                           | Layers                                         | Advantages                                                                                 | Disadvantages                                                          | Remarks                                                          |
|----------------------------------------|------------------------------------------------|--------------------------------------------------------------------------------------------|------------------------------------------------------------------------|------------------------------------------------------------------|
| Convolutional neural network (CNN)     | Convolutional, pooling, fully connected layers | Excellent for spatial feature extraction; robust against noise with adequate preprocessing | Requires large labeled datasets; performance degrades on noisy data    | Suitable for clean, structured ECG datasets                      |
| Recurrent neural network (RNN)         | Recurrent layers (e.g., GRU, LSTM)             | Captures sequential dependencies; handles temporal features effectively                    | Prone to vanishing gradient problems; slow training process            | Useful for time-series analysis but requires gradient management |
| Long short-term memory (LSTM)          | Input, forget, output gates                    | Solves vanishing gradient problem; excellent for long-term dependencies                    | Computationally expensive; sensitive to hyperparameter tuning          | Ideal for sequential tasks such as ECG signal interpretation     |
| Bi-directional LSTM (BiLSTM)           | Forward and backward LSTM layers               | Captures past and future temporal features                                                 | High computational overhead; challenging for large datasets            | Performs well in tasks requiring bidirectional dependencies      |
| Hybrid CNN-LSTM                        | Convolutional layers + LSTM layers             | Combines spatial and temporal feature learning                                             | High computational cost; pre-processing (e.g., QRS detection) required | Effective for complex ECG signal classification                  |
| Dense convolutional network (DenseNet) | Dense convolutional blocks, transition layers  | Efficient feature reuse; excellent for structured data                                     | High memory requirement due to dense connections                       | Ideal for datasets with detailed temporal-spatial patterns       |
| Residual networks (ResNet)             | Residual blocks with skip connections          | Solves degradation problem; enables deeper networks                                        | Requires careful tuning for small datasets                             | Useful for capturing subtle variations in ECG signals            |
| Autoencoders                           | Encoder-decoder architecture                   | Learns unsupervised representations; useful for anomaly detection                          | Limited interpretability; prone to over-fitting                        | Ideal for anomaly detection in ECG signals                       |
| Attention mechanism                    | Attention layers on top of RNN or CNN          | Focuses on critical parts of the ECG signal                                                | Computationally expensive; complex to train                            | Effective for applications requiring interpretability            |
| Transformer networks                   | Multi-head attention, positional encoding      | Handles long-range dependencies; scalable for large datasets                               | Requires large datasets and high computational resources               | Promising for real-time, large-scale ECG classification          |

during training. The models can discover and classify novel patterns and variations by learning from diverse examples. This adaptability makes deep learning models suitable for dynamic environments where ECG data may evolve.

The remainder of this article is organized as follows. Section 2 discusses the main objectives and methodology of the proposed study. Section 3 refers to the publicly available databases for experimentation on ECG beat classification. The background and significance of ECG beat classification are discussed in Section 4, with pre-processing, feature extraction techniques, and ECG beat classification algorithms, along with their performance, elaborated on in Sections 4.1–4.3. A general discussion of machine learning and deep learning for the detection of ECG beats is presented in Section 5. Finally, limitations, future directions, recommendations from the state-of-the-art review, and conclusions are presented in Sections 6, 8, and 9.

## 2 Systematic review protocol

The primary focus of this research was to thoroughly explore the extensive range of machine learning and deep learning methodologies employed in the context of ECG beat classification. Arrhythmia is a condition characterized by irregular heartbeats, either too fast or too slow, that disrupt the normal functioning of the heart (2). The heart has a natural pacemaker that sets the rhythm of the heartbeats. However, any disturbances in electrical impulses can result in arrhythmia. There are many cardiac arrhythmias, each with unique characteristics and potential complications. Early detection and treatment of arrhythmias can help prevent serious complications such as stroke or sudden cardiac arrest. Significant cardiac arrhythmias include (i) atrial fibrillation (AF), (ii) ventricular

tachycardia (VT), (iii) sinus bradycardia (SB), (iv) atrial flutter (AFT), (v) ventricular fibrillation (VF), and (vi) supraventricular tachycardia (SVT) (34). An irregularity in the upper chambers of the heart causes AF, which in turn causes blood clots, stroke, and heart failure (35). A fast heart rate that starts in the ventricles (the lower chambers of the heart) will produce VT beats. It can cause dizziness, chest pain, and fainting. If left untreated, it can also lead to sudden cardiac arrest. SB is a slow heart rate that originates in the sinoatrial node (SA). It can cause fatigue, dizziness, and fainting. AF is an irregular heart rhythm caused by electrical activity in the atrial chambers of the heart. The atrial rate is generally between 250 and 350 beats per minute and is quick and regular. Since the atrioventricular (AV) node slows down electrical impulses, the ventricles (the lower chambers of the heart) can also beat rapidly (34).

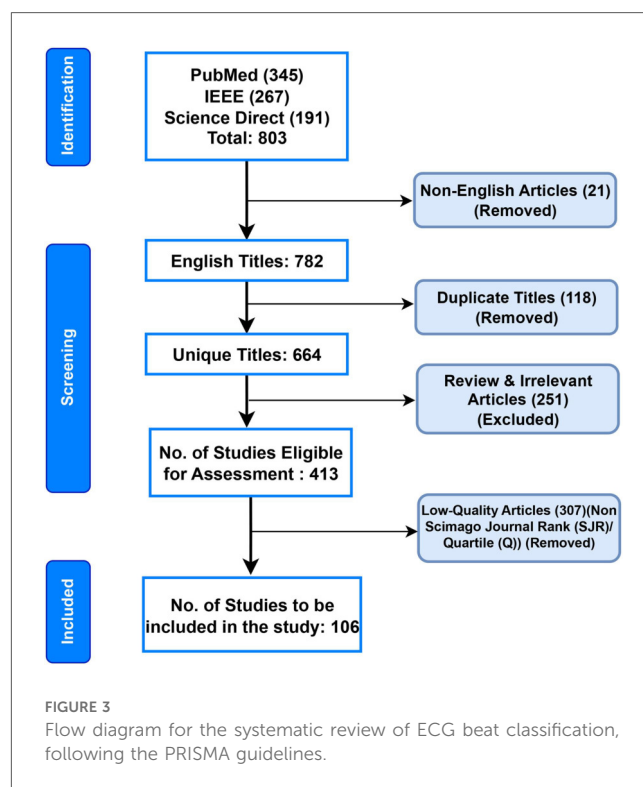
The study investigates ECG signal analysis to identify and categorize various ECG beats with better accuracy and efficiency. The investigation emphasizes vital pre-processing techniques for denoising the raw ECG data to achieve this objective. The removal of unwanted noise ensures that subsequent classification algorithms can work with high-quality input, ultimately leading to more robust and reliable results. These pre-processing steps are essential in enhancing the performance of the classification models reviewed in this study (21, 29, 36, 37). Furthermore, the review study in this work explores various feature extraction and selection techniques (11, 35, 36). These methods are fundamental in transforming the raw ECG signals into a set of discernible and informative features that the classification algorithms can effectively utilize. By studying and comparing the various feature extraction methods available in the literature, this study aimed to identify the most relevant and influential features in the classification of ECG beats, with the aim of optimizing the precision and efficiency of the classification

process (19, 20, 33). This study also surveys and discusses the existing literature, critically assessing the performance of the different machine and deep learning techniques employed in ECG beat classification. By evaluating the strengths and weaknesses of these methods, the research aims to provide valuable information on the most promising approaches and their potential applications in real-world scenarios, such as arrhythmia detection and cardiac health monitoring (2). Moreover, this investigation seeks to contribute to the broader field of biomedical signal processing by presenting a comprehensive overview of the state-of-the-art techniques for ECG beat classification (18, 19, 25–27, 38–40). By consolidating and presenting this knowledge, researchers and practitioners can better understand the most effective methodologies available and further advance the field's capabilities.

## 2.1 Search strategy: inclusion and exclusion criteria

We used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) criteria to find studies pertinent to the classification of ECG beats (41). Articles published up until April 2024 utilizing the terms “Electrocardiogram,” “ECG beat Classification,” “Artificial Intelligence,” “Machine Learning,” and “Deep Learning” in their respective Boolean combinations were searched for in the PubMed, Institute of Electrical and Electronics Engineers Association (IEEE), and Science Direct databases. The authors excluded non-English publications, duplicate titles, irrelevant works, review articles, pilot studies, non-accessible articles, and articles published before 2014 (42). Thus, this study consisted of 106 articles that focused on AI-based ECG beat classification. Figure 3 shows the detailed search method, which was compliant with the PRISMA criteria (42). In Figure 4, it can be observed that the yearly article publication trend (2014–2024) and yearly citation trend (2014–2024) for ECG beat classification indicate a strong and growing research interest in this domain. The publication trend shows a steady increase in the number of articles, which peaked in recent years (2022–2024), highlighting the expanding focus on deep learning, signal processing, and patient-specific classification approaches. This growth suggests an increasing number of researchers contributing to advancements in ECG analysis. A peak in citations often follows high publication periods, indicating that time is required for newer methods to be widely cited. These trends emphasize the need for continued innovation and the adoption of novel approaches to sustain research in ECG beat classification.

To ensure the inclusion of high-quality and peer-reviewed research, the study employed a quality-filtering criterion based on the Scimago Journal Rank (SJR) and Journal Citation Reports (JCR) quartile (Q) indexing systems. In this review, only articles published in journals indexed in SJR or those with JCR quartile rankings (Q1–Q4) were considered eligible for inclusion. Studies published in journals not indexed in either SJR or JCR, or lacking identifiable quality metrics, were classified as “Non-SJR/Q” and consequently excluded. The SJR metric reflects the scientific



influence of scholarly journals by accounting for both the number of citations received and the prestige of the citing journals, while the JCR-Q system ranks journals from Q1 (highest impact) to Q4 (lowest) based on citation distributions. This quality screening ensured that the included literature represented peer-reviewed, credible, and widely recognized sources within the scientific community. Furthermore, conference papers, pilot studies, non-English articles, and inaccessible manuscripts were excluded to maintain methodological rigor and focus on reproducible, peer-reviewed work.

Figure 5 summarizes the most commonly reported performance indicators, namely, accuracy, precision, recall, and F-Score. It is acknowledged that these metrics can be misleading under severe class imbalance. Accuracy, in particular, tends to overestimate performance when normal beats dominate the dataset. Although several reviewed studies reported more robust and threshold-independent metrics, such as the Matthews correlation coefficient (MCC) and area under the precision-recall curve (AUPRC), these measures were not consistently available across all publications, preventing their inclusion in the aggregated summary plots. To maintain comparability across the 106 reviewed works, we therefore only visualized the universally reported metrics. Nevertheless, we recognize that MCC and AUPRC provide a more balanced and informative assessment of classifier performance, especially for minority arrhythmic classes such as supraventricular ectopic beat (SVEBs) and ventricular ectopic beat (VEB). This limitation highlights the need for future ECG classification studies to adopt standardized, imbalance-aware metrics to enable more equitable and clinically meaningful performance comparisons. Initially, machine learning methods

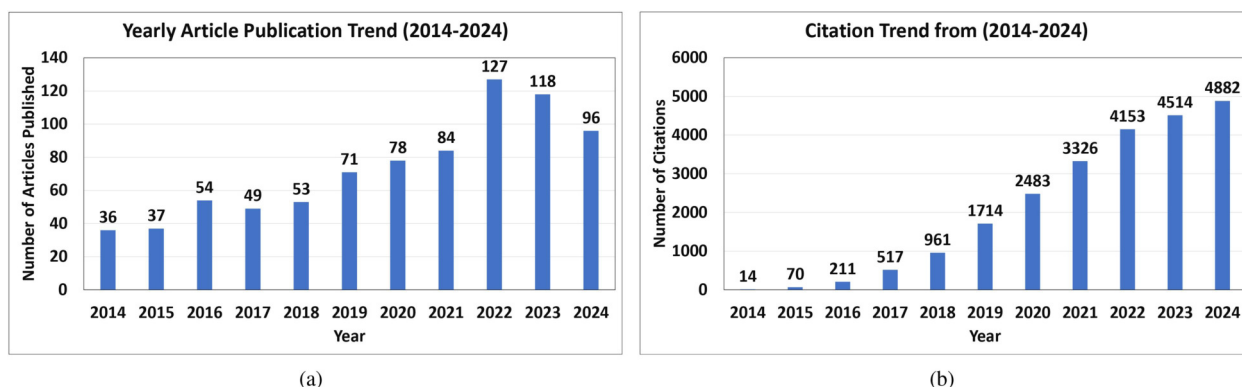


FIGURE 4  
(a) Article and (b) citation trends for ECG beat classification from 2014 to 2024.

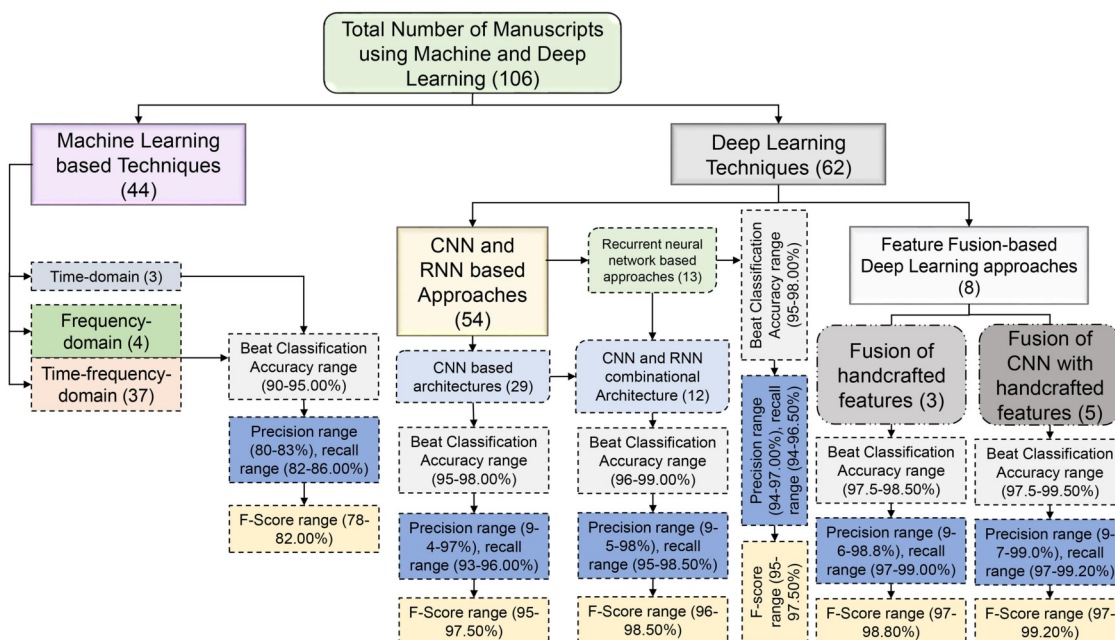


FIGURE 5  
Manuscript details and performance ranges of different machine and deep learning techniques.

relied heavily on manual feature extraction, where domain experts would identify relevant characteristics of ECG signals, such as QRS complexes, heart rate variability, and waveform shapes (29, 33, 43–46). These handcrafted features were then fed into classifiers, such as support vector machine (SVM), random forest (RF), and k-nearest neighbors (kNN), which, despite their effectiveness, were often limited by the quality and completeness of the extracted features. With the advent of deep learning, the process became more automated and data-driven, allowing models such as CNNs and RNNs to learn directly from raw ECG data (25–27). These networks captured complex temporal dependencies and subtle morphological variations without explicit feature engineering, resulting in more robust and accurate ECG

beat classification systems. This shift improved classification performance and opened new possibilities for real-time and patient-specific ECG analysis.

### 3 Data sources

In the literature, the majority of the studies have used three different ECG databases to verify how effectively ECG beat classification methods perform, namely, the Massachusetts Institute of Technology—Beth Israel Hospital (MIT-BIH) arrhythmia (8), St. Petersburg Institute of Cardiological Technics (INCART) (8), and MIT-BIH supraventricular

arrhythmia (3) databases. A summary of these three databases is shown in Table 2. Table 3 presents a detailed breakdown of the manuscripts that utilized various ECG beat classification databases. It further highlights the use of the three prominent databases—MIT-BIH Arrhythmia, INCART, and MIT-BIH supraventricular—across three major publishers: Elsevier, IEEE, and Springer. In addition, it includes lesser-known databases such as the American Heart Association (AHA) and MIT-BIH Long-Term ECG, which are used for specialized research in arrhythmia detection. This table helps to identify trends in database usage for ECG classification in the scientific community.

**MIT-BIH Arrhythmia Database (8):** The MIT-BIH arrhythmia database includes both typical and abnormal cardiac rhythms. This database is widely considered the gold standard in heartbeat detection and classification. There are 48 ambulatory ECG recordings, each lasting 30 min and collected from 47 subjects (8). The first 23 recordings are connected to standard clinical recordings, whereas the rest feature potentially fatal cardiac arrhythmias. Each ECG signal is collected at 360 Hz in these data (8). There are two information streams in every recording, with the first channel's signal (MLII) being of higher quality than the second (V5). There are around 1,09,000 heartbeats in the collection, annotated with 16 distinct labels. Four records (102, 104, 107, and 217) are of inadequate quality out of 48 ECG records (8). As a result, the classification efficiency is calculated without including these files. The ECG signals in the MIT-BIH Arrhythmia Database were recorded from different patients, and even for the same patient, the signal characteristics can vary over time. This can make it challenging to develop a generalized classification system that can perform well on new patients (8).

**St Petersburg INCART Arrhythmia Database (47):** One popular dataset for training and testing ECG classification

algorithms is the 12-lead arrhythmia dataset from the St. Petersburg INCART database. Arrhythmias, ischaemic heart disease, myocardial infarction, and other cardiac diseases are represented in the seventy-five 15 min ECG recordings from 25 patients (53). Two cardiologists have manually annotated each recording with beat annotations in the database. The reference annotation files have more than 175,000 beat annotations, which are useful for testing and developing ECG classification systems. Although the patient group is diverse in age, gender, and disease, it may not represent all patient populations (53).

**MIT-BIH Supraventricular Arrhythmia Database (3):** This dataset includes 78 two-lead ECG recordings, each lasting for 30 min. The recordings were collected from 14 patients, 11 men and three women, with different types of supraventricular arrhythmias (3). These included normal, ventricular, fusion, and unknown beats. The ECG signals were sampled at 128 Hz and digitized with 11-bit resolution. Physicians and researchers have extensively utilized this database to develop and validate algorithms to detect and classify various arrhythmias (3). It has been used in several studies to compare the performance of different algorithms, including machine learning and deep learning algorithms (28).

### 3.1 Data imbalance issues in the above databases

Data imbalance is a significant challenge in ECG beat classification, particularly when using the MIT-BIH Arrhythmia (8), St. Petersburg INCART (8, 47), and MIT-BIH Supraventricular Arrhythmia (47) databases. These databases contain varied distributions of ECG beat types, often resulting

TABLE 2 Details of the databases utilized for ECG beat classification.

| S.no | Name of the database                        | Number of subjects | Sampling frequency (in Hz) | Resolution | Total number of ECG segments | No. of classes |
|------|---------------------------------------------|--------------------|----------------------------|------------|------------------------------|----------------|
| 1    | MIT-BIH Arrhythmia (8)                      | 47                 | 360                        | 11-bits    | 109,000                      | 17             |
| 2    | St Petersburg INCART (8, 47)                | 25                 | 257                        | 12-bits    | 175,840                      | 15             |
| 3    | MIT-BIH Supraventricular Arrhythmia (3, 47) | 14                 | 128                        | 11-bits    | 184,508                      | 9              |

TABLE 3 Estimated number of manuscripts using various ECG beat classification databases from different publishers.

| Database                                          | Elsevier manuscripts | IEEE manuscripts | Springer manuscripts |
|---------------------------------------------------|----------------------|------------------|----------------------|
| MIT-BIH Arrhythmia Database(8)                    | 22                   | 18               | 19                   |
| INCART (8, 47)                                    | 5                    | 7                | 7                    |
| MIT-BIH Supraventricular Database(47)             | 11                   | 9                | 8                    |
| PTB diagnostic ECG (47)                           | 2                    | 3                | 9                    |
| European ST-T (47, 48)                            | 4                    | 3                | 5                    |
| LTAF (long-term AF) (47, 49)                      | 5                    | 6                | 3                    |
| Atrial Fibrillation Database (AFDB) (47, 50)      | 6                    | 8                | 4                    |
| CPSC (China Physiological Signal Challenge) (51)  | 4                    | 5                | 2                    |
| CINC (Computers in Cardiology Challenge) (47, 52) | 3                    | 4                | 7                    |
| Others                                            | 6                    | 9                | 9                    |

in an overrepresentation of normal or common beats and an underrepresentation of rare arrhythmias, which may degrade the model performance and reduce generalizability. In the MIT-BIH Arrhythmia database, N beats vastly outnumber abnormal arrhythmias such as VEB or SVEB (8). As a result, classifiers trained on this database often achieve high accuracy by predominantly predicting the majority class (N beats), while failing to correctly identify rarer arrhythmias. This class imbalance can lead to a high false-negative rate for life-threatening conditions such as ventricular tachycardia, as the classifier may not learn sufficient patterns for these rare events. Techniques, such as the oversampling of minority classes and undersampling of the majority class, and applying synthetic data generation methods, such as the synthetic minority oversampling technique (SMOTE), can help mitigate this imbalance (54). In addition, cost-sensitive learning, where higher penalties are assigned to misclassifying rare classes, can also improve classification performance.

The INCART database, which includes recordings from 25 subjects with different types of arrhythmias, also suffers from data imbalance (8, 47). The majority of the beats in this dataset are NSR, while pathological beats such as ischaemic events are underrepresented. Similar to the MIT-BIH Arrhythmia database, this imbalance can bias classifiers towards predicting normal beats. Addressing this issue is crucial for the real-world application of ECG beat classifiers, as it ensures the model can detect critical events such as ischaemia. To address this imbalance, data augmentation techniques, such as adding noise or time-shifting the ECG signals of rare arrhythmias, can be used, as well as balancing the dataset through stratified sampling (55). The MIT-BIH Supraventricular Arrhythmia database also presents a unique challenge due to its smaller dataset size and the imbalance between normal and supraventricular arrhythmias (47). Given the relatively low number of supraventricular beats compared to normal beats, classifiers can become biased towards predicting normal rhythms, further complicating the detection of less frequent but clinically important supraventricular ectopic beats.

Bias in machine learning refers to systematic errors that cause a model to make consistently incorrect predictions, often favoring certain patterns while neglecting others (56). Bias can arise due to

several factors, such as imbalanced training data, poor feature selection, or overly simplistic model assumptions that fail to capture the true complexity of the data (19). Bias in ECG beat classification is a critical issue, often stemming from imbalanced datasets where certain heartbeat types are overrepresented while others are underrepresented. This imbalance can lead to classifiers favoring the majority class, resulting in poor sensitivity for minority classes, such as supraventricular and fusion beats (57). One study demonstrated that standard classification models trained on imbalanced datasets tend to overfit the majority class while failing to recognize less frequent arrhythmias effectively (58). This issue is exacerbated by linear dependencies in ECG data, which further skew the model's learning process and introduce classification bias (57). Moreover, the impact of inter-patient variability, where models perform well on seen patients but fail to generalize to unseen cases, further contributes to biased ECG classification (59). A detailed overview of certain issues due to bias is presented in Table 4.

To address bias in ECG beat classification, several state-of-the-art methods have been proposed (19, 58–60). One promising approach is the use of a dynamic minority-biased batch weighting loss function, which enhances the learning process for minority classes while maintaining the model's ability to classify the majority classes accurately (59). In addition, feature fusion neural networks, which integrate multiple ECG representations, have been shown to improve classification fairness by extracting diverse feature sets that reduce bias (59). Another strategy involves differential beat accuracy (DBA), a metric that optimizes classifier performance by adjusting the learning process based on the statistical distribution of different beat types, ensuring a more balanced classification (58). Ensemble-based techniques, such as multiple-classifier architectures, have also been effective in reducing bias by leveraging diverse model outputs to correct misclassifications (60). Collectively, these approaches contribute to improving the fairness and generalization of ECG beat classification models, making them more suitable for real-world clinical applications. In addition to traditional methods for addressing imbalance, transfer learning can be a beneficial approach in this case (55). A model pre-

TABLE 4 Bias issues and solutions in ECG classification.

| Literature | Issue                            | Impact                                                                                                                                | Solution                                                                                                           |
|------------|----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|
| (57)       | Imbalanced datasets              | Certain heartbeat types are overrepresented, leading to classifiers favoring majority classes and poor sensitivity for minority beats | Dynamic minority-biased batch weighting loss function to prioritize underrepresented classes                       |
| (58)       | Overfitting to majority class    | Standard models fail to recognize less frequent arrhythmias effectively, resulting in classification errors                           | Feature fusion neural networks to extract diverse feature sets and improve classification fairness                 |
| (57)       | Linear dependencies in data      | Skews the model's learning process and increases classification bias, making it harder to detect minority classes                     | Differential beat accuracy (DBA) metric to optimize classifier learning based on data distribution                 |
| (59)       | Inter-patient variability        | Models perform well on seen patients but fail to generalize to new, unseen cases, reducing real-world effectiveness                   | Ensemble-based classification techniques to improve generalization by leveraging diverse model outputs             |
| (98)       | Cost-sensitive misclassification | Traditional classifiers may misclassify minority beats due to non-optimized decision boundaries, reducing clinical reliability        | Cost-sensitive classifiers that adjust decision thresholds based on the misclassification cost of minority classes |
| (175)      | Feature selection bias           | Feature selection methods can introduce bias if not carefully designed, leading to suboptimal classification performance              | Hybrid deep learning models combining rule-based and data-driven feature selection to minimize bias                |
| (60)       | Limited training data            | ECG classification models struggle with limited labeled data, resulting in poor generalization to new datasets                        | Data augmentation techniques such as synthetic ECG generation to enhance training dataset diversity                |

trained on a larger, more balanced ECG dataset can be fine-tuned using the smaller MIT-BIH Supraventricular database, allowing the model to better generalize across beat types. In all these cases, proper evaluation metrics should be used to assess the classifier's performance under imbalance conditions. Accuracy alone may not be a reliable metric, as it can be misleading in imbalanced datasets. Instead, metrics such as the F1-score, sensitivity, specificity, and AUPRC provide a more comprehensive understanding of model performance on both the majority and minority classes (61). By carefully considering these strategies and metrics, researchers can better handle the class imbalance problem in ECG beat classification, leading to more robust and clinically useful models.

## 4 Overview of included studies

The ECG beat classification method is depicted as a block diagram in Figure 6. The three primary stages of ECG beat classification systems are pre-processing, feature extraction, and classification. In automated machine learning (AutoML) algorithms (11, 36, 62), feature extraction and classification are often treated as separate stages. This means the algorithm typically involves a dedicated process for extracting relevant features from the raw data, followed by a classification module that uses these features to make predictions. In contrast, deep learning algorithms often integrate feature extraction and classification into a single module. This is due to the hierarchical structure of neural networks, particularly in CNNs and other deep learning architectures, where the network learns to extract features and classify data in an end-to-end manner automatically (18, 19, 25–28). This integration is one of the key strengths of deep learning, as it eliminates the need

for manual feature engineering and allows the model to optimize the feature extraction process as part of the overall learning task.

### 4.1 Pre-processing

Pre-processing is the initial and critical stage in ECG beat classification systems, aiming to enhance the quality of the raw ECG signals (Table 5). This stage is essential to mitigate the impact of noise and artifacts that can obscure the true physiological information in the ECG data. Pre-processing technique details and their categorization are illustrated in Figure 7. Common sources of noise include baseline wander, powerline interference (PLI), and muscle artifacts (MAs) (36). Techniques such as bandpass filtering, wavelet transforms, and normalization are employed to remove or reduce these interferences. In addition, this stage may involve the segmentation of the ECG signal into individual beats, setting the foundation for accurate feature extraction and classification in subsequent phases. First, the raw ECG signal is filtered to eliminate unwanted noise and artifacts in the preprocessing stage (35). The next step is to separate the signal into individual heartbeats, often identified by the R-peaks in the ECG. In the feature extraction phase, data from each heartbeat are parsed for various features based on its time domain, frequency domain, and morphology. Heartbeat rhythms and shapes are characterized using these properties, which can be utilized to identify certain arrhythmias. Classification involves applying machine learning methods to the retrieved attributes of individual heartbeats (90). The basic techniques for pre-processing ECG signals can be summarized as follows: classical filtering techniques, transform-based techniques, statistical and adaptive techniques, modern machine learning approaches, and advanced techniques.

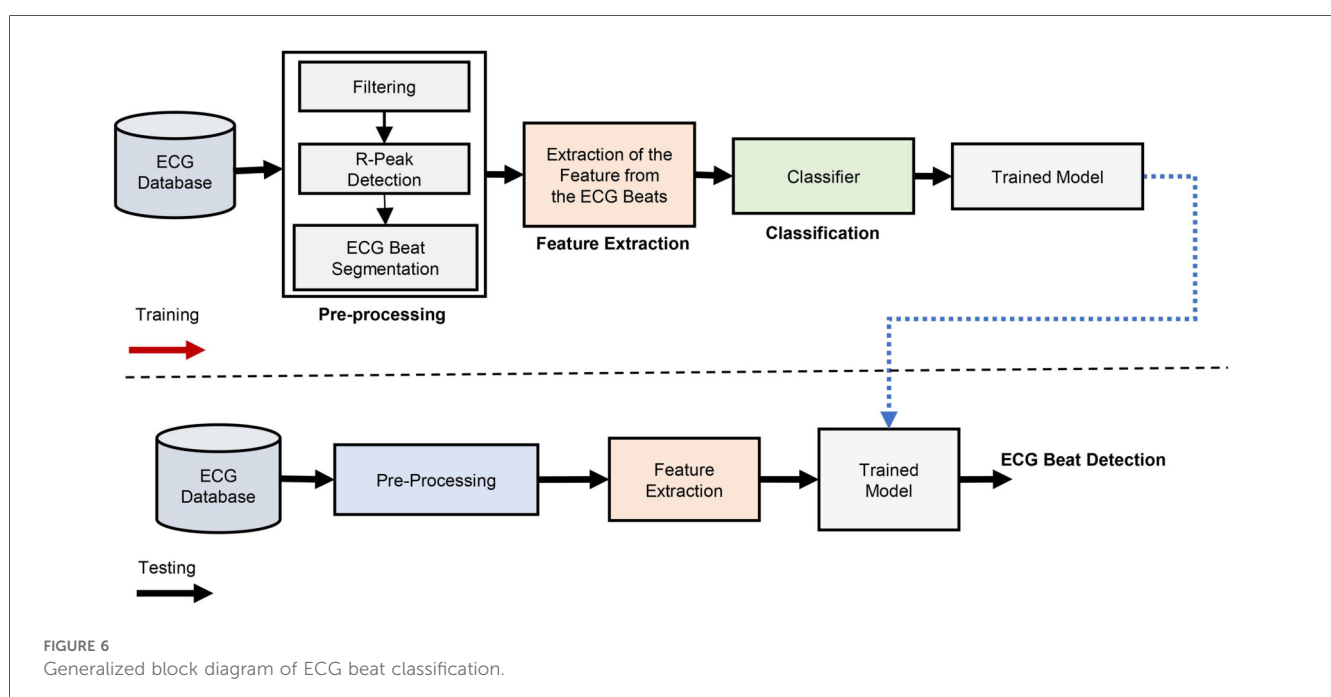


TABLE 5 ECG signal pre-processing methods: summary, advantages, and disadvantages.

| Method                                                         | Summary                                                                                                                                                                                      | Advantages                                                                                                                                     | Disadvantages                                                                                                                 |
|----------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| Bandpass filtering (63–65)                                     | Utilizes high-pass (HP) and low-pass filters (LPFs) to remove frequencies outside the heart's signal range (e.g., 0.5–150 Hz). Commonly implemented using Butterworth or Chebyshev filters.  | Effectively removes baseline wander and high-frequency noise (e.g., muscle artifacts).                                                         | May distort the original ECG signal, especially the QRS complex.                                                              |
| Wavelet transform (66–68)                                      | Decomposes the ECG signal into different frequency components using wavelet functions such as Daubechies or Symlets.                                                                         | Can localize noise in both time and frequency domains, making it ideal for isolating noise such as baseline wander and powerline interference. | Computationally intensive; selecting the appropriate wavelet function and level of decomposition can be challenging.          |
| Principal component analysis (PCA) (69–71)                     | Reduces dimensionality by extracting the principal components, which retain the most significant features of the ECG signal, while discarding noise components.                              | Efficient in noise reduction and works well for multi-channel ECG data                                                                         | Can lead to loss of important information if principal components are not correctly identified; assumes linearity             |
| Adaptive filtering (72–74)                                     | Uses a reference noise signal (e.g., from a secondary channel) and adapts the filter coefficients to remove the correlated noise. Common algorithms include LMS (least mean squares).        | Can dynamically adjust to varying noise conditions; effective in reducing powerline interference and motion artifacts.                         | Requires a reference signal; may not perform well if noise is highly non-stationary.                                          |
| Empirical mode decomposition (EMD) (75–77)                     | Breaks down the ECG signal into intrinsic mode functions (IMFs) to separate noise from the underlying signal.                                                                                | Adaptive method that works well with non-linear and non-stationary signals, such as ECG data.                                                  | Can introduce mode mixing, where noise and signal are not clearly separated; sensitive to noise in the decomposition process. |
| Notch filtering (56, 78, 79)                                   | A narrow-band filter is used to eliminate specific frequencies, typically used to remove powerline interference (e.g., 50/60 Hz).                                                            | Highly effective in removing powerline interference without affecting the rest of the signal.                                                  | Can distort the ECG signal around the notch frequency; it does not address other noise sources.                               |
| Sparse representation (80–82)                                  | Represents the ECG signal as a sparse linear combination of basis functions. Noise is filtered out by removing components that do not contribute significantly to the sparse representation. | Effective in removing various types of noise while retaining signal morphology; adaptable to low-SNR ECG data.                                 | Requires careful selection of the dictionary; computationally expensive for large datasets.                                   |
| Non-local means (NLM) filtering (76, 83, 84)                   | Averages similar ECG signal patches to reduce random noise. The similarity between patches is determined based on their intensity and spatial distance.                                      | Highly effective in reducing random noise while preserving the sharp features of ECG signals.                                                  | Computationally intensive, especially for large datasets, can blur features if parameter selection is improper.               |
| Deep learning-based denoising (85–87)                          | Uses deep learning models such as autoencoders, CNNs, or RNNs to learn noise patterns and denoise ECG signals in a supervised or unsupervised manner.                                        | Capable of removing complex noise patterns while preserving ECG features; adaptable to real-time applications.                                 | Requires a large dataset for training; computationally intensive and can overfit if not properly regularized.                 |
| Total variation denoising (TVD) (88)                           | Reduces noise by minimizing total variation in the signal, preserving edges (e.g., sharp ECG features).                                                                                      | Preserves sharp transitions such as R-peaks, effectively removing low-level noise.                                                             | May lead to over-smoothing if the regularization parameter is not properly chosen.                                            |
| Empirical wavelet transform (EWT) (89)                         | A variant of the wavelet transform that adapts the filter bank to the specific frequency bands of the input signal.                                                                          | Automatically adapts to the signal's characteristics, providing better decomposition for non-stationary signals like ECG.                      | Still computationally expensive and requires careful parameter tuning.                                                        |
| Generative adversarial networks (GAN)-based denoising (85, 86) | Employs GANs where the generator learns to denoise ECG signals while the discriminator ensures that the denoised signal is close to the true ECG.                                            | Can learn complex noise patterns and adapt to various datasets; highly flexible.                                                               | Training GANs is computationally expensive and requires careful balancing to avoid mode collapse.                             |
| Convolutional neural networks (CNN)-based denoising (85, 86)   | CNNs are trained to identify and reduce noise in ECG signals automatically.                                                                                                                  | Efficient in removing structured noise while retaining ECG morphology; suitable for real-time applications.                                    | Requires a large amount of labeled training data; computationally intensive.                                                  |
| Recurrent neural networks (RNN)-based denoising (85, 87)       | Utilizes RNNs, particularly LSTMs, to handle time-series ECG data and remove temporal noise.                                                                                                 | Effective in handling temporal dependencies and preserving important features such as R-peaks.                                                 | Training is computationally expensive and requires careful hyperparameter tuning to avoid overfitting.                        |

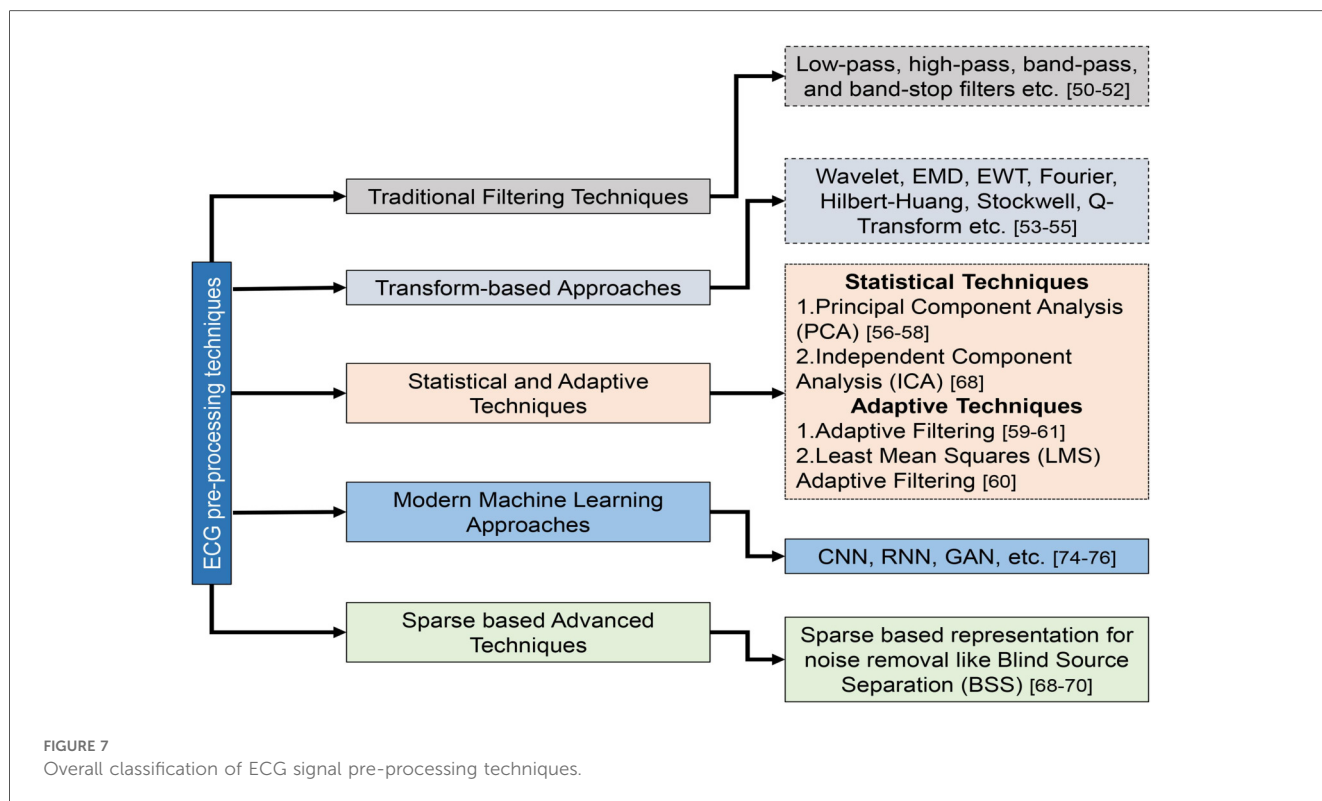
#### 4.1.1 Classical filtering techniques (63–65)

Low-pass, high-pass, bandpass, and band-stop filters are the classical filtering techniques in ECG signal pre-processing. Bandpass filtering, which combines high-pass and low-pass filters, effectively removes baseline wander and high-frequency noise, preserving the heart's signal within a specific frequency range. Notch filtering targets particular frequencies, such as 50/60 Hz powerline interference, to eliminate sinusoidal noise

without affecting other aspects of the ECG signal. These methods are straightforward and widely used but can distort the ECG signal, especially in complex cases.

#### 4.1.2 Transform-based techniques (66–68)

Transform-based methods such as wavelet transform and empirical mode decomposition (EMD) offer advanced noise reduction by analyzing the ECG signal in different frequency



components or intrinsic mode functions. Wavelet Transform is particularly useful for isolating noise from baseline wander and powerline interference by decomposing the signal into various frequency bands. EMD, on the other hand, adapts to non-linear and non-stationary signals by breaking the ECG signal into intrinsic mode functions, which can be selectively processed to reduce noise. These methods, while powerful, can be computationally intensive and may require careful parameter tuning.

#### 4.1.3 Statistical and adaptive techniques (69–71)

Statistical methods such as PCA and adaptive filtering focus on removing noise through statistical analysis and adaptive adjustments. PCA reduces dimensionality and noise by extracting significant components from multi-channel ECG data. Adaptive filtering, which uses reference noise signals to adjust filter coefficients dynamically, effectively addresses varying noise conditions, including powerline interference and motion artifacts (MAs). These techniques effectively reduce noise but may be limited in rapidly changing noise characteristics.

#### 4.1.4 Deep learning-based technique (85–87)

Recent advancements in machine learning have introduced sophisticated methods such as deep learning-based denoising, CNN denoising, and RNN denoising. These approaches leverage deep learning models to learn and remove complex noise patterns while preserving essential features of the ECG signal. For example, through adversarial training, GAN-based denoising techniques are used to separate noise from the ECG signal, enhancing adaptability to various noise types. CNN-based

denoising uses convolutional layers to filter structured noise effectively. At the same time, RNN-based techniques, particularly those employing LSTM networks, manage temporal noise by capturing dependencies over time. Although these modern methods offer superior performance and adaptability, they require substantial computational resources and extensive training data to achieve optimal results.

#### 4.1.5 Advanced techniques (80–82)

Emerging methods such as sparse representation and transfer learning-based denoising provide innovative solutions for ECG signal pre-processing. Sparse representation filters out noise by representing the signal as a sparse linear combination of basis functions, which helps preserve signal morphology. Transfer learning-based denoising, utilizing pre-trained models, reduces the need for extensive data collection and training, making it efficient for small datasets. In addition, blind source separation (BSS) methods, such as independent component analysis (ICA), separate the ECG signal from noise sources by exploiting statistical independence, offering effective noise reduction in multi-lead ECGs. These advanced techniques continue to push the boundaries of ECG signal processing, addressing challenges that traditional methods may struggle with.

In ECG signal processing, significant noise artefacts pose a considerable challenge to the accurate analysis and interpretation of ECG signals. These extraneous disturbances can stem from various sources, including muscular activity, electromagnetic interference, electrode impedance, and baseline drift. Consequently, faithful extraction of relevant physiological information from noisy ECG recordings becomes critical,

demanding innovative signal processing techniques (29). ECG recordings are generally mixed with different noises during acquisition from cardiac patients. Pre-processing techniques are crucial for the design of better classification systems. Identifying different fiducial points in ECG signals, such as P-onset and off-set, Q-onset and off-set, and T-onset and off-set, are difficult in noisy environments (91). Hence, the filtering of ECG signals without losing important information is a challenging task. The primary noise sources in ECG signal acquisition include MAs, caused by muscle contractions and tension, leading to irregular, high-frequency oscillations that distort the cardiac signal (7); electromagnetic interference (EMI) from electronic devices and power lines, causing voltage fluctuations and noise spikes (7); BW, gradual baseline shifts due to respiration or movement, obscuring low-amplitude cardiac components (92); PLI, visible as sinusoidal noise from electrical systems (9); MArS, high-frequency noise resulting from patient movements such as coughing or shifting positions (7); and electrode contact noise (ECN), resulting from poor electrode-skin contact, leading to signal distortion, especially during motion (2).

In addition, minor noise sources such as drift and offset, lead misplacement, sweat, moisture, and external interference also affect ECG morphology. Several researchers have developed different digital signal processing techniques to remove noise from the ECG signal (32, 38, 90, 93–96). In (93), an empirical mode decomposition was developed to effectively eliminate noise by using significant intrinsic mode functions of the ECG signal. A deep score-based diffusion model for ECG BW and noise removal and a multishot averaging strategy were developed to improve signal reconstructions in (96). A denoising autoencoder (DAE) was designed to remove the BW and PLI from the ECG signal in (95). A four-stage adaptive noise canceller was designed to remove the noise artefacts from the ECG signal in (94). In the dual stage, a different approach, namely singular value decomposition (SVD), was developed to improve the signal-to-noise ratio (SNR) of the ECG signal (38). Authors have proposed a sequential Monte Carlo algorithm combining the Wavelet transform to handle MA noise in the ECG signals (21). Other authors have proposed an R-peak detection and denoising algorithm using Shannon-energy and Hilbert transform (97). In the literature, several techniques have been developed to remove the noise from the ECG signals, even though there are still many challenges in real-time ECG acquisition. Some of the challenges are (i) patient movement during data acquisition can introduce noise and distortion to the ECG signal, making it difficult to extract the underlying cardiac information accurately; (ii) distinguishing between various types of noise and genuine cardiac signals is a critical step in effective noise removal as developing accurate algorithms for artifact identification is essential for preserving diagnostic integrity; and (iii) noise removal techniques developed in controlled laboratory settings may not always translate effectively to diverse clinical environments. Adapting and validating these techniques for real-world conditions are a challenge.

The pre-processing phase plays a pivotal role in ensuring that the ECG signal fed to machine learning or deep learning models is

free from artifacts and retains physiologically relevant information. Raw ECG signals are typically contaminated by several types of noise, including baseline wander, PLI, MArS, and muscle (electromyogram) noise. Baseline wander, often caused by respiration or electrode movement, leads to low-frequency drift that can distort wave boundaries. Powerline interference introduces sinusoidal noise at 50 or 60 Hz, while MArS and muscle activity generate high-frequency components overlapping with the QRS complex, thereby degrading the diagnostic quality of the signal (38, 93). To mitigate these artifacts, various denoising approaches have been adopted. The Butterworth band-pass filter, commonly configured to between 0.5 and 40 Hz, effectively removes baseline drift and high-frequency disturbances while preserving critical cardiac information. Adaptive filters, such as the least mean square (LMS) (98) and recursive least squares (RLS) (48) algorithms, dynamically adjust their parameters to cancel correlated noise, especially powerline components and electrode movement artifacts (90). In addition, wavelet-based filtering has gained widespread use because it offers multi-resolution analysis. The discrete wavelet transform (DWT) (62) decomposes the ECG into different frequency bands, allowing selective thresholding to suppress noise without distorting QRS morphology (4). Normalization techniques, such as Z-score normalization and min-max scaling, are also applied to standardize ECG amplitude across subjects and devices, ensuring consistent model convergence. Across the reviewed literature, studies that employed multi-stage denoising pipelines—typically combining wavelet filtering with adaptive or Butterworth filtering—reported significant improvements in R-peak detection and classification accuracy. Overall, wavelet-adaptive hybrid pipelines consistently yielded superior signal quality and classification performance compared with single-stage filtering approaches (25, 57).

## 4.2 Feature extraction methods

Feature extraction techniques help to identify the different patterns and characteristics of the ECG signal, which can then be used to diagnose specific cardiac conditions. For example, feature extraction techniques can be used to identify abnormalities in the various segments of the acquired ECG signal, which can be indicative of different cardiac disorders. Therefore, accurate feature extraction and classification techniques are crucial for effective diagnosis and treatment. Section 4.2 reviews the existing techniques for feature extraction and classification of ECG signals in the literature.

### 4.2.1 Handcrafted feature extraction

The performance of the classifier is dependent upon the extracted features. Feature extraction techniques are crucial in ECG signal processing and analysis, as they help identify important signal characteristics, such as amplitude, frequency, duration, and shape. Handcrafted feature extraction techniques are mainly classified into the following three types: time-domain

(33), frequency-domain (33), and time-frequency domain (11) methods. These methods have shown promising results in identifying essential features of ECG signals that are useful for diagnosing different ECG beats.

#### 4.2.2 Time-domain features

Time-domain features refer to the signal characteristics based on the time each observation is measured. In other words, these features describe how the signal changes over time. Time-domain features include a signal's mean, variance, and standard deviation and measures of the signal's shape, such as its skewness and kurtosis. Some standard time domain features utilized in ECG analysis are RR interval, P-wave duration, QRS duration, QT interval, heart rate variability (HRV), PR interval, and ST segment. Clinicians use these time-domain features to diagnose and monitor a range of heart conditions, including arrhythmias, heart failure, and myocardial infarction. They can also be used in machine learning algorithms to develop automated classification models for detecting different types of ECG beats (4, 35, 91, 99–104). Time-domain features alone do not allow the model to better interpret the ECG signal and they provide limited information about the underlying physiological processes, as these features do not capture the complex patterns and dynamics of the ECG signal. These are sensitive to noise and artifacts, which can affect the classification's performance. In addition, these are vulnerable to variations in signal morphology, such as changes in heart rate, respiration, and electrode placement. This can affect the accuracy and reliability of classification results. Several limitations need to be considered when developing a classification system. Combining time-domain features with other features or techniques is often necessary to improve classification performance (31).

#### 4.2.3 Frequency-domain features

Frequency-domain features are a group of traits or characteristics that characterize a signal's frequency content. In signal processing, it is expected to work with signal representations in both the time and frequency domains. In contrast to time-domain features, which characterize the signal over a given time interval, frequency-domain features characterize the signal's spectral characteristics throughout a range of frequencies. Some of the critical frequency-domain features include spectral power density (105), spectral entropy (106), spectral bandwidth (107), spectral flatness (108), and spectral skewness (109). Frequency-domain features can be extracted using various signal processing techniques such as the Fourier transform (FT) (105, 110), wavelet transforms (36), and spectrogram analysis (9). FT is a widely used technique for analyzing the frequency-domain characteristics of signals. In ECG analysis, the FT (110) can identify the frequency components in the ECG signal characteristic waves, such as the QRS complex, T wave, and P wave. The extracted features can be used for various applications, such as arrhythmia detection, heart rate variability analysis, and heart disease diagnosis. Frequency-domain features alone are not effective in classification due to the following reasons: some important

features of the signal, such as the shape and duration of the QRS complex, may not be fully captured in the frequency domain; in some cases, different features of the ECG signal may have similar frequency components. This can make it difficult to distinguish between these features based on frequency-domain analysis alone. Frequency-domain analysis assumes that the signal is stationary over time, meaning its statistical properties do not change. However, ECG signals are often non-stationary, with features that change over time. In such cases, time-frequency analysis techniques may be more appropriate for ECG signal interpretation (62).

#### 4.2.4 Time-frequency domain features

Sections 4.2.2 and 4.2.3 individually covered the characteristics of the time and frequency domains for ECG beat classification systems. Classifying ECG beats accurately requires information from both the time and frequency domains. These features help capture the temporal and spectral characteristics of the ECG signal, which are crucial for distinguishing between different heartbeats. In ECG beat classification, time-domain features, such as the RR interval, QRS duration, and QT interval, are commonly used to extract information about the duration and amplitude of various segments of the ECG signal. However, these features do not provide information about the signal's spectral content, which can be important when identifying specific types of heartbeats. In contrast, frequency-domain features provide information about the frequency content of the ECG signal. For example, the power spectrum of the ECG signal can be used to identify different frequency bands that correspond to specific physiological phenomena, such as the QRS complex, T wave, and P wave. In addition to time-domain and frequency-domain features, time-frequency features, such as wavelet transforms and spectrograms, are commonly used in ECG beat classification. These features provide a more comprehensive representation of the ECG signal, capturing both the temporal and spectral characteristics. The wavelet transform (111) is another technique that can be used to analyze the frequency domain characteristics of ECG signals. It provides a more localized frequency analysis than the Fourier transform and is useful in identifying transient features in the signal. DWT (29) is a signal processing technique that decomposes a signal into different frequency sub-bands. It is useful for identifying different frequency components in the ECG signal. These techniques can be used alone or in combination to extract frequency domain features from ECG signals. A number of transformation techniques can be utilized, i.e., dual-tree complex wavelet transform (DTCWT) (11), and Stockwell transform (ST) (29), to extract the time-frequency-based features from the pre-processed data. The short-time Fourier transform (STFT) uses a window function to analyze the signal in short-time intervals, which can lead to spectral leakage and reduced resolution. The resolution of the STFT is limited by the window size and the sampling rate, making it difficult to simultaneously analyze signals with high temporal and high-frequency content. Interpreting STFT results can be challenging, especially when analyzing complex signals with overlapping frequency content.

The DTCWT requires careful selection of wavelet filters, which can be challenging and subjective. The DTCWT also requires significant computational power to process and analyze signals, especially for high-resolution applications or long signals. Finally, while the DTCWT offers improved shift-invariance compared to other wavelet transforms, it is not completely shift-invariant. This can cause issues in specific applications where shift invariance is critical. The ST provides high time-frequency resolution, allowing for a more accurate analysis of non-stationary signals. In addition, the S-transform is shift-invariant, meaning that it is not affected by signal translations or shifts in time, and it produces a two-dimensional (2D) representation of the signal in time-frequency space, which is easy to interpret and analyze. The ST can handle non-uniformly sampled data, making it suitable for applications where data are not uniformly sampled. Therefore, the ST is better compared to the STFT (31) and WT (11) for ECG beat classification (112).

#### 4.2.5 Deep learning-based feature set

Handcrafted feature extraction is a traditional method of extracting relevant features from raw ECG signals for ECG beat classification. The handcrafted feature engineering process is often time-consuming and requires significant effort and resources, which can hinder the development of large-scale systems (20). Another limitation is the potential for human bias. The features are designed by human experts, who may have inherent biases and subjective judgements. Finally, handcrafted feature extraction may not be suitable for complex ECG signals. The features are designed based on prior knowledge and assumptions about the data, which may not always hold in practice (18). This can result in features that are not representative of the underlying data distribution, leading to poor generalization and performance. Hence, automatic feature extraction based on deep-learning techniques from the ECG database has been introduced for ECG beat classification. Deep learning-based feature extraction has become an increasingly popular alternative to handcrafted feature extraction in recent years (113). Unlike handcrafted features, deep learning-based features are automatically learned from raw data, eliminating the need for human expertise and domain knowledge. Deep learning-based feature extraction involves training a neural network to learn a hierarchy of features from raw data. The first layers of the network learn simple, low-level features, such as edges and corners, while deeper layers learn more complex and abstract features. One of the main advantages of deep learning-based feature extraction is its ability to learn features tailored to the specific task. This contrasts with handcrafted features, which are designed based on prior knowledge and assumptions about the data (114). Once a deep neural network (DNN) has been trained on a large dataset, the learned features can be reused for other tasks or applied to new datasets. This can significantly reduce the time and resources required for feature engineering and model development.

Deep learning methods can be classified into several categories based on their architecture, learning mechanisms, and applications. Deep learning methods are generally

categorized into three types, namely, discriminative, representative, and generative models (99). Discriminative deep learning methods are a class of deep learning algorithms designed to learn a mapping between inputs and outputs directly. Unlike generative models that learn the underlying probability distribution of the data, discriminative models learn to discriminate between different classes of data based on their features. Some popular discriminative deep learning methods include CNNs for classification, RNNs (25), and DNNs (9). These methods have achieved state-of-the-art performance in ECG beat classification. Discriminative deep learning methods typically involve many parameters learned through backpropagation. Backpropagation involves computing the gradient of a loss function for the model parameters and using it to update the parameters to minimize the loss. One of the advantages of discriminative deep learning methods is their ability to learn complex decision boundaries between classes, which can lead to high accuracy on classification tasks (26). However, they are often data-hungry and require large amounts of labeled training data to perform well.

Representative deep learning methods are essential for advancing deep learning because they are the foundation for developing new and innovative deep learning models. By understanding the underlying principles of these methods and the techniques used to optimize them, researchers can build upon them to create even more powerful and effective deep learning algorithms. Some examples of representative deep learning methods include GANs (115), autoencoders (AEs) (116), and deep belief networks (DBNs) (117). GANs are generative models that generate new data samples from a given input. They consist of two neural networks, a generator and a discriminator, trained in a minimax game (115). GANs have been successfully applied to tasks such as image generation, data augmentation, and anomaly detection. AEs are unsupervised deep learning models that are used for feature learning and dimensionality reduction (116). They consist of an encoder and a decoder network that learn to compress and reconstruct the input data. Autoencoders have been successfully applied to tasks such as image denoising, anomaly detection, and data compression. DBNs are deep generative models with multiple restricted Boltzmann machines (RBM) layers (117).

Generative deep learning models are a class of artificial neural networks (ANN) designed to generate new, synthetic data similar to data from a training set (115). These models can learn complex patterns and structures from the training data and then use that knowledge to generate new examples similar to the original data. Several generative deep learning models include variational autoencoders (VAEs) (118), GANs (115), and autoregressive models (AMs) (119). Discriminative models are designed to learn the boundary between different classes of data, while representative models aim to learn the underlying structure of the data. In contrast, generative models learn to generate new data similar to the training data. Overall, each of these three types of deep learning models has its strengths and weaknesses, and the choice of model depends on the specific task at hand and the nature of the dataset being used.

Feature extraction converts pre-processed ECG signals into compact and discriminative representations that can effectively describe cardiac morphology and rhythm. The extracted features generally fall into five categories, namely, time-domain, morphological, frequency-domain, time-frequency, and non-linear descriptors, each capturing different aspects of the ECG waveform (4). Time-domain features quantify temporal variations between successive heartbeats. Parameters such as the R–R interval, HRV indices [standard deviation of normal-to-normal intervals (SDNN), root mean square of successive differences (RMSSD), and percentage of successive normal-to-normal (NN) intervals that differ by more than 50 milliseconds (pNN50)], and mean or variance of beat-to-beat intervals reflect autonomic regulation and rhythm irregularities. These measures are computationally efficient and remain the foundation for arrhythmia detection in wearable and real-time systems (23). Morphological features describe the geometric and amplitude characteristics of individual ECG waves. Metrics including QRS width, P–R and Q–T intervals, R-peak amplitude, and area under the QRS complex capture structural deformation associated with ventricular and supraventricular ectopic activity. Derivative-based slopes and amplitude ratios between successive waves further enhance discrimination among beat classes (1).

Frequency-domain features, derived using the fast Fourier transform (FFT) (110) or power spectral density (PSD) (120), provide information on periodic energy distribution within specific frequency bands (0–40 Hz). However, because ECG signals are non-stationary, time-frequency representations such as the DWT (62), STFT (23), or wavelet packet transform (WPT) (36) are preferred. These methods capture transient spectral changes and localize abnormalities more accurately than pure spectral analysis. Finally, non-linear descriptors such as sample entropy, approximate entropy, and fractal dimension quantify signal complexity and chaotic behavior, while PCA (58) and ICA (121) are employed to reduce dimensionality and highlight salient features. Across the reviewed studies, hybrid feature sets combining wavelet coefficients with entropy-based complexity measures consistently achieved over 97% accuracy on the MIT-BIH Arrhythmia Database, underscoring the advantage of multi-domain representation for robust ECG beat classification.

## 4.3 Classification methods

In the ECG beat classification system, a classifier automatically classifies different heartbeats based on their ECG waveform features. Various machine and deep learning techniques have been reported for identifying different types of heartbeats (18–20, 30, 112, 114, 122–129). These classifiers utilized extracted features from the ECG signal to distinguish between different types of heartbeats. The classifier's performance depends on the quality of the ECG signal, the feature extraction quality, and the choice of the classification algorithm. There are three different types of classification algorithms, namely, (i) unsupervised, (ii) semi-supervised, and (iii) supervised.

### 4.3.1 Unsupervised

Machine learning can take the form of unsupervised learning, in which the algorithm can learn patterns and relationships in data without being explicitly supervised. In unsupervised learning, the algorithm receives a set of input data but no labels to indicate the desired results. The algorithm is trained to recognize patterns and connections in the data and cluster items with similar characteristics (130). The k-means clustering, hierarchical clustering, and PCA methods are all examples of popular unsupervised learning techniques (130). A machine learning clustering algorithm groups related data elements. Clustering partitions a dataset so that data points in the same cluster are more similar than those in other clusters. Clustering methods such as k-means, hierarchical, and density-based spatial are popular. Using the mean of data points inside each cluster, k-means clustering divides data into K clusters, iteratively assigning data points to the nearest cluster centroid and recalculating centroids until convergence. An algorithm in hierarchical clustering arranges data points into a tree-like structure of clusters, with each node representing a cluster. Each data point is initially treated as its own cluster by the algorithm, which then iteratively merges the two closest clusters together until all data points belong to the same cluster. The algorithm in density-based clustering organizes data into clusters according to their density, with points closer together having a greater density than those further away. To make a dataset more manageable, dimensionality reduction is another common unsupervised learning method. This can aid in making the data simpler and, therefore, easier to analyze. The standard methods include PCA and *t*-distributed stochastic neighbor embedding (*t*-SNE). PCA is an algorithm that projects data into a lower-dimensional space after determining the directions in the data with the most variance. This results in fewer features that can be used to represent the data. The *t*-SNE algorithm can reduce the dimensions in a dataset without losing the information about the relationships between the data points. This can be helpful when visualizing data at a higher level in a lower dimension. In (35), three independent, unsupervised techniques, namely, linear discriminant analysis (LDA), PCA, and ICA, were utilized to classify ECG beats, as per the AAMI standard. K-means clustering was used for the classification of premature ventricular contraction (PVC), normal (N), left bundle branch block (LBBB), paced beats (P), and right bundle branch block (RBBB) ECG beats in (131).

### 4.3.2 Semi-supervised

Semi-supervised learning is a machine learning paradigm that falls between supervised and unsupervised learning. In the context of ECG beat classification, semi-supervised learning can be used to enhance the performance of the classification model by leveraging a large amount of unlabeled data (132). Several methods can be used for semi-supervised learning in ECG beat classification. One approach is to use a combination of unsupervised and supervised learning methods. In this approach, an unsupervised learning algorithm is used to cluster the unlabeled data, and

then a supervised learning algorithm is used to classify the labeled data using the clusters as features. This can improve the accuracy of the classification model, as the unsupervised learning algorithm can identify underlying patterns in the data that may not be apparent to the supervised learning algorithm. Another approach is to use a generative model, such as a GAN, to generate synthetic labeled data from the unlabeled data. The synthetic data can be used to augment the labeled data and improve the accuracy of the classification model. In addition to these approaches, there are also active learning methods to select the most informative unlabeled samples for labeling (132). This can be particularly useful in scenarios where labeling the entire unlabeled dataset is not feasible due to time or resource constraints. Overall, semi-supervised learning is promising for ECG beat classification since it can use enormous amounts of unlabeled data to enhance the classification model's accuracy (132). However, it is essential to thoroughly assess the model's efficacy and ensure it can withstand shifts in the input data. To distinguish between SVEBs (also known as S beats) and VEBs (also known as V beats), Zahi et al. (37) proposed a semi-supervised iterative label update method. Semi-supervised strategies for the categorization of paroxysmal atrial fibrillation (PAF) using CNNs and LSTMs are reported in (133).

### 4.3.3 Supervised

Unsupervised and semi-supervised learning are not typically the best approaches for ECG beat classification because they rely on clustering or dimensionality reduction techniques, which may not capture the complex and diverse patterns in ECG signals (134). ECG signals can contain various beat types and subtypes, each with distinct characteristics. Unsupervised and semi-supervised learning techniques may have difficulty accurately identifying and separating these different beat classes. Unsupervised and semi-supervised learning techniques are often used when labeled data are limited or expensive. The effectiveness of semi-supervised learning depends heavily on the quality of the unlabeled data. If the unlabeled data contain a lot of noise or irrelevant information, it can decrease the model's performance. In addition, semi-supervised learning requires a subset of the data to be labeled, which can be time-consuming and expensive. Determining which data to label can also be challenging to maximize the model's performance. ECG beat classification is a critical task that requires a high degree of accuracy (134). Supervised learning approaches are better suited to this task as they can be trained to optimize for accuracy and can leverage a larger number of labeled data points. However, in the case of ECG beat classification, a substantial amount of labeled data is available, making supervised learning approaches a more appropriate choice. The algorithm is trained in supervised learning using a labeled dataset containing examples of ECG signals and their corresponding beat types. The algorithm aims to learn a function that maps the input ECG signal to the correct beat type (134).

Some of the supervised machine learning classifiers are ANNs (122), SVMs (135), Hidden Markov models (127) and self-organizing maps (SOMs) (136). Hidden Markov models are

used to detect cardiac arrhythmias, as reported in (4). DTCWT is utilized to extract the morphological features and merge them with the temporal features. Five different types of ECG arrhythmias have been categorized using a multi-layer back propagation (MLP-BP) neural network (137). Regarding neurons in the deep layers, MLP-BP is extremely sensitive. Underfit occurs in MLP when the number of neurons in the hidden layer is low. Too many neurons in the hidden layer may cause the fitting curve to oscillate erratically due to overfitting. The network model will stop functioning if the weights are high. Although DTCWT shows merit as a feature extraction strategy, the method's final classification performance suffers from the limitations of the MLP-BP algorithm (9). To improve SVM's generalization capability in the identification of various ECG beats, particle swarm optimization (PSO) is employed (138). In (29), five distinct types of ECG beats were classified using an algorithm based on bacteria foraging optimization (BFA) and SVM. Using SVMs, Pawel et al. (122) proposed an ensemble classifier for categorising arrhythmias. In this case, a genetic algorithm was used to optimize the characteristics acquired by Weich and the discrete Fourier transform (DFT). In (123), various classifiers, including naive Bayes, linear and quadratic discriminating functions, and J48 classifiers based on majority voting, were used to categorize five distinct heartbeats according to the AAMI standard. To extract information from an ECG signal, (139) used a DWT in conjunction with a novel one-dimensional hexadecimal local pattern (1D-HLP) approach and then used a single nearest-neighborhood (1NN) classifier to categorize 17 different types of arrhythmias. The genetic ensemble of classifiers optimized by sets (GECS) was used to categorize 17 myocardial dysfunctions in (124). We estimate power spectral density features to improve the quality of the ECG signal. Feature extraction, the process of choosing and extracting valuable features from the ECG signal for use in machine learning methods, is performed manually. However, the classifier's precision may suffer if inappropriate features are used. Knowing which features are the most important when performing a classification task can be difficult. There have been several proposals for classifying cardiac arrhythmias, but many of the efforts that have been reported to date have at least one of the following limitations: (i) accuracy was only good for a few carefully chosen ECG recordings; (ii) feature extraction methods were overly complicated; (iii) classifier performance was suboptimal; (iv) fewer output classes; and (v) beat loss when the ECG signal was filtered for noise.

The use of deep learning algorithms for ECG beat classification has increased in recent years. CNNs (18), RNNs (25), DBNs (117), AEs (116), and attention-based models are only some of the deep-learning methods that can be applied to the problem of ECG beat classification. Automatic feature extraction is the main advantage of deep learning models. In recent years, deep-learning models have modified their structure in ECG beat classification to improve accuracy. During the initial stages of ECG beat classification, the prevailing models relied on handcrafted features, such as QRS duration, heart rate, and T-wave amplitude. These models exhibited a restricted level

of accuracy, prompting researchers to initiate investigations into deep learning models. CNNs have achieved notable success in classifying ECG beats due to their ability to directly extract relevant attributes from the raw ECG signal, minimizing the requirement for manually extracted features. In an initial study by (127), a one-dimensional CNN (1D CNN) was introduced. This CNN is capable of accurately identifying ECG beats without the need for manually extracted features. However, this research employed an FFT to preprocess the ECG beats. In (18), a 1D-CNN was introduced to process raw ECG signals without pre-processing. In (140), a parallel configuration of CNN is described as an efficient method for classifying ECG beats. Following the emergence of CNNs and RNNs, these models were developed to process data sequences effectively. RNNs can effectively capture the sequential dependencies and extended patterns found in ECG data.

A deep neural LSTM including spectral features for ECG data classification has been suggested by Grzegorz et al. (113). In (26), different ECG beats were detected simultaneously using a combination of a CNN and LSTM. In (141), four types of ECG beats were classified using a dense convolutional network (DenseNet) and bi-directional long short-term memory (Bi-LSTM) architecture that combines the wavelet transform. Furthermore, the research on ECG beat classification has also included other combinations, such as CNN-LSTM (27), LSTM-CNN (27), CNN-BiLSTM (142), and Bi-LSTM (25). DL algorithms have the ability to acquire sophisticated attributes and comprehend intricate patterns in ECG signals. ECG beat classification is especially crucial when significant deviations in beat morphology exist. Therefore, DL-based methods (142–148) are outperforming previously published techniques such as template matching, rule-based, and ML-based methods (12–17, 22–24). Most of the methods outlined in the current literature have limitations, such as feature extraction that requires human intervention, issues with class imbalance, requiring a large amount of training data, and the need for powerful GPUs.

A proficient hybrid deep learning architecture is required for efficient automatic feature extraction and ECG beat classification. 1D-CNNs are efficient but require more depth and parameters to achieve higher accuracy. ResNet was designed to address this problem, but overfitting may arise when the model's structure is very intricate, requiring high-performance hardware and extended training duration. Furthermore, ResNet is unsuccessful in capturing the long-term dependencies in a sequence. In the majority of the studies in the literature, researchers convert one-dimensional ECG beats into grey-scale images (39) or spectrograms (112) to improve performance. However, this approach can be computationally expensive, particularly when working with images rather than signals. Thus, a bi-directional gated recurrent unit (Bi-GRU) is implemented to capture the long-term dependencies in an ECG signal. Bi-GRU has the ability to process the input sequence in both forward and backward directions. This enables the model to effectively capture contextual information from past and future inputs, making it highly valuable for tasks that involve analyzing the relationships within the input ECG beat segment.

Bi-GRU is more efficient in training and converges faster than ResNet, particularly when working with smaller datasets (28). This is due to its reduced parameter count. Although Bi-GRU may not be able to extract complex features from input data like CNNs and ResNet, it still has its strengths. Using a CNN, ResNet, or Bi-GRU alone fails to improve ECG beat classification performance. To address these problems, the utilization of dual-stream or multi-stream (40) networks can be beneficial for achieving precise classification of ECG beats. Utilizing dual or multi-stream deep learning techniques to combine information from multiple sources or modalities can greatly enhance the performance of deep learning models (28). In dual-stream deep learning, the model receives data from two sources, each representing a distinct input or feature. In one stage in the model's design, these streams are usually combined to produce a forecast. Multi-stream deep learning is quite similar but uses three or more data streams. This approach may be particularly beneficial when dealing with complex data that can be broken down into numerous modalities. It is now common practice to employ dual or multi-stream deep learning approaches to combine data from many sources to enhance models' predictive abilities (28).

An optimal fit in a machine learning or deep learning model is considered good because it strikes a balance between underfitting and overfitting, allowing the model to generalize well to unseen data (144). Unlike underfitting, where the model is too simple to capture important patterns, or overfitting, where it memorizes noise from the training data, an optimal fit ensures that the model learns meaningful relationships without excessive complexity (163). This results in better generalization, reduced bias and variance, improved accuracy on both training and validation datasets, and enhanced robustness across different data distributions. By maintaining this balance, an optimally fitted model provides reliable and stable predictions, making it suitable for real-world applications where consistency and adaptability are essential.

Traditional machine learning classifiers continue to play an essential role in ECG beat classification owing to their interpretability and computational efficiency. Among these, SVMs and ensemble models, such as random forest and gradient boosting, remain dominant, achieving accuracies of 98%–99% when combined with optimized DWT-PCA features. Simpler methods, such as kNN, decision tree, and naïve Bayes, provide lightweight alternatives for embedded or real-time systems. Addressing class imbalance through SMOTE or cost-sensitive learning further enhances reliability. Overall, ensemble-based and kernel-optimized SVM frameworks deliver a strong balance between accuracy, speed, and interpretability, confirming that well-engineered ML systems remain competitive with deep-learning models in ECG beat classification.

## 5 ECG beat classification using advanced machine learning

We thoroughly analyzed several articles that use machine learning, deep learning, and explainable AI (XAI) techniques to

classify ECG beats. To address this, we have included a summary note on AI-based methods for ECG diagnosis in Section 5.

## 5.1 Machine learning approaches

An overview of previous studies on the classification of ECG beats through the application of machine learning techniques is presented in Table 6. The table reviews various ECG beat classification techniques using machine learning models, highlighting their applications, databases, and performance metrics. Techniques such as radial basis function neural networks (RBFNNs), SVMs, and ensemble classifiers have been applied in datasets such as MITDB, PTBDB, and INCART, achieving accuracy rates from 84.60% to 99.99%. Each approach has specific advantages, such as resilience to noise and improved specificity, but also faces limitations, including increased system complexity or computational demands. Notably, classifiers such as SVM and RF consistently showed high performance, making them suitable for real-world applications; however, challenges in feature selection and model generalization were critical issues in different studies. After extracting the features from the ECG beats, artificial intelligence methods from machine learning can be used to build models from these data to classify arrhythmia heartbeats (43, 43–46). SVMs (126), ANNs (164), KNNs (24), and LDAs (165) are some of the most popular machine learning methods used for ECG beat classification. Despite noise and outliers in the ECG data, SVM continues to produce effective results because it finds the ideal hyperplane that maximizes the margin between distinct classes. Basic artificial neural network structures with fewer layers and parameters could improve interpretability, making it easier for clinicians to comprehend and have confidence in the model's decisions when classifying ECG beats. The KNN algorithm is beneficial due to its simplicity and non-parametric nature, which makes it suited for small-to-medium-sized datasets and circumstances with extremely non-linear decision boundaries (24). However, when the classes in the ECG data are well-separated, LDA emerges as a linear classifier that strives to maximize the between-class variation while minimizing the within-class variance (165). Moreover, it is important to consider the dataset's size, the task's difficulty level, and the available processing resources before selecting an approach. Each algorithm has both advantages and disadvantages.

ECG beat classification using machine learning classification algorithms increases cardiac health assessment efficiency, accuracy, and automation. Initially, these methods offered rapid and automated ECG data analysis, saving healthcare workers time and effort. This may improve heart abnormality diagnosis and treatment outcomes and lead to lower healthcare expenditures. In addition, machine learning classifiers can efficiently process significant ECG data, making scalable and cost-effective analysis possible in clinical settings. Furthermore, these methods can understand intricate patterns and correlations within ECG readings, which enables the detection of minor anomalies that may not be evident to traditional human observers.

However, machine learning algorithms for ECG beat classification have been associated with notable challenges and limitations. These models depend on the quality and quantity of training data, which may be noisy, artifact-filled, and variable among populations. In addition, the sophisticated decision-making processes of machine learning models in ECG analysis may be difficult for clinicians to understand. This misinterpretability may limit the clinical acceptance of machine learning-based diagnostic technologies.

## 5.2 Deep learning approach-based ECG beat classification

A comprehensive literature review on ECG beat classification using deep learning models is presented in Table 7. The review of deep learning models for ECG beat classification revealed various approaches, such as CNNs, LSTMs, and hybrid models, applied to datasets such as MITDB, INCART, and SVDB. CNNs show robust performance with accuracies up to 99.90%, while LSTM-based methods excel in capturing temporal features, improving sensitivity and specificity. The advantages of these models include the automation of feature extraction and their potential for real-time applications. However, challenges remain, such as computational intensity, model complexity, and limited generalization across diverse datasets. Ensuring robustness in noisy, real-world conditions and addressing resource-intensive training remains critical for their wider clinical adoption. Deep learning has become an essential tool in ECG beat classification, with enormous significance in medical diagnostics and healthcare (18, 26). A key advantage of deep learning over more conventional machine learning approaches for ECG beat classification is its ability to automatically derive hierarchical features from raw ECG data, eliminating the need for features to be manually extracted (18, 26). In contrast, deep learning models, specifically CNNs and RNNs, are exceptionally proficient at automatically extracting complex features from raw ECG data, which allows for the accurate classification of a wide range of cardiac disorders and arrhythmias (18, 19).

DL algorithms can improve diagnostic ability by utilizing extensive datasets, allowing them to generalize across various cardiac diseases and adjust to differences in patient demographics and recording settings (18, 26). Over the past few years, CNNs have been utilized extensively in ECG diagnosis and attained remarkable performance. CNNs are superior to other methods for extracting spatial characteristics from input signals; this allows them to detect local patterns that may indicate different cardiac problems. In contrast, RNNs can understand the temporal correlations between subsequent ECG samples since they specialize in modeling sequential data (25–27). By utilizing the combined advantages of convolutional neural networks and recurrent neural networks, particularly in hybrid structures such as CNN-RNN, ECG classification algorithms can attain exceptional performance in terms of accuracy and robustness (26–28).

DL methods have various advantages over standard ECG beat classification algorithms, including automatically learning

TABLE 6 Review of the ECG beat classification techniques that use ML models.

| Study | Application                           | Database                           | ML approach                               | Performance (in %)                  | Advantages                                                                                        | Disadvantages                                                                     |
|-------|---------------------------------------|------------------------------------|-------------------------------------------|-------------------------------------|---------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| (149) | Arrhythmia classification             |                                    | RBFNN                                     | Sen 94.54                           | Optimized classification performance with minimal network size.                                   | A minimal number of morphological features are extracted from ECG beat.           |
| (150) | ECG beat classification               |                                    | LDA and MLPNN                             | Acc 84.60<br>Acc 89.00              | Non-linear classification capabilities were improved.                                             | Generalization capability of the system needs to be improved.                     |
| (151) | ECG beat classification               |                                    | SVM                                       | Acc 97.20                           | Resilient to noise and variability.                                                               | Limited generalization capability.                                                |
| (152) | ECG beat classification               |                                    | LR                                        | Acc 98.43<br>Spe 97.75              | A high average specificity of 97.75% with minimal false positives.                                | Implementing and fine-tuning Reservoir Computing algorithms are complex.          |
| (153) | AF classification                     | MITDB                              | SVM                                       | Acc 90.27<br>F1 84.00               | Suited for real-world applications.                                                               | AF rhythm detection has a lower F1 score.                                         |
| (154) | Abnormalities detection               |                                    | Ensemble classifier                       | Acc 90.25                           | Improves the classification system's performance.                                                 | May require enormous computational resources.                                     |
| (155) | ECG classification                    |                                    | CNN with SVM                              | Acc 97.10<br>Sen 96.50              | Improved performance with CNN and SVM.                                                            | Complex model requiring extensive tuning.                                         |
| (156) | Cardiac arrhythmia detection          |                                    | EEMD with KNN                             | Acc 93.40<br>Sen 95.40<br>F1 96.30  | Supports efficient automatic diagnosis in clinical settings.                                      | Advanced signal processing increases computational demands.                       |
| (157) | ECG beat classification               |                                    | Decision tree<br>Gradient boosting        | Acc 98.62<br>Acc 99.13              | Good accuracy for early heart disease detection.                                                  | Noise and artifacts may affect performance.                                       |
| (35)  | Five types of ECG beat classes        |                                    | PCA, LDA, ICA, DWT                        | Acc 99.97                           | DWT offers energy compaction; choice of dimensionality reduction depends on data characteristics. | Not tested with 10-fold validation; lacks generalization as tested only on MITDB. |
| (9)   | Five beat classification              |                                    | RF classifier                             | Acc 98.50                           | Higher accuracy with 10-fold validation.                                                          | DTCWT for feature extraction is hard to implement on hardware.                    |
| (29)  | ECG beat classification               | INCART                             | SVM                                       | Sen 91.70                           | Improved classification performance, and generalization capability of the model is verified.      | System complexity increases with various connected components.                    |
| (158) | Arrhythmia classification             | MITDB<br>SVEB                      | KNN and SVM with PSO optimization         | Acc 94.50<br>Acc 85.10              | Maintains high accuracy for 15 heartbeat classes.                                                 | Higher computational cost due to PSO optimization.                                |
| (159) | Heart diseases                        | MITSA                              | RF<br>GDB                                 | Acc 97.98<br>Acc 96.95              | Acceptable for large datasets.                                                                    | Performance may drop with poor feature selection.                                 |
| (24)  | Myocardial infarction & heart failure | NSR, PHR, MIT-BIH sinus arrhythmia | KNN                                       | Acc 98.40                           | Effective with small datasets.                                                                    | Memory-intensive for large datasets.                                              |
| (126) | ECG anomaly detection                 | Physionet long-term ECG            | Gaussian kernel-based SVM classifier      | Acc 99.99                           | Handles high-dimensional spaces effectively.                                                      | Computationally expensive with large datasets.                                    |
| (160) | Myocardial infarction                 | PTBD                               | SVM                                       | Acc 95.30<br>Sen 94.60<br>Spe 96.00 | Consistently better results.                                                                      | Computationally intensive model.                                                  |
| (161) | Myocardial infarction                 | PTB                                | SVM, KNN, and RF                          | Sen 92.60<br>Sen 92.30<br>Sen 91.43 | Automated MI diagnosis.                                                                           | Multiple-instance learning complicates classification.                            |
| (92)  | ECG arrhythmia detection              | SVEB                               | SVM with Kruskal-Wallis Feature selection | Acc 98.06<br>MCC 91.51              | Robust on imbalanced datasets.                                                                    | Preprocessing may limit generalizability.                                         |
| (162) | Apnea detection                       | Private                            | CHMM                                      | Sen 93.98<br>Spe 95.38              | Works well with real-time data.                                                                   | Limited training data reduces performance.                                        |

MLPNN, Multilayer Perceptron Neural Network; EEMD, Ensemble Empirical Mode Decomposition; CHMM, Coupled Hidden Markov Model; IraNet, Inter- and Intra-patient Representation Aggregation Network.

discriminative features from raw ECG data without manual feature extraction (18). With this automated feature extraction technique, DL models can better classify cardiac diseases by detecting intricate patterns and tiny variations in the data. Further, the temporal and sequential nature of ECG signals is well-suited to DL methods, especially RNNs and CNNs, enabling them to detect the long-range relationships and

temporal patterns in heartbeats (25). Furthermore, DL models have also shown excellent generalization abilities, which means they can adapt to different patient groups and recording situations, which is very important for real-world clinical applications.

In the field of ECG beat classification, deep learning algorithms have some significant limitations and challenges.

TABLE 7 Review of the studies on ECG beat classification using deep learning models.

| Study | Application                               | Database                | DL approach                                                               | Performance (in %)                  | Advantages                                                                                                                                                                                  | Disadvantages                                                                                                                                                           |
|-------|-------------------------------------------|-------------------------|---------------------------------------------------------------------------|-------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (127) | ECG beat classification                   |                         | CNN                                                                       | Acc 97.40<br>Sen 60.31              | DL replaces time-consuming, error-prone manual feature extraction.                                                                                                                          | Training CNNs for each patient can be computationally intensive, especially with more patients.                                                                         |
| (136) | ECG heartbeat classification              |                         | CNN (nine-layered)                                                        | Acc 94.30<br>Acc 84.07 (Noisy ECG)  | Accurately classifies heartbeats in noise-free ECGs, indicating ability to detect abnormal rhythms.                                                                                         | Training dataset quality significantly impacts performance. A biased dataset can hinder model generalization.                                                           |
| (137) | ECG arrhythmia detection                  |                         | DCNN (seven-layered)                                                      | Spe 99.83<br>Acc 99.68              | More efficient and accurate than conventional methods.                                                                                                                                      | Limited generalization to other datasets or real-life clinical scenarios.                                                                                               |
| (135) | ECG arrhythmia classification             |                         | 1D-CNN<br>2D-CNN                                                          | Acc 90.93<br>Acc 99.00              | 2D-CNN model classified five arrhythmias accurately.                                                                                                                                        | High computational resources and time needed for training, especially for 2D-CNN.                                                                                       |
| (166) | ECG heartbeat classification              |                         | LSTM with AE                                                              | Sen 98.63<br>Spe 99.66              | LSTM-AE learns features without manual extraction.                                                                                                                                          | LSTM networks are computationally expensive to train and deploy.                                                                                                        |
| (129) | ECG beat classification                   | MITDB                   | LSTM-based RNN                                                            | Acc 99.10<br>F1 95.00               | Patient-specific analysis achieved by training on 5 min segments.                                                                                                                           | Complex implementation hinders adoption, especially with limited resources.                                                                                             |
| (167) | ECG arrhythmia classification             |                         | Multiscale convolution and FCBA (frequency convolutional block attention) | Acc 95.60<br>Sen 93.17              | CNN with new multiscale blocks and attention modules classify arrhythmia with high sensitivity and accuracy. Effectively resolves data imbalance using oversampling and noise augmentation. | Complex model architecture and preprocessing methods may need huge computational resources. The method's performance may change with increasingly diversified datasets. |
| (39)  | Normal atrial fibrillation                |                         | Encoded deep CNN with input ECG beats as image                            | Acc 99.52<br>F1 Score 95.64         | The morphological variations can be captured more effectively using images compared to one-dimensional signals.                                                                             | Complex models with high computational power (GPU-based systems) are required to handle the large input image data.                                                     |
| (40)  | Beat classification as per AAMI standard  |                         | BiLSTM and random forest with PCA                                         | Acc 98.30                           | Ability to extract more in-depth features from the ECG signal.                                                                                                                              | As it is an ensemble learning model, it requires more training and testing time.                                                                                        |
| (28)  | Beat classification as per AAMI standard  |                         | Multi-stream Bi-GRU network with random forest                            | Acc 99.93                           | Deep features are extracted from the input, which are very helpful to accurately assess the type of beat.                                                                                   | The reliance on multiple deep-learning models can lead to high computational costs. Further research is needed to evaluate scalability for real-time applications.      |
| (141) | ECG arrhythmia classification             |                         | DenseNet with Bi-LSTM                                                     | Acc ~99.44                          | BiLSTM is integrated to enhance the model's capacity for extracting local features and capturing temporal features of ECG signals.                                                          | Bi-LSTM requires more computational power and needs to process input twice due to the parallel deep learning architecture.                                              |
| (27)  | ECG arrhythmia classification (six types) |                         | CNN and LSTM                                                              | Acc 99.32                           | The proposed model exhibits strong generalization capabilities and could serve as a valuable tool for clinicians in diagnosing arrhythmia.                                                  | Requires QRS detection, introducing additional computational overhead. Imbalanced dataset challenges, particularly with limited AFL class instances.                    |
| (10)  | Beat classification as per AAMI standard  |                         | An explainable deep transfer learning approach                            | Sen 99.00                           | Explainable AI enhances the accuracy and reliability of heartbeat classification by making DL models more transparent and understandable for physicians.                                    | Deployment feasibility for real-time monitoring must consider computational resources, integration with clinical workflows, and regulatory standards.                   |
| (168) | ECG arrhythmia classification             | MITDB<br>AFDB           | Customized CNN                                                            | Acc 97.31<br>Sen 96.50<br>F1 98.30  | Classifies numerous arrhythmias accurately utilizing short ECG segments, making them appropriate for continuous monitoring using wearable devices.                                          | Real-world wearable data noise and artifacts may affect performance. Single-lead ECG may lower arrhythmia detection accuracy.                                           |
| (143) | Detection of atrial fibrillation          | AFDB                    | CNN and RNN combination                                                   | Acc 89.30                           | Extract high-level features from segments of RR intervals (RRIs) to classify them as either AF or NSR.                                                                                      | Analysis of the model's performance on noisy ECG segments revealed a higher number of false positives, as anticipated.                                                  |
| (169) | ECG signal Classification                 | MITDB<br>ICCAD          | CLINet (Conv+LSTM+Involution)                                             | Acc 99.94<br>Acc 99.90              | The architecture is lightweight, making it appropriate for deployment on wearable devices, and it achieves very high accuracy across multiple datasets.                                     | A lack of comprehensive preprocessing may influence robustness in noisy real-world datasets.                                                                            |
| (170) | ECG beat classification                   | MITDB<br>INCART<br>SVDB | DNN                                                                       | Acc 91.30<br>Acc 92.40<br>Acc 90.61 | Enhanced system robustness and generalizability.                                                                                                                                            | Stacking autoencoders and DNNs is computationally expensive and resource-intensive.                                                                                     |

(Continued)

TABLE 7 Continued

| Study | Application                       | Database          | DL approach                              | Performance (in %)                  | Advantages                                                                      | Disadvantages                                                                       |
|-------|-----------------------------------|-------------------|------------------------------------------|-------------------------------------|---------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| (171) | Premature ventricular contraction | MITDB INCART      | Hybrid BiLSTM                            | Acc 97.20<br>Sen 96.00              | Achieves high performance in classifying premature ventricular contractions.    | May be less effective in detecting rare arrhythmias or artifacts.                   |
| (172) | ECG beat classification           | LUDB              | VGG16-based CNN                          | Acc 99.90                           | Accurately classifies beats using time-frequency representation.                | Impressive accuracy may indicate overfitting, especially on smaller datasets.       |
| (173) | Abnormalities detection           | MITDB Real-time   | Hybrid DNN                               | Acc 99.28<br>Acc 99.12              | High accuracy ensures reliable abnormality detection for clinical applications. | Heavily reliant on input data quality; noise and artifacts can reduce accuracy.     |
| (170) | ECG beat classification           | MITDB INCART SVDB | DNN                                      | Acc 91.30<br>Acc 92.40<br>Acc 90.61 | Enhanced system robustness and generalizability.                                | Stacking autoencoders and DNNs is computationally expensive and resource-intensive. |
| (174) | Heartbeat classification          | SVEB              | IraNet (residual attention with Bi-LSTM) | Acc 95.48<br>Sen 95.75              | Competitive accuracy and sensitivity across multiple classes.                   | Complex model may require extensive computational resources.                        |

Overfitting is a significant challenge in ECG beat classification, particularly in deep learning-based models, where the model captures noise and irrelevant patterns from training data, leading to poor generalization to new, unseen data. One of the primary causes of overfitting in ECG classification is the use of complex models with high-dimensional feature spaces, especially when datasets are imbalanced. Studies have shown that utilizing all 12 ECG leads in classification models can introduce redundancy and unnecessary complexity, thereby reducing generalizability and increasing the risk of overfitting (176). In addition, the presence of linear dependencies within ECG data has been identified as a major issue that skews the learning process towards the majority class, leading to biased model predictions and lower accuracy in minority class detection (57). This issue is particularly evident in patient-specific ECG classification, where models tend to overfit to the training patients, failing to generalize effectively across different individuals (29). Moreover, deep learning models trained on high-dimensional spectro-temporal ECG features have been found to suffer from overfitting due to their inability to leverage beat-to-beat variations effectively (120).

To mitigate overfitting in ECG beat classification, researchers have proposed various strategies. Feature selection and dimensionality reduction techniques, such as selecting an optimal subset of ECG leads instead of using all 12 leads, have been found to enhance classification accuracy while maintaining interpretability (176). Sample reduction techniques, such as QR and SVD, have also been effective in eliminating redundant data and addressing class imbalance, thereby reducing bias and improving generalization (57). In addition, the use of ensemble methods, such as multiple classifier architectures and hybrid machine learning approaches, has shown promise in improving classification robustness and interpretability (60, 175). Regularization techniques, including dropout and L2 regularization, are widely used to constrain deep learning models and prevent overfitting. Cost-sensitive learning approaches, such as modifying the classification threshold based on class imbalance, have also been shown to significantly enhance ECG beat classification performance by reducing bias

towards majority classes (98). Furthermore, data augmentation strategies, such as adding noise, applying transformations, or generating synthetic ECG beats using generative models, have been successful in increasing the diversity of training data and improving model generalization (120). Collectively, these approaches ensure the reliability and robustness of ECG beat classification models, making them more effective for real-world medical applications.

In addition to the overfitting, large annotated datasets are also necessary for effective deep neural network training, which is a significant challenge. Generating high-quality annotated ECG data, particularly for rare cardiac disorders, can be costly and time-consuming (9). Furthermore, DL models are computationally demanding and require enormous computational resources, particularly during the training and optimization phases. This may limit their scalability and practical implementation in contexts with limited resources, such as point-of-care settings. Despite these challenges, continuous research is endeavoring to overcome these constraints by creating interpretable DL architectures, data-efficient learning methods, and hardware optimizations (24). The ultimate goal is to maximize the benefits of DL for ECG beat classification and simultaneously reduce its potential drawbacks.

Deep learning has become the leading paradigm for ECG beat classification by automatically learning spatial-temporal features from raw signals. Among various models, CNN-LSTM hybrids and attention-enhanced networks consistently demonstrate superior performance, achieving accuracies of 98%–99% across multiple ECG databases. CNNs capture morphological patterns, while LSTM and GRU layers model temporal dependencies. Transformer-based models further improve interpretability and long-range context handling. Data augmentation with GANs and representation learning via autoencoders effectively address class imbalance and noise. In addition, transfer-learning and lightweight architectures, such as MobileNet, enable real-time deployment on wearable devices. Overall, hybrid and attention-driven frameworks represent the state of the art in robust, generalizable ECG beat classification systems.

### 5.3 Explainable AI for ECG beat classification

ECG beat classification using XAI leverages advanced machine learning techniques to enhance transparency in automated diagnostic systems. Traditional classification methods often function as “black boxes,” making it difficult for clinicians to interpret or trust decisions, especially in critical applications such as arrhythmia detection. Explainable AI addresses this limitation by employing methods such as saliency maps, feature importance visualization, and decision boundary analysis. These approaches allow for better understanding and verification of the features contributing to a model’s predictions. Techniques such as layer-wise relevance propagation (LRP) and gradient-weighted class activation mapping (Grad-CAM) have been applied to visualize critical regions of ECG signals, helping practitioners align model decisions with medical insights (177, 178). This not only improves diagnostic reliability but also aids in identifying potential errors in classification. Incorporating XAI into ECG beat classification aligns automated systems with real-world clinical requirements. Explainable methods enable the assessment of model predictions in handling class imbalances and morphological complexities in ECG datasets. Studies integrating XAI with neural networks, such as convolutional and recurrent architectures, have shown improved interpretability without compromising classification accuracy (58, 121). For instance, combining ICA with neural networks has demonstrated robust classification results with enhanced feature interpretability, achieving over 98% accuracy on benchmark datasets (98, 179). By bridging the gap between machine intelligence and clinician trust, explainable AI ensures that automated ECG diagnostic tools are not only effective but also transparent and clinically viable. A detailed comparison of the different features is summarized in Table 8.

XAI has emerged as a crucial frontier in ECG beat classification research, bridging the gap between high-performing deep-learning models and clinical interpretability.

While deep neural networks, such as CNNs, LSTMs, and transformers, have demonstrated outstanding diagnostic accuracy, their “black-box” nature continues to limit real-world clinical trust and regulatory acceptance. Among the 106 reviewed studies, approximately 14% (15 papers) explicitly incorporated XAI methods to interpret model predictions and visualize decision reasoning. Most of these studies applied saliency-based techniques, including Grad-CAM (180) and LRP (60), to highlight the waveform segments that contribute most strongly to the classification outcomes. Such visualization maps frequently revealed correspondence between algorithmic focus regions and clinically significant components of the ECG—particularly the QRS complex, P-wave, and ST-T segments—demonstrating the potential of XAI to validate model reasoning in physiological terms.

Beyond saliency methods, a smaller number of studies utilized attention mechanisms within hybrid CNN-LSTM architectures to provide interpretability by weighting critical temporal regions. Others employed the SHapley Additive exPlanations (SHAP) (181) and Local Interpretable Model-Agnostic Explanations (LIME) to quantify feature importance in models trained on handcrafted attributes. However, the overall adoption of such explainability frameworks remains limited. Approximately 86% of the analyzed deep-learning studies focused solely on predictive performance metrics—accuracy, precision, recall, or F1-score—without providing interpretive insight into model behavior or physiological relevance. This significant gap highlights that the ECG-AI community remains primarily performance-driven, with interpretability often treated as an auxiliary consideration rather than a core design principle. The limited integration of XAI tools underscores an urgent need for standardization in interpretability reporting. Future research should emphasize quantitative explainability benchmarks, such as region relevance overlap with annotated ECG segments, and clinician-in-the-loop validation to assess whether the model explanations align with expert reasoning. Integrating XAI at the design stage can enhance clinical transparency, regulatory

TABLE 8 Comparison of XAI, traditional ML, and deep learning.

| Feature                          | XAI                                                                                     | Traditional ML                                                                                           | Deep learning                                                      |
|----------------------------------|-----------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------|
| Transparency                     | High transparency; models provide insights into decision-making processes.              | Moderate transparency; some models (e.g., decision trees) are interpretable, others (e.g., SVM) are not. | Low transparency; often acts as a “black box.”                     |
| Accuracy                         | Slight trade-off in accuracy due to constraints of explainability in some cases.        | Moderate accuracy depending on the model and dataset.                                                    | High accuracy, especially in complex tasks.                        |
| Interpretability                 | Clear reasoning behind predictions, improving trust and adoption.                       | Varies; interpretable in simple models (e.g., linear regression) but not in ensemble models.             | Poor interpretability due to complexity.                           |
| Scalability                      | Dependent on the complexity of explainable methods; may require additional computation. | Good scalability; lightweight and efficient.                                                             | Highly scalable for large-scale data.                              |
| Application areas                | Suitable for trust-sensitive fields like healthcare and finance.                        | General-purpose applications like classification and regression.                                         | Ideal for unstructured data such as images, text, and audio.       |
| User trust                       | Builds trust through interpretable results.                                             | Moderate; depends on model interpretability.                                                             | Limited due to black-box nature.                                   |
| Complexity of implementation     | Higher due to integration of methods like SHAP, LIME, or Grad-CAM.                      | Moderate; simpler than XAI and DL.                                                                       | High; requires expertise in network architecture and optimization. |
| Regulatory compliance            | Aligned with regulations requiring transparency and accountability.                     | Limited compliance where explainability is required.                                                     | Poor compliance due to lack of interpretability.                   |
| Performance on unstructured data | Performs well but may lag behind DL in raw accuracy.                                    | Moderate; pre-processing required for unstructured data.                                                 | Excels with unstructured data such as images and signals.          |

compliance, and physician confidence, ultimately enabling safer and more explainable deployment of AI-driven cardiac diagnostic systems.

## 6 Limitations, solutions, and future directions

Though this review provides a comprehensive overview of ECG beat classification systems, it is important to acknowledge several limitations inherent in the current ECG beat classification approaches.

### 6.1 Minor challenges

- As the traditional methods use handcrafted features, they are unable to provide reliable and adequate information from similar ECG beats due to the action of morphological patterns, which may lead to misclassification (11, 43–46).
- Deep learning models involve many computations, particularly those with large architectures such as CNNs or RNNs. High-end GPUs offer more powerful processors and parallel processing capabilities, enabling faster and more efficient execution of these computations than CPUs (central processing units) (20, 25, 28, 114).
- Deep learning algorithms require extensive training datasets to train the models (9, 19).

*Solutions for minor challenges:* Implementing deep learning models, such as CNNs and RNNs, that automatically learn relevant features from raw ECG data can significantly improve classification accuracy (127). These models capture complex patterns and morphological variations without relying on handcrafted features. To optimize neural network architectures for computational efficiency, techniques such as model pruning and quantization can be adopted, or lightweight models such as MobileNet can be utilized (182). Leveraging cloud-based GPU resources or distributed computing helps meet computational demands without compromising performance (183). Employing data augmentation techniques to artificially expand the training dataset by introducing variations, such as scaling, rotation, or adding noise to existing ECG signals, enhances the model's robustness (184). In addition, one could use transfer learning from models pre-trained on large datasets or explore semi-supervised learning methods that can effectively use unlabeled data (185).

### 6.2 Major challenges

- ECG signals exhibit significant intra- and inter-patient variability in beat morphology due to differences in heart anatomy, electrode placement, noise, artifacts, and physiological factors. This variability makes it challenging for deep learning models to generalize across different individuals and accurately classify ECG beats (18, 20, 28, 116).

- Most of the pre-processing techniques in the literature depend on the signal sampling frequency. If the sampling frequency of the input signal varies, it is very difficult to achieve better classification accuracy (113, 114).
- Deep learning models generally require a large amount of labeled data to achieve high accuracy. Obtaining a large and accurately annotated ECG dataset for training deep learning models can be challenging, as ECG data collection and annotation are time-consuming and require expertise (10, 25, 136).
- ECG beat classification models trained on one dataset may not generalize well to unseen data from different sources. Ensuring the robustness and generalizability of deep learning models is a challenge in ECG beat classification (9, 18, 19).
- The performance of the classification algorithms is significantly degraded in the presence of similar morphological patterns with minor variations from different classes (9, 19, 26, 27).

*Solutions for major challenges:* Collecting a diverse dataset that represents different patient populations is crucial for training models that generalize well (186). Techniques such as domain adaptation can adjust the model to new patient data, and personalization strategies can fine-tune the model based on individual patient characteristics (187). Standardizing the sampling frequency by resampling all ECG signals to a common rate prior to processing ensures consistency (188). Developing pre-processing methods that are robust to sampling frequency variations or designing models that handle inputs with different sampling rates using adaptive algorithms can mitigate issues related to varied sampling frequencies (189). The use of semi-supervised learning and active learning approaches reduces the amount of labeled data required (190). Collaborating with medical institutions to share and pool annotated datasets expands the available training data (191). Exploring synthetic data generation methods, such as the use of generative adversarial networks to create realistic ECG signals, provides additional training resources (192). Training models on combined datasets from multiple sources improves generalization capabilities (163). Implementing cross-validation strategies across different datasets and using domain generalization techniques ensure that models are robust to variations in data distribution from different sources (193). Enhancing the model's ability to detect subtle differences by incorporating attention mechanisms and multi-scale feature extraction techniques improves discrimination between classes with similar morphological patterns (194). Using ensemble methods that combine multiple models further enhances classification performance (195), and fine-tuning the model using specialized loss functions, such as focal loss, emphasizes hard-to-classify examples (196).

### 6.3 Synthesis of the proven solutions and best practices

From the 106 selected articles, we identified consistently successful practices for ECG beat classification, particularly regarding data imbalance, generalization, model design, and

evaluation (55). A key observation across numerous studies is that models trained and tested on the same patient cohort (intra-patient splits) tend to exhibit inflated performance, as they inadvertently learn subject-specific morphology (127). In contrast, inter-patient validation—where all beats from a given subject appear in only one data split—offers a more realistic assessment by ensuring generalization to unseen individuals. Therefore, inter-patient (or leave-one-subject-out) evaluation should be considered the primary reporting standard, with any intra-patient results clearly labeled as ablation or sensitivity analyses. The synthesis of reviewed literature highlights that rigorous data curation and proper dataset partitioning contribute significantly more to trustworthy and generalizable results than methodological novelty alone. These best practices address four major recurring challenges in ECG beat classification: data imbalance, architectural design, training and evaluation methodology, and reproducibility. Therefore, it is increasingly accepted that inter-patient evaluation should serve as the primary reporting protocol, with intra-patient results presented only as supplementary ablation tests. Equally important is the transparent documentation of any pre-processing, i.e., band-pass filtering (BPF) parameters, baseline-wander removal, R-peak detection methods, and beat segmentation window lengths, since small differences in these steps can significantly alter results (57). Explicitly reporting such settings is now recognized as best practice for reproducibility.

The most persistent problem identified across datasets such as MIT-BIH Arrhythmia (8), INCART (53), and MIT-BIH Supraventricular Arrhythmia (3) is the severe class imbalance between normal and abnormal beats. To mitigate this, studies have increasingly combined algorithmic and data-level solutions. Among the algorithmic solutions, loss-function re-weighting or dynamic minority-biased batch weighting consistently improved sensitivity to rare arrhythmias without harming majority class precision. These methods dynamically scale gradients according to class frequency, preventing the network from converging toward trivial majority predictions. Complementary strategies such as feature-fusion architectures, which integrate temporal, frequency-domain, and time-frequency representations, were also effective, as they expose the classifier to richer discriminative cues and mitigate bias towards dominant waveform shapes (197). A smaller but influential group of studies used DBA or F1-based early-stopping criteria instead of overall accuracy, thereby explicitly optimizing for balanced recognition across classes. Collectively, these practices demonstrate that metric-aware optimization is more faithful to clinical relevance than simple accuracy maximization. At the data level, the combination of ECG-specific augmentation and transfer learning emerged as another reliable route to improved minority-class recall. Augmentations such as controlled time-shifting, amplitude scaling, additive Gaussian noise, and mild temporal warping expand training diversity while preserving physiological plausibility. Overly aggressive distortions, such as heavy frequency modulation, were found to degrade performance by altering beat morphology. Consequently, augmentations that respect signal physiology are now favored.

For smaller or domain-specific datasets, transfer learning—pre-training on a large balanced corpus and fine-tuning on the target data—consistently reduced overfitting and variance, especially for rare arrhythmias. Ensemble learning, though computationally heavier, repeatedly provided robustness against minority misclassifications by averaging the decisions of differently initialized or architected models (21). Together, these results establish a hierarchy of preferred imbalance-handling practices, namely, loss-level weighting, physiologically consistent augmentation, transfer learning, and lightweight ensembles, in order of practicality. Architectural trends in the literature show a clear evolution from manual feature engineering to end-to-end deep networks capable of automatically learning morphological and temporal patterns. Among these, 1D-CNNs remain the dominant backbone for local feature extraction because they effectively capture QRS-complex morphology and achieve high efficiency. When CNNs are extended with recurrent layers—especially LSTM or Bi-LSTM units—they gain temporal awareness, enabling them to exploit rhythm dependencies between consecutive beats. Hybrid CNN-LSTM architectures, sometimes referred to as hierarchical temporal models, consistently occupied the upper performance tier, with F-scores typically between 95% and 99% in benchmark datasets. These combinations proved particularly successful in distinguishing morphologically similar beats that differ primarily in timing rather than shape.

A further architectural enhancement repeatedly associated with improved interpretability and accuracy is the attention mechanism. Attention layers help the network emphasize diagnostically important regions of the ECG, such as the QRS onset and offset, while down-weighting redundant background segments. Across multiple studies, attention modules increased sensitivity to subtle morphological differences between ventricular and supraventricular ectopic beats, and they provided intuitive saliency maps that facilitate clinical interpretability. Where data and computational resources permitted, transformer-style multi-head attention further improved long-range context modeling, capturing dependencies across entire cardiac cycles rather than single beats. In parallel, modern convolutional refinements such as residual connections, squeeze-and-excitation blocks, and frequency-channel attention enhanced optimization stability and feature discrimination. These structural elements have therefore become standard components of competitive ECG classifiers. Beyond architecture, the review reveals consensus on several training and evaluation practices that determine whether good models generalize in clinically realistic conditions. Regularization through dropout, weight decay, and batch normalization is nearly universal among successful deep networks, preventing overfitting to small patient subsets. Early stopping based on validation F1-score or AUPRC proved more reliable than using loss reduction alone, because it guards against bias towards majority classes. Many articles also emphasize aligning model capacity with hardware constraints. For example, for wearable or edge deployment, compact 1D-CNNs with attention heads offer an attractive trade-off between interpretability, accuracy, and latency.

Reporting inference time per beat or per signal window is increasingly viewed as essential, since real-time performance is critical for embedded medical systems.

## 6.4 Future directions

The proposed ECG beat classification techniques in the literature provide promising performance, but there is a need to develop such methods to overcome the limitations of the current techniques. In this section, we discuss some of the open possible extensions of the techniques proposed in this article.

- *Personalized, adaptive models for individualized ECG classification:* One significant future direction is the development of personalized and adaptive models that tailor the ECG classification to the individual patient's data. Traditional models often struggle with variability in ECG signals across different patients due to factors such as age, physiology, and comorbid conditions. By incorporating patient-specific data into the training process, models can adjust their parameters to better reflect individual cardiac patterns. Transfer learning and online learning can fine-tune models using a patient's historical ECG data, leading to more accurate and reliable diagnoses (198, 199).
- *Explainable and interpretable deep learning models:* Deep learning models must be both explainable and interpretable for widespread clinical acceptance. Clinicians must understand the reasoning behind a model's prediction to trust and effectively use it in decision-making. Future research should focus on integrating explainable AI techniques, highlighting which parts of the ECG signal contribute the most to a classification decision. Methods such as attention mechanisms, saliency maps, and layer-wise relevance propagation can provide insight into the inner workings of the model (181, 200). By transparentizing the decision process, we can bridge the gap between complex algorithms and clinical practice, ensuring that these tools support rather than hinder healthcare professionals.
- *Efficient, real-time processing for wearable and mobile devices:* As wearable technology and mobile health applications become more prevalent, there is a growing need for algorithms that can process ECG data efficiently and in real-time. Future developments should aim to optimize models to run on devices with limited computational resources without sacrificing accuracy. This can be achieved through model compression techniques, such as quantization and pruning, and the design of lightweight architectures specifically tailored for edge computing (182, 201). Efficient algorithms enable continuous monitoring and prompt detection of cardiac anomalies, which is essential for timely medical responses and for improving patient care in everyday settings.
- *Unlabeled ECG data with advanced learning techniques:* A vast amount of ECG data remains unlabeled due to the time and expertise required for annotation. Unsupervised, semi-

supervised, and self-supervised learning techniques offer ways to utilize this untapped resource. Future research should focus on developing models that can learn meaningful representations from unlabeled data, thereby reducing the dependence on large labeled datasets. Techniques such as autoencoders, contrastive learning, and generative adversarial networks can uncover underlying patterns in the data, which can then be fine-tuned with minimal labeled examples (202, 203). Leveraging unlabeled data not only enhances model performance but also accelerates the development of robust ECG classification systems.

- *Performance with federated learning:* Patient privacy is of paramount concern in healthcare, limiting data sharing across institutions. Federated learning presents a solution by allowing models to be trained on decentralized data without transferring sensitive information. Future directions involve implementing federated learning frameworks for ECG analysis, enabling collaboration between hospitals and research centers while maintaining patient confidentiality (180, 197). This approach can lead to the creation of more generalized models trained on diverse datasets, improving accuracy and reliability across different populations. Addressing challenges such as communication efficiency and model convergence in federated settings will be essential for practical deployment.
- *Multi-modal learning:* Combining ECG data with other physiological signals or medical information through multi-modal learning can significantly improve diagnostic precision. Future work should explore integrating data from sources such as blood pressure monitors, oxygen saturation sensors, and patient medical histories. By providing a more comprehensive view of a patient's health status, models can make more informed predictions about cardiac events (204, 205). Multi-modal approaches can help identify complex conditions that may not be detectable through ECG analysis alone, leading to more holistic and effective patient care. Developing algorithms capable of processing and synthesizing information from multiple modalities will be a key area of focus.

## 7 Limitations of the review methodology

Despite the systematic and comprehensive approach adopted in this review paper, several methodological limitations must be acknowledged to ensure transparency and contextual accuracy in the interpretation of our findings. These limitations arise from the nature of systematic reviews themselves, the search and selection procedures, the heterogeneity of the included studies, and the diversity in experimental designs, databases, and performance metrics used in the studies on ECG beat classification. Recognizing these constraints provides a balanced understanding of the review outcomes and outlines potential directions for improving future systematic reviews in this field.

## 7.1 Scope and search limitations

The review process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure methodological rigor and reproducibility. Nevertheless, one inherent limitation stems from the restriction to peer-reviewed, English-language publications indexed in major databases such as PubMed, IEEE Xplore, ScienceDirect, and Springer. While this criterion ensured quality and accessibility, it potentially excluded significant research reported in non-English journals, regional repositories, theses, and conference proceedings not indexed in the selected databases. Excluding non-English studies may have introduced a language bias, especially given that ECG research is conducted globally, with active contributions from Asian, European, and South American institutions. Consequently, important innovations, particularly those published in national medical journals or non-indexed conference proceedings, may have been overlooked. This review paper provides a detailed analysis of machine learning and deep learning techniques for the classification of ECG beats, focusing on the advancements from 2014 to 2024, but it may have excluded earlier pioneering work that laid the groundwork for more advanced ECG analysis methods. A systematic approach was adopted to analyze the 106 studies, offering a comprehensive evaluation of methodologies and their applications in ECG classification. Another limitation relates to the deliberate exclusion of grey literature, including technical reports, dissertations, white papers, and non-peer-reviewed conference abstracts. While this approach enhanced the credibility of the included studies by ensuring they met peer-review standards, it may have resulted in the omission of innovative but unpublished work or early-stage algorithmic implementations. Grey literature often contains valuable methodological insights, comparative evaluations, or negative findings that are less likely to appear in journal publications due to publication bias.

A major methodological limitation of this review arises from the heterogeneity across the datasets used in the included studies. The three most frequently employed databases—MIT-BIH Arrhythmia (8), St. Petersburg INCART (53), and MIT-BIH Supraventricular Arrhythmia (3)—differ substantially in sampling frequency, signal resolution, lead configuration, and patient demographics. Consequently, performance metrics such as accuracy, sensitivity, and F1-score cannot be directly compared across studies, as the underlying data distributions vary considerably. For instance, the MIT-BIH Arrhythmia database contains 48 half-hour two-lead ECG recordings from 47 subjects sampled at 360 Hz, whereas the INCART dataset (53) includes 75 recordings at 257 Hz from 25 patients with different arrhythmia profiles. These discrepancies introduced variability in the reported results, even when identical algorithms are used. Furthermore, some studies utilized patient-specific evaluation protocols (intra-patient validation), while others employed inter-patient cross-validation strategies. The choice between these validation schemes has a profound effect on reported accuracy. Models evaluated using intra-patient splits

typically yield inflated performance metrics since the test data shares morphological characteristics with the training data. Conversely, inter-patient validation better represents real-world generalization but usually yields lower accuracy. The lack of standardized evaluation protocols across the studies makes it difficult to provide an entirely uniform assessment of algorithmic performance.

## 8 Recommendations from the critical review

Based on our comprehensive systematic review of 106 studies on the state-of-the-art machine and deep learning techniques for ECG beat classification, we propose the following recommendations to advance the field and address existing limitations. Addressing the prevalent issue of data imbalance, techniques such as SMOTE, GANs, and transfer learning should be employed to enhance model performance on underrepresented arrhythmia classes. Robust pre-processing pipelines using modern denoising methods, such as deep learning-based filtering and wavelet transforms, are crucial for mitigating noise artifacts such as baseline wander and motion interference. Hybrid and ensemble models that combine handcrafted and deep learning-based features can leverage domain knowledge while improving classification robustness. The adoption of transformer architectures and attention mechanisms can better capture long-term dependencies in ECG signals, while patient-specific and adaptive models, utilizing transfer learning and federated learning, can address inter-patient variability. To enhance real-world applicability, optimizing models for deployment on wearable and Internet of Things (IoT) devices, with a focus on lightweight architectures and edge computing, is essential. Furthermore, standardizing evaluation metrics and benchmarking datasets, fostering interdisciplinary collaborations, integrating multimodal data sources, and emphasizing ethical considerations such as data privacy and XAI will collectively ensure the development of clinically relevant and scalable ECG classification systems. These recommendations aim to address current challenges in ECG beat classification and pave the way for innovative, clinically relevant, and scalable solutions. By focusing on these areas, researchers and practitioners can enhance the diagnostic accuracy and real-world applicability of automated ECG analysis systems.

## 9 Conclusions

This systematic review highlights growing trends in deep learning for ECG classification, with CNN and hybrid models showing consistently high performance across benchmarks. Despite advancements, clinical translation is limited by bias, data imbalance, and lack of interpretability. While the field has witnessed significant progress, particularly with the adoption of deep learning methods such as CNNs and RNNs, challenges

remain. These include noise and artifacts in ECG signals, data imbalance, the need for extensive annotated datasets, computational resource constraints, and the lack of generalizability to unseen data. In addition, many traditional approaches rely on handcrafted features, which may not fully capture the complexity and variability of ECG signals. Deep learning models have demonstrated superior performance in ECG beat classification tasks, offering the ability to automatically learn features from raw data and effectively capture temporal and morphological patterns. However, the “black-box” nature of these models raises concerns about interpretability, which is critical for clinical applications. Addressing these limitations requires the development of explainable AI algorithms that provide transparent decision-making while maintaining the high accuracy of deep learning models. This review underscores the need for robust pre-processing pipelines, advanced data augmentation strategies, and the integration of multimodal data to improve classification performance. Furthermore, interdisciplinary collaboration between researchers, clinicians, and engineers is essential to ensure that the developed systems align with clinical needs. By addressing these challenges and embracing emerging techniques, such as transfer learning, explainable AI, and real-time processing on wearable devices, future research can pave the way for more accurate, interpretable, and scalable ECG beat classification systems that improve patient care and diagnostic efficiency.

## 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

AJ: Investigation, Conceptualization, Software, Formal analysis, Writing – review & editing, Methodology, Writing – original draft. AN: Supervision, Project administration, Writing – review & editing. IE: Funding acquisition, Supervision, Writing – review & editing, Project administration.

## References

1. Strik M, Ploux S, Weigel D, van der Zande J, Velraeds A, Racine HP, et al. The use of smartwatch electrocardiogram beyond arrhythmia detection. *Trends Cardiovasc Med.* (2024) 34:174–80. doi: 10.1016/j.tcm.2022.12.006
2. Mirvis DM, Goldberger AL. Electrocardiography. *Heart Dis.* (2001) 1:82–128.
3. Data from: The MIT-BIH supraventricular arrhythmia database. (2012). Available online at: <http://www.physionet.org/physiobank/database/svdb/> (Accessed March 1, 2012).
4. Chandra S, Sharma A, Singh GK. Feature extraction of ECG signal. *J Med Eng Technol.* (2018) 42:306–16. doi: 10.1080/03091902.2018.1492039
5. Gaidai O, Cao Y, Loginov S. Global cardiovascular diseases death rate prediction. *Curr Probl Cardiol.* (2023) 48:101622. doi: 10.1016/j.cpcardiol.2023.101622
6. Ismail SNMS, Razak SFA, Ab Aziz NA. ECG-based transfer learning for cardiovascular disease: a scoping review. *Int J Cogn Comput Eng.* (2025) 6:280–97. doi: 10.1016/j.ijcce.2025.01.007
7. Prakash AJ, Samantray S, Sahoo SP, Ari S. A deformable CNN architecture for predicting clinical acceptability of ECG signal. *Biocybern Biomed Eng.* (2023) 43:335–51. doi: 10.1016/j.bbe.2023.01.006
8. Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. *IEEE Eng Med Biol Mag.* (2001) 20:45–50. doi: 10.1109/51.932724
9. Prakash AJ, Ari S. 35—a system for automatic cardiac arrhythmia recognition using electrocardiogram signal. In: Pal K, Kraatz HB, Khasnobish A, Bag S, Banerjee I, Kuruganti U, editors. *Bioelectronics and Medical Devices*. Cambridge:

HJ: Project administration, Writing – review & editing, Supervision. MA: Supervision, Writing – review & editing, Writing – original draft, Resources, Funding acquisition.

## Funding

The author(s) declare financial support was received for the research and/or publication of this article. This work is supported by the KU-UAEU Joint Research Program Award No. KU-UAEU-2023-004 with UAEU grant code G00004680.

## 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.

## Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. In the preparation of this manuscript, AI-assisted technologies, specifically Grammarly, were used to enhance language clarity, grammar, and readability. No generative AI tools were employed to generate original content, data, or analysis.

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.

Woodhead Publishing (2019), Woodhead Publishing Series in Electronic and Optical Materials; p. 891–911.

10. Peimankar A, Ebrahimi A, Will UK. xECG-Beats: an explainable deep transfer learning approach for ecg-based heartbeat classification. *Netw Model Anal Health Inform Bioinform.* (2024) 13:1–13. doi: 10.1007/s13721-024-00481-2
11. Prakash AJ, Ari S. Aami standard cardiac arrhythmia detection with random forest using mixed features. In: *2019 IEEE 16th India Council International Conference (INDICON)*. IEEE (2019). p. 1–4.
12. Tsipouras MG, Fotiadis DI, Sideris D. Arrhythmia classification using the RR-interval duration signal. In: *Computers in Cardiology*. IEEE (2002). p. 485–8.
13. Al-Ani M, Rawi AA. A rule-based expert system for automated ecg diagnosis. *Int J Adv Eng Technol.* (2014) 6:1480–92.
14. Xu M, Wei S, Qin X, Zhang Y, Liu C. Rule-based method for morphological classification of st segment in ECG signals. *J Med Biol Eng.* (2015) 35:816–23. doi: 10.1007/s40846-015-0092-x
15. Bansal G, Gera P, Bathula DR. Template based classification of cardiac arrhythmia in ECG data. In: *2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)*. (2015). p. 337–41.
16. Silva-Filander MD, Marzbanrad F. Combining template-based and feature-based classification to detect atrial fibrillation from a short single lead ECG recording. In: *2017 Computing in Cardiology (CinC)*. (2017). p. 1–4.
17. Medhi K, Hussain MI. Lightweight arrhythmia detection using cross-correlation. In: *2021 4th International Conference on Recent Trends in Computer Science and Technology (ICRTCTST)*. (2022). p. 296–301.
18. Ferretti J, Randazzo V, Cirrincione G, Pasero E. 1-D convolutional neural network for ECG arrhythmia classification. *Progr Artif Intell Neural Syst.* (2021) 184:269–79. doi: 10.1007/978-981-15-5093-5\_25
19. Sakli N, Ghabri H, Soufiene BO, Almalki F, Sakli H, Ali O, et al. Resnet-50 for 12-lead electrocardiogram automated diagnosis. *Comput Intell Neurosci.* (2022) 2022:1.16 doi: 10.1155/2022/7617551
20. Özal Y, Plawiak P, Tan RS, Acharya UR. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Comput Biol Med.* (2018) 102:411–20. doi: 10.1016/j.combiomed.2018.09.009
21. Banerjee S, Singh GK. A new moving horizon estimation based real-time motion artifact removal from wavelet subbands of ECG signal using particle filter. *J Signal Process Syst.* (2023) 11:1–15. doi: 10.1007/s11265-023-01887-3
22. Rabee A, Barhumli I. ECG signal classification using support vector machine based on wavelet multiresolution analysis. In: *2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)*. (2012). p. 1319–23.
23. Pan G, Xin Z, Shi S, Jin D. Arrhythmia classification based on wavelet transformation and random forests. *Multimed Tools Appl.* (2018) 77:21905–22. doi: 10.1007/s11042-017-5225-5
24. Rangappa VG, Prasad S, Agarwal A. Classification of cardiac arrhythmia stages using hybrid features extraction with k-nearest neighbour classifier of ECG signals. *Learning.* (2018) 11:21–32. doi: 10.22266/ijies2018.1231.03
25. Rahul J, Sharma LD. Artificial intelligence-based approach for atrial fibrillation detection using normalised and short-duration time-frequency ECG. *Biomed Signal Process Control.* (2022) 71:103270. doi: 10.1016/j.bspc.2021.103270
26. Zheng Z, Chen Z, Hu F, Zhu J, Tang Q, Liang Y. An automatic diagnosis of arrhythmias using a combination of CNN and LSTM technology. *Electronics.* (2020) 9:121. doi: 10.3390/electronics9010121
27. Chen C, Hua Z, Zhang R, Liu G, Wen W. Automated arrhythmia classification based on a combination network of CNN and LSTM. *Biomed Signal Process Control.* (2020) 57:101819. doi: 10.1016/j.bspc.2019.101819
28. Allam JP, Sahoo SP, Ari S. Multi-stream bi-gru network to extract a comprehensive feature set for ECG signal classification. *Biomed Signal Process Control.* (2024) 92:106097. doi: 10.1016/j.bspc.2024.106097
29. Das MK, Ari S. Patient-specific ECG beat classification technique. *Health Technol Lett.* (2014) 1:98–103. doi: 10.1049/htl.2014.0072
30. Sahoo S, Dash M, Behera S, Sabut S. Machine learning approach to detect cardiac arrhythmias in ECG signals: a survey. *IRBM.* (2020) 41:185–94. doi: 10.1016/j.irbm.2019.12.001
31. Gupta V, Mittal M, Mittal V, Saxena NK. A critical review of feature extraction techniques for ECG signal analysis. *J Inst Eng (India) Ser B.* (2021) 102:1049–60. doi: 10.1007/s40031-021-00606-5
32. Chatterjee S, Thakur RS, Yadav RN, Gupta L. Sparsity-based modified wavelet de-noising autoencoder for ECG signals. *Signal Processing.* (2022) 198:108605. doi: 10.1016/j.sigpro.2022.108605
33. De Chazel P, Reilly R. A comparison of the ECG classification performance of different feature sets. In: *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)*. IEEE (2000). p. 327–30.
34. Wechsler SB, Wernovsky G. Cardiac disorders. *Man Neonatal Care.* (2008) 5:407.
35. Martis RJ, Acharya UR, Min LC. ECG beat classification using PCA, LDA, ICA and discrete wavelet transform. *Biomed Signal Process Control.* (2013) 8:437–48. doi: 10.1016/j.bspc.2013.01.005
36. Ye C, Vijaya Kumar BVK, Coimbra MT. Heartbeat classification using morphological and dynamic features of ECG signals. *IEEE Trans Biomed Eng.* (2012) 59:2930–41. doi: 10.1109/TBME.2012.2213253
37. Zhai X, Zhou Z, Tin C. Semi-supervised learning for ECG classification without patient-specific labeled data. *Expert Syst Appl.* (2020) 158:113411. doi: 10.1016/j.eswa.2020.113411
38. Al-Zuhairi DT, Hameed AS, Hameed IS. Dual-stage SVD basis approach for ECG signal associated noise removal. *Signal Image Video Process.* (2022) 16:1489–96. doi: 10.1007/s11760-021-02102-1
39. Mewada H. 2D-wavelet encoded deep CNN for image-based ECG classification. *Multimed Tools Appl.* (2023) 82:1–17. doi: 10.1007/s11042-022-14302-z
40. Varalakshmi P, Sankaran AP. An improved hybrid AI model for prediction of arrhythmia using ECG signals. *Biomed Signal Process Control.* (2023) 80:104248. doi: 10.1016/j.bspc.2022.104248
41. Rauf A, Muhammad N, Mahmood H, Aftab M. Healthcare service quality: a systematic review based on prisma guidelines. *Int J Qual Reliab Manag.* (2025) 42 (3):837–50. doi: 10.1108/IJQRM-12-2023-0403
42. Sarkis-Onofre R, Catalá-López F, Aromataris E, Lockwood C. How to properly use the prisma statement. *Syst Rev.* (2021) 10:1–3. doi: 10.1186/s13643-021-01671-z
43. Jambukia SH, Dabhi VK, Prajapati HB. ECG beat classification using machine learning techniques. *Int J Biomed Eng Technol.* (2018) 26:32–53. doi: 10.1504/IJBET.2018.089255
44. Jambukia SH, Dabhi VK, Prajapati HB. Classification of ECG signals using machine learning techniques: a survey. In: *2015 International Conference on Advances in Computer Engineering and Applications*. IEEE (2015). p. 714–21.
45. Li Q, Rajagopalan C, Clifford GD. Ventricular fibrillation and tachycardia classification using a machine learning approach. *IEEE Trans Biomed Eng.* (2013) 61:1607–13. doi: 10.1109/TBME.2013.2275000
46. Tang X, Ma Z, Hu Q, Tang W. A real-time arrhythmia heartbeats classification algorithm using parallel delta modulations and rotated linear-kernel support vector machines. *IEEE Trans Biomed Eng.* (2019) 67:978–86. doi: 10.1109/TBME.2019.2926104
47. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation.* (2000) 101:e215–20. doi: 10.1161/01.CIR.101.23.e215
48. Taddei A, Distant G, Emdin M, Pisani P, Moody G, Zeelenberg C, et al. The European ST-T database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. *Eur Heart J.* (1992) 13:1164–72. doi: 10.1093/oxfordjournals.eurheartj.a060332
49. Petrutiu S, Sahakian A, Swiryn S. Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans. *Europace.* (2007) 9:466–70. doi: 10.1093/europace/eum096
50. Moody G. Spontaneous termination of atrial fibrillation: a challenge from physionet and computers in cardiology 2004. In: *Computers in Cardiology, 2004. IEEE (2004)*. Vol. 31. p. 101–4.
51. Li X, Han H, Lu H, Niu X, Yu Z, Dantcheva A, et al. The 1st challenge on remote physiological signal sensing (REPS). In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. (2020). p. 314–5.
52. Silva I, Moody GB, Celi L. Improving the quality of ECGs collected using mobile phones: the physionet/computing in cardiology challenge 2011. In: *2011 Computing in Cardiology*. IEEE (2011). p. 273–6.
53. Berrahou N, El Alami A, Mesbah A, El Alami R, Berrahou A. Arrhythmia detection in inter-patient ECG signals using entropy rate features and RR intervals with CNN architecture. *Comput Methods Biomech Biomed Engin.* (2024) 1:1–20. doi: 10.1080/10255842.2024.2378105
54. Fernández A, García S, Herrera F, Chawla NV. Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *J Artif Intell Res.* (2018) 61:863–905. doi: 10.1613/jair.1.11192
55. Thabtah F, Hammoud S, Kamalov F, Gonsalves A. Data imbalance in classification: experimental evaluation. *Inf Sci (Nij).* (2020) 513:429–41. doi: 10.1016/j.ins.2019.11.004
56. Biswas U, Maniruzzaman M. Removing power line interference from ECG signal using adaptive filter and notch filter. In: *2014 International Conference on Electrical Engineering and Information & Communication Technology*. IEEE (2014). p. 1–4.
57. Hintermüller C, Hirschrodt M, Blessberger H, Steinwender C. ECG beat classification: impact of linear dependent samples. *Curr Dir Biomed Eng.* (2023) 9:23–6. doi: 10.1515/cdbme-2023-1207
58. Vadillo-Valderrama A, Goya-Esteban R, Caulier-Cisterna R, García-Alberola A, Rojo-Álvarez J. Differential beat accuracy for ECG family classification using machine learning. *IEEE Access.* (2022) 10:129362–81. doi: 10.1109/ACCESS.2022.3227219

59. Cai J, Song J, Peng B. Enhancing ECG heartbeat classification with feature fusion neural networks and dynamic minority-biased batch weighting loss function. *Physiol Meas.* (2024) 45:075002. doi: 10.1088/1361-6579/ad5cc0
60. Palreddy S, Hu Y, Mani VP, Tompkins W. A multiple-classifier architecture for ECG beat classification. In: *Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop.* (1997). p. 172–81.
61. Behera B, Kumaravelan G, Kumar P. Performance evaluation of deep learning algorithms in biomedical document classification. In: *2019 11th International Conference on Advanced Computing (ICoAC).* IEEE (2019). p. 220–4.
62. Addison PS. Wavelet transforms and the ECG: a review. *Physiol Meas.* (2005) 26:R155. doi: 10.1088/0967-3334/26/5/R01
63. Altay Y, Kremlev A, Zimenko K, Margun A. The effect of filter parameters on the accuracy of ECG signal measurement. *Biomed Eng.* (2019) 53:176–80. doi: 10.1007/s10527-019-09903-2
64. Caref EB, Turitto G, Ibrahim BB, Henkin R, El-Sherif N. Role of bandpass filters in optimizing the value of the signal-averaged electrocardiogram as a predictor of the results of programmed stimulation. *Am J Cardiol.* (1989) 64:16–26. doi: 10.1016/0002-9149(89)90646-2
65. Ahlstrom M, Tompkins W. Digital filters for real-time ECG signal processing using microprocessors. *IEEE Trans Biomed Eng.* (1985) BME-32:708–13. doi: 10.1109/TBME.1985.325589
66. Al-Qawasmī AR, Daqrouq K. ECG signal enhancement using wavelet transform. *WSEAS Trans on Biol Biomed.* (2010) 7:62–72.
67. Kumar A, Tomar H, Mehla VK, Komaragiri R, Kumar M. Stationary wavelet transform based ECG signal denoising method. *ISA Trans.* (2021) 114:251–62. doi: 10.1016/j.isatra.2020.12.029
68. Khaja B, Kalluci E, Nikolla L. Wavelet transform applied in ECG signal processing. *Eur Sci J.* (2015) 11.
69. Castells F, Laguna P, Sörnmo L, Bollmann A, Roig JM. Principal component analysis in ECG signal processing. *EURASIP J Adv Signal Process.* (2007) 2007:1–21. doi: 10.1155/2007/74580
70. Martis RJ, Acharya UR, Mandana K, Ray AK, Chakraborty C. Application of principal component analysis to ECG signals for automated diagnosis of cardiac health. *Expert Syst Appl.* (2012) 39:11792–800. doi: 10.1016/j.eswa.2012.04.072
71. Savitha R, Breesha S, Joseph XF. Pre processing the abdominal ECG signal using combination of fir filter and principal component analysis. In: *2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015].* IEEE (2015). p. 1–4.
72. Pongponsri S, Yu XH. An adaptive filtering approach for electrocardiogram (ECG) signal noise reduction using neural networks. *Neurocomputing.* (2013) 117:206–13. doi: 10.1016/j.neucom.2013.02.010
73. Thakor NV, Zhu YS. Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection. *IEEE Trans Biomed Eng.* (1991) 38:785–94. doi: 10.1109/10.83591
74. Romero I, Geng D, Berset T. Adaptive filtering in ECG denoising: a comparative study. In: *2012 Computing in Cardiology.* IEEE (2012). p. 45–8.
75. Chang KM, Liu SH. Gaussian noise filtering from ECG by wiener filter and ensemble empirical mode decomposition. *J Signal Process Syst.* (2011) 64:249–64. doi: 10.1007/s11265-009-0447-z
76. Kumar S, Panigrahy D, Sahu P. Denoising of electrocardiogram (ECG) signal by using empirical mode decomposition (EMD) with non-local mean (NLM) technique. *Biocybern Biomed Eng.* (2018) 38:297–312. doi: 10.1016/j.bbe.2018.01.005
77. Han G, Lin B, Xu Z. Electrocardiogram signal denoising based on empirical mode decomposition technique: an overview. *J Instrum.* (2017) 12:P03010. doi: 10.1088/1748-0221/12/03/P03010
78. Bai YW, Chu WY, Chen CY, Lee YT, Tsai YC, Tsai CH. Adjustable 60 Hz noise reduction by a notch filter for ECG signals. In: *Proceedings of the 21st IEEE Instrumentation and Measurement Technology Conference (IEEE Cat. No. 04CH37510).* IEEE (2004). Vol. 3. p. 1706–11.
79. Verma A, Singh Y. Adaptive tunable notch filter for ECG signal enhancement. *Procedia Comput Sci.* (2015) 57:332–7. doi: 10.1016/j.procs.2015.07.347
80. Raj S, Ray KC. Sparse representation of ECG signals for automated recognition of cardiac arrhythmias. *Expert Syst Appl.* (2018) 105:49–64. doi: 10.1016/j.eswa.2018.03.038
81. Zhou Y, Hu X, Tang Z, Ahn AC. Sparse representation-based ECG signal enhancement and QRS detection. *Physiol Meas.* (2016) 37:2093. doi: 10.1088/0967-3334/37/12/2093
82. He W, Wang G, Hu J, Li C, Guo B, Li F. Simultaneous human health monitoring and time-frequency sparse representation using EEG and ECG signals. *IEEE Access.* (2019) 7:85985–94. doi: 10.1109/ACCESS.2019.2921568
83. Singh P, Pradhan G. Variational mode decomposition based ECG denoising using non-local means and wavelet domain filtering. *Australas Phys Eng Sci Med.* (2018) 41:891–904. doi: 10.1007/s13246-018-0685-0
84. Singh P, Shah Nawazuddin S, Pradhan G. An efficient ECG denoising technique based on non-local means estimation and modified empirical mode decomposition. *Circuits Syst Signal Process.* (2018) 37:4527–47. doi: 10.1007/s00034-018-0777-9
85. Pravin C, Ojha V. A novel ECG signal denoising filter selection algorithm based on conventional neural networks. In: *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA).* IEEE (2020). p. 1094–100.
86. Nurmaini S, Darmawahyuni A, Sakti Mukti AN, Rachmatullah MN, Firdaus F, Tutuko B. Deep learning-based stacked denoising and autoencoder for ECG heartbeat classification. *Electronics.* (2020) 9:135. doi: 10.3390/electronics9010135
87. Hou Y, Liu R, Shu M, Xie X, Chen C. Deep neural network denoising model based on sparse representation algorithm for ECG signal. *IEEE Trans Instrum Meas.* (2023) 72:1–11. doi: 10.1109/TIM.2023.3251408
88. Kumar SS, Mohan N, Prabakaran P, Soman K. Total variation denoising based approach for R-peak detection in ECG signals. *Procedia Comput Sci.* (2016) 93:697–705. doi: 10.1016/j.procs.2016.07.268
89. Singh O, Sunkaria RK. Powerline interference reduction in ECG signals using empirical wavelet transform and adaptive filtering. *J Med Eng Technol.* (2015) 39:60–8. doi: 10.3109/03091902.2014.979954
90. Malik SA, Parah SA, Malik BA. Power line noise and baseline wander removal from ECG signals using empirical mode decomposition and lifting wavelet transform technique. *Health Technol.* (2022) 12:745–56. doi: 10.1007/s12553-022-00662-x
91. Ramakrishnan S, Yogeswaran R. Design and analysis of feature extraction algorithm for ECG signals using adaptive threshold method. In: *2017 Trends in Industrial Measurement and Automation (TIMA).* IEEE (2017). p. 1–8.
92. Fatimah B, Singhal A, Singh P. ECG arrhythmia detection in an inter-patient setting using fourier decomposition and machine learning. *Med Eng Phys.* (2024) 124:104102. doi: 10.1016/j.medengphy.2024.104102
93. Hussein A, Mohammed WR, Musa Jaber M, Ibrahim Khalaf O. An adaptive ECG noise removal process based on empirical mode decomposition (EMD). *Contrast Media Mol Imaging.* (2022) 2022. doi: 10.1155/2022/3346055
94. Faiz MMU, Kale I. Removal of multiple artifacts from ECG signal using cascaded multistage adaptive noise cancellers. *Array.* (2022) 14:100133. doi: 10.1016/j.array.2022.100133
95. Kollem S, Baig MR, Lasya D, Kalyan EA, Varma KN. ECG noise removal using FCN DAE method. In: *2022 2nd International Conference on Intelligent Technologies (CONIT).* IEEE (2022). p. 1–8.
96. Li H, Ditzler G, Roveda J, Li A. Descod-ECG: deep score-based diffusion model for ECG baseline wander and noise removal. *IEEE J Biomed Health Inform.* (2023) 28:5081–91. doi: 10.1109/JBHI.2023.3237712
97. Nivitha V V, Cao H, Peyrodie L. Variational mode decomposition-based simultaneous R peak detection and noise suppression for automatic ECG analysis. *IEEE Sens J.* (2023) 23:8703–13. doi: 10.1109/JSEN.2023.3257332
98. Zidelmal Z, Amirou A, Abdeslam D, Mercklé J. ECG beat classification using a cost-sensitive classifier. *Comput Methods Programs Biomed.* (2013) 111:570–7. doi: 10.1016/j.cmpb.2013.05.011
99. Mert A, Kılıç N, Akan A. Evaluation of bagging ensemble method with time-domain feature extraction for diagnosing of arrhythmia beats. *Neural Comput Appl.* (2014) 24:317–26. doi: 10.1007/s00521-012-1232-7
100. Kuila S, Dhanda N, Joardar S. Feature extraction of electrocardiogram signal using machine learning classification. *Int J Electr Comput Eng.* (2020) 10:6598–605. doi: 10.11591/ijece.v10i6.pp6598-6605
101. Li H, Feng X, Cao L, Li E, Liang H, Chen X. A new ECG signal classification based on WPD and apen feature extraction. *Circuits Syst Signal Process.* (2016) 35:339–52. doi: 10.1007/s00034-015-0068-7
102. Khan TT, Sultana N, Reza RB, Mostafa R. ECG feature extraction in temporal domain and detection of various heart conditions. In: *2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT).* IEEE (2015). p. 1–6.
103. Cai J, Zhou G, Dong M, Hu X, Liu G, Ni W. Real-time arrhythmia classification algorithm using time-domain ECG feature based on FFNN and CNN. *Math Probl Eng.* (2021) 2021:1–17. doi: 10.1155/2021/6648432
104. Sridhar C, Acharya UR, Fujita H, Bairy GM. Automated diagnosis of coronary artery disease using nonlinear features extracted from ECG signals. In: *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC).* IEEE (2016). p. 000545–9.
105. Lin CH. Frequency-domain features for ECG beat discrimination using grey relational analysis-based classifier. *Comput Math Appl.* (2008) 55:680–90. doi: 10.1016/j.camwa.2007.04.035
106. Asgharzadeh-Bonab A, Amirani MC, Mehri A. Spectral entropy and deep convolutional neural network for ECG beat classification. *Biocybern Biomed Eng.* (2020) 40:691–700. doi: 10.1016/j.bbe.2020.02.004
107. Fattah SA, Jameel ASMM, Goswami R, Saha SK, Syed N, Akter S, et al. An approach for human identification based on time and frequency domain features extracted from ECG signals. In: *TENCON 2011-2011 IEEE Region 10 Conference.* IEEE (2011). p. 259–63. doi: 10.1186/s12911-020-01337-1

108. Hu Y, Zhao Y, Liu J, Pang J, Zhang C, Li P. An effective frequency-domain feature of atrial fibrillation based on time-frequency analysis. *BMC Med Inform Decis Mak.* (2020) 20:1–11.
109. Chen S, Hua W, Li Z, Li J, Gao X. Heartbeat classification using projected and dynamic features of ECG signal. *Biomed Signal Process Control.* (2017) 31:165–73. doi: 10.1016/j.bspc.2016.07.010
110. Mitra S, Mitra M, Chaudhuri B. Generation of digital time database from paper ECG records and fourier transform-based analysis for disease identification. *Comput Biol Med.* (2004) 34:551–60. doi: 10.1016/j.compbimed.2003.08.001
111. Banerjee S, Mitra M. Application of cross wavelet transform for ECG pattern analysis and classification. *IEEE Trans Instrum Meas.* (2013) 63:326–33. doi: 10.1109/TIM.2013.2279001
112. Jaya Prakash A, Ari S. Spec: a system for patient specific ECG beat classification using deep residual network. *Biocybern Biomed Eng.* (2020) 40:1446–57. doi: 10.1016/j.bbe.2020.08.001
113. Kłosowski G, Rymarczyk T, Wójcik D, Skowron S, Cieplak T, Adamkiewicz P. The use of time-frequency moments as inputs of LSTM network for ECG signal classification. *Electronics.* (2020) 9:1452. doi: 10.3390/electronics9091452
114. Özal Y. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Comput Biol Med.* (2018) 96:189–202. doi: 10.1016/j.compbimed.2018.03.016
115. Shaker AM, Tantawi M, Shedeed HA, Tolba MF. Generalization of convolutional neural networks for ECG classification using generative adversarial networks. *IEEE Access.* (2020) 8:35592–605. doi: 10.1109/ACCESS.2020.2974712
116. Siouda R, Nemissi M, Seridi H. ECG beat classification using neural classifier based on deep autoencoder and decomposition techniques. *Prog Artif Intell.* (2021) 10:333–47. doi: 10.1007/s13748-021-00243-5
117. Sayantan G, Kien P, Kadambari K. Classification of ECG beats using deep belief network and active learning. *Med Biol Eng Comput.* (2018) 56:1887–98. doi: 10.1007/s11517-018-1815-2
118. van de Leur RR, Bos MN, Taha K, Sammani A, Yeung MW, van Duijvenboden S, et al. Improving explainability of deep neural network-based electrocardiogram interpretation using variational auto-encoders. *Eur Heart J Digit Health.* (2022) 3:390–404. doi: 10.1093/ehjdh/ztac038
119. Ge D, Srinivasan N, Krishnan SM. Cardiac arrhythmia classification using autoregressive modeling. *Biomed Eng Online.* (2002) 1:1–12. doi: 10.1186/1475-925X-1-5
120. Hao C, Wibowo S, Majmudar M, Rajput KS. Spectro-temporal feature based multi-channel convolutional neural network for ECG beat classification. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* (2019). p. 5642–5.
121. Yu SN, Chou KT. Integration of independent component analysis and neural networks for ECG beat classification. *Expert Syst Appl.* (2008) 34:2841–6. doi: 10.1016/j.eswa.2007.05.006
122. Plawiak P, Acharya UR. Novel deep genetic ensemble of classifiers for arrhythmia detection using ECG signals. *Neural Comput Appl.* (2020) 32:11137–61. doi: 10.1007/s00521-018-03980-2
123. Kandala RNVPS, Dhuli R, Plawiak P, Naik GR, Moeinzadeh H, Gargiulo GD, et al. Towards real-time heartbeat classification: evaluation of nonlinear morphological features and voting method. *Sensors.* (2019) 19:5079. doi: 10.3390/s19235079
124. Plawiak P. Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals. *Swarm Evol Comput.* (2018) 39:192–208. doi: 10.1016/j.swevo.2017.10.002
125. Acharya UR, Fujita H, Oh SL, Hagiwara Y, Tan JH, Adam M, et al. Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals. *Appl Intell.* (2019) 49:16–27. doi: 10.1007/s10489-018-1179-1
126. Kampouraki A, Manis G, Nikou C. Heartbeat time series classification with support vector machines. *IEEE Trans Inf Technol Biomed.* (2009) 13:512–8. doi: 10.1109/TITB.2008.2003323
127. Kiranyaz S, Ince T, Gabbouj M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Trans Biomed Eng.* (2015) 63:664–75. doi: 10.1109/TBME.2015.2468589
128. Zhang Y, Li J, Wei S, Zhou F, Li D. Heartbeats classification using hybrid time-frequency analysis and transfer learning based on ResNet. *IEEE J Biomed Health Inform.* (2021) 25:4175–84. doi: 10.1109/JBHI.2021.3085318
129. Boda S, Mahadevappa M, Dutta PK. An automated patient-specific ECG beat classification using lstm-based recurrent neural networks. *Biomed Signal Process Control.* (2023) 84:104756. doi: 10.1016/j.bspc.2023.104756
130. Barlow HB. Unsupervised learning. *Neural Comput.* (1989) 1:295–311. doi: 10.1162/neco.1989.1.3.295
131. Kaur M, Arora A. Unsupervised analysis of arrhythmias using k-means clustering. *Int J Comput Sci Inform Technol.* (2010) 1:417–9.
132. Van Engelen JE, Hoos HH. A survey on semi-supervised learning. *Mach Learn.* (2020) 109:373–440. doi: 10.1007/s10994-019-05855-6
133. Zhang P, Chen Y, Lin F, Wu S, Yang X, Li Q. Semi-supervised learning for automatic atrial fibrillation detection in 24-hour holter monitoring. *IEEE J Biomed Health Inform.* (2022) 26:3791–801. doi: 10.1109/JBHI.2022.3173655
134. Suthaharan S. Supervised learning algorithms. In: *Machine Learning Models and Algorithms for Big Data Classification. Integrated Series in Information Systems.* Vol 36. Boston, MA: Springer (2016). doi: 10.1007/978-1-4899-7641-3\_8
135. Huang J, Chen B, Yao B, He W. ECG arrhythmia classification using stft-based spectrogram and convolutional neural network. *IEEE Access.* (2019) 7:92871–80. doi: 10.1109/ACCESS.2019.2928017
136. Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adam M, Gertych A, et al. A deep convolutional neural network model to classify heartbeats. *Comput Biol Med.* (2017) 89:389–96. doi: 10.1016/j.compbimed.2017.08.022
137. Sannino G, De Pietro G. A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. *Future Gener Comput Syst.* (2018) 86:446–55. doi: 10.1016/j.future.2018.03.057
138. Melgani F, Bazi Y. *Evolutionary Computation Approach to ECG Signal Classification.* Setúbal: SciTePress (2008). p. 19–24.
139. Tuncer T, Dogan S, Plawiak P, Acharya UR. Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals. *Knowl Based Syst.* (2019) 186:104923. doi: 10.1016/j.knosys.2019.104923
140. Allam JP, Samantray S, Ari S. Patient-specific ECG beat classification using emd and deep learning-based technique. In: *Advanced Methods in Biomedical Signal Processing and Analysis.* New York, NY: Elsevier (2023). p. 87–108.
141. Gan Y, Shi J, He W, Sun F. Parallel classification model of arrhythmia based on DenseNet-BiLSTM. *Biocybern Biomed Eng.* (2021) 41:1548–60. doi: 10.1016/j.bbe.2021.09.001
142. Rahul J, Sharma LD. Automatic cardiac arrhythmia classification based on hybrid 1-D CNN and Bi-LSTM model. *Biocybern Biomed Eng.* (2022) 42:312–24. doi: 10.1016/j.bbe.2022.02.006
143. Andersen RS, Peimankar A, Puthusserypady S. A deep learning approach for real-time detection of atrial fibrillation. *Expert Syst Appl.* (2019) 115:465–73. doi: 10.1016/j.eswa.2018.08.011
144. Oh SL, Ng EY, San Tan R, Acharya UR. Automated beat-wise arrhythmia diagnosis using modified U-Net on extended electrocardiographic recordings with heterogeneous arrhythmia types. *Comput Biol Med.* (2019) 105:92–101. doi: 10.1016/j.compbimed.2018.12.012
145. Jangra M, Dhull SK, Singh KK. ECG arrhythmia classification using modified visual geometry group network (mVGGNet). *J Intell Fuzzy Syst.* (2020) 38:3151–65. doi: 10.3233/JIFS-191135
146. Arun Kumar M, Chakrapani A. Classification of ECG signal using FFT based improved AlexNet classifier. *PLoS One.* (2022) 17:e0274225. doi: 10.1371/journal.pone.0274225
147. Degirmenci M, Ozdemir MA, Izci E, Akan A. Arrhythmic heartbeat classification using 2D convolutional neural networks. *IRBM.* (2022) 43:422–33. doi: 10.1016/j.irbm.2021.04.002
148. Zhang Y, Liu S, He Z, Zhang Y, Wang C. A CNN model for cardiac arrhythmias classification based on individual ECG signals. *Cardiovasc Eng Technol.* (2022) 13:1–10. doi: 10.1007/s13239-021-00599-8
149. Korürek M, Doğan B. ECG beat classification using particle swarm optimization and radial basis function neural network. *Expert Syst Appl.* (2010) 37:7563–9. doi: 10.1016/j.eswa.2010.04.087
150. Mar T, Zaunseder S, Martínez JP, Llamado M, Poll R. Optimization of ECG classification by means of feature selection. *IEEE Trans Biomed Eng.* (2011) 58:2168–77. doi: 10.1109/TBME.2011.2113395
151. Zidelmal Z, Amirou A, Ould-Abdeslam D, Merckle J. ECG beat classification using a cost sensitive classifier. *Comput Methods Prog Biomed.* (2013) 111:570–7. doi: 10.1016/j.cmpb.2013.05.011
152. Escalona-Morán MA, Soriano MC, Fischer I, Mirasso CR. Electrocardiogram classification using reservoir computing with logistic regression. *IEEE J Biomed Health Inform.* (2014) 19:892–8. doi: 10.1109/JBHI.2014.2332001
153. Liu N, Sun M, Wang L, Zhou W, Dang H, Zhou X. A support vector machine approach for af classification from a short single-lead ECG recording. *Physiol Meas.* (2018) 39:064004. doi: 10.1088/1361-6579/aa7ca
154. Yang H, Wei Z. A novel approach for heart ventricular and atrial abnormalities detection via an ensemble classification algorithm based on ECG morphological features. *IEEE Access.* (2021) 9:54757–74. doi: 10.1109/ACCESS.2021.3071273
155. Zou C, Müller A, Wolfgang U, Rückert D, Müller P, Becker M, et al. Heartbeat classification by random forest with a novel context feature: a segment label. *IEEE J Transl Eng Health Med.* (2022) 10:1–8. doi: 10.1109/JTEHM.2022.3202749
156. Pham VT, Pham VS, Nguyen MT, Le HC. Efficient electrocardiogram-based arrhythmia detection utilizing R-peaks and machine learning. In: 2023

International Conference on System Science and Engineering (ICSSE). IEEE (2023). p. 604–8.

157. Mishra A, Bhusnur S, Mishra SK, Singh P. Exploring a new frontier in cardiac diagnosis: ECG analysis enhanced by machine learning and parametric quartic spline modeling. *J Electrocardiol.* (2024) 85:19–24. doi: 10.1016/j.jelectrocard.2024.05.086

158. Nasim A, Nchekwube DC, Munir F, Kim YS. An evolutionary-neural mechanism for arrhythmia classification with optimum features using single-lead electrocardiogram. *IEEE Access.* (2022) 10:99050–65. doi: 10.1109/ACCESS.2022.3203586

159. Alarsan FI, Younes M. Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. *J Big Data.* (2019) 6:1–15. doi: 10.1186/s40537-019-0244-x

160. Padhy S, Dandapat S. Third-order tensor based analysis of multilead ECG for classification of myocardial infarction. *Biomed Signal Process Control.* (2017) 31:71–8. doi: 10.1016/j.bspc.2016.07.007

161. Sun L, Lu Y, Yang K, Li S. ECG analysis using multiple instance learning for myocardial infarction detection. *IEEE Trans Biomed Eng.* (2012) 59:3348–56. doi: 10.1109/TBME.2012.2213597

162. Ghahjaverestan NM, Masoudi S, Shamsollahi MB, Beuchee A, Pladys P, Ge D, et al. Coupled hidden Markov model-based method for apnea bradycardia detection. *IEEE J Biomed Health Inform.* (2015) 20:527–38. doi: 10.1109/JBHI.2015.2405075

163. Strodtthoff N, Wagner P, Schaeffter T, Samek W. Deep learning for ECG analysis: benchmarks and insights from PTB-XL. *IEEE J Biomed Health Inform.* (2020) 25:1519–28. doi: 10.1109/JBHI.2020.3022989

164. Baptista FD, Rodrigues S, Morgado-Dias F. Performance comparison of ann training algorithms for classification. In: *2013 IEEE 8th International Symposium on Intelligent Signal Processing.* IEEE (2013). p. 115–20.

165. Tharwat A, Gaber T, Ibrahim A, Hassanien AE. Linear discriminant analysis: a detailed tutorial. *AI Commun.* (2017) 30:169–90. doi: 10.3233/AIC-170729

166. Hou B, Yang J, Wang P, Yan R. LSTM-based auto-encoder model for ECG arrhythmias classification. *IEEE Trans Instrum Meas.* (2019) 69:1232–40. doi: 10.1109/TIM.2019.2910342

167. Zhou F, Sun Y, Wang Y. Inter-patient ECG arrhythmia heartbeat classification network based on multiscale convolution and FCBA. *Biomed Signal Process Control.* (2024) 90:105789. doi: 10.1016/j.bspc.2023.105789

168. Liu LR, Huang MY, Huang ST, Kung LC, Lee C, Yao WT, et al. An arrhythmia classification approach via deep learning using single-lead ECG without QRS wave detection. *Heliyon.* (2024) 10:e27200. doi: 10.1016/j.heliyon.2024.e27200

169. Mantravadi A, Saini S, Mittal S, Shah S, Devi S, Singhal R, et al. Clinet: a novel deep learning network for ECG signal classification. *J Electrocardiol.* (2024) 83:41–8. doi: 10.1016/j.jelectrocard.2024.01.004

170. Al Rahhal MM, Bazi Y, AlHichri H, Alajlan N, Melgani F, Yager RR. Deep learning approach for active classification of electrocardiogram signals. *Inf Sci (Nj).* (2016) 345:340–54. doi: 10.1016/j.ins.2016.01.082

171. Wang Z, Wang K, Chen X, Zheng Y, Wu X. A deep learning approach for inter-patient classification of premature ventricular contraction from electrocardiogram. *Biomed Signal Process Control.* (2024) 94:106265. doi: 10.1016/j.bspc.2024.106265

172. Rashed-Al-Mahfuz M, Moni MA, Lio' P, Islam SMS, Berkovsky S, Khushi M, et al. Deep convolutional neural networks based ECG beats classification to diagnose cardiovascular conditions. *Biomed Eng Lett.* (2021) 11:147–62. doi: 10.1007/s13534-021-00185-w

173. Patro KK, Prakash AJ, Samantray S, Pławiak J, Tadeusiewicz R, Pławiak P. A hybrid approach of a deep learning technique for real-time ECG beat detection. *Int J Appl Math Comput Sci.* (2022) 32:455–65. doi: 10.34768/amcs-2022-0033

174. Wang Y, Zhou G, Yang C. Interpatient heartbeat classification using modified residual attention network with two-phase training and assistant decision. *IEEE Trans Instrum Meas.* (2022) 72:1–15. doi: 10.1109/TIM.2022.3232646

175. Samaratunga PS, Silva T, Karunananda A. ECG beat classification using capsule networks. In: *2023 7th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI).* (2023). p. 1–6.

176. Lai C, Zhou S, Trayanova N. Abstract 15042: heart rhythm classification from an optimal lead subset of the 12-lead electrocardiogram by deep learning. *Circulation.* (2020) 142. doi: 10.1161/circ.142.suppl\_3.15042

177. Dokur Z, Ölmez T. ECG beat classification by a novel hybrid neural network. *Comput Methods Programs Biomed.* (2001) 66:167–81. doi: 10.1016/S0169-2607(00)00133-4

178. Dong X, Wang C, Si W. ECG beat classification via deterministic learning. *Neurocomputing.* (2017) 240:1–12. doi: 10.1016/j.neucom.2017.02.056

179. Yu SN, Chou KT. Selection of significant independent components for ECG beat classification. *Expert Syst Appl.* (2009) 36:2088–96. doi: 10.1016/j.eswa.2007.12.016

180. Yang Q, Liu Y, Chen T, Tong Y. Federated machine learning: concept and applications. *ACM Trans Intell Syst Technol.* (2019) 10:12. doi: 10.1145/3298981

181. Samek W, Wiegand T, Müller KR. Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. *arXiv [Preprint]. arXiv:1708.08296* (2017).

182. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications. *arXiv [Preprint]. arXiv:1704.04861* (2017).

183. Xiao Q, Lee K, Mokhtar SA, Ismail I, Pauzi ALM, Zhang Q, et al. Deep learning-based ECG arrhythmia classification: a systematic review. *Appl Sci.* (2023) 13:4964. doi: 10.3390/app13084964

184. Um TT, Pfister FMJ, Pichler D, Endo S, Lang M, Hirche S, et al. Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks. In: *Proceedings of the 19th ACM International Conference on Multimodal Interaction.* (2017). p. 216–20.

185. Gajendran MK, Khan MZ, Khattak MAK. ECG classification using deep transfer learning. In: *2021 4th International Conference on Information and Computer Technologies (ICICT).* IEEE (2021). p. 1–5.

186. Clifford GD, Liu C, Moody B, Lehman LH, Silva I, Li Q, et al. AF classification from a short single lead ECG recording: the physionet/computing in cardiology challenge 2017. In: *2017 Computing in Cardiology Conference (CinC).* IEEE (2017). p. 1–4.

187. Shadmand S, Mashoufi B. A new personalized ECG signal classification algorithm using block-based neural network and particle swarm optimization. *Biomed Signal Process Control.* (2016) 25:12–23. doi: 10.1016/j.bspc.2015.10.008

188. Mir HY, Singh O. ECG denoising and feature extraction techniques—a review. *J Med Eng Technol.* (2021) 45:672–84. doi: 10.1080/03091902.2021.1955032

189. Islam M, Alajlan N, Malek S. Resampling of ECG signal for improved morphology alignment. *Electron Lett.* (2012) 48:427–9. doi: 10.1049/el.2012.0421

190. Zhu XJ, Goldberg AB. *Introduction to Semi-Supervised Learning.* Princeton, NJ: Morgan & Claypool Publishers (2009), Vol. 3.

191. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, physiotookit, and physionet. *Circulation.* (2000) 101:e215–20. doi: 10.1161/01.CIR.101.23.e215

192. Delaney AM, Brophy E, Ward TE. Synthesis of realistic ECG using generative adversarial networks. *arXiv [Preprint]. arXiv:1909.09150* (2019).

193. Shaker AM, Tantawi M, Shedeed HA, Tolba MF. Generalization of convolutional neural networks for ECG classification using generative adversarial networks. *IEEE Access.* (2020) 8: 35592–605. doi: 10.1109/ACCESS.2020.2974712

194. Al Rahhal MM, Bazi Y, AlHichri H, Alajlan N, Melgani F, Yager RR. Deep learning approach for active classification of electrocardiogram signals. *Inf Sci (Nj).* (2016) 345:340–54. doi: 10.1016/j.ins.2016.01.082

195. Li D, Zhang J, Zhang Q, Wei X. Classification of ECG signals based on 1D convolution neural network. In: *2017 IEEE 19th International Conference on E-health Networking, Applications and Services (Healthcom).* IEEE (2017). p. 1–6.

196. Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection. In: *Proceedings of the IEEE International Conference on Computer Vision.* (2017). p. 2980–8.

197. Li X, Gu Y, Dvornek NC, Staib LH, Ventola P, Duncan JS. FedHealth: a federated transfer learning framework for wearable healthcare. *IEEE Intell Syst.* (2020) 35:83–93. doi: 10.1109/MIS.2020.2972533

198. Weimann K, Conrad TO. Transfer learning for ECG classification. *Sci Rep.* (2021) 11:5251. doi: 10.1038/s41598-021-84374-8

199. Goodfellow I, Bengio Y, Courville A. *Deep Learning.* Cambridge, MA: MIT Press (2016).

200. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Attentive recurrent neural networks for medical concept extraction. *arXiv [Preprint]. arXiv:1608.05745* (2016).

201. Han S, Mao H, Dally WJ. Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding. *arXiv [Preprint]. arXiv:1510.00149* (2016).

202. Kingma DP, Welling M. Auto-encoding variational bayes. In: *2nd International Conference on Learning Representations (ICLR).* (2014).

203. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. *arXiv [Preprint]. arXiv:2002.05709* (2020).

204. Loh BC, Then PH. Deep learning for cardiac computer-aided diagnosis: benefits, issues & solutions. *Mhealth.* (2017) 3:45. doi: 10.21037/mhealth.2017.09.01

205. De D, Bharti P, Das SK, Chellappan S. Multimodal wearable sensing for fine-grained activity recognition in healthcare. *IEEE Internet Comput.* (2015) 19:26–35. doi: 10.1109/MIC.2015.72

## Glossary

|          |                                                                   |         |                                                                          |
|----------|-------------------------------------------------------------------|---------|--------------------------------------------------------------------------|
| ACC      | Accuracy                                                          | IoMT    | Internet of Medical Things                                               |
| AE       | Autoencoder                                                       | IoT     | Internet of Things                                                       |
| AF       | Atrial fibrillation                                               | JCR     | Journal Citation Reports                                                 |
| AFDB     | Atrial Fibrillation Database                                      | kNN     | K-nearest neighbor                                                       |
| AHA      | American Heart Association                                        | LMS     | Least mean squares                                                       |
| AI       | Artificial intelligence                                           | LOSO    | Leave-one-subject-out validation                                         |
| AM       | Autoregressive model                                              | LPF     | Low-pass filter                                                          |
| ANOVA    | Analysis of variance                                              | LR      | Learning rate                                                            |
| API      | Application programming interface                                 | LSTM    | Long short-term memory                                                   |
| AUC      | Area under the curve                                              | MA      | Muscle artifacts                                                         |
| AUPRC    | Area under the Precision–Recall curve                             | MAR     | Motion artifact                                                          |
| AV       | Atrioventricular                                                  | MAE     | Mean absolute error                                                      |
| BPF      | Band-pass filter                                                  | MFCC    | Mel-frequency cepstral coefficients                                      |
| BN       | Batch normalization                                               | MIT-BIH | Massachusetts Institute of Technology–Beth Israel Hospital Database      |
| BP       | Blood pressure                                                    | ML      | Machine learning                                                         |
| BSS      | Blind source separation                                           | MLP-BP  | Multilayer perceptron-backpropagation                                    |
| BW       | Baseline wander                                                   | MSE     | Mean squared error                                                       |
| BWF      | Baseline wander filter                                            | NSTEMI  | Non-ST-segment elevation myocardial infarction                           |
| CINC     | Computers in Cardiology Challenge                                 | NPV     | Negative predictive value                                                |
| CI       | Confidence interval                                               | PAC     | Premature atrial contraction                                             |
| CNN      | Convolutional neural network                                      | PCA     | Principal component analysis                                             |
| CNN-LSTM | Convolutional neural network combined with long short-term memory | PLI     | Powerline interference                                                   |
| CPU      | Central processing unit                                           | PPG     | Photoplethysmography                                                     |
| CPSC     | China Physiological Signal Challenge                              | PPV     | Positive predictive value                                                |
| CSV      | Comma-separated values                                            | PRISMA  | Preferred Reporting Items for Systematic Reviews and Meta-Analyses       |
| CV       | Cross-validation                                                  | PTB     | Physikalisch-Technische Bundesanstalt Diagnostic ECG Database            |
| CVDs     | Cardiovascular diseases                                           | PVC     | Premature ventricular contraction                                        |
| CWT      | Continuous wavelet transform                                      | PWA     | Pulse wave analysis                                                      |
| DBA      | Differential beat accuracy                                        | Q       | Quartile (Journal Citation Reports Ranking, Q1–Q4)                       |
| DBN      | Deep belief network                                               | QOL     | Quality of life                                                          |
| DNN      | Deep neural network                                               | QRS     | Q-wave, R-wave, and S-wave complex                                       |
| DL       | Deep learning                                                     | RMS     | Root mean square                                                         |
| DOI      | Digital object identifier                                         | RMSE    | Root mean square error                                                   |
| DTCWT    | Dual-tree complex wavelet transform                               | RMSSD   | Root mean square of successive differences                               |
| DWT      | Discrete wavelet transform                                        | RNN     | Recurrent neural network                                                 |
| ECG      | Electrocardiogram                                                 | ROC     | Receiver operating characteristic                                        |
| ECN      | Electrode contact noise                                           | RR      | R–R interval (interval between consecutive R-peaks)                      |
| EEG      | Electroencephalogram                                              | SA      | Sinoatrial (node)                                                        |
| EMD      | Empirical mode decomposition                                      | SAE     | Stacked autoencoder                                                      |
| EMG      | Electromyogram                                                    | SB      | Sinus bradycardia                                                        |
| EWT      | Empirical wavelet transform                                       | SD      | Standard deviation                                                       |
| FCN      | Fully convolutional network                                       | SDG     | Sustainable development goal                                             |
| FFT      | Fast Fourier transform                                            | SDNN    | Standard deviation of normal-to-normal intervals                         |
| FN       | False negative                                                    | SJR     | Scimago Journal Rank                                                     |
| FP       | False positive                                                    | SJR/Q   | Scimago Journal Rank / Journal Citation Reports Quartile (Quality Index) |
| FPGA     | Field programmable gate array                                     | SMOTE   | Synthetic minority oversampling technique                                |
| GAN      | Generative adversarial network                                    | SNR     | Signal-to-noise ratio                                                    |
| GRU      | Gated recurrent unit                                              | SOM     | Self-organizing map                                                      |
| GPU      | Graphics processing unit                                          | SoC     | System on chip                                                           |
| HMM      | Hidden Markov model                                               | Softmax | Softmax activation function                                              |
| HPF      | High-pass filter                                                  | SPE     | Specificity                                                              |
| HR       | Heart rate                                                        | ST      | Stockwell transform                                                      |
| HRV      | Heart rate variability                                            | STEMI   | ST-segment elevation myocardial infarction                               |
| ICA      | Independent component analysis                                    | STFT    | Short-time Fourier transform                                             |
| IMF      | Intrinsic mode function                                           | SVEB    | Supraventricular ectopic beat                                            |
| INCART   | St. Petersburg Institute of Cardiological Technics Database       | SVT     | Supraventricular tachycardia                                             |

|      |                                    |     |                          |
|------|------------------------------------|-----|--------------------------|
| SVM  | Support vector machine             | VAE | Variational autoencoder  |
| SWT  | Stationary wavelet transform       | VEB | Ventricular ectopic beat |
| TL   | Transfer learning                  | VF  | Ventricular fibrillation |
| TCN  | Temporal convolutional network     | VT  | Ventricular tachycardia  |
| TN   | True negative                      | WT  | Wavelet transform        |
| TP   | True positive                      | ZCR | Zero-crossing rate       |
| TQWT | Tunable Q-factor wavelet transform |     |                          |