A Literature Review: ECG-Based Models for Arrhythmia Diagnosis Using Artificial Intelligence Techniques

Bioinformatics and Biology Insights
Volume 17: 1–29

DOI: 10.1177/11779322221149600

Abir Boulif[1], Bouchra Ananou[1], Mustapha Ouladsine[1] and Stéphane Delliaux[2]

[1]Aix-Marseille University, CNRS, LIS, Marseille, France. [2]Aix-Marseille University, INSERM, INRAE, C2VN, Marseille, France.

ABSTRACT: In the health care and medical domain, it has been proven challenging to diagnose correctly many diseases with complicated and interferential symptoms, including arrhythmia. However, with the evolution of artificial intelligence (AI) techniques, the diagnosis and prognosis of arrhythmia became easier for the physicians and practitioners using only an electrocardiogram (ECG) examination. This review presents a synthesis of the studies conducted in the last 12 years to predict arrhythmia's occurrence by classifying automatically different heartbeat rhythms. From a variety of research academic databases, 40 studies were selected to analyze, among which 29 of them applied deep learning methods (72.5%), 9 of them addressed the problem with machine learning methods (22.5%), and 2 of them combined both deep learning and machine learning to predict arrhythmia (5%). Indeed, the use of AI for arrhythmia diagnosis is emerging in literature, although there are some challenging issues, such as the explicability of the Deep Learning methods and the computational resources needed to achieve high performance. However, with the continuous development of cloud platforms and quantum calculation for AI, we can achieve a breakthrough in arrhythmia diagnosis.

KEYWORDS: Arrhythmia, artificial intelligence, deep learning, diagnosis, prediction, health care

RECEIVED: April 21, 2022. **ACCEPTED:** December 18, 2022.

TYPE: Review

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Abir Boulif, Aix-Marseille University, CNRS, LIS, 13397 Marseille, France. Email: abir.boulif@univ-amu.fr

Introduction

Neuro-cardiovascular diseases are the leading cause of death in the world. Arrhythmias represent a category of these diseases associated with medical issues that can range from a minor inconvenience or discomfort to a fatal problem. An arrhythmia is an abnormality of the heart's rhythm which is controlled by electrical signals. It may beat too slow, too quick, or irregular.[1] The electrocardiogram (ECG) is an effective tool for arrhythmia diagnosis, it measures the heart's electrical activity. Other ambulatory devices can be used for the same aim, such as Holter Monitor.[2] However, the diagnosis of arrhythmias is not always obvious, especially for atrial fibrillation (AF) that can be related to asymptomatic and transient forms. Moreover, there are some limitations in the extraction methods and time series analysis from ECG singularities and their dynamics. To address these limitations, artificial intelligence (AI) is applied to the diagnosis and prognosis of diseases, such as arrhythmia. For this end, we focused in our previous works on the diagnosis of AF with machine learning (ML) methods. For instance, we conducted a multi-dynamics analysis of QRS complex with support vector machine (SVM) and multiple kernel learning (MKL) in Trardi et al,[3] which reached respective sensitivity values of 96.54% and 95.47%. Other works were mainly based on the extraction of different features from R-wave derivatives for automatic medical decision-making, especially for AF detection as in literature.[4-6] In addition, the use of univariate and multivariate methods plays a major role in the analysis of the ECG time series.

On future work, the objective is to process ECG signals and classify different categories of heartbeat to detect different types of arrhythmia and thus help health care professionals. For this aim, we realized a literature review on ECG-based models for arrhythmia detection using AI techniques in the last 12 years.

This article is organized as follows: Section "Methods of Search and Selection" presents an overview on search strategy and the criteria of studies' selection. Section "Results of Studies' Exploration" emphasizes the exploration of the studies selected and the collected information from these studies. Section "Discussion and Interpretation" is dedicated to the interpretation and discussion of the obtained results, the contribution of this review, and comparison to other literature reviews. To sum up, the final section "Conclusions" contains a summary of the strengths and limitations of the used deep learning (DL) techniques.

Methods of Search and Selection

This section presents the search strategy, the criteria of selection, and the extraction of study characteristics.

Search strategy

To conduct this review, multiple academic research databases were selected to gather relevant articles that were published from January 2010 to September 2022. These open source databases are PubMed, IEEE Xplore, Springer, ScienceDirect, and ResearchGate. PubMed and IEEE Xplore are considered as 2 of the leading databases in biological sciences and engineering, respectively.[7] Springer is one of the leading research publishers that provides a large number of resources for literature in different fields. ScienceDirect was used for its several and various

Table 1. Inclusion criteria.

CRITERION	DESCRIPTION
Date	2010-2022
Type	Journal article and conference
Domain	Bioinformatics and computer science
Participants	With no other disease symptoms/ medications effects
Review	Peer-reviewed
Evaluation metrics	Accuracy /sensitivity /specificity/F1 score / AUC /confusion matrix

Abbreviation: AUC, area under curve.

peer-reviewed journals and articles. For ResearchGate, it allows access to a large number of free papers because it is the largest academic social network in terms of active users.[8]

To realize a targeted search, we identify articles from their titles and abstracts using the following keywords: "artificial intelligence for arrhythmia," "arrhythmia diagnosis," "arrhythmia classification," "heartbeat classification," "ECG classification."

Collection of data sources was based on the process below:

1. Targeted web-based search.
2. Classification of sources by level of relevance.
3. Studies' selection based on the inclusion criteria listed in Table 1.

The selected databases provide peer-reviewed articles except ResearchGate that does not require a peer-review for the published articles.

Study selection

The scope of the studies selected comprises the following conditions:

- ➢ Studies conducted on the diagnosis of arrhythmia in the last 12 years;
- ➢ Studies addressing the classification of any type of arrhythmia to,
  - Present a variety of databases,
  - Identify a large number of arrhythmia in both subclasses and super-classes;
- ➢ Studies handling the diagnosis of arrhythmia with no other cardiovascular disease interference to encircle tightly the scope and circumstances of occurrence of this abnormality and thus realize a precise and accurate diagnosis and prognosis.

Moreover, the studies can handle either beat classification or category classification. The former contains many subclasses of heart rhythm as shown in Table A3 (Appendix 1) and the latter addresses the classification topic based on the main arrhythmia categories defined by the American National Standards Institute/Association for the Advancement of Medical Instrumentation (ANSI/AAMI). Different heartbeat categories will be found in Table A2 (Appendix 1).

Many criteria were considered in the first articles' selection, but only articles with the inclusion criteria sorted in Table 1 were retained.

Study mining: extraction of study characteristics

With a view to exploring the selected publications, retrieving information, and identifying patterns, we extracted the following characteristics:

- Study perimeter: defines year of publication and authors;
- Input information: includes datasets used in the study, number of participants, and number of arrhythmia classes to predict;
- ECG signal information: contains ECG recording format and signal duration;
- Feature set: defines the extraction approach and the extracted features from each study. The extraction methodology depends on the learning structure, hand-crafted methods, or end-to-end learning (where the selection, extraction, and classification are embedded in one stage);
- Methods: define the pre-processing and prediction methods used in each study to implement the AI algorithms;
- Evaluation: presents the metrics and key performance to evaluate the prediction.

The objective is to extract a large number of characteristics to analyze deeply each study.

Results of Studies' Exploration

Selection

To search for the adequate papers, we rely on the process shown in Figure 1.

First, we realized a quick search for the topic by keywords which results in 730 records. Second, we removed duplicates given that we used 5 research databases, then a paper inventory was held by sorting publications by abstract.

When applying the inclusion criteria in Table 1, we focused on selecting papers that do not deal with other cardiovascular diseases. Although the study[9] includes the treatment of myocardial ischemia, it was selected because there was no interference with arrhythmia diagnosis; the 2 diseases were independently addressed. Irfan et al[10] used a dataset with 13 types of heartbeats, including arrhythmias and myocardial infarctions, which added more variety to the dataset without affecting the performance of the model for arrhythmia diagnosis.

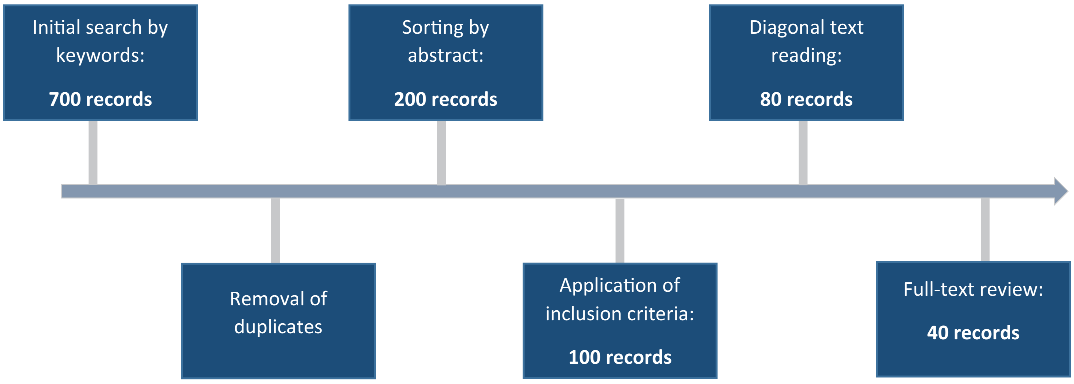

Figure 1. The process of study selection.

We included the study[11] although they conducted the classification of congestive heart failure rhythm with the normal and arrhythmia rhythms because this allowed the classifier to have a higher recognition ability in classification.

In this stage, 100 records were selected. We adopted a diagonal-text reading and found that some papers focused on the analysis of the ECG signals without addressing the arrhythmia prediction, so we excluded 18 studies.

After a full-text review, we selected 40 studies that handle the arrhythmia classification with various methods of pre-processing and several approaches of diagnosis.

Study design

Various datasets were used for the selected studies. In total, 95% of the studies used open access datasets from PhysioNet repository[12] mainly the Massachusetts Institute of Technology-Boston's Beth Israel Hospital (MIT-BIH) arrhythmia database. Besides, Li et al[13] used data collected from Fluke ProSim2 vital sign simulator which is a paying portable solution that transforms physiological simulation by adding multi-parameter functionalities, including ECG simulation and arrhythmia waveform selections.[14] Ribeiro et al[15] used the Telehealth Network of Minas Gerais (TNMG) dataset which was obtained from one of the largest telehealth services in Brazil. Hannun et al[16] used an own collected data recorded by Zio monitor which is a portable device described by physicians to diagnose irregular heart rhythms for up to 14 days[17] unlike Holter monitor, used in Park and Kang,[18] which can be worn only from 24 to 28 hours.

For the studies relying on the PhysioNet repository databases, some of them used small samples of individuals (between 14 and 78 participants). Hannun et al[16] included 53 877 participants aged 69 ± 16 years and Ribeiro et al[15] collected data from 1 676 384 individuals older than 16 years. The remainder of the studies did not report the exact number of participants (Table 2).

Figure 2 shows that 9 studies used augmented data to balance the datasets and enhance the AI models. It should be observed that most of the selected studies are published in the last 5 years, and this can be explained by the recent emergence of AI and the growth of literature sources lately.

Ullah et al[39] used 2 types of datasets from PhysioNet: MIT-BIH arrhythmia database and PTB Diagnostic ECG Database. In addition, they used generative adversarial network (GAN) model to generate new artificial signal for classes with small amount of data. The same technique was used to augment data in Ma et al.[41]

Irfan et al[10] used the publicly available MIT-BIH arrhythmia database and UCI arrhythmia dataset available in the University of California Irvine ML Repository.

Synthetic minority oversampling technique (SMOTE) is used in literature[10,50] to handle the problem of the imbalanced data in the MIT-BIH arrhythmia database and MIT-BIH AF database. SMOTE relies on a k-nearest neighbor algorithm to create new synthetic data. Whereas, Shahin et al[40] upsampled the training data by randomly duplicating the samples resulting in relatively equal classes but not as varied as it should be.

Two public databases from China Physiological Signal Challenge 2018 (CPSC-2018) and Computing in Cardiology Challenge 2017 (CinC-2017) were used in Wang et al;[43] the data were augmented by applying, respectively, flipping and random erasure techniques.

Hu et al[45] used MIT-BIH arrhythmia database for classifying 4 classes following the AAMI annotation and 8 classes following the widely used classification in literature. Different label classifications comprising 41, 20, and 5 classes are also reported in Feyisa et al[46] with the use of PTB-XL dataset which is a 12-lead database with various types of arrhythmia. Wang et al[48] used the CinC-2017 database and applied a data augmentation with the Mix-Up operation in the training stage to reduce the data imbalance and thus the overfitting; the method generates more training data without extra computational resources.

Table 2 shows in details the input information for each study.

The databases in the table above can contain a higher number of classes than what was reported in this review, but since there are some studies that focus on specific rhythms, we mentioned only the classes that were actually used for classification.

Feature set

Given that some studies used DL techniques, the end-to-end structure was implemented, in which selection, extraction, and

Table 2. Input information.

YEAR	STUDY	DATABASES	NO. OF CLASSES	NO. OF PARTICIPANTS	ECG RECORDING FORMAT	SIGNAL DURATION	TYPE OF DATA
2019	Chen et al[19]	MIT-BIH arrhythmia database,	6	47	2-lead	30 mn	Native
		QT database,	6	NR	2-lead	NR	
		MIT-BIH supraventricular arrhythmia database,	4	NR	2-lead	30 mn	
		INCART database	4	NR	12-lead	NR	
2017	Acharya et al[20]	MIT-BIH arrhythmia database	5	47	1-lead (lead II)	30 mn	Native, Augmented
2019	Yildirim et al[21]	MIT-BIH arrhythmia database	5	NR	1-lead (lead II)	30 mn	Native
2019	Yang et al[22]	MIT-BIH arrhythmia database	15	NR	2-lead (II and V1)	30 mn	Native
		INCART database	7	NR	2-lead (II and V1) 2-lead (V1 and V5) 2-lead (II and V5) 2-lead (II and V1)		
2018	Yildirim[23]	MIT-BIH arrhythmia database	5	47	NR	30 mn	Native
2014	Sumathi et al[9]	MIT-BIH arrhythmia Database, MIT-BIH AF database, MIT-BIH malignant ventricular ectopy database	5	NR	NR	NR	Native
2019	Gao et al[24]	MIT-BIH arrhythmia database	8	47	NR	30 mn	Native
2013	Martis et al[25]	MIT-BIH arrhythmia database	5	NR	NR	NR	Native
2016	Li et al[13]	MIT-BIH arrhythmia database, Fluke ProSim2 vital sign simulator	5	NR	NR	NR	Native, Simulated
2018	Anwar et al[26]	MIT-BIH arrhythmia database, MIT-BIH supraventricular arrhythmia database	18 5	47 NR	NR	30 mn	Native
2013	Liu[27]	MIT-BIH arrhythmia database	5	NR	2-lead (II and VI)	30 mn	Native
2015	Elhaj et al[28]	MIT-BIH arrhythmia database	5	NR	NR	30 mn	Native
2019	Kim et al[29]	MIT-BIH arrhythmia database	5	44	NR	NR	Native
2018	Yıldırım[30]	MIT-BIH arrhythmia database	13 15 17	45 (19 F, 26 M)	1-lead (MLII)	10 seconds	Native
2018	Oh et al[31]	MIT-BIH arrhythmia database	5	47	1-lead (MLII)	NR	Native

(Continued)

Table 2. (Continued)

YEAR	STUDY	DATABASES	NO. OF CLASSES	NO. OF PARTICIPANTS	ECG RECORDING FORMAT	SIGNAL DURATION	TYPE OF DATA
2018	Oh et al[32]	MIT-BIH arrhythmia database	5	47	1-lead (MLII)	Variable length	Native
2018	Raj and Ray[33]	MIT-BIH arrhythmia database	16 5	47	NR	30 mn	Native
2020	Ribeiro et al[15]	Telehealth Network of Minas Gerais (TNMG) dataset	6	1 676 384 (60.3% F, 39.7% M) (>16 years)	12-lead	7-10 seconds	Native
2017	Qin et al[34]	MIT-BIH arrhythmia database	6	NR	1-lead (II)	30 mn	Native
2017	Rajagopal and Ranganathan[35]	MIT-BIH arrhythmia database	5	47	NR	30 mn	Native
2010	Benali et al[36]	MIT-BIH arrhythmia database	5	NR	NR	NR	Native
2017	Rajesh and Dhuli[37]	MIT-BIH arrhythmia database, INCART	5	NR	NR	30:06 mn	Native
2018	Yang et al[38]	MIT-BIH arrhythmia database	5	NR	2-lead (MLII and V5)	30 mn	Native
2019	Hannun et al[16]	Original ECG dataset recorded by Zio monitor	12	53 877 (43% F, 57% M) (69 ± 16 year old)	1-lead (MLII)	30 seconds	Native
2014	Park and Kang[18]	MIT-BIH arrhythmia database, Holter ECG monitoring data	17	47	1-lead (MLII)	30 mn	Native
2021	Ullah et al[39]	MIT-BIH arrhythmia database, PTB diagnostic ECG database	5 2	NR NR	NR	NR	Native, Augmented
2022	Irfan et al[10]	MIT-BIH arrhythmia database, UCI arrhythmia dataset	5 13	47 (25 F, 22 M), (44.9% M, 55.1% F)	2-lead, NR	30 mn NR	Augmented Native
2020	Shahin et al[40]	MIT-BIH arrhythmia database	5	47	2-lead	30 mn	Augmented
2022	Ma et al[41]	MIT-BIH arrhythmia database	5	47	NR	NR	Augmented
2021	Sabut et al[42]	CU ventricular tachyarrhythmia database, MIT-BIH malignant ventricular ectopy database	3, 3	NR	NR	8 minutes, 30 mn	Native
2020	Wang et al[43]	China Physiological Signal Challenge 2018 database, Computing in Cardiology Challenge 2017 database	9, 3	NR	12-lead, 1-lead	[6,60] seconds, ⩾9 seconds	Augmented
2022	Anbarasi et al[11]	MIT-BIH arrhythmia database, MIT-BIH NSR database, BIDMC database	2, 2, 2	10 18, 15	NR, NR, 2-lead	1 mn, NR, 20 hours	Native
2022	Zubair and Yoon[44]	MIT-BIH arrhythmia database	5	48	1-lead (MLII)	30 mn	Native

(Continued)

Table 2. (Continued)

YEAR	STUDY	DATABASES	NO. OF CLASSES	NO. OF PARTICIPANTS	ECG RECORDING FORMAT	SIGNAL DURATION	TYPE OF DATA
2022	Hu et al[45]	MIT-BIH arrhythmia database, MIT-BIH AF database	4, 8 2	NR	1-lead (MLII), 1-lead (MLII)	30 mn, 10 hours	Native
2022	Feyisa et al[46]	PTB-XL dataset	41, 20, 5	(52% M, 48% F)	12-lead	10 seconds	Native
2019	Ju et al[47]	MIT-BIH arrhythmia database	13	NR	NR	NR	Native
2022	Wang et al[48]	Computing in Cardiology Challenge 2017 database	4	NR	1-lead	⩾ 9 seconds	Augmented
2021	Wang[49]	MIT-BIH arrhythmia database, China Physiological Signal Challenge 2018 database	2, 2	NR	NR	30 mn, [6,60] seconds	Native
2021	Luo et al[50]	MIT-BIH AF database	9	NR	2-lead	30 mn	Augmented
2022	Iftene et al[51]	MIT-BIH arrhythmia database, PTB diagnostic ECG database	2, 2	47, 290	2-lead, NR	30 mn, NR	Augmented

Abbreviations: BIDMC, Beth Israel Deaconess Medical Center; CU, Creighton University; INCART, St. Petersburg Institute of Cardiological Techniques; MIT-BIH, Massachusetts Institute of Technology-Boston's Beth Israel Hospital; MLII, Modified Limb lead II; NR, Not Reported; NSR, Normal Sinus Rhythm; PTB, Physikalisch-Technische Bundesanstalt; UCI, University of California Irvine.

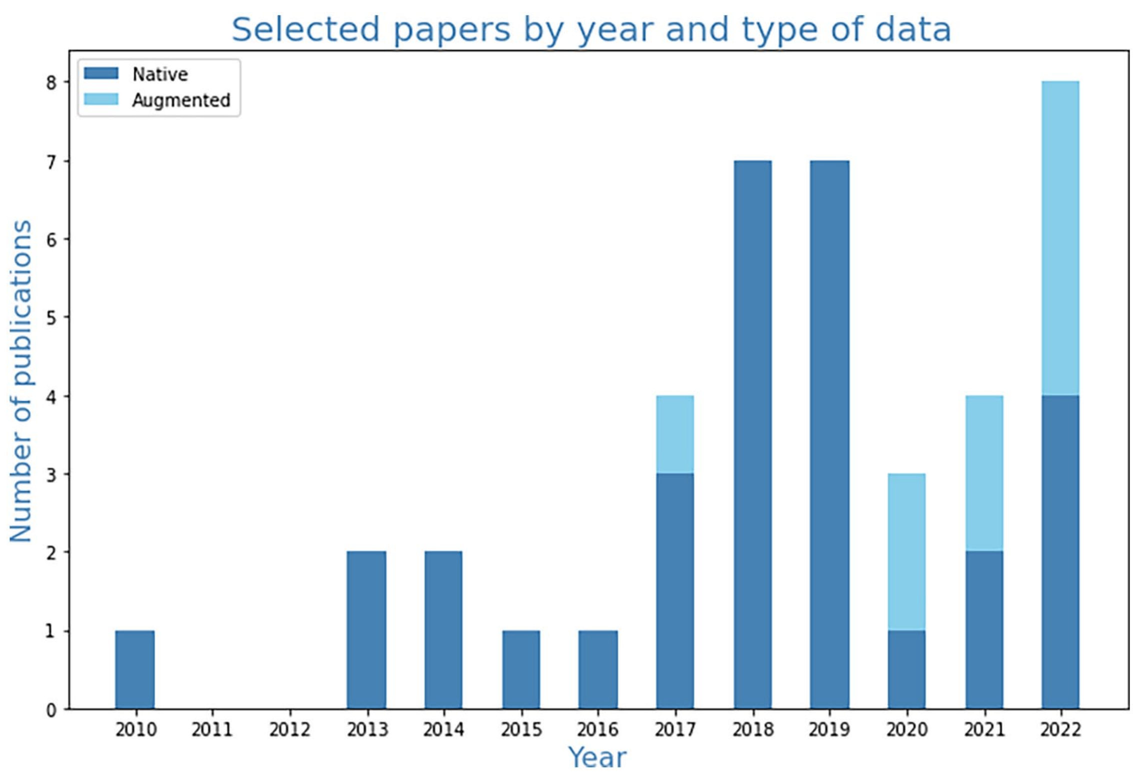

Figure 2. Selected publications from 2010 to 2022.

classification are embedded in one stage. However, the hand-crafted methods process the extraction of features independently from other learning stages.

Raw ECG signal is fed as input to models where no feature extraction phase is required; Table 3 reports the hand-crafted studies where feature extraction was realized. In addition, the

Table 3. Feature set.

YEAR	STUDY	EXTRACTION APPROACH/DATA INPUT	EXTRACTED FEATURES	METHODOLOGY
2019	Chen et al[19]	RR interval, DWT, FFT	2 time-domain 3 frequency-domain 5 space-domain	12-lead feature fusion
2019	Yildirim et al[21]	Raw ECG signal, Deep-coded ECG signals by CAE	NR	CAE for coded data
2019	Yang et al[22]	Dual-lead raw ECG signal, Triple-lead raw ECG signal	NR	DL-CCANet, TL-CCANet
2014	Sumathi et al[9]	Wavelet transform	5 space-domain	NR
2013	Martis et al[25]	HOS cumulants + PCA, DWT + HOS + PCA	12 nonlinear cumulants	NR
2016	Li et al[13]	PCA + KICA Db 2 DWT + PCA + LDA	20 nonlinear 4 frequency-domain	NR
2018	Anwar et al[26]	RR interval Discrete Meyer Wavelet Transform + ICA Teager energy operator	4 time-domain 12 time-frequency domain 1 time-domain	NR
2015	Elhaj et al[28]	Meyer DWT + PCA HOS cumulants + ICA	12 time-frequency domain 16 nonlinear	NR
2018	Raj and Ray[33]	Sparse decomposition	5 time-domain	Composite dictionary (CD): DOST, DST, DCT dictionaries
2017	Qin et al[34]	Biorthogonal 6.8 Wavelet multi-resolution analysis + PCA	12 time-frequency domain	NR
2017	Rajagopal and Ranganathan[35]	Db 4 DWT + PCA	12 time-frequency domain	NR
2010	Benali et al[36]	Raw data	1 space-domain 1 frequency-domain 3 time-domain	NR
2017	Rajesh and Dhuli[37]	Intrinsic mode functions (IMFs) decomposition	4 nonlinear 4 nonlinear	Ensemble empirical mode decomposition (EEMD) Empirical mode decomposition (EMD)
2018	Yang et al[38]	PCANet (CNN with PCA filters)	NR	NR
2014	Park and Kang[18]	Raw data	2 space-domain 4 time-domain	NR
2021	Sabut et al[42]	Data decomposition	24 time-frequency domain	Db6 DWT, EMD,VMD

Abbreviations: CAE, Convolutional Auto-encoder; CNN, Convolutional Neural Network; DCT, Discrete Cosine Transform; DL-CCANet, Dual-Lead Canonical Correlation Analysis Network; DOST, Discrete Orthogonal Stockwell Transform; DST, Discrete Sine Transform; DWT, Discrete Wavelet Transform; HOS, Higher-Order Spectra; ICA, Independent Component Analysis; KICA, Kernel-Independent Component Analysis; LDA, Linear Discriminant Analysis; NR, Not Reported; PCA, Principal Component Analysis; PCANet, Principal Component Analysis Network.

most employed technique for data extraction is discrete wavelet transform (DWT); 17% of the studies used this technique either separately as in literature[9,19] or with other extraction methods, as in literature,[13,25,26,28,34,35] such as principal component analysis (PCA) that was used in 15% of the studies. Other methods were used, such as Fast Fourier Transform (FFT) in Chen et al[19] and Higher-Order Spectra (HOS) in literature.[25,28] Other studies used personalized DL techniques for feature extraction. For instance, Yildirim et al[21] used a convolutional auto-encoder (CAE) while Yang et al[22] used a canonical correlation analysis network (CCANet) which combines canonical correlation analysis and cascaded convolution network, and Yang et al[38] used a principal component analysis network (PCANet) which is a convolutional neural network (CNN) with PCA filters. Wang et al[43] realized a multi-scale feature learning with CNN kernels to extract features from segments with different size.

Another type of signal decomposition is the intrinsic mode functions (IMFs) decomposition that can be characterized by empirical model decomposition (EMD) and ensemble empirical model decomposition (EEMD) as in Rajesh and Dhuli[37] or by variable model decomposition (VMD) as in Sabut et al.[42]

Depending on the used approach, the features may be related to time-, frequency-, or space-domain and can be linear or nonlinear features (Table 3). In addition, 4 studies reported the use of MATLAB software to realize the feature extraction phase where the other studies did not fill in this information.

Pre-processing and prediction methods

All selected studies used pre-processing methods to handle the data except Hannun et al[16] that proceeded directly to the classification. In the selected studies, we found out that some methods were used once for feature extraction and other times for data pre-processing. The pre-processing methods include noise removal, data segmentation, data normalization, data reduction, signal compression, and signal detection. Wavelet transform (WT) method, including different types of wavelets, was used for noise removal in literature[20,24,28,35,38] and with improved versions in literature.[13,23] Sumathi et al[9] used Symlet WT for QRS detection as shown in Table 4. The Pan–Tompkins algorithm proposed by Pan and Tompkins[52] was used for segmentation and QRS detection in literature[18,25,28] and for R-peak detection in literature.[20,33,35,38] More than half of the studies used data normalization. Some studies used ML methods for data processing as in Yildirim et al[21] and Liu[27] where they used, respectively, CAE for signal compression and SVM for QRS marking. Ullah et al[39] mentioned segmentation and pre-processing of data with no more details on the used techniques. Irfan et al[10] applied standardization of data (standard scalar unit) and feature reduction with PCA on the UCI dataset, and noise removal with DWT and normalization on the MIT-BIH arrhythmia dataset.

In addition, data padding was reported in Wang et al[43] as a processing operation to fix the input length.

Using continuous WT, Anbarasi et al[11] transformed 1-D signal to 2-D colored images to feed the CNN network. The transfer learning was introduced in Hu et al[45] to overcome the imbalance data problem.

Ju et al[47] proposed a bidirectional gated recurrent unit (GRU) network where the output is linked to the forward and backward states resulting in a better fit than unidirectional GRU and simpler structure than LSTM. To alleviate the issue of redundancy in bidirectional GRU, Wang[49] used an improved version of the aforementioned technique by adding a scale parameter to the model and combining it with CNN for feature extraction.

As shown in Table 4, the selected studies used several AI methods:

- ML methods: SVM, random forest, decision tree, feed-forward NN, residual NN, K-nearest neighbors.
- DL methods: CNN, long short-term memory (LSTM), GAN, GRU.
- Statistical AI methods: CCA, linear discriminant analysis.
- Artificial evolutionary algorithms: Genetic algorithm.
- Mathematics algorithms: Fuzzy logic, directed acyclic graph.

Some studies used the methods above either separately, combined, or in personalized view adapted to the application context to enhance the model performance. For instance, Sumathi et al[9] combined fuzzy logic with NN, and Ullah et al[39] combined CNN with LSTM, and Attention method which uses a weighted sum of all the encoder hidden states to flexibly focus the attention of the decoder to the most relevant parts of the input sequence. Feyisa et al[46] relied on a multi-receptive CNN where the receptive field can be obtained by either using multiple kernels of different sizes or using a fixed-size kernel with a varying dilation rate.

Most of the studies reported the use of k-fold cross-validation method for evaluation.

When there are various classes/categories in the dataset, the mentioned metrics refer to the overall performance on the ensemble of classes or databases. For more details, Table A1 (Appendix 1) shows different metrics for evaluation.

In the case of multi-class classification, we adopt averaging methods for some metrics calculation, resulting in a set of different average scores (macro, weighted, micro) in the classification report.

Discussion and Interpretation

In this review, we synthetize some literature studies addressing the ECG diagnostic approaches and the arrhythmia classification methods. We establish a comparison between the selected studies by discussing the following topics.

Used datasets and ECG signal information

The set of databases used in the selected studies is listed below. There are some studies that tested some of these databases separately or combined to provide high amount of data. The BIDMC database used in Anbarasi et al[11] for congestive heart failure was excluded from this analysis because we want to focus only on databases with arrhythmias.

Table 4. Pre-processing and prediction methods in the selected studies.

YEAR	STUDY	PRE-PROCESSING METHODS	PREDICTION METHODS	EVALUATION METHODS	OVERALL ACCURACY (%)
2019	Chen et al[19]	PCA for dimensionality reduction	Cascaded classifier composed of random forest and multilayer perceptron	NR	99.80
2017	Acharya et al[20]	Db 6 WT for noise removal, Pan–Tompkins algorithm for R-peak detection, ECG heartbeat segmentation, Z-score normalization	CNN	10-fold cross-validation	94.03
2019	Yildirim et al[21]	ECG heartbeat segmentation, CAE for ECG signal compression	Long short-term memory (LSTM)	NR	99.23
			CAE with LSTM	NR	99.11
2019	Yang et al[22]	ECG heartbeat segmentation, min-max normalization	SVM with DL-CCANet (MIT-BIH)	10-fold cross-validation	99.40
			SVM with DL-CCANet (INCART lead II and V1)	5-fold cross-validation	98.31
			SVM with DL-CCANet (INCART lead V1 and V5)	5-fold cross-validation	98.26
			SVM with DL-CCANet (INCART lead II and V5)	5-fold cross-validation	98.31
			SVM with TL-CCANet (INCART lead II, V1, and V5)	5-fold cross-validation	98.76
2018	Yildirim[23]	ECG heartbeat segmentation, Daubechies Wavelet sequence for multi-resolution analysis	Deep unidirectional LSTM-WS	NR	99.25
			Deep bidirectional LSTM-WS	NR	99.39
2014	Sumathi et al[9]	Noise removal, Symlet WT for QRS detection	Adaptive neuro-fuzzy inference system (ANFIS) model	NR	98.24
2019	Gao et al[24]	Db 6 DWT for noise removal, ECG heartbeat segmentation, Z-score normalization	LSTM with focal loss (noise-free data)	NR	99.26
			LSTM with focal loss (noisy data)	NR	99.07
2013	Martis et al[25]	Noise removal, Pan–Tompkins algorithm for QRS detection, ECG heartbeat segmentation	Feedforward NN (with HOS + PCA)	10-fold cross-validation	94.52
			Least-square SVM (with HOS + PCA)	10-fold cross-validation	94.30
			Feedforward NN (with DWT + HOS + PCA)	10-fold cross-validation	93.61
			Least-Square SVM (with DWT + HOS + PCA)	10-fold cross-validation	93.76
2016	Li et al[13]	Improved wavelet threshold method for noise removal	SVM + genetic algorithm (with MIT-BIH data)	NR	98.80
			SVM + genetic algorithm (with personalized ECG acquisition platform)	NR	97.30
2018	Anwar et al[26]	Noise removal, ECG heartbeat segmentation	Feedforward NN (18-class scheme)	3-fold cross-validation	99.75
			Feedforward NN (5-category scheme)	3-fold cross-validation	99.80
2013	Liu[27]	Noise removal, data normalization, SVM for QRS detection and marking	Self-constructing neural-fuzzy inference network (SoNFIN)	NR	96.40

(Continued)

Table 4. (Continued)

YEAR	STUDY	PRE-PROCESSING METHODS	PREDICTION METHODS	EVALUATION METHODS	OVERALL ACCURACY (%)
2015	Elhaj et al[28]	Db 6 DWT for noise removal, Pan–Tompkins algorithm for QRS detection, ECG segmentation	Feed-forward NN	10-fold cross-validation	98.90
			SVM with RBF kernel	10-fold cross-validation	98.91
2019	Kim et al[29]	ECG heartbeat segmentation	GoogleNet deep NN with 1-inception	NR	95.30
			GoogleNet deep NN with 2-inception	NR	96.30
			GoogleNet deep NN with 1-inception + CNN	NR	95.90
2018	Yıldırım[30]	Data normalization	1-D CNN (13-class scheme)	NR	95.20
			1-D CNN (15-class scheme)	NR	92.51
			1-D CNN (17-class scheme)	NR	91.33
2018	Oh et al[31]	Heterogeneous ECG segmentation, Z-score normalization	Modified U-net architecture	10-fold cross-validation	97.32
2018	Oh et al[32]	ECG segmentation, Z-score normalization	CNN + LSTM (without dropout regularization)	10-fold cross-validation	98.42
			CNN + LSTM (2-dropout)	10-fold cross-validation	97.88
			CNN + LSTM (3-dropout)	10-fold cross-validation	98.10
2018	Raj and Ray[33]	Noise removal, Pan–Tompkins algorithm for R-peak detection, ECG heartbeat segmentation	ABC-DAG-LSTSVMs classifier with 16-class scheme	14-fold cross-validation	99.21
			ABC-DAG-LSTSVMs classifier with 5-class scheme	22-fold cross-validation	90.08
2020	Ribeiro et al[15]	Vector cardiogram linear transformation for dimensionality reduction	Unidimensional residual NN	NR	92.55
2017	Qin et al[34]	ECG heartbeat segmentation	One-vs-one SVM (beat-based scheme)	10-fold cross-validation	99.70
			One-vs-one SVM (record-based scheme)	10-fold cross-validation	81.47
2017	Rajagopal and Ranganathan[35]	Db 8 DWT for noise removal, Pan–Tompkins algorithm for R-peak detection, ECG heartbeat segmentation	K-nearest neighbors + SVM	10-fold cross-validation	99.78
2010	Benali et al[36]	Noise removal, QRS detection (original algorithm by GBM laboratory at Tlemcen university)	Wavelet neural network (WNN)	NR	98.78
2017	Rajesh and Dhuli[37]	Noise removal, ECG heartbeat segmentation	Sequential minimal optimization (SMO)–SVM (cubic kernel) with EMD for MIT-BIH data	10-fold cross-validation	99.20
			SMO–SVM (RBF kernel) with EEMD for MIT-BIH data	10-fold cross-validation	96.45
			SMO–SVM (cubic kernel) with EEMD for INCART data	10-fold cross-validation	97.57

(Continued)

Table 4. (Continued)

YEAR	STUDY	PRE-PROCESSING METHODS	PREDICTION METHODS	EVALUATION METHODS	OVERALL ACCURACY (%)
2018	Yang et al[38]	Db 8 WT for noise removal, Pan–Tompkins algorithm for R-peak detection, ECG heartbeat segmentation, min-max normalization	Linear SVM, K-nearest neighbors, Random Forest, Backpropagation NN (Noisy data)	10-fold cross-validation	97.77 97.10 96.01 96.95
			Linear SVM, K-nearest neighbors, Random Forest, Backpropagation NN (Noise-free data)	10-fold cross-validation	97.08 96.27 95.22 95.89
2019	Hannun et al[16]	NR	Deep CNN with sequence level, Deep CNN with set level	NR	97.8 97.7
2014	Park and Kang[18]	Pan–Tompkins algorithm for QRS detection	Decision tree with J4.8 algorithm (personalized scheme)	10-fold cross-validation	85.26
			Decision tree with J4.8 algorithm (non-personalized scheme)	10-fold cross-validation	89.95
2021	Ullah et al[39]	ECG heartbeat segmentation	CNN	NR	99.12
			CNN + LSTM	NR	99.3
			CNN + LSTM + attention method	NR	99.29
2022	Irfan et al[10]	DWT for noise removal, data standardization and normalization, PCA for feature reduction	CNN + LSTM with MIT-BIH database	NR	99.35
			CNN + LSTM with UCI arrhythmia dataset	NR	99.05
2020	Shahin et al[40]	ECG segmentation, Z-score normalization	Multi-task adversarial network	NR	86
2022	Ma et al[41]	Db 6 WT for noise removal, Pan–Tompkins algorithm for R-peak detection, ECG heartbeat segmentation	ResNet + Bi-LSTM + attention method	NR	99.4
2021	Sabut et al[42]	Noise removal, Z-score normalization, heartbeat segmentation with 5s window	Deep NN	NR	99.2
2020	Wang et al[43]	Data normalization	Multi-scale fusion CNN	5-fold cross-validation	NR
2022	Anbarasi et al[11]	CWT for noise removal, ECG segmentation	Combined CNN and LSTM	10-fold cross-validation	98.7
2022	Zubair and Yoon[44]	Noise removal, Pan–Tompkins algorithm for peaks detection, ECG heartbeat segmentation	CNN for inter-patient classification	10-fold cross-validation	96.36
			CNN for intra-patient classification	10-fold cross-validation	99.81
2022	Hu et al[45]	Z-score normalization, wavelet, and Pan–Tompkins for QRS detection, heartbeat segmentation	Transformer-based CNN for 8 classes	10-fold cross-validation	99.12
			Transformer-based CNN for 4 classes	10-fold cross-validation	99.49
			Transformer-based CNN for 2 classes	10-fold cross-validation	99.23
2022	Feyisa et al[46]	Standard normalization, 2.5-s segmentation	Multi-receptive field CNN for 41 classes	NR	98
			Multi-receptive field CNN for 20 classes	NR	96.2
			Multi-receptive field CNN for 5 classes	NR	89.7

(Continued)

Table 4. (Continued)

YEAR	STUDY	PRE-PROCESSING METHODS	PREDICTION METHODS	EVALUATION METHODS	OVERALL ACCURACY (%)
2019	Ju et al[47]	Noise removal, dimensionality reduction with PCA	Deep bidirectional GRU network	5-fold cross-validation	99.51
2022	Wang et al[48]	NR	Dual-path recurrent neural network (RNN)	5-fold cross-validation	84.5
2021	Wang[49]	2-s segmentation	CNN + improved BGRU for MIT-BIH data	NR	97.9
			CNN + improved BGRU for CPSC data	NR	98.3
2021	Luo et al[50]	0.65-s segmentation, Z-score normalization	Hybrid convolutional RNN	10-fold cross-validation	99.01
2022	Iftene et al[51]	PQRST detection, data segmentation, data normalization	1-D CNN with pre-processing	NR	98
			1-D CNN	NR	95
			Bayesian NN	NR	90
			GRU network	NR	94

Abbreviations: BGRU, Bidirectional Gated Recurrent Unit; CPSC, China Physiological Signal Challenge; CNN, Convolutional Neural Network; CWT, Continuous Wavelet Transform; DL-CCANet, Dual-Lead Canonical Correlation Analysis Network; DWT, Discrete Wavelet Transform; EEMD, Ensemble Empirical Mode Decomposition; EMD, Empirical Mode Decomposition; GBM, Génie Bio-médical; GRU, Gated Recurrent Unit; HOS, Higher-Order Spectra; INCART, St. Petersburg Institute of Cardiological Techniques; LSTM, Long Short-Term Memory; LSTM-WS, Long Short-Term Memory Wavelet Sequence; MIT-BIH, Massachusetts Institute of Technology-Boston's Beth Israel Hospital; NN, Neural Network; NR, Not Reported; RBF, Radial Basis Function; SVM, Support Vector Machine; UCI, University of California Irvine.

- Open access databases: MIT-BIH databases, QT database, INCART database, PTB diagnostic ECG database, CU ventricular tachyarrhythmia database, Computing in Cardiology Challenge 2017, China Physiological Signal Challenge 2018, PTB-XL dataset. More details about these databases are found on PhysioNet Bank.[12]
- UCI arrhythmia dataset: An open access database available on the ML repository of the UCI university.[53]
- Telehealth Network of Minas Gerais (TNMG) dataset: Data collected under the scope of the CODE (Clinical Outcomes in Digital Electrocardiology) study in the Telehealth Network of Minas Gerais which is a public telehealth system in Minas Gerais, Brazil. Publicly available on TNMG dataset.[54]
- ProSim simulator dataset: An industry-leading patient simulator for monitoring and preventive testing, developed by Fluke Biomedical. It is a commercial paid solution.[14]
- Zio monitor dataset: Non-free ambulatory monitoring solution developed by iRhythm Technologies Inc, San Francisco, CA. The solution provides FDA-cleared, single-lead, patch-based ECG monitor that continuously records data from a single vector, the recording can be up to 14 days.[17]
- Holter monitor dataset: Private data collected from wearable device which records heartbeats for diagnosis. It is a noninvasive solution that can be worn up to 2 days.

As shown in Table 2, 33 out of 40 studies used the MIT-BIH arrhythmia database.

To diagnose arrhythmia, the studies relied on the multi-class prediction. Most of the studies predicted the occurrence of more than 2 types of the arrhythmic heartbeat, yet not all of the studies using 12-class prediction and more recorded the highest performances as in literature.[16,18,22,26,30,33] This can be explained by the imbalance of datasets; some heartbeat types have a small number of records which affect negatively the classification rate.

However, 33 out of the 40 selected studies performed the classification with input signal equal to 30 minutes' length (Table 2). Only 1 study used variable length duration,[31] and a special U-net architecture was developed for this purpose to handle the variable-size data.

As of the ECG recording format, only literature[15,19,43,46] used 12-lead ECG signal which is the standard technique in the real clinical settings. Although literature[15,16] provided the largest datasets among all the selected studies which can improve the model ability of generalization, they did not reach the highest performances due to the imbalance of data. Also, Hannun et al[16] did not apply any pre-processing methods on the data which can increase the error rate.

Data augmentation is used to tackle the issue of data imbalance. While some techniques help to mitigate the over-fitting in the training stage, such as SMOTE technique and

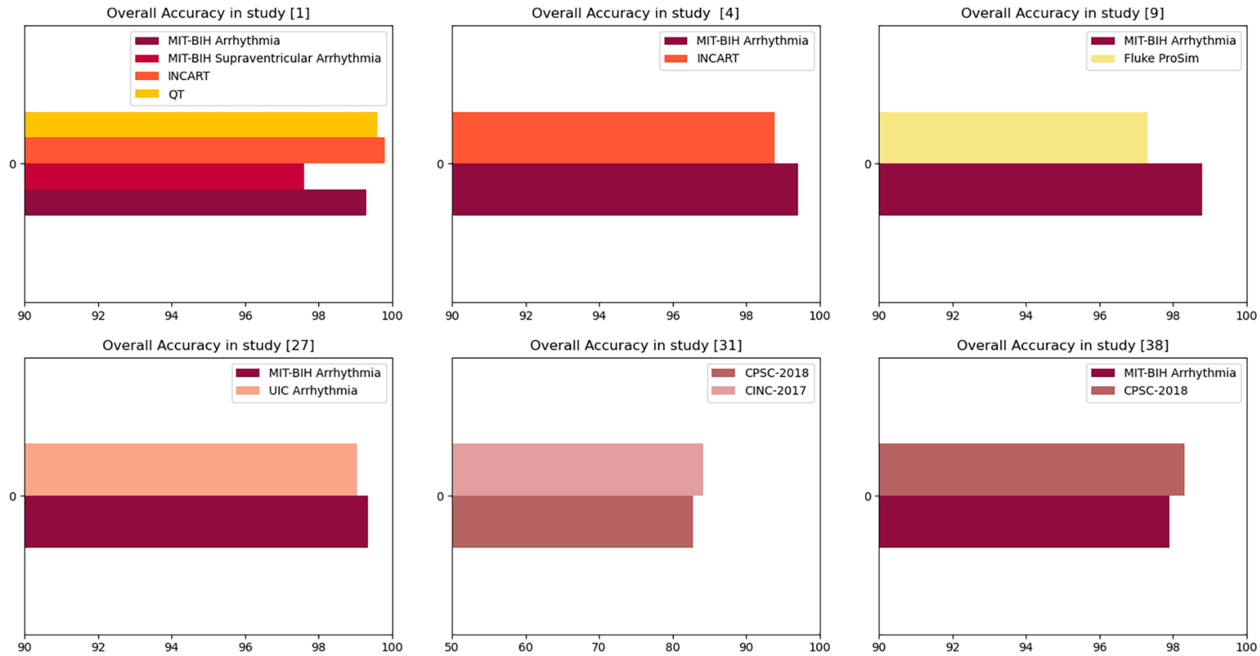

Figure 3. Overall accuracy recorded in some datasets.

GAN network, other methods allow only to increase the volume of data without having a measurable effect on the performance and the variance of the dataset since they rely on a simple resampling or the addition of Gaussian noise and interpolation as in Iftene et al.[51]

It is logical to analyze the use of different datasets in the same study since it used the same pre-processing and prediction methods. The comparison of the same database used in different studies will not be relevant.

Figure 3 shows that INCART database in Chen et al[19] reached the highest accuracy among the other databases, given that all of them were imbalanced. This good performance can be explained by the fact that INCART is 12-lead and the study combined features from all these leads to ensure classification. However, Yang et al[22] showed better accuracy for MIT-BIH database than INCART, from which were extracted only 2 leads (II and V1). Irfan et al[10] and Wang[49] recorded better results on MIT-BIH database because the other databases were highly imbalanced.

Taking everything discussed above into account, we assume that:

- MIT-BIH still is one of the best and most complete databases used in arrhythmia classification as it provides annotations, signal characteristics, and different lead recordings.
- The combination of 12-leads can help increasing the accuracy because the model will be fed with various information.
- It is essential to tackle the imbalance data issue because it can hinder good pre-processing and prediction techniques from achieving higher performances. SMOTE technique is recommended for this end.

Feature selection and extraction

Table 5 indicates the types of some extracted features from the selected studies. The most used features are RR intervals, which represent the time-domain, and the amplitude of R wave, which represents the space-domain. Moreover, the WT and the PCA methods are the most used in the feature extraction stage, given that PCA provides low-dimension features while preserving as much of the data variation as possible and WT allows to capture both frequency and time information.

Sabut et al[42] extracted various features having temporal, statistical, and spectral information, such as filter leakage measure, covariance, kurtosis, skewness, threshold crossing interval, Shannon Entropy, etc to improve the accuracy of classification.

For the studies based on DL models, such as CNN as in literature,[39,40,41,45] the extraction is held by the DL model itself, by sliding multiple convolutional windows over the ECG and performing multiple convolutional operations on the local features.

There is no doubt that feature extraction allows a better understanding of the model as it helps setting an explicit feature design of the ML model but when it is embedded in the DL model, it decreases the consumption of resources and time. For instance, we can rely on the strength of CNN for dealing with the extraction stage even if CNN can be time-consuming when a high number of layers are used.

Pre-processing methods

According to Table 4, only 1 study was not subject to data pre-processing. The most used techniques are ECG heartbeat segmentation (17 studies), noise removal (13 studies), data

Table 5. Types of extracted data.

DOMAIN	FEATURE	STUDIES
Frequency	Signal phase angle of the FFT	Chen et al[19]
	Signal wave power spectrum of the FFT	Chen et al[19]
	DWT frequency	Chen et al[19]
	Mean of wavelet coefficient	Li et al[13]
	Min of wavelet coefficient	Li et al[13]
	Max of wavelet coefficient	Li et al[13]
	SD of wavelet coefficient	Li et al[13]
	QRS duration	Benali et al[36]
Time	Skewness of RR intervals	Chen et al[19]
	Kurtosis of RR intervals	Chen et al[19] and Raj and Ray[33]
	SD of RR intervals	Raj and Ray[33]
	Interval RR	Anwar et al[26],Raj and Ray[33], Elhaj et al[28] and Park and Kang[18]
	Interval PR	Park and Kang[18]
	Position of R point	Park and Kang[18]
	Position of P point	Park and Kang[18]
	Local RR interval	Anwar et al[26] and Raj and Ray[33]
	Average RR interval	Anwar et al[26] and Raj and Ray[33]
	Energy	Anwar et al[26] and Raj and Ray[33]
	Ratio between the distance RR following the previous one	Benali et al[36]
Space	Amplitude of R wave	Chen et al[19], Sumathi et al[9], Benali et al[36] and Park and Kang[18]
	Amplitude of Q wave	Sumathi et al[9]
	Amplitude of P wave	Park and Kang[18]
	Amplitude of S wave	Sumathi et al[9]
	Amplitude of K1	Sumathi et al[9]
	Amplitude of K2	Sumathi et al[9]
	Highest voltage value	Chen et al[19]
	Lowest voltage value	Chen et al[19]
	Average amplitude	Chen et al[19]
	Variance of amplitudes	Chen et al[19]
Nonlinear	Variance	Chen et al[19]
	Permutation entropy	Raj and Ray[33]

Abbreviations: DWT, Discrete Wavelet Transform; SD, Standard Deviation.

normalization (8 studies), and QRS detection (6 studies). However, 4 studies relied on R-peak detection and this detection reached an accuracy of 99.3% in Oh et al.[31] Furthermore, the most used algorithms in the pre-processing phase are the Pan–Tompkins algorithm to detect accurately R peaks and QRS complexes, and the WT to reduce the cost of continuous wavelet computation.

Table 6 below summarizes the pre-processing methods, their application, and the objective from their usage.

The ECG signal segmentation is applied with different sample-long segments that vary between 100- and 500-sample long. The samples are centered either around the detected R peaks or the detected QRS complexes. The segmentation can

Table 6. Pre-processing methods and their applications.

PRE-PROCESSING METHOD	APPLICATION/TYPE	OBJECTIVE	TECHNIQUE
Signal segmentation	[100, 140, 150, 200, 250, 252, 256, 260, 300, 360, 400, 500] sample long.	• Infer the hidden states of signal at each time, • Subsequent signal classification.	Annotated R peaks, annotated QRS complexes, cardiologists' annotations.
Noise removal	Power line interference, muscle noise, motion artifact, baseline wander, high-frequency artifacts.	• Improve the interpretability and perception of multi-dimension information, • Reduce the probability of error in QRS detection, • Enhance the classification accuracy.	DWT with its different distributions, band-pass filter, median filter, low-pass filter, mathematics equations.
Data normalization	Signal rescaling	• Eliminate the offset effect, • Standardize the ECG signal amplitude, • Improve the backpropagation process by speeding up the convergence rate.	Min-max normalization, Z-score normalization
QRS detection	QRS mid-point, RR markers	• Subsequent rhythm classification, • Identify features characteristics.	Pan–Tompkins algorithm, SVMs, Symlet WT, original algorithm by GBM at Tlemcen University.
R-peak detection	R point recognition	• Facilitate features' extraction	Pan–Tompkins algorithm
Signal compression	Segmented ECG data	• Reduce the signal size of beats with the minimum loss, • Reduce the storage cost of the large amount of data.	Deep CAE
Dimensionality reduction	Feature space reduction	• Reduce the overhead of computing, • Improve accuracy.	PCA, vector cardiogram linear transformation.

Abbreviation: GBM, Génie Bio-médical.

also rely on the extraction of T-to-T segments as in Zubair and Yoon,[44] or can simply rely on the database annotation files.

Noise removal method is applied to remove different types of noise that can result from patient motion or respiration, power line interference, muscle artifacts, baseline drift, electrode motion artifact or data-collecting device noise. To the fact that each noise source resides in a characteristic frequency band, different filters and techniques are used depending on the type of noise.

Data normalization can be considered one of the most interesting methods due to its important influence on the classification process. Namely, signal rescaling improves significantly the backpropagation process by speeding up the convergence rate.

Most of the studies where pre-processing was applied to data showed a better performance on the classification as in Iftene et al[51] where CNN model reached an accuracy of 95% without pre-processing vs 98% when applying data augmentation and normalization.

The other pre-processing methods used in the selected studies are shown in Table 6.

To show the correlation between the use of pre-processing methods and the obtaining of better accuracies, we plot the performance corresponding to different pre-processing methods.

We compare between noisy data and noise-free data. Figure 4 shows that when cleaning data from noise, better accuracy can be obtained. Yet, Yang et al[38] demonstrated its ability to detect successfully noisy heartbeats with different ML methods. This is due to the use of PCA filters when extracting features, which can remove implicitly unwanted noise.

We notice that all the studies which recorded accuracies lower than 90% did not undergo noise removal.

As result, we affirm that the use of one or many combined pre-processing methods can decrease the error rate. It is highly recommended to realize data augmentation and noise removal to avoid misclassifications and improve the detection ability of arrhythmia.

Prediction methods

As shown in Table 4, CNN and LSTM network are the most used techniques, followed by the SVM in 22% of the studies. Indeed, more than half of the studies used DL techniques to improve the accuracy. CNNs are used with different variants in the convolutional blocks as in literature.[44-46]

Some studies reported the computation time in the learning phase; Yang et al[22] recorded the lowest training time with a value equal to 68.8 seconds using leads II and V1. The use of CCANet in the feature extraction phase has definitely reduced the computation cost and improved the accuracy and specificity which reached, respectively, 99.4% and 99.6%.

Shahin et al[40] reported a very interesting DL technique; the architecture of the adversarial multi-task model consists of 3 networks: the generator network, the heartbeat-type discriminator which discriminates between 5 types of heartbeats, and the subject discriminator which discriminates between 39 different subjects. This design has increased the performance allowing double discrimination and forcing the system to take into account only the heartbeat variations. Yet, it can be improved by changing the method of synthetic data generation; generating new data with GAN network or SMOTE technique instead of upsampling which generates duplicated data.

Sabut et al[42] used a fusion of 2 CNN branches with different scales and an Attention module to mine the discriminative features. In fact, the attention mechanism boosts the classification performance as shown in Hu et al[45] where the attention helped to capture the inter-beat dependencies.

The combination of residual convolutional blocks and bidirectional LSTM model with Attention method in Ma et al[41] seems to be effective since it allows a local and global feature extraction, and high accuracy that reached 99.4%. Zubair and Yoon[44] mitigated the problem of imbalanced data in CNNs by designing a novel cost-sensitive loss function in the network. This learning strategy is based on training efficiently the model without changing the distribution of the data. Besides, the aforementioned study highlighted the use of 2 different paradigms: the intra-patient and inter-patient classifications to show how the latter achieves better generalization capability.

Luo et al[50] used a hybrid model combining CNN layers, LSTM, and GRU networks. Indeed, the authors took advantage of every network's strength: the high ability of temporal and spatial information extraction of CNN, acquiring sequential information by LSTM, retaining only relevant information by the GRU, and avoiding the gradient disappearance issue.

As for the development tools, Python was used with its different ML libraries, such as TensorFlow, PyTorch, Scikit-learn, and Keras. MATLAB is also employed in some studies. Iftene et al[51] developed the prediction technique in the Amazon Web Services platform using an integrated DL model.

We gather the prediction methods used in this review in the scheme below (Figure 5).

General AI can be divided into 2 categories:

- **Symbolic AI** which is based on a system of "rules," the machine therefore does not improvise by itself, it acts according to the rules it has received. One of the most important algorithms in symbolic AI is the genetic algorithm used in Li et al.[13]
- **ML** is a form of AI where based on more data and computers can learn without being explicitly programmed to do so instead of programmers teaching the machines what tasks they need to perform.

We visualize the accuracy of CNN and SVM networks in Figure 6.

As it can be shown, the average accuracy over all the studies that used CNN is 97.82% vs 98.41% for SVM. When running through literature, we find that SVMs when preceded with feature extraction stage can achieve promising results. The selected studies in this review used PCA filters, DWT, and convolutional layers for the extraction which definitely have boosted the SVM performance.

Taking everything into consideration:

- The CNN and LSTM are the most used techniques in the last years; they allow the extraction of temporal, spatial, and sequential information from the ECG signal and they analyze deeply the extracted features which result in high accuracy.
- The attention method can boost the classification performance and the generalization ability.
- Pre-processing combined with DL techniques can help achieving promising results.
- To achieve high performance, DL methods need high-performance computing.
- When high computational resources are not provided, the use of SVM can be a good alternative for arrhythmia detection but should be preceded with relevant feature extraction methods which can be time-consuming.

Evaluation and performances

First, we compare the studies that used the same datasets. We sort 7 different datasets available on the PhysioNet repository:[23] MIT-BIH arrhythmia database, QT database, MIT-BIH supraventricular arrhythmia database (SV), INCART, MIT-BIH atrial arrhythmia database, Malignant ventricular arrhythmia database, and PTB diagnostic ECG database, besides 4 non-open datasets collected during studies[13,15,16,18] either from simulation devices or Holter monitor or ECG recorder as indicated in Table 2. However, 2 studies[19,26] used SV database, Anwar et al[26] combined the data from both MIT-BIH arrhythmia database (MIT-BIH) and SV database to apply class-oriented prediction (based on different sub-categories of beats depending on the used datasets) with 18 classes and subject-oriented prediction (main category classification with beat's annotation according to ANSI/AAMI standard) with 5 classes while Chen et al[19] applied class-oriented scheme with 4 classes. As shown in Figure 4, the model combining MIT-BIH and SV databases achieved a high accuracy of 99.8%, whereas the model relying only on SV database achieved an accuracy of 97.6% (Figure 7). Nevertheless, both the results were promising.

Studies[19,22,37] enrolled the learning phase using INCART database for 4-class, 7-class, and 5-class prediction, respectively. The highest accuracy of 99.8% was achieved by the first study where they used 12-lead ECG recording format vs an accuracy of 98.76% with 2-lead format and 97.57% for the 2 other studies. For the rest of the studies where they used MIT-BIH arrhythmia database, all of them reached an accuracy above 90%.

Second, we carry on a comparison between studies using the same prediction methods. Regarding the studies applying SVM,

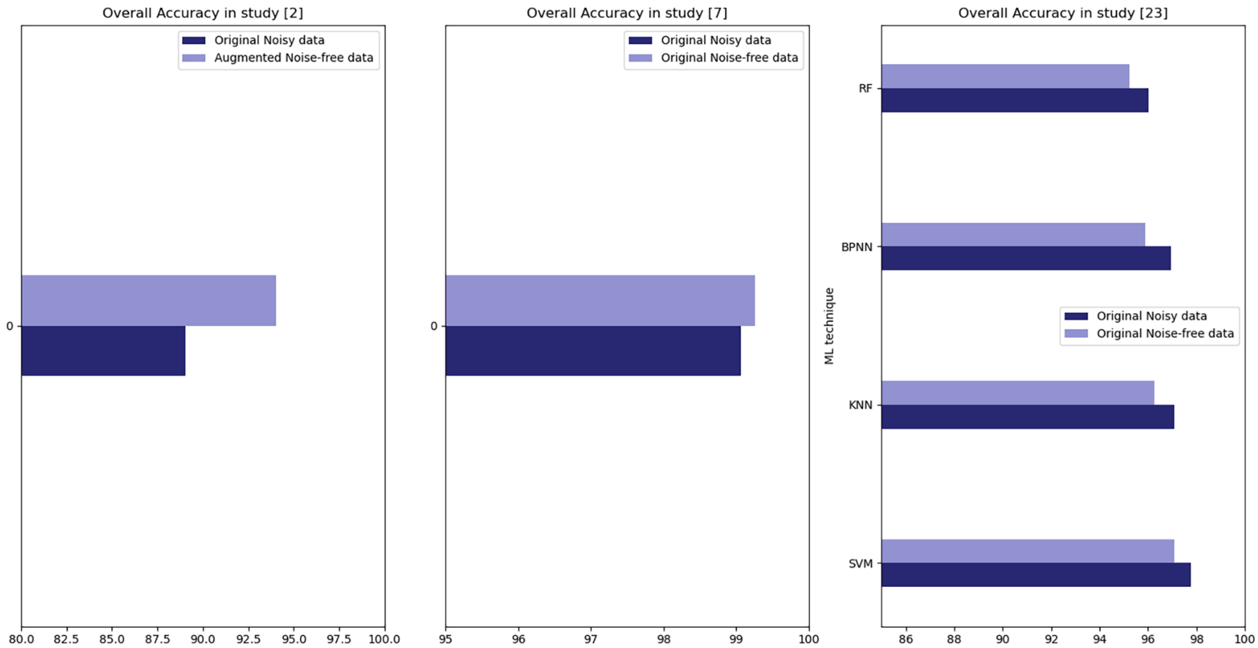

Figure 4. Accuracy of noisy data vs noise-free data.

they used different kernel functions and some of them were combined with other ML algorithms, but most of them yielded an accuracy greater than 94%. This can be explained by the powerful methods used for feature extraction and data pre-processing, including the use of DL techniques,[20,22,38] in these studies. When comparing the studies that applied CNN, all of them attained high accuracy rates above 94%. The lowest metrics (accuracy, specificity, and sensitivity) were obtained by Qin et al[34] that performed SVM on record-based training scheme where the classifier was trained and tested on separate records from different individuals.

Regarding the studies with smaller signal durations (between 7 and 30 seconds), they achieved good F1-score values but the highest scores were obtained by 30-minute duration studies. And yet, the increase in ECG signal length does not guarantee the highest accuracy rates. Indeed, in this review, the studies with the lowest signal duration[15,16,30] could perform better, especially when they applied deep CNNs; however, they either did not proceed data pre-processing[16] or they used imbalanced data[30] for the classification.

In Irfan et al,[10] the DL model achieved better results on the second dataset, this is due to the highly imbalanced data in the first dataset. Only the accuracy of the best model was reported in Table 4 (an overall accuracy of 99.35% for balanced data vs 93.33% for imbalanced data).

For Shahin et al,[40] the adversarial multi-task model achieved an overall accuracy of 86% on the validation set and 87% on the test set, which are lower comparing to other techniques, due to the imbalanced data.

Zubair and Yoon[44] achieved a high accuracy of 99.81% in the intra-patient paradigm with CNN with different size kernels and cost-sensitive function. Hu et al[45] reached an accuracy of 99.49% for 4 class-categorization with transformer encoder–decoder network with CNN layers and attention mechanism. The use of CNN with different kernel sizes (to capture different segment and interval lengths) allowed to obtain an accuracy of 98% for 41 arrhythmias classification.

Wang[49] used a novel method of premature ventricular contraction (PVC) detection where they modified a GRU network to avoid the redundancy of information in the forward and backward connections. This improved version of GRU yielded an accuracy of 97.9% on MIT-BIH data and 98.3% on CPDB.

Most of methods relying on DL, ML, statistical AI techniques, or a combination of them had performed high accuracies because all of the selected studies in this review realized rigorously the feature extraction phase and the pre-processing phase.

Among the studies selected, there are many that have used variety of approaches/databases/methods. Depending on each criterion, we linked the use of pre-processing and prediction methods to the accuracies which they are shown in Table 4.

Contributions and comparison to other literature reviews results

We compare our review to other review papers in literature that focus on reviewing studies with ML methods for arrhythmia classification.

Some papers focused only on describing the DL techniques and neglected the effect of the pre-processing stage and the type of datasets on the performance as in review[55] which conducted a shallow description of the papers. Unlike Ebrahimi et al,[56] where they realized a well-organized overview to the existing papers in literature starting from 2017. Yet, they basically selected papers using the public PhysioNet databases which can be useful when producing and comparing works between

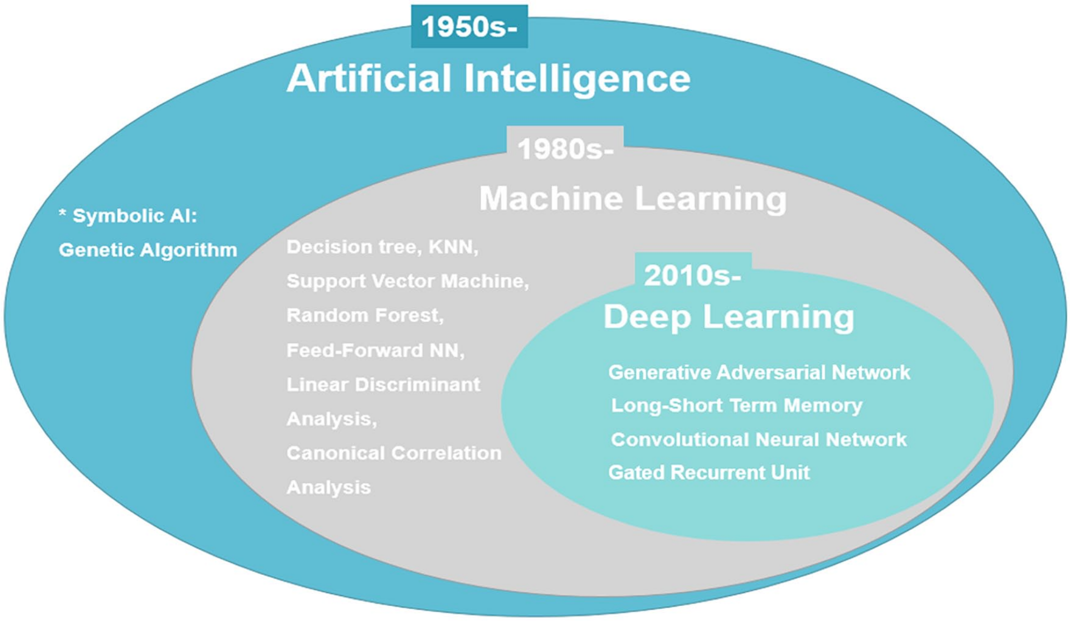

Figure 5. Used AI methods.

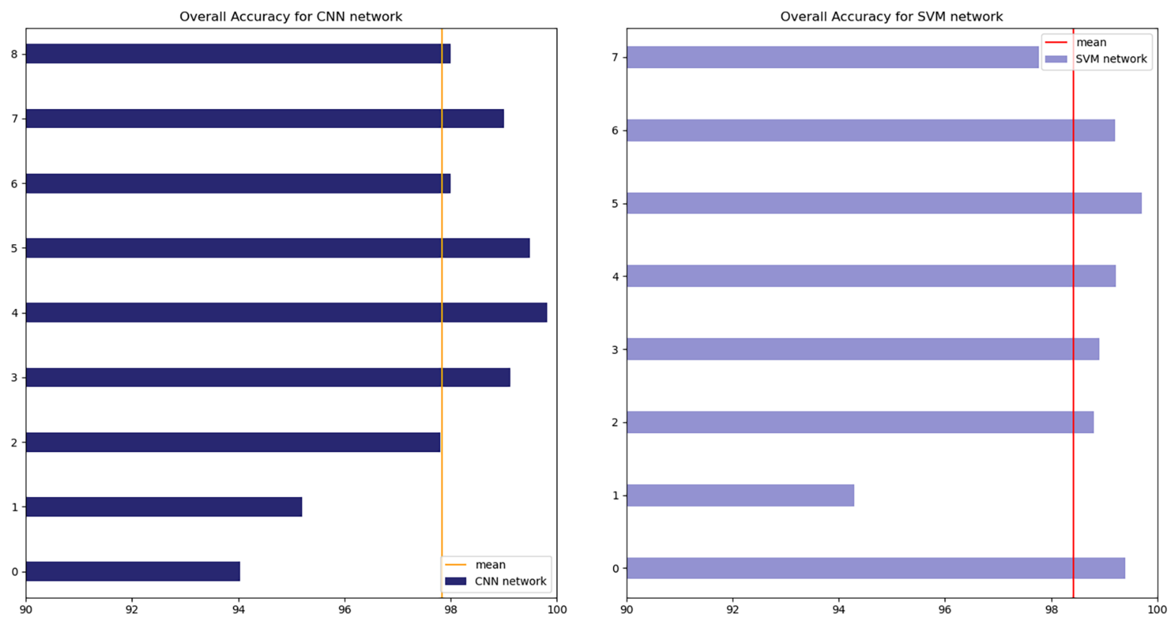

Figure 6. Accuracy of CNN and SVM models.
CNN, Convolutional Neural Network; SVM, Support Vector Machine.

researchers, but it neglects the DL performance that can be recorded on wearable monitoring devices. In the same review, they presented papers that used variants of GRU, RNN, and CNN: models with very promising results in the literature.

One of the strong points of Annam et al[57] is the presentation of the inter-patient vs the intra-patient paradigms in heartbeat classification with both DL and ML techniques. However, they did not discuss the pre-processing methods used in the selected papers. Jensen et al[58] and Tamariz et al[59] focused on the study of papers handling the validation of, respectively, AF occurrence and ventricular arrhythmias while focusing on the validation metrics and the used datasets without dedicating special attention to the classification methods which were reported as administrative codes. Jensen et al[58] reported only 16 studies which can question the relevance of this review. Sanamdikar[60] reported feature extraction, pre-processing, and prediction techniques for arrhythmia classification with description of the used databases. However, the review was limited in terms of the reported techniques, especially for the pre-processing where they mentioned only noise removal.

One of the most interesting reviews in literature is Luz et al.[61] They relied on a good search strategy and succeeded to report information about used databases, feature extraction, pre-processing, and prediction methods.

Parvaneh et al[62] presented an overview on arrhythmia detection with respect to the following aspects: used datasets, type of

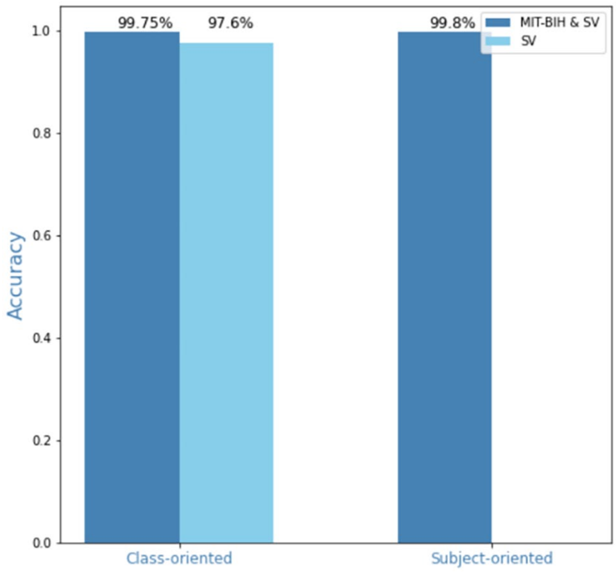

Figure 7. Average classification accuracy on SV database. SV, supraventricular arrhythmia database.

input data, model architecture, and evaluation metrics. Due to the shallow analysis of the selected papers, this review is considered to be conceptual. The DL architecture and feature extraction were briefly stated. Another gap is the absence of pre-processing methods which should be discussed because they affect the performance of the DL model. However, Houssein et al[63] focused on the studies related to arrhythmia classification by artificial neural network (ANN) and SVMs. The review presented the 3 main stages prior to classification: pre-processing, feature extraction, and feature selection. Detailed information about every phase was given while relating to the used methods in every study. Thus, this can be an interesting review for reference but since they focused only on 2 models, ANN and SVM, more papers should have been included to the analysis.

The strengths of our review can be mentioned as follows:

- We presented the search method and the inclusion criteria that we rely on, to select the studies analyzed in our review.
- We reported and analyzed papers using either DL or ML or both, to emphasize the good performance that can be reached when combining different techniques. Moreover, we want to provide the reader with other alternatives when high computing resources are not provided.
- We established a deep description of the papers; exploration of used datasets, feature extraction, pre-processing, prediction methods, and performance evaluation.
- We pointed out the advantages and limitations of the used methods.
- We analyzed the relationship between the high performance and the use of pre-processing methods, especially noise removal and data augmentation which help avoiding misclassifications.

We believe that this review can help in defining the scope of future research work when planning to apply ML or DL techniques for arrhythmia classification to given datasets.

In the future, we plan to follow up this literature by the developing an AI model to classify ECG heartbeats and predict the occurrence of arrhythmia.

Conclusions

This review synthetizes and interprets some of the papers in the literature that deal with arrhythmia detection using ECG-based models.

Taking everything into account, we summarize the findings of this review as follows:

- The selected studies relied on a multi-class prediction of arrhythmia with no other cardiovascular disease diagnosis, to keep the focus on the irregularity of heartbeat types related to the arrhythmic aspect. Most of the studies used a 30-minute signal length and a single- or dual-lead ECG recording format.
- ECG heartbeat segmentation relies on the signal sliding based on the position of R peaks with equal-size segment. Therefore, variable-size segments should be used more frequently, especially when detecting arrhythmias to capture the intra-beat and inter-beat irregularities.
- Most of databases contain imbalanced data which result in heartbeat misclassification for the minority classes. Therefore, strong methods for data augmentation should be used as SMOTE or GAN network.
- It is found that the use of data augmentation technique is proportional to the use of DL techniques which need balanced data to emphasize their performance.
- The most performing models used arrhythmia databases from the PhysioNet repository mainly the MIT-BIH databases because they are properly annotated and organized. Moreover, the most used features were RR intervals and the amplitude of R waves which indicate the importance of these time-domain and space-domain features, respectively, in the prediction of arrhythmia.
- Overall, 96% of the selected studies applied pre-processing methods among which there are noise removal, normalization, and QRS detection. These methods demonstrated their efficiency in decreasing the computing cost and increasing the accuracy rate.
- All selected studies used either ML techniques or DL techniques, indicating that AI is becoming an important twist in the health care and telemedicine field. The most used technique is CNNs followed by SVM and the combination of CNN and LSTM. The use of SVM, with the combination of DL techniques in the feature extraction and the pre-processing phases, recorded very important results.

Table 7. Advantages and limitations of arrhythmia classification methods.

STUDY	CLASSIFICATION METHOD	ADVANTAGES	DRAWBACKS/ LIMITATIONS	PERSPECTIVES
Chen et al[19]	Cascaded classifier composed of random forest (RF) and multilayer perceptron (MLP)	• Time-saving method • No need for signal pre-processing	The absence of the dynamic property extraction from 1-D signal	LSTM network will be applied to extract more features from ECG dynamic
Acharya et al[20]	CNN	• Self-removal of the unwanted noise • Fully automatic algorithm: No need for additional feature extraction or selection • Insensitive to the ECG signal quality	• Training is computationally expensive and time-consuming (≈hours) • Require specialized hardware to efficiently train (GPU) Require very large number of images	Training the CNN to recognize temporal sequences: normal, abnormal, and potentially life-threatening conditions of heart electrical activity
Yildirim et al[21]	CAE and LSTM	• Reduce training time cost • Secure transmission of patient data due to coded structure Low feature loss during compression	• Complex DL model for compression	The use of coded features with traditional classifiers
Yang et al[22]	CCANet and SVM	• High overall accuracy Classification of detailed classes (15 classes on MIT-BIH database and 7 classes on INCART database) • Use of the correlation of multi-lead ECG signals A small number of parameters to be adjusted	• Lower recognition performance for classes with minimal heartbeats The size of ECG matrix needs to be adjusted for different databases due to the different sample rates	Using more ECG leads and adopting a resampling algorithm
Yildirim[23]	Unidirectional and bidirectional LSTM with wavelet-based layer	• Providing more distinguishing and size-reduced features The input length does not affect the network storage requirements since LSTMs are located in space and time	Extra time cost due to the wavelet layer	Combining LSTMs with other DL methods
Sumathi et al[9]	ANFIS: combination of NN and fuzzy logic	• The advantage of a fuzzy set is the depiction of prior knowledge into a set of constraints to reduce the optimization research space Easy to implement	NR	NR
Gao et al[24]	LSTM with focal loss	• Addressing the issue of imbalanced dataset classification • Extracting the timing characteristics of ECG Robust model with noisy data	• Time cost of the training phase	Incorporating more beat types and adding different types of noise
Martis et al[25]	NN and SVM	Obtaining high accuracies due to the use of nonlinear features	NR	Evaluating a combination of several nonlinear features
Li et al[13]	SVM with genetic algorithm	Improving the classifier performance by the GA algorithm	NR	NR

(Continued)

Table 7. (Continued)

STUDY	CLASSIFICATION METHOD	ADVANTAGES	DRAWBACKS/ LIMITATIONS	PERSPECTIVES
Anwar et al[26]	Feedforward NN	Computationally efficient for arrhythmia classification: 18 types	NR	The use of automatic patient customization scheme allowing the heartbeat classification method to be able to adjust to individual physiological features using wearable sensors
Liu[27]	SoNFIN: NN with fuzzy logic	• Obtaining high performance with the fuzzy logic Suitable model for a portable system	Recognition and classification are more difficult for 12-lead signal	NR
Elhaj et al[28]	Feedforward NN and SVM	High capability for generalization	NR	NR
Kim et al[29]	GoogleNet Deep NN	Accuracy enhanced by inception structure	The computational complexity increases exponentially as the layer becomes deeper	NR
Yıldırım[30]	1-D CNN	• Efficient and non-complex • Highly accurate and fast (real-time classification) • End-to-end structure • Reducing computational complexity	• Small number of ECG signal fragments are analyzed. No possibility for classifying fragments of ECG signal containing more than 1 class	Testing the efficiency of developed 1D-CNN using other physiological signals and classifying fragments of the ECG signal that containing more than 1 class
Oh et al[31]	Modified U-net architecture: CAE with skipped connections	• End-to-end solution, requires minimal processing • High capacity of handling the heterogeneity of beats (ECG segments with mixed arrhythmia types) • Producing localized outputs of higher resolution Good generalization ability without overfitting	• The training phase is computationally intensive and slow • The model is trained and tested using an imbalanced dataset • Predictions for R peak are not precise	The model can be tested on variable length signals for analysis
Oh et al[32]	Combination of CNN and LSTM	• CNN is good at picking up spatial features while the LSTM is better at learning temporal features • Predictions made by the system are reproducible with no inter- and intra-observer biases • Noise filtering, feature extraction and selection techniques are not required	• Model is computationally intensive and learning is slow • Algorithm is not robust in detecting the atrial premature beat from normal ECG segment • The system is developed using imbalanced dataset	The use of auto-encoder network on the ECG data for element-wise analysis, by associating each pixel with a class label High-end graphic cards to accelerate the training process of the model Data augmentation to balance data
Raj and Ray[33]	Least-square twin SVM	• The method is fast since the classification time is very low • The implementation of cross-validation scheme makes the method highly robust and reliable The proposed method is highly efficient and completely automatic	• More memory is required for implementation • More optimization time is taken by the optimization technique to tune the classifier The use of fixed window for the heartbeats	More classes of cardiac arrhythmias can be detected using the proposed method

(Continued)

Table 7. (Continued)

STUDY	CLASSIFICATION METHOD	ADVANTAGES	DRAWBACKS/ LIMITATIONS	PERSPECTIVES
Ribeiro et al[15]	Unidimensional residual NN	• Model more robust to noise Recognize accurately ECG rhythm and morphological abnormalities in clinical examinations	The presence of relatively infrequent classes leading to few erroneous classifications	Diagnosis of less common forms of arrhythmia Testing the algorithm in a controlled real-life situation
Qin et al[34]	One-vs-one SVM	Beats accurately recognized in beat-based scheme	Ineffective recognition of record-based scheme due to the lack of comprehensive knowledge over the training beats	Using the state-of-the-art DL method for the classification of more types of heartbeats
Rajagopal and Ranganathan[35]	Combination of KNN and SVM	• Successfully trained to classify overlapped classes Complex tasks can be learned using simple procedures by local approximation for the KNN	Limitation in speed and size during both training and testing for SVM	NR
Benali et al[36]	WNN	• Provide faster training times and multi-resolution analysis capabilities High discrimination between cardiac rhythms	NR	NR
Rajesh and Dhuli[37]	SMO–SVM	High ability to discriminate different types of heartbeats even under noisy conditions	The probability to obtain biased results	Feature-based disease identification
Yang et al[38]	Linear SVM	Robust to noise and skewed data	The imbalance of data	The use of dimension reduction and data augmentation techniques
Hannun et al[16]	CNN	• High diagnostic performance similar to that of cardiologists High ability of recognizing patterns from raw ECG signal	Some prevalence-dependent metrics, such as the F1 score would not be expected to generalize to the broader population since the extraction of rhythms was enrolled in targeted patients	NR
Park and Kang[18]	Decision tree	• The model can accommodate complicated patterns, and considers many more types of beats. Accurate classification method that reduces the number of false alarms and missing events by considering more types of heartbeats	Low specificity due to the overfitting problem associated with decision tree	• The use of harmonic classification forest Adding the "live mode" to the system for the use in mobile Holter monitoring
Ullah et al[39]	Combined CNN and LSTM and Attention method	High accuracy due to combination of CNN and LSTM	Possibility of overfitting the data because the design of the proposed model includes 10 residual blocks	• Application of the model in binding domains, such as cloud and mobile systems Develop wearable technologies with integrated low-power consumption model
Irfan et al[10]	Combined CNN and LSTM	• Impervious to the time-step order • Cost-effective time	The addition of more DL networks will increase computational cost	• Deployment on embedded systems Deal with real-time data

(Continued)

Table 7. (Continued)

STUDY	CLASSIFICATION METHOD	ADVANTAGES	DRAWBACKS/ LIMITATIONS	PERSPECTIVES
Shahin et al[40]	Adversarial multi-task model	Improve system generalization by forcing the model to discriminate based only on variations between heartbeat types and not on variations between subjects	High rate of misclassification in test set	• Extend the experiments to other types of heartbeats Use CNN and LSTM techniques
Ma et al[41]	Fusion of ResNet and Bi-LSTM with Attention method	• Approach the periodicity of ECG by the fusion of ResNet spatial information and Bi-LSTM temporal information High recognition performance	NR	Extend the model to other arrhythmia databases
Sabut et al[42]	Deep NN	Detect shockable ventricular tachycardia	NR	• Use CNN to improve the computational complexity Use the model in AED devices for automatic delivery of shocks
Wang et al[43]	Multi-scale fusion CNN	Employ correlations among features of different scales	NR	Apply the model to other physiological signal
Anbarasi et al[11]	Combined CNN and LSTM	Effective model with high extraction techniques	Computationally very expensive	Apply the model to disorders, such as gastrointestinal ailments, the differentiation of neoplastic and non-neoplastic tissues
Zubair and Yoon[44]	CNN	• Incorporate both the short-term and long-term morphological characteristics of ECG Increased classification rate of minority classes due to cost-sensitive function	NR	Use the cost-sensitive learning in “Point of Sale” systems
Hu et al[45]	Transformer-based CNN	• Competitive heartbeat positioning and classification due to inter-beat dependencies High ability of generalization	NR	• Detection of special signals Deploy the algorithm on wearable devices
Feyisa et al[46]	Multi-receptive field CNN	Facilitate the ability to look into multiple fields simultaneously and capture various features to discriminate the ECG classes	NR	• Tackle the data imbalance issue with GAN network Use WT for feature representation
Ju et al[47]	Deep bidirectional GRU Network	Provide more powerful expression and learning ability	NR	NR
Wang et al[48]	Dual-path RNN	Realize both intra-segmental and inter-segmental modeling	Perform poorly for longer sequences	Develop multi-channel dual-path RNN
Wang[49]	Improved bidirectional GRU	Alleviate the problem of information redundancy	Only devoted to PVC detection	Detect other signal types using more diverse data
Luo et al[50]	Hybrid convolutional RNN	Combine the strengths of 3 DL methods Strong ability to extract deep features	Time-consuming	Extend the model to other disease detection
Iftene et al[51]	CNN, BNN, RNN	High accuracy when using pre-processing	Lack of balanced databases	Develop an automated solution for AF detection on mobile devices

Abbreviations: AED, Automatic External Defibrillator; BNN, Bayesian Neural Network; GA, Genetic Algorithm; GPU, Graphics Processing Unit; INCART, St. Petersburg Institute of Cardiological Techniques; KNN, k-nearest neighbors; LSTM, Long Short-Term Memory; MIT-BIH, Massachusetts Institute of Technology-Boston's Beth Israel Hospital; NR, Not Reported; SVM, Support Vector Machine.

In Table 7, shown all techniques that are used in arrhythmia classification in this review. We present the advantages and limitations of each classification method as they are identified by the authors of each study.

We notice that the most common limitations for the use of DL methods are that they are time-consuming and computationally expensive and require very efficient hardware resources. Otherwise, they can perform accurately the classification of heartbeats with the end-to-end learning besides they can be robust to noise. For the traditional ML methods, they can be simply implemented and are computationally efficient and provide faster training time.

To sum up, we cannot give a decisive recommendation of the best model, based on the analysis made in the "Discussion" section because none of the 40 selected studies applied the exact same feature extraction, pre-processing, or prediction techniques in all the stages. Besides, the input information, the ECG signal information, the development tools, and the computing capacities vary from one study to another. However, when taking into consideration all these variants and the results of the studies' analysis, we can presume that the usage of DL solely or the usage of ML combined with DL techniques can achieve very promising results.

Author Contributions

BA, MO, and SD contributed to conceptualization, resources, and supervision.

REFERENCES

1. Arrhythmia—NHS. Accessed January 13, 2021. https://www.nhs.uk/conditions/arrhythmia/
2. Mubarik A, Iqbal AM. Holter Monitor. In: *StatPearls* [Internet]. Treasure Island, FL: StatPearls Publishing; Updated July 25, 2022. https://www.ncbi.nlm.nih.gov/books/NBK538203/
3. Trardi Y, Ananou B, Haddi Z, Ouladsine M. Multi-dynamics analysis of QRS complex for atrial fibrillation diagnosis. Paper presented at: 2018 5th International Conference on Control, Decision and Information Technologies (CoDIT); April 10-13, 2018; Thessaloniki, Greece. doi:10.1109/CoDIT.2018.8394935
4. Trardi Y, Ananou B, Haddi Z, Ouladsine M. A novel method to identify relevant features for automatic detection of atrial fibrillation. Paper presented at: 2018 26th Mediterranean Conference on Control and Automation (MED); June 19-22, 2018; Zadar, Croatia. doi:10.1109/MED.2018.8442479
5. Trardi Y, Ananou B, Ouladsine M. An advanced arrhythmia recognition methodology based on R-waves time-series derivatives and benchmarking machine-learning algorithms. Paper presented at: 2020 European Control Conference (ECC); May 12-15, 2020, St. Petersburg, Russia. doi:10.23919/ECC51009.2020.9143678
6. Trardi Y, Ananou B, Ouladsine M. Computationally efficient algorithm for atrial fibrillation detection using linear and geometric features of RR time-series derivatives. Paper presented at: 2022 International Conference on Control, Automation and Diagnosis (ICCAD); July 13-15, 2022; Lisbon, Portugal. doi:10.1109/ICCAD55197.2022.9853910
7. The best academic research databases [2019 update]—Paperpile. Accessed January 13, 2021. https://paperpile.com/g/academic-research-databases/
8. ResearchGate. In: Wikipedia. 2020. https://en.wikipedia.org/wiki/ResearchGate
9. Sumathi S, Beaulah HL, Vanithamani R. A wavelet transform based feature extraction and classification of cardiac disorder. *J Med Syst*. 2014;38:98. doi:10.1007/s10916-014-0098-x
10. Irfan S, Anjum N, Althobaiti T, Alotaibi AA, Siddiqui AB, Ramzan N. Heartbeat classification and arrhythmia detection using a multi-model deep-learning technique. *Sensors*. 2022;22:5606. doi:10.3390/s22155606
11. Anbarasi A, Ravi T, Manjula VS, et al. A modified deep learning framework for arrhythmia disease analysis in medical imaging using electrocardiogram signal. *Biomed Res Int*. 2022;2022:5203401. doi:10.1155/2022/5203401
12. PhysioNet Databases. Accessed January 13, 2021. https://physionet.org/about/database/#open
13. Li H, Yuan D, Wang Y, Cui D, Cao L. Arrhythmia classification based on multi-domain feature extraction for an ECG recognition system. *Sensors*. 2016;16:1744. doi:10.3390/s16101744
14. Patient Monitor Simulators Fluke Biomedical. Accessed January 13, 2021. https://www.flukebiomedical.com/products/biomedical-test-equipment/patient-monitor-simulators
15. Ribeiro AH, Ribeiro MH, Paixão GMM, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun*. 2020;11:1760. doi:10.1038/s41467-020-15432-4
16. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med*. 2019;25:65-69. doi:10.1038/s41591-018-0268-3
17. Why Zio iRhythm. Accessed January 13, 2021. https://www.irhythmtech.com/patients/why-zio
18. Park J, Kang K. PcHD: personalized classification of heartbeat types using a decision tree. *Comput Biol Med*. 2014;54:79-88. doi:10.1016/j.compbiomed.2014.08.013
19. Chen G, Hong Z, Guo Y, Pang C. A cascaded classifier for multi-lead ECG based on feature fusion. *Comput Methods Programs Biomed*. 2019;178:135-143. doi:10.1016/j.cmpb.2019.06.021
20. Acharya UR, Oh SL, Hagiwara Y, et al. A deep convolutional neural network model to classify heartbeats. *Comput Biol Med*. 2017;89:389-396. doi:10.1016/j.compbiomed.2017.08.022
21. Yildirim O, Baloglu UB, Tan RS, Ciaccio EJ, Acharya UR. A new approach for arrhythmia classification using deep coded features and LSTM networks. *Comput Methods Programs Biomed*. 2019;176:121-133. doi:10.1016/j.cmpb.2019.05.004
22. Yang W, Si Y, Wang D, Zhang G. A novel approach for multi-lead ECG classification using DL-CCANet and TL-CCANet. *Sensors*. 2019;19:3214. doi:10.3390/s19143214
23. Yildirim Ö. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Comput Biol Med*. 2018;96:189-202. doi:10.1016/j.compbiomed.2018.03.016
24. Gao J, Zhang H, Lu P, Wang Z. An effective LSTM recurrent network to detect arrhythmia on imbalanced ECG dataset. *J Healthc Eng*. 2019;2019:6320651. doi:10.1155/2019/6320651
25. Martis RJ, Acharya UR, Lim CM, Mandana KM, Ray AK, Chakraborty C. Application of higher order cumulant features for cardiac health diagnosis using ECG signals. *Int J Neural Syst*. 2013;23:1350014. doi:10.1142/S0129065713500147
26. Anwar SM, Gul M, Majid M, Alnowami M. Arrhythmia classification of ECG signals using hybrid features. *Comput Math Methods Med*. 2018;2018:1380348. doi:10.1155/2018/1380348
27. Liu S-H. Arrhythmia identification with two-lead electrocardiograms using artificial neural networks and support vector machines for a portable ECG monitor system. *Sensors*. 2013;13:813-828.
28. Elhaj FA, Salim N, Harris AR, Swee TT, Ahmed T. Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals. *Comput Methods Programs Biomed*. 2016;127:52-63. doi:10.1016/j.cmpb.2015.12.024
29. Kim JH, Seo SY, Song CG, Kim KS. Assessment of electrocardiogram rhythms by GoogLeNet deep neural network architecture. *J Healthc Eng*. 2019;2019:2826901. doi:10.1155/2019/2826901
30. Yıldırım Ö. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Comput Biol Med*. 2018;10:411-420.
31. Oh SL, Ng EYK, Tan RS, Acharya UR. Automated beat-wise arrhythmia diagnosis using modified U-net on extended electrocardiographic recordings with heterogeneous arrhythmia types. *Comput Biol Med*. 2019;105:92-101. doi:10.1016/j.compbiomed.2018.12.012
32. Oh SL, Ng EYK, Tan RS, Acharya UR. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Comput Biol Med*. 2018;102:278-287. doi:10.1016/j.compbiomed.2018.06.002
33. Raj S, Ray KC. Automated recognition of cardiac arrhythmias using sparse decomposition over composite dictionary. *Comput Methods Programs Biomed*. 2018;165:175-186. doi:10.1016/j.cmpb.2018.08.008
34. Qin Q, Li J, Zhang L, Yue Y, Liu C. Combining low-dimensional wavelet features and support vector machine for arrhythmia beat classification. *Sci Rep*. 2017;7:6067. doi:10.1038/s41598-017-06596-z
35. Rajagopal R, Ranganathan V. Design of a hybrid model for cardiac arrhythmia classification based on Daubechies wavelet transform. *Adv Clin Exp Med*. 2018;27:727-734. doi:10.17219/acem/68982

36. Benali R, Bereksi Reguig F, Hadj Slimane Z. Automatic classification of heartbeats using wavelet neural network. *J Med Syst*. 2012;36:883-892. doi:10.1007/s10916-010-9551-7
37. Rajesh KNVPS, Dhuli R. Classification of ECG heartbeats using nonlinear decomposition methods and support vector machine. *Comput Biol Med*. 2017;87:271-284. doi:10.1016/j.compbiomed.2017.06.006
38. Yang W, Si Y, Wang D, Guo B. Automatic recognition of arrhythmia based on principal component analysis network and linear support vector machine. *Comput Biol Med*. 2018;101:22-32. doi:10.1016/j.compbiomed.2018.08.003
39. Ullah W, Siddique I, Zulqarnain RM, Alam MM, Ahmad I, Raza UA. Classification of arrhythmia in heartbeat detection using deep learning. *Comput Intell Neurosci*. 2021;2021:2195922. doi:10.1155/2021/2195922
40. Shahin M, Oo E, Ahmed B. Adversarial multi-task learning for robust end-to-end ECG-based heartbeat classification. Paper presented at: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); July 20-24, 2020; Montreal, QC, Canada. doi:10.1109/EMBC44109.2020.9175640
41. Ma S, Cui J, Xiao W, Liu L. Deep learning-based data augmentation and model fusion for automatic arrhythmia identification and classification algorithms. *Comput Intell Neurosci*. 2022;2022:1577778. doi:10.1155/2022/1577778
42. Sabut S, Pandey O, Mishra BSP, Mohanty M. Detection of ventricular arrhythmia using hybrid time–frequency-based features and deep neural network. *Phys Eng Sci Med*. 2021;44:135-145. doi:10.1007/s13246-020-00964-2
43. Wang R, Fan J, Li Y. Deep multi-scale fusion neural network for multi-class arrhythmia detection. *IEEE J Biomed Health Inform*. 2020;24:2461-2472. doi:10.1109/JBHI.2020.2981526
44. Zubair M, Yoon C. Cost-sensitive learning for anomaly detection in imbalanced ECG data using convolutional neural networks. *Sensors*. 2022;22:4075. doi:10.3390/s22114075
45. Hu R, Chen J, Zhou L. A transformer-based deep neural network for arrhythmia detection using continuous ECG signals. *Comput Biol Med*. 2022;144:105325. doi:10.1016/j.compbiomed.2022.105325
46. Feyisa DW, Debelee TG, Ayano YM, Kebede SR, Assore TF. Lightweight multireceptive field CNN for 12-lead ECG signal classification. *Comput Intell Neurosci*. 2022;2022:8413294. doi:10.1155/2022/8413294
47. Ju Y, Zhang M, Zhu H. Study on a new deep bidirectional GRU network for electrocardiogram signals classification. Paper presented at: Proceedings of the 3rd International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2019); May 30-31, 2019; Chongqing, China. doi:10.2991/iccia-19.2019.54
48. Wang M, Rahardja S, Fränti P, Rahardja S. Single-lead ECG recordings modeling for end-to-end recognition of atrial fibrillation with dual-path RNN. *Biomed Signal Process Control*. 2022;79:104067. doi:10.1016/j.bspc.2022.104067
49. Wang J. Automated detection of premature ventricular contraction based on the improved gated recurrent unit network. *Comput Methods Programs Biomed*. 2021;208:106284. doi:10.1016/j.cmpb.2021.106284
50. Luo X, Yang L, Cai H, Tang R, Chen Y, Li W. Multi-classification of arrhythmias using a HCRNet on imbalanced ECG datasets. *Comput Methods Programs Biomed*. 2021;208:106258. doi:10.1016/j.cmpb.2021.106258
51. Iftene A, Burlacu A, Gifu D. Atrial fibrillation detection based on deep learning models. *Procedia Comput Sci*. 2022;207:3752-3760. doi:10.1016/j.procs.2022.09.436
52. Pan J, Tompkins WJ. A real-time QRS detection algorithm. *IEEE T Biomed Eng*. 1985; BME-32:230-236. doi:10.1109/TBME.1985.325532
53. Guvenir H, Acar B, Muderrisoglu H. *Arrhythmia*. UCI Machine Learning Repository; 1998. https://archive.ics.uci.edu/ml/datasets/arrhythmia
54. TNMG dataset. Accessed October 20, 2022. https://zenodo.org/record/3765780
55. Gupta D, Bajpai B, Dhiman G, Soni M, Gomathi S, Mane D. Review of ECG arrhythmia classification using deep neural network [published online ahead of print May 22, 2021]. *Mater Today Proc*. doi:10.1016/j.matpr.2021.05.249
56. Ebrahimi Z, Loni M, Daneshtalab M, Gharehbaghi A. A review on deep learning methods for ECG arrhythmia classification. *Expert Syst Appl*. 2020;7:100033. doi:10.1016/j.eswax.2020.100033
57. Annam JR, Kalyanapu S, Ch S, Somala J, Raju SB. Classification of ECG heartbeat arrhythmia: a review. *Procedia Comput Sci*. 2020;171:679-688. doi:10.1016/j.procs.2020.04.074
58. Jensen PN, Johnson K, Floyd J, Heckbert SR, Carnahan R, Dublin S. A systematic review of validated methods for identifying atrial fibrillation using administrative data: detection of atrial fibrillation in claims. *Pharmacoepidemiol Drug Saf*. 2012;21:141-147. doi:10.1002/pds.2317
59. Tamariz L, Harkins T, Nair V. A systematic review of validated methods for identifying ventricular arrhythmias using administrative and claims data: detection of ventricular arrhythmia in claims. *Pharmacoepidemiol Drug Saf*. 2012;21:148-153. doi:10.1002/pds.2340
60. Sanamdikar ST. A literature review on arrhythmia analysis of ECG signal. *IRJET*. 2015;2(3):1-6.
61. Luz EJ, Schwartz WR, Cámara-Chávez G, Menotti D. ECG-based heartbeat classification for arrhythmia detection: a survey. *Comput Methods Programs Biomed*. 2016;127:144-164. doi:10.1016/j.cmpb.2015.12.008
62. Parvaneh S, Rubin J, Babaeizadeh S, Xu-Wilson M. Cardiac arrhythmia detection using deep learning: a review. *J Electrocardiol*. 2019;57S:S70-S74. doi:10.1016/j.jelectrocard.2019.08.004
63. Houssein EH, Kilany M, Hassanien AE. ECG signals classification: a review. *Int J Intell Eng Inform*. 2017;5:376. doi:10.1504/IJIEI.2017.087944

Appendix 1

Table A1. Evaluation metrics.

STUDY	DATABASE/ DATA TYPE/ METHOD	ACCURACY (%)	SENSITIVITY (RECALL) (%)	SPECIFICITY (%)	PRECISION (%)	F1-SCORE (%)	AUC (%)
Chen et al[19]	SV, MIT-BIH, QT, INCART	97.6, 99.3, 99.6, 99.8	97.6, 99.3, 99.6, 99.8	NR	97.6, 99.3, 99.6, 99.8	97.6, 99.3, 99.6, 99.8	NR
Acharya et al[20]	Augmented noisy data, Augmented free-noise data, Original noisy data, Original free-noise data	93.47 94.03 89.07 89.03	96.01 96.71 95.90 95.90	91.64 91.54 88.35 88.39	NR	NR	NR
Yildirim et al[21]	LSTM, CAE-LSTM	99.23 99.11	99.0 99.0	NR NR	99.0 99.0	99.0 99.0	NR
Yang et al[22]	MIT-BIH, INCART (II and V1), INCART (V1 and V5), INCART (II and V5), INCART (II, V1, and V5)	99.4, 98.31, 98.26, 98.31, 98.76	94.6, 90.89, 90.74, 90.38, 92.71	99.6, 98.85, 98.84, 98.87, 99.16	NR	NR	NR
Yildirim[23]	DULSTM-WS, DBLSTM-WS	99.25, 99.39	99.0, 100	NR	100, 100	100, 100	NR
Sumathi et al[9]	(ANFIS) model	98.24	NR	NR	NR	NR	NR
Gao et al[24]	LSTM with focal loss, LSTM with cross-entropy	99.26, 98.70	99.26, 98.70	99.14, 98.05	99.30, 98.75	99.27, 98.36	NR
	LSTM-FL with noise-free data, LSTM-FL with noisy data	99.26, 99.07	99.26, 99.07	99.14, 98.99	99.30, 99.13	99.27, 99.09	NR
Martis et al[25]	Feedforward NN (with HOS + PCA), Least-square SVM (with HOS + PCA), Feedforward NN (DWT + HOS + PCA), Least-square SVM (DWT + HOS + PCA)	94.52, 94.30, 93.61, 93.76	98.61, 99.72, 98.51, 99.46	98.41, 96.69, 97.80, 97.36	NR	NR	NR
Li et al[13]	MIT-BIH arrhythmia database, ECG acquisition experimental platform	98.80, 97.30	98.50, 97.50	99.69, 99.32	NR	NR	NR
Anwar et al[26]	Class-oriented scheme with 18 classes, Subject-oriented scheme with 5 classes	99.75, 99.8	98.7, 99.7	99.9, 99.9	NR	NR	NR
Liu[27]	MIT-BIH arrhythmia database	96.4	NR	NR	NR	NR	NR
Elhaj et al[28]	SVM-RBF with PCA-DWT + ICA-HOS, NN with PCA-DWT + ICA-HOS, SVM-RBF with PCA-DWT, NN with PCA-DWT, SVM-RBF with ICA-HOS, NN with ICA-HOS	98.91, 98.90, 88.04, 93.48, 97.83, 94.57	98.91, 98.90, NR NR NR NR	97.85, 98.90, NR NR NR NR	NR	NR	NR
Kim et al[29]	GoogleNet deep NN with 1-inception, GoogleNet deep NN with 2-inception, GoogleNet deep NN with CNN + 1-inception	95.3, 96.3, 95.9	NR	NR	NR	NR	NR

(Continued)

Table A1. (Continued)

STUDY	DATABASE/ DATA TYPE/ METHOD	ACCURACY (%)	SENSITIVITY (RECALL) (%)	SPECIFICITY (%)	PRECISION (%)	F1-SCORE (%)	AUC (%)
Yıldırım[30]	CNN (13-classes), CNN (15-classes), CNN (17-classes)	95.2, 92.51, 91.33	93.52, 88.57, 83.91	99.61, 99.39, 99.41	92.52, 90.48, 89.52	92.45, 89.28, 85.38	NR
Oh et al[31]	Modified U-net architecture	97.32	94.44	98.26	NR	NR	NR
Oh et al[32]	CNN + LSTM without dropout, CNN + LSTM with 2-dropout, CNN + LSTM with 3-dropout	98.42, 97.88, 98.10	98.07, 97.26, 97.50	98.76, 98.50, 98.70	NR	NR	NR
Raj and Ray[33]	Category-based scheme, Personalized scheme	99.21, 90.08	99.21, NR	NR	NR	99.21, NR	NR
Ribeiro et al[15]	Unidimensional residual NN	NR	93.48	98.4	92.36	92.55	NR
Qin et al[34]	Beat-based training scheme, Record-based training scheme	99.70, 81.47	99.82, 44.40	99.82, 88.88	NR	NR	NR
Rajagopal and Ranganathan[35]	Combined KNN and SVM	99.78	92.56	99.53	NR	94.5	NR
Benali et al[36]	WNN	98.78	NR	NR	NR	NR	NR
Rajeshand Dhuli[37]	SMO–SVM with EMD (linear, RBF, cubic kernels) for MIT-BIH data	97.44, 98.58, 99.2	93.06, 96.48, 98.01	98.66, 99.00, 99.49	NR	NR	NR
	SMO–SVM with EEMD (linear, RBF, cubic kernels) for MIT-BIH data	89.86, 96.45, 94.68	72.18, 85.80, 86.71	93.10, 98.81, 96.67	NR	NR	NR
	SMO–SVM with EEMD (linear, RBF, cubic kernels) for INCART	96.59, 97.46, 97.57	93.2, 94.92, 95.15	97.73, 98.30, 98.37	NR	NR	NR
Yang et al[38]	Linear SVM, KNN, BP-NN, RF (noisy data)	97.77, 97.10, 96.95, 96.01	NR	NR	NR	NR	NR
	Linear SVM, KNN, BP-NN, RF (noise-free data)	97.08, 96.27, 95.89, 95.22	NR	NR	NR	NR	NR
Hannun et al[16]	CNN with sequence-level, CNN with set level	NR	NR	NR	NR	80.7, 83.7	97.8, 97.7
Park and Kang[18]	Decision tree for non-personalized scheme, Decision tree for personalized scheme	89.95, 85.26	94.61, 97.99	85.28, 72.52	NR	NR	NR
Ullah et al[39]	CNN, CNN + LSTM, CNN + LSTM + Attention method	NR	99.12, 99.3, 99.29	NR	99.12, 99.3, 99.29	99.12, 99.3, 99.29	NR
Irfan et al[10]	CNN + LSTM for MIT-BIH database, CNN + LSTM for UCI arrhythmia dataset	99.35, 99.05	98.37, 89.11	99.59, 99.40	NR	NR	NR
Shahin et al[40]	Multi-task adversarial network	86	NR	NR	NR	NR	NR

(Continued)

Table A1. (Continued)

STUDY	DATABASE/ DATA TYPE/ METHOD	ACCURACY (%)	SENSITIVITY (RECALL) (%)	SPECIFICITY (%)	PRECISION (%)	F1-SCORE (%)	AUC (%)
Ma et al[41]	ResNet + Bi-LSTM + Attention method	99.4	98.4	99.3	NR	NR	NR
Sabut et al[42]	DNN	99.2	98.8	99.3	NR	NR	NR
Wang et al[43]	Deep multi-scale fusion NN (CPSC dataset, CINC dataset)	NR	82.2, 82.9	NR	83.8, 85.6	82.8, 84.1	NR
Anbarasi et al[11]	Combined CNN and LSTM	98.7	98	98	NR	NR	NR
Zubair and Yoon[44]	CNN (intra-patient, inter-patient)	99.81, 96.36	88.82, 70.60	99.54, 96.16	NR	NR	NR
Hu et al[45]	Transformer-based CNN (8, 4, 2 classes)	99.12, 99.49, 99.23	97.53, 92.51, 99.23	99.83, 99.84, 99.23	98.54, 95.38, 99.23	98.03, 93.88, 99.23	NR
Feyisa et al[46]	Multi-receptive CNN (41, 20, 5 classes)	98, 96.2 89.7	31, 56, 76	NR	28, 42, 73	29, 46, 72	92, 92, 93
Ju et al[47]	Deep bidirectional GRU	99.51	99	NR	98	98	NR
Wang et al[48]	Dual-path RNN	84.5	NR	NR	NR	82.91	NR
Wang[49]	CNN + improved BGRU (MIT-BIH, CPDB)	97.9, 98.3	98, 98.4	97.8, 98.2	NR	NR	NR
Luo et al[50]	Hybrid convolutional RNN	99.01	99.58	NR	NR	99.51	NR
Iftene et al[51]	1-D CNN with pre-processing, 1-D CNN, Bayesian NN, GRU network	98, 95, 90, 94	NR	NR	NR	NR	NR

Abbreviations: ANFIS, Adaptive Neuro-Fuzzy Inference System; AUC, Area Under Curve; BP-NN, Back-Propagation Neural Network; CAE, Convolutional Auto-encoder; CINC, Computing in Cardiology Challenge; CPDB, China Physiological Signal Challenge database; CPSC, China Physiological Signal Challenge; CNN, Convolutional Neural Network; DBLSTM-WS, Deep Bidirectional Long Short Term Memory network based Wavelet-Sequences; DNN, Deep Neural Network; DULSTM-WS, Deep Unidirectional Long Short Term Memory network based Wavelet-Sequences; DWT, Discrete Wavelet Transform; EEMD, Ensemble Empirical Mode Decomposition; EMD, Empirical Mode Decomposition; GRU, Gated Recurrent Unit; HOS, Higher-Order Spectra; ICA, Independent Component Analysis; INCART, St. Petersburg Institute of Cardiological Techniques; KNN, k-nearest neighbors; LSTM, Long Short-Term Memory; LSTM-FL, Long Short-Term Memory with Focal Loss; MIT-BIH, Massachusetts Institute of Technology-Boston's Beth Israel Hospital; NR, Not Reported; PCA, Principal Component Analysis; RBF, Radial Basis Function; SV, supraventricular arrhythmia database; SVM, Support Vector Machine; UCI, University of California Irvine.

Table A2. ECG beats categorized as per ANSI/AAMI EC57; 2012 standard.

N	S	V	F	Q
• Normal • Left bundle branch block • Right bundle branch block • Atrial escape • Nodal (junctional) escape	• Atrial premature • Aberrant atrial premature • Nodal (junctional) premature • Supraventricular premature	• PVC • Ventricular escape	• Fusion of ventricular and normal	• Paced • Fusion of paced and normal • Unclassifiable

Table A3. Beat annotations by PhysioBank.

CODE	DESCRIPTION
N	Normal beat (displayed as "·" by the PhysioBank ATM, LightWAVE, pschart, and psfd)
L	Left bundle branch block beat
R	Right bundle branch block beat
B	Bundle branch block beat (unspecified)
A	Atrial premature beat
a	Aberrated atrial premature beat
J	Nodal (junctional) premature beat
S	Supraventricular premature or ectopic beat (atrial or nodal)
V	Premature ventricular contraction
r	R-on-T premature ventricular contraction
F	Fusion of ventricular and normal beat
e	Atrial escape beat
j	Nodal (junctional) escape beat
n	Supraventricular escape beat (atrial or nodal)
E	Ventricular escape beat
/	Paced beat
f	Fusion of paced and normal beat
Q	Unclassifiable beat
?	Beat not classified during learning

Appendix 2

Table of acronyms.

ACRONYM	SIGNIFICATION
ABC	Artificial bee colony
AI	Artificial intelligence
AUC	Area under curve
CAE	Convolutional auto-encoder
CNN	Convolutional neural network
DAG	Directed acyclic graph
Db	Daubechies
DL-CCANet	Dual-lead canonical correlation analysis network
DCT	Discrete cosine transform
DOST	Discrete orthogonal stockwell transform
DST	Discrete sine transform
DWT	Discrete wavelet transform
FFT	Fast Fourier transform
GBM	Génie Bio-médical
HOS	Higher-order spectrum
ICA	Independent component analysis
INCART	St. Petersburg Institute of Cardiological Techniques
KICA	Kernel-independent component analysis
LDA	Linear discriminant analysis
LSTM	Long short-term memory
LSTM-WS	Long short-term memory wavelet sequence
LS-TSVM	Least-square twin support vector machine
MIT-BIH	Massachusetts Institute of Technology-Boston's Beth Israel Hospital
MLP	Multilayer perceptron
NN	Neural network
PCA	Principal component analysis
PCANet	Principal component analysis network
RBF	Radial basis function
RF	Random forest
SD	Standard deviation
SV	Supraventricular
SVM	Support vector machine
TL-CCANet	Triple-lead canonical correlation analysis network