applied sciences

MDPI

Review

Deep Learning-Based ECG Arrhythmia Classification: A Systematic Review

Qiao Xiao [1], Khuan Lee [2], Siti Aisah Mokhtar [1], Iskasymar Ismail [3,4], Ahmad Luqman bin Md Pauzi [3,4], Qiuxia Zhang [2] and Poh Ying Lim [1,*]

[1] Department of Community Health, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang 43400, Malaysia
[2] Department of Nursing, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang 43400, Malaysia
[3] Department of Medicine, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang 43400, Malaysia
[4] RESQ Stroke Emergency Unit, Hospital Sultan Abdul Aziz Shah, Universiti Putra Malaysia, Serdang 43400, Malaysia
* Correspondence: pohying_my@upm.edu.my

Abstract: Deep learning (DL) has been introduced in automatic heart-abnormality classification using ECG signals, while its application in practical medical procedures is limited. A systematic review is performed from perspectives of the ECG database, preprocessing, DL methodology, evaluation paradigm, performance metric, and code availability to identify research trends, challenges, and opportunities for DL-based ECG arrhythmia classification. Specifically, 368 studies meeting the eligibility criteria are included. A total of 223 (61%) studies use MIT-BIH Arrhythmia Database to design DL models. A total of 138 (38%) studies considered removing noise or artifacts in ECG signals, and 102 (28%) studies performed data augmentation to extend the minority arrhythmia categories. Convolutional neural networks are the dominant models (58.7%, 216) used in the reviewed studies while growing studies have integrated multiple DL structures in recent years. A total of 319 (86.7%) and 38 (10.3%) studies explicitly mention their evaluation paradigms, i.e., intra- and inter-patient paradigms, respectively, where notable performance degradation is observed in the inter-patient paradigm. Compared to the overall accuracy, the average F_1 score, sensitivity, and precision are significantly lower in the selected studies. To implement the DL-based ECG classification in real clinical scenarios, leveraging diverse ECG databases, designing advanced denoising and data augmentation techniques, integrating novel DL models, and deeper investigation in the inter-patient paradigm could be future research opportunities.

Keywords: electrocardiogram (ECG); arrhythmia; deep learning; convolutional neural network (CNN); inter-patient paradigm; systematic review

Citation: Xiao, Q.; Lee, K.; Mokhtar, S.A.; Ismail, I.; Pauzi, A.L.b.M.; Zhang, Q.; Lim, P.Y. Deep Learning-Based ECG Arrhythmia Classification: A Systematic Review. *Appl. Sci.* **2023**, *13*, 4964. https://doi.org/10.3390/app13084964

Academic Editor: Leo K. Cheng

Received: 4 March 2023
Revised: 7 April 2023
Accepted: 8 April 2023
Published: 14 April 2023

1. Introduction

Cardiovascular diseases (CVDs) are common chronic diseases that pose major threats to human health [1]. Electrocardiogram (ECG) is a kind of noninvasive technique that records the fluctuation of the heart's bio-electric activities. The phenomena of cyclical contractions and relaxations of the heart could be tracked by an ECG machine through electrodes placed on the patient's skin surface. Normal ECG signals consist of different types of waves, including T wave, P wave, and QRS complex. The statistical and morphological characteristics of those ECG waves are important health indicators that could reveal symptoms of heart-related health issues [2]. For example, the absence of P-waves and an irregular ventricular rate in ECG signals could relate to atrial fibrillation (AF) [3]. In daily medical routine, to identify heart abnormalities and provide effective treatment for those issues, cardiologists usually perform ECG screening for patients, which requires significant human efforts and expensive medical procedures. Due to the population aging, the number of patients having cardiovascular diseases is expected to increase explosively, which calls for efficient, accurate, and low-cost automatic ECG diagnosis [4]. In this review, we focus on the classification of heart arrhythmias, i.e., irregular heartbeats, which is a common medical procedure to identify CVDs.

Deep learning (DL) has shown remarkable success in medical diagnosis and has been exploited for automatic heart abnormality classification with ECG signals in recent years. The mapping from ECG features to their corresponding medical categories is learned, which can be characterized by DL models consisting of multiple perception neural layers. The inference capability of the DL model is optimized by a training process with training datasets [5], where the neuron weights are optimized to minimize the mismatch between the inferred and the ground-truth categories of the training data. Compared to traditional machine learning-based classification methods such as clustering and support vector machine (SVM), the DL-based ECG classification could better map the characteristics of ECG signals to their corresponding categories thanks to its powerful multi-level abstraction capability of feature extraction [6]. In this work, the studies which consider DL-based arrhythmia classification with ECG signals are reviewed. The diagnosis for different arrhythmia types is a different clinical problem for cardiologists in practice. However, from the perspective of the classification task with DL, the classification methods for arrhythmia categories share an identical context and goal, i.e., establishing the accurate mapping from ECG characteristics to corresponding categories. Hence, this survey focuses on the current research status, challenges, and research opportunities for deep learning-based arrhythmia classification overall.

According to [7], clinical trials of artificial intelligence-enhanced ECG (AI-ECG) diagnosis have been conducted at the Mayo Clinic for the detection of various cardiovascular diseases, which has demonstrated the potential benefit of AI-ECG. However, they conclude that the implementation of the AI-ECG diagnosis is still in its infancy. Hence, although DL techniques have proven their effectiveness for ECG classification in the research community, their applications in the practical clinical process have been limited due to challenges both from the perspectives of DL techniques and ECG data. For example, in the inter-patient paradigm, DL models need to infer arrhythmia types based on ECG signals from patients who are not included in the training process, which is more challenging than the intra-patient paradigm where the models could experience the same patients during both of the training and inference stages. Hence, as the DL techniques significantly rely on the distribution of data in the feature space while ECG signals vary considerably from person to person, the models trained based on particular ECG datasets may not be applied reliably in practice. As many existing reviews [5,6,8] concentrate mainly on DL algorithms, we consider various factors across the whole DL workflow for the ECG arrhythmia classification. Specifically, our major contributions are as follows.

- We perform a systematic review for DL-based arrhythmia classification with ECG signals from perspectives of ECG database, preprocessing, DL methodology, evaluation paradigm, and performance metrics in the complete DL workflow as well as the code availability of the reviewed studies;
- The trend of techniques in each perspective in recent years is analyzed to summarize the historical road map and illustrate possible future research directions;
- We present the detailed performance gap between the ECG arrhythmia classification under intra- and inter-patient paradigms.

To the best of our knowledge, there is no systematic review on the comparison of DL-based ECG arrhythmia classification under different evaluation paradigms, i.e., intra-patient paradigm vs. inter-patient paradigm. Most existing works consider the intra-patient paradigm; while the investigation in the inter-patient paradigm is limited but more desirable in clinical applications, a thorough comparison between the two paradigms could shed light on future research opportunities.

2. Materials and Methods

2.1. Search Strategy

This systematic review for DL-based ECG arrhythmia diagnosis is performed based on the literature search with four major scholar databases, i.e., Google Scholar, PubMed, Scopus, and the Digital Bibliography and Library Project, focusing on studies published until December 2022. As many studies do not explicitly mention their classification tasks, we first implement a coarse search to include more candidate studies and avoid overlooking studies for arrhythmia classification. Hence, the searching keywords for the literature search are set as (Deep learning OR deep neural network OR convolutional neural network OR CNN OR recurrent neural network OR RNN OR LSTM) AND (ECG OR electrocardiogram).

The detailed paper search and refinement process is shown in Figure 1. A total of 3910 studies are obtained in the initial identification step. After removing the duplicates, 2265 unique studies remain. We then perform the refining process to extract studies that are more relevant to arrhythmia classification with DL. After the initial identification step, the obtained studies go through further screening and eligibility evaluation according to the inclusion and exclusion criteria. Specifically, 1694 studies are excluded by screening their titles and abstracts, and 203 papers are also removed based on the full-text assessment. The inclusion criterion enforces that the studies should be published in English and leverage DL to classify arrhythmia with ECG signals. The studies dealing with other tasks, such as emotion detection and drug and alcohol assessment, should be removed. Those studies which do not have full-text available are also excluded. Hence, a total of 368 studies were selected to be included in this review. The whole process is completed by two independent reviewers (QX and QXZ) in order and rechecked by PYL to ensure fair results of paper search and refinement.

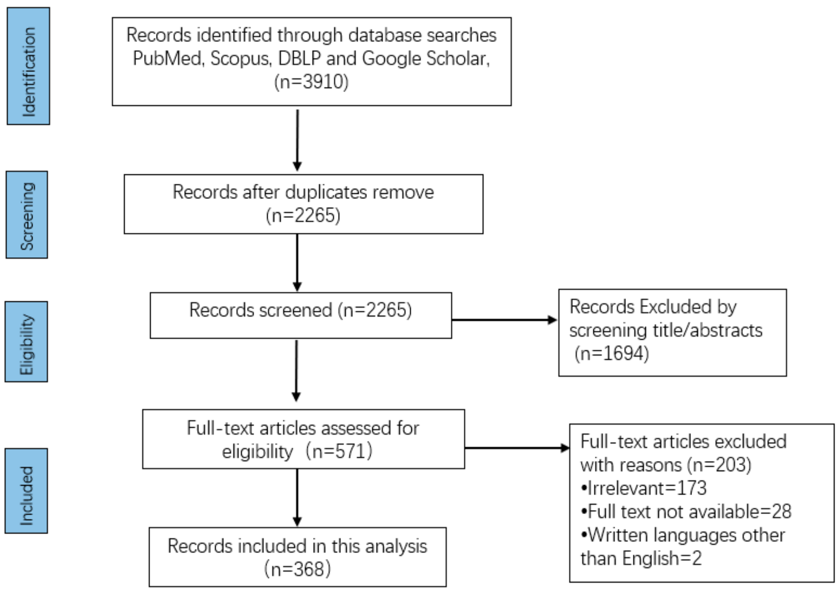


Figure 1. Paper search and refinement process.

2.2. Data Extraction

Table 1 summarizes the data items that are further extracted from the 368 selected studies. This review focuses on diverse aspects, including the general information, ECG database, preprocessing, DL methodology, evaluation paradigm, performance metric, and code availability. A detailed description of the extracted information from those aspects is as follows:

A. **General Information**: An overview of the origin of the selected studies, i.e., the conference proceedings or journals in which they are published and their publication years, is provided;

B. **ECG Database**: The publication information, ECG signal information, and demographic information are analyzed for popular ECG databases employed for arrhythmia classification;
C. **Preprocessing**: Two types of commonly used preprocessing techniques, i.e., denoising to remove artifacts and data augmentation to deal with imbalanced datasets, are summarized;
D. **DL Methodology**: The DL algorithms from all the selected studies are investigated and summarized. The information about the types of DL models, optimization techniques, and classification categories for arrhythmia is presented;
E. **Evaluation Paradigm**: The data-driven ECG diagnosis can be categorized into intra- and inter-patient paradigms depending on how the training and testing ECG data from patients are organized;
F. **Performance Metric**: In addition to the widely used performance metrics such as overall accuracy, other metrics such as sensitivity (Sen), positive predictivity (Ppv), false positive rate (FPR) and F_1 score of the selected studies is discussed;
G. **Code Availability**: Detailed information about studies that publish their code and the source of the code is listed.

Table 1. Extracted information from papers.

	Extracted Items		Definition
A	*General Information*		
		Origin	Journal/conference where the articles were published.
		Publication year	Years of selected studies were published in.
B	*Database*		
		Publication information	Source, release year, and whether the database is public available or not.
		Signal information	Number of channels, sampling rate, signal duration, subject, and records.
		Demographic information	Information about the characteristics of subjects, e.g., gender, age.
C	*Preprocessing*		
		Denoising	Denoising techniques.
		Data Augmentation	Methods to balance data categories.
D	*DL Methodology*		
		Model	Deep learning architecture or framework.
		Optimization	Techniques to optimize the model learnable weights.
		Category	Number of categories of the DL models.
E	*Evaluation Paradigm*		Whether training and testing datasets contain ECG data from the same patients or not.
F	*Performance Metrics*		Metrics to evaluate the classification performance, e.g., F_1, Sp.
G	*Code Availability*		Whether the code is shared online or not.

3. Results

The selected 368 studies consist of 290 journal papers and 78 conference papers focusing on DL-based arrhythmia classification, where 347 (94%) studies were published after the year 2017. Specifically, the number of published works in 2022 is almost four times more than that in 2017 (increasing from 21 works in 2017 to 99 works in 2022), which indicates that the research interest in DL-based ECG arrhythmia classification has been growing significantly in recent years. The top three journals where the selected studies are published are *Computers in Biology and Medicine* (22 studies), *Biomedical Signal Processing and Control* (18 studies), and *IEEE Access* (18 studies). We provide the detailed information of selected papers at Supplementary Table S1.

3.1. Database

DL models require a large amount of ECG signals as the training data to learn the relation between ECG characteristics and the corresponding types of arrhythmias. However, ECG data are considered highly private and sensitive health information, which in general, is difficult to collect from a large group of patients and form a comprehensive database. For the ease of access and the sake of fairly comparing developed DL methods in existing works, the majority of selected studies (89%, 326 out of 368) have established and evaluated their DL models based on ECG datasets from open-source or publicly available databases such as MIT-BIH Arrhythmia Database (MITDB) [9] and MIT-BIH Atrial Fibrillation Database (AFDB) [10]. Table 2 presents the ECG datasets used in the selected works, including their publication information, signal information, demographic information, and the number of selected works that use them for arrhythmia classification. As can be seen in Table 2, MITDB is the most popular, as about 61% (223 out of 368) of works use it for arrhythmia classification. Other popular databases used by more than ten selected works are AFDB, PTB [11], PTB-XL [12], NSRDB, and INCART databases.

Table 2. Popular ECG databases used by the selected studies.

Database	Publicly Available	Release Year	No. Channels	Sampling Rate (Hz)	Duration	Subjects	Collection Place	No. of Records	Demographic Information	Papers
MIT-BIH Arrhythmia Database (MITDB)	Yes	2005	2	360	30 min	47	USA	48	Male: 25, female: 22 (23–89 years old)	223
MIT-BIH Atrial Fibrillation Database (AFDB)	Yes	2000	2	250	10 h	25	USA	23	Subjects are suffering from atrial fibrillation	26
PTB Diagnostic ECG Database (PTB)	Yes	2004	15	1000	2 min	290	Germany	549	Male: 377, female: 139	24
PTB-XL ECG dataset (PTBXL)	Yes	2020	12	500	10 s	18,885	Germany	21,837	Male: 9820; female: 9065	15
MIT-BIH Normal Sinus Rhythm (NSRDB)	Yes	1999	2	/	24 h	18	USA	18	Male:5 (26–45 years old), female:13 (20–50 years old), both no significant arrhythmia;	15
St Petersburg INCART 12-lead Arrhythmia (INCART)	Yes	2008	12	257	30 min	32	Russia	75	Male: 17, female: 15 (18–80 years old). None of the patients had pacemakers; most had ventricular ectopic beats	11
BIDMC Congestive Heart Failure (BIDMC)	Yes	2000	2	250	20 h	15	USA	15	Male:11 (22–710 years old), female: 4 (54–63 years old); subjects with severe congestive heart failure.	10

Table 2. *Cont.*

Database	Publicly Available	Release Year	No. Channels	Sampling Rate (Hz)	Duration	Subjects	Collection Place	No. of Records	Demographic Information	Papers
MIT-BIH Malignant Ventricular Ectopy Database (VFDB)	Yes	1999	2	250	30 min	16	USA	22	Subjects with episodes of sustained ventricular tachycardia, ventricular flutter, and ventricular fibrillation.	9
Chapman University and Shaoxing People's Hospital Dataset (Chapman)	Yes	2020	12	500	10 s	10,646	China	10,646	Male: 5956, female 4690 (4–98 years old). 17% of subjects had normal sinus rhythm, and 83% had at least one abnormality.	8
Fantasia	Yes	2003	12	250	2 h	40	USA	40	20 young (21–34 years old) and 20 elderly (68–86 years old) subjects	7
UCI Machine Learning Repository Arrhythmia Dataset (UCI)	Yes	1998	12	/	/	279	USA	452	203 instances correspond to male subjects; 249 are from female subjects	4
MIT-BIH Noise Stress Test Database (NSTDB)	Yes	1984	2	360	30 min	12	USA	15	Subjects are physically active volunteers	4
Creighton University Ventricular Tachyarrhythmia Database (CUDB)	Yes	2007	1	250	8 min	/	USA	35	Subjects who experienced episodes of sustained ventricular tachycardia, ventricular flutter, and ventricular fibrillation	4
The American Heart Association database short/long (AHA)	No (2 samples available)	1982	2	250	30/150 min	/	USA	10/67	/	4
Chinese Cardiovascular Disease Database (CCDD)	Yes	2012	12	500	10 s	/	China	90	/	4
The QT Dataset (QT)	Yes	1999	2	250	15 min	15	USA and Europe	105	Chosen primarily from among existing ECG databases	3
European ST-T	Yes	2009	2	250	2 h	79	Europe	90	Male: 70 (30–84 years old), female: 8 (55–71 years old); Myocardial ischemia was diagnosed or suspected for each subject	3

In addition, most datasets contain 12-lead ECG signals where ten electrodes are placed in different locations of the human body, such as V1 for the fourth intercostal space on the right sternum and RA (right arm) for anywhere between the right shoulder and right elbow [13]. It results in 12-channel ECG signals where signals of aVR (augmented vector right), aVL (augmented vector left), and aVF (augmented vector foot) channels are obtained based on combinations of ECG signal measurements from other electrodes. The multi-channel ECG signals could better capture additional heart status information based on a greater number of simultaneous measurements. Furthermore, the sampling rates of ECG signals range from 128 Hz to 1000 Hz, and about half of ECG databases (8/17) have a sample rate of 250 Hz, as can be seen in Table 2. Based on the signal duration, the ECG signals can be categorized into long- and short-term measurements, from 10 s to 2 h. Most databases provide gender and age information, and the numbers of females and males tested are generally balanced. Some datasets, such as MITDB, contain both normal and abnormal ECG signals, while most ECG signals in datasets, such as the INCART database, are from patients having ventricular ectopic beats. Compared to the widely-used MIT-BIH series of databases collected in the USA, such as MITDB, NSRDB AFDB, recent ECG databases such as PTB/PTB-XL and Chapman collected in Germany and China, respectively, emerged in the research community, which is considered by increasing studies considered for DL-based ECG arrhythmia classification.

Figure 2 shows the trend of major databases used by selected works each year from 2017 to 2022. One can observe that every year, MITDB is still the dominant database in the research community. The proportion of studies that consider PTB/PTBXL has been increasing in recent years. The diversity of databases is improved as the number of databases used by more than ten studies increases from 6 in 2017 to 9 in 2022.

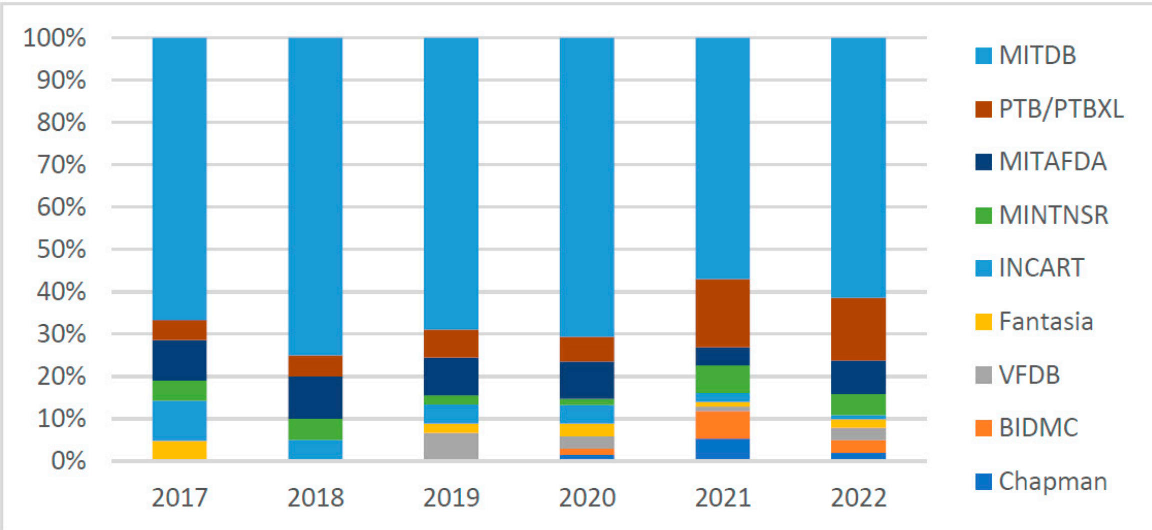


Figure 2. Trend of different ECG databases used by selected studies.

Being the most popular ECG database, MITDB contains 2-lead ECG signals with a sampling rate of 360 Hz and a duration of 30 min. The ECG signals are collected from 47 patients. Recordings 201 and 202 are collected from the same patient, resulting in 48 recordings in total. The age of patients ranges from 23 to 89 years. The duration of each recording is about 30 min. However, the ECG dataset from MITDB is an *imbalanced dataset* where most ECG recordings are normal while abnormal recordings are much less than the normal ones. As the abnormal signals are more difficult to collect, most ECG databases encounter the issue of data imbalance, which could potentially introduce learning bias in the DL-based classification frameworks [14].

Besides ECG datasets obtained specifically for arrhythmia classification, to improve the robustness of the DL models, noisy but normal ECG records can be added to the training dataset. For example, the MIT-BIH Noise Stress Test Database, which is collected from physically active volunteers to mimic ambulatory ECG, acts as a category of noisy ECG signals [15–17]. In this way, the real situation in clinical practice can be emulated.

Among the selected works, 165 (45%) of studies consider more than one ECG dataset by combining multiple different ECG databases. For example, [18] exploits five public ECG datasets, i.e., AFDB, MITDB, NSRDB, the 2017 PhysioNet/CinC Challenge Database, and the first China Physiological Signal Challenge 2018 Database (CPSC2018), where AFDB is used for training and evaluation while other four datasets are used to test the generalization performance of the proposed DL model. This mechanism of training and testing DL models with ECG signals from two non-overlapping groups of patients, respectively, is a typical case of *inter-patient diagnosis*. However, as those datasets have different attributes such as categories and numbers of channels, a smaller number of classification categories, such as categories of Atrial fibrillation (AF)/non-AF and categories of AF, Normal, Premature Atrial Contractions (PAC), Premature Ventricular Contractions (PVC), Ventricular fibrillation (VF), and Noise are often considered [19,20]. In [21], ECG signals from MITDB, MIT-BIH AFDB, CUDB, and MIT-BIH VFDB are fused to form one dataset where the training and testing datasets are obtained by randomly selecting ECG data from the combined dataset. Hence, the *intra-patient diagnosis* is performed where the DL model has the possibility to train and test based on ECG information from the same patient. By mixing multiple ECG datasets, the issue of imbalance in data categories can also be alleviated [22]. Regardless of inter- or intra-patient diagnosis, it shows a clear trend over the last few years that increasing studies exploit combined ECG datasets for DL-based ECG arrhythmia analysis [18,21,23–26].

3.2. Preprocessing

Before inputting ECG signals into DL models, a preprocessing step is often applied to those signals, which could improve the learning efficiency and reduce the computational complexity of DL models [27]. In this review, the preprocessing step is reviewed from two aspects, i.e., denoising [28] and data augmentation [29]. The two deal with noisy ECG signals and imbalanced datasets, respectively, which are common cases in real clinical scenarios.

3.2.1. Denosing

The ECG signals are prone to be contaminated by background noise and bioelectrical inference, such as power-line noise and muscle movement. The denoising step could clean the ECG signals to prevent overwhelming micro features in signals and help DL models focus more on the ECG features [30]. Based on the selected studies, only about 38% of selected works (138 out of 368) specifically mentioned their denoising methods, and those methods can be mainly categorized into three types, i.e., traditional filter-based denoising methods (45.9%, 62 out of 138), wavelet-based denoising methods (38.4%, 53 out of 138), and hybrid denoising methods (16.7%, 23 out of 138). The traditional denoising filters, such as lowpass, bandpass, and notch filters, assume that the noise and useful signals lie in different frequency bands. Other denoising filters include smoothing filters such as the median filter and the Savitzky–Golay (S-G) [23,31–33] and adaptive filters [34,35]. The discrete wavelet transform (DWT) could project ECG signals onto the time-frequency domain based on wavelet basis functions [36]. To remove the noise, the wavelet coefficients at high-frequency bands can be simply set to zero or apply a thresholding process to set the modest wavelet coefficients to zero [19–22] based on the assumption that the useful ECG signal is similar to the selected wavelet basis function. A combination of different types of denoising methods can be applied for noise removal, e.g., [37,38] combines DWT, median filters, or S-G filters for denoising. However, this type of method will induce higher processing latency.

The frequency counts of the three types of methods in each year are presented in Figure 3a. The number of works considering denoising has been increasing in recent years. The traditional filter-based methods are more popular than the other two denoising methods because of their effectiveness but easier implementation. In addition, there have been increasing works that consider wavelet-based methods for ECG signal denoising in recent years.

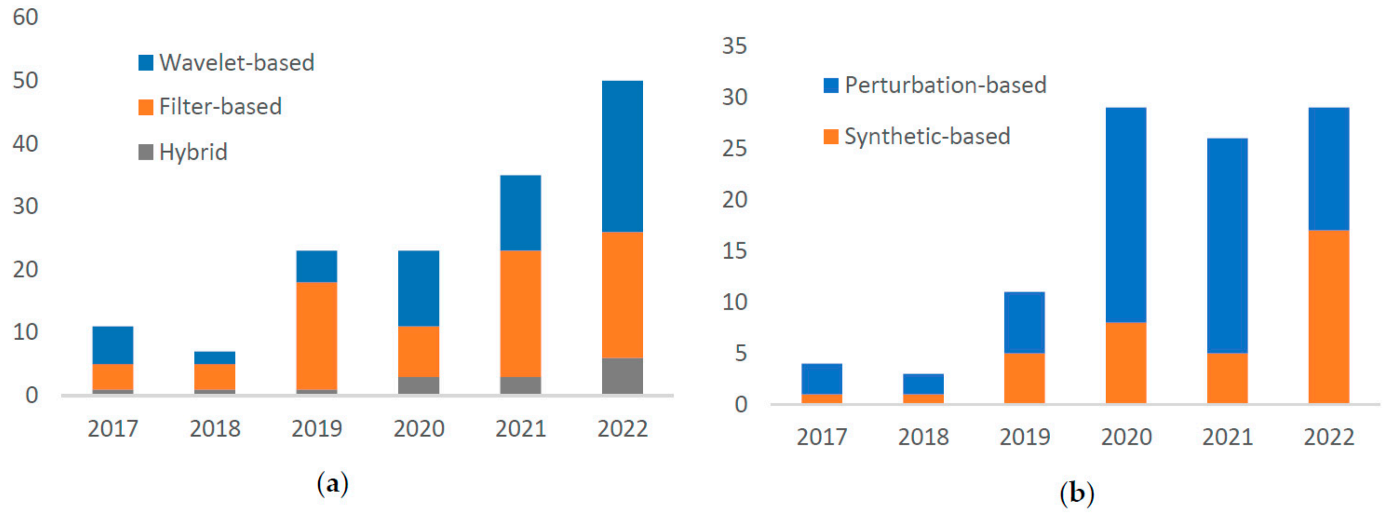


Figure 3. Distribution of preprocessing methods used by selected studies in each year. (**a**) Trend of different types of denoising methods; (**b**) Trend of different types of data augmentation methods.

3.2.2. Data Augmentation

ECG data often has biased distributions of abnormal categories much less than normal categories, as the abnormal signals are more difficult to obtain. DL models trained with the imbalanced ECG dataset will, in nature, put more attention to majority categories and overlook the minority categories leading to biased learning. In this survey, we focus on the data argumentation technologies [39], which take effect during the data preprocessing step to gain more training samples. From the selected studies, 102 (28%) studies explicitly claim the use of data augmentation techniques in their work. The augmentation techniques can be categorized into two types, i.e., perturbation-based methods (64%, 65 out of 102) and synthetic-based methods (36%). Specifically, for perturbation-based methods, extra data samples can be added to ECG dataset by adjusting or perturbating the original samples from the same dataset, such as scaling and shifting ECG waveforms [40] or adding artificial noise to existing ECG signals [41]. The perturbation of data samples is essentially acquiring new data samples from the neighborhood of corresponding original data samples in the feature space. Hence, the new data samples could be highly correlated to the original samples based on which the new data is perturbated. On the other hand, synthetic-based methods generate synthetic ECG data either based on the linear combination of real data samples or the construction of ECG signals by imitating real ECG features. The synthetic minority oversampling technique (SMOTE) and its variants, such as SMOTENN [42], Borderline SMOTE [43], and SVM-SMOTE [42–45], are often used to extend the minority categories. Just recently, DL techniques have also been used for synthetic data generation, e.g., the convolutional neural style transfer network [46], the generative adversarial network (GAN) [47], and the ACGAN consists of variational auto-encoder model [14].

Figure 3b shows frequency counts of the two types of augmentation methods each year. One can see that the synthetic-based strategies have drawn more attentions in recent years as the number of works considering this type of data augmentation method increased from 1 in 2017 to 17 in 2022.

3.3. DL Methodology

3.3.1. Model

The design of DL models is crucial to the pipeline of DL-based ECG arrhythmia classification. The DL models have multi-level or multi-layer structures, and each level or layer can be regarded as a feature extractor that can learn how to better summarize signal characteristics [48]. Based on the intrinsic property of the major feature extractor within the neural networks, the DL classification models considered in the selected studies can mainly be categorized into the following types: convolutional neural networks (CNNs), recurrent neural networks (RNNs), including the long short-term memory (LSTM) and bidirectional LSTM (BiLSTM), transformer, "hybrid" which refers to combinations of different DL models, and "others" corresponding to less popular models such as restricted Boltzmann machines and deep-belief networks. The detailed analysis of those DL models for ECG arrhythmia classification is as follows.

- CNN

CNN is a DL model widely used in image classification, signal analysis, and natural language processing [48]. Each layer of CNN usually contains a convolutional filter followed by pooling operations to extract both local and global features [49]. Depending on the number of filtering directions of the convolutional filters in the spatial domain, the CNN can be further categorized into 1D CNN and 2D CNN. Specifically, the filters in 1D CNN and 2D CNN move along one and two filtering directions, i.e., feature dimensions, respectively. In general, 1D CNN is suitable for raw or denoised ECG signals, which only have one single feature dimension. For instance, in [50], an adaptive 1D CNN is proposed for ECG classification and anomaly detection at any sampling rate of ECG signals to avoid hand-crafted feature extraction. In [51], a lightweight 1D CNN considering channel shuffle over the group and depth-wise convolutions is designed, where 2-s ECG signal segments are considered as model input [37]. In [38], the 1D CNN is leveraged to classify 2, 5, and 20 types of heart diseases where few-shot learning is considered to deal with the small-size of the dataset. On the other hand, 2DCNN mainly takes into account the image-like input, such as the spectrogram and scalogram of ECG signals. In [52], the 2D scalogram is obtained by transforming the 1D ECG signals having 500 samples to the wavelet domain using continuous wavelet transform. Then the 2D scalogram is regarded as a 3-channel color image with a size of 227×227 in the spatial domain. A classic 2D CNN, i.e., AlexNet [53], is used to classify ECG signals. In [54], the plot of 1D ECG recordings is directly transformed to 2D gray-level images with a size of 15×15 which are then fed as input for the 2D CNN. In [55], a multi-lead CNN takes multi-lead ECG as the matrix input, where the sub-2D convolution and lead asymmetrical pooling are exploited to extract multi-scale features. Due to simpler operation compared to 2D convolution, 1D CNN often contains fewer learnable parameters and has higher computation speed, making it suitable for real-time ECG classification and is often easier to be deployed in hardware.

- RNN/LSTM/BiLSTM

Taking into account the temporal correlation of feature sequences, RNN is a type of DL structure that considers the input as a time series. As ECG signals are time series in nature, their temporal correlation within the signals could potentially better reveal the sign of their categories. For typical RNNs, the information in their hidden layer at the current moment does not only depend on the current input but relies on the information at the previous time instance [56]. In this way, the RNN is more sensitive to the temporal features of the input sequence and is advantageous in capturing hidden temporal information in ECG features [56]. Furthermore, the improved RNN, i.e., the long short-term memory (LSTM), gains higher popularity than the conventional RNN because of its higher capability to analyze time series. Specifically, the LSTM has three gate structures to control the output, input, and forget information flow in stored memory cells [57]. Compared to the RNN, the LSTM could deal with longer signal sequences as it selectively acquires useful information from historical inputs. In [58], a 6-layer LSTM is developed to automatically identify PVC beats based on ECG sequences. Furthermore, bidirectional LSTM (BiLSTM) is a special type of LSTM consisting of two LSTMs that go through the input sequence along the temporal direction forwardly and reversely, respectively [32]. Hence, it could capture both the causal and noncausal time dependency information of signals to pursue potential better classification performance. In [59], the BiLSTM model is used for ECG classification based on the extracted ECG wave statistics along the temporal dimension, including RR interval, QR interval, ST segment starting point, and amplitudes of Q- and R-waves. In [60], a 2D

BiLSTM is used for AF detection based on the spectrogram of ECG signals, where the input features are the frequency components at each time instance. In [61], the BiLSTM taking the sequence of RR intervals as input, is proposed for AF detection. To summarize, the input sequences for RNNs can be the raw ECG sequences, time-varying wave statistics, and time-frequency representation of ECG.

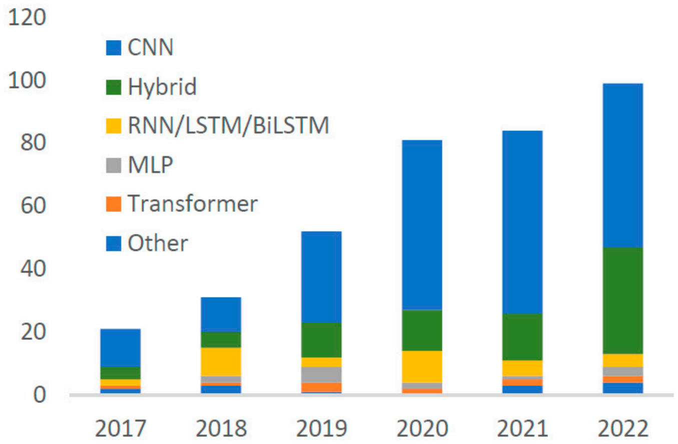
- Transformer

The attention mechanism gained more popularity in recent DL research communities as it is capable of learning how to assign higher learning weights to significant features [62]. The transformer is an encoder–decoder structure that consists of only attention mechanisms and fully connected layers [63]. It was originally designed for natural language processing (NLP) but has been extended to other applications since it could achieve better performance than RNN/LSTM [64]. In [65], the encoder part of the transformer is used for heartbeat classification with ECG signals where the heart beat sequences are considered as input. In addition, RR intervals are concatenated with the features extracted by the attention module for final classification. In [66], a random forest model is first used to select 22 important features, such as RR interval median and P-wave correlation coefficients. Then the encoder of the transformer is exploited to extract features directly from ECG signals. The combination of the hand-crafted features and the features automatically extracted by the transformer is used for ECG classification. A waveform transformer is proposed in [67] in which the input ECG segments are first projected to a 1D vector through a multi-layer perceptron. Then the embedded segments, together with positional embedding and learnable class embedding, are taken as the input for the transformer encoder. The extracted features from the transformer are combined with 22 static features together for the final ECG classification. The transformer was developed in 2017, and its application to ECG signal is still in its early stage; however, more results with the transformer are expected in the future.

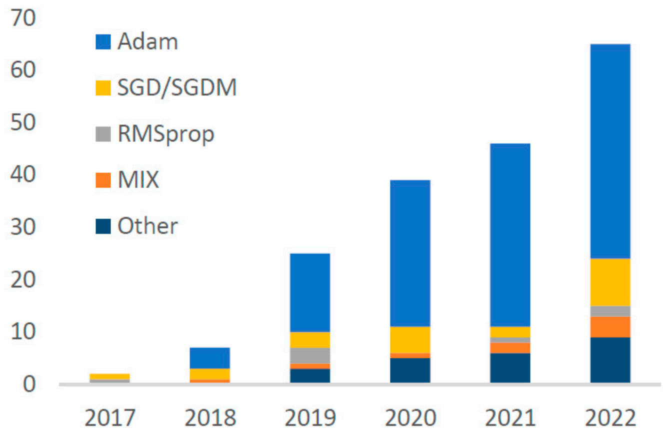
- Hybrid DL model

Many selected studies consider integrating multiple DL models into one DL network for ECG arrhythmia classification. For example, in [68], it combines the CNN and the RNN to form an encoder–decoder structure for heartbeat classification. CNN is used for feature extraction, and RNNs are used to translate the extracted features to their corresponding categories. More examples of the combination of CNN with LSTM and BiLSTM can be seen in [69–71], where CNNs are stacked in front of LSTM/BiLSTM modules for feature extraction. In [63], 1D CNN is first used to extract the features from ECG sequences. Then the CNN features are added with the positional encoding to further serve as the input for a transformer to finally detect the ECG arrhythmia. In [72], a 1D CNN is exploited for local attention embedding, and the encoder of the transformer is used for further feature extraction. In [23], shallow-domain knowledge-injection attention is first to extract the ECG signal feature. Then the attention outputs from the original and smoothed ECG data are regarded as the multivariate input for the 2D classification CNN. More works considering combining CNNs with transformers can refer to [35,73]. The CNNs are also combined with attention mechanisms in [74–76]. In the selected studies, 82 studies take advantage of hybrid models, which assemble different types of DL models to classify ECG arrhythmia. The top 3 hybrid models include CNN+LSTM (24 studies), CNN+BiLSTM (15 studies), and CNN+RNN (8 studies). In most hybrid models, the CNNs often serve as feature extractors, followed by other models which perform further feature extraction.

As shown in Figure 4a, the proportions of different DL models used in the selected studies are presented. Overall, the CNN (58.7%, 216 out of 368), RNN/LSTM/BiLSTM (9%, 33 out of 368), and hybrid (22.3%, 82 out of 368) are the most popular DL models for arrhythmia classification. Each year, there are more selected works considering CNN models than those considering other models each year but the number of works considering the hybrid model has been increasing in recent years.



(a)



(b)

Figure 4. Distribution of architectures used by selected studies in each year. (**a**) Trend of different DL models used by the selected studies; (**b**) Trend of optimization techniques.

3.3.2. Optimizer

The way to optimize the DL models' learnable weights through backpropagation is another important control knob for classification performance. Figure 4b shows the trend of optimization techniques mentioned in the selected studies within years. A total of 50% (184 out of 368) studies did not explicitly report their optimization method. Three most frequently used optimizes are adaptive moment estimation (Adam) (66.8%, 123 out of 184), Stochastic gradient descent (SGD)/ SGD with momentum (SGDM) (12%, 22 out of 184), and root mean square propagation (RMSProp) (3.8%, 7 out of 184).

3.3.3. Classification Categories

Out of 368 selected studies, 118 (32%) studies categorized ECG signals into five classes. The large proportion stems from the fact that most studies utilize MITDB as their ECG databases, where ECG signals have been categorized into five essential groups (N: Normal beat; S: Supraventricular ectopic beat; V: Ventricular ectopic beat; F: Fusion beat; Q: unknown beat) following the American Association of Medical Instrumentation (AAMI) standards [77]. Some studies [68,78] follow the AAMI standards but calculate the classification performance of categories of N, S, V, and F, which account for major categories in the ECG dataset. Binary classification (19%, 73 out of 368) is mostly used to identify one certain arrhythmia type, such as AF [79] and left ventricular dysfunction [80]. The conclusions from many studies suggest that accurate multi-class arrhythmia classification is more challenging [19,20].

3.4. Evaluation Paradigm

The model generalization performance of DL models is a crucial perspective to be considered in the step of model evaluation. The generalization performance refers to the capability of classification models to infer categories of previously unseen and new data. For ECG classification, two evaluation paradigms have been investigated to evaluate the classification capability of DL models, i.e., the intra- and inter-patient paradigms, as depicted in Figure 5a. Specifically, in the inter-patient paradigm, the learning model trained on ECG signals from one group of patients is evaluated with different groups of patients which do not overlap with the training group. The intra-patient paradigm refers to the case that the DL mode could be trained and evaluated based on ECG signals from the same patients.

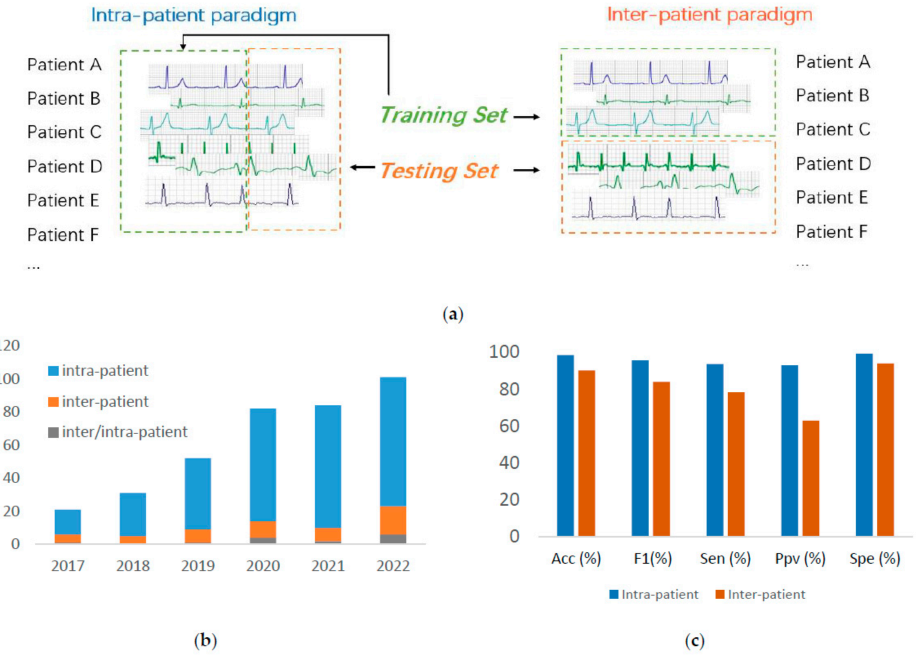


Figure 5. Comparison of the intra- and inter-patient paradigms. (**a**) An illustration of the intra- and inter-patient paradigms; (**b**) Trend of the intra- and inter-patient paradigms; (**c**) Comparison of classification performance achieved in the inter- and intra-patient paradigms.

Among all the selected studies, 27 studies focus on the inter-patient paradigm, while a significant number of studies (319) consider the intra-patient diagnosis. In addition, a total of 11 studies consider both paradigms, while few studies do not describe their paradigm explicitly. As can be seen in Figure 5b, the proportion of the selected studies considering the inter-patient paradigm has been increasing in recent years as it is more desirable for clinical applications in practice.

Detailed information about the selected studies, which consider the inter-patient paradigm, is presented in Table 3. It summarizes the specific ECG data used for training/validation and testing, deep learning algorithm, classification category, and classification performance. The MITDS1/DS2 method (82%, 31 of 38) is the most popular evaluation method for the inter-patient paradigm. Specifically, the ECG data in MITDB is divided into two groups, i.e., DS1 and DS2, where 22 records are included. The details about how to obtain the standard DS1 and DS2 are illustrated in [77]. Please note that the MITDS1/DS2 method is modified in some studies [63] for ECG analysis, where different recordings are included in DS1 and DS2. Additionally, some works consider leveraging ECG data from one database for training and testing the trained models with different ECG databases. For example, the model in [81] trains based on AFDB and then tests the model with MITDB. The number of classification categories in those selected works varies from 2 to 9 categories.

Table 3. The inter-patient paradigm studies in the selected articles.

Paper	Train/Validate Data	Test Data	Algorithm	Class	Performance				
					Acc (%)	F_1 (%)	Sen (%)	Ppv (%)	Spe (%)
[82]	MITDB DS1	MITDB DS2	DenseNet-BiLSTM	5: N, S, V, F, Q	92.37	63.49	68.29	60.35	94.51
[83]	MITDB DS1	MITDB DS2	CNN	5: N, S, V, F, Q	88.34	/	90.9	48.25	88.51
[84]	Fantasia + INCART	Fantasia + INCART	CNN-LSTM	2: N, CAD	95.76	95.57	95.7	/	95.76
[85]	MITDB DS1	MITDB DS2	O-WCNN	5: N, S, V, F, Q	99.43	92.05	91.06	93.5	99.69

Table 3. *Cont.*

Paper	Train/Validate Data	Test Data	Algorithm	Class	Performance				
					Acc (%)	F_1 (%)	Sen (%)	Ppv (%)	Spe (%)
[81]	AFDB AFDB	MITDB NSRDB	CNN+RNN	2:AF; NoAF	89.3 /	/ /	99.82 /	51.71 /	87.94 95.01
[86]	MITDB DS1	MITDB DS2	CNN+BLSTM	5: N, S, V, F, Q	96.77	77.84	74.89	81.24	95.16
[28]	PTB-XL (by subject)	PTB-XL (by subject)	CNN-FWS	2: N; abnormal	90.05	90.2	88.9	91.5	/
[87]	PTB	PTB	MLA-CNN-BiGRU	6	62.94	/	63.97	63	/
[88]	MITDB DS1	MITDB DS2	CNN	5: N, S, V, F, Q	96.36	/	70.6	48.1	96.16
[89]	Tongji Hospital, China	CPSC 2018	DCNN	6	/	84.2	80	/	98
[90]	MITDB DS1	MITDB DS2	CNN	5: N, S, V, F, Q	90.22	/	35.64	27.71	87.87
[78]	MITDB DS1 MITDB DS1 MITDB DS1	MITDB DS2 MIT-BIH-SUP INCART	DDCNN + CLSM	5: N, S, V, F, Q	95.1 88.2 91.6	84 / /	87.2 77.7 88	82.4 64.7 65.3	/ / /
[91]	MITDB DS1	MITDB DS2	CNN-LSTM	5: N, S, V, F, Q	95.81	71.06	69.20	74.94	94.56
[92]	MITDB DS1	MITDB DS2	DRCNN	5: N, S, V, F, Q	88.99	/	52.10	56.82	94.75
[93]	MITDB DS1	MITDB DS2	OptRPC	5: N, S, V, F, Q	98.48	/	98.45	98.43	98.06
[94]	UVA Holter Recordings	MITDB	CNN+RNN	4: NSR; AF; Other; noise	/	75.5	/	/	/
[95]	MITDB DS1	MITDB DS2	CNN	2: PVC; no PVC	/	98.90	99.20	98.60	/
[29]	MITDB DS1	MITDB DS2	SE-ResNet	5: N, S, V, F, Q	99.61	/	93.78	/	/
[96]	MITDB DS1 MITDB MITDB	MITDB DS2 SVDB INCARTDB	CNN	5: N, S, V, F, Q	84,62 84.17 95.36	/ / /	/ / /	/ / /	/ / /
[97]	MITDB DS1	MITDB DS2	RBM	5: N, S, V, F, Q	98.61	/	87.31	/	98.76
[98]	MITDB DS1	MITDB DS2	Faster R-CNN	5: N, S, V, F, Q	95.68	/	72.8	90	/
[99]	MITDB DS1	MITDB DS2	CNN	5: N, S, V, F, Q	94.70	88.9	89	93.7	/
[100]	MITDB DS1	MITDB DS2	FNN+CNN	5: N, S, V, F, Q	94.2	/	58.2	53.6	/
[101]	MITDB DS1	MITDB DS2	CWT+CNN	5: N, S, V, F, Q	98.74	68.76	67.47	70.75	/
[102]	MITDB DS1	MITDB DS2	WaveNet-LSTM	5: N, S, V, F, Q	96.80	/	/	/	/
[42]	MITDB DS1	MITDB DS2	DHCAF	5: N, S, V, F, Q	93.0	/	75.1	70.4	/
[103]	MITDB DS1	MITDB DS2	CraftNet	5: N, S, V, F, Q	89.24	/	89.25	61.84	95.79
[48]	MITDB DS1	MITDB DS2	BiLSTM	5: N, S, V, F, Q	97.3	/	77.9	/	/
[27]	MITDB DS1	MITDB DS2	CNN	5: N, S, V, F, Q	92.3	/	73.50	68.33	/
[104]	MITDB DS1	MITDB DS2	CNNs	5: N, S, V, F, Q	98.6	/	93	/	/
[4]	MITDB DS1 MITDB DS1	MITDB DS2 MITDB DS2	DNN	2: S, non-S 2: V, non-V	/ /	/ /	61.4 91.8	 /	98.3 99.5
[105]	MITDB DS1	MITDB DS2	FE-CNN	5: N, S, V, F, Q	98.6	88.0	84.2	92.3	99.45
[106]	MITDB DS1	MITDB DS2	LSTM	5: N, S, V, F, Q	/	/	74.91	76.16	/
[107]	MITDB DS1	MITDB DS2	RBM	5: N, S, V, F, Q	95.20	/	83.07	51.42	/
[31]	MITDB DS1	MITDB DS2	DNN	5: N, S, V, F, Q	97.5	/	85.9	84.4	/
[68]	MITDB DS1	MITDB DS2	BiLSTM	5: N, S, V, F, Q	99.53	/	96.19	97.21	98.58
[108]	CPSC	MITDB	RBNN	3: L; R; O	78.58	/	78.57	/	/
[109]	MITDB DS1 MITDB MITDB DS1; SVDB	MITDB DS2 SVDB MITDB DS2	CRNN	5: N, S, V, F, Q 5: N, S, V, F, Q 5: N, S, V, F, Q	93.66 75.33 93.04	75.85 42.11 75.18	76.98 47.36 80.53	74.76 61.66 70.89	95.59 89.12 94.57

Acc: overall accuracy; F_1: F_1 score; Sen: Sensitivity; Ppv: Positive predictivity; Spe: Specifically.

As the motivations, tasks, datasets, and classification methods of all the reviewed studies are different, it is unfair and not straightforward to compare the classification performance across all the reviewed studies. However, we still intuitively compare the average performance metrics of all the selected studies in the inter-patient paradigm, regardless of the number of categories to be classified. The averaged Acc, F_1 score, Sen, Ppv, and Spe are 92.62%, 79.48%, 79.25%, 71.74%, and 95.26%, respectively. Although those statistics are biased due to the way they are calculated, it intuitively shows that the classification performance can still be improved as some performance metrics, i.e., F_1 score, Sen, and Ppv are significantly lower on average than the other two.

Furthermore, Table 4 illustrates similar information about the 11 studies which investigate both the inter- and intra-patient paradigms. The averaged values of classification accuracy of the 11 studies are 98.39% and 90.15% in the intra- and inter-patient paradigms, respectively. As shown in Figure 5c, the averaged values of F_1 score, Sen, Ppv, and Spe for 11 studies in the intra-patient paradigm (inter-patient paradigm) are 95.52% (83.89%), 93.51% (78.16%), 92.78% (62.82%), and 99.19% (93.86%), respectively. The differences in the F_1 score, Sen, and Ppv between the two paradigms are 11.63%, 15.35%, and 29.96%, which are much higher than the difference in terms of accuracy. Therefore, the inter-patient paradigm is a more challenging scenario, which calls for more research attention and effort.

Table 4. Comparative summary table of inter/intra-patient paradigms studies.

Paper	Paradigm	Train/Validate Data	Test Data	Algorithm	Class	Cross-Validation	Performance				
							Acc	F_1	Sen	Ppv	Spe
[82]	Intra- Inter-	MITDB MITDB DS1	MITDB MITDB DS2	DenseNet-BiLSTM	4: N, S, V, F	10-fold	99.44 92.37	95.89 63.49	95.69 68.29	96.11 60.35	99.32 94.51
[83]	Intra- Inter-	MITDB MITDB DS1	MITDB MITDB DS2	CNN	5: N, S, V, F, Q	10-fold	99.48 88.34	/ /	96.97 90.90	98.83 48.25	99.87 88.51
[84]	Intra-	Fantasia+ INCART	Fantasia+ INCART	CNN-LSTM	2: N, CAD	/	99.85	99.52	99.85	/	99.84
	Inter-	Fantasia+ INCART (by Subjects)	Fantasia + INCART (by Subjects)				95.76	95.57	95.70	/	95.76
[85]	Intra- Inter-	MITDB MITDB DS1	MITDB MITDB DS2	O-WCNN	4: N, S, V, F	10-fold	99.58 99.43	99.28 92.05	99.2 91.06	/ 93.50	99.15 99.69
[81]	Intra- Inter- Inter-	AFDB AFDB AFDB	AFDB MITDB NSRDB	CNN+RNN	2:AF; NoAF	5-fold	97.80 89.30 /	/ / /	98.98 99.82 /	95.76 51.71 /	96.95 87.94 95.01
[86]	Intra- Inter-	MITDB MITDB DS1	MITDB MITDB DS2	CNN+BLSTM	5: N, S, V, F, Q	10-fold	99.56 96.77	96.40 77.84	95.90 74.89	97.14 81.24	99.47 95.16
[28]	Intra- Inter-	PTB-XL PTB-XL (by subject)	PTB-XL PTB-XL (by subject)	CNN-FWS	2: N; abnormal	/	89.92 90.05	90.70 90.20	91.40 88.90	90.01 91.50	/ /
[87]	Intra- Inter-	PTB PTB	PTB PTB	MLA-CNN-BiGRU	6	5-fold	99.11 62.94	/ /	99.02 63.97	99.10 63.00	/ /
[88]	Intra- Inter-	MITDB MITDB DS1	MITDB MITDB DS2	CNN	5: N, S, V, F, Q	/	99.81 96.36	/ /	88.82 70.6	95.68 48.10	99.54 96.16
[89]	Intra-	Tongji Hospital, Database China	Tongji Hospital, Database China	DCNN	26	/	/	91.3	89.1	/	99.7
	Inter-	Tongji Hospital, China	CPSC 2018		6		/	84.2	80.0	/	98.0
[90]	Intra- Inter-	MITDB MITDB DS1	MITDB MITDB DS2	CNN	5: N, S, V, F, Q	/	99.31 90.22	/ /	73.66 35.64	69.6 27.71	98.83 87.87
Average performance compartment of above studies					Intra-patient Inter-patient Differences		98.39 90.15 8.24	95.52 83.89 11.63	93.51 78.16 15.35	92.78 62.82 29.96	99.19 93.86 5.33

Acc: overall accuracy; F_1: F_1 score; Sen: Sensitivity; Ppv: Positive predictivity; Spe: Specifically.

3.5. Performance Metrics

To evaluate the classification performance of DL models, the commonly used performance metrics are overall accuracy (Acc), sensitivity (Sen), positive predictivity (Ppv), false positive rate (FPR), and F_1 score. Sen and Ppv correspond to recall and precision rates, respectively. It can be seen that Acc (82.1%, 302), Sen (72%, 265), F_1 score (59.8%, 220) are the three most used performance metrics to evaluate the classification performance of the DL models for arrhythmia classification.

Regardless of the evaluation paradigm and other conditions, such as task and number categories to be classified, we simply calculate the average of those metrics and find that the classification accuracy of the DL models in the selected studies is already above 95%, while other metrics such as F_1 score, sensitivity, positive predictivity, and specificity are relatively lower, within the range of 80–95%. Furthermore, interesting comparisons between the DL models and professional cardiologists are performed in [110,111]. It can be concluded that the DL models are very competitive to cardiologists, which exhibits their great potential for clinical ECG classification.

3.6. Code Availability

Sharing the code of DL models online is a way for researchers to reproduce the performance results of existing works. However, only around 6% (20 out of 368) provide the code information directly in their papers, and most of them shared the code through the GitHub platform. In Table 5, to help researchers access the relevant codes conveniently, we list detailed information about the studies whose codes are publicly available. Few studies mention that the codes are available upon request, e.g., [85,112] but are not listed in the table.

Table 5. The details information of code available studies.

Paper	Code Platform	Database	Evaluation Paradigm	Class Types	Model
[23]	Website	UEA; UCR; PhysioNet. AF-D1, AF-D2, and Two Lead	Intra-patient	2	SDK-CNN
[29]	GitHub	MITDB	Intra-patient	5	CNN (ResNet)
[113]	GitHub	MITDB; CPSC2018	Intra-patient	5	DNN
[114]	GitHub	MNIST and Fashion MNIST; MITBIH	Intra-patient	2	D-RBFDD
[115]	GitHub	PTBXL; CPSC2018	Intra-patient	2	SPN-V2
[116]	GitHub	UCI Machine-learning repository-based arrhythmia dataset; MITDB	Intra-patient	13	CDNN
[75]	GitHub	2021 PhysioNet Challenge data	Intra-patient	30	d-RINCA
[20]	GitHub	2020 Physionet Challenge data	Intra-patient	24	CNN (ResNet)
[117]	GitHub	Chapman University and Shaoxing People’s Hospital Dataset	Intra-patient	4	Hybrid DNN
[118]	GitHub	Physionet Challenge database; AFDB; AF Termination database	Intra-patient	2	CNN+ BiLSTM
[40]	GitHub	CPSC2018	Intra-patient	9	DNN
[98]	GitHub	MITDB	Intra-/Inter-patient	5	Faster R-CNN-DNN
[70]	Gitlab	MITDB	Intra-patient	5	CNN-BLSTM

Table 5. *Cont.*

Paper	Code Platform	Database	Evaluation Paradigm	Class Types	Model
[101]	GitHub	MITDB	Intra-patient	5	CWT+CNN
[111]	GitHub	Telehealth Network of Minas Gerais (TNMG)	Intra-patient	6	DNN
[119]	GitHub	MITDB; CPSC2018	Intra-patient	17	CNN+ BiLSTM
[120]	Website	CPSC2018	Intra-patient	9	CNN
[121]	Website	2017 PhysioNet Challenge data	Intra-patient	12	DNN
[68]	GitHub	MITDB	Intra-patient	5	Hybrid CNN
[122]	Website	2017 PhysioNet Challenge data	Intra-patient	4	CNN

4. Discussion

In this section, the findings from the selected papers in the field of DL-based ECG arrhythmia classification are summarized from the perspectives of the ECG database, preprocessing, DL methodology, evaluation paradigm, and performance metric. In addition, future challenges and possible directions are also discussed accordingly.

4.1. ECG Database

The quality of data plays a vital role in achieving high classification performance [37]. DL techniques highly depend on the training data from which they learn the relationship between data characteristics and corresponding categories. As ECG signals are considered private medical information, they are, in general, difficult to collect from a broad range of patients having different genders and ages. In addition, the measurement conditions should be kept unified for every patient, and the annotation for ECG signal samples should be precise, which all require standard facilities and significant annotation effort from cardiologists [5]. Hence, at the current stage, the publicly available ECG databases are the main data resources for DL-based ECG arrhythmia classification and support the progress of the research. However, the symptoms of arrhythmia in ECG signals vary from person to person. Exploiting diverse ECG databases to help DL models experience a greater amount of data samples could significantly improve their inference performance in practical clinical applications.

According to this review, among the multiple ECG databases, the MITDB is the most popular database for DL-based arrhythmia classification, while it was collected about 40 years ago [9]. The MITDB actually acts as the data baseline to help compare newly designed DL methods to well-established models. Nowadays, as the model complexity of DL models increases to pursue higher classification performance, it inevitably consumes more ECG data for training. Hence, there have been growing works that consider combining datasets from multiple different public ECG databases [16,123]. Another example is that, in PhysioNet Computing in Cardiology Challenge 2020, seven public databases are provided to participants. However, the differences in the number of leads, signal duration, measurement condition, and patient demographic distribution of those different ECG databases should not be simply ignored.

Another issue with the existing arrhythmia-related ECG database is that the data categories are significantly imbalanced. The amount of ECG data in the normal category is often dominant in those databases [10]. Although various methods, such as data augmentation [124] and focal loss [92], are used to address the issue of imbalanced datasets, collecting more data in abnormal categories is the ideal way to entirely resolve the issue. However, collecting the specific data requires patients who exactly have the diseases to be classified, which often is difficult in reality. Hence, the imbalance of the ECG dataset is one of the challenges that researchers should expect to confront in the long term.

4.2. Preprocessing

Real clinical ECG signals often contain diverse noise and interference. However, many existing works do not consider ECG signal denoising, which could raise risks when they are implemented for real clinical scenarios. Thorough studies about the impact of ECG denoising on the DL-based ECG classification methods in clinical applications are in demand. Furthermore, the existing denoising methods assume that the characteristics of noise and interference differ from that of the useful signals in a predetermined domain, such as the frequency domain and the wavelet domain [96]. In general, the DWT-based denoising methods could better retain the details of higher frequency signal components compared to traditional filtering-based denoising methods, which rely on discrete Fourier transform [36]. However, the DWT-based denoising methods are offline algorithms that cannot apply to real-time ECG signal denoising [125]. Design for advanced denoising algorithms specifically considering ECG characteristics is needed to be explored.

As ECG datasets are often imbalanced, data augmentation could be a necessary step to further level up the performance of DL models. It can be concluded that synthetic-based methods have gained higher popularity in recent years. In particular, the GAN-style model actually learns to generate synthetic ECG signals based on the characteristics of ECG signals [126]. However, this type of method is still a data-driven process that relies heavily on the quality of ECG data. Research [46] proposes a DL model which jointly considers synthetic ECG signals generated by a mathematical model and real clinical ECG data. It could be a better solution for imbalanced data as the generated data set as the generated new data samples are not simply the linear combinations of other data samples but contain the theoretical a priori knowledge from cardiologists.

4.3. DL Methodology

This review clearly shows that CNNs are the most popular DL models for ECG classification thanks to their excellent capability for feature extraction [8]. As ECG signals are time series in nature, RNNs are another popular type of DL model that has been widely adopted. The transformer is a type of relatively new DL model with the emergence of the attention mechanism and has been used in some recent works. In addition, it clearly shows that more studies leverage hybrid DL models for the arrhythmia classification. Specifically, the CNNs often served as feature extractors right after the input layer of the hybrid model [28]. Other DL structures, such as RNNs and transformers, are exploited to further extract refined features. Their results show that in most cases, the hybrid model could achieve better classification performance but induce higher computational complexity [26,84,127–130]. However, as most selected studies consider traditional DL models such as CNNs and RNNs, the investigation into incorporating novel DL models or structures for arrhythmia classification with ECG signals is still limited. With the emergence of novel DL models such as ViT [131] and MLPMixer [132], the adaption of those novel DL models is expected to be introduced for ECG classification to pursue better performance improvement. In addition, most selected works focus on the improvement in classification performance as much as possible, while the interpretability of DL models is generally not discussed. The interpretable DL models [45] are highly desired to make the ECG classification results trusted in real clinical scenarios and could potentially further help cardiologists relate the heart abnormalities to possible hidden features of ECG signals, such as ECG phenotyping discussed in [7].

In most of the reviewed studies, the DL models are exploited under the supervised learning framework. However, how to leverage DL models in other artificial intelligence frameworks, such as active learning [133] and reinforcement learning [134], to improve the accuracy of ECG diagnosis could be one future research direction. In addition, how to systematically optimize the DL model structures, such as the size of convolutional kernels and hyperparameters, such as the minibatch size and learning rate, could be another crucial control knob for ECG classification.

The number of categories considered in the reviewed studies generally ranges from 2 to 9. As the number of categories to be classified in the DL models increases, learning the mapping for arrhythmia classification becomes more challenging [20]. In MITDB, the total number of categories is 26. However, the amount of ECG data in some categories are very limited. Hence, to achieve accurate classification performance with a higher number of categories to be classified, improving the learning capability of the DL process requires research effort, which could overcome the challenges such as the high complexity of the mapping relationship and lack of data in minority categories.

4.4. Evaluation Paradigm

Based on how the training and testing datasets are organized, the ECG classification can be categorized into two paradigms, i.e., inter- and intra-patient paradigms [109]. Most of the selected studies consider the intra-patient paradigm, while more recent works consider the inter-patient paradigm as it is highly desirable in clinical applications [86]. However, under the inter-patient paradigm, we found that the values of Acc and Spe are more than 10% higher than those of F_1 score, Sen, and Ppv. In addition, we also compare the classification performance of the same models in the inter- and intra-patient paradigms from the existing works which consider both of the paradigms at the same time. It shows that the F_1 score, Sen, and Ppv achieved in the inter-patient paradigm are about 15% lower than the performance metrics achieved in the intra-patient paradigm. Hence, further research effort is required to fill the performance gaps between the inter- and intra-patient paradigms, which brings research opportunities.

4.5. Performance Metrics

A wide variety of performance metrics are used for comparison. The most common metrics are overall Acc, Sen, Ppv, FPR, and F_1 score. In general, the classification accuracy of the proposed DL models in most of the selected studies is already above 95%, while other metrics such as Sen, Ppv, FPR and F_1 score were relatively lower in the range of 80–95%, which also calls for the research effort.

5. Conclusions

DL techniques have been extensively investigated for arrhythmia diagnosis with ECG signals, which exhibit the great potential to be implemented in clinical applications. However, this survey shows that some essential aspects of the DL pipeline require further research efforts before reliably applying it in clinical ECG arrhythmia classification. Specifically, leveraging diverse ECG databases for training and testing, design of advanced denoising and data augmentation techniques, developing novel integrated DL models, and deeper investigation in the inter-patient paradigm could be future research directions and opportunities to ensure trusted DL-based arrhythmia classification and promoting its application in real clinical scenarios.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/app13084964/s1, Table S1: Extraction information from selected papers.

Author Contributions: Conceptualization, P.Y.L.; methodology, Q.X., S.A.M. and K.L.; software, Q.X., I.I. and A.L.b.M.P.; validation, Q.Z. and P.Y.L.; formal analysis, Q.X. and K.L.; investigation, I.I., S.A.M. and A.L.b.M.P.; writing—original draft preparation, Q.X. and P.Y.L.; writing—review and editing, K.L. and P.Y.L.; visualization, Q.X. and Q.Z.; supervision, P.Y.L., K.L., S.A.M., I.I. and A.L.b.M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to acknowledge the Universiti Putra Malaysia for all the assistances and funding support for publication fees. Futhermore, we would like to to thank to Yue Li from University of South China, School of Computer School, Hengyang, China for his contribution on this review article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ebrahimi, Z.; Loni, M.; Daneshtalab, M.; Gharehbaghi, A. A review on deep learning methods for ECG arrhythmia classification. *Expert Syst. Appl. X* **2020**, *7*, 100033. [CrossRef]
2. Hong, S.; Zhou, Y.; Shang, J.; Xiao, C.; Sun, J. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Comput. Biol. Med.* **2020**, *122*, 103801. [CrossRef] [PubMed]
3. Xia, Y.; Wulan, N.; Wang, K.; Zhang, H. Detecting atrial fibrillation by deep convolutional neural networks. *Comput. Biol. Med.* **2018**, *93*, 84–92. [CrossRef] [PubMed]
4. Xu, S.S.; Mak, M.-W.; Cheung, C.-C. Towards End-to-End ECG Classification With Raw Signal Extraction and Deep Neural Networks. *IEEE J. Biomed. Heal Inform.* **2018**, *23*, 1574–1584. [CrossRef] [PubMed]
5. Liu, X.; Wang, H.; Li, Z.; Qin, L. Deep learning in ECG diagnosis: A review. *Knowl.-Based Syst.* **2021**, *227*, 107187. [CrossRef]
6. Egger, J.; Gsaxner, C.; Pepe, A.; Pomykala, K.L.; Jonske, F.; Kurz, M.; Li, J.; Kleesiek, J. Medical deep learning—A systematic meta-review. *Comput. Methods Programs Biomed.* **2022**, *221*, 106874. [CrossRef]
7. Siontis, K.C.; Noseworthy, P.A.; Attia, Z.I.; Friedman, P.A. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat. Rev. Cardiol.* **2021**, *18*, 465–478. [CrossRef] [PubMed]
8. Roy, Y.; Banville, H.; Albuquerque, I.; Gramfort, A.; Falk, T.H.; Faubert, J. Deep learning-based electroencephalography analysis: A systematic review. *J. Neural Eng.* **2019**, *16*, 051001. [CrossRef]
9. Moody, G.B.; Mark, R.G. The impact of the MIT-BIH Arrhythmia Database. *IEEE Eng. Med. Biol. Mag.* **2001**, *20*, 45–50. [CrossRef] [PubMed]
10. Mark, R.G.; Moody, G.B. A new method for detecting atrial fibrillation using R-R intervals. *Comput. Cardiol.* **1983**, *10*, 227–230.
11. Bousseljot, R.; Kreiseler, D.; Schnabel, A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. *Biomed. Eng./Biomed. Tech.* **1995**, *40*, 317–318. [CrossRef]
12. Wagner, P.; Strodthoff, N.; Bousseljot, R.-D.; Kreiseler, D.; Lunze, F.I.; Samek, W.; Schaeffter, T. PTB-XL, a large publicly available electrocardiography dataset. *Sci. Data* **2020**, *7*, 154. [CrossRef] [PubMed]
13. Berkaya, S.K.; Uysal, A.K.; Gunal, E.S.; Ergin, S.; Gunal, S.; Gulmezoglu, M.B. A survey on ECG analysis. *Biomed. Signal Process. Control* **2018**, *43*, 216–235. [CrossRef]
14. Du, C.; Liu, P.X.; Zheng, M. Classification of Imbalanced Electrocardiosignal Data using Convolutional Neural Network. *Comput. Methods Programs Biomed.* **2022**, *214*, 106483. [CrossRef] [PubMed]
15. Wu, Z.; Feng, X.; Yang, C. A Deep Learning Method to Detect Atrial Fibrillation Based on Continuous Wavelet Transform. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019. [CrossRef]
16. Nurmaini, S.; Darmawahyuni, A.; Mukti, A.N.S.; Rachmatullah, M.N.; Firdaus, F.; Tutuko, B. Deep Learning-Based Stacked Denoising and Autoencoder for ECG Heartbeat Classification. *Electronics* **2020**, *9*, 135. [CrossRef]
17. Pestana, J.; Belo, D.; Gamboa, H. Detection of Abnormalities in Electrocardiogram (ECG) using Deep Learning. In Proceedings of the BIOSIGNALS 2020—13th International Conference on Bio-Inspired Systems and Signal Processing, Part of 13th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC. Valletta, Malta, 24–26 February 2020; pp. 236–243. [CrossRef]
18. Liu, S.; Wang, A.; Deng, X.; Yang, C. MGNN: A multiscale grouped convolutional neural network for efficient atrial fibrillation detection. *Comput. Biol. Med.* **2022**, *148*, 105863. [CrossRef]
19. Liu, Z.; Zhou, B.; Jiang, Z.; Chen, X.; Li, Y.; Tang, M.; Miao, F. Multiclass Arrhythmia Detection and Classification From Photoplethysmography Signals Using a Deep Convolutional Neural Network. *J. Am. Hear Assoc.* **2022**, *11*, 023555. [CrossRef]
20. Zhao, Z.; Murphy, D.; Gifford, H.; Williams, S.; Darlington, A.; Relton, S.D.; Fang, H.; Wong, D.C. Analysis of an adaptive lead weighted ResNet for multiclass classification of 12-lead ECGs. *Physiol. Meas.* **2022**, *43*, 034001. [CrossRef]
21. Mathunjwa, B.M.; Lin, Y.-T.; Lin, C.-H.; Abbod, M.F.; Shieh, J.-S. ECG arrhythmia classification by using a recurrence plot and convolutional neural network. *Biomed. Signal Process. Control* **2020**, *64*, 102262. [CrossRef]
22. Tao, Y.; Li, Z.; Gu, C.; Jiang, B.; Zhang, Y. ECG-based expert-knowledge attention network to tachyarrhythmia recognition. *Biomed. Signal Process. Control* **2022**, *76*, 103649. [CrossRef]
23. Oh, S.; Lee, M. A Shallow Domain Knowledge Injection (SDK-Injection) Method for Improving CNN-Based ECG Pattern Classification. *Appl. Sci.* **2022**, *12*, 1307. [CrossRef]
24. Lee, B.T.; Kong, S.T.; Song, Y.; Lee, Y. Self-Supervised Learning with Electrocardiogram Delineation for Arrhythmia Detection. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, Guadalajara, Mexico, 1–5 November 2021; pp. 591–594. [CrossRef]

25. Nurmaini, S.; Tondas, A.E.; Darmawahyuni, A.; Rachmatullah, M.N.; Partan, R.U.; Firdaus, F.; Tutuko, B.; Pratiwi, F.; Juliano, A.H.; Khoirani, R. Robust detection of atrial fibrillation from short-term electrocardiogram using convolutional neural networks. *Futur. Gener. Comput. Syst.* **2020**, *113*, 304–317. [CrossRef]
26. Mathunjwa, B.M.; Lin, Y.-T.; Lin, C.-H.; Abbod, M.F.; Sadrawi, M.; Shieh, J.-S. ECG Recurrence Plot-Based Arrhythmia Classification Using Two-Dimensional Deep Residual CNN Features. *Sensors* **2022**, *22*, 1660. [CrossRef] [PubMed]
27. Niu, L.; Chen, C.; Liu, H.; Zhou, S.; Shu, M. A Deep-Learning Approach to ECG Classification Based on Adversarial Domain Adaptation. *Healthcare* **2020**, *8*, 437. [CrossRef] [PubMed]
28. Zhu, J.; Lv, J.; Kong, D. CNN-FWS: A Model for the Diagnosis of Normal and Abnormal ECG with Feature Adaptive. *Entropy* **2022**, *24*, 471. [CrossRef]
29. Li, X.; Zhang, F.; Sun, Z.; Li, D.; Kong, X.; Zhang, Y. Automatic heartbeat classification using S-shaped reconstruction and a squeeze-and-excitation residual network. *Comput. Biol. Med.* **2021**, *140*, 105108. [CrossRef]
30. Wang, J. A deep learning approach for atrial fibrillation signals classification based on convolutional and modified Elman neural network. *Futur. Gener. Comput. Syst.* **2019**, *102*, 670–679. [CrossRef]
31. Luo, K.; Li, J.; Wang, Z.; Cuschieri, A. Patient-Specific Deep Architectural Model for ECG Classification. *J. Healthc. Eng.* **2017**, *2017*, 4108720. [CrossRef]
32. Alkhodari, M.; Apostolidis, G.; Zisou, C.; Hadjileontiadis, L.J.; Khandoker, A.H. Swarm Decomposition Enhances the Discrimination of Cardiac Arrhythmias in Varied-Lead ECG Using ResNet-BiLSTM Network Activations. In Proceedings of the 2021 Computing in Cardiology (CinC), Brno, Czech Republic, 13–15 September 2021. [CrossRef]
33. Lui, H.W.; Chow, K.L. Multiclass classification of myocardial infarction with convolutional and recurrent neural networks for portable ECG devices. *Inform. Med. Unlocked* **2018**, *13*, 26–33. [CrossRef]
34. Eltrass, A.S.; Tayel, M.B.; Ammar, A.I. Automated ECG multi-class classification system based on combining deep learning features with HRV and ECG measures. *Neural. Comput. Appl.* **2022**, *34*, 8755–8775. [CrossRef]
35. Eltrass, A.S.; Tayel, M.B.; Ammar, A.I. A new automated CNN deep learning approach for identification of ECG congestive heart failure and arrhythmia using constant-Q non-stationary Gabor transform. *Biomed. Signal Process. Control* **2020**, *65*, 102326. [CrossRef]
36. Zhang, G.; Si, Y.; Yang, W.; Wang, D. A Robust Multilevel DWT Densely Network for Cardiovascular Disease Classification. *Sensors* **2020**, *20*, 4777. [CrossRef] [PubMed]
37. Degirmenci, M.; Ozdemir, M.; Izci, E.; Akan, A. Arrhythmic Heartbeat Classification Using 2D Convolutional Neural Networks. *IRBM* **2021**, *43*, 422–433. [CrossRef]
38. Jin, Y.; Liu, J.; Liu, Y.; Qin, C.; Li, Z.; Xiao, D.; Zhao, L.; Liu, C. A Novel Interpretable Method Based on Dual-Level Attentional Deep Neural Network for Actual Multilabel Arrhythmia Detection. *IEEE Trans. Instrum. Meas.* **2021**, *71*, 2500311. [CrossRef]
39. Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; Yang, Y. Random Erasing Data Augmentation. In Proceedings of the AAAI 2020—34th AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 13001–13008. [CrossRef]
40. Zhang, D.; Yang, S.; Yuan, X.; Zhang, P. Interpretable deep learning for automatic diagnosis of 12-lead electrocardiogram. *iScience* **2021**, *24*, 102373. [CrossRef] [PubMed]
41. Prabhakararao, E.; Dandapat, S. Multi-Scale Convolutional Neural Network Ensemble for Multi-Class Arrhythmia Classification. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 3802–3812. [CrossRef]
42. He, J.; Rong, J.; Sun, L.; Wang, H.; Zhang, Y.; Ma, J. A framework for cardiac arrhythmia detection from IoT-based ECGs. *World Wide Web* **2020**, *23*, 2835–2850. [CrossRef]
43. Nankani, D.; Baruah, R.D. Ventricular Arrhythmia Classification and Interpretation Using Residual Neural Network with Guided Backpropagation. In Proceedings of the IEEE Region 10 Annual International Conference, Proceedings/TENCON, Auckland, New Zealand, 7–10 December 2021; pp. 574–579. [CrossRef]
44. Li, X.; Zhang, J.; Chen, W.; Chen, Y.; Zhang, C.; Xiang, W.; Li, D. Inter-patient automated arrhythmia classification: A new approach of weight capsule and sequence to sequence combination. *Comput. Methods Programs Biomed.* **2021**, *214*, 106533. [CrossRef]
45. Singh, P.; Sharma, A. Interpretation and Classification of Arrhythmia using Deep Convolutional Network. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 2518512. [CrossRef]
46. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image Style Transfer Using Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423.
47. Shaker, A.M.; Tantawi, M.; Shedeed, H.A.; Tolba, M.F. Generalization of Convolutional Neural Networks for ECG Classification Using Generative Adversarial Networks. *IEEE Access* **2020**, *8*, 35592–35605. [CrossRef]
48. Ganguly, B.; Ghosal, A.; Das, A.; Das, D.; Chatterjee, D.; Rakshit, D. Automated Detection and Classification of Arrhythmia From ECG Signals Using Feature-Induced Long Short-Term Memory Network. *IEEE Sens. Lett.* **2020**, *4*, 6001604. [CrossRef]
49. Pandey, S.K.; Janghel, R.R. Automatic detection of arrhythmia from imbalanced ECG database using CNN model with SMOTE. *Australas. Phys. Eng. Sci. Med.* **2019**, *42*, 1129–1139. [CrossRef] [PubMed]
50. Kiranyaz, S.; Ince, T.; Gabbouj, M. Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. *IEEE Trans. Biomed. Eng.* **2015**, *63*, 664–675. [CrossRef] [PubMed]
51. Tesfai, H.; Saleh, H.; Al-Qutayri, M.; Mohammad, M.B.; Tekeste, T.; Khandoker, A.; Mohammad, B. Lightweight Shufflenet Based CNN for Arrhythmia Classification. *IEEE Access* **2022**, *10*, 111842–111854. [CrossRef]

52. Mostayed, A.; Luo, J.; Shu, X.; Wee, W. Classification of 12-Lead ECG Signals with Bi-Directional LSTM Network. 2018. Available online: http://arxiv.org/abs/1811.02090 (accessed on 26 December 2022).
53. Mashrur, F.R.; Roy, A.D.; Saha, D.K. Automatic Identification of Arrhythmia from ECG Using AlexNet Convolutional Neural Network. In Proceedings of the 2019 4th International Conference on Electrical Information and Communication Technology, EICT 2019, Khulna, Bangladesh, 20–22 December 2019. [CrossRef]
54. de Santana, J.R.G.; Costa, M.G.F.; Filho, C.F.F.C. A New Approach to Classify Cardiac Arrythmias Using 2D Convolutional Neural Networks. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, Auckland, New Zealand, 7–10 December 2021; pp. 566–570. [CrossRef]
55. Liu, W.; Zhang, M.; Zhang, Y.; Liao, Y.; Huang, Q.; Chang, S.; Wang, H.; He, J. Real-Time Multilead Convolutional Neural Network for Myocardial Infarction Detection. *IEEE J. Biomed. Heal Inform.* **2017**, *22*, 1434–1444. [CrossRef]
56. Singh, S.; Pandey, S.K.; Pawar, U.; Janghel, R.R. Classification of ECG Arrhythmia using Recurrent Neural Networks. *Procedia Comput. Sci.* **2018**, *132*, 1290–1297. [CrossRef]
57. Khan, M.A.; Kim, Y. Cardiac Arrhythmia Disease Classification Using LSTM Deep Learning Approach. *Comput. Mater. Contin.* **2021**, *67*, 427–443. [CrossRef]
58. Zhou, X.; Zhu, X.; Nakamura, K.; Mahito, N. Premature Ventricular Contraction Detection from Ambulatory ECG Using Recurrent Neural Networks. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018. [CrossRef]
59. Xu, W.; Wang, L.; Wang, B.; Cheng, W. Intelligent Recognition Algorithm of Multiple Myocardial Infarction Based on Morphological Feature Extraction. *Processes* **2022**, *10*, 2348. [CrossRef]
60. Rahul, J.; Sharma, L.D. Artificial intelligence-based approach for atrial fibrillation detection using normalised and short-duration time-frequency ECG. *Biomed. Signal Process. Control* **2021**, *71*, 103270. [CrossRef]
61. Faust, O.; Shenfield, A.; Kareem, M.; San, T.R.; Fujita, H.; Acharya, U.R. Automated detection of atrial fibrillation using long short-term memory network with RR interval signals. *Comput. Biol. Med.* **2018**, *102*, 327–335. [CrossRef]
62. Jiang, M.; Gu, J.; Li, Y.; Wei, B.; Zhang, J.; Wang, Z.; Xia, L. HADLN: Hybrid Attention-Based Deep Learning Network for Automated Arrhythmia Classification. *Front. Physiol.* **2021**, *12*, 683025. [CrossRef]
63. 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. [CrossRef] [PubMed]
64. He, K.; Gan, C.; Li, Z.; Rekik, I.; Yin, Z.; Ji, W.; Gao, Y.; Wang, Q.; Zhang, J.; Shen, D. Transformers in medical image analysis. *Intell. Med.* **2023**, *3*, 59–78. [CrossRef]
65. Yan, G.; Liang, S.; Zhang, Y.; Liu, F. Fusing Transformer Model with Temporal Features for ECG Heartbeat Classification. In Proceedings of the—2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019, San Diego, CA, USA, 18–21 November 2019; pp. 898–905. [CrossRef]
66. Natarajan, A.; Chang, Y.; Mariani, S.; Rahman, A.; Boverman, G.; Vij, S.; Rubin, J. A Wide and Deep Transformer Neural Network for 12-Lead ECG Classification. *Comput. Cardiol.* **2020**, *2020*, 20350657. [CrossRef]
67. Natarajan, A.; Boverman, G.; Chang, Y.; Antonescu, C.; Rubin, J. Convolution-Free Waveform Transformers for Multi-Lead ECG Classification. In Proceedings of the 2021 Computing in Cardiology (CinC), Brno, Czech Republic, 13–15 September 2021. [CrossRef]
68. Mousavi, S.; Afghah, F. Inter-and intra-patient ecg heartbeat classification for arrhythmia detection: A sequence to sequence deep learning approach. In Proceedings of the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 1308–1312.
69. Swapna, G.; Soman, K.P.; Vinayakumar, R. Automated detection of cardiac arrhythmia using deep learning techniques. *Procedia Comput. Sci.* **2018**, *132*, 1192–1201. [CrossRef]
70. Shoughi, A.; Dowlatshahi, M.B. A practical system based on CNN-BLSTM network for accurate classification of ECG heartbeats of MIT-BIH imbalanced dataset. In Proceedings of the 26th International Computer Conference, Computer Society of Iran, CSICC 2021, Tehran, Iran, 3–4 March 2021; pp. 1–6. [CrossRef]
71. Xiong, Z.; Stiles, M.K.; Gillis, A.M.; Zhao, J. Enhancing the detection of atrial fibrillation from wearable sensors with neural style transfer and convolutional recurrent networks. *Comput. Biol. Med.* **2022**, *146*, 105551. [CrossRef]
72. Meng, L.; Tan, W.; Ma, J.; Wang, R.; Yin, X.; Zhang, Y. Enhancing dynamic ECG heartbeat classification with lightweight transformer model. *Artif. Intell. Med.* **2022**, *124*, 102236. [CrossRef]
73. Yoo, J.; Jun, T.J.; Kim, Y.-H. xECGNet: Fine-tuning attention map within convolutional neural network to improve detection and explainability of concurrent cardiac arrhythmias. *Comput. Methods Programs Biomed.* **2021**, *208*, 106281. [CrossRef]
74. Li, J.; Pang, S.-P.; Xu, F.; Ji, P.; Zhou, S.; Shu, M. Two-dimensional ECG-based cardiac arrhythmia classification using DSE-ResNet. *Sci. Rep.* **2022**, *12*, 216. [CrossRef]
75. Srivastava, A.; Pratiher, S.; Alam, S.; Hari, A.; Banerjee, N.; Ghosh, N.; Patra, A. A deep residual inception network with channel attention modules for multi-label cardiac abnormality detection from reduced-lead ECG. *Physiol. Meas.* **2022**, *43*, 064005. [CrossRef]
76. Nejedly, P.; Ivora, A.; Smisek, R.; Viscor, I.; Koscova, Z.; Jurak, P.; Plesinger, F. Classification of ECG Using Ensemble of Residual CNNs with Attention Mechanism. *Comput. Cardiol.* **2021**, *48*, 9662723. [CrossRef]

77. de Chazal, P.; O'Dwyer, M.; Reilly, R. Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1196–1206. [CrossRef] [PubMed]
78. He, J.; Rong, J.; Sun, L.; Wang, H.; Zhang, Y. An Advanced Two-Step DNN-Based Framework for Arrhythmia Detection. *Lect. Notes Comput. Sci.* **2020**, *12085*, 422–434. [CrossRef]
79. Tutuko, B.; Nurmaini, S.; Tondas, A.E.; Rachmatullah, M.N.; Darmawahyuni, A.; Esafri, R.; Firdaus, F.; Sapitri, A.I. AFibNet: An implementation of atrial fibrillation detection with convolutional neural network. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 216. [CrossRef]
80. Katsushika, S.; Kodera, S.; Nakamoto, M.; Ninomiya, K.; Inoue, S.; Sawano, S.; Kakuda, N.; Takiguchi, H.; Shinohara, H.; Matsuoka, R.; et al. The Effectiveness of a Deep Learning Model to Detect Left Ventricular Systolic Dysfunction from Electrocardiograms. *Int. Hear J.* **2021**, *62*, 1332–1341. [CrossRef]
81. Andersen, R.S.; Peimankar, A.; Puthusserypady, S. A deep learning approach for real-time detection of atrial fibrillation. *Expert Syst. Appl.* **2018**, *115*, 465–473. [CrossRef]
82. Gan, Y.; Shi, J.-C.; He, W.-M.; Sun, F.-J. Parallel classification model of arrhythmia based on DenseNet-BiLSTM. *Biocybern. Biomed. Eng.* **2021**, *41*, 1548–1560. [CrossRef]
83. Sellami, A.; Hwang, H. A robust deep convolutional neural network with batch-weighted loss for heartbeat classification. *Expert Syst. Appl.* **2018**, *122*, 75–84. [CrossRef]
84. Tan, J.H.; Hagiwara, Y.; Pang, W.; Lim, I.; Oh, S.L.; Adam, M.; Tan, R.S.; Chen, M.; Acharya, U.R. Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals. *Comput. Biol. Med.* **2018**, *94*, 19–26. [CrossRef]
85. Jangra, M.; Dhull, S.K.; Singh, K.K.; Singh, A.; Cheng, X. O-WCNN: An optimized integration of spatial and spectral feature map for arrhythmia classification. *Complex Intell. Syst.* **2021**, *2021*, 1–14. [CrossRef]
86. Chen, A.; Wang, F.; Liu, W.; Chang, S.; Wang, H.; He, J.; Huang, Q. Multi-information fusion neural networks for arrhythmia automatic detection. *Comput. Methods Programs Biomed.* **2020**, *193*, 105479. [CrossRef] [PubMed]
87. Fu, L.; Lu, B.; Nie, B.; Peng, Z.; Liu, H.; Pi, X. Hybrid Network with Attention Mechanism for Detection and Location of Myocardial Infarction Based on 12-Lead Electrocardiogram Signals. *Sensors* **2020**, *20*, 1020. [CrossRef] [PubMed]
88. Zubair, M.; Yoon, C. Cost-Sensitive Learning for Anomaly Detection in Imbalanced ECG Data Using Convolutional Neural Networks. *Sensors* **2022**, *22*, 4075. [CrossRef]
89. Ran, S.; Yang, X.; Liu, M.; Zhang, Y.; Cheng, C.; Zhu, H.; Yuan, Y. Homecare-Oriented ECG Diagnosis with Large-Scale Deep Neural Network for Continuous Monitoring on Embedded Devices. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–13. [CrossRef]
90. Rizqyawan, M.I.; Siradj, Y.; Amri, M.F.; Pratondo, A. Re-implementation of Convolutional Neural Network for Arrhythmia Detection. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2022**, *12*, 1319–1326. [CrossRef]
91. Essa, E.; Xie, X. An Ensemble of Deep Learning-Based Multi-Model for ECG Heartbeats Arrhythmia Classification. *IEEE Access* **2021**, *9*, 103452–103464. [CrossRef]
92. Li, Y.; Qian, R.; Li, K. Inter-patient arrhythmia classification with improved deep residual convolutional neural network. *Comput. Methods Programs Biomed.* **2022**, *214*, 106582. [CrossRef] [PubMed]
93. Labib, M.I.; Nahid, A.-A. OptRPC: A novel and optimized recurrence plot-based system for ECG beat classification. *Biomed. Signal Process. Control* **2021**, *72*, 103328. [CrossRef]
94. Pokaprakarn, T.; Kitzmiller, R.R.; Moorman, J.R.; Lake, D.E.; Krishnamurthy, A.K.; Kosorok, M.R. Sequence to Sequence ECG Cardiac Rhythm Classification Using Convolutional Recurrent Neural Networks. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 572–580. [CrossRef]
95. Sarshar, N.T.; Mirzaei, M. Premature Ventricular Contraction Recognition Based on a Deep Learning Approach. *J. Heal Eng.* **2022**, *2022*, 1450723. [CrossRef]
96. Wang, G.; Chen, M.; Ding, Z.; Li, J.; Yang, H.; Zhang, P. Inter-patient ECG arrhythmia heartbeat classification based on unsupervised domain adaptation. *Neurocomputing* **2021**, *454*, 339–349. [CrossRef]
97. Pandey, S.K.; Janghel, R.R.; Dev, A.V.; Mishra, P.K. Automated arrhythmia detection from electrocardiogram signal using stacked restricted Boltzmann machine model. *SN Appl. Sci.* **2021**, *3*, 624. [CrossRef]
98. Qiu, X.; Liang, S.; Meng, L.; Zhang, Y.; Liu, F. Exploiting feature fusion and long-term context dependencies for simultaneous ECG heartbeat segmentation and classification. *Int. J. Data Sci. Anal.* **2021**, *11*, 181–193. [CrossRef]
99. Zhang, J.; Liu, A.; Liang, D.; Chen, X.; Gao, M. Interpatient ECG Heartbeat Classification with an Adversarial Convolutional Neural Network. *J. Health Eng.* **2021**, *2021*, 9946596. [CrossRef]
100. Cai, J.; Zhou, G.; Dong, M.; Hu, X.; Liu, G.; Ni, W. Real-Time Arrhythmia Classification Algorithm Using Time-Domain ECG Feature Based on FFNN and CNN. *Math. Probl. Eng.* **2021**, *2021*, 6648432. [CrossRef]
101. Wang, T.; Lu, C.; Sun, Y.; Yang, M.; Liu, C.; Ou, C. Automatic ECG Classification Using Continuous Wavelet Transform and Convolutional Neural Network. *Entropy* **2021**, *23*, 119. [CrossRef] [PubMed]
102. Qu, Y.; Zhang, N.; Meng, Y.; Qin, Z.; Lu, Q.; Liu, X. ECG Heartbeat Classification Detection Based on WaveNet-LSTM. In Proceedings of the 2020 IEEE 4th International Conference on Frontiers of Sensors Technologies (ICFST), Shanghai, China, 6–9 November 2020. [CrossRef]
103. Li, Y.; He, Z.; Wang, H.; Li, B.; Li, F.; Gao, Y.; Ye, X. CraftNet: A deep learning ensemble to diagnose cardiovascular diseases. *Biomed. Signal Process. Control* **2020**, *62*, 102091. [CrossRef]

104. Zhao, W.; Hu, J.; Jia, D.; Wang, H.; Li, Z.; Yan, C.; You, T. Deep Learning Based Patient-Specific Classification of Arrhythmia on ECG signal. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 1500–1503. [CrossRef]
105. Xie, Q.; Tu, S.; Wang, G.; Lian, Y.; Xu, L. Feature Enrichment Based Convolutional Neural Network for Heartbeat Classification From Electrocardiogram. *IEEE Access* **2019**, *7*, 153751–153760. [CrossRef]
106. Zhou, R.; Li, X.; Yong, B.; Shen, Z.; Wang, C.; Zhou, Q.; Cao, Y.; Li, K.-C. Arrhythmia recognition and classification through deep learning-based approach Arrhythmia recognition and classification through deep learning-based approach 507. *Arrhythmias* **2019**, *19*, 506.
107. Mathews, S.M.; Kambhamettu, C.; Barner, K.E. A novel application of deep learning for single-lead ECG classification. *Comput. Biol. Med.* **2018**, *99*, 53–62. [CrossRef]
108. Hu, J.; Zhao, W.; Jia, D.; Yan, C.; Wang, H.; Li, Z.; Fang, J.; Yang, M. Deep Multi-instance Networks for Bundle Branch Block Detection from Multi-lead ECG. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020. [CrossRef]
109. Guo, L.; Sim, G.; Matuszewski, B. Inter-patient ECG classification with convolutional and recurrent neural networks. *Biocybern. Biomed. Eng.* **2019**, *39*, 868–879. [CrossRef]
110. Rajpurkar, P.; Hannun, A.Y.; Haghpanahi, M.; Bourn, C.; Ng, A.Y. Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. 2017. Available online: http://arxiv.org/abs/1707.01836 (accessed on 6 July 2017).
111. Ribeiro, A.H.; Ribeiro, M.H.; Paixão, G.M.M.; Oliveira, D.M.; Gomes, P.R.; Canazart, J.A.; Ferreira, M.P.S.; Andersson, C.R.; Macfarlane, P.W.; Meira, W.; et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat. Commun.* **2020**, *11*, 1760. [CrossRef] [PubMed]
112. Bollepalli, S.C.; Sevakula, R.K.; Au-Yeung, W.M.; Kassab, M.B.; Merchant, F.M.; Bazoukis, G.; Boyer, R.; Isselbacher, E.M.; Armoundas, A.A. Real-Time Arrhythmia Detection Using Hybrid Convolutional Neural Networks. *J. Am. Hear Assoc.* **2021**, *10*, 023222. [CrossRef] [PubMed]
113. Ma, L.; Liang, L. A regularization method to improve adversarial robustness of neural networks for ECG signal classification. *Comput. Biol. Med.* **2022**, *144*, 105345. [CrossRef] [PubMed]
114. Bazargani, M.H.Z.; Pakrashi, A.; Mac Namee, B. The Deep Radial Basis Function Data Descriptor (D-RBFDD) Network: A One-Class Neural Network for Anomaly Detection. *IEEE Access* **2022**, *10*, 70645–70661. [CrossRef]
115. Huang, Y.; Yen, G.G.; Tseng, V.S. Snippet Policy Network V2: Knee-Guided Neuroevolution for Multi-Lead ECG Early Classification. *IEEE Trans. Neural. Netw. Learn. Syst.* **2022**, *99*, 3187741. [CrossRef] [PubMed]
116. Irfan, S.; Anjum, N.; Althobaiti, T.; Alotaibi, A.A.; Siddiqui, A.B.; Ramzan, N. Heartbeat Classification and Arrhythmia Detection Using a Multi-Model Deep-Learning Technique. *Sensors* **2022**, *22*, 5606. [CrossRef] [PubMed]
117. Murat, F.; Yildirim, O.; Talo, M.; Demir, Y.; Tan, R.-S.; Ciaccio, E.J.; Acharya, U.R. Exploring deep features and ECG attributes to detect cardiac rhythm classes. *Knowl.-Based Syst.* **2021**, *232*, 107473. [CrossRef]
118. Radhakrishnan, T.; Karhade, J.; Ghosh, S.; Muduli, P.; Tripathy, R.; Acharya, U.R. AFCNNet: Automated detection of AF using chirplet transform and deep convolutional bidirectional long short term memory network with ECG signals. *Comput. Biol. Med.* **2021**, *137*, 104783. [CrossRef]
119. Liang, Y.; Yin, S.; Tang, Q.; Zheng, Z.; Elgendi, M.; Chen, Z. Deep Learning Algorithm Classifies Heartbeat Events Based on Electrocardiogram Signals. *Front. Physiol.* **2020**, *11*, 569050. [CrossRef]
120. Chen, T.-M.; Huang, C.-H.; Shih, E.S.; Hu, Y.-F.; Hwang, M.-J. Detection and Classification of Cardiac Arrhythmias by a Challenge-Best Deep Learning Neural Network Model. *iScience* **2020**, *23*, 100886. [CrossRef] [PubMed]
121. Hannun, A.Y.; Rajpurkar, P.; Haghpanahi, M.; Tison, G.H.; Bourn, C.; Turakhia, M.P.; Ng, A.Y. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat. Med.* **2019**, *25*, 65–69. [CrossRef] [PubMed]
122. Plesinger, F.; Nejedly, P.; Viscor, I.; Halamek, J.; Jurak, P. Parallel use of a convolutional neural network and bagged tree ensemble for the classification of Holter ECG. *Physiol. Meas.* **2018**, *39*, 094002. [CrossRef] [PubMed]
123. Meng, Y.; Liang, G.; Yue, M. Deep Learning-Based Arrhythmia Detection in Electrocardiograph. *Sci. Program.* **2021**, *2021*, 9926769. [CrossRef]
124. 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. [CrossRef] [PubMed]
125. Mishra, A.; Dharahas, G.; Gite, S.; Kotecha, K.; Koundal, D.; Zaguia, A.; Kaur, M.; Lee, H.-N. ECG Data Analysis with Denoising Approach and Customized CNNs. *Sensors* **2022**, *22*, 1928. [CrossRef] [PubMed]
126. Rath, A.; Mishra, D.; Panda, G.; Satapathy, S.C. Heart disease detection using deep learning methods from imbalanced ECG samples. *Biomed. Signal Process. Control* **2021**, *68*, 102820. [CrossRef]
127. Petmezas, G.; Haris, K.; Stefanopoulos, L.; Kilintzis, V.; Tzavelis, A.; Rogers, J.A.; Katsaggelos, A.K.; Maglaveras, N. Automated Atrial Fibrillation Detection using a Hybrid CNN-LSTM Network on Imbalanced ECG Datasets. *Biomed. Signal Process. Control* **2020**, *63*, 102194. [CrossRef]
128. Obeidat, Y.; Alqudah, A.M. A Hybrid Lightweight 1D CNN-LSTM Architecture for Automated ECG Beat-Wise Classification. *Traitement Signal* **2021**, *38*, 1281–1291. [CrossRef]

129. Chen, C.; Hua, Z.; Zhang, R.; Liu, G.; Wen, W. Automated arrhythmia classification based on a combination network of CNN and LSTM. *Biomed. Signal Process. Control* **2019**, *57*, 101819. [CrossRef]
130. Oh, S.L.; Ng, E.Y.; San Tan, R.; Acharya, U.R. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Comput. Biol. Med.* **2018**, *102*, 278–287. [CrossRef] [PubMed]
131. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. A Survey on Vision Transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 87–110. [CrossRef] [PubMed]
132. Tolstikhin, I.; Houlsby, N.; Kolesnikov, A.; Beyer, L.; Zhai, X.; Unterthiner, T.; Yung, J.; Steiner, A.; Keysers, D.; Uszkoreit, J.; et al. MLP-Mixer: An all-MLP Architecture for Vision. *Adv. Neural. Inf. Process. Syst.* **2021**, *34*, 24261–24272.
133. He, Z.; Yuan, S.; Zhao, J.; Du, B.; Yuan, Z.; Alhudhaif, A.; Alenezi, F.; Althubiti, S.A. A novel myocardial infarction localization method using multi-branch DenseNet and spatial matching-based active semi-supervised learning. *Inf. Sci.* **2022**, *606*, 649–668. [CrossRef]
134. Khajuria, R.; Sarwar, A. Reinforcement Learning in Medical Diagnosis: An Overview. *Lect. Notes Electr. Eng.* **2022**, *832*, 179–188. [CrossRef]