Computers in Biology and Medicine 196 (2025) 110594

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A comprehensive review of AI-Based detection of Arrhythmia using Electrocardiogram (ECG)

Ahtisham Ayyub *, Christos Politis, Muhammad Arslan Usman

Faculty of Engineering, Computing and the Environment (ECE), Kingston University London, KT1 2EE, United Kingdom

ARTICLE INFO

Keywords:
ECG
Arrhythmia detection
AI-based decision making

ABSTRACT

The success of AI-assisted decision-making systems over traditional methods has driven extensive research across various real-world applications. In the past decade, the application of AI systems for analysing physiological signals, particularly electrocardiograms (ECG), has attracted considerable attention. While several survey papers have explored this domain, they often face limitations, for instance outdated research coverage, narrow scope, inadequate evaluation of study quality and publication credibility or a lack of statistical insights. To address these gaps, this review rigorously selected research articles from high-impact journals and top-tier conferences, ensuring reliable and validated findings. We comprehensively reviewed 219 research articles, making this paper a valuable resource for researchers interested in the intersection of AI and ECG analysis. Our review provides an in-depth analysis of employed techniques, obtained results, and emerging trends, offering insights beneficial to researchers at all levels. Additionally, we present a statistical analysis of the reviewed studies to offer a broader understanding of this research area. A key contribution of this paper is the application of Pearson's correlation to examine relationships among performance metrics such as accuracy, sensitivity, specificity, and F1-score. This analysis highlights how these metrics interact and influence each other across various methodologies, offering deeper insights into model performance and optimisation strategies in ECG analysis. Finally, we address existing challenges and propose new research directions for further exploration.

1. Introduction

Arrhythmia is a term which has been derived from Greek roots meaning without rhythm. It refers to irregularities in the heart's rhythm, characterized by abnormal electrical activity. Everyone has a heartbeat rate, this heartbeat rate can vary slightly from person to person depending on multiple factors including age, sex, and other physical characteristics. A deviation of the heart rate from the normal range, is called arrhythmia [1]. Arrhythmia can disturb the coordinated contraction of the heart muscle, which may result in compromising its ability to pump blood effectively throughout the body. This disruption can lead to various symptoms, including palpitations, dizziness, fainting, or, in severe cases, cardiac arrest. Timely detection and accurate diagnosis of arrhythmia are critical for the patient's well-being. If left untreated, arrhythmia can significantly increase the risk of stroke, heart failure, and sudden cardiac death. One reasons why biomedical studies have focused extensilvely on arrhythmia is that it is among the most common reasons for deaths [2].

Artificial intelligence (AI) is a term used to describe technologies and systems capable of replicating human-like intelligence. AI is a constantly evolving field [3], one of its key subsets is Machine Learning (ML), which enables systems to learn from data, enhance their capabilities through experience, recognise patterns within data, and make predictions. Neural Networks are a relatively new concept in ML. As the name suggests, these are inspired by the human brain's complex structure. The human brain is a fascinating system with processing capabilities beyond our imaginations. Its remarkable processing power has attracted researchers to simulate real world problems using Neural Networks. Within ML, there is an even more specialised subset known as Deep Learning (DL). DL is characterised by highly complex architectures that differ from traditional ML techniques in that they do not require manual feature engineering, furthermore, there is no need for extensive data preprocessing. These systems are capable of building end-to-end models, taking raw data as input, processing it thorugh multiple hidden layers and producing outputs. The greater the number of hidden layers, the deeper a network is considered [4]. AI has been applied in number of areas including pattern recognition [5], medical image classification [6], iris-based human recognition [7] and gait-based person recognition [8–10]. Within the healthcare sector, substantial research is dedicated to exploring the use of AI-based algorithms and their potential clinical benefits. These AI-powered systems

* Corresponding author.
E-mail addresses: a.ayyub@kingston.ac.uk (A. Ayyub), c.politis@kingston.ac.uk (C. Politis), m.usman@kingston.ac.uk (M.A. Usman).

https://doi.org/10.1016/j.compbiomed.2025.110594
Received 27 September 2024; Received in revised form 14 April 2025; Accepted 16 June 2025
Available online 8 July 2025

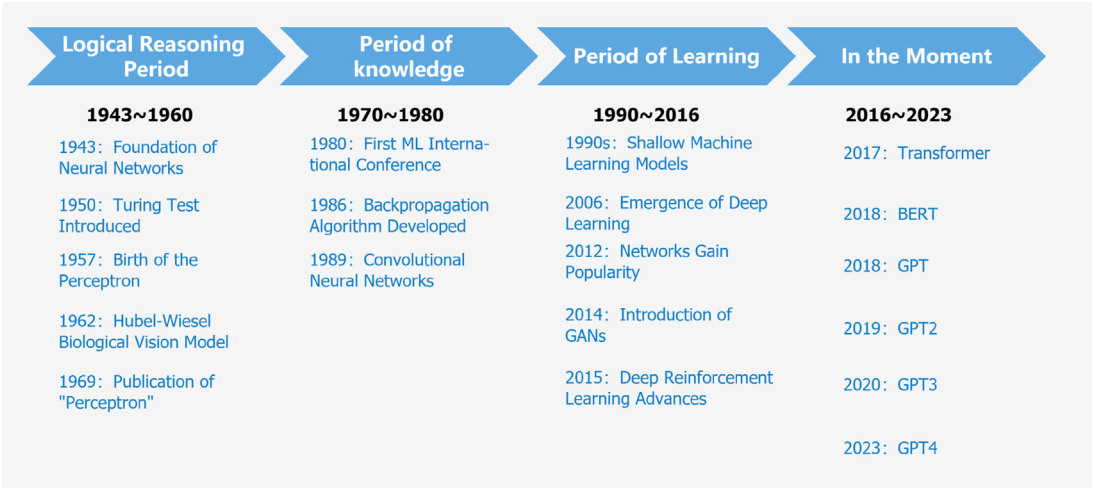


Fig. 1. Evolution of AI [3].

have the capacity to revolutionise healthcare practices and improve patient outcomes through advanced data analysis and decision-making processes. Fig. 1 presents an overview of the evolution of AI.

An electrocardiogram (ECG) is a non-invasive medical test that records the electrical activity of the heart over a period of time [11]. It is commonly used to assess the heart's rhythm and electrical conduction, providing valuable information about cardiac health and function. The ECG has been used for over a hundred years to assess heart rhythm and function, having first been recorded by Augustus Desire Waller [12].

During an ECG, small electrodes are placed on the skin of the chest, arms, and legs, which detect the electrical signals generated by the heart as it beats. These signals are then amplified and recorded as a series of waves on graph paper or displayed digitally on a monitor.

ECGs are widely used in clinical practice, emergency settings, and routine health screenings due to their effectiveness, and ability to provide valuable insights into cardiac function with minimal discomfort to the patient. The physicians identify arrhythmia by interpreting ECG taken from the patient. But due to the complex nature of ECG graphs along with a lot of noise; this manual interpretation can be time-consuming and prone to human error, which invites AI to come in and play its role. As mentioned above, AI can introduce better accuracy, scalability and time efficiency while minimising human intervention in ECG interpretation.

The integration of AI into ECG has significantly advanced the detection and classification of cardiac arrhythmias. Traditional ECG interpretation, while effective, is often limited by human variability and the complexity of arrhythmic patterns. AI, through ML and DL techniques, offers enhanced accuracy and efficiency in analysing ECG data, leading to improved diagnostic outcomes [13].

1.1. Challenges in arrhythmia detection using AI-based methods

Although AI has demonstrated significant promise in ECG analysis, several key challenges continue to limit its broader application. A major concern is the lack of standardisation in ECG data formats and preprocessing methods. Currently, there is no universally accepted ECG input type or preprocessing protocol. While many studies use 10-second, 12-lead ECG recordings taken in the supine position, others rely on segmented Holter monitor data or non-standard sources such as single-lead wearable devices and patches. Preprocessing techniques also differ widely; some models use raw ECG image matrices [14–16], whereas others apply signal processing methods like band-pass filtering, discrete wavelet transforms [13], short-time Fourier transforms [17], or other approaches [18]. These inconsistencies in data input and signal processing can significantly affect model performance.

Despite these technical advancements, ECG data standardisation remains under emphasised. Few studies clearly specify the exact form of input data; whether it is images, raw signal matrices, or other representations; or provide detailed descriptions of preprocessing steps used. As a result, it is difficult to compare model performance across studies, making it unclear which data format or preprocessing technique is most effective for leveraging the full potential of AI.

Furthermore, one inherent limitation of both ML and DL-based models is their lack of interpretability. In many cases, the decision-making process behind model predictions is not transparent, making it challenging to understand how and why certain classifications are made [19]. This black-box nature of predictive models poses a barrier to clinical adoption, where explainability is crucial for trust and ensuring validation.

1.2. Contributions of the study

Our goal is to perform a systematic literature review on the application of AI-based decision making systems for the detection of arrhythmia.

Major contributions of this paper can be summarised as follows.

1. We have presented detailed statistics to assist researchers, derived from an in-depth analysis of the selected articles.
2. We have examined more than 200 research papers in this study. Tto the best of our knowledge, there in no recent study which specifically targets application of AI-based decision making systems in detection of arrhythmia, in such a detailed manner.
3. Only studies published in Q1\Q2 journals or top-tier conferences have been selected to ensure the integrity and credibility of the results
4. The inclusion of the latest publications from 2023 and 2024 enables researchers to identify current trends in this research domain
5. We discuss detailed results obtained through different methods, which can significantly help researchers choose appropriate techniques for their own studies.
6. We apply Pearson's correlation analysis to various evaluation metrics to explore relationships among them in two ways: first, based on individual methods; and second, across all methods combined
7. We present a graphical bibliographic analysis of the existing research in this area.
8. We include a forest plot to visually compare the performance metrics of different methods, enhancing interpretability and trend analysis.

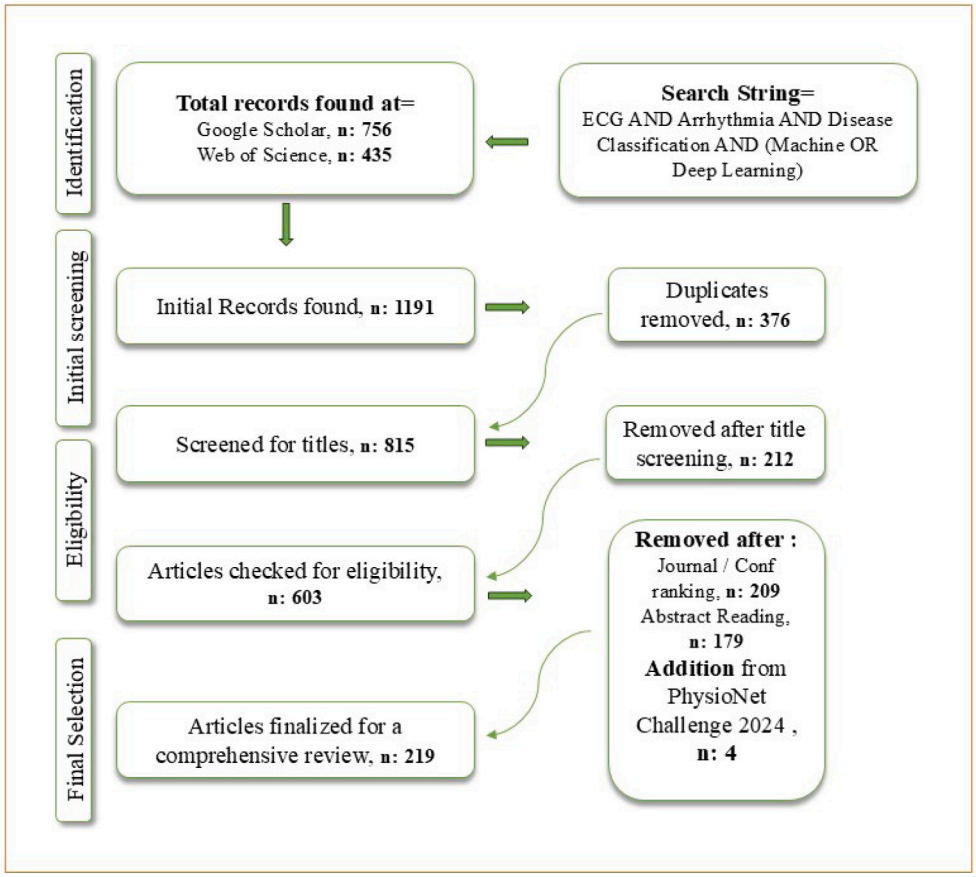


Fig. 2. The step by step procedure of selection of the studies.

Rest of the paper is organised as follows. The "Methodology" is described in the next section, followed by an over view of related literature in the "Literature Review". Subsequently, we present detailed statistics in the "Presentation of Findings" section. We then discuss the limitations of existing studies along with research directions for future work in "Limitations and Research Directions".Finally, the paper concludes with the "Conclusion" section.

2. Methodology

In this section, we describe the step-by-step process used to conduct the systematic literature review. This process follows the framework defined by [20].

- First the research questions were formulated.
- Various electronic repositories were searched to identify relevant papers using carefully defined search strings and keywords.
- Clear inclusion and exclusion criteria were established, according to which papers were selected or rejected.
- The results were then presented in both textual and graphical form.

2.1. Research questions

Keeping in mind the research objectives, following research questions were formulated.

1. How many studies were conducted each year within this research domain?
2. Which types of ML\DL architectures have been adopted by researchers in this area?
3. How many studies focused specifically on ML\DL architectures rather than traditional methods?
4. Which datasets were most commonly used by researchers?
5. Which methods produced the best results?
6. What potential research gaps can be identified from the analysis of contemporary researches?

2.2. Selection of articles

In this section we present a step-by-step overview of the selection procedure of articles for this review in the following, also presented graphically in Fig. 2.

2.2.1. Databases used

Following e-libraries were used to extract the articles.

- Google scholar (https://scholar.google.com/).
- Web of science (https://www.webofscience.com/wos/woscc/basic-search).

2.2.2. Search strings

Keywords for searching relevant articles are as follows.

- "ECG" AND "Arrhythmia" AND "Disease Classification" AND ("Machine" OR "Deep Learning")
- "ECG" AND "Arrhythmia" AND "Disease Classification" AND ("CNN" OR "LSTM" OR "RNN" OR "Deep Neural Networks" OR "Transformers" OR "MLP")

2.2.3. Inclusion criteria

Following criteria were used while including papers in our study.

- Papers which were in English language.
- Papers which were published during period of 2013 to 2024.
- Papers which had relevant titles.
- Papers which had abstract, which related to our research questions.
- Papers which were published in journals, which are either Q1 or Q2 only.
- Conference papers which were either published in a category A or B conference. Or had more than 50 citations.

2.2.4. Exclusion criteria

Papers were not included in the study, if they fulfilled any criteria given below:

- Paper was published in a journal which was in lower quartile than Q2 i.e. Q3, Q4 or in a lower ranked conference.
- Titles of the papers did not match the keywords.
- Paper focused on other diseases than Arrhythmia.
- Where language was not English.

As shown in Fig. 2, the initial query returned a total of 1191 papers. Of these, 376 duplicates were identified as duplicates and subsequently removed, leaving 815 unique papers. The next stage involved filtering the papers based on their titles. Papers with titles irrelevant to the focus of this review, such as those addressing diseases other than arrhythmia, were excluded, resulting in 603 papers.

In the following step, we assessed the quality of the journals and conferences where these papers were published. We set a quality threshold, excluding papers from journals ranked lower than Q2 and conferences not recognised with a strong ranking, ensuring the integrity and high standard of our review. This step was critical in maintaining the relevance and scholarly value of the included studies. Where a final decision could not be reached, the number of citations was considered; papers with fewer than 50 citations were excluded.

Finally, the abstracts of the remaining papers were carefully examined. Any papers with abstracts that did not align with the scope of our review or indicated that the research was not directly related to our area of focus were removed. Additionally, four recent studies from Physio Net Challenge 2024 were included to incoporate contemporary research. This rigorous selection process resulted in 219 articles, each meeting the criteria for inclusion in this comprehensive review.

2.3. Present findings

The final step of our research was presentation of findings, for which we employed various graphical methods.

- Graphs illustrating the bibliographic coupling of journals in which articles related to our research domain were published.
- Graphs showing the co-citation analysis of articles within our domain.
- Pie charts displaying the proportion of conference and journal publications.
- Bar charts indicating the frequency of usage of each dataset across different studies.
- A detailed review of the papers included in the study.
- Multiple correlation matrices generated using Pearson's correlation coefficient, showing the relationships among different performance metrics.
- A forest plot for analysing evaluation metric scores across different models.

3. Literature review

In this section we are going to discuss results for arrhythmia detection using different ML\DL architectures.

3.1. Machine learning models

Support Vector Machine (SVM) — is a popular machine learning approach used for linear and non-linear classification. The idea was initially presented by [21]. It works on the principal of construction of a decision boundary between classes and later classifying accordingly. SVM has been applied in diverse real life applications; for example internet traffic monitoring [22], pattern recognition [5], medical image classification [6] and iris based recognition of human subjects [7].

K Nearest neighbour (KNN) — is a classical machine learning algorithm which is very efficient yet very simple to implement. Its also called non-parametric classifier, as it does not have to tune any parameters [23]. It has been implemented for web usage and recommended systems [24], for forestry [25], student performance prediction [26] health monitoring [27] and finance-based time series forecasting [28].

Random Forest (RF) — classifiers are ensemble learning models [29]. That have gained popularity in diverse fields. They are renowned for their accuracy and rely on the creation of multiple decision trees, each trained on a randomly selected subset of training samples and features. This randomness reduces overfitting and boosts the model's capacity for generalised predictions. RF employs bagging, generating multiple decision trees from bootstrapped training data, and the final prediction is a result of averaging (for regression) or voting (for classification) across all the trees. RF is well-regarded for its ability to handle noisy data and high-dimensional feature spaces, and its capacity to deliver superior classification performance when compared to individual decision trees and other machine learning algorithms [rf base paper]. Random forests have been used in the fields of ecology [30], classification of Unmanned Arial Vehicle (UAV) data [31], map irrigated areas [32] and credit risk assessment [33].

Multi Layer Perceptron (MLP) — is one of the earliest forms of an artificial neural network. Here input are fed to the input layer, processed, and are passed to output layer. Also referred as a feed forward network [34]. It's simple yet effective structure has enabled it to be used in diverse fields such as pattern classification [35], speech emotion recognition [36], recognition of driving postures [37] and seizure prediction [38]. A single layer perceptron can be computed through following equation.

$$h(x_i) = \begin{cases} +1 & w \cdot x + b \geq 0 \\ -1 & w \cdot x + b < 0 \end{cases} \tag{1}$$

Ensemble — is a technique which works on the creation of a predictive model that combines the strengths of multiple individual models. The idea behind the ensemble methodology is to assign weight to multiple individual classifiers and integrate them to create a classifier that performs better than each of them individually [39]. Ensembles have been extensively used in screening of diabetic retinoscopy [40], predicting the number of software faults [41] and prediction of bankruptcy [42].

3.2. Deep learning models

Recurrent Neural Network (RNN) — can keep track of recently used inputs. Which makes them suitable for the time-based data. That is why it is used in many areas successfully; for example speech recognition [43,44], machine translation [45], machine comprehension [46], sentence summarizing [47], word representation [48,49], bio informatics [50] and gait recognition. RNN have a flaw i.e. when time lap between a given input and relevant output is too large, its efficiency decreases gradually. RNN computes output from following equation:

$$s_t = \tanh(U.x_t + W.s_{t-1}) \tag{2}$$

$$\hat{y}_t = softmax(V.s_t) \tag{3}$$

Long Short Term Memory (LSTM) — LSTM architectures have been one of most popular techniques used in ECG related-research, as it

can effectively deal with time-based data. LSTM was presented by [51] in 1997. It has rapidly become popular among researchers due to its ability to handle and memorize temporal information [52]. It works in the form of gates. Each gate has its own functionality. First gate is forget gate. It helps LSTM in deciding whether to keep\discard a certain input.

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{4}$$

Next is input gate. Here values are normalised in range of 1 and −1 to decide whether to pass\stop a certain input.

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{5}$$

$$\tilde{C}_t = \tanh(W_C.[h_{t-1}, x_t] + b_C) \tag{6}$$

Then comes the cell state and last one is output gate, which provides final output of the architecture.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C} \tag{7}$$

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{8}$$

$$h_t = o_t * \tanh(C_t) \tag{9}$$

Convolutional Neural Network (CNN) — is a well known variant of neural networks. It has gained significant popularity in research community rapidly in recent times; due to certain characteristics. [53] stated that CNN has reduced the importance of hand crafted feature techniques, as it has the capability to do the same automatically. CNN has been successfully used in several applications; for example face detection [54], object detection [55], human pose estimation [56], image classification [57], traffic signal classification [58], natural language processing [59], fire detection in video [60] and visual recognition [61].

Generative adversarial network (GAN) — is a technique used to overcome the issue of lack of training data. It is a data augmentation approach. It is an approach in which two models are trained at the same time; the generative model tires to produce fake data according to the probability distribution of the given data. On the other hand, the discriminative model tries to detect \identify the fake data [62]. It has been used to solve various problems; for instance description-based image generation [63], pattern-based image retrieval [64], converting low into high resolution images [65] and image to image translation [66].

Autoencoder — serves the purpose of acquiring data encodings through unsupervised learning. The primary objective of an Autoencoder is to compute a lower-dimensional representation (referred to as an encoding) for data that initially resides in a higher-dimensional space. Typically employed for dimensionality reduction, the network is trained to extract the most critical features of the input image. Autoencoder were introduced by Rumelhart and his associates during the decade of 1980 [67]. Autoencoder have been used in financial analysis [68], cyber security applications [69] and medical image reconstruction [70].

Attention-Based Methods- The attention mechanism is merely a mechanism introduced to improve the encoder–decoder model, by introducing an additional layer which helps the model to focus on the most relevant parts of input. It was introduced by [71]. It has been used for speech recognition [72], sentiment analysis [73], medical image segmentation [74], video description [75] and facial expression recognition [76].

Transfer learning — is a method where pre-trained models are used on unseen datasets. It increases time efficiency as time for training is eliminated. The idea was presented initially by [77]. It has been applied in the fields of drug discovery [78], smart buildings [79], visual categorization [80] and medical imaging [81].

Abbreviations used in Table 1:

- acc = Accuracy
- spec = Specificity
- sen = Sensitivity
- f1 = F1-Score
- OAC = Overall Accuracy
- prec = Precision
- Pp = Positive Predicted Value
- Npv = Negative Predicted Value
- avg = average
- ROC = Receiver-operating characteristic curve
- Kappa = Cohen's Kappa
- AC1 = First-order Agreement Coefficient
- AUC = Area Under Curve
- Classes = Each class represents a normal\disease instance of data.

Table 1 provides a comprehensive overview of the papers included in this study, serving as a key resource for researchers interested in this field. The table contains critical information, making it a valuable information source for those looking to explore or build upon existing work. Each paper is cited in the first column to ensure easy and direct access to the original articles, streamlining the literature review process. Subsequent columns capture essential details such as the datasets used, the number of classes targeted, feature preprocessing and extraction techniques employed, the results obtained, and key findings from each study. By compiling this information in a single table, researchers can quickly identify relevant studies, compare methodologies, and evaluate the performance of different approaches. The rows are organised according to the methodologies used, ensuring a clear and logical flow of information. This arrangement not only helps readers to understand the nuances of each approach but also enables them to easily track patterns and trends in the data. Such detailed categorisation aids in identifying gaps in current research and paves the way for future investigations, making this table an invaluable asset for scholars seeking to deepen their understanding of the field or contribute new insights.

4. Presentation of findings

In this section we will discuss different findings uncovered in this study. First, we provide details of the datasets which were used in the reviewed research. Next, we present graphical representation of the bibliographic analysis. We then examine the distribution of the papers included in this review and provide an overview of year-wise publication trends for the reader's insight. Finally, we explore the various evaluation metrics employed by researchers to assess their models' performance.

4.1. Datasets details

Several datasets have been utilized by researchers for the detection of arrhythmia using ECG signals. Table 2 presents a comparative summary of prominent ECG datasets, highlighting key aspects such as the number and length of recordings, lead configurations, sampling frequencies, classification granularity (number of arrhythmia classes), and demographic details where available.

One of the most widely used datasets is the MIT-BIH Arrhythmia Database (2001), which contains 48 recordings, each 30 minutes long, collected using 2 leads at a frequency of 360 Hz. It supports 15 arrhythmia classes and includes demographic data for 47 subjects,(25 men and 22 women). This dataset remains a cornerstone in ECG-based research due to its accessibility and detailed annotations.

Other datasets from the MIT-BIH collection, such as MIT-BIH AFDB, SVDB, ST Change DB, and Malignant Ventricular Ectopy DB, offer recordings with varying characteristics (e.g., durations ranging from 30 minutes to 10 hours) and fewer arrhythmia classes, typically between 2 to 4. These datasets are also open-source, facilitating reproducibility in research.

Table 1
Summary of key findings from reviewed studies.

Method, Reference	Dataset	Classes	Feature preprocessing\Extraction	Results (%) Acc, Spec, Sen, F1	Findings
Machine learning models					
SVM, [82]	2011 CinC, MIT-BIH NSTD	11	Low pass FIR filter	acc = 88.07	Trained on a simulated dataset, which may not capture real-world noise variations. Performance drops significantly on real-world datasets with arrhythmic signals.
SVM, [83]	MIT-BIH, CUDB, VFDB	5	Mean subtraction, five-order moving average filtering, high-pass filtering, low-pass Butterworth filtering	acc = 95.0, spec = 99.0	The SVM classifier's performance is affected by data imbalance, requiring careful metric selection like BER and AUC, also for VF classification it has high variance and bias, indicating the need for more training data and additional features.
SVM, [84]	MIT-BIH	4	Median filter	acc = 86.7	Dataset dependence limits applicability to other ECG datasets. Class imbalance affects classification accuracy.
SVM, [85]	MIT-BIH	5	Spectral correlation metric, PCA, fisher score	acc = 98.6, spec = 99.7, sen = 99.2	Spectral correlation computation is resource intensive. There is certain information loss due to PCA and Fisher Score for feature reduction.
SVM, [86]	MIT-BIH	16, 5	Band-pass filter, Pan and Tompkins	acc = 99.1	The model performs lower in patient-based assessments, suggesting generalisation issues across different patients. Real-world implementation is lacking, limiting its practical use in healthcare settings.
SVM, [87]	MIT-BIH, INCART	5	Moving average filter, High-pass filter, low-pass filter, segmentation, ensemble empirical mode decomposition (EEMD), empirical mode decomposition (EMD)	acc = 99.2, spec = 99.5, sen = 98.0	Focuses on only five heartbeat types, limiting generalisability. Fixed-length segmentation may not accommodate varying heart rhythms.
SVM, [88]	MIT-BIH	6	Segmentation, wavelet multi-resolution analysis, PCA	acc = 99.7, spec = 99.8, sen = 99.1	Overfitting may occur due to beat-based cross validation with data from the same patient. The dataset's imbalance reduces the classifier's generalisation, especially for rare beats.
SVM, [89]	MIT-BIH	16	First order derivative, segmentation, Discrete Cosine Transform-Based Discrete Orthogonal Stockwell Transform, PCA	acc = 98.8, sen = 98.8	Computational constraints due to high processing time on general-purpose processors. R-peak detection accuracy dependency.
SVM, [90]	MIT-BIH	4	Wavelet transform	acc = 98.4	Noise sensitivity affects diagnosis accuracy due to interference in ECG signals. Limited beat classification focuses on only four types of beats, missing other arrhythmia types.
SVM, [91]	MIT-BIH	5	Wavelet transform, PCA	acc = 98.2	Assumes QRS morphology is sufficient, limiting generalisation. Sparse matrix computing may reduce accuracy.
SVM, [92]	MIT-BIH	5	Discrete Wavelet Transform (DWT), Independent Component Analysis (ICA), Principal Component Analysis (PCA)	acc = 94.1 (dataset D)	The OPF classifier is less generalisable, especially for clinically significant classes like V and S. The study evaluated only a limited number of classifiers (OPF, SVM, and Bayesian), potentially missing better alternatives.
SVM, [93]	MIT-BIH	4	PCA, dynamic time warping (DTW)	acc = 97.8, sen = 99.3	Lower sensitivity for fusion beats leads to potential misclassifications. The method's adaptability to real-world noisy data remains unclear.
SVM, [94]	MIT-BIH	5	WReliefF-GA-SVM	acc = 99.7, spec = 99.8, sen = 99.4	High computational complexity due to high dimensional data. Performance is tested only on the MIT-BIH dataset, limiting generalisability.
SVM, [95]	MIT-BIH	5	Local binary pattern LBP, wavelet, higher-order statistical HOS	acc = 98.3, spec = 99.3, sen = 97.4, f1 = 97.5	Computationally intensive feature extraction and optimisation may cause problems in real-time deployment.
SVM, [96]	MIT-BIH, NSDB	2	Multi-resolution wavelet-based approach, gabor filters	acc = 96.9, spec = 95.8, sen = 99.0	High computational complexity. Sensitive to parameter selection.
SVM, [97]	Chapman University and Shaoxing People's hospital	11	PCA	acc = 87.0, spec = 95.6, sen = 81.7, f1 = 82.4	The study relied on Qiskit simulation rather than actual quantum hardware, limiting the potential benefits of quantum computation. The use of PCA for dimensionality reduction leads to feature loss, which may degrade classification performance.

(*continued on next page*)

Table 1 (*continued*).

Method, Reference	Dataset	Classes	Feature preprocessing\Extraction	Results (%) Acc, Spec, Sen, F1	Findings
SVM, [98]	MIT-BIH	4	Multirate processing, wavelet decomposition	acc = 97.35 ± 0.72	Limited dataset scope affects generalisability to other ECG datasets. Computational complexity increases due to multirate processing and wavelet decomposition.
KNN, [99]	MIT-BIH	5	Discrete Wavelet Transform	acc = 98.7	The classifier's effectiveness relies heavily on selecting the most relevant features for ECG beat characterisation. The model only classifies five heartbeat types, excluding other potential arrhythmias.
KNN, [100]	MIT-BIH	17	Welch's method, Discrete Fourier Transform	acc = 98.9, spec = 99.4, sen = 90.2	Insufficient subject-oriented validation due to the limited number of ECG signals in the MIT-BIH database. The system is limited to analysing single class types per fragment, reducing its applicability in mixed heart conditions.
KNN, [101]	MIT-BIH	17	Discrete wavelet transform, 1-dimensional hexadecimal local pattern (1D-HLP), neighbourhood component analysis (NCA)	acc = 95.0	Fixed signal duration constraint limits real-time applicability. High dimensionality of features introduces computational overhead.
KNN, [102]	MIT-BIH	17	Q wavelet transform, Chi2 feature selection, neighbourhood component analysis	acc = 97.3, f1 = 96.1, prec = 98.3, recall = 94.3	The study uses a small dataset of 1000 ECG signals, limiting its ability to capture the full variability of real world arrhythmias. The model relies on feature selection methods like NCA, which may not generalise well to other datasets.
KNN, [103]	PTB-ECG, MIT-BIH NST DB	2	Normalisation, segmentation	acc = 50.0-75.0 (for highly noisy data)	Model focuses on traditional ML models rather than DL approaches. No real-time deployment feasibility analysis.
KNN, [104]	MIT-BIH, Shaoxing and Ningbo People's Hospital	5	DERMA, FrlFT	acc = 99.9, spec = 99.9, sen = 99.9	Sampling frequency differences between MIT-BIH and SPNH datasets affect feature precision. Limited class scope as the method focuses only on five heartbeat types, reducing generalisability.
KNN, [105]	MIT-BIH	5	Median filter, low pass filter, PCA	acc = 99.4, f1 = 98.9, prec = 99.1, recall = 98.8	Sensitive to noise, affecting real world classification performance. Limited to five heartbeat classes, missing other clinically relevant arrhythmias.
RF, [106]	MIT-BIH, St-Petersburg	5, 4	PCA, Discrete Wavelet Transform (DWT)	RF: acc = 99.3 (MIT-BIH), 99.9 (St-Petersburg), C 4.5: acc 98.4 = (MIT-BIH), 99.8 (St-Petersburg), CART: acc = 98.6 (MIT-BIH), 99.8 (St-Petersburg)	Dataset size limitation may not capture real world ECG variability. Computational complexity due to feature extraction and de-noising.
RF, [107]	MIT-BIH	10	Low-order polynomial, normalisation, segmentation	acc = 97.3	Limited to 10 out of 16 arrhythmia types, excluding rare conditions. Performance is dependent on the MIT-BIH dataset, affecting generalisability to other data sources.
RF, [108]	MIT-BIH, MIT-BIH SVDB	4, 2	Discrete Wavelet Transform (DWT)	acc = 98.0	Limited to binary classification, restricting multi-class capability. Single node setup may not scale for large datasets or real-time applications.
RF, [109]	MIT-BIH, CinC 2017, QT database	5	High-pass filter, low-pass filter	acc = 99.6, spec = 99.8, sen = 96.7, f1 = 97.1	Relies on MIT-BIH dataset, limiting generalisability to other datasets. Low sampling rate (250 Hz) may affect classification performance for high-frequency features.
RF, [110]	MIT-BIH	3	The Discrete Wavelet Transform (DWT), median filters,	acc = 99.5, spec = 99.6, sen = 99.3	Imbalance handling still under performs for minority classes. High computational complexity impacts real-time deployment.
RF, [111]	MIT-BIH	4	Median filter, wavelet transform coefficients, higher order statistics (HOS), and 1D-Local binary patterns (1D-LBP)	overall acc = 98.1, avg Pp = 93.9	Finding an optimal feature space for different heartbeat classes is challenging. Class imbalance affects classification results for minority classes.
RF, [112]	MIT-BIH NSR, MIT-BIH AF, MIT-BIH ST change DB	2, 3	Resampling, segmentation, band pass filter, probability statistics for feature extraction, PCA,	prec = 89.0, f1 = 89.1, recall = 89.3	Iterative algorithms (SVM, LS-SVM) show reduced stability with increased feature complexity. The study focuses on a limited range of arrhythmia types, excluding broader classifications.
RF, [113]	Own DS	2	Normalisation	AUC = 92.0 (class no, class yes)	Dataset limitations impact generalisation. Lack of clinical validation affects practical applicability.

(*continued on next page*)

Table 1 (*continued*).

Method, Reference	Dataset	Classes	Feature preprocessing\Extraction	Results (%) Acc, Spec, Sen, F1	Findings
RF, [114]	MIT-BIH , MIT-BIH SV DB, INCART	5	Pan–Tompkins algorithm, CNN	acc = 99.8, spec = 99.7, sen = 99.7, f1 = 99.9	The model's performance was evaluated on only three publicly available datasets, which may not cover the full range of patient populations and ECG variability. The fusion of features from CNN, residuals, and Bi-GRU might lead to increased model complexity.
MLP, [115]	MIT-BIH	5	Wavelet based denoising, Pan–Tompkins algorithm, Discrete Cosine Transform (DCT) coefficients, PCA	acc = 99.5, spec = 99.9, sens = 98.6 (avg)	The system classifies only five beat types, limiting broader arrhythmia detection. Heavy reliance on PCA and PNN may limit generalisation to different datasets.
MLP, [116]	MIT-BIH	5	Wavelet transform	acc = 90.9 (highest)	The classification performance for class Q is negligible due to its extremely low representation in the dataset. The OPF classifier is not significantly faster than SVM during learning, being up to five times slower in some cases.
MLP, [117]	MIT-BIH, MIT-BIH AF DB	3	DWT based denoising, Daubechies D6 ('db6') wavelet basis function, Discrete wavelet transform, discrete cosine transform	acc = 99.4, spec = 100.0, sen = 99.6	Limited dataset size may lead to overfitting and poor generalisation. The study only classifies a few arrhythmia types, missing clinically significant ones.
MLP, [118]	MIT-BIH	6	WPD	acc = 97.7	Limited dataset size may lead to problems in generalisation. Feature dependency affects model robustness.
MLP, [119]	MIT-BIH	2	PCA	acc = 99.8	Simplification trade-offs may limit model expressiveness and performance. Dataset limitation restricts generalisability as it is evaluated only on the MIT-BIH dataset.
MLP, [120]	MIT-BIH	2	Butterworth band pass filter, Pan–Tompkins algorithm	acc = 99.0 (proposed), 94.6 (NN), 93.4 (MLP), 93.3 (SVM), 91.3 (KNN)	The algorithm's accuracy is sensitive to the quality of the input ECG signals. The model only distinguishes normal vs. abnormal signals, missing specific abnormalities.
MLP, [121]	MIT-BIH	2	Low pass, high pass and butter worth filter, mean, SD, root mean square, pulse transit time, pulse rate variability	acc = 99.7 (Nbayes, highest), 94.0 (ANN), 93.0 (Adaboost), 87.5 (SVM)	Noise sensitivity, as filtering techniques may not handle all artifacts. Limited generalisation, with performance tested on a specific dataset.
MLP, [122]	MIT-BIH	5	Fourier, Goertzel, Higher Order Statistics (HOS), and Structural Co-Occurrence Matrix	acc = 94.0 (highest Bayes)	Effectiveness of SCM in noisy environments not fully explored. Impact of ECG lead configurations on performance not analysed.
MLP, [123]	MIT-BIH	6	Pan–Tompkins algorithm	acc = 97.7 (KNN)(highest)	Focuses solely on QRS complex, potentially missing abnormalities in P and T waves. Marginal improvement over existing methods indicates that further enhancements are needed.
MLP, [124]	MIT-BIH	2	Multivariate Empirical Mode Decomposition (MEMD) for denoising,	acc = 89.8	The method only classifies arrhythmia into two categories, limiting its utility for detecting a broader range of arrhythmias. The MEMD technique for denoising may not be ideal for noisy or complex multichannel ECG signals.
Bayes, [125]	CCDD	10	Automatic, PPNN	acc = 74.2, spec = 73.9, f1 = 76.2, recall = 75.2	Use of DTW for feature extraction increases computational overhead. Robustness against noisy ECG signals not extensively analysed.
Bayes, [126]	MIT-BIH	8	Normalisation, PCA, LDA	acc = 99.71 (avg)	Performance not tested on multiple datasets. Potential overfitting due to high accuracy.
Bayes, [127]	MIT-BIH	5	DWT, PCA, LDA, ICA	acc = 99.3, spec = 99.8, sen = 99.9	Difficulty in detecting subtle ECG changes. Limited classifier selection and untested performance with other classifiers.
Ensembles, [128]	UCI repository	2	Feature elimination methods	acc = 91.1	Accuracy depends on the effectiveness of feature elimination techniques. Data dependency limits generalisability to other datasets.
Ensembles, [129]	MIT-BIH	5	Moving average filter, High-pass filter, low-pass filter, ensemble empirical mode decomposition (ICEEMD), intrinsic mode functions (IMFs)	acc = 99.1 ± .08, spec = 99.4 ± .12, sen = 97.9 ± .27	Results may not generalise well beyond the MIT-BIH dataset. Fixed heartbeat segmentation may not be ideal for varying heart rhythms.

(*continued on next page*)

Table 1 (*continued*).

Method, Reference	Dataset	Classes	Feature preprocessing\Extraction	Results (%) Acc, Spec, Sen, F1	Findings
Ensembles, [130]	MIT-BIH, AHA	5	ECG resampling, Butterworth high-pass filter	acc = 98.6, spec = 99.7, sen = 84.4	Uses single lead ECG, which might limit generalisation to multi lead settings. Does not optimise classifier parameters, which might reduce accuracy.
Ensembles, [131]	MIT-BIH	4	Wavelets. LBP, HOS	None	Limited generalisation to other datasets. High computational cost due to ensemble of classifiers.
Ensembles, [132]	MIT-BIH	4	Wavelet transform, Higher order statistics. HOS	acc = 89.2, sen = 89.3, prec = 95.8	The approach relies on handcrafted features, limiting the automation potential of deep learning models. The model's generalisation is unverified beyond the MIT-BIH dataset.
Ensembles, [133]	CCDB	7	Median filter, high pass filter, WFDB tool box	acc = 75.2, f1 = 75.2, prec = 80.8, recall = 71.6,	Dataset dependency limits generalisability to other ECG datasets. Class imbalance issue may lead to biased classification results.
Ensembles, [134]	MIT-BIH	17	Standardization, rescaling, welsh method, Discrete Fourier Transform	sen = 98.3	Imbalanced data affects model robustness due to underrepresented ECG classes. The fixed input duration of 10 s ECG segments restricts flexibility for real-world applications.
Ensembles, [135]	MIT-BIH	5	Normalisation, segmentation	acc = 83.0 (highest)	Focuses on the MIT-BIH database, limiting generalisability. Does not explore the impact of noisy ECG signals and real-world artifacts.
Ensembles, [136]	Chapman University and Shaoxing People's hospital	7	Butterworth lowpass filter, local polynomial regression smoother curve fitting, non-local means (NLM) technique	acc = 94.0, sen = 94.0, f1 = 93.6, prec = 93.7	Limited dataset scope affects generalisability to other ECG databases. Class imbalance issue impacts performance for underrepresented arrhythmias.
Ensembles, [137]	MIT-BIH	16	Band-pass-filter, median filter	acc = 99.4	Dataset dependency restricts generalisability to other ECG datasets. Sensitivity to random initialisation can cause unstable classification.
Ensembles, [138]	MIT-BIH, MIT-BIH AF	3	Adaptive filter, sliding window	acc = 99.0, spec = 100.0, sen = 100.0	Limited ECG leads (only Lead 1 and Lead 2) may reduce the model's ability to capture a full range of cardiac abnormalities.
Deep learning models					
RNN, [139]	MIT-BIH AR, SVDB	5	Median filter, moving average	acc = 93.7, spec = 94.8, sen = 91.3, f1 = 89.8	Significant inter-patient ECG variability affects classification accuracy. No feature engineering, potentially missing domain-specific insights.
GRU, [140]	MIT-BIH, PTB	5,27	High pass filter, MPOA, Hyp-GRNN	acc = 99.0 (MIT-BIH), acc = 99.1 (PTB)	Low accuracy in hybrid approaches, limiting reliability for clinical applications. Prolonged processing time due to multiple stages, which could delay real-time diagnosis.
GRU, [141]	CPSC 2018	9	Morphological filtering, DWT	acc = 99.5, spec = 99.7, f1 = 98.8	Incomplete ECG segments may affect model performance due to its reliance on complete ECG beats for accurate detection. Dataset imbalance may under-represent certain arrhythmia types, affecting the model's generalisation.
LSTM, [142]	MIT-BIH	5	LSTM	f1 = 99.3, pre = 99.3, recall = 99.8	Focused solely on anomaly detection, limiting its use for specific arrhythmia classification. Dependence on normal ECG data only, which may affect performance on novel or rare abnormalities.
LSTM, [143]	MIT-BIH	5	Wavelet-based de noising, median filtering	acc = 99.3, spec = 99.7, sen = 85.9	The model depends on patient-specific labelled data, which may not always be readily available. It performs poorly in detecting supraventricular ectopic beats (SVEB) due to under-representation in the training data.
LSTM, [144]	MIT-BIH	5	High-pass (HP) and low-pass (LP) filters, Wavelet transform	acc = 99.4	Impact of different wavelet families and decomposition levels on performance not explored. Comparison with CNN-based or transformer-based models is limited.
LSTM, [145]	MIT-BIH	8	Discrete Wavelet Transform, segmentation	acc = 99.3, spec = 99.1, f1 = 99.3	Limited to only eight ECG beat types, restricting generalisability. High computational time required for LSTM training.
LSTM, [146]	MIT-BIH	7	Automatic, LSTM	acc = 35.7 (highest)	The second-stage LSTM model did not perform well because summary features are not time-series data, making LSTM unsuitable. Class imbalance and a limited amount of labelled data negatively affect the model's predictive performance.
LSTM, [147]	MIT-BIH	5	Discrete wavelet transform, u-MDFA	acc = 97.3, sen = 77.9	Limited feature set for classification. Dependence on multifractal detrended fluctuation analysis (u-MDFA).

(*continued on next page*)

Table 1 (*continued*).

Method, Reference	Dataset	Classes	Feature preprocessing\Extraction	Results (%) Acc, Spec, Sen, F1	Findings
LSTM, [148]	MIT-BIH	2	Normalisation, Butterworth bandpass filter	acc = 90.1	Limited to binary classification, not distinguishing between different arrhythmia types. No signal quality assessment, which may affect performance with noisy ECG data.
LSTM, [149]	UCI repository	16	Normalisation, PCA, LSTM	acc = 93.5, f1 = 91.7, prec = 92.8, recall = 90.7	Class imbalance affects performance, especially for less frequent arrhythmias. PCA for noise removal may discard important features.
LSTM, [150]	MIT-BIH, BIDMC Congestive Heart Failure	3	Multi-stage filter, AT method	acc = 98.9 (highest from LSTM model 3)	Relies on angle transformation, adding computational complexity. Focuses only on MIT-BIH and BIDMC datasets, limiting generalisation.
LSTM, [151]	MIT-BIH	5	Resampling, median filters	acc = 99.3, spec = 93.7, f1 = 91.7	Model lacks noise reduction techniques, which may affect real-world performance. No comparison with advanced architectures.
LSTM, [152]	MIT-BIH	5	Discrete Wavelet Transform (DWT), Delta Sigma Modulation (DSM)	acc = 99.6, sen = 99.8, f1 = 98.2	Impact of different lead configurations on classification not explored. Model's robustness against motion artifacts in wearable ECG devices not tested.
LSTM, [153]	MIT-BIH	5	Normalisation, median filters, low-pass FIR filter, wavelet transform	acc = 99.1, spec = 99.8, sen = 92.4, f1 = 95.1	Model's real-time deployment feasibility on wearable devices is not explored. Misclassification of SVEB as normal beats due to their under-representation in training data.
LSTM, [154]	CPSC 2018	9	Deformable CNN	f1 = 81.6 (avg), weighted-f1 = 86.0 (avg)	Poor interpretability of the deep learning model, hindering its integration into clinical settings. Potential overfitting, with performance improvements not guaranteed to generalise well to diverse ECG datasets.
LSTM, [155]	Chapman University and Shaoxing People's Hospital dataset	7	Butterworth lowpass filter, local polynomial regression smoother curve fitting, non-local means	acc = 98.6 (Lead II), acc = 98.3 (Lead aVF)	Single-lead limitation may affect performance when applied to multi-lead ECG data, which contains richer information. Model complexity (combining CNN, BILSTM, and BiGRU) reduces interpretability and may make deployment in clinical settings challenging.
LSTM, [156]	MIT-BIH, Own DS, PTB-XL	4, 6	SJTU-ECGNet	acc = 93.7 (avg), f1 = 83.5 (avg)	Dataset limited to a Chinese population, potentially limiting generalisation to other populations or regions. Model performance was compared with cardiologists only for a specific subset of arrhythmias, possibly neglecting others.
LSTM, [157]	MIT-BIH	2	CNN	acc = 99.6, f1 = 99.3, recall = 98.87	The model is computationally intensive and requires a GPU for deployment. The performance is heavily dependent on the MIT-BIH dataset, limiting its generalisation.
CNN, [158]	MIT-BIH	5	CNN	acc = 99.0, spec = 98.9, sen = 93.9	The system struggles with misclassifying rare or anomalous beats, particularly supraventricular ectopic beats. The model does not support incremental learning, preventing it from adapting to evolving patient-specific ECG patterns over time.
CNN, [159]	MIT-BIH	5	Daubechies wavelet 6 filters	acc = 94.0 (h)	Performance is affected by imbalanced datasets. Noise in raw ECG signals can affect classification accuracy.
CNN, [160]	MIT-BIH, AFDB, CUDB	4	Resampling, Daubechies wavelet 6, normalisation, CNN	acc = 94.9, spec = 81.4, sen = 99.1 (for 5 s duration)	Requirement for large datasets can limit the model's applicability to rare arrhythmias. Long training time makes real-time or resource-constrained deployment challenging.
CNN, [161]	own DS	14	Automatic, CNN	f1 = 80.9, prec = 80.9, recall = 82.7,	Limited detection of certain arrhythmias and heart conditions. Requires clinical validation to confirm real-world effectiveness.
CNN, [162]	MIT-BIH	17	Normalisation	acc = 91.3, spec = 99.4, f1 = 85.4, prec = 89.5	Small dataset size limits generalisability and may not capture diverse arrhythmia types. Single lead ECG signals may miss important features, reducing robustness for real-world use.
CNN, [163]	MIT-BIH	5	Heart beat segmentation, segment length scaling	acc = 98.6, spec = 99.2, sen = 93.8 (24 records)	Limited performance in handling multiple adjacent heartbeats. Reliance on a small training dataset limits generalisation.
CNN, [164]	MIT-BIH	8	Automatic, CNN	acc = 99.1, spec = 99.5, sen = 97.6 (avg)	Trained only on MIT-BIH, limiting generalisability. High computational complexity for real-time use.

(*continued on next page*)

Table 1 (*continued*).

Method, Reference	Dataset	Classes	Feature preprocessing\Extraction	Results (%) Acc, Spec, Sen, F1	Findings
CNN, [165]	MIT-BIH	5	Discrete wavelet transform, mean, variance, sum, root mean square, mean absolute deviation, normalisation	acc = 98.0 (highest)	Limited dataset validation reduces generalisability. Majority voting-based fusion may not be optimal compared to advanced methods.
CNN, [166]	CinC 2017	4	Automatic, CNN	f1 = 82.0	Limited data (1000 ECGs with AF) restricts model generalisation. Severe class imbalance challenges accurate classification.
CNN, [167]	MIT-BIH	5	CNN	acc = 98.8, spec = 99.1, sen = 95.5	The method requires fine tuning on small datasets, limiting its generalisation to new individuals. The model's computational complexity could restrict deployment on low-power wearable devices.
CNN, [168]	MIT-BIH	5	Automatic, CNN	acc = 99.5, spec = 99.8, sen = 96.9	Batch-weighted loss may not generalise well to real-world imbalanced datasets. Robustness against noisy ECG signals and artifacts not extensively tested.
CNN, [169]	MIT-BIH	5	WFDB Toolbox for segmentation, CNN	acc = 97.3	The model struggles with detecting Fusion arrhythmia due to dataset imbalance and requires more balanced datasets or augmentation techniques. Limited to five arrhythmia types, which may not suffice for comprehensive clinical diagnosis.
CNN, [13]	Own DS, PhysionNet Challenge	12	Automatic, CNN	f1 = 83.7, ROC = 97.0	Single-lead ECGs limit diagnostic capability compared to 12-lead ECGs. Small test dataset limits generalisability and subgroup analysis.
CNN, [170]	MIT-BIH, MIT-BIH(AFDB, VFDB)	4	Continuous wavelet transformation (CWT), CNN	acc = 97.8, spec = 98.8, sen = 99.8	The model requires a large amount of training data to perform effectively. The lack of input normalisation can lead to invalid classification due to inconsistent data.
CNN, [171]	MIT-BIH	5	Segmentation, STFT, CNN	acc = 99.0 (avg)	Limited generalisability due to dependency on the MIT-BIH database. High computational cost from time–frequency spectrograms.
CNN, [172]	MIT-BIH	6	GLCM	avg : f1 = 92.0, prec = 92.0, recall = 92.0	Computational complexity limits real-time implementation. Fixed parameter selection may not be optimal for all arrhythmia types.
CNN, [173]	Own dataset	Not given	Notch filter, bandpass filter, adaptive filters	spec = 98.2, sen = 90.8, f1 = 83.4	The model struggles with detecting rare arrhythmias due to insufficient training data. Noise misclassification is a concern, as noise can be misidentified as ventricular tachycardia.
CNN, [174]	MIT-BIH	5	Wavelet transform, segmentation	acc = 97.4, spec = 99.4, sen = 97.1	Model's robustness against rare arrhythmia types not explored. No analysis of noisy ECG signals on classification accuracy.
CNN, [175]	MIT-BIH	5	Wavelet packet decomposition	acc = 98.8 (avg), f1 = 97.2, prec = 99.4, recall = 95.2	Data preprocessing steps may lose valuable information. Limitations in handling distributed heartbeat features with 1D convolution.
CNN, [176]	NHANES	2	Mean, CNN	acc = 81.8, spec = 81.8, recall = 77.3	Relies on the NHANES dataset, limiting generalisability to other datasets. Does not analyse computational feasibility for real-time healthcare deployment.
CNN, [177]	Alibaba Tianchi Cloud Competition	32	Wavelet transformation db6, automatic	f1 = 92.4, prec = 93.7, recall = 91.1	Dependence on dataset quality from the Alibaba Tianchi Cloud Competition may not fully reflect real-world clinical ECG variations. The use of multiple deep neural networks increases computational cost.
CNN, [178]	MIT-BIH	8	Wavelet based thresholding, reconstruction algorithm of wavelet decomposition	acc = 99.1 (highest)	Limited to eight arrhythmia types, which restricts its diagnostic capabilities. The model's generalisability is constrained by reliance on the MIT-BIH dataset.
CNN, [179]	MIT-BIH AF, MIT-BIH NSR	2	Wavelet basis function, segmentation, CNN	acc = 99.2, spec = 98.7, sen = 99.7 (filtered signals)	The simple CNN architecture may limit performance compared to more complex models. Single-lead ECG signals may not fully capture comprehensive heart activity.
CNN, [180]	MIT-BIH, PTB dataset	5	WAVELET BASED DENOISING, normalisation	acc = 99.3, sen = 99.1, f1 = 99.1	Data imbalance may still impact the model's performance on less frequent arrhythmia classes. The model's high computational complexity could limit its real-time application.
CNN, [181]	CPSC 2018	9	Automatic	f1 = 84.0 (median)	The model achieved high accuracy on the CPSC2018 dataset but may require refinement to perform equally well on other ECG datasets. Transfer learning is needed to adapt the model to different datasets.

(*continued on next page*)

Table 1 (*continued*).

Method, Reference	Dataset	Classes	Feature preprocessing\Extraction	Results (%) Acc, Spec, Sen, F1	Findings
CNN, [182]	MIT-BIH	5	Automatic, CNN	acc = 99.43, sen = 99.2	Limited to MLII lead, reducing adaptability to multi-lead ECG. Overfitting concerns, with long-term stability untested in diverse environments.
CNN, [183]	MIT-BIH	5	Segmentation, automatic, CNN	acc = 98.4, prec = 98.4, f1 = 98.4	Tested only on specific dataset, limiting generalisation. Fixed segmentation method may not handle irregular rhythms.
CNN, [184]	MIT-BIH	11	CNN(DL)	acc = 99.0	Increased training time due to 2D conversion of ECG signals. High computational complexity limits real-time applications.
CNN, [185]	MIT-BIH	5	CNN	acc = 98.7, spec = 99.8, sen = 78.3	Relies on patient-specific ECG data, which may not be available. Assumes patient identity is known, which may not be feasible in real-world scenarios.
CNN, [186]	MIT-BIH	5	MPCNN	acc = 96.4, f1 = 89.7	Computational complexity due to symbolic representation and multi-perspective CNN. Class imbalance issue, affecting minority heartbeat types.
CNN, [187]	MIT-BIH	5	Heart beat segmentation, segment length scaling, CNN	acc = 97.8, spec = 98.8, sen = 75.6	The system still requires some manual labelling, particularly for normal beats from testing patients. The system's performance is affected by the availability and accuracy of normal beats for training, impacting its overall reliability.
CNN, [188]	MIT-BIH	5	Wavelet threshold and reconstruction	acc = 99.0	The model relies on a single ECG lead, which may limit generalisability. The dataset (MIT-BIH) is relatively small, potentially impacting model robustness.
CNN, [189]	MIT-BIH	17	Segmentation, rescaling, CNN	acc = 98.6, f1 = 94.0, prec = 95.0 (for majority)	Limited generalisation to real-world clinical data. Imperfect performance on minority classes.
CNN, [190]	MIT-BIH	5	Median filters, segmentation, Continuous Wavelet Transform (CWT), CNN	acc = 98.7, sen = 67.5, f1 = 68.4	The model struggles with accurately classifying the F class (fusion of ventricular and normal beats) due to its low representation in the dataset. Generalisation to real-world clinical settings remains untested.
CNN, [191]	MIT-BIH, MIT-BIH AFDB, MIT-BIH VFDB, CUDB	6	Automatic	acc = 95.1 (second stage)	Imbalanced dataset with fewer PAC episodes. High computational cost, limiting feasibility.
CNN, [192]	HEAT DS	2	Noise removal, wavelet transform, STFT	acc = 82.8, AUC = 90.0	Poor performance in distinguishing atrial fibrillation from atrial flutter due to process discontinuity. Limited labelled data for AFib and AFlu discrimination reduces classification accuracy.
CNN, [193]	MIT-BIH, MIT-BIH SVDB, INCART	4	CNN	sen = 90.1, f1 = 90.0	Domain shift issues lead to degraded performance on new patient data. Dependence on unlabelled data, which may not always be available.
CNN, [194]	CPSC 2018, China 12-Lead ECG Challenge, INCART, PTB, PTB-XL, Georgia 12-Lead ECG Challenge	24	Wavelet transform analysis	validation score = 42.6, 32.5, time : 1664 min, 35 min(GoogleNet, rule-based)	The model suffers from limited training data diversity. There is a significant class imbalance, affecting classification accuracy for underrepresented conditions.
CNN, [195]	CPSC 2018, PTB-XL, Georgia	9	Inception-ResNet-V2	f1 = 85.2	The model struggles with classifying premature atrial contraction (PAC) accurately due to its complexity. The approach relies on specific ECG leads (lead II and lead aVR), potentially missing crucial information from other leads.
CNN, [196]	MIT-BIH, MIT-BIH Atrial fibrillation	8,4,2	Normalisation, segmentation	acc = 99.1	Limited dataset evaluation, affecting generalisability. Transformer-based architecture requires high computational resources.
CNN, [197]	MIT-BIH, MIT-BIH NSR	3	Resampling, denoising, normalisation, ImageNet	acc = 98.8, spec = 99.0, prec = 97.1	Dependence on noise-free ECG signals affects performance in real-world scenarios. Computational constraints and hardware implementation challenges limit real-time and wearable device applications.
CNN, [198]	MIT-BIH	5	Median filter, low pass filter, temporal transition module for feature extraction	acc = 99.8, spec = 99.5, sen = 88.8 (intra patient)	The cost-sensitive loss function's effectiveness is dataset-specific and may not generalise well. The temporal transition module needs further validation across diverse datasets.

(*continued on next page*)

Table 1 (*continued*).

Method, Reference	Dataset	Classes	Feature preprocessing\Extraction	Results (%) Acc, Spec, Sen, F1	Findings
CNN, [199]	MIT-BIH, IEEE DataPort	17	Butterworth bandpass filter moving average filter, segmentation, Short-time Fourier transform (STFT)	acc = 99.1 ± 0.25, spec = 99.8 ± 0.16, sen = 99.3 ± 0.18, f1 = 99.4 ± 0.25 (highest)	The model was trained on a small dataset, limiting generalisability. Single beat analysis may overlook important contextual information from longer ECG sequences.
CNN, [200]	MIT-BIH	2	Automatic, CNN	acc = 98.8, spec = 98.8, sen = 98.6	Tested on small dataset, limiting generalisation. Not tested on multilead ECG or other signal types.
CNN, [201]	AECG	2	Fetal heartbeat detection, CNN	acc = 96.2, spec = 100, recall = 100	Accuracy is affected by heartbeat detection errors due to maternal heartbeat interference and fetal movements. Limited dataset with only 26 subjects, reducing generalisability to larger populations.
CNN, [202]	MIT-BIH, MIT-BIH SUP, CPSC, CPSC Extra, G12EC, INCART	3	Moving avg filter, band-pass filter	f1 improves by 8% and 4% for SVEB and VEB.	Imprecise labelling of data results in increased noise. Limited fine grained annotations.
CNN, [203]	MIT-BIH	4	Empirical mode decomposition (EMD)	Not given	Sensitivity to hyperparameters in preprocessing. Performance can be affected by noise and motion artifacts.
CNN, [204]	MIT-BIH	5	Relative position matrix	acc = 99.3, spec = 99.8, sen = 98.9	Focuses on the MIT-BIH dataset, limiting application to diverse datasets. High computational cost due to Gam-ResNet18 and image transformation.
CNN, [205]	MIT-BIH	4	Scaling, segmentation, CNN	acc = 99.0, spec = 99.0, sen = 94.0	The dataset's imbalance affects model generalisation, especially for under-represented arrhythmia types. The model may be overfitting to the MIT-BIH dataset, given its high accuracy.
CNN, [206]	MIT-BIH, PTB Diagnostic	5, 2	IIR Notch Filter, FIR Filterfeature extractor	acc = 98.7, prec = 98.9, f1 = 96.3, recall = 93.9 (MIT-BIH)	Computational complexity challenges real-time deployment. Segmentation sensitivity due to reliance on R-peaks, affected by noise or irregular signals.
CNN, [207]	MIT-BIH, European ST-T DS	4	Automatic, CNN	acc = 97.9, sen = 97.9	Local optima issue affects performance. High computational cost limits real-time applications.
CNN, [208]	MIT-BIH	5	CNN	acc = 97.3, sen = 99.3, f1 = 99.6	Inter-patient variability affects classification, requiring robust domain adaptation. Computational complexity limits real-time deployment.
CNN, [209]	INCART	7	DWT	prec = 99.73, recall = 99.62 (class R)	Public ECG datasets lack sufficient data for rare diseases, limiting the model's generalisability. The use of specific loss functions like Contrastive Loss limits the flexibility of the model in some applications.
CNN, [210]	Shaoxing and Ningbo People's Hospital, CPSC 2018	45, 9	CNN(global channel attention block)	acc = 95.6, f1 = 96.2, prec = 39.4, recall = 53.3 (45 classes)	Insufficient data for certain arrhythmia classes, affecting prediction performance for these specific classes. Limited dataset diversity, restricting the model's generalisation to a broader range of arrhythmia types.
CNN, [211]	MIT-BIH	4	Butterworth filter, min–max normalisation	acc = 97.9 (avg), f1 = 90.3 (avg)	Dataset imbalance remains a concern despite attempts to address it. Dependence on accurate segmentation methods, with errors in segmentation potentially impacting classification performance.
CNN, [212]	MIT-BIH	12	1D CNN	acc = 98.6, spec = 97.8, f1 = 98.4	The model's validation is limited to the MIT-BIH and PhysioNet databases, potentially affecting its ability to generalise to other patient populations. The hybrid architecture of CNNs and Transformers may require significant computational resources.
CNN, [213]	MIT-BIH, CPSC 2018	5	Continuous wavelet transform (CWT)	acc = 98.53, 99.38; f1 = 97.57, 98.65 (CPSC 2018, MIT-BIH)	The experimental dataset is small, which may not reflect the complexity of real-world data. The model requires improvement in handling ECG signal noise while retaining key features.
DNN, [214]	MIT-BIH, SVDB, INCART	4	Denoising	overall acc = 100, spec = 99.9 (MIT-BIH)	The approach depends on large ECG datasets, which may not be readily available. Active learning may not always yield the most efficient or accurate results.
DNN, [215]	MIT-BIH	2	Median filters, low pass filter, WaveForm DataBase (WFDB) Toolbox	acc = 99.7, spec = 99.8, sen = 99.5	Requires large amounts of training data for good generalisation. Computationally expensive due to deep network architecture.

(*continued on next page*)

Table 1 (*continued*).

Method, Reference	Dataset	Classes	Feature preprocessing\Extraction	Results (%) Acc, Spec, Sen, F1	Findings
DNN, [216]	MIT-BIH	5	DNN	acc = 93.1, 94.7 (Exp1, Exp2)	The method may not exceed the performance of patient-specific classifiers, which benefit from personalized data. The testing was limited to the MIT-BIH arrhythmia database, which may not fully represent diverse patient populations.
DNN, [217]	TNMG	6	Removing stop-words, generating n grams	spec = >99, f1 = >80	Lack of statistical confidence in the superiority of the model over medical residents raises concerns about its reliability. The study has not tested the model on abnormalities related to acute coronary syndromes or other complex conditions, limiting its generalisability.
DNN, [218]	Chapman University and Shaoxing People's Hospital dataset	7,4	CDNN(Cascade DNN)	acc = 99.9, spec = 99.9, sen = 99.9, f1 = 99.9	Dataset representativeness may limit the model's generalisability to diverse patients and conditions in clinical settings. Potential dataset biases could impact performance, especially for rare arrhythmias or under-represented conditions.
GAN, [219]	MIT-BIH	15	Low/high pass filter, segmentation	f1 = 91.4, AUC = 94.7	Limited abnormal ECG data affects model performance and generalisation to rare arrhythmias. GANs may lead to overfitting if not carefully balanced, impacting robustness.
Encoders, [220]	MIT-BIH	5	Auto encoder	acc = 97.0	The RDDL method has a high feature extraction time, making it slower than other deep learning methods. The effectiveness of the robustness criterion in other domains remains untested, limiting its broader application.
Encoders, [221]	MIT-BIH	6	Stacked sparse auto encoders	acc = 99.5	Noisy ECG impact not fully explored. Assumes similar distributions for training and test data
Attention, [222]	Physiological signal challenge	8	Normalisation, CNN	f1 = 81.2, prec = 82.6, recall = 80.1	Sensitivity to class imbalance affects classification accuracy. Difficulty detecting subtle ECG changes due to noise masking small waveform deviations.
Attention, [223]	MIT-BIH, MIT-BIH (SVDB) Supraventricular Arrhythmia	5	Median filters, normalisation, heart beat segmentation	acc = 98.7, f1 = 92.0	Struggles with highly imbalanced data, underperforming in rare heartbeat classes. May face computational constraints for real-time wearable device deployment.
Attention, [224]	MIT-BIH	3	Sparse low rank filter, HSDI	acc = 96.0, spec = 9, f1 = 89.14	The evaluation is based on the MIT-BIH dataset, limiting generalisability to diverse ECG data. The model involves complex processing and requires significant computational resources.
Attention, [225]	MIT-BIH	5	Normalisation, segmentation	acc = 98.9, spec = 99.8, sen = 96.6	Underperformance on SVEB detection due to class imbalance. Limited evaluation of query functions for active learning.
Transfer-learning, [226]	MIT-BIH, AFDB, VFDB, European STT DB	4	DenseNet	acc = 97.2	Small ECG datasets lead to overfitting risks. Dependence on spectrogram transformation may miss some relevant features.
Transfer-learning, [227]	MIT-BIH, PTB	29	Band pass filter by combining a low pass and a high pass filter, segmentation, DenseNet	acc = 98.6	Transfer learning from DenseNet may limit the model's adaptability to ECG-specific features. The dataset augmentation may introduce synthetic variations not reflective of real-world ECG patterns.
Transfer-learning, [228]	CinC Challenge 2017	Not given	CNN	Beat classification: f1 = 77.9 ± .014, AUC = 96.2 ± .006, 96.1 ± .004 (PTB-XL, ICBEB2018)	Limited pretraining data usage, which may hinder the model's ability to capture comprehensive features. Dependency on annotated ECG data for supervised pretraining, limiting the availability of large datasets.
Transfer-learning, [229]	MIT-BIH	5	Median filters, low-pass filter, adaptive filter, ECG segmentation, continuous wavelet transform	acc = 99.1, sen = 96.2, prec = 96.5	Data imbalance may still affect sensitivity despite using a cost-sensitive loss function. Fixed feature representation (200-dimensional) may not capture all variations in ECG signals, affecting classification accuracy.
Transfer-learning, [230]	MIT-BIH	4	Butterworth filter, segmentation	acc = 99.2 (LC-CNN, avg), acc = 98.7 (MobileNet-V2, avg)	Limited to ECG plot images, which may overlook features that could be extracted from raw ECG signals. Lack of real-world data testing, as performance on the MIT-BIH dataset may not generalise well to real-world ECG data.
Transfer-learning, [231]	MIT-BIH SV DB, INCART	5	Segmentation, VGG, ResNet	acc = 97.0, 98.0 (VGG, ResNet)	The model was evaluated on specific datasets, limiting its generalisability to other datasets. The high computational demands due to large model parameters pose challenges for deployment.

(*continued on next page*)

Table 1 (*continued*).

Method, Reference	Dataset	Classes	Feature preprocessing\Extraction	Results (%) Acc, Spec, Sen, F1	Findings
Transfer-learning, [232]	PTB-XL	11	Resize, augmentation, CNN	f1 = 50.2 (on hidden dataset)	Poor performance on low-quality mobile phone ECG images. Limited generalisation across different ECG image types.
Transfer-learning, [233]	PTB-XL, CPSC 2018	18	Resampling frequency, image resizing, ConvNetXt	f1 = 64.5 (on hidden dataset)	Struggles with varying image quality, especially black-and-white scans. Poor performance on under-represented labels.
Transfer-learning, [234]	PTB-XL	11	Image resizing, ConvNeXt	f1 = 81.7 (on hidden dataset)	Risk of overfitting despite early stopping. Limited generalisation to other datasets.
Transfer-learning, [235]	PTB-XL, PhysioNet Challenge 2021	11	Segmentation (U-Net), EfficientNet-B0	f1 = 73.0	Inefficient due to large input dataset. Rule-based digitization is prone to exceptions.
Hybrid techniques					
Hybrid, [236]	MIT-BIH	5	Daubechies wavelet 8 filter, normalisation, PCANet	acc = 97.8 (imbalanced original noisy DS)	The model's performance is inconsistent, with slightly lower accuracy on noise-free data compared to noisy data. Its real-world applicability remains uncertain, as it was only tested on the MIT-BIH arrhythmia database.
Hybrid, [237]	MIT-BIH	4	ResNet50 + VGGNet16	acc = 99.8, prec = 99.7, recall = 99.3	ECG signals must be converted into images for feature extraction, adding extra preprocessing steps. The method's reliance on pretrained CNNs designed for images may not be well-suited for 1D ECG signals.
Hybrid, [238]	MIT-BIH	5	Denoising, segmentation, CNN	acc = 99.9	Convolutional feature maps are blurred, affecting classification accuracy. Imbalanced data impacts performance, especially for minority classes.
Hybrid, [239]	MIT-BIH AFDB	2	Automatic, CNN	acc = 99.4, spe = 99.3, sen = 99.5	Performance dependent on ECG quality; noise or poor signals hinder detection. Limited generalisation to other datasets and clinical variations.
Hybrid, [240]	MIT-BIH	5	Normalisation, resizing of each heart beat	acc = 99.9, spec = 98.8, sen = 100 (class N, intra-patient paradigm), acc = 99.53, sen = 99.68, spec = 96.05, (class N, inter-patient paradigm)	Class imbalance sensitivity due to SMOTE oversampling. Inter-patient performance variability, reducing real-world reliability.
Hybrid, [241]	CPSC 2018	9	Downsampling, rearranging the length of ECG, CNN	f1 = 83.5	Limited to 12-lead ECG, reducing applicability to other setups. High computational complexity, potentially hindering real-time applications.
Hybrid, [242]	CPSC 2018	9	Downsampling, wavelet transformation	acc = 91.0	Limited ECG signal duration reduces context for detecting certain arrhythmias. Signal down-sampling reduces the quality, affecting the detection of high-frequency anomalies.
Hybrid, [243]	Physionet Challenge 2017	5	Automatic, CL3	f1 = 83.1 ± 1.5	The dataset imbalance affects the ability to classify less common arrhythmias accurately. The model may overfit, especially when trained on smaller datasets, and could benefit from an ensemble approach to improve generalisation.
Hybrid, [244]	PhysioNet/CinC Challenge	1	CNN	f1 = 82.1 (highest, 2nd architecture)	CRNN model complexity may be due to its larger number of parameters rather than inherent advantages. Data augmentation assumes a constant heart rate, limiting its applicability to real-world ECG variations.
Hybrid, [245]	MIT-BIH	5	Normalisation, CNN	acc = 98.1, spec = 98.7, sen: 97.5	Dependence on R-peak detection limits the model's robustness. Poor performance in APB detection affects classification accuracy.
Hybrid, [246]	MIT-BIH, INCARTDB, MIT-BIH(SVDB)	5	LSTM	acc = 99.7, spec = 99.8, sen = 99.1	High cost of labelling clinical ECG records. Needs frequent updates to improve generalisation.
Hybrid, [247]	INCART	4	Beat resuction, SOMTE algorithm	acc = 99.3, spec = 98.5, sen = 99.3	Relies on the INCART database, limiting generalisability. Increased computational complexity makes real-time deployment challenging.
Hybrid, [248]	CPSC 2018	9	Synchronizing ECG lengths	f1 = 80.6	Fixed input length (30-s windows) may lead to errors with recordings of varying lengths. Potential bias from data augmentation methods (simple duplication), not fully capturing natural ECG diversity.

(*continued on next page*)

Table 1 (*continued*).

Method, Reference	Dataset	Classes	Feature preprocessing\Extraction	Results (%) Acc, Spec, Sen, F1	Findings
Hybrid, [249]	Chapman University and Shaoxing People's Hospital dataset	7, 4	Butterworth low pass filter, local polynomial regression smoother (LOESS) and Non-Local Means (NLM)	acc = 96.1, spec = 98.7, sen = 98.7, f1 = 95.6	Limited generalisability due to focus on a single dataset. No exploration of real-world noise and artifacts on model performance.
Hybrid, [250]	MIT-BIH, MIT-BIH NSRDB, MIT-BIH AFDB, CinC challenge 2017	6	Daubechies wavelet, WFDB Toolbox, normalisation	acc = 97.2, sens = 97.2, spec = 97.0	High computational cost from QRS detection. Imbalanced dataset affecting accuracy.
Hybrid, [251]	MIT-BIH	15, 5	CNN	acc = 99.3 (class-oriented)	Performance on tiny classes is significantly lower, impacting real-world applicability. Limited generalisation to other ECG datasets restricts model effectiveness beyond MIT-BIH.
Hybrid, [252]	MIT-BIH	5	Automatic, CNN	acc = 99.6, f1 = 96.4	Poor performance on minority classes due to imbalanced data. High computational cost and limited real-world testing restrict real-time deployment and clinical applicability.
Hybrid, [253]	MIT-BIH	5	Automatic, CNN	spec = 98.9, sen = 98.9	Optimised for the MIT-BIH dataset, reducing applicability to other datasets. High computational complexity may hinder real-time processing.
Hybrid, [254]	MIT-BIH, CinC 2017, QT database	5	Median filters, low-pass filter, HOS	acc = 95.8 (avg)	Verification network reduces false positives but decreases sensitivity. No exploration of the impact of ensemble model count on performance.
Hybrid, [255]	MIT-BIH Normal Sinus Rhythm, MIT-BIH Arrhythmia	4	Baseline wander removal, split sequencing, ECG normalisation	acc = 97.9	Publicly available datasets have missing information, and certain conditions overlap, leading to data inconsistencies. Overfitting may occur with small training datasets, affecting generalisation.
Hybrid, [256]	MIT-BIH AR, MIT-BIH normal sinus rhythm, BIDMC	4	Continuous Wavelet Transformation (CWT)	acc = 99.0, spec = 99.0, sen = 96.0, f1 = 96.0	The study relies on a limited dataset (MIT-BIH), which may not generalise well. The model's performance is not tested on live ECG signals, making real-world applicability uncertain.
Hybrid, [257]	MIT-BIH AFDB, MIT-BIH VFDB, CUDB	4	Stationary wavelet transforms (SWT), median filter with Savitzky–Golay (SG) filter, segmentation, z-score normalisation	acc = 99.4	Class imbalance in datasets may bias classification performance towards normal beats. Limited validation techniques, lacking leave-one-subject-out validation for better generalisation.
Hybrid, [258]	MIT-BIH, St-Petersburg DS(INCART)	5	WFDB package for ECG signals extraction, normalisation, resampling, CNN	acc = 98.0	Model performance is limited by its generalisability, performing poorly on the St-Petersburg dataset. Resampling techniques to handle imbalanced data may introduce biases.
Hybrid, [259]	MIT-BIH	5	Discrete Wavelet Transform (DWT), segmentation, resampling, PCA, BLSTM	acc = 98.3 (BILSTM, highest)	Data imbalance affects model performance, especially for rare arrhythmia types. Lack of clinical validation for practical healthcare adoption.
Hybrid, [260]	MIT-BIH	5	Low pass Butterworth filter, wavelet transformation	acc = 97.0	Block-based Neural Network increases computational complexity. Focuses primarily on the MIT-BIH dataset, limiting generalisation.
Hybrid, [261]	MIT-BIH, AFDB, SVDB, CUDB	4	VDCNN	acc = 97.4 (with fusion), 94.8 (without fusion)	Absence of preprocessing techniques, potentially affecting robustness to noisy or incomplete ECG data. Limited class focus, only classifying four types of arrhythmias, which could restrict clinical applicability.
Hybrid, [262]	MIT-BIH, EBD, INCART	4	1D CNN	acc = 99.3, spec = 99.6, sen = 98.6, f1 = 98.7, prec = 98.68	Focuses on limited benchmark datasets, which may not generalise well. Model's effectiveness in real-time or noisy environments is not thoroughly evaluated.
Hybrid, [263]	MIT-BIH	5	Discrete wavelet transform (DWT), segmentation, normalisation, PCA	acc = 98.4 (avg)	Feature selection may result in the loss of important sample information. Imbalanced data may lead to model overfitting, affecting generalisation.
Hybrid, [264]	MIT-BIH, PTB	5, 1	Gramian Angular Field (GAF), Recurrence Plot (RP) and Markov Transition Field (MTF)	acc = 99.7 (MIT), 99.2 (PTB)	Data limitations, with training restricted to MIT-BIH and PTB datasets. Computational constraints due to the need for three separate AlexNets.
Hybrid, [265]	MIT-BIH	4	Normalisation, segmentation,1D-CNN	acc = 99.5, sen = 98.2	Fixed feature selection may not adapt well to new ECG patterns, limiting generalisability. The study does not discuss the feasibility of real-time arrhythmia detection, which is crucial for practical use in clinical settings.

(*continued on next page*)

Table 1 (*continued*).

Method, Reference	Dataset	Classes	Feature preprocessing\Extraction	Results (%) Acc, Spec, Sen, F1	Findings
Hybrid, [266]	own data	–	Statistical + frequency domain methods	acc = 95.0	Computational complexity due to hybrid optimisation and deep CNN. IoT network constraints may cause latency and connectivity issues.
Hybrid, [267]	MIT-BIH	5	Automatic, CNN	acc = 98.9	Real-time feasibility on low-power devices not explored. Noisy ECG data not extensively evaluated.
Hybrid, [268]	MIT-BIH	5	Butterworth high-pass filter, Pan–Tompkins algorithm	acc = 99.1 (highest)	Limited ECG class coverage, restricting its scope. Lack of large-scale clinical validation for real-world applications.
Hybrid, [269]	MIT-BIH	16	Median filters, 12-tap FIR low-pass filter	acc = 96.3 (highest)	Real-time hardware implementation on an ARM-based platform not tested in clinical settings. No comparison with modern deep learning-based classification approaches.
Hybrid, [270]	MIT-BIH, SVDB, INCART	4,5,12	Automatic	acc = 99.4, sen = 94.1 (VEB)	The effectiveness of focal loss in addressing class imbalance is not extensively compared to other balancing techniques. The generative module introduces computational overhead by converting 1D ECG signals into 2D images.
Hybrid, [271]	MIT-BIH, CPSC 2018	5	Residual network and LSTM	acc = 99.8, spec = 99.9, sen = 99.5	Limited ECG class coverage, restricting broader applicability. Validation only on MIT-BIH, lacking real-world clinical validation.
Hybrid, [272]	MIT-BIH	15	CNN(DL)	acc = 98.0, spec = 97.4, sen = 97.7	Mode collapse neglects minor classes. High computational overhead.
Hybrid, [273]	MIT-BIH	5	Discrete Wavelet Transform, segmentation	acc = 98.7, spec = 98.5, sen = 98.2	Traditional resampling methods may lead to data loss or overfitting. ECG-DCGAN model requires further refinement for smoother generated data.
Hybrid, [274]	PhysioNet 2017, MIT-BIH	5	Chebyshev Type II filtering, Daubechies wavelet, Z-normalisation	acc = 99.9, f1 = 98.4, prec = 97.8, recall = 99.9	Dataset constraints due to noise in the MIT-BIH database and limited availability of gold-standard labelled datasets. Real-world testing has not been conducted, limiting its clinical applicability.
Hybrid, [275]	MIT-BIH	5	Wavelet threshold denoising,	acc = 95.5	GAN-based approach may struggle with mode collapse, reducing ECG diversity. No evaluation on highly noisy ECG signals.
Hybrid, [276]	MIT-BIH	4	Conv + bi-lstm (DL)	acc = 94.7 (inter-patient)	Data imbalance affects model performance. Risk of overfitting with synthetic data.
Hybrid, [277]	MIT-BIH	5	z-score normalisation, filtering, denoising, segmentation	acc = 99.6, f1 = 98.2, prec = 97.6, recall = 99.6	Limited dataset scope restricts generalisation. High computational intensity limits real-time deployment.
Hybrid, [278]	MIT-BIH	4	Modified frequency slice wavelet transform (MFSWT)	acc = 97.5 (overall)	The model requires additional patient-specific data for training, limiting its scalability. Computational demands are high, requiring around an hour for training, which may hinder real-time applications.
sffamilyHybrid, [279]	MIT-BIH,2011 PhysioNet/Computing in Cardiology Challenge	5	ECG compression	acc = 99.4	Additional preprocessing step increases computational overhead. Focuses solely on the MIT-BIH dataset, limiting generalisation.
Hybrid, [280]	MIT-BIH SVDB, NSTDB	5	WFDB (WaveForm Database)	acc = 98.9, spec = 99.8, sen = 96.3	Data imbalance may still affect sensitivity despite using a cost-sensitive loss function. Fixed feature representation (200-dimensional) may not capture all variations in ECG signals, affecting classification accuracy.
Hybrid, [281]	MIT-BIH	5	LSTM	average (acc = 99.7, spec = 99.8, sen = 99.4) (beat-based approach)	Class imbalance affects accuracy for minority classes. Lower record-based accuracy suggests potential overfitting.
Hybrid, [282]	MIT-BIH	–	Bandpass filter, normalisation	f1 = 92.6 ± .003, prec = 92.3 ± 0.003, recall = 93.0 ± .003	Limited anomaly data may affect the algorithm's performance when dealing with more anomalies than tested. Generalisation to other time series data beyond ECG is untested, limiting its applicability in other domains.

(*continued on next page*)

Table 1 (*continued*).

Method, Reference	Dataset	Classes	Feature preprocessing\Extraction	Results (%) Acc, Spec, Sen, F1	Findings
Hybrid, [283]	MIT-BIH, SVDB	5	Low-pass filter, adaptive filter	acc = 98.6, sen = 97.9	Limited feature set that may not capture all relevant ECG characteristics. No real-time testing, which limits practical application in continuous ECG monitoring.
Hybrid, [284]	MIT-BIH, NSRDB, AFDB, CinC challenge 2017	6	Dual tree complex wavelet transform (DTCWT),	acc = 97.2, spec = 96.2, sen = 99.4	Limited arrhythmia types classified. Overfitting to the MIT-BIH dataset.
Hybrid, [285]	MIT-BIH, ICENTIA 11 K	5	CNN + autoencoder	acc = 97.7 (transformer + CNN), 97.9 (transformer + autoencoder)	Simple data augmentation techniques limit model's generalisation for minority classes. Underperformance on minority classes reduces practical usability in diverse real-world settings.
Hybrid, [286]	ECG5000	2	Segmentation autoencoder	acc = 98.0, f1 = 96.0	Scarcity of annotated ECG data limits supervised learning. Model trained only on normal ECG signals, limiting anomaly detection.
Hybrid, [287]	CPSC 2018, PTB	9, 8	Butterworth low pass filter, CNN	avg f1 = 87.4, 90.2 (CPSC, PTB)	High computational cost limits real-time applicability. Model may struggle with overfitting, affecting generalisability to real-world clinical scenarios.
Hybrid, [288]	MIT-BIH	5	Wavelet sub-band coefficients and HOS cumulants, PCA	acc = 99.3, spec = 98.1, sen = 99.6	Model complexity and performance are affected by the number of hidden neurons and convolutional kernels. Computational efficiency was prioritised over maximising performance, potentially limiting accuracy.
Hybrid, [289]	MIT-BIH	8	Dual-Tree Complex Wavelet Transform	acc = 98.5 (highest from 8, MSE = 0.1)	High memory usage may limit real-world application. Training procedure was limited to two batch sizes, and dynamic sampling could improve performance.
Hybrid, [290]	PhysioNet 2017	4	ResNet +Bi-lstm	acc = 86.7, f1 = 88.0	Limited dataset variability impacts generalisation. High computational cost limits real-time applicability.
Hybrid, [291]	MIT-BIH (NSTDB)	Not given	Normalisation, CNN	acc = 98.9, prec = 98.8, recall = 98.5	Limited real-world noise evaluation (motion artifacts). No real-time validation under varied conditions.
Hybrid, [292]	MIT-BIH	5	Bandpass Butterworth filter, PT algorithm, normalisation	acc = 98.9, spec = 99.3, sen = 96.5	Computational complexity of CNN-LSTM-Attention model challenges real-time use. Effect of time representation on ECG lead configurations not explored.
Hybrid, [293]	MIT-BIH	5	Wavelet Transform (WT), soft-thresholding filtering	acc = 99.8, spec = 99.9, perc = 99.6 (for intra patient)	Imbalance handling still underperforms for minority classes. High computational complexity impacts real-time deployment.
Hybrid, [294]	MIT-BIH	5	Pan–Tompkins algorithm, CNN	f1 = 90.8	Imbalanced dataset leads to challenges in classifying rare heartbeat types (e.g., SVEB). Fixed input length constraint may cause information loss for heartbeats of varying durations.
Hybrid, [295]	MIT-BIH	5	Denoising, segmentation, normalisation, CNN	acc = 95.7, sen = 88.1, f1 = 82.6	Scarcity of annotated ECG data limits supervised learning. Model trained only on normal ECG signals, limiting anomaly detection.
Hybrid, [296]	MIT-BIH	4	Median filter, subtraction	acc = 95.5, sen = 96.6, f1 = 97.7	Inter-patient variability impacts classification accuracy. Class imbalance handling may still struggle with minority class representation.
Hybrid, [297]	MIT-BIH	3	Mean removal, moving average filter, high pass filter, AlexNet	acc = 92.4 (testing)	The model only classifies ECG signals into three categories, limiting its applicability to a broader range of arrhythmias. The reliance on pre-trained AlexNet may reduce effectiveness for ECG-specific patterns, and the model may overlook clinically important handcrafted features.
Hybrid, [298]	MIT-BIH	5	Automatic	acc = 94.5 (noisy data), 98.9 (clean data)	Federated learning increases the risk of data poisoning attacks, requiring mechanisms for data integrity and authentication. The framework assumes homogeneous data and devices, which may affect model performance in real-world scenarios.
Hybrid, [299]	MIT-BIH	2	1D-CNN	acc = 96.7, f1 = 96.5	Single-lead ECG analysis limits clinical applicability. Motion artifacts may affect signal quality despite noise reduction.

Table 2
Datasets comparison.

Database & Reference	Year	Recording conditions		Leads	Frequency	Classes	Demographic Info		Open source
		No. of recordings	Length of recordings				No. of subjects	Gender	
MIT-BIH DB [300]	2001	48	30 min	2	360 Hz	15	47	25 men, 22 women	Yes
MIT-BIH AF DB [301]	1983	25	10 h	2	250 Hz	4	25	Not given	Yes
MIT-BIH SV DB [302]	1990	78	30 min	2	360 Hz	4	Not given	Not given	Yes
MIT-BIH ST Change DB [303]	1983	28	varies	2	360 Hz	4	Not given	Not given	Yes
MIT-BIH Malignant Ventricular Ectopy DB [304]	1986	22	30 min	2	250 Hz	2	Not given	Not given	Yes
MIT-BIH NST DB [305]	1984	12	30 min	2	360 Hz	1	Not given	Not given	Yes
MIT-BIH NSR DB [306]	2000	18	24 h	2	128 Hz	2	18	5 men, 13 women	Yes
CinC 2011 [307]	2011	Not given	10 s	12	500 Hz	6	Not given	Not given	Yes
CinC 2017 [308]	2017	12186	30 s	1	300 Hz	4	Not given	Not given	Yes
CU DB [309]	1986	35	8 min	5	250 Hz	3	35	Not given	Yes
UCI Repository [310]	1997	Not given	Not given	12	Not given	16	452	Not given	Yes
INCART [311]	2008	75	30 min	12	257 Hz	11	32	17 men, 15 women	Yes
PAF prediction challenge [312]	2001	50	30 min	2	128 Hz	2	100	Not given	Yes
AHA [313]	1977	154	3 h	2	250 Hz	8	Not given	Not given	No
CCDB [314]	2012	120	10 s	12	500 Hz	7	Not given	Not given	Yes
CPSC 2018 [315]	2018	9831	7–60 min	12	500 Hz	9	9458	Not given	Yes
PTB DB [316]	2004	549	2 min	15	1K Hz	9	290	209 men, 81 women	Yes
PTB-XL DB [317]	2022	21837	10 s	12	1K Hz	71	18885	Not given	Yes
Georgia 12-Lead [318]	2020	66361	Not given	12	500 Hz	24	Not given	Not given	No
Chapman & Shaoxing & Ningbo [319]	2020	45152	10 s	12	500 Hz	11	45152	Not given	Yes
Chapman & Shaoxing [320]	2019	10646	10 s	12	500 Hz	11	10646	5956 men, 4960 women	Yes
ICENTIA 11 K DS [321]	2019	11000	70 min	1	250 Hz	3 rhythm	11000	Not given	No
European ST-T [322]	2009	78	varies	2	250 Hz	2	78	70 men, 8 women	Yes

Recent large-scale datasets, such as PTB-XL (2020), Chapman & Shaoxing (2020), and Georgia 12-Lead (2020) provide substantial volume and diversity. For instance PTB-XL offers over 21,000 recordings sampled at 500 Hz with 12-lead ECGs, covering 71 arrhythmia classes. Similarly, the Chapman & Shaoxing dataset contains over 45,000 recordings, making it well-suited for deep learning approaches that require extensive data.

Shorter-duration datasets like CinC 2011 and CinC 2017 provide brief 10-second and 30-second segments, respectively, but are annotated with a reasonable number of arrhythmia types (up to 6 classes), making them suitable for real-time and portable ECG analysis applications.

In terms of lead configuration, datasets vary significantly. Some, like the CU DB and MIT-BIH series, use 2–5 leads, while most modern large-scale datasets (e.g., PTB-XL, CPSC 2018, Chapman & Shaoxing) employ the full 12-lead configuration. Higher lead counts allow for more comprehensive cardiac activity analysis but also increase the complexity of data handling and model design.

Regarding sampling frequency, older datasets tend to use 250–360 Hz, sufficient for traditional signal processing. Newer datasets often use 500 Hz or even 1 kHz (e.g., PTB DB), which better preserves signal fidelity for advanced AI models, particularly CNNs.

Demographic information is inconsistently reported across datasets. For example, the INCART dataset includes specific gender details (17 men, 15 women), whereas many others, including large datasets like Georgia and Chapman & Shaoxing, do not provide such metadata. The inclusion of age and gender data is crucial for ensuring the generalisability and fairness of AI models across population subgroups.

Finally, open-source availability remains a critical factor for the widespread adoption of datasets in research. While most MIT-BIH variants, PTB-XL, and Chapman databases are openly accessible, others like AHA and Georgia are not, potentially limiting their utility within the broader research communities.

4.2. Frequency of datasets used

Understanding which datasets are most commonly used for arrhythmia detection research is essential for selecting the most appropriate dataset for future studies. Fig. 3 provides a visual summary of the frequency of dataset usage across the reviewed literature. It is evident from the figure that the MIT-BIH Arrhythmia database is by far the most widely used, appearing in 124 out of the 219 studies included in this review. This popularity can be attributed to its comprehensive annotations, open accessibility, and long-standing use as a benchmark for arrhythmia detection.

Following MIT-BIH, the MIT-BIH AF dataset (a specialised subset focusing on atrial fibrillation) and the INCART Database are the second most frequently utilised, with each dataset being employed by 14 studies. Notably, several other variants of the MIT-BIH collection also appear among the top ten most frequently used datasets, further underscoring their widespread acceptance and relevance within the research community.

It is important to note that some studies used multiple datasets to enhance the robustness and generalisability of their results. Fig. 3 illustrates all datasets employed across the surveyed studies, providing a comprehensive overview of dataset popularity within the domain of arrhythmia detection.

4.3. Graphical representation of bibliographic analysis

Figs. 4 and 5 were generated using VOS Viewer [323]. Fig. 4 illustrates the bibliographic coupling of journals where articles related to our research domain have been published. Various colours indicate distinct clusters, which are formed based on bibliographic coupling. In contrast, Fig. 5 displays a graph constructed using citations and co-citation analysis of articles within our research domain.

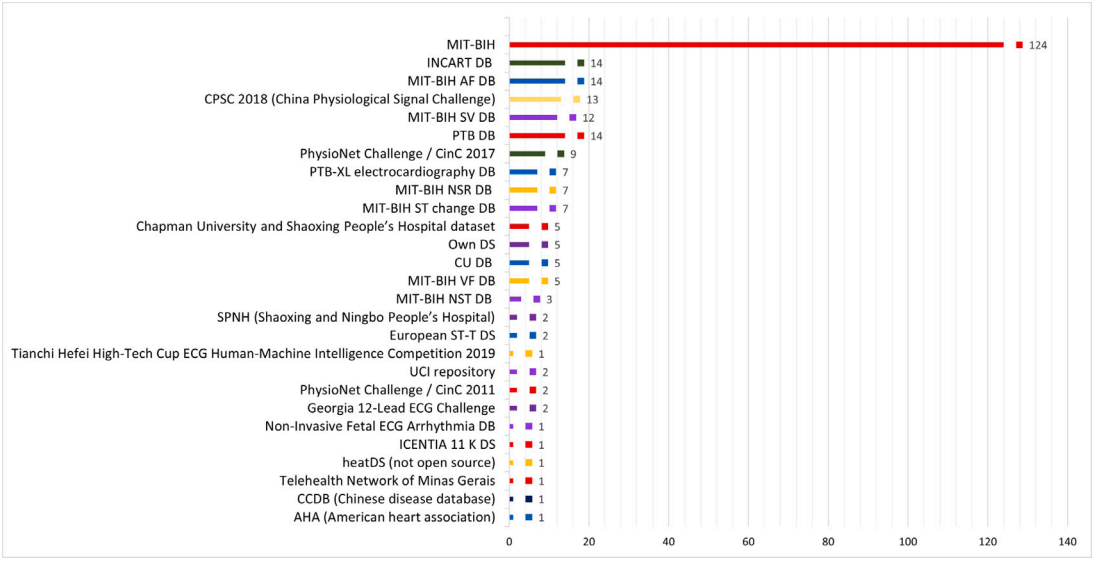


Fig. 3. Frequency of usage of datasets.

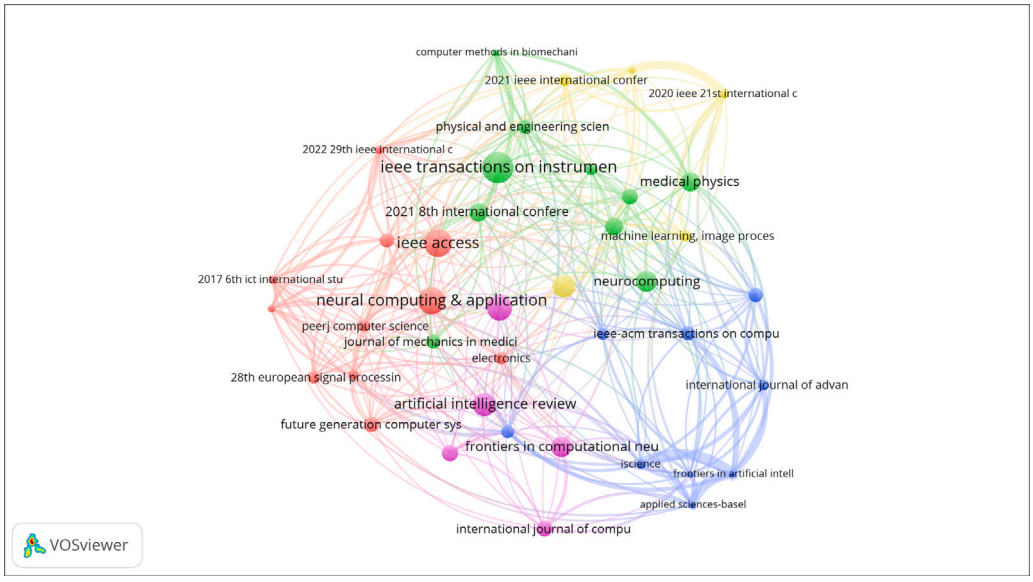


Fig. 4. Journals with most publications related to Arrhythmia.

4.4. Distribution of publications

Our primary objective throughout this study was to create a comprehensive collection of articles from esteemed and reputable journals, to ensure the integrity of research material. As previously explained, our selection criteria were strict, encompassing papers exclusively from journals categorized as Q1 or Q2 in terms of their academic standing. Conversely, we applied a similarly rigorous standard when considering papers presented at renowned conferences within our research domain. Fig. 6 provides a visual representation of the composition of our review, revealing that only 14 conference papers, constituting only 6% of the total (14 out of 219), were included. In contrast, the majority of the articles, comprising a substantial 94% (205 out of 219), were sourced from publications in reputable journals. If there was a paper, about which we were unable to decide upon selection, we set the criteria for it to have at least 50 citations in order to be considered for including in this survey. This selective approach aimed to ensure the inclusion of high-quality and influential sources in our analysis.

4.5. Year wise distribution of articles

In this review paper, we have included papers published during the time period of 2013 to 2024. One of our main focuses was to pinpoint the latest trends in the area of arrhythmia detection using ECG signals, thus we tried to select as many as of the current papers available. As illustrated in Fig. 7, a major portion i.e. 75% of the studies included in this survey are drawn from the year 2019 and onward. This deliberate choice is justified by the rapid advancements in AI techniques and their application to ECG-based arrhythmia detection. In recent years, significant progress has been made with the introduction of novel architectures such as Transformers, Graph Neural Networks, and enhanced DL models that outperform traditional approaches. Additionally, the availability of large, high quality public datasets and improved benchmarking practices have enabled researchers to achieve more accurate and generalisable results. Moreover, the field has increasingly focused on real-world applicability, including clinical validation, robustness, and interpretability. These are some areas that have gained prominence only in recent studies. By prioritising these recent works, our

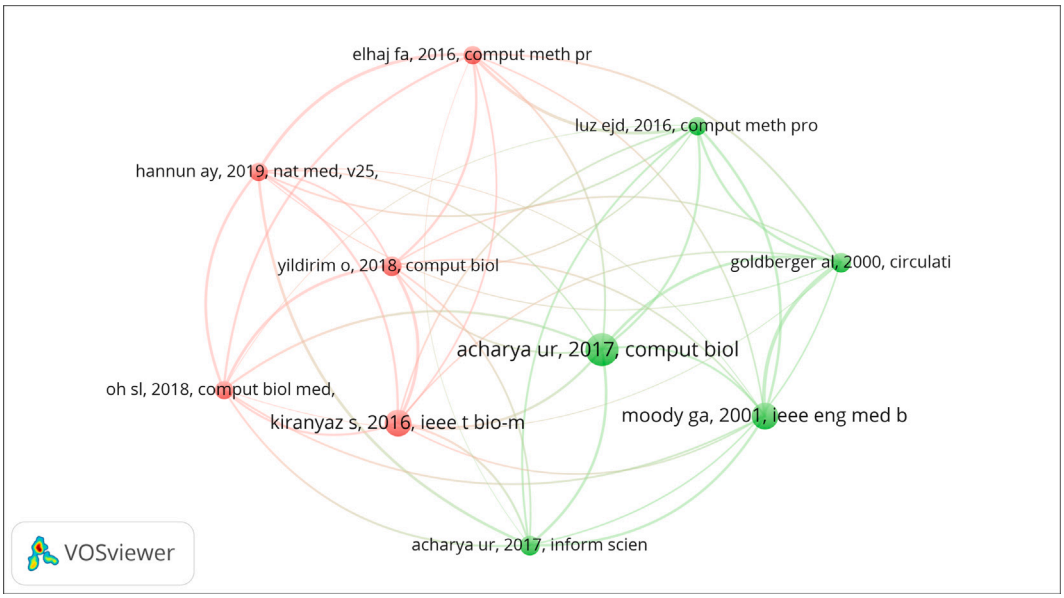


Fig. 5. Papers with most citations.

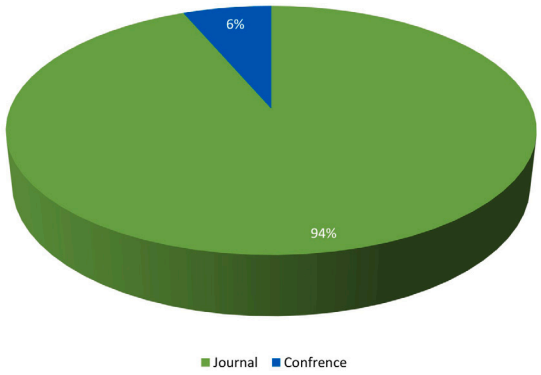


Fig. 6. Journal and conference paper proportion.

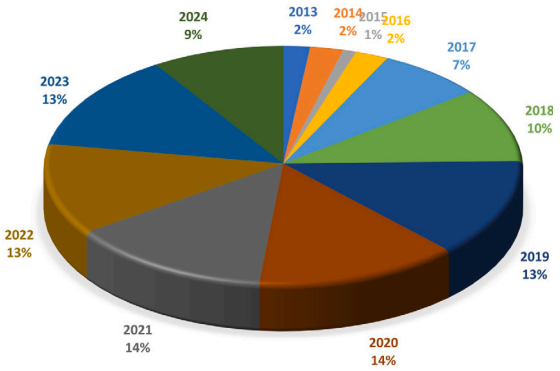


Fig. 7. Year-wise paper proportion.

review aims to provide an up-to-date, comprehensive overview of the current state-of-the-art and emerging trends, identifying both existing challenges and potential future directions in AI-based ECG analysis.

4.6. Evaluation metrics

The performance of models used for arrhythmia detection is evaluated using several statistical metrics, which are described below.

Accuracy: Accuracy is a statistical measure that indicates how well a binary classification model correctly identifies a given condition. It is calculated as:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{10}$$

where TP, TN, FP, and FN refer to true positive, true negative, false positive, and false negative, respectively.

Sensitivity (recall): Sensitivity, also known as Recall, is the ratio of true positives that have been correctly identified by a certain test [324]. It is calculated as:

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{11}$$

Specificity: Specificity measures the ratio of true negatives correctly identified by a test [325]. It is computed using:

$$Specificity = \frac{TN}{(TN + FP)} \tag{12}$$

Precision: Precision measures how many of the positive identifications were actually correct. It is defined as:

$$Precision = \frac{TP}{(TP + FP)} \tag{13}$$

F1-score: The F1-Score is the harmonic mean of Precision and Recall, providing a balanced measure that accounts for both false positives and false negatives.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{14}$$

4.7. Comparative analysis for different methods

In this section, we provide a comparative analysis of AI methods for arrhythmia detection, evaluating their strengths, weaknesses, and practical applicability based on the dataset and performance metrics analysed in this study.

Fig. 8 presents a comparative analysis of different methods on basis of four key evaluation metrics, i.e. accuracy, sensitivity, specificity and F1-score. Based on the results summarised in the findings and visualised in the performance comparison chart, the top-performing methods include RF, Deep Neural Networks (DNN), Hybrid models, CNN, and Ensemble approaches. Each of these methods demonstrates

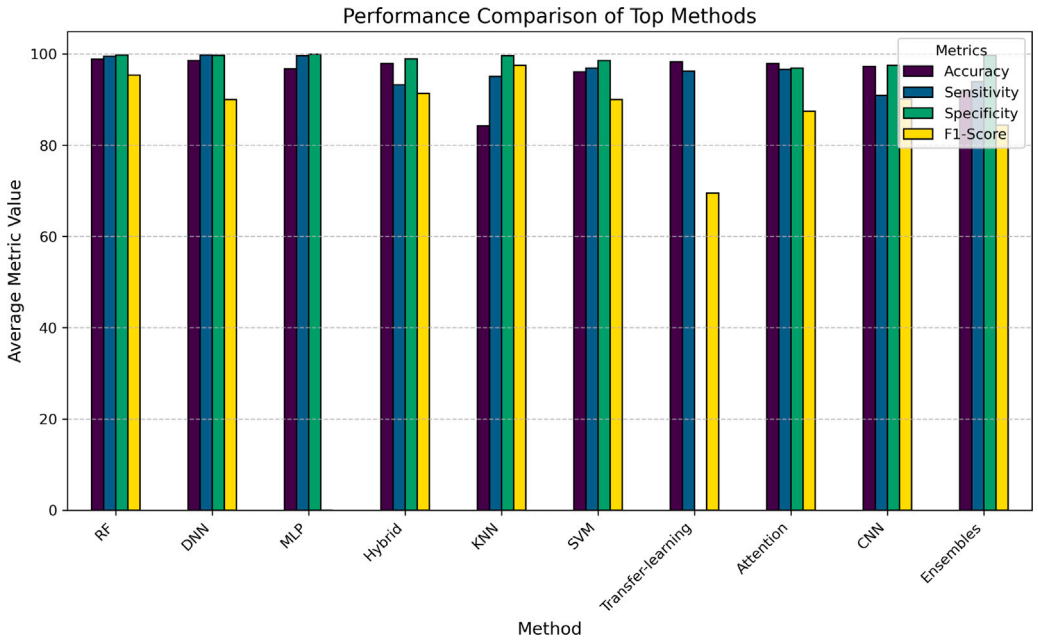


Fig. 8. Performance comparison of methods.

high accuracy and specificity, but variations in sensitivity and F1-score provide key insights into their relative strengths and limitations.

Hybrid Models consistently show strong performance across all metrics, making them one of the most balanced approaches. They deliver near-perfect accuracy and specificity while maintaining high sensitivity and F1-score, indicating low false negative rates and robust overall classification performance.

CNN and Ensemble methods also exhibit high accuracy and specificity; however they tend to show slightly lower sensitivity compared to Hybrid Models and DNNs. This suggests that while they are highly effective at general classification tasks, they may have some limitations in correctly detecting certain positive cases, potentially leading to higher false negative rates.

According to the analysed data, KNN and Attention-based models demonstrate comparatively lower performance in sensitivity and F1-score, despite acheiving relatively high accuracy. This disparity indicates potential overfitting or challenges in correctly identifying minority class instances.

Finally RF and DNNs achieved nearly 100% accuracy and specificity, making them the most reliable classifiers in terms of correctly identifying both positive and negative cases. However, RF shows a slight drop in sensitivity, suggesting a tendency to miss some positive cases.

Overall, among the reviewed approaches, Hybrid models stand out as the most consistent and effective, achieving a balance between accuracy, sensitivity, specificity, and F1-score. In contrast, approaches like KNN and Attention-based models reval limitations in detecting positive cases, which could affect their practical deployment for clinical decision support.

Fig. 9. Accuracy-based forest plot.

4.8. Forest plot analysis of model performance

To strengthen the comparative analysis of model performance, we constructed a forest plot shown in Fig. 9, which illustrates the distribution of accuracy values across various AI-based approaches in ECG arrhythmia detection. Each point on the plot represents the accuracy reported by an individual study for a given model. The models are grouped by algorithm type (e.g., SVM, CNN, LSTM), with each group displayed along the *y*-axis and corresponding accuracy values plotted along the *x*-axis.

This visual representation enables immediate identification of performance trends and variability within and across model types. For example, models such as CNNs, LSTMs, and Hybrid approaches edmosntrate consistently high accuracy, often clustering between 95%–100%, indicating robust and reliable classification performance. In contrast, traditional models such as Naïve Bayes and KNN display wider variability and lower peak accuracy, reflecting their more limited generalisability and potential sensitivity to dataset characteristics.

Importantly, the forest plot highlights both the central tendency and the dispersion of accuracy values for each method, providing clear insights into their reliability and consistency. This level of detail supports a deeper understanding of how performance varies across methodological categories and offers valuable evidence for selecting appropriate models in future research and potential clinical applications.

4.9. Individual pearson correlation analysis by method

In Fig. 10, we present Pearson's correlation coefficients between different evaluation metrics to assess the degree of interdependence among them. These correlation coefficients are complemented by additional statistics, including p-values, r-values, confidence intervals(CI), and effect sizes,providing a comprehensive understanding of how strongly and reliably these metrics relate to each other. Pearson's correlation coefficient quantifies the strength and direction of a linear relationship between two continuous variables, with values ranging from −1 to 1, where values closer to 1 or −1 indicate a stronger linear relationship, and values near 0 suggest a weaker or no linear relationship. The *p*-value tests the significance of the correlation, with smaller values (typically less than 0.05) suggesting that the observed relationship is statistically significant, while larger p-values imply that the relationship may have occurred by chance. The confidence intervals around these coefficients provide a range of values within which the true correlation likely falls, offering insight into the reliability of the correlation estimate. Effect size, on the other hand, measures the magnitude of the correlation, providing a sense of how strong the relationship is in practical terms, which is particularly useful in understanding the practical significance of the findings.

Understanding these correlations, along with their statistical significance and effect size, can provide valuable insights to researchers when analysing their results. It helps them determine whether their findings align with existing studies in the field and assess the practical relevance of the correlations. Additionally, this analysis can guide researchers in selecting the most suitable evaluation metrics based on the characteristics of their dataset and the specific objectives of their study, ensuring that they are using metrics that not only correlate well but also have meaningful statistical significance and effect sizes.

Pearson's correlation was chosen over Spearman's correlation because it measures the strength and direction of a linear relationship between two continuous variables, making it particularly useful for evaluating metrics that exhibit a linear dependency. Unlike Spearman's correlation, which assesses monotonic relationships and is less sensitive to linearity, Pearson's correlation provides a more precise measure when the relationship between variables follows a straight-line pattern.

4.9.1. SVM

The Pearson's correlation analysis of evaluation metrics for the SVM model provides valuable insight into the internal consistency of performance measures in arrhythmia detection. A strong and statistically significant positive correlation was observed between accuracy and sensitivity ($r = 0.97$, $p < 0.001$), with a 95% confidence interval of [0.86, 0.99] and a large effect size. Suggesting that as the model becomes more accurate, its ability to correctly identify arrhythmic cases (true positives) improves proportionally. Similarly, the relationship between accuracy and specificity was moderately strong ($r = 0.76$, $p = 0.028$), indicating that improvements in accuracy are also associated with better identification of non-arrhythmic instances (true negatives), although the broader confidence interval [0.13, 0.95] implies greater variability and potentially less consistency across datasets or sampling conditions.

Interestingly, perfect correlations ($r = 1.00$) were found between F1-score and the other three metrics (accuracy, sensitivity, and specificity). While this might initially suggest exceptionally robust performance, the corresponding p-values of 1.000 and fixed confidence intervals [1.00, 1.00] raise concerns about overfitting or metric redundancy, particularly in datasets where class imbalance or uniform model behaviour may distort the independence of evaluation metrics. Researchers are advised to investigate whether these results stem from deterministic relationships inherent to the dataset or metric formulation rather than meaningful model behaviour.

The moderate correlation between sensitivity and specificity ($r = 0.66$, $p = 0.108$) did not reach statistical significance and featured a wide confidence interval [−0.19 to 0.94], highlighting a possible trade-off in the model's ability to balance true positive and true negative predictions. This suggests that while the SVM model performs well in individual metrics, achieving an optimal balance between sensitivity and specificity remains a challenge; a critical consideration for clinical deployment where both false negatives and false positives carry substantial risk.

Overall, these findings indicate that while the SVM model can deliver strong and consistent performance across key evaluation metrics, researchers should be cautious when interpreting highly correlated or perfect metric values. Such insights can guide future model development by emphasising the importance of independent and diverse evaluation metrics, particularly when addressing imbalanced or complex medical datasets.

4.9.2. KNN

The Pearson's correlation analysis for the KNN model reveals a series of perfect correlations ($r = 1.00$) across most pairs of evaluation metrics, including accuracy, sensitivity, specificity, and F1-score. Even though these results might first seem to show that the model performs very consistently, the statistics behind them suggest we should be careful when interpreting them. All observed p-values were equal to 1.000, and the corresponding 95% confidence intervals were fixed at [1.00, 1.00], which is statistically atypical and indicative of potential redundancy or deterministic dependencies among the metrics rather than meaningful variability.

Such perfect correlations may arise from insufficient diversity in the underlying data, uniform prediction behaviour, or structural issues such as class imbalance or overfitting. In these scenarios, the model's performance metrics fail to reflect independent aspects of classification quality, reducing the practical interpretability of the results.

Notably, the correlations between sensitivity and F1-score, as well as specificity and F1-score, could not be computed due to insufficient valid data points. This further highlights the limitations of using KNN in this context, emphasising the need for more robust evaluation setups and diverse data to ensure reliable performance assessment.

4.9.3. RF

The evaluation of metric correlations for the RF model reveals a mix of perfect correlations and undefined relationships, highlighting a potentially constrained performance profile. Specifically, accuracy showed a perfect correlation with both sensitivity and F1-score ($r = 1.00$, $p = 1.00$, 95% CI = [1.00, 1.00]), with a maximum effect size of 1.00. These results, while seemingly ideal, are likely reflective of deterministic behaviour in the model's predictions rather than a natural relationship between the evaluation metrics. The fixed confidence intervals and non-significant p-values indicate a lack of meaningful variability in the model's outputs across samples, which reduces the interpretability and generalisability of the observed performance. Its worth keeping in mind that the correlation involving specificity with all other metrics could not be computed due to insufficient valid data points.

Overall, these findings suggest that while the RF model may deliver consistently high performance under certain conditions, researchers should ensure diverse and representative data and carefully monitor for metric redundancy to avoid overestimating its practical effectiveness.

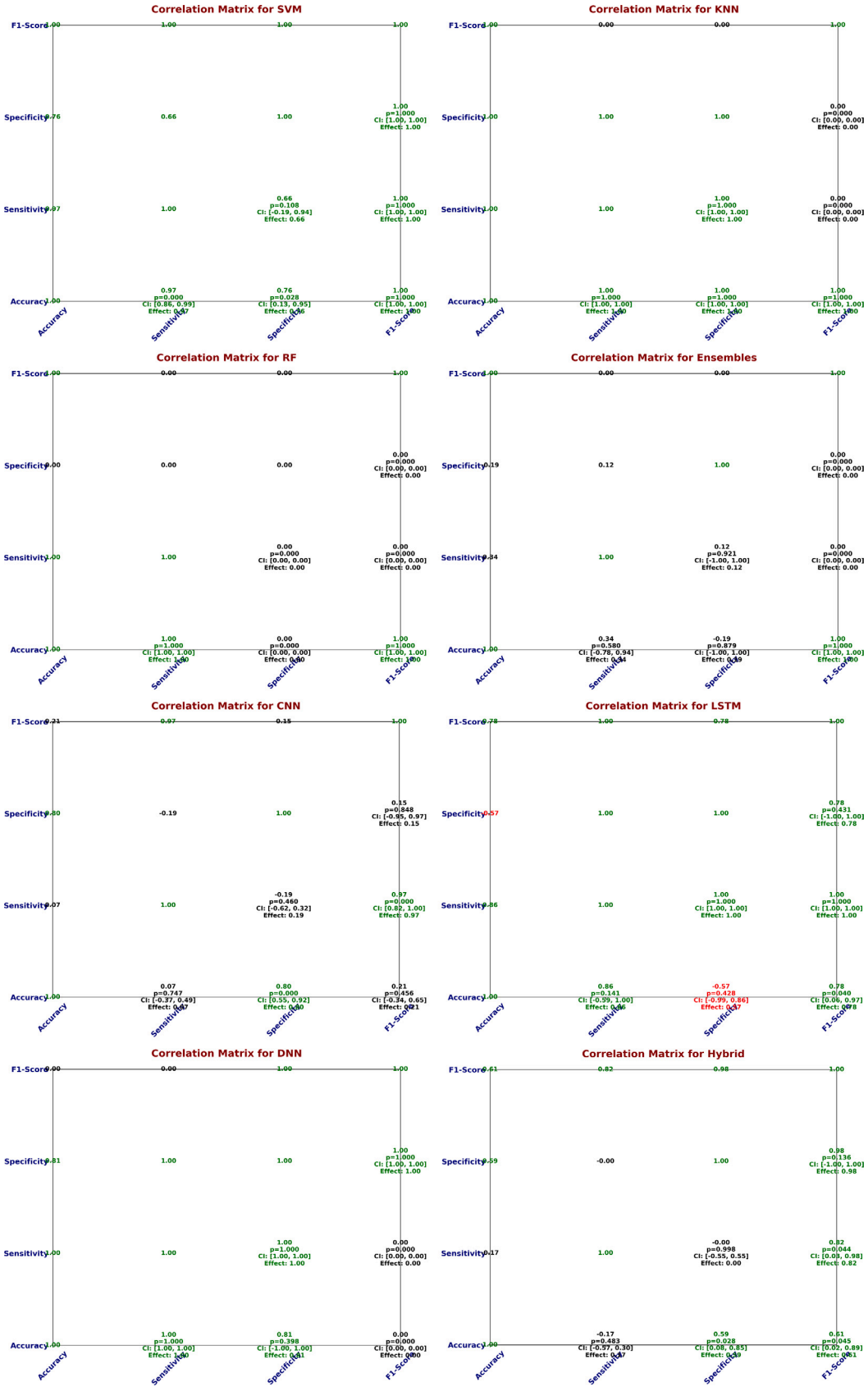


Fig. 10. Pearson correlation heatmaps for various methods.

4.9.4. Ensembles

The correlation analysis for the Ensemble model reveals a distinct pattern compared to more deterministic approaches, characterised by weaker and often non-significant associations among key evaluation metrics. The correlation between accuracy and sensitivity was weak and statistically non-significant (r = 0.34, p = 0.580), with a wide 95% confidence interval ranging from −0.78 to 0.94, indicating considerable uncertainty and inconsistency in how improvements in accuracy translate to better sensitivity. Similarly, accuracy and specificity were negatively correlated (r = −0.19, p = 0.879), again without statistical significance and with an even broader confidence interval of [−1.00, 1.00], suggesting a lack of stable relationship between correctly identifying negative cases and overall classification performance.

Interestingly, accuracy and F1-score showed a perfect correlation (r = 1.00), yet with a *p*-value of 1.00 and a fixed confidence interval [1.00, 1.00], implying possible metric redundancy or a computational artifact rather than genuine performance synergy.

Furthermore, most pairwise relationships involving F1-score, sensitivity, and specificity could not be reliably calculated due to insufficient valid data points. This limitation further complicates the interpretability of the Ensemble model's performance consistency and raises caution about over-reliance on any single metric.

Overall, these findings suggest that while Ensemble methods can offer robust aggregate performance, their internal metric relationships may vary substantially, underscoring the need for comprehensive and representative validation to ensure reliable real-world deployment.

4.9.5. CNN

CNN model exhibited a nuanced and varied correlation profile across evaluation metrics, suggesting differential sensitivity to specific performance dimensions. A strong and statistically significant correlation was observed between accuracy and specificity (r = 0.80, p = 0.000, 95% CI = [0.55, 0.92]), indicating that CNN's overall correctness is closely tied to its ability to accurately identify negative cases. This could imply that the model is especially effective in reducing false positives; an important consideration in medical diagnostics where overtreatment or misclassifications of healthy patients can have significant implications.

Conversely, the relationship between accuracy and sensitivity was significantly weak and non-significant (r = 0.07, p = 0.747), as was the link between accuracy and F1-score (r = 0.21, p = 0.456), suggesting that improvements in accuracy are not reliably associated with enhancements in identifying true positive cases or overall balance between precision and recall. This finding underscores a potential imbalance in model performance, where gains in specificity do not translate proportionally to improvements in sensitivity; a known challenge in CNN-based classification when trained on imbalanced datasets.

A particularly strong correlation was seen between sensitivity and F1-score (r = 0.97, p = 0.000, 95% CI = [0.82, 1.00]), highlighting that F1-score in this model is predominantly driven by sensitivity, possibly at the expense of specificity. This behaviour may reflect the CNN's deeper feature extraction capability favouring true positive predictions, which improves recall but may not consistently preserve balance across all classes.

For researchers, these insights suggest that CNN models in arrhythmia detection require careful metric-specific tuning, particularly when high sensitivity is desired without compromising specificity. Techniques such as cost-sensitive learning, data augmentation for minority classes, and threshold optimisation may help mitigate the observed imbalance. Additionally, reliance on a single metric like accuracy may lead to misleading conclusions, emphasizing the importance of a multi-metric evaluation framework when deploying deep learning models in critical healthcare applications.

4.9.6. LSTM

The LSTM model displayed a robust correlation structure, with some key relationships demonstrating both significant and nuanced dynamics across performance metrics. A strong positive correlation was observed between accuracy and sensitivity (r = 0.86, p = 0.141), though the *p*-value indicates that this relationship did not reach statistical significance. This suggests that, while the model performs well overall, increases in sensitivity may not always guarantee proportional accuracy, hinting at potential challenges in balancing true positives with overall classification performance. A negative correlation was observed between accuracy and specificity (r = −0.57, p = 0.428), reflecting a trade-off where improvements in overall accuracy may be at the expense of correctly identifying negative cases.

Interestingly, accuracy and F1-score exhibited a strong and statistically significant correlation (r = 0.78, p = 0.040), suggesting that the model's ability to maintain a balance between precision and recall is positively linked with overall correctness. However, the relatively modest *p*-value and confidence intervals require caution in interpreting this as a robust and reliable association across all datasets.

The correlation between sensitivity and specificity was perfectly positive (r = 1.00, p = 1.000), indicating that improvements in one metric lead directly to improvements in the other. This behaviour likely points to a model tendency to optimise for both true positives and true negatives simultaneously, which is advantageous in real-world medical applications where both false positives and false negatives have clinical consequences.

Moreover, sensitivity and F1-score were perfectly correlated (r = 1.00, p = 1.000), suggesting that the model's performance in identifying true positives directly governs its overall balance of precision and recall. Finally, the correlation between specificity and F1-score (r = 0.78, p = 0.431) shows a moderate positive relationship, though with a broader confidence interval, suggesting that specificity also plays an important role in achieving balanced classification performance.

For researchers, these results highlight the importance of fine-tuning LSTM models to maintain an appropriate balance between sensitivity and specificity, especially in cases where both types of errors (false positives and false negatives) carry significant consequences. Additionally, while accuracy and F1-score are often used as benchmark metrics, it is critical to consider multi-metric optimisation to ensure that LSTM models can generalise well to diverse and imbalanced clinical datasets.

4.9.7. DNN

The Pearson's correlation analysis for the DNN model revealed a distinctive pattern, marked by several perfect correlations among its core evaluation metrics. Notably, accuracy and sensitivity showed a strong and perfect correlation (r = 1.00, p = 1.000), suggesting that increases in sensitivity are perfectly aligned with overall accuracy in the model's predictions. This indicates that DNN models are highly efficient at correctly identifying both positive and negative cases, leading to excellent overall performance.

Additionally, accuracy and specificity were moderately correlated (r = 0.81, p = 0.398), although the *p*-value suggests that this correlation did not reach statistical significance, hinting at some variability in the relationship. This could imply that while specificity plays a role in improving accuracy, the overall relationship may not always be linear. The perfect correlation between sensitivity and specificity (r = 1.00, p = 1.000) further emphasises the model's tendency to optimise for both true positives and true negatives simultaneously, which could be a significant advantage when working with datasets where both false positives and false negatives have clinical repercussions.

In terms of F1-score, it showed perfect correlation with specificity (r = 1.00, p = 1.000), indicating that the DNN model's balance between precision and recall is directly driven by its ability to accurately identify negative cases.

Overall, these findings suggest that the DNN model can deliver highly consistent performance across metrics, particularly for applications where both false positives and false negatives carry significant

clinical implications. However, the presence of perfect correlations warrants further investigation to ensure that the results stem from genuine model generalisability rather than artefacts of the dataset or evaluation procedure.

4.9.8. Hybrid models

The Hybrid model displayed a mixed correlation structure across its evaluation metrics, with some strong correlations between accuracy, specificity, and F1-score, while sensitivity showed weaker interactions with the other metrics. The accuracy and sensitivity relationship was found to be weak and not statistically significant ($r = -0.17$, $p = 0.483$), suggesting that changes in sensitivity do not consistently affect overall accuracy. This is in contrast to other models where accuracy and sensitivity often exhibited strong positive correlations.

However, the accuracy and specificity relationship exhibited a moderate but statistically significant positive correlation ($r = 0.59$, $p = 0.028$), indicating that higher accuracy in this model is moderately associated with better specificity. This is an important finding for applications that prioritise correctly identifying negatives (true negatives) while maintaining good overall performance. Similarly, accuracy and F1-score also showed a moderate positive correlation ($r = 0.61$, $p = 0.045$), suggesting that improving the balance between precision and recall tends to result in better overall model accuracy.

The relationship between sensitivity and F1-score was stronger ($r = 0.82$, $p = 0.044$), highlighting that higher sensitivity (ability to correctly identify positives) leads to better F1-scores, an important factor in imbalanced datasets where both false positives and false negatives need to be minimized. On the other hand, sensitivity and specificity showed an extremely weak and non-significant correlation ($r = -0.00$, $p = 0.998$), implying that these two metrics do not have a meaningful relationship in this model, which might suggest potential trade-offs between identifying positives and negatives that do not align consistently.

Finally, specificity and F1-score exhibited a very strong correlation ($r = 0.98$, $p = 0.136$), indicating that the model's ability to correctly classify negatives is a key factor in balancing precision and recall. This strong correlation suggests that improving specificity will likely have a significant positive impact on the model's overall F1-score.

For researchers, these findings emphasise the importance of optimising both accuracy and specificity to enhance the overall performance of Hybrid models, especially in scenarios where the detection of both positive and negative cases is critical. The strong relationship between specificity and F1-score suggests that improving specificity will likely improve the model's ability to balance precision and recall, ultimately improving diagnostic accuracy.

4.10. Global Pearson's correlation analysis

Following the presentation of individual Pearson's correlation coefficients for each evaluation metric, we conduct a comprehensive analysis that considers all evaluation methods collectively. This analysis is based on data extracted from all 219 papers included in our study and aims to identify overarching patterns and methodological trends commonly observed in the literature.

The analysis of all models combined reveals (see Fig. 11) significant insights into the relationships between key evaluation metrics. The accuracy metric exhibited a strong positive correlation with specificity ($r = 0.86$, $p = 0.001$), indicating that across all models, higher specificity tends to correlate with better overall accuracy. Similarly, accuracy and F1-score also showed a moderate positive correlation ($r = 0.60$, $p = 0.049$), suggesting that enhancing the balance between precision and recall positively impacts the accuracy of the models.

Sensitivity was found to have a particularly strong positive correlation with F1-score ($r = 0.98$, $p = 0.000$), emphasising that the ability to correctly identify positive cases is highly influential in improving the F1-score. However, the correlation between sensitivity and specificity was moderate ($r = 0.50$, $p = 0.118$), suggesting that while there is some relationship between these two metrics, it is not as strong as between other pairs of metrics. This moderate correlation indicates potential trade-offs between maximising sensitivity and specificity in model optimisation.

Notably, the relationship between accuracy and sensitivity was also moderate ($r = 0.56$, $p = 0.072$), revealing that increasing sensitivity often results in better accuracy, but the relationship is not as robust across all methods. Finally, the specificity and F1-score correlation ($r = 0.56$, $p = 0.070$) further confirms the importance of specificity in achieving a balanced F1-score.

Overall, the findings suggest that maximising specificity and sensitivity simultaneously, while maintaining accuracy, is crucial for enhancing F1-scores across diverse machine learning models. The strong relationship between sensitivity and F1-score highlights the importance of focusing on correctly identifying positive cases to achieve better performance, especially in imbalanced datasets.

4.11. Comparison of global and individual Pearson's correlation

The global correlation matrix serves as an aggregate representation of the relationships between evaluation metrics across all the models tested, offering a holistic view of how different performance metrics interact across a diverse set of approaches. This combined perspective helps us identify general trends and patterns that may not be immediately identified when analysing individual models in isolation.

However, each model has distinct characteristics, learning mechanisms, and strengths that can lead to variations in the correlations between the metrics. For example, some models might exhibit strong correlations between accuracy and sensitivity, while others may show weaker or no such correlations. By comparing individual models with the global correlation, we are able to highlight these differences and demonstrate the variability in how the evaluation metrics behave across different methods. This provides valuable insights into the models' strengths and limitations in the context of the evaluation metrics used in this study.

4.11.1. Global vs. SVM

To better understand the interrelationships between evaluation metrics in arrhythmia detection using AI, we compared correlation patterns from individual models with the global (combined) correlation results across all methods. Notably, the SVM model demonstrated significantly stronger and more consistent correlations among its performance metrics compared to the global trends. For example, accuracy and sensitivity in SVM were highly correlated ($r = 0.97$, $p = 0.000$, 95% CI [0.86, 0.99]), whereas the global correlation for the same pair was notably weaker ($r = 0.56$, $p = 0.072$, 95% CI [−0.06, 0.87]). Moreover, SVM showed perfect alignment between accuracy, F1-score, sensitivity, and specificity ($r = 1.00$), indicating a high level of internal consistency in its performance evaluation. In contrast, the global correlations, while still moderate to strong, displayed greater variability and less statistical significance.

This analysis suggests that evaluating models individually rather than relying solely on global trends can uncover important insights about metric behaviour. Researchers should be cautious when interpreting performance based on a single metric, as models like SVM show that certain metrics (e.g., F1-score) may inherently align with others, making them more reliable indicators of overall performance. Furthermore, models exhibiting stronger internal metric correlations may offer more stable and interpretable behaviour in clinical settings. Therefore, we recommend that future studies not only report standard evaluation metrics but also perform correlation analyses to identify models with tightly coupled and robust metric relationships, particularly when deploying AI systems for critical tasks such as arrhythmia detection.

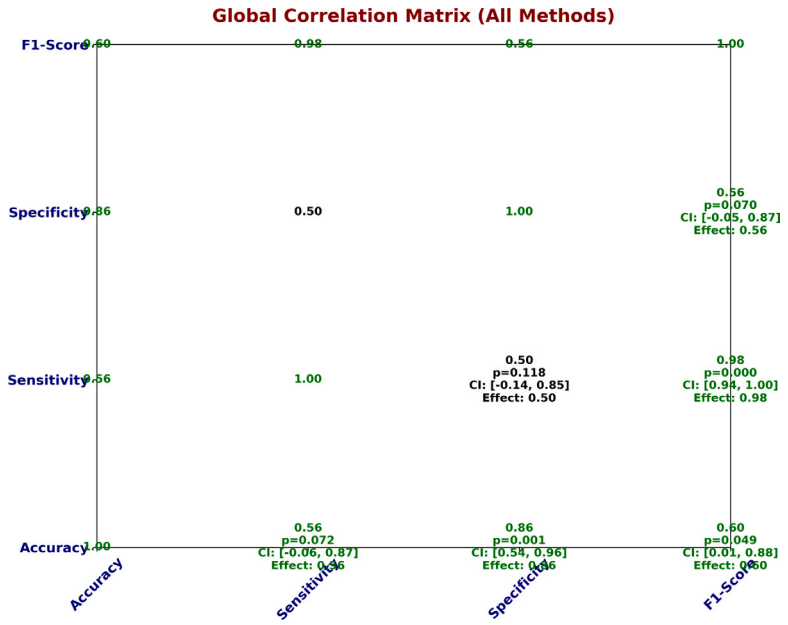


Fig. 11. Global Pearson correlation matrix.

4.11.2. Global vs. KNN

In contrast to global trends, the KNN model exhibited extreme correlations across most evaluation metrics, with Pearson correlation coefficients of 1.00 for accuracy, sensitivity, and specificity (allp = 1.000, 95% CI [1.00, 1.00]). These perfect correlations suggest that, within this model, changes in one metric directly mirror changes in the others, indicating no metric-level variability across test runs or data partitions. However, a striking deviation was observed in the relationship of sensitivity and specificity with F1-score, both showing a complete lack of correlation (r = 0.0), which led to the exclusion of these pairs due to insufficient valid data points.

When compared with the global correlation values, which were generally moderate to strong but varied (e.g., accuracy vs. sensitivity: r = 0.56; sensitivity vs. F1-score: r = 0.98), KNN's results suggest an over-simplified or potentially overfitted model behaviour. Such perfect correlations often indicate a lack of robustness or metric independence, which may reduce a model's reliability in real-world clinical settings where trade-offs between sensitivity and specificity are critical.

Based on this, we recommend that researchers exercise caution when evaluating models that show unusually high inter-metric correlations, such as KNN in this case. These patterns may indicate that the model is not capturing enough variability in the data, leading to inflated performance metrics. Future work should include deeper analyses; such as metric variance, data distribution effects, and external validation; to ensure that models generalise well. Moreover, researchers are encouraged to use correlation analyses not only to interpret model performance but also as a diagnostic tool to detect potential overfitting or metric redundancy in arrhythmia detection systems.

4.11.3. Global vs. RF

The evaluation of the RF model revealed distinct patterns in metric relationships that differ notably from global trends. Within the RF model, accuracy showed a perfect positive correlation with both sensitivity and F1-score (r = 1.00, p = 1.000, 95% CI [1.00, 1.00]), indicating that these metrics moved in complete unison across different evaluations. However, correlations involving specificity were either undefined or statistically invalid due to a lack of variability in specificity scores, not because of missing data. This lack of correlation suggests that specificity remained constant across model runs, offering little contribution to performance differentiation in RF.

When compared with the global correlation matrix; which showed a strong accuracy and specificity relationship (r = 0.86, p = 0.001) and a near-perfect sensitivity vs. F1-score correlation (r = 0.98, p = 0.000); RF appears to capture only a narrow slice of the overall metric dynamics. The absence of measurable relationships between specificity and the other metrics in RF suggests an imbalance in the model's classification behaviour, potentially favouring true positives at the expense of false positive control.

4.11.4. Global vs. Ensembles

The correlation analysis of Ensemble models reveals mixed and somewhat inconsistent relationships between evaluation metrics. accuracy and F1-score show a perfect correlation (r = 1.00, p = 1.000, 95% CI [1.00, 1.00]), suggesting that improvements in accuracy directly and consistently enhance F1-score in these models. However, other metric pairs display weak or even negative correlations. For instance, accuracy and sensitivity are only moderately correlated (r = 0.34, p = 0.580), with a wide confidence interval, while accuracy and specificity are negatively correlated (r = −0.19, p = 0.879), which is in stark contrast to the strong positive global trend (r = 0.86, p = 0.001). Additionally, sensitivity– specificity correlation is negligible (r = 0.12), and other pairs like sensitivity–F1 and specificity–F1 were statistically undefined due to insufficient variability in the results.

These inconsistencies suggest that the performance dynamics in Ensemble models are not as robust or interlinked as seen in other methods like SVM or in the global analysis. While the perfect accuracy–F1 alignment indicates that Ensembles can effectively balance precision and recall, the weak or undefined correlations with sensitivity and specificity raise concerns about the model's ability to consistently detect true positives and avoid false positives or negatives.

For researchers, this implies that while Ensemble models may achieve strong aggregate performance (as reflected in high accuracy and F1-score), they may not always provide balanced behaviour across all key metrics. Especially in medical applications like arrhythmia detection; where both sensitivity (to avoid missing positive cases) and specificity (to reduce false alarms) are critical; this lack of consistent correlation could limit clinical reliability.

4.11.5. Global vs. CNN

The CNN model shows mixed correlations when compared to the global results. While it demonstrates a strong relationship between sensitivity and F1-Score (r = 0.97), similar to the global trend (r = 0.98), it deviates in other areas. The correlation between accuracy and sensitivity is weak (r = 0.07), far lower than the global correlation (r = 0.56), suggesting that improvements in accuracy do not strongly correlate with sensitivity in CNN. The CNN model also shows a strong accuracy– specificity correlation (r = 0.80), which is slightly weaker than the global correlation (r = 0.86), but both indicate that increasing accuracy tends to reduce false positives. However, the correlation between accuracy and F1-Score in CNN is weak (r = 0.21), contrasting with the global dataset's moderate correlation (r = 0.60), highlighting that CNN may struggle with balancing accuracy and F1-Score.

The CNN model also exhibits a weak negative correlation between sensitivity and specificity (r = −0.19), unlike the global dataset, which shows a mild positive correlation (r = 0.50). This suggests a trade-off in CNN, where improving sensitivity may reduce specificity, a pattern not observed in other methods. Additionally, CNN has a minimal and statistically insignificant correlation between specificity and F1-Score (r = 0.15), in contrast to the global correlation (r = 0.56), which shows a moderate positive relationship. Overall, while CNN excels at improving sensitivity and F1-Score, it demonstrates less consistent performance in balancing accuracy, specificity, and F1-Score compared to other models.

Based on the comparison of CNN with the global correlation values, researchers working on arrhythmia detection should consider several key insights when selecting models. First, while CNN shows strong correlations between sensitivity and F1-Score, it struggles with balancing accuracy and specificity. This suggests that CNN may be more suited for tasks where high sensitivity (detecting positive cases) is crucial, but researchers should be mindful of its potential limitations in minimising false positives (specificity).

4.11.6. Global vs. LSTM

When comparing LSTM model's performance with the global correlation matrix, several key insights emerge. LSTM demonstrates a relatively strong correlation between accuracy and sensitivity (r = 0.86), which is higher than the global correlation (r = 0.56), suggesting that LSTM performs better in maintaining a balance between these two metrics. However, LSTM's correlation between accuracy and specificity (r = −0.57) is notably weaker than the global correlation (r = 0.86), indicating that LSTM may not balance accuracy and specificity as effectively as other models. The correlation between accuracy and F1-Score in LSTM (r = 0.78) is also stronger than the global average (r = 0.60), showing that LSTM aligns better with F1-Score compared to the broader trends observed across methods.

Additionally, LSTM has perfect correlations between sensitivity and both specificity (r = 1.00) and F1-Score (r = 1.00), which stands out compared to the global correlations of r = 0.50 for sensitivity–specificity and r = 0.98 for sensitivity–F1-Score. These perfect correlations suggest that LSTM might be highly efficient in balancing sensitivity with specificity and F1-Score, whereas other models in the global matrix show more varied relationships.

Overall, the LSTM model exhibits stronger and more direct correlations between accuracy, sensitivity, and F1-Score, but its relationship with specificity stands out as an area of difference from the global trends. Researchers working on arrhythmia detection may consider LSTM for tasks where a strong relationship between sensitivity and F1-Score is crucial but should be cautious of its potential challenges in balancing accuracy and specificity.

4.11.7. Global vs. DNN

In comparison to the global correlation matrix, the DNN model displays some notable trends. The correlation between accuracy and sensitivity in DNN (r = 1.00) is significantly stronger than the global correlation (r = 0.56), indicating that DNN maintains a perfect relationship between these two metrics, which is uncommon across other models. The DNN's accuracy–specificity correlation (r = 0.81) is also higher than the global correlation (r = 0.86), suggesting that DNN better aligns with specificity than other methods in the global matrix. However, its relationship between accuracy and F1-Score is lacking due to insufficient valid data points, making it difficult to compare directly with the global matrix, where the correlation is r = 0.60.

The perfect correlations in DNN, specifically between sensitivity, specificity, and F1-Score (r = 1.00 for all pairs), contrast with the weaker global correlations, particularly for sensitivity–specificity (r = 0.50) and sensitivity–F1-Score (r = 0.98). This suggests that DNN might offer a very strong balance between these metrics, particularly in the ability to achieve high sensitivity and specificity, which could be valuable in arrhythmia detection tasks that require accurate identification of both true positives and true negatives. However, lack of a strong correlation with F1-Score due to data limitations may pose challenges in scenarios where F1-Score optimisation is critical. Researchers should consider these relationships when selecting models, especially in situations where a balanced performance across multiple metrics is essential.

4.11.8. Global vs. Hybrid models

When comparing the Hybrid model to the global correlation matrix, some distinct trends emerge. The Hybrid model shows a negative correlation between accuracy and sensitivity (r = −0.17), which contrasts sharply with the global positive correlation (r = 0.56). This suggests that the Hybrid model may struggle to simultaneously optimise both accuracy and sensitivity, whereas the global analysis indicates a more balanced relationship between these two metrics. Additionally, the Hybrid model's accuracy–specificity correlation (r = 0.59) is lower than the global correlation (r = 0.86), which highlights that the Hybrid model may not be as effective at balancing accuracy and specificity as other models in the global matrix.

On the other hand, the Hybrid model's correlation between sensitivity and F1-Score (r = 0.82) is significantly higher than the global correlation (r = 0.98), indicating a strong relationship between these two metrics, which is beneficial in applications requiring precise detection of both positive and negative cases, such as arrhythmia detection. The correlation between specificity and F1-Score (r = 0.98) in the Hybrid model is very similar to the global correlation (r = 0.56), suggesting that the Hybrid model might be more sensitive to specificity in terms of improving F1-Score.

These findings suggest that researchers may want to consider using the Hybrid model in cases where sensitivity and F1-Score optimisation are prioritised, while also being mindful of its challenges with accuracy and sensitivity trade-offs.

5. Limitations and research directions

After conducting an exhaustive and in-depth review of contemporary research focused on arrhythmia detection using ECG signals, several significant challenges and research gaps have emerged. These findings highlight critical areas where further investigation and innovation are necessary to advance both the scientific understanding and practical capabilities in arrhythmia detection and classification.

- **Enhancing Generalisation and Validation in ECG Analysis** — Generalisation in the context of ECG analysis, both at the inter-patient and intra-patient levels, is a fundamental challenge. Creating models that can generalise effectively across a diverse population of patients is vital for the real-world deployment of

ECG-based diagnostic and monitoring tools. External validation is crucial to ensure this generalisability. Many existing studies have relied on small, single-source datasets, limiting the ability to assess the model's performance on data from different populations or settings. For example, studies such as [105,208,240,246,265] used datasets that were limited in size or scope, preventing meaningful cross-patient or cross-population validation. Inter-patient generalisation demands the development of models that can adapt and provide accurate predictions for patients with varying demographics, medical histories, and physiological characteristics. Intra-patient generalisation, on the other hand, requires models that can adapt to the dynamic nature of an individual's ECG data, accounting for changes over time due to factors like lifestyle, health conditions, and medication. To address these challenges, future research should prioritise testing models on independent datasets or unseen data, especially datasets that represent diverse patient populations. By utilising external validation techniques, models can be more rigorously tested for their robustness and reliability across different clinical settings. Addressing these generalisation challenges calls for the creation and adoption of large, well-annotated, publicly available datasets that include a wide range of arrhythmias, patient profiles, and acquisition conditions to improve both benchmarking and real-world performance. Striking the right balance between flexibility, specificity, and external validation is crucial to ensure that these models can accommodate both inter-patient variability and intra-patient dynamics while delivering reliable and precise results.

- **Technical Implementation Challenges** — Despite promising results in controlled settings, many models face technical barriers to real-world deployment due to high computational complexity, sensitivity to signal noise, and dependence on manual feature engineering. Papers such as [103,119,179,186], and [143] highlight challenges in adapting current architectures for real-time performance, especially in mobile or wearable environments. Future work should prioritise lightweight, noise-resilient models with minimal preprocessing requirements that can operate efficiently on low-power devices, facilitating broader integration in telehealth and continuous monitoring systems.
- **Methodological Recommendations** — Several studies lack rigorous methodological practices, which undermines the reproducibility and interpretability of results. Papers [193,260,274,285], and [97] demonstrate shortcomings such as missing ablation studies, dataset-specific preprocessing steps, and insufficient transparency in decision-making processes. Researchers should move toward standardised pipelines that include interpretable model components, explainability techniques, and detailed performance breakdowns. In addition, expanding classification beyond simple detection and validating against multiple datasets will foster more robust and clinically meaningful outcomes.
- **Clinical Integration Barriers** — Clinical adoption remains limited due to the absence of interpretability, clinician validation, and real-world testing. Studies like [97,108,152,266], and [194] reveal that many models are not aligned with clinical workflows or lack trustworthiness from a healthcare perspective. To bridge this gap, future research should emphasise clinician collaboration from model design through validation, incorporate domain-informed explanations, and conduct pilot studies in real clinical settings. These steps are crucial to ensure that AI models not only perform well technically but also integrate smoothly into existing healthcare systems and gain user trust.
- **Complex Nature of an ECG Signal** — The interpretation of an ECG signal poses a significant challenge due to its complex nature. This complexity becomes even more when we encounter scenarios where the boundaries or thresholds differentiating distinct classes are too narrow. In such cases, distinguishing between subtle variations in the signal and making accurate classifications can be particularly challenging. This nature of ECG waveform and the fine distinctions between different cardiac conditions require a high level of expertise and precision in analysis. Researchers and healthcare professionals must counter this complexity to ensure accurate diagnoses and effective medical decision-making.
- **Addressing Data Imbalance in ECG Datasets** — Addressing data imbalance is a critical challenge in ECG analysis. In most of the existing ECG databases; the normal class has a lot more samples than the abnormal classes. Consequently, classes with fewer training samples often demonstrate considerable inaccuracies when compared to other classes. Hence there is a need for new datasets, having more samples for the abnormal classes. Researchers have applied various techniques to address this issue. These approaches include data augmentation, leveraging unsupervised learning methods and implementing focal loss for the detection of abnormalities in ECG signals. Nevertheless, the issue remains a challenging one, as the results obtained thus so far are not entirely promising. Therefore, there is a continued need for innovative and effective strategies to rectify the data imbalance problem in ECG analysis, which is vital for the development of clinically applicable solutions.
- **Dealing with Noise in ECG Data** — The presence of noise from various sources in real-world ECG data is another challenge in ECG analysis. This noise may be a result of different sources, for instance baseline wander; which can be caused by any inappropriate physical move from the patient, powerline interference; which could be result of any power line issue, and electrode motion artifact noise; quite like baseline wander. Traditionally, researchers have tackled these challenges by applying preprocessing techniques to remove or reduce noise from the data. While these preprocessing methods effectively enhance data quality, they also come at a cost, namely an increase in computational overhead and a reduction in efficiency. Therefore, there is a growing need to develop more robust models capable of handling noisy ECG data without the need for extensive prior processing. Such models would not only streamline the analysis process but also hold the potential to improve the accuracy and practicality of ECG-based diagnostics and monitoring, particularly in real-world scenarios where data may be inherently noisy and challenging to preprocess effectively.
- **Managing Label Noise** — Moreover, it is essential to consider the issue of label noise. Experts involved in the manual labelling process may inadvertently introduce errors, and this factor needs to be addressed in the analysis.
- **Challenges in Model Interpretability** — One of the prominent challenges in the domain of ML and DL is poor interpretability. As these models grow increasingly complex and powerful, they also become more blurred in their decision-making processes. This lack of transparency poses significant concerns, especially in applications where understanding the rationale behind a model's predictions is crucial, such as in ECG analysis. While these models can achieve remarkable accuracy, their inner workings remain hidden, making it challenging to comprehend how they arrive at specific conclusions. This lack of interpretability can result in lack of trust, accountability, and the ability to detect biases or errors. Hence there is a significant need for research and development focused on enhancing the transparency and interpretability of ML and DL models to ensure their trust worth clinical deployment for ECG analysis.
- **Choosing the Right Model** — ML algorithms offer certain advantages in terms of speed and memory efficiency when compared to their counterparts. However, they often require manual handcrafted feature engineering, where domain-specific knowledge is essential to design the right set of features. In contrast, DL stands out as a paradigm where the model can autonomously learn the

underlying feature representations from the data, eliminating the need for manual feature engineering. To address this challenge, there is a growing need for lightweight variants of DL models that can achieve similar performance with reduced computational demands.

- **Overcoming Complexity in Deep Neural Networks** — In case of deep models, it is worth highlighting that the increased complexity associated with deep neural networks can be a significant hurdle in their practical implementation. This complexity becomes particularly critical in the context of wearable ECG monitoring devices, where efficiency and real-time processing are of great importance.

6. Conclusion

In this review, we conducted a comprehensive and systematic analysis of AI-based techniques for arrhythmia detection using ECG signals.Unlike previous review papers in the same research area, we contributed a performance oriented synthesis of 219 peer-reviewed studies. We extracted and standardised key evaluation metrics; including accuracy, sensitivity, specificity, and F1-score; and presented them in a harmonized format to ensure consistency and comparability.

Beyond metric-wise reporting, we evaluated the correlation between these metrics using Pearson correlation analysis, both at the individual model level (e.g., SVM, CNN, LSTM) and globally across all methods. This statistical exploration, supported by p-values, confidence intervals, and effect size calculations, revealed important interdependencies among metrics, offering practical insights for model evaluation and selection.

To enhance visual interpretability, we introduced forest plots, bar charts and comparative tables to clearly illustrate performance variations across ML, DL, and hybrid approaches. A detailed dataset analysis further highlighted class distributions, demographic attributes, and recording conditions, underscoring the challenges of achieving generalisability in real-world settings.

Importantly, we synthesized specific and actionable future research directions by aligning common limitations; for instance data imbalance, lack of external validation, and limited interpretability; with methodological gaps and barriers to clinical integration. These are detailed in the section "Limitations and Research Directions".

We believe this work serves as both a critical synthesis of the current state-of-the-art and a data-driven guide for future exploration. By combining rigorous literature coverage, analytical depth, and clear visualisation, our review provides meaningful insights to help researchers and practitioners develop robust, interpretable, and clinically viable AI-based arrhythmia detection systems; ultimately contributing to earlier diagnosis, improved patient outcomes, and safer deployment of AI in healthcare.

Glossary

- AI = Artificial Intelligence
- ECG = Electrocardiogram
- ML = Machine learning
- DL = Deep learning
- SVM = Support vector machine
- KNN = K Nearest Neighbour
- RF = Random Forest
- MLP = Multi Layer Perceptron
- RNN = Recurrent Neural Network
- LSTM = Long Short Term Memory
- CNN = Convolutional Neural Network
- GAN = Generative Adversarial Network
- DNN = Deep Neural Network
- CI = confidence interval

CRediT authorship contribution statement

Ahtisham Ayyub: Writing – original draft, Software, Resources, Methodology, Investigation, Data curation, Conceptualisation. **Christos Politis:** Writing – review & editing, Supervision. **Muhammad Arslan Usman:** Writing – review & editing, Visualisation, Supervision.

Ethics statement on the use of generative AI

Generative AI tools were used exclusively for language and grammar correction during the preparation of this manuscript. No AI tools were used for content generation, data interpretation, or analysis. The authors affirm that all intellectual content is their own original work.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Q.A. Rahman, L.G. Tereshchenko, M. Kongkatong, T. Abraham, M.R. Abraham, H. Shatkay, IEEE Trans. Nanobiosci. 14 (5) (2015) 505–512, http://dx.doi.org/10.1109/TNB.2015.2426213.

[2] R.G. Afkhami, G. Azarnia, M.A. Tinati, Pattern Recognit. Lett. 70 (2016) 45–51, http://dx.doi.org/10.1016/j.patrec.2015.11.018.

[3] F. Jiang, K. Chao, J. Xiao, Q. Liu, K. Gu, J. Wu, Y. Cao, Electronics 12 (9) (2023) 2046, http://dx.doi.org/10.3390/electronics12092046.

[4] G. Marcus, 2018, http://dx.doi.org/10.48550/arxiv.1801.00631, arXiv preprint arXiv:1801.00631.

[5] C.J. Burges, Data Min. Knowl. Discov. 2 (2) (1998) 121–167, http://dx.doi.org/10.1023/A:1009715923555.

[6] C.-S. Lo, C.-M. Wang, Comput. Math. Appl. 64 (5) (2012) 1153–1162, http://dx.doi.org/10.1016/j.camwa.2012.03.033.

[7] H. Rai, A. Yadav, Expert Syst. Appl. 41 (2) (2014) 588–593, http://dx.doi.org/10.1016/j.eswa.2013.07.083.

[8] H. Ling, J. Wu, P. Li, International Conference on Artificial Intelligence and Security, Springer, 2019, pp. 442–451, http://dx.doi.org/10.1007/978-3-030-24265-7_38.

[9] X. Wang, W.Q. Yan, Int. J. Neural Syst. 30 (01) (2020) 1950027, http://dx.doi.org/10.1142/s0129065719500278.

[10] Y. Su, S. An, Z. Feng, M. Xing, J. Zhang, J. Vis. Commun. Image Represent. 67 (2020) 102753, http://dx.doi.org/10.1016/j.jvcir.2020.102753.

[11] A. Alberdi, A. Aztiria, A. Basarab, J. Biomed. Inform. 59 (2016) 49–75, http://dx.doi.org/10.1016/j.jbi.2015.11.007.

[12] D.M. Krikler, Cardiol. Clin. 5 (3) (1987) 349–355, http://dx.doi.org/10.1016/s0733-8651(18)30525-3.

[13] A.Y. Hannun, P. Rajpurkar, M. Haghpanahi, G.H. Tison, C. Bourn, M.P. Turakhia, A.Y. Ng, Nature Med. 25 (1) (2019) 65–69, http://dx.doi.org/10.1038/s41591-018-0268-3.

[14] Z.I. Attia, S. Kapa, F. Lopez-Jimenez, P.M. McKie, D.J. Ladewig, G. Satam, P.A. Pellikka, M. Enriquez-Sarano, P.A. Noseworthy, T.M. Munger, et al., Nature Med. 25 (1) (2019) 70–74, http://dx.doi.org/10.1038/s41591-018-0240-2.

[15] Z.I. Attia, P.A. Noseworthy, F. Lopez-Jimenez, S.J. Asirvatham, A.J. Deshmukh, B.J. Gersh, R.E. Carter, X. Yao, A.A. Rabinstein, B.J. Erickson, et al., Lancet 394 (10201) (2019) 861–867, http://dx.doi.org/10.1016/S0140-6736(19)31721-0.

[16] C.D. Galloway, A.V. Valys, J.B. Shreibati, D.L. Treiman, F.L. Petterson, V.P. Gundotra, D.E. Albert, Z.I. Attia, R.E. Carter, S.J. Asirvatham, et al., JAMA Cardiol. 4 (5) (2019) 428–436, http://dx.doi.org/10.1001/jamacardio.2019.0640.

[17] U. Erdenebayar, Y.J. Kim, J.-U. Park, E.Y. Joo, K.-J. Lee, Comput. Methods Programs Biomed. 180 (2019) 105001, http://dx.doi.org/10.1016/j.cmpb.2019.105001.

[18] J.-Y. Sun, Y. Qiu, H.-C. Guo, Y. Hua, B. Shao, Y.-C. Qiao, J. Guo, H.-L. Ding, Z.-Y. Zhang, L.-F. Miao, et al., J. Cardiovasc. Electrophysiol. 32 (4) (2021) 1095–1102, http://dx.doi.org/10.1111/jce.14936.

[19] M.M. Ahsan, R. Nazim, Z. Siddique, P. Huebner, Healthcare, Vol. 9, MDPI, 2021, p. 1099, http://dx.doi.org/10.3390/healthcare9091099.

[20] B. Kitchenham, Keele UK Keele Univ. 33 (2004) (2004) 1–26.

[21] B.E. Boser, I.M. Guyon, V.N. Vapnik, Proceedings of the Fifth Annual Workshop on Computational Learning Theory, 1992, pp. 144–152, http://dx.doi.org/10.1145/130385.130401.

[22] Z. Li, R. Yuan, X. Guan, 2007 IEEE International Conference on Communications, IEEE, 2007, pp. 1373–1378, http://dx.doi.org/10.1109/icc.2007.231.

[23] G. Chirici, M. Mura, D. McInerney, N. Py, E.O. Tomppo, L.T. Waser, D. Travaglini, R.E. McRoberts, Remote Sens. Environ. 176 (2016) 282–294, http://dx.doi.org/10.1016/j.rse.2016.02.001.

[24] D.A. Adeniyi, Z. Wei, Y. Yongquan, Appl. Comput. Inform. 12 (1) (2016) 90–108, http://dx.doi.org/10.1016/j.aci.2014.10.001.

[25] R.E. McRoberts, E. Næsset, T. Gobakken, Remote Sens. Environ. 163 (2015) 13–22, http://dx.doi.org/10.1016/j.rse.2015.02.026.

[26] H. Al-Shehri, A. Al-Qarni, L. Al-Saati, A. Batoaq, H. Badukhen, S. Alrashed, J. Alhiyafi, S.O. Olatunji, 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering, CCECE, IEEE, 2017, pp. 1–4, http://dx.doi.org/10.1109/ccece.2017.7946847.

[27] J. Vitola, F. Pozo, D.A. Tibaduiza, M. Anaya, Sensors 17 (2) (2017) 417, http://dx.doi.org/10.3390/s17020417.

[28] N. Zhang, A. Lin, P. Shang, Phys. A 477 (2017) 161–173, http://dx.doi.org/10.1016/j.physa.2017.02.072.

[29] L. Breiman, Mach. Learn. 45 (2001) 5–32, http://dx.doi.org/10.1023/A:1010933404324.

[30] D.R. Cutler, T.C. Edwards Jr., K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, J.J. Lawler, Ecology 88 (11) (2007) 2783–2792, http://dx.doi.org/10.1890/07-0539.1.

[31] L. Ma, L. Cheng, M. Li, Y. Liu, X. Ma, ISPRS J. Photogramm. Remote Sens. 102 (2015) 14–27, http://dx.doi.org/10.1016/j.isprsjprs.2014.12.026.

[32] J. Magidi, L. Nhamo, S. Mpandeli, T. Mabhaudhi, Remote. Sens. 13 (5) (2021) 876, http://dx.doi.org/10.3390/rs13050876.

[33] N. Arora, P.D. Kaur, Appl. Soft Comput. 86 (2020) 105936, http://dx.doi.org/10.1016/j.asoc.2019.105936.

[34] M.-C. Popescu, V.E. Balas, L. Perescu-Popescu, N. Mastorakis, WSEAS Trans. Circuits Syst. 8 (7) (2009) 579–588.

[35] B. Chaudhuri, U. Bhattacharya, Neurocomputing 34 (1–4) (2000) 11–27, http://dx.doi.org/10.1016/S0925-2312(00)00305-2.

[36] A.A. Alnuaim, M. Zakariah, P.K. Shukla, A. Alhadlaq, W.A. Hatamleh, H. Tarazi, R. Sureshbabu, R. Ratna, et al., J. Heal. Eng. 2022 (2022) http://dx.doi.org/10.1155/2022/6005446.

[37] C.H. Zhao, B.L. Zhang, X.Z. Zhang, S.Q. Zhao, H.X. Li, Neural Comput. Appl. 22 (2013) 175–184, http://dx.doi.org/10.1007/s00521-012-1057-4.

[38] E. Bou Assi, L. Gagliano, S. Rihana, D.K. Nguyen, M. Sawan, Sci. Rep. 8 (1) (2018) 15491, http://dx.doi.org/10.1038/s41598-018-33969-9.

[39] R. Polikar, IEEE Circuits Syst. Mag. 6 (3) (2006) 21–45.

[40] B. Antal, A. Hajdu, Knowl.-Based Syst. 60 (2014) 20–27, http://dx.doi.org/10.1016/j.knosys.2013.12.023.

[41] S.S. Rathore, S. Kumar, Expert Syst. Appl. 82 (2017) 357–382.

[42] A. Verikas, Z. Kalsyte, M. Bacauskiene, A. Gelzinis, Soft Comput. 14 (2010) 995–1010, http://dx.doi.org/10.1007/s00500-009-0490-5.

[43] H. Mei, M. Bansal, M.R. Walter, 2015, http://dx.doi.org/10.1609/aaai.v30i1.10364, arXiv preprint arXiv:1506.04089.

[44] W. Chan, N. Jaitly, Q.V. Le, O. Vinyals, 2015, http://dx.doi.org/10.48550/arXiv.1508.01211, arXiv preprint arXiv:1508.01211.

[45] D. Bahdanau, K. Cho, Y. Bengio, 2014, http://dx.doi.org/10.48550/arXiv.1409.0473, arXiv preprint arXiv:1409.0473.

[46] K.M. Hermann, T. Kocisky, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, P. Blunsom, Adv. Neural Inf. Process. Syst. 28 (2015) 1693–1701, http://dx.doi.org/10.48550/arxiv.1506.03340.

[47] F. Meng, Z. Lu, M. Wang, H. Li, W. Jiang, Q. Liu, 2015, http://dx.doi.org/10.48550/arXiv.1503.01838, arXiv preprint arXiv:1503.01838.

[48] W. Ling, C. Dyer, A.W. Black, I. Trancoso, Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2015, pp. 1299–1304, http://dx.doi.org/10.3115/v1/n15-1142.

[49] W. Ling, Y. Tsvetkov, S. Amir, R. Fermandez, C. Dyer, A.W. Black, I. Trancoso, C.-C. Lin, Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015, pp. 1367–1372, http://dx.doi.org/10.18653/v1/d15-1161.

[50] S.K. Sønderby, C.K. Sønderby, H. Nielsen, O. Winther, International Conference on Algorithms for Computational Biology, Springer, 2015, pp. 68–80, http://dx.doi.org/10.1007/978-3-319-21233-3_6.

[51] S. Hochreiter, J. Schmidhuber, Neural Comput. 9 (8) (1997) 1735–1780, http://dx.doi.org/10.1162/neco.1997.9.8.1735.

[52] A. Zhao, L. Qi, J. Li, J. Dong, H. Yu, Ninth International Conference on Graphic and Image Processing, ICGIP 2017, Vol. 10615, International Society for Optics and Photonics, 2018, p. 106155B, http://dx.doi.org/10.1117/12.2305277.

[53] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Proc. IEEE 86 (11) (1998) 2278–2324, http://dx.doi.org/10.1109/5.726791.

[54] Y.-N. Chen, C.-C. Han, C.-T. Wang, B.-S. Jeng, K.-C. Fan, 18th International Conference on Pattern Recognition, ICPR'06, Vol. 3, IEEE, 2006, pp. 552–555.

[55] Z. Cai, N. Vasconcelos, IEEE Trans. Pattern Anal. Mach. Intell. (2019) http://dx.doi.org/10.1109/TPAMI.2019.2956516.

[56] A. Bulat, G. Tzimiropoulos, European Conference on Computer Vision, Springer, 2016, pp. 717–732, http://dx.doi.org/10.1007/978-3-319-46478-7_44.

[57] D. Ciregan, U. Meier, J. Schmidhuber, 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 3642–3649, http://dx.doi.org/10.1109/cvpr.2012.6248110.

[58] D. CireşAn, U. Meier, J. Masci, J. Schmidhuber, Neural Netw. 32 (2012) 333–338, http://dx.doi.org/10.1016/j.neunet.2012.02.023.

[59] R. Collobert, J. Weston, Proceedings of the 25th International Conference on Machine Learning, 2008, pp. 160–167, http://dx.doi.org/10.1145/1390156.1390177.

[60] S. Frizzi, R. Kaabi, M. Bouchouicha, J.-M. Ginoux, E. Moreau, F. Fnaiech, IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society, IEEE, 2016, pp. 877–882, http://dx.doi.org/10.1109/iecon.2016.7793196.

[61] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, T. Darrell, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2625–2634, http://dx.doi.org/10.21236/ada623249.

[62] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Commun. ACM 63 (11) (2020) 139–144.

[63] A. Brock, J. Donahue, K. Simonyan, 2018, http://dx.doi.org/10.48550/arXiv.1809.11096, arXiv preprint arXiv:1809.11096.

[64] A. Bansal, S. Ma, D. Ramanan, Y. Sheikh, Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 119–135, http://dx.doi.org/10.1007/978-3-030-01228-1_8.

[65] A. Bulat, J. Yang, G. Tzimiropoulos, Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 185–200, http://dx.doi.org/10.1007/978-3-030-01231-1_12.

[66] X. Yu, X. Cai, Z. Ying, T. Li, G. Li, Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part V 14, Springer, 2019, pp. 341–356, http://dx.doi.org/10.1007/978-3-030-20873-8_22.

[67] D.E. Rumelhart, G.E. Hinton, R.J. Williams, et al., 1985, http://dx.doi.org/10.21236/ada164453.

[68] W. Bao, J. Yue, Y. Rao, PloS One 12 (7) (2017) e0180944, http://dx.doi.org/10.1371/journal.pone.0180944.

[69] M. Yousefi-Azar, V. Varadharajan, L. Hamey, U. Tupakula, 2017 International Joint Conference on Neural Networks, IJCNN, IEEE, 2017, pp. 3854–3861, http://dx.doi.org/10.1109/ijcnn.2017.7966342.

[70] H.-C. Shin, M. Orton, D.J. Collins, S. Doran, M.O. Leach, 2011 10th International Conference on Machine Learning and Applications and Workshops, Vol. 1, IEEE, 2011, pp. 259–264, http://dx.doi.org/10.1109/ICMLA.2011.38.

[71] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Adv. Neural Inf. Process. Syst. 30 (2017) http://dx.doi.org/10.48550/arXiv.1706.03762.

[72] J. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, Y. Bengio, 2015, http://dx.doi.org/10.48550/arXiv.1506.07503, arXiv preprint arXiv:1506.07503.

[73] L. Li, Y. Liu, A. Zhou, Proceedings of the 22nd Conference on Computational Natural Language Learning, 2018, pp. 181–189, http://dx.doi.org/10.18653/v1/k18-1018.

[74] D. Nie, Y. Gao, L. Wang, D. Shen, Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV 11, Springer, 2018, pp. 370–378, http://dx.doi.org/10.1007/978-3-030-00937-3_43.

[75] Y. Bin, Y. Yang, F. Shen, N. Xie, H.T. Shen, X. Li, IEEE Trans. Cybern. 49 (7) (2018) 2631–2641, http://dx.doi.org/10.1109/TCYB.2018.2831447.

[76] J. Li, K. Jin, D. Zhou, N. Kubota, Z. Ju, Neurocomputing 411 (2020) 340–350, http://dx.doi.org/10.1016/j.neucom.2020.06.014.

[77] S. Bozinovski, A. Fulgosi, Proceedings of Symposium Informatica, Vol. 3, 1976, pp. 121–126.

[78] C. Cai, S. Wang, Y. Xu, W. Zhang, K. Tang, Q. Ouyang, L. Lai, J. Pei, J. Med. Chem. 63 (16) (2020) 8683–8694, http://dx.doi.org/10.1021/acs.jmedchem.9b02147.s002.

[79] G. Pinto, Z. Wang, A. Roy, T. Hong, A. Capozzoli, Adv. Appl. Energy 5 (2022) 100084, http://dx.doi.org/10.1016/j.adapen.2022.100084.

[80] L. Shao, F. Zhu, X. Li, IEEE Trans. Neural Netw. Learn. Syst. 26 (5) (2014) 1019–1034, http://dx.doi.org/10.1109/TNNLS.2014.2330900.

[81] M. Raghu, C. Zhang, J. Kleinberg, S. Bengio, 2019, http://dx.doi.org/10.48550/arXiv.1902.07208, arXiv preprint arXiv:1902.07208.

[82] Q. Li, C. Rajagopalan, G.D. Clifford, Comput. Methods Programs Biomed. 117 (3) (2014) 435–447, http://dx.doi.org/10.1016/j.cmpb.2014.09.002.

[83] F. Alonso-Atienza, E. Morgado, L. Fernández-Martínez, A. García-Alberola, J.L. Rojo-Álvarez, IEEE Trans. Bio- Med. Eng. 61 (3) (2014) 832–840, http://dx.doi.org/10.1109/TBME.2013.2290800.

[84] Z. Zhang, J. Dong, X. Luo, K.-S. Choi, X. Wu, Comput. Biol. Med. 46 (2014) 79–89, http://dx.doi.org/10.1016/j.compbiomed.2013.11.019.

[85] A.F. Khalaf, M.I. Owis, I.A. Yassine, Expert Syst. Appl. 42 (21) (2015) 8361–8368, http://dx.doi.org/10.1016/j.eswa.2015.06.046.

[86] S. Raj, K.C. Ray, O. Shankar, Comput. Methods Programs Biomed. 136 (2016) 163–177, http://dx.doi.org/10.1016/j.cmpb.2016.08.016.

[87] K.N.V.P.S. Rajesh, R. Dhuli, Comput. Biol. Med. 87 (2017) 271–284, http://dx.doi.org/10.1016/j.compbiomed.2017.06.006.

[88] Q. Qin, J. Li, L. Zhang, Y. Yue, C. Liu, Sci. Rep. 7 (1) (2017) 6067, http://dx.doi.org/10.1038/s41598-017-06596-z.

[89] S. Raj, K.C. Ray, IEEE Trans. Instrum. Meas. 66 (3) (2017) 470–478, http://dx.doi.org/10.1109/TIM.2016.2642758.

[90] S. Sahoo, B. Kanungo, S. Behera, S. Sabut, Measurement 108 (2017) 55–66, http://dx.doi.org/10.1016/j.measurement.2017.05.022.

[91] Z. Chen, J. Luo, K. Lin, J. Wu, T. Zhu, X. Xiang, J. Meng, IEEE Trans. Circuits Syst. II: Express Briefs 65 (7) (2018) 948–952, http://dx.doi.org/10.1109/TCSII.2017.2747596.

[92] V.H.C. de Albuquerque, T.M. Nunes, D.R. Pereira, E.J.S. Da Luz, D. Menotti, J.P. Papa, J.M.R.S. Tavares, Neural Comput. Appl. 29 (3) (2018) 679–693, http://dx.doi.org/10.1007/s00521-016-2472-8.

[93] W. Zhu, X. Chen, Y. Wang, L. Wang, IEEE/ ACM Trans. Comput. Biol. Bioinform. 16 (1) (2019) 131–138, http://dx.doi.org/10.1109/TCBB.2018.2846611.

[94] J. Yang, R. Yan, IEEE Sens. J. 21 (13) (2021) 14180–14190, http://dx.doi.org/10.1109/JSEN.2020.3047962.

[95] E.H. Houssein, I.E. Ibrahim, N. Neggaz, M. Hassaballah, Y.M. Wazery, Expert Syst. Appl. 181 (2021) 115131, http://dx.doi.org/10.1016/j.eswa.2021.115131.

[96] V. Bhagyalakshmi, R.V. Pujeri, G.D. Devanagavi, J. King Saud Univ. - Comput. Inf. Sci. 33 (1) (2021) 54–67, http://dx.doi.org/10.1016/j.jksuci.2018.02.005.

[97] Z. Ozpolat, M. Karabatak, Diagn. (Basel, Switzerland) 13 (6) (2023) http://dx.doi.org/10.3390/diagnostics13061099.

[98] S.M. Qaisar, A. Mihoub, M. Krichen, H. Nisar, Sens. (Basel, Switzerland) 21 (4) (2021) http://dx.doi.org/10.3390/s21041511.

[99] F. Bouaziz, D. Boutana, H. Oulhadj, 2018 International Conference on Applied Smart Systems, ICASS, IEEE, 2018, pp. 1–6, http://dx.doi.org/10.1109/ICASS.2018.8652020.

[100] P. Pławiak, Expert Syst. Appl. 92 (2018) 334–349, http://dx.doi.org/10.1016/j.eswa.2017.09.022.

[101] T. Tuncer, S. Dogan, P. Pławiak, U. Rajendra Acharya, Knowl.-Based Syst. 186 (2019) 104923, http://dx.doi.org/10.1016/j.knosys.2019.104923.

[102] M.A. Kobat, O. Karaca, P.D. Barua, S. Dogan, Symmetry 13 (10) (2021) 1914, http://dx.doi.org/10.3390/sym13101914.

[103] M. Sraitih, Y. Jabrane, A. Hajjam El Hassani, J. Clin. Med. 11 (17) (2022) http://dx.doi.org/10.3390/jcm11174935.

[104] Q.-U.-A. Mastoi, T.Y. Wah, M.A. Mohammed, U. Iqbal, S. Kadry, A. Majumdar, O. Thinnukool, Life (Basel, Switzerland) 12 (6) (2022) http://dx.doi.org/10.3390/life12060842.

[105] X. Dong, W. Si, IEEE Access (2023) http://dx.doi.org/10.1109/access.2023.3305473.

[106] E. Alickovic, A. Subasi, J. Med. Syst. 40 (4) (2016) 108, http://dx.doi.org/10.1007/s10916-016-0467-8.

[107] J. Park, M. Kang, J. Gao, Y. Kim, K. Kang, J. Med. Syst. 41 (1) (2017) 11, http://dx.doi.org/10.1007/s10916-016-0660-9.

[108] F.I. Alarsan, M. Younes, J. Big Data 6 (1) (2019) http://dx.doi.org/10.1186/s40537-019-0244-x.

[109] B.-H. Kung, P.-Y. Hu, C.-C. Huang, C.-C. Lee, C.-Y. Yao, C.-H. Kuan, IEEE J. Biomed. Heal. Inform. 25 (6) (2021) 1904–1914, http://dx.doi.org/10.1109/JBHI.2020.3035191.

[110] J. Rahul, M. Sora, L.D. Sharma, V.K. Bohat, Biocybern. Biomed. Eng. 41 (2) (2021) 656–666, http://dx.doi.org/10.1016/j.bbe.2021.04.004.

[111] P. Yang, D. Wang, W.-B. Zhao, L.-H. Fu, J.-L. Du, H. Su, Biomed. Signal Process. Control. 63 (2021) 102138, http://dx.doi.org/10.1016/j.bspc.2020.102138.

[112] Y. Zhang, Z. Ma, J. Song, X. Kong, Z. Guo, B. Jiang, Methods (San Diego, Calif.) 202 (2022) 144–151, http://dx.doi.org/10.1016/j.ymeth.2021.04.006.

[113] S. Matin Malakouti, Biomed. Signal Process. Control. 84 (2023) 104796, http://dx.doi.org/10.1016/j.bspc.2023.104796.

[114] J.P. Allam, S.P. Sahoo, S. Ari, Biomed. Signal Process. Control. 92 (2024) 106097, http://dx.doi.org/10.1016/j.bspc.2024.106097.

[115] R.J. Martis, U.R. Acharya, C.M. Lim, J.S. Suri, Knowl.-Based Syst. 45 (2013) 76–82, http://dx.doi.org/10.1016/j.knosys.2013.02.007.

[116] E.J.S. Da Luz, T.M. Nunes, V.H.C. de Albuquerque, J.P. Papa, D. Menotti, Expert Syst. Appl. 40 (9) (2013) 3561–3573, http://dx.doi.org/10.1016/j.eswa.2012.12.063.

[117] R.J. Martis, U.R. Acharya, H. Adeli, H. Prasad, J.H. Tan, K.C. Chua, C.L. Too, S.W.J. Yeo, L. Tong, Biomed. Signal Process. Control. 13 (2014) 295–305, http://dx.doi.org/10.1016/j.bspc.2014.04.001.

[118] H. Li, D. Yuan, X. Ma, D. Cui, L. Cao, Sci. Rep. 7 (2017) 41011, http://dx.doi.org/10.1038/srep41011.

[119] M. Wess, P.D. Sai Manoj, A. Jantsch, 2017 IEEE International Symposium on Circuits and Systems, ISCAS, IEEE, 2017, pp. 1–4, http://dx.doi.org/10.1109/ISCAS.2017.8050805.

[120] M. Hammad, A. Maher, K. Wang, F. Jiang, M. Amrani, Measurement 125 (2018) 634–644, http://dx.doi.org/10.1016/j.measurement.2018.05.033.

[121] S. Celin, K. Vasanth, J. Med. Syst. 42 (12) (2018) 241, http://dx.doi.org/10.1007/s10916-018-1083-6.

[122] L.B. Marinho, N.d.M. Nascimento, J.W.M. Souza, M.V. Gurgel, P.P. Rebouças Filho, V.H.C. de Albuquerque, Future Gener. Comput. Syst. 97 (2019) 564–577, http://dx.doi.org/10.1016/j.future.2019.03.025.

[123] H. Yang, Z. Wei, IEEE Access 8 (2020) 47103–47117, http://dx.doi.org/10.1109/ACCESS.2020.2979256.

[124] S. Murawwat, H.M. Asif, S. Ijaz, M. Imran Malik, K. Raahemifar, Alex. Eng. J. 61 (4) (2022) 2807–2823, http://dx.doi.org/10.1016/j.aej.2021.08.014.

[125] N. Feng, S. Xu, Y. Liang, K. Liu, IEEE Access 7 (2019) 50431–50439, http://dx.doi.org/10.1109/ACCESS.2019.2910880.

[126] J.-S. Wang, W.-C. Chiang, Y.-L. Hsu, Y.-T.C. Yang, Neurocomputing 116 (2013) 38–45, http://dx.doi.org/10.1016/j.neucom.2011.10.045.

[127] R.J. Martis, U.R. Acharya, L.C. Min, Biomed. Signal Process. Control. 8 (5) (2013) 437–448, http://dx.doi.org/10.1016/j.bspc.2013.01.005.

[128] S. Jadhav, S. Nalbalwar, A. Ghatol, Soft Comput. 18 (3) (2014) 579–587, http://dx.doi.org/10.1007/s00500-013-1079-6.

[129] K.N. Rajesh, R. Dhuli, Biomed. Signal Process. Control. 41 (2018) 242–254, http://dx.doi.org/10.1016/j.bspc.2017.12.004.

[130] M. Alfaras, M.C. Soriano, S. Ortín, Front. Phys. 7 (2019) http://dx.doi.org/10.3389/fphy.2019.00103.

[131] V. Mondéjar-Guerra, J. Novo, J. Rouco, M.G. Penedo, M. Ortega, Biomed. Signal Process. Control. 47 (2019) 41–48, http://dx.doi.org/10.1016/j.bspc.2018.08.007.

[132] Y. Li, Z. He, H. Wang, B. Li, F. Li, Y. Gao, X. Ye, Biomed. Signal Process. Control. 62 (2020) 102091, http://dx.doi.org/10.1016/j.bspc.2020.102091.

[133] Z. Sun, C. Wang, Y. Zhao, C. Yan, IEEE Access 8 (2020) 117986–117996, http://dx.doi.org/10.1109/ACCESS.2020.3004908.

[134] P. Pławiak, U.R. Acharya, Neural Comput. Appl. 32 (15) (2020) 11137–11161, http://dx.doi.org/10.1007/s00521-018-03980-2.

[135] M. Sraitih, Y. Jabrane, A. Hajjam El Hassani, J. Clin. Med. 10 (22) (2021) http://dx.doi.org/10.3390/jcm10225450.

[136] T. Yoon, D. Kang, J. Pers. Med. 13 (2) (2023) http://dx.doi.org/10.3390/jpm13020373.

[137] Y. Xu, L. Liu, S. Zhang, W. Xiao, Soft Comput. (2023) http://dx.doi.org/10.1007/s00500-023-07861-2.

[138] S. Mandala, A. Rizal, Adiwijaya, S. Nurmaini, S. Suci Amini, G. Almayda Sudarisman, Y. Wen Hau, A. Hanan Abdullah, Plos One 19 (4) (2024) e0297551, http://dx.doi.org/10.1371/journal.pone.0297551.

[139] L. Guo, G. Sim, B. Matuszewski, Biocybern. Biomed. Eng. 39 (3) (2019) 868–879, http://dx.doi.org/10.1016/j.bbe.2019.06.001.

[140] R. Pashikanti, C. Patil, S.A. Anirudhe, Biomed. Signal Process. Control. 94 (2024) 106328, http://dx.doi.org/10.1016/j.bspc.2024.106328.

[141] R. Sarankumar, M. Ramkumar, K. Vijaipriya, R. Velselvi, Knowl.-Based Syst. 294 (2024) 111696, http://dx.doi.org/10.1016/j.knosys.2024.111696.

[142] S. Chauhan, L. Vig, 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA, IEEE, 2015, pp. 1–7, http://dx.doi.org/10.1109/DSAA.2015.7344872.

[143] C. Zhang, G. Wang, J. Zhao, P. Gao, J. Lin, H. Yang, 2017 13th IASTED International Conference on Biomedical Engineering, BioMed, IEEE, 2017, pp. 63–67, http://dx.doi.org/10.2316/p.2017.852-029.

[144] Ö. Yildirim, Comput. Biol. Med. 96 (2018) 189–202, http://dx.doi.org/10.1016/j.compbiomed.2018.03.016.

[145] J. Gao, H. Zhang, P. Lu, Z. Wang, J. Heal. Eng. 2019 (2019) 6320651, http://dx.doi.org/10.1155/2019/6320651.

[146] S. Chauhan, L. Vig, S. Ahmad, Comput. Biol. Med. 109 (2019) 14–21, http://dx.doi.org/10.1016/j.compbiomed.2019.04.009.

[147] B. Ganguly, A. Ghosal, A. Das, D. Das, D. Chatterjee, D. Rakshit, IEEE Sens. Lett. 4 (8) (2020) 1–4, http://dx.doi.org/10.1109/LSENS.2020.3006756.

[148] A. Sharma, N. Garg, S. Patidar, R. San Tan, U.R. Acharya, Comput. Biol. Med. 120 (2020) 103753, http://dx.doi.org/10.1016/j.compbiomed.2020.103753.

[149] M. Ashfaq Khan, Y. Kim, Comput. Mater. Contin. 67 (1) (2021) 427–443, http://dx.doi.org/10.32604/cmc.2021.014682.

[150] Y. Kaya, F. Kuncan, R. Tekin, Arab. J. Sci. Eng. 47 (8) (2022) 10497–10513, http://dx.doi.org/10.1007/s13369-022-06617-8.

[151] M. Yang, W. Liu, H. Zhang, Front. Physiol. 13 (2022) 982537, http://dx.doi.org/10.3389/fphys.2022.982537.

[152] M. Karri, C.S.R. Annavarapu, Expert Syst. Appl. 214 (2023) 119221, http://dx.doi.org/10.1016/j.eswa.2022.119221.

[153] S. Boda, M. Mahadevappa, P. Kumar Dutta, Biomed. Signal Process. Control. 84 (2023) 104756, http://dx.doi.org/10.1016/j.bspc.2023.104756.

[154] X. Yang, H. Yang, M. Dou, Signal Image Video Process. (2024) 1–12, http://dx.doi.org/10.1007/s11760-024-03057-9.

[155] S. Chopannejad, A. Roshanpoor, F. Sadoughi, Digit. Heal. 10 (2024) 20552076241234624, http://dx.doi.org/10.1177/20552076241234624.

[156] Y. Jin, Z. Li, M. Wang, J. Liu, Y. Tian, Y. Liu, X. Wei, L. Zhao, C. Liu, Commun. Med. 4 (1) (2024) 31, http://dx.doi.org/10.1038/s43856-024-00464-4.

[157] S. Din, M. Qaraqe, O. Mourad, K. Qaraqe, E. Serpedin, Artif. Intell. Med. 150 (2024) 102818, http://dx.doi.org/10.1016/j.artmed.2024.102818.

[158] S. Kiranyaz, T. Ince, M. Gabbouj, IEEE Trans. Bio- Med. Eng. 63 (3) (2016) 664–675, http://dx.doi.org/10.1109/TBME.2015.2468589.

[159] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, M. Adam, A. Gertych, R.S. Tan, Comput. Biol. Med. 89 (2017) 389–396, http://dx.doi.org/10.1016/j.compbiomed.2017.08.022.

[160] U.R. Acharya, H. Fujita, O.S. Lih, Y. Hagiwara, J.H. Tan, M. Adam, Inform. Sci. 405 (2017) 81–90, http://dx.doi.org/10.1016/j.ins.2017.04.012.

[161] P. Rajpurkar, A.Y. Hannun, M. Haghpanahi, C. Bourn, A.Y. Ng, 2017, http://dx.doi.org/10.48550/arXiv.1707.01836, arXiv preprint arXiv:1707.01836.

[162] Ö. Yıldırım, P. Pławiak, R.-S. Tan, U.R. Acharya, Comput. Biol. Med. 102 (2018) 411–420, http://dx.doi.org/10.1016/j.compbiomed.2018.09.009.

[163] X. Zhai, C. Tin, IEEE Access 6 (2018) 27465–27472, http://dx.doi.org/10.1109/ACCESS.2018.2833841.

[164] T.J. Jun, H.M. Nguyen, D. Kang, D. Kim, D. Kim, Y.-H. Kim, 2018, http://dx.doi.org/10.48550/arXiv.1804.06812, arXiv preprint arXiv:1804.06812.

[165] Z. Golrizkhatami, A. Acan, Expert Syst. Appl. 114 (2018) 54–64, http://dx.doi.org/10.1016/j.eswa.2018.07.030.

[166] Z. Xiong, M.P. Nash, E. Cheng, V.V. Fedorov, M.K. Stiles, J. Zhao, Physiol. Meas. 39 (9) (2018) 094006, http://dx.doi.org/10.1088/1361-6579/aad9ed.

[167] Y. Li, Y. Pang, J. Wang, X. Li, Neurocomputing 314 (2018) 336–346, http://dx.doi.org/10.1016/j.neucom.2018.06.068.

[168] A. Sellami, H. Hwang, Expert Syst. Appl. 122 (2019) 75–84, http://dx.doi.org/10.1016/j.eswa.2018.12.037.

[169] E. Izci, M.A. Ozdemir, M. Degirmenci, A. Akan, 2019 Medical Technologies Congress, TIPTEKNO, IEEE, 2019, pp. 1–4, http://dx.doi.org/10.1109/TIPTEKNO.2019.8895011.

[170] H. Fujita, D. Cimr, Appl. Intell. 49 (9) (2019) 3383–3391, http://dx.doi.org/10.1007/s10489-019-01461-0.

[171] J. Huang, B. Chen, B. Yao, W. He, IEEE Access 7 (2019) 92871–92880, http://dx.doi.org/10.1109/ACCESS.2019.2928017.

[172] W. Sun, N. Zeng, Y. He, IEEE Access 7 (2019) 67123–67129, http://dx.doi.org/10.1109/ACCESS.2019.2918361.

[173] K.S. Rajput, S. Wibowo, C. Hao, M. Majmudar, 2019, http://dx.doi.org/10.48550/arXiv.1904.00138, arXiv preprint arXiv:1904.00138.

[174] M. Wu, Y. Lu, W. Yang, S.Y. Wong, Front. Comput. Neurosci. 14 (2020) 564015, http://dx.doi.org/10.3389/fncom.2020.564015.

[175] J.-S. Huang, B.-Q. Chen, N.-Y. Zeng, X.-C. Cao, Y. Li, J. Ambient. Intell. Humaniz. Comput. (2020) http://dx.doi.org/10.1007/s12652-020-02110-y.

[176] A. Dutta, T. Batabyal, M. Basu, S.T. Acton, Expert Syst. Appl. 159 (2020) 113408, http://dx.doi.org/10.1016/j.eswa.2020.113408.

[177] D. Wang, Q. Meng, D. Chen, H. Zhang, L. Xu, Sensors (Basel, Switzerland) 20 (6) (2020) http://dx.doi.org/10.3390/s20061579.

[178] A. Ullah, S.M. Anwar, M. Bilal, R.M. Mehmood, Remote. Sens. 12 (10) (2020) 1685, http://dx.doi.org/10.3390/rs12101685.

[179] M.-L. Huang, Y.-S. Wu, Biomed. Eng. Lett. 10 (2) (2020) 183–193, http://dx.doi.org/10.1007/s13534-020-00146-9.

[180] T. Mahmud, S.A. Fattah, M. Saquib, IEEE Access 8 (2020) 104788–104800, http://dx.doi.org/10.1109/ACCESS.2020.2998788.

[181] T.-M. Chen, C.-H. Huang, E.S.C. Shih, Y.-F. Hu, M.-J. Hwang, IScience 23 (3) (2020) 100886, http://dx.doi.org/10.1016/j.isci.2020.100886.

[182] X. Xu, H. Liu, IEEE Access 8 (2020) 8614–8619, http://dx.doi.org/10.1109/ACCESS.2020.2964749.

[183] T.F. Romdhane, H. AlHichri, R. Ouni, M. Atri, Comput. Biol. Med. 123 (2020) 103866, http://dx.doi.org/10.1016/j.compbiomed.2020.103866.

[184] Z. Dokur, T. Ölmez, Neural Comput. Appl. 32 (16) (2020) 12515–12534, http://dx.doi.org/10.1007/s00521-020-04709-w.

[185] S.S. Xu, M.-W. Mak, C.-C. Cheung, IEEE J. Biomed. Heal. Inform. 24 (3) (2020) 717–727, http://dx.doi.org/10.1109/JBHI.2019.2919732.

[186] J. Niu, Y. Tang, Z. Sun, W. Zhang, IEEE J. Biomed. Heal. Inform. 24 (5) (2020) 1321–1332, http://dx.doi.org/10.1109/JBHI.2019.2942938.

[187] X. Zhai, Z. Zhou, C. Tin, Expert Syst. Appl. 158 (2020) 113411, http://dx.doi.org/10.1016/j.eswa.2020.113411.

[188] A. Ullah, S.U. Rehman, S. Tu, R.M. Mehmood, Fawad, M. Ehatisham-Ul-Haq, Sensors (Basel, Switzerland) 21 (3) (2021) http://dx.doi.org/10.3390/s21030951.

[189] Y. Lu, M. Jiang, L. Wei, J. Zhang, Z. Wang, B. Wei, L. Xia, Biomed. Signal Process. Control. 69 (2021) 102843.

[190] T. Wang, C. Lu, Y. Sun, M. Yang, C. Liu, C. Ou, Entropy (Basel, Switzerland) 23 (1) (2021) http://dx.doi.org/10.3390/e23010119.

[191] B.M. Mathunjwa, Y.-T. Lin, C.-H. Lin, M.F. Abbod, J.-S. Shieh, Biomed. Signal Process. Control. 64 (2021) 102262, http://dx.doi.org/10.1016/j.bspc.2020.102262.

[192] S. Sager, F. Bernhardt, F. Kehrle, M. Merkert, A. Potschka, B. Meder, H. Katus, E. Scholz, PloS One 16 (12) (2021) e0261571, http://dx.doi.org/10.1371/journal.pone.0261571.

[193] G. Wang, M. Chen, Z. Ding, J. Li, H. Yang, P. Zhang, Neurocomputing 454 (2021) 339–349, http://dx.doi.org/10.1016/j.neucom.2021.04.104.

[194] G. Bortolan, I. Christov, I. Simova, Diagn. (Basel, Switzerland) 11 (9) (2021) http://dx.doi.org/10.3390/diagnostics11091678.

[195] H. Zhang, C. Liu, Z. Zhang, Y. Xing, X. Liu, R. Dong, Y. He, L. Xia, F. Liu, Front. Physiol. 12 (2021) 648950, http://dx.doi.org/10.3389/fphys.2021.648950.

[196] R. Hu, J. Chen, L. Zhou, Comput. Biol. Med. 144 (2022) 105325, http://dx.doi.org/10.1016/j.compbiomed.2022.105325.

[197] A.S. Eltrass, M.B. Tayel, A.I. Ammar, Neural Comput. Appl. 34 (11) (2022) 8755–8775, http://dx.doi.org/10.1007/s00521-022-06889-z.

[198] M. Zubair, C. Yoon, Sensors (Basel, Switzerland) 22 (11) (2022) http://dx.doi.org/10.3390/s22114075.

[199] A.M. Alqudah, A. Alqudah, Soft Comput. 26 (3) (2022) 1123–1139, http://dx.doi.org/10.1007/s00500-021-06555-x.

[200] M. Hammad, S. Meshoul, P. Dziwiński, P. Pławiak, I.A. Elgendy, Sensors (Basel, Switzerland) 22 (23) (2022) http://dx.doi.org/10.3390/s22239347.

[201] S. Nakatani, K. Yamamoto, T. Ohtsuki, Bioeng. (Basel, Switzerland) 10 (1) (2022) http://dx.doi.org/10.3390/bioengineering10010048.

[202] Y. Liu, Q. Li, R. He, K. Wang, J. Liu, Y. Yuan, Y. Xia, H. Zhang, Front. Physiol. 13 (2022) 850951, http://dx.doi.org/10.3389/fphys.2022.850951.

[203] Y. Li, J.-h. Luo, Q.-y. Dai, J.K. Eshraghian, B.W.-K. Ling, C.-y. Zheng, X.-l. Wang, Biomed. Signal Process. Control. 79 (2023) 104188, http://dx.doi.org/10.1016/j.bspc.2022.104188.

[204] M. Zhang, H. Jin, B. Zheng, W. Luo, Entropy 25 (9) (2023) 1264, http://dx.doi.org/10.3390/e25091264.

[205] A.A. Ahmed, W. Ali, T.A.A. Abdullah, S.J. Malebary, Mathematics 11 (3) (2023) 562, http://dx.doi.org/10.3390/math11030562.

[206] S. Kumar, A. Mallik, A. Kumar, J. Del Ser, G. Yang, Comput. Biol. Med. 153 (2023) 106511, http://dx.doi.org/10.1016/j.compbiomed.2022.106511.

[207] X. He, W. Shan, R. Zhang, A.A. Heidari, H. Chen, Y. Zhang, Biomimetics 8 (3) (2023) 268, http://dx.doi.org/10.3390/biomimetics8030268.

[208] Z. Ma, J. Wang, J. Yue, Y. Lin, Comput. Methods Programs Biomed. (2023) 107740, http://dx.doi.org/10.1016/j.cmpb.2023.107740.

[209] M.M. Eduardo Vasconcellos, B.G. Ferreira, J.S. Leandro, B.F.S. Neto, F.R. Cordeiro, I.A. Cestari, M.A. Gutierrez, A. Sobrinho, T.D. Cordeiro, IEEE Access 11 (2023) 5365–5376, http://dx.doi.org/10.1109/ACCESS.2023.3236189.

[210] H.K. Kim, M.H. Sunwoo, IEEE Access (2024) http://dx.doi.org/10.1109/access.2024.3380892.

[211] J. Kwak, J. Jung, PeerJ Comput. Sci. 10 (2024) e2299, http://dx.doi.org/10.7717/peerj-cs.2299.

[212] H.A. Shah, F. Saeed, M. Diyan, N.A. Almujally, J.-M. Kang, CAAI Trans. Intell. Technol. (2024) http://dx.doi.org/10.1049/cit2.12293.

[213] C. Qiu, H. Li, C. Qi, B. Li, Heliyon 10 (5) (2024) http://dx.doi.org/10.1016/j.heliyon.2024.e26147.

[214] M.A. Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, R.R. Yager, Inform. Sci. 345 (2016) 340–354, http://dx.doi.org/10.1016/j.ins.2016.01.082.

[215] G. Sannino, G. de Pietro, Future Gener. Comput. Syst. 86 (2018) 446–455, http://dx.doi.org/10.1016/j.future.2018.03.057.

[216] S.S. Xu, M.-W. Mak, C.-C. Cheung, IEEE J. Biomed. Heal. Inform. 23 (4) (2019) 1574–1584, http://dx.doi.org/10.1109/JBHI.2018.2871510.

[217] A.H. Ribeiro, M.H. Ribeiro, G.M.M. Paixão, D.M. Oliveira, P.R. Gomes, J.A. Canazart, M.P.S. Ferreira, C.R. Andersson, P.W. Macfarlane, W. Meira, T.B. Schön, A.L.P. Ribeiro, Nat. Commun. 11 (1) (2020) 1760, http://dx.doi.org/10.1038/s41467-020-15432-4.

[218] W. Zeng, L. Shan, C. Yuan, S. Du, Appl. Soft Comput. (2024) 112056, http://dx.doi.org/10.1016/j.asoc.2024.112056.

[219] D.-H. Shin, R.C. Park, K. Chung, IEEE Access 8 (2020) 108664–108674, http://dx.doi.org/10.1109/ACCESS.2020.3000638.

[220] A. Majumdar, R. Ward, 2017 International Joint Conference on Neural Networks, IJCNN, IEEE, 2017, pp. 4400–4407, http://dx.doi.org/10.1109/IJCNN.2017.7966413.

[221] J. Yang, Y. Bai, F. Lin, M. Liu, Z. Hou, X. Liu, Int. J. Mach. Learn. Cybern. 9 (10) (2018) 1733–1740, http://dx.doi.org/10.1007/s13042-017-0677-5.

[222] Q. Yao, R. Wang, X. Fan, J. Liu, Y. Li, Inf. Fusion 53 (2020) 174–182, http://dx.doi.org/10.1016/j.inffus.2019.06.024.

[223] H. Xie, H. Liu, S. Zhou, T. Gao, M. Shu, Appl. Intell. (2022) http://dx.doi.org/10.1007/s10489-022-04303-8.

[224] S. Karthikeyani, S. Sasipriya, M. Ramkumar, Biomed. Signal Process. Control. 92 (2024) 105997, http://dx.doi.org/10.1016/j.bspc.2024.105997.

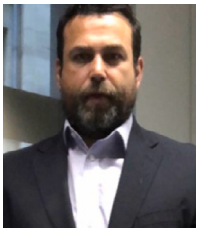
[225] W. Fan, Y. Si, W. Yang, G. Zhang, Measurement 185 (2021) 110040, http://dx.doi.org/10.1016/j.measurement.2021.110040.

[226] M. Salem, S. Taheri, J.-S. Yuan, 2018 IEEE Biomedical Circuits and Systems Conference, BioCAS, Ieee, 2018, pp. 1–4, http://dx.doi.org/10.1109/biocas.2018.8584808.

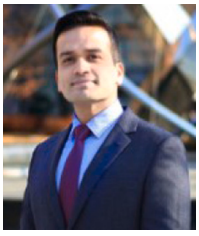
[227] A. Pal, R. Srivastva, Y.N. Singh, Big Data Res. 26 (2021) 100271, http://dx.doi.org/10.1016/j.bdr.2021.100271.
[228] K. Weimann, T.O.F. Conrad, Sci. Rep. 11 (1) (2021) 5251, http://dx.doi.org/10.1038/s41598-021-84374-8.
[229] X. Sun, P. Liu, Z. He, Y. Han, B. Su, Ecol. Inform. 69 (2022) 101628, http://dx.doi.org/10.1016/j.ecoinf.2022.101628.
[230] H. Bechinia, D. Benmerzoug, N. Khlifa, IEEE Access 12 (2024) 40827–40841, http://dx.doi.org/10.1109/ACCESS.2024.3378730.
[231] A. Peimankar, A. Ebrahimi, U.K. Wiil, Netw. Model. Anal. Heal. Inform. Bioinform. 13 (1) (2024) 1–13, http://dx.doi.org/10.1007/s13721-024-00481-2.
[232] A.M. Gitau, V. Ruto, Y. Njathi, L. Mugambi, V.A. Sitati, A. Kaburia, Comput. Cardiol. (2024) http://dx.doi.org/10.22489/CinC.2024.498.
[233] E.B. L'ubomır Antoni, P. Bugata, Peter Bugata Jr., D. Gajdoš, D. Hudák, V. Kmecová, M.G. Shridhara, M. Stanková, G. Vozáriková, Comput. Cardiol. (2024) http://dx.doi.org/10.22489/CinC.2024.231.
[234] F.M. Dias, E. Ribeiro, Q.B. Soares, J.E. Krieger, M.A. Gutierrez, Comput. Cardiol. (2024) http://dx.doi.org/10.22489/CinC.2024.398.
[235] H.-C. Yoon, D.-K. Kim, H.-S. Kim, W.-Y. Seo, C.-H. Heo, S.-H. Kim, Comput. Cardiol. (2024) http://dx.doi.org/10.22489/CinC.2024.227.
[236] W. Yang, Y. Si, Di Wang, B. Guo, Comput. Biol. Med. 101 (2018) 22–32, http://dx.doi.org/10.1016/j.compbiomed.2018.08.003.
[237] T. Khatibi, N. Rabinezhadsadatmahaleh, Australas. Phys. Eng. Sci. Med. (2019) http://dx.doi.org/10.1007/s13246-019-00814-w.
[238] W. Lu, H. Hou, J. Chu, Biomed. Signal Process. Control. 41 (2018) 152–160, http://dx.doi.org/10.1016/j.bspc.2017.11.010.
[239] S.S. Mousavi, F. Afghah, A. Razi, U.R. Acharya, IEEE-EMBS International Conference on Biomedical and Health Informatics. IEEE-EMBS International Conference on Biomedical and Health Informatics, Vol. 2019, 2019, http://dx.doi.org/10.1109/BHI.2019.8834637.
[240] S. Mousavi, F. Afghah, ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2019, pp. 1308–1312, http://dx.doi.org/10.1109/ICASSP.2019.8683140.
[241] J. Zhang, A. Liu, M. Gao, X. Chen, X. Zhang, X. Chen, Artif. Intell. Med. 106 (2020) 101856, http://dx.doi.org/10.1016/j.artmed.2020.101856.
[242] S. Dhyani, A. Kumar, S. Choudhury, Biomed. Signal Process. Control. 79 (2023) 104160, http://dx.doi.org/10.1016/j.bspc.2022.104160.
[243] P. Warrick, M.N. Homsi, 2017 Computing in Cardiology, CinC, IEEE, 2017, pp. 1–4, http://dx.doi.org/10.22489/CinC.2017.161-460.
[244] M. Zihlmann, D. Perekrestenko, M. Tschannen, 2017 Computing in Cardiology, CinC, IEEE, 2017, pp. 1–4, http://dx.doi.org/10.22489/CinC.2017.070-060.
[245] S.L. Oh, E.Y.K. Ng, R.S. Tan, U.R. Acharya, Comput. Biol. Med. 102 (2018) 278–287, http://dx.doi.org/10.1016/j.compbiomed.2018.06.002.
[246] G. Wang, C. Zhang, Y. Liu, H. Yang, D. Fu, H. Wang, P. Zhang, Inform. Sci. 501 (2019) 523–542, http://dx.doi.org/10.1016/j.ins.2018.06.062.
[247] F. Liu, X. Zhou, J. Cao, Z. Wang, H. Wang, Y. Zhang, ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2019, pp. 1303–1307.
[248] R. He, Y. Liu, K. Wang, N. Zhao, Y. Yuan, Q. Li, H. Zhang, IEEE Access 7 (2019) 102119–102135, http://dx.doi.org/10.1109/ACCESS.2019.2931500.
[249] O. Yildirim, M. Talo, E.J. Ciaccio, R.S. Tan, U.R. Acharya, Comput. Methods Programs Biomed. 197 (2020) 105740, http://dx.doi.org/10.1016/j.cmpb.2020.105740.
[250] C. Chen, Z. Hua, R. Zhang, G. Liu, W. Wen, Biomed. Signal Process. Control. 57 (2020) 101819, http://dx.doi.org/10.1016/j.bspc.2019.101819.
[251] H. Shi, C. Qin, D. Xiao, L. Zhao, C. Liu, Knowl.-Based Syst. 188 (2020) 105036, http://dx.doi.org/10.1016/j.knosys.2019.105036.
[252] A. Chen, F. Wang, W. Liu, S. Chang, H. Wang, J. He, Q. Huang, Comput. Methods Programs Biomed. 193 (2020) 105479, http://dx.doi.org/10.1016/j.cmpb.2020.105479.
[253] M. Hammad, A.M. Iliyasu, A. Subasi, E.S.L. Ho, A.A.A. El-Latif, IEEE Trans. Instrum. Meas. 70 (2021) 1–9, http://dx.doi.org/10.1109/TIM.2020.3033072.
[254] E. Essa, X. Xie, IEEE Access 9 (2021) 103452–103464, http://dx.doi.org/10.1109/ACCESS.2021.3098986.
[255] M.S. Haleem, R. Castaldo, S.M. Pagliara, M. Petretta, M. Salvatore, M. Franzese, L. Pecchia, Biomed. Signal Process. Control. 70 (2021) 102968, http://dx.doi.org/10.1016/j.bspc.2021.102968.
[256] P. Madan, V. Singh, D.P. Singh, M. Diwakar, B. Pant, A. Kishor, Bioeng. (Basel, Switzerland) 9 (4) (2022) http://dx.doi.org/10.3390/bioengineering9040152.
[257] J. Rahul, L.D. Sharma, Biocybern. Biomed. Eng. 42 (1) (2022) 312–324, http://dx.doi.org/10.1016/j.bbe.2022.02.006.
[258] S.U. Hassan, M.S. Mohd Zahid, T.A. Abdullah, K. Husain, Digit. Heal. 8 (2022) 20552076221102766, http://dx.doi.org/10.1177/20552076221102766.
[259] P. Varalakshmi, A.P. Sankaran, Biomed. Signal Process. Control. 80 (2023) 104248, http://dx.doi.org/10.1016/j.bspc.2022.104248.
[260] S. Shadmand, B. Mashoufi, Biomed. Signal Process. Control. 25 (2016) 12–23, http://dx.doi.org/10.1016/j.bspc.2015.10.008.
[261] M. Amrani, M. Hammad, F. Jiang, K. Wang, A. Amrani, Neural Comput. Appl. 30 (7) (2018) 2047–2057, http://dx.doi.org/10.1007/s00521-018-3616-9.
[262] E.H. Houssein, D.S. Abdelminaam, I.E. Ibrahim, M. Hassaballah, Y.M. Wazery, IEEE Access 9 (2021) 86194–86206, http://dx.doi.org/10.1109/ACCESS.2021.3088783.
[263] J. Cui, L. Wang, X. He, V.H.C. de Albuquerque, S.A. AlQahtani, M.M. Hassan, Neural Comput. Appl. (2021) http://dx.doi.org/10.1007/s00521-021-06487-5.
[264] Z. Ahmad, A. Tabassum, L. Guan, N.M. Khan, IEEE Access 9 (2021) 100615–100626, http://dx.doi.org/10.1109/ACCESS.2021.3097614.
[265] M.K. Ojha, S. Wadhwani, A.K. Wadhwani, A. Shukla, Phys. Eng. Sci. Med. 45 (2) (2022) 665–674, http://dx.doi.org/10.1007/s13246-022-01119-1.
[266] A. Kumar, S. Kumar, V. Dutt, A.K. Dubey, V. García-Díaz, Biomed. Signal Process. Control. 76 (2022) 103638, http://dx.doi.org/10.1016/j.bspc.2022.103638.
[267] E. Kıymaç, Y. Kaya, Expert Syst. Appl. 213 (2023) 119162, http://dx.doi.org/10.1016/j.eswa.2022.119162.
[268] N. Sinha, R. Kumar Tripathy, A. Das, Biomed. Signal Process. Control. 78 (2022) 103943, http://dx.doi.org/10.1016/j.bspc.2022.103943.
[269] S. Raj, K.C. Ray, Sci. Rep. 8 (1) (2018) 11395, http://dx.doi.org/10.1038/s41598-018-29690-2.
[270] M.M. Al Rahhal, Y. Bazi, H. Almubarak, N. Alajlan, M. Al Zuair, IEEE Access 7 (2019) 182225–182237, http://dx.doi.org/10.1109/ACCESS.2019.2960116.
[271] P. Wang, B. Hou, S. Shao, R. Yan, IEEE Access 7 (2019) 100910–100922, http://dx.doi.org/10.1109/ACCESS.2019.2930882.
[272] A.M. Shaker, M. Tantawi, H.A. Shedeed, M.F. Tolba, IEEE Access 8 (2020) 35592–35605, http://dx.doi.org/10.1109/ACCESS.2020.2974712.
[273] S. Ma, J. Cui, C.-L. Chen, X. Chen, Y. Ma, Measurement 203 (2022) 111978, http://dx.doi.org/10.1016/j.measurement.2022.111978.
[274] M.S. Islam, M.N. Islam, N. Hashim, M. Rashid, B.S. Bari, F.A. Farid, IEEE Access 10 (2022) 58081–58096, http://dx.doi.org/10.1109/ACCESS.2022.3178710.
[275] J. Qin, F. Gao, Z. Wang, D.C. Wong, Z. Zhao, S.D. Relton, H. Fang, Artif. Intell. Med. 136 (2023) 102489, http://dx.doi.org/10.1016/j.artmed.2023.102489.
[276] Y. Xia, Y. Xu, P. Chen, J. Zhang, Y. Zhang, Biomed. Signal Process. Control. 80 (2023) 104276, http://dx.doi.org/10.1016/j.bspc.2022.104276.
[277] M.S. Islam, K.F. Hasan, S. Sultana, S. Uddin, P. Lio', J.M.W. Quinn, M.A. Moni, Neural Netw. : Off. J. Int. Neural Netw. Soc. 162 (2023) 271–287, http://dx.doi.org/10.1016/j.neunet.2023.03.004.
[278] K. Luo, J. Li, Z. Wang, A. Cuschieri, J. Heal. Eng. 2017 (2017) 4108720, http://dx.doi.org/10.1155/2017/4108720.
[279] O. Yildirim, U.B. Baloglu, R.-S. Tan, E.J. Ciaccio, U.R. Acharya, Comput. Methods Programs Biomed. 176 (2019) 121–133, http://dx.doi.org/10.1016/j.cmpb.2019.05.004.
[280] E.K. Wang, X. Zhang, L. Pan, IEEE Access 7 (2019) 182873–182880, http://dx.doi.org/10.1109/ACCESS.2019.2936525.
[281] B. Hou, J. Yang, P. Wang, R. Yan, IEEE Trans. Instrum. Meas. 69 (4) (2020) 1232–1240, http://dx.doi.org/10.1109/TIM.2019.2910342.
[282] M. Thill, W. Konen, H. Wang, T. Bäck, Appl. Soft Comput. 112 (2021) 107751, http://dx.doi.org/10.1016/j.asoc.2021.107751.
[283] P. Liu, X. Sun, Y. Han, Z. He, W. Zhang, C. Wu, Biomed. Signal Process. Control. 71 (2022) 103228, http://dx.doi.org/10.1016/j.bspc.2021.103228.
[284] M. Ramkumar, R. Sarath Kumar, A. Manjunathan, M. Mathankumar, J. Pauliah, Biomed. Signal Process. Control. 77 (2022) 103826, http://dx.doi.org/10.1016/j.bspc.2022.103826.
[285] Y. Xia, Y. Xiong, K. Wang, Biomed. Signal Process. Control. 86 (2023) 105271, http://dx.doi.org/10.1016/j.bspc.2023.105271.
[286] M. Roy, S. Majumder, A. Halder, U. Biswas, Eng. Appl. Artif. Intell. 124 (2023) 106484, http://dx.doi.org/10.1016/j.engappai.2023.106484.
[287] Z. Li, H. Zhang, Biomed. Signal Process. Control. 85 (2023) 104849, http://dx.doi.org/10.1016/j.bspc.2023.104849.
[288] F. Liu, X. Zhou, T. Wang, J. Cao, Z. Wang, H. Wang, Y. Zhang, 2019 International Joint Conference on Neural Networks, IJCNN, IEEE, 2019, pp. 1–8, http://dx.doi.org/10.1109/IJCNN.2019.8852037.
[289] N. Mangathayaru, P. Rani, V. Janaki, K. Srinivas, B. Mathura Bai, G. Sai Mohan, B. Lalith Bharadwaj, Comput. Mater. Contin. 69 (2) (2021) 2425–2443, http://dx.doi.org/10.32604/cmc.2021.016534.
[290] M. Jiang, J. Gu, Y. Li, B. Wei, J. Zhang, Z. Wang, L. Xia, Front. Physiol. 12 (2021) 683025, http://dx.doi.org/10.3389/fphys.2021.683025.
[291] P. Singh, A. Sharma, IEEE Trans. Instrum. Meas. 71 (2022) 1–10, http://dx.doi.org/10.1109/TIM.2022.3197757.
[292] Y. Huang, H. Li, X. Yu, Biomed. Signal Process. Control. 83 (2023) 104628, http://dx.doi.org/10.1016/j.bspc.2023.104628.
[293] Y. Zhao, J. Ren, B. Zhang, J. Wu, Y. Lyu, Biomed. Signal Process. Control. 80 (2023) 104337, http://dx.doi.org/10.1016/j.bspc.2022.104337.
[294] F. Liu, H. Li, T. Wu, H. Lin, C. Lin, G. Han, ISA Trans. (2023) http://dx.doi.org/10.1016/j.isatra.2023.02.028.
[295] W. Wu, Y. Huang, X. Wu, Biomed. Signal Process. Control. (2023) 105017, http://dx.doi.org/10.1016/j.bspc.2023.105017.
[296] Y. Wang, G. Zhou, C. Yang, IEEE Trans. Instrum. Meas. 72 (2023) 1–15, http://dx.doi.org/10.1109/TIM.2022.3232646.

[297] A. Isin, S. Ozdalili, Procedia Comput. Sci. 120 (2017) 268–275, http://dx.doi.org/10.1016/j.procs.2017.11.238.
[298] A. Raza, K.P. Tran, L. Koehl, S. Li, Knowl.-Based Syst. 236 (2022) 107763, http://dx.doi.org/10.1016/j.knosys.2021.107763.
[299] Z. Wang, S. Stavrakis, B. Yao, Comput. Biol. Med. 155 (2023) 106641, http://dx.doi.org/10.1016/j.compbiomed.2023.106641.
[300] G.B. Moody, R.G. Mark, IEEE Eng. Med. Biol. Mag. 20 (3) (2001) 45–50, http://dx.doi.org/10.1109/51.932724.
[301] G. Moody, Proc. Comput. Cardiol. 10 (1983) 227–230.
[302] S.D. Greenwald, R.S. Patil, R.G. Mark, Improved Detection and Classification of Arrhythmias in Noise-Corrupted Electrocardiograms Using Contextual Information, IEEE, 1990.
[303] P. Albrecht, ST Segment Characterization for Long Term Automated ECG Analysis (Ph.D. thesis), Massachusetts Institute of Technology, Department of Electrical Engineering . . . , 1983.
[304] S.D. Greenwald, The Development and Analysis of a Ventricular Fibrillation Detector (Ph.D. thesis), Massachusetts Institute of Technology, 1986.
[305] G.B. Moody, W. Muldrow, R.G. Mark, Comput. Cardiol. 11 (3) (1984) 381–384.
[306] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, Circulation 101 (23) (2000) e215–e220, http://dx.doi.org/10.1161/01.cir.101.23.e215.
[307] I. Silva, G.B. Moody, L. Celi, 2011 Computing in Cardiology, IEEE, 2011, pp. 273–276.
[308] G.D. Clifford, C. Liu, B. Moody, H.L. Li-wei, I. Silva, Q. Li, A. Johnson, R.G. Mark, 2017 Computing in Cardiology, CinC, IEEE, 2017, pp. 1–4.
[309] F. Nolle, F. Badura, J. Catlett, R. Bowser, M. Sketch, Comput. Cardiol. 13 (1) (1986) 515–518.
[310] H.A. Guvenir, B. Acar, G. Demiroz, A. Cekin, Computers in Cardiology 1997, IEEE, 1997, pp. 433–436, http://dx.doi.org/10.1109/cic.1997.647926.
[311] E. Yakushenko, 2008, URL: https://physionet.org/content/incartdb/1.0.0/.
[312] G. Moody, A. Goldberger, S. McClennen, S. Swiryn, Computers in Cardiology 2001. Vol. 28 (Cat. No. 01CH37287), IEEE, 2001, pp. 113–116, http://dx.doi.org/10.1109/cic.2001.977604.
[313] A.H. Association, 1982, URL: https://physionet.org/content/ahadb/1.0.0/.
[314] J.-W. Zhang, X. Liu, J. Dong, Int. J. Artif. Intell. Tools 21 (05) (2012) 1240020, http://dx.doi.org/10.1142/s0218213012400209.
[315] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He, et al., J. Med. Imaging Heal. Inform. 8 (7) (2018) 1368–1373, http://dx.doi.org/10.1166/jmihi.2018.2442.
[316] R. Bousseljot, D. Kreiseler, A. Schnabel, Biomed. Eng./ Biomed. Eng. 40 (s1) (1995) 317–318.
[317] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F.I. Lunze, W. Samek, T. Schaeffter, Sci. Data 7 (1) (2020) 154, http://dx.doi.org/10.1038/s41597-020-0495-6.
[318] E.A.P. Alday, A. Gu, A.J. Shah, C. Robichaux, A.-K.I. Wong, C. Liu, F. Liu, A.B. Rad, A. Elola, S. Seyedi, et al., Physiol. Meas. 41 (12) (2020) 124003, http://dx.doi.org/10.22489/cinc.2020.236.
[319] J. Zheng, H. Guo, H. Chu, PhysioNet (2022) http://dx.doi.org/10.13026/92ks-sq55.
[320] J. Zheng, J. Zhang, S. Danioko, H. Yao, H. Guo, C. Rakovski, Sci. Data 7 (1) (2020) 48, http://dx.doi.org/10.1038/s41597-020-0386-x.
[321] S. Tan, G. Androz, A. Chamseddine, P. Fecteau, A. Courville, Y. Bengio, J.P. Cohen, 2019, http://dx.doi.org/10.48550/arXiv.1910.09570, arXiv preprint.
[322] A. Taddei, G. Distante, M. Emdin, P. Pisani, G. Moody, C. Zeelenberg, C. Marchesi, Eur. Heart J. 13 (9) (1992) 1164–1172, http://dx.doi.org/10.1093/oxfordjournals.eurheartj.a060332.
[323] N. Van Eck, L. Waltman, Scientometrics 84 (2) (2010) 523–538, http://dx.doi.org/10.1007/s11192-009-0146-3.
[324] D.G. Altman, J.M. Bland, BMJ: Br. Med. J. 308 (6943) (1994) 1552, http://dx.doi.org/10.1136/bmj.308.6943.1552.
[325] A.G. Lalkhen, A. McCluskey, Contin. Educ. Anaesth. Crit. Care Pain 8 (6) (2008) 221–223, http://dx.doi.org/10.1093/bjaceaccp/mkn041.

Mr. Ahtisham Ayyub He is a PhD Student and an early career researcher at the School of Computer Science and Mathematics, Kingston University, Kingston Upon Thames, London. His contributions in this paper consist of conceptualisation of the study, data collection, information processing and the drathing of the manuscript.



Professor Christos Politis is a Professor (Chair) of Digital Technologies at Kingston University London, School of Computer Science (CS). He is the Director of the Digital Information Research Centre (DIRC). He is a senior member of the IEEE and UK chartered engineer.



Dr. Muhammad Arslan Usman He is a senior lecturer at the School of Computer Science and Mathematics of Kingston University London. He also has academic responsibilities in Research, Teaching (FHEA) and Knowledge Exchange. His contributions in this paper are validation of the study and proofreading of the manuscript.