



# eFuseNet: A deep ensemble fusion network for efficient detection of Arrhythmia and Myocardial Infarction using ECG signals

Amitesh Kumar Dwivedi<sup>1</sup> · Gaurav Srivastava<sup>2</sup> · Sakshi Tripathi<sup>3</sup> · Nitesh Pradhan<sup>4</sup>

Received: 20 April 2023 / Revised: 10 May 2024 / Accepted: 20 June 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

## Abstract

Myocardial infarction (MI) and cardiac arrhythmias remain leading causes of mortality within cardiovascular diseases. Accurate and timely diagnosis using electrocardiogram (ECG) signals is crucial, yet manual ECG interpretation is prone to error and time-consuming. Addressing these challenges, the authors propose a novel deep neural network architecture for efficient and accurate detection and classification of both MI and arrhythmia from ECG signals. Our model strategically fuses a modified AlexNet for robust feature extraction with Long Short-Term Memory and Gated Recurrent Units networks to exploit the temporal dependencies inherent in ECG data. Fusion techniques combine the insights from these diverse sub-models, aiming to reduce generalization errors and surpass the performance of single-model approaches. Extensive experimentation on the MIT-BIH and PTB datasets demonstrates superior accuracy (98.51% and 99.97%, respectively) and reduced training time compared to other pre-trained networks. Robustness is validated through 10-fold cross-validation, yielding average accuracies of 98.76% and 99.48%. Our findings highlight the potential of this computationally efficient ensemble model for practical clinical implementation, enabling more accurate and timely diagnosis of life-threatening cardiovascular conditions.

**Keywords** Arrhythmia · Myocardial infarction · ECG signals · Ensemble learning · Feature fusion · Long short term memory · Gated recurrent units

---

✉ Nitesh Pradhan  
nitesh.pradhan@lnmiit.ac.in

Gaurav Srivastava  
mailto:gaurav2001@gmail.com

<sup>1</sup> Johns Hopkins Whiting School of Engineering, Johns Hopkins University, 21218 Baltimore, MD, United States

<sup>2</sup> Dell International Services India Pvt. Ltd., Dell Technologies, 500081 Hyderabad, Telangana, India

<sup>3</sup> Department of Computer Science and Engineering, Manipal University Jaipur, 303007 Jaipur, Rajasthan, India

<sup>4</sup> Department of Computer Science and Engineering, The LNM Institute of Information Technology, Jaipur, 302031 Jaipur, Rajasthan, India

# 1 Introduction

Cardiovascular disease (CVD) encompasses various conditions affecting the heart and blood vessels [1]. It is a leading cause of global mortality, responsible for approximately 17.9 million deaths each year [2]. Myocardial infarction (MI), also known as a heart attack, and arrhythmia are among the most common forms of CVD. MI occurs due to blockage in the heart's arteries, cutting off blood flow to the heart muscle [3]. This can cause rapid tissue damage and, if not addressed quickly, may be fatal. Notably, MI constitutes a significant portion of CVD cases, approximately 85% [4]. Furthermore, MI can lead to cardiac arrhythmia, an irregular heartbeat potentially resulting in major complications like stroke or heart failure [5]. Early diagnosis and treatment are essential for managing CVDs and minimizing their risks.

Arrhythmia is a medical condition where the heart's normal rhythm is disrupted. This disruption stems from irregular electrical impulses, causing the heart to beat too slowly, too quickly, or erratically. If severe, arrhythmia can hinder the heart's ability to pump blood effectively, potentially leading to organ damage or failure [6]. As a serious health concern, arrhythmia requires prompt medical attention for diagnosis, management, and prevention of complications. While the condition's prevalence and severity vary, effective treatments exist. Continued research is crucial to deepen our understanding of arrhythmia's causes, risk factors, and optimal treatment approaches.

Electrocardiograms (ECGs) are essential for diagnosing cardiac arrhythmia and myocardial infarction. However, accurate ECG interpretation demands expertise, and manual analysis is time-consuming. Practitioners may require additional tests (e.g., stress tests, tilt-table tests, electrophysiological tests) for uncertain cases, further increasing time and labor demands on medical institutions. Since the risk of a second MI increases after the first, there's a clear need for automated ECG diagnostic tools that reduce the time and labor burden [7]. Despite the potential benefits of automated ECG interpretation, challenges remain. Accuracy can be hindered by the variability of ECG waveforms across diverse patient populations. Successful clinical integration also requires user-friendly software interfaces for medical practitioners. While automated ECG diagnosis could significantly streamline the diagnosis of cardiac abnormalities like arrhythmia and myocardial infarction, and AI/machine learning algorithms show promise, further research is needed to overcome existing challenges and enable widespread clinical adoption.

Researchers have explored machine learning techniques for automated cardiac arrhythmia and myocardial infarction detection using ECGs [8–10]. However, ECG signals vary between and within individuals, and effective diagnosis may require extracting different features depending on the specific cardiovascular disorder. Due to the heterogeneity of these conditions, it may even be necessary to extract distinct features for the same diagnosis. Arrhythmia detection is a crucial focus for biomedical researchers, given the condition's prevalence and economic burden (affecting millions in the US alone, with costs up to \$67.4 billion annually [11]). Traditional machine learning algorithms like modified Support Vector Machines (SVM) and Random Forest (RF) have been used for automated ECG diagnosis. However, these algorithms have limitations: SVM scales poorly with large datasets [12], and RF is susceptible to overfitting with noisy data [13]. Deep learning approaches have emerged as a more powerful and accurate alternative for automated ECG diagnosis [14, 15].

Deep learning models excel in learning ECG signal characteristics through weight and bias optimization via backpropagation. This enables them to isolate individual ECG wave embeddings, enhancing classification accuracy. Their potential for improving the speed and precision of automated ECG diagnosis is crucial for effective cardiac arrhythmia

management. Convolutional Neural Networks (CNNs) are popular for automated feature extraction in ECG diagnosis [16]. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), are also frequently used due to the temporal dependencies within ECG signals [17–19]. While complex pre-trained models (e.g., VGG-based [20], deep belief networks [21], dual attention mechanisms [22]) can boost accuracy, their computational cost and training time make them less suitable for clinical settings [23].

Computationally efficient methods are essential in biomedical engineering, especially where resources may be constrained [24]. To address this, we propose the eFuseNet model, built on the traditional AlexNet architecture. eFuseNet uses fewer parameters than other pre-trained networks for efficient detection of cardiac arrhythmia and myocardial infarction. Our goal is to provide a cost-effective and automated diagnostic aid to medical professionals. We evaluated eFuseNet on the MIT-BIH and PTB databases, employing 10-fold cross-validation for robustness. We used an Ensemble Fusion Technique (EFT) to mitigate bias due to dataset imbalance (a common issue in MIT-BIH and PTB). EFT also helps us achieve superior accuracy and classification metrics. Our study's key contributions include:

- The authors have optimized the traditional architectures such as AlexNet, Xception, and ResNet50 specifically for 1-dimensional ECG signals, significantly reducing computational overhead compared to larger pre-trained models while preserving classification accuracy.
- The authors designed an ensemble fusion model uniquely combining adapted AlexNet, LSTM, and GRU. This approach effectively extracts deep features and models the temporal dependencies inherent in ECG data.
- The ensemble model - eFuseNet generates a comprehensive feature set, integrating the strengths of CNNs and RNNs. This richer representation has the potential to boost classification performance in diagnosing arrhythmia and myocardial infarction.

The rest of the paper is divided as the following: Section 2 discusses the related works performed, Section 3 talks about materials and methods used in the experimentation, Section 4 discusses the proposed method and Section 5 discusses the experimental results.

## 2 Related works

ECG signals are time-based, and Arrhythmia and Myocardial Infarction are time-series-based classification problems. Concerning the time-series-based classification of both Myocardial Infarction and Arrhythmia, we have reviewed literature that uses deep learning and machine learning techniques to detect and classify Cardiac Arrhythmia and Myocardial Infarction. Therefore, this section has explored methodologies that encompass both Deep Learning and Machine Learning techniques.

In a study presented by Pal et al., CardioNet is proposed. CardioNet uses both the MIT-BIH Arrhythmia dataset, PTB dataset, and a pre-trained DenseNet on ImageNet and adds Transfer Learning to the model. A ratio of 80%:20% for training and testing was maintained during experimentation. Additionally, the results of VGG, ResNet, ResNetV2, and Densenet, were compared before selecting DenseNet as the focused model for this experimentation. Furthermore, the weights obtained during DenseNet's training on ImageNet are fine-tuned to give optimal results in 10 epochs. CardioNet gives an accuracy of 98.92%, which proved to be better than its competitive studies [25]. Hammad et al. propose a Deep Neural Network (DNN) technique for digitalized classification and detection Arrhythmia. In their proposed

DNN technique, first comes a learning stage where the accuracy is improved through effective feature extraction methods. After the learning stage, a Genetic Algorithm is used to aggregate and combine the best results from feature extraction and classification. A Support Vector Machine (SVM) classifier is used to classify the various types of Arrhythmia during the experimentation. MIT-BIH Arrhythmia dataset is used during this experimentation, and an average accuracy of 94% F1 score of 0.953 is obtained after the result [26].

Additionally, Zhang et al. pointed out the scope of improving classification results of Arrhythmia since all the 12 leads in the ECG signal detection apparatus have varied contributions to the detection of cardiac Arrhythmia. Thus, a Spatio-temporal Attention-based Convolutional Recurring Neural Network (STA-CRNN) is presented by Zhang et al., which gives close focus to the features along the temporal and spatial axis. The STA-CRNN consists of a convolutional neural network subnetwork, Spatio-Temporal Attention Modules, and an RNN subnetwork. The model mentioned above in the study by Zhang et al. shows remarkable classification improvements achieving an average F1 score of 0.835 and proving to be a promising model for automating Cardiac Arrhythmia Detection [27]. Hu et al. found the ignorance of inter-heartbeat dependencies to detect and classify Cardiac Arrhythmia. Thus, to overcome the issue, a novel transformer-based Deep Neural Network: ECG DETR is proposed to perform cardiac Arrhythmia detection and classification on continuous ECG wave signals. The study uses MIT-BIH Arrhythmia and MIT-BIH Atrial Fibrillation datasets used in the scope of experimentation. Hu et al. experimentation yielded an average accuracy of 99.12% on the 8-way classification of Cardiac Arrhythmia, which proved to be better than other comparative studies [28].

Furthermore, Natarajan et al. presented a unique arrhythmia classification technique that amalgamates 22 static hand-crafted features, automatically extracted by a neural network comprising a CNN and a transformer [29]. Stradthoof et al. trained an ensemble of multiple neural networks on the PTB dataset and obtained a sensitivity of 93.3% and a specificity of 89.7%. The results were evaluated with a 10-fold cross-validation based on real-time patient-based sampling. The sampling used to cross-validate the results took input from the standardized 12-lead input. During the experimentation, recurring and convolutional neural networks are trained and observed before deciding on the final networks to be ensembled together [30]. Wu et al. suggested a novel deep feature model based on the classification and detection of Myocardial Infarction [31]. The proposed model's approach is to learn an illustration of the extracted feature that makes the classification process more efficient. Furthermore, multiscale wavelet transformation is also used in feature learning to facilitate the extraction of Myocardial Infarction features. A softmax regressor is used for multi-class classification, and a PTB dataset is used for experimentation. This method proved to be better than its comparative studies. Wang et al. proposed MENN: a multi-lead-based ensemble neural network to classify and differentiate Anterior Myocardial Infarction, Inferior Myocardial Infarction from Healthy Control. MENN is a combination of three sub-networks. This ensembling method provides better classification accuracy when compared to other baseline methods mentioned in the experimentation manuscript. The PTB dataset is used for the scope of this research. Furthermore, the results have been cross-validated five times via cross-fold validation [32].

The application of deep neural networks (DNNs) extends beyond ECG analysis to other medical image classification tasks. For instance, Ali et al. employed a hybrid CNN model for diabetic retinopathy detection [42]. Additionally, Ahmad et al. demonstrated the efficacy of pre-trained CNNs for autism spectrum disorder detection from facial images [43]. Securing medical image transmission is also paramount, as addressed by Alarood et al., who used deep neural networks in their encryption approach [44]. These studies highlight the versatility of

deep learning in addressing a spectrum of healthcare challenges. Additionally, deep learning is actively being applied to brain tumor segmentation. Aslam et al. proposed an attention-based, lightweight model (AML-Net) specifically designed for this task within the Internet of Medical Things context [45]. Their model emphasizes efficiency and overcoming limitations in existing encoder-decoder architectures.

DNNs require sizeable computational power to obtain highly accurate results. Furthermore, the number of parameters in a network is directly proportional to the computational power required. Though the depth of the network may be directly proportional to its results, it may not be viable to deploy in small-scale medical institutions. Hence, this paper introduces a deep neural network (DNN) that exhibits reduced input parameters and computational expenses, while still delivering comparable accuracy to the cutting-edge techniques. The comparative analysis of existing techniques can be seen in Tables 1 and 2.

### 3 Materials and methods

This section gives a brief description of both the MIT-BIH Arrhythmia Database and the PTB Dataset. Furthermore, the section talks about the method of pre-processing and segmentation in the datasets used and explains the training parameters and the deep learning techniques used for the experimentation.

#### 3.1 Dataset description

##### 3.1.1 ECG waves

The results of an ECG are displayed as a graph with peaks and valleys. The spots on the ECG Diagnosis graph represent different waves of electrical activity. The waves in the graph can be broken down into three distinct waves.

The following are the three waves:

- **P wave:** This wave depicts the electrical activity traveling from the ventricles to the atria and running through the heart. These are the heart's two chambers located at the top.
- **QRS complex:** This represents the electrical activity that occurs from the atrium to the ventricles. These are the two chambers of the heart that are located lower down. The largest wave in the QRS complex is the R wave.
- **T wave:** This illustrates the electrical activity that is taking place during the ventricular repolarization of the heart. This indicates that it displays the electrical reset that the heart goes through as it gets ready for the next cardiac cycle.
- **ST segment:** It is a flat, isoelectric section of an ECG signal that follows the QRS complex and precedes the T wave. It represents the interval between ventricular depolarization and repolarization and is used in clinical diagnosis to assess myocardial ischemia and infarction. Any deviation from the baseline or a shift in the ST segment can indicate underlying cardiac abnormalities, making it a crucial feature in ECG analysis.

The ST segment can be seen in the results of the ECG. The part of the waveform may be seen between the end of the QRS complex and the beginning of the T wave. A person is diagnosed with ST depression if the ST segment is shallow and falls below the baseline as shown in Figs. 1 and 2 [46, 47]. ST depression is associated with a range of medical problems [48].

**Table 1** Summary of Literature Employing Deep Learning Approaches for Arrhythmia and Myocardial Infarction

Study	Methodology	Dataset(s)	Metrics	Strengths	Limitations
Pal et al. [25]	DenseNet + Transfer Learning	MIT-BIH, PTB	Accuracy	High accuracy	Computationally intensive
Hammad et al. [26]	DNN + Feature Extraction + GA	MIT-BIH	Acc., F1	Multi-stage approach for accuracy	Complex pipeline, potential for overfitting
Zhang et al. [27]	STA-CRNN	MIT-BIH	Acc., F1	Spatio-temporal focus, good for specific arrhythmias	May not generalize to all arrhythmia types
Hu et al. [28]	Transformer (ECG DETR)	MIT-BIH	Accuracy	Handles long-term dependencies well	May require large datasets for training
Natarajan et al. [29]	CNN + Transformer + Hand-crafted Features	MIT-BIH, PTB	Acc., F1	Combines diverse feature representations	Complex architecture, feature engineering needed
Strothhoff et al. [30]	Ensemble of RNNs and CNNs	PTB	Sens., Spec.,	Robust validation, handles 12-lead input	May still be computationally demanding
Wu et al. [31]	Deep feature model + Wavelet Transform	PTB	Accuracy	Focus on MI-specific features	May not generalize well to arrhythmias
Wang et al. [32]	Multi-lead ensemble network (MENN)	PTB	Accuracy	Designed for MI, outperforms baselines	Specialized for MI, not all arrhythmias

**Table 2** Comparison from the state-of-the-art

Author	Methods/models	Accuracy on MIT-BIH	Sensitivity on MIT-BIH	Accuracy on PTB	Sensitivity on PTB
Hammad et al. [26]	Multi-Stage DNN, Genetic Algorithm and SVM	98	99.7	—	—
Hu et al. [28]	Transformer based DNN	99.12	97.60	—	—
Mousavi et al. [33]	ELP + CNN	97.62	81.39	—	—
Teijeiro et al. [34]	Clustering	97.98	83.45	—	—
Das et al. [35]	MLPNN	97.50	74.83	—	—
Yang et al. [36]	PCA + SVM	97.76	83.41	—	—
Wu et al. [31]	Deep feature	—	—	—	99.64
Arif et al. [37]	Back Propagation Neural Networks	—	—	97.75	97.50
Liu et al. [38]	Novel ECG parameterization algorithm	—	—	98.33	96.66
Padhy et al. [39]	Third-order tensor based analysis	—	—	95.30	94.60
Banerjee et al. [40]	Cross Wavelet Transform	—	—	97.60	97.30
Acharya et al. [16]	Deep convolutional neural network	—	—	95.22	95.49
Sharma et al. [41]	Optimal Features Based Lead Specific Approach	—	—	99.05	98.34
Proposed Method	<b>eFuseNet</b>	<b>98.76</b>	<b>99.96</b>	<b>99.48</b>	<b>100</b>

The bold results are the results from the proposed approach and better from past works



**Fig. 1** ST segment of an ECG signal abnormally below baseline [46]

### 3.1.2 MIT-BIH arrhythmia database

The MIT-BIH Arrhythmia dataset contains 48 half-hour fragments of dual-channel ambulatory recordings from 47 subjects. Twenty-three recordings were selected randomly from a massive set of a mixed population of inpatient and outpatient subjects. The other twenty-five selected from the same set but included abnormal and clinically significant arrhythmias [49]. Depiction of various classes of the MIT-BIH Arrhythmia dataset is shown in Table 3 and Fig. 3.

### 3.1.3 PTB database

The PTB data set was collected using a unique PTB prototype recorder. The database has 549 different recordings from 290 subjects aged between 17-87, with a mean of 57.2. The dataset includes recordings of 15 simultaneous signals: 12 of them were the conventional leads together, and the rest 3 were Frank lead ECG signals. The digitization rate was 1000 samples/sec with a resolution of 16 bit over a range of  $\pm 16.384$  mV [50]. The various classes of the PTB dataset are depicted in Table 4 and Fig. 4.

## 3.2 Convolutional neural networks

Convolutional Neural Networks (CNNs) is one of the most popular types of Deep Neural Networks. It derives its name from the mathematical operation between vectors called convolution [51]. Convolutional Neural Networks consist of neurons analogous to the neurons in our brain where the output of the last neuron is the input of the neuron ahead [52]. A CNN consists of weights that are the parameters inside a neural network which converts the input



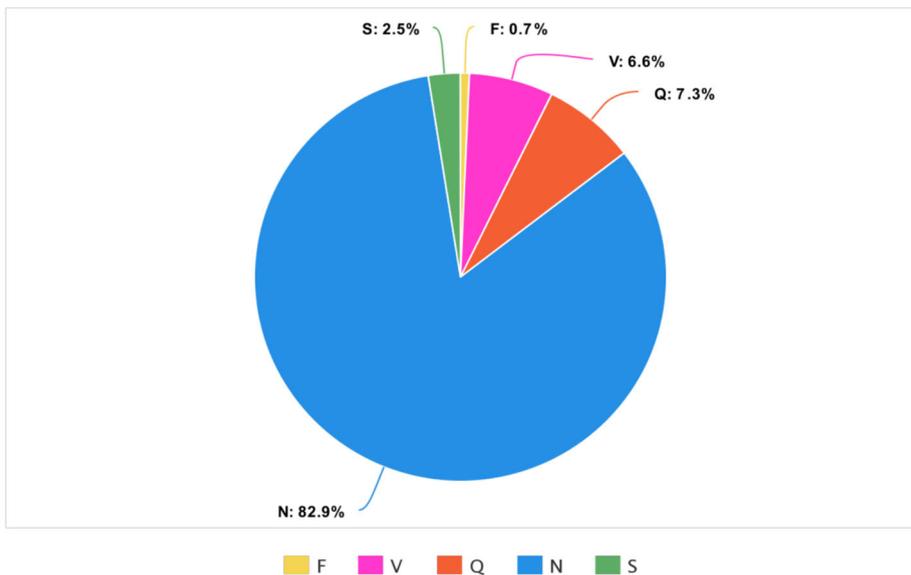
**Fig. 2** ST segment of an ECG signal abnormally above baseline [47]

**Table 3** Description of MIT-BIH database heartbeat classes grouped according to the AAMI standard

AAMI groups	MIT-BIH class	No. of samples
N	Normal beat	75052
	Left bundle branch block	7259
	Right bundle branch block	8075
	Atrial escape beat	16
	Nodal (junctional) escape beat	229
S	Atrial premature contraction beat	2546
	Aberrated atrial premature beat	150
	Nodal (junctional) premature beat	83
	Supraventricular premature beat	2
V	Premature ventricular contraction	7130
	Ventricular escape beat	106
F	Fusion of ventricular and normal beat	803
Q	Paced beat	7028
	Fusion of paced and normal beat	982
	Unclassified beat	33

data inside the network's hidden layer. Weights are usually encompassed within the hidden layers of the network.

It is a multi-layer deep neural network architecture consisting of many layers such as a convolutional layer, activation layer, pooling layer, etc. [53]

**Fig. 3** Graphical Representation of imbalance in the MIT-BIH Arrhythmia Dataset

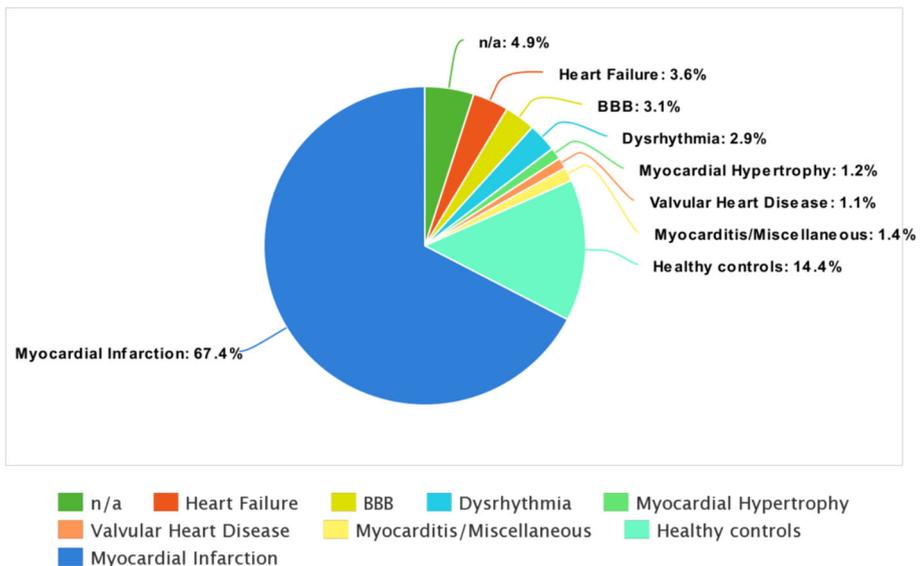
**Table 4** Description of PTB Diagnostic ECG Database Data

PTB Classes	No. of subjects	No. of records
Myocardial infarction (MI)	148	369
Cardiomyopathy/Heart failure	18	20
Bundle branch block	15	17
Dysrhythmia	14	16
Myocardial hypertrophy	7	7
Valvular heart disease	6	6
Myocarditis	4	4
Miscellaneous	4	4
Healthy controls (HC)	52	79
n/a	22	27
Total	290	549

1. **Convolutional Layer:** This layer consists of a rectangular grid of neurons. Since the convolutional layer takes inputs from previous layers, the previous layers must be rectangular. Here the weights are the same for each neuron. A CNN may have multiple convolutional layers which take inputs from the past layers and apply the same or different filters. Mathematically a convolutional layer may be represented as shown in (1).

$$G[m, n] = (f * h)[m, n] = \sum_j \sum_k h[j, k] f[m - j, n - k] \quad (1)$$

The above equation represents a convolution operation between an input image “f” and a kernel “h” to produce an output matrix “G”. The indexes of the resulting matrix are denoted by “m” and “n” to represent the rows and columns, respectively.

**Fig. 4** Graphical Representation of imbalance in the PTB Dataset

2. **Max-pooling layer:** Ahead of every convolutional layer, there may be a pooling layer that takes Input from the rectangular block and subsamples the output to an operation determining the maximum value in each block.
3. **Activation layer:** The activation layer is the final component of the convolutional layer. It aims to enhance the non-linearity in the Input. In this experiment, ReLu has been used as the activation function in the convolutional layer. Mathematically, ReLu can be described as which is the maximum value between zero and the input value from the neuron.

CNN is a famous go-to architecture for many machine learning problems, especially in image classification, computer vision, and natural language processing applications [54].

### 3.3 AlexNet

AlexNet is a convolution neural network, i.e., A neural network where the unseen layers are composed of convolutional layers, pooling layers, and normalization layers [55]. Primarily designed by Alex Krizhevsky in the paper [56], it was the first network that used a GPU to aid its performance. The architecture comprises eight layers, of which the first 5 are convolutional layers, and the last three are fully connected [57]. The input dataset consisted of  $256 \times 256$  RGB images from the ImageNet Dataset. A non-linear ReLU activation function was used since it gave better results than the Sigmoid and the tanh function [58]. In addition, ReLu normalized the data input to prevent saturation using their Local Response Normalisation (LRN) technique. By doing so, top-1 and top-5 error rates were reduced by 1.4% and 1.2%, respectively. AlexNet also allows for multi-parallel training methods by segregating the neurons equally on each GPU. However, when AlexNet is compared with VGG, ResNet, and other models, it stands less deep and struggles to absorb features from data sets. AlexNet has been chosen in the scope of this experimentation because of its ability to parallel segmentation which helps in the classification of signals. In addition to the previous point, AlexNet takes less time to train and gives a robust output.

### 3.4 Long short term memory

Long Short Term Memory is a type of Recurrent Neural Network architecture that was designed keeping in mind a network that can remember previous data and relate the previous data to make appropriate and suitable predictions. It was invented by Schmidhuber in 1997 [59]. LSTM averts the vanishing gradient issue by adding three gated units: forget gate, the output gate, and the input gate. These gates are used to effectively preserve the memory of previous states [60].

The vector equations presented below provide a mathematical representation of Long Short-Term Memory (LSTM) networks. The LSTM takes an input vector  $\mathbf{x}(t)$ , which can either be the output of a Convolutional Neural Network (CNN) or the input sequence itself. The LSTM also receives the hidden state vector  $\mathbf{h}(t-1)$  and cell state vector  $\mathbf{c}(t-1)$  from the previous timestep. The output vector  $\mathbf{o}(t)$  for the current timestep is generated by the LSTM, which also updates the cell state vector and hidden state vector for the next timestep, represented by  $\mathbf{c}(t)$  and  $\mathbf{h}(t)$ , respectively. The equations below illustrate the different stages of an LSTM cell, and (2) to (7) provide the mathematical equations for an LSTM cell.

$$f_t = \sigma_g (W_f \times x_t + U_f \times h_{t-1} + b_f) \quad (2)$$

$$i_t = \sigma_g (W_i \times x_t + U_i \times h_{t-1} + b_i) \quad (3)$$

$$o_t = \sigma_g (W_o \times x_t + U_o \times h_{t-1} + b_o) \tag{4}$$

$$c'_t = \sigma_c (W_c \times x_t + U_c \times h_{t-1} + b_c) \tag{5}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot c'_t \tag{6}$$

$$h_t = o_t \cdot \sigma_c (c_t) \tag{7}$$

where,  $f_t$  is forget gate,  $i_t$  is input gate,  $o_t$  is output gate,  $c_t$  is cell state,  $h_t$  is hidden state,  $\sigma_g$  : sigmoid,  $\sigma_c$  : tanh.

Similar to CNN, LSTM is used in a variety of machine learning application fields, particularly in the areas requiring pattern recognition applications such as Speech Recognition and Natural Language Processing [61]. The working of a LSTM cell is shown in Fig. 5.

### 3.5 Gated recurring units

Similar to LSTMs, Gated Recurrent Unit (GRU) is a type of Recurrent Neural Network architecture that solves the vanishing gradient problem by adding two gates such that the update gate and reset gate [60]. This reduces the number of gating signals when compared to the LSTMs. The GRU RNN can mathematically be represented as shown in (8).

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{8}$$

$$\tilde{h}_t = g (W_h x_t + U_h (r_t \odot h_{t-1}) + b_h)$$

And the two gates can be represented mathematically as depicted in the (9).

$$z_t = \sigma (W_z x_t + U_z h_{t-1} + b_z) \tag{9}$$

$$r_t = \sigma (W_r x_t + U_r h_{t-1} + b_r)$$

Like LSTM, GRU is extensively used in the areas used for predictions by pattern recognition applications [62]. The working of a GRU cell is shown in Fig. 6.

Now that we have reviewed the foundational concepts of Convolutional Neural Networks (CNNs), AlexNet, Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU), and gained an understanding of our dataset, we will now deep dive into the proposed methodology and the eFuseNet model.

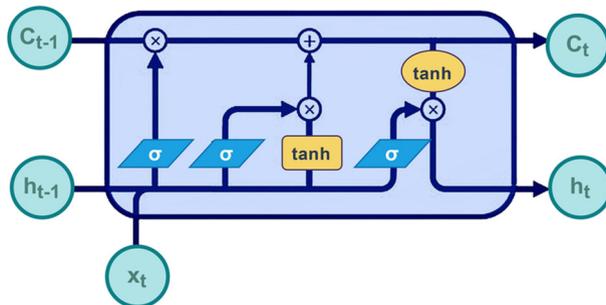


Fig. 5 Graphical Representation of an LSTM Cell

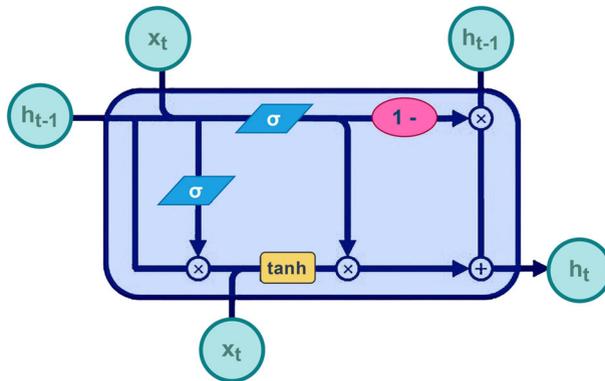


Fig. 6 Graphical Representation of an GRU Cell

## 4 Proposed method

### 4.1 Dataset preprocessing and segmentation

It is complicated to derive meaningful data from the unfiltered ECG signal in the MIT-BIH dataset due to extraneous features such as baseline, electro-myography disturbance, wandering, and power-line interference, which is why a step of filtration is necessary prior to advance processing procedures [63]. Since over-filtration will cause the loss of meaningful data, in the scope of this experiment, the authors eliminated the noise-baseline wandering that considerably impacts ECG classification. The movement and respiration of the patient cause baseline were wandered. Following earlier works, two median filters are elected to achieve baseline wandering, subtracting it from the signal to yield correct ECG signals without baseline wandering.

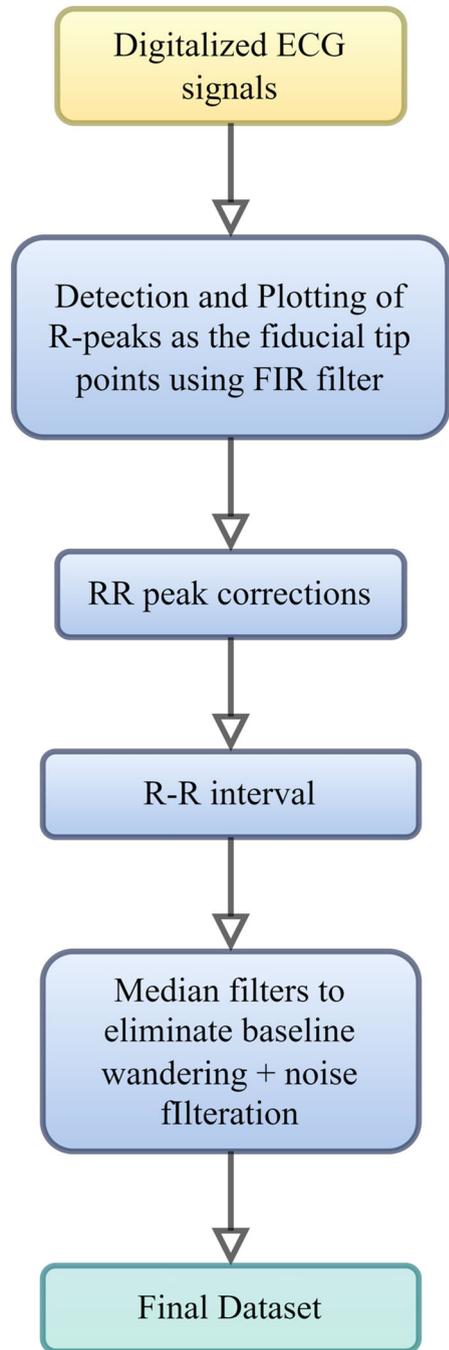
Before classifying the ECG signals, we need to segment individual heartbeat components from the Electrocardiogram samples [64]. Also, precise detection of QRS waves is required along with the fiducial tips of the heartbeats.

The heartbeats contain a variety of high-precision QRS waves and fiducial tips. Using specific annotated R-peak locations as fiducial tip points, the authors segmented the following ECG signal into a sequence of heartbeats which can be directly compared with other works. We acquired 190 samples before and 110 samples after the peak of the R-wave for every heartbeat to obtain a fixed size of 300 sample ECG signals. As a result, the sample peaks have been taken as the foremost heartbeat waves. The complete flowchart for data preprocessing can be seen in Fig. 7.

### 4.2 Modified alexnet architecture

AlexNet was originally designed to address the image classification problem using deep convolutional neural networks. However, ECG signals are a type of one-dimensional time series data, which requires modifications to the traditional AlexNet architecture, which accepts two-dimensional input. In this study, a modified version of the AlexNet architecture has been utilized to accept one-dimensional input data. Additionally, to prevent the model from overfitting, a dropout layer has been added between the convolution layer and the fully connected

**Fig. 7** Flow Chart of Data Preprocessing



layer. This dropout layer with a dropout rate of 0.5 helps eliminate 50% of the neurons in every iteration, resolving the issue of exploding gradients. Moreover, the size of the convolutional layer strides and the number of nodes in the fully connected layer have also been adjusted. The proposed modified AlexNet architecture, including all modifications, is illustrated in Fig. 8.

### 4.3 Training parameters

#### 4.3.1 Loss function: categorical cross-entropy

Loss Functions are computational methods to learn the error between actual and predicted results. There are various Loss functions available [65, 66]. Our study employs two variants of cross-entropy loss, chosen based on the specific classification tasks. For detecting myocardial infarction (a binary classification problem), the authors utilized the binary cross-entropy loss function as shown in (10). For arrhythmia classification (a multi-class problem), the authors used the categorical cross-entropy loss function as shown in (11).

$$J(\mathbf{w}) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \tag{10}$$

where,

\*  $\mathbf{w}$ : Model parameters (weights and biases) \*  $N$ : Number of samples \*  $y_i$ : True label (0 or 1) for sample  $i$  \*  $\hat{y}_i$ : Predicted probability of sample  $i$  belonging to the MI class

$$J(\mathbf{w}) = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^M y_{i,c} \log(\hat{y}_{i,c}) \tag{11}$$

where,

\*  $\mathbf{w}$ : Model parameters (weights and biases) \*  $N$ : Number of samples \*  $M$ : Number of arrhythmia classes \*  $y_{i,c}$ : True label (1 if sample  $i$  belongs to class  $c$ , otherwise 0) \*  $\hat{y}_{i,c}$ : Predicted probability of sample  $i$  belonging to class  $c$

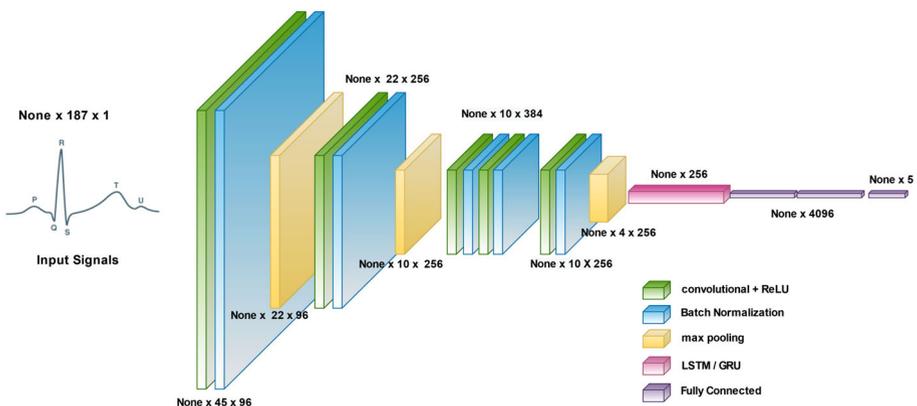


Fig. 8 Modified AlexNet Architecture

### 4.3.2 Optimizer: adam

In the presented research, the authors employed the Adam optimization algorithm to optimize our deep convolutional neural network model during training. This algorithm utilizes an exponential decay scheme to calculate the past gradients ( $m_t$ ) and past squared gradients ( $v_t$ ), as demonstrated by (12) and (13) respectively. The hyperparameters  $\beta_1$  and  $\beta_2$  dictate the rate at which the mean and non-centered variance of the gradient are forgotten, respectively. Through keeping track of these two measures, the Adam optimizer estimates the first and second moments of the gradients and utilizes them to adaptively adjust the learning rate.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \left[ \frac{\delta L}{\delta w_t} \right] \quad (12)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) \left[ \frac{\delta L}{\delta w_t} \right]^2 \quad (13)$$

where,

1.  $\epsilon$  = a small positive constant to avoid denominator becoming zero when ( $v_t - > 0$ )  $\cdot (10^{-8})$
2.  $\beta_1$  &  $\beta_2$  = decay rates of the average of gradients in the above two methods. ( $\beta_1 = 0.9$  &  $\beta_2 = 0.999$ )
3.  $\alpha$  denotes the step size or learning rate, which is typically set to 0.001. It determines the magnitude of adjustment made to the model's parameters during each iteration of the optimization process.

### 4.3.3 Classifier - softmax

In order to classify the extracted features, a Softmax classifier has been employed. The Softmax function is defined as:

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad \text{for } i = 1, \dots, K \quad (14)$$

where  $\mathbf{z} = (z_1, \dots, z_K)$  represents the vector of class scores and  $\sigma(\mathbf{z})_i$  represents the predicted probability of class  $i$ . This function converts a vector of real-valued scores into a probability distribution over  $K$  classes. The denominator in the equation ensures that the output is a valid probability distribution, where all the values range between 0 and 1 and the sum of all values is equal to 1.

## 4.4 Proposed ensemble fusion model - eFuseNet

To address the challenges of prediction variance and generalization in ECG classification, we propose eFuseNet, a novel ensemble fusion model. eFuseNet integrates a modified AlexNet architecture, LSTM, and GRU networks. We tailor the pre-trained AlexNet CNN to effectively extract discriminative features from 1-dimensional ECG waveforms, making modifications to suit our input data and classification task. These extracted features, representing higher-level abstractions of the ECG signal, are then fed as input to both the LSTM and GRU sub-models in parallel. These RNN variants excel in capturing the long-range temporal dependencies prevalent in ECG time-series data. eFuseNet fuses the outputs of the AlexNet+LSTM and

AlexNet+GRU sub-models through concatenation, feeding them into a final fully-connected layer. This approach leverages the complementary strengths of CNN-based feature extraction and RNN-based temporal modeling, with the goal of mitigating biases and improving classification robustness.

eFuseNet employs parallel AlexNet+LSTM and AlexNet+GRU sub-models, strategically combining their outputs to mitigate potential biases and exploit the complementary strengths of CNNs and RNNs. We optimize eFuseNet using Adam, a robust adaptive gradient descent algorithm, with Categorical Cross-Entropy as the loss function to guide multi-class arrhythmia classification. A final Softmax layer generates probabilistic outputs. To enhance the efficiency of the model, the authors used the different hyperparameters as shown in Table 5.

While computationally complex, with 38,955,535 total parameters (517 trainable), eFuseNet's architecture is designed to prioritize accurate arrhythmia classification. Our primary objective is to surpass the performance of single-model approaches in classifying diverse arrhythmia types within ECG time-series data. Figure 9 illustrates the proposed model - eFuseNet architecture.

## 5 Experimental results and discussion

### 5.1 Dataset division

Reliability is a crucial factor to consider when assessing a medical decision-making model. To ensure efficient evaluation, it is recommended to split the data into three parts - Training Set, Validation Set, and Testing Set. This practice helps prevent overfitting and model selection bias. The ratio in which the dataset is divided depends on the dataset's size. For instance, the authors divided the MIT-BIH dataset into 64% for training, and 16-20% for validation and testing, respectively. However, the authors evaluated model's performance on the PTB dataset to test its robustness. For this, we divided the data into 72% for training, 8% for validation, and 20% for testing. Table 6 illustrates the distribution of the PTB dataset for the 72-8-20 split, while Table 7 shows the distribution for the five classes of the MIT-BIH Arrhythmia Dataset for the 64-16-20 split.

**Table 5** Hyperparameters Set for the Model Training

Hyperparameters	Values
1st momentum Decay Rate ( $\beta_1$ )	0.9
2nd momentum Decay Rate ( $\beta_2$ )	0.999
Epsilon ( $\epsilon$ )	1e-7
Starting Learning rate ( $\alpha$ )	0.001
Factor	0.1
Patience	10
Total Epochs	250-300
Optimizer	Adam
Dropout value	0.5
Batch Size	32

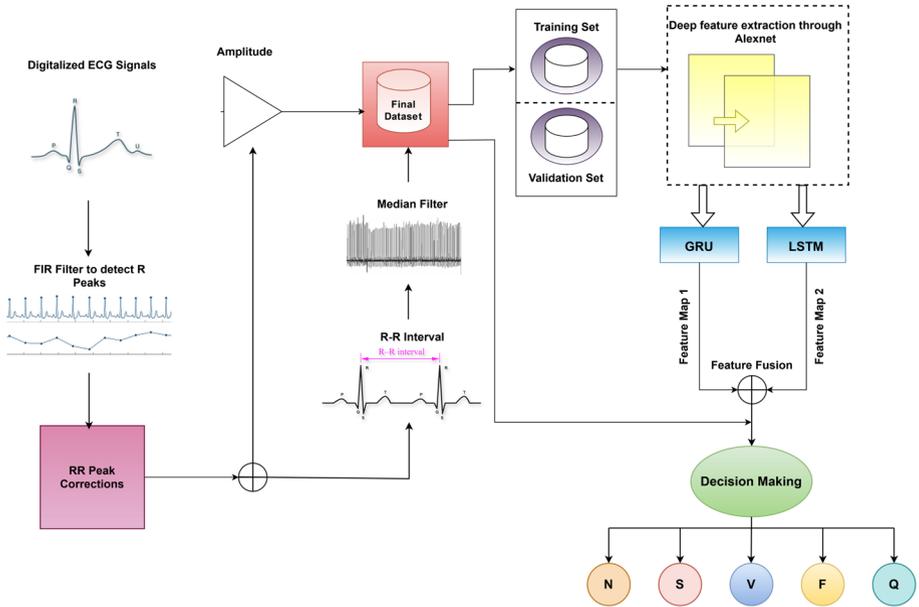


Fig. 9 Graphical abstract of the proposed work

### 5.2 Experimental setup

The present investigation utilized the Tensorflow framework for coding implementation, and the Models were trained on a workstation equipped with Intel® Core™ i7-10610U Processor and Nvidia GeForce® GTX 1650 GPU. Training was conducted for a range of 100-500 epochs to achieve optimal accuracy in training, validation, and testing, for both the MIT-BIH Arrhythmia and PTB datasets.

### 5.3 Evaluation metrics

To assess the performance, the authors used the models’ accuracy and F1-score as evaluation metrics.

1. **Precision:** It is a statistical metric used to evaluate the performance of a model. It is calculated by dividing the number of accurately predicted positive instances (True Positives) by the sum of the True Positive and False Positive instances. This metric is essential in

**Table 6** Distribution of PTB Dataset

	Myocardial Infarction	Healthy controls
Training Set	106	38
Validation Set	12	4
Testing Set	30	10

**Table 7** Distribution of MIT-BIH Arrhythmia Dataset

	N	S	V	F	Q
Training Set	19794	984	1302	9	850
Validation Set	4949	246	326	2	214
Testing Set	6186	308	407	3	265

assessing how well a model performs in predicting positive outcomes.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{15}$$

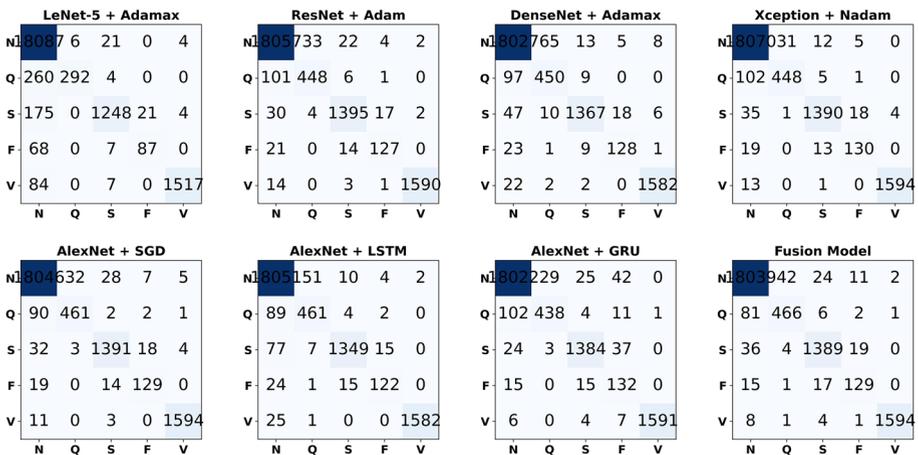
2. **Sensitivity (Recall):** It is another performance metric used to evaluate the performance of a model, especially in binary classification problems. Sensitivity is the ability of the model to correctly identify positive instances, and it is calculated as the ratio of true positives to the sum of true positives and false negatives:

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{16}$$

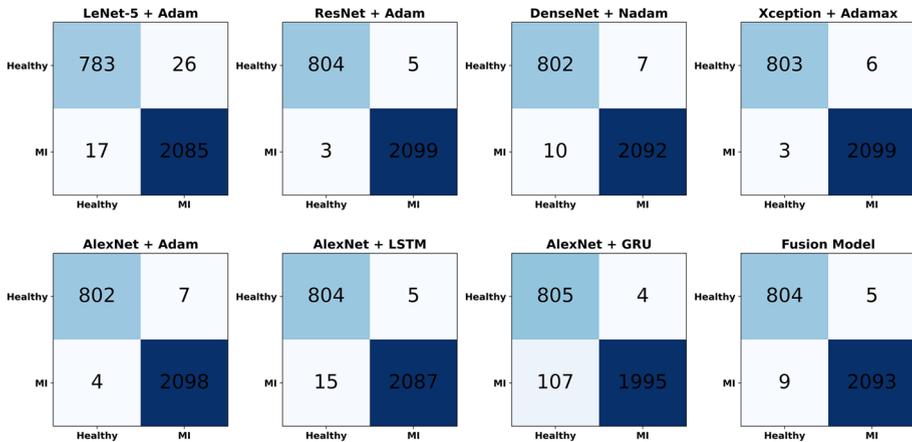
3. **F1 Accuracy:** It is a combined metric of Precision and Recall of a particular classifier which is calculated through the harmonic mean of Precision and Recall

$$F1 \text{ score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

4. **Confusion Matrix:** It is a metric to evaluate the performance of a classification model, and it can help visualize how the model is performing in terms of classifying positive and negative cases. It represents the number of true positives, false positives, true negatives, and false negatives for each class in a tabular format. The rows and columns of the matrix correspond to the predicted and actual class labels, respectively. The confusion matrices for different models, including EfuseNet, and their performance are shown in Figs. 10 and 11.



**Fig. 10** Graphical representation of the Confusion Matrices of best performing architectures and the EfuseNet when tested on the MIT-BIH dataset



**Fig. 11** Graphical representation of the Confusion Matrices of best performing architectures and the EfuseNet when tested on the PTB dataset

5. **10-fold cross-validation:** Cross-validation is a resampling technique used to assess machine learning models on a small data sample. The main purpose of cross-validation in applied machine learning is to assess the strength and resilience of a model. The authors have used k-fold cross-validation using the following steps:

- The authors trained proposed model arbitrarily ten times.
- The authors evaluated the accuracy of model 10 times and calculated the mean of all accuracies to test the robustness of the model.

Table 8 depicts the experimental results obtained on various architectures and the Proposed model EfuseNet on performing 10 Folds.

## 5.4 Results

Our experimentation examined various pre-trained deep neural networks on the MIT-BIH Arrhythmia and the PTB Myocardial datasets. Since most pre-trained models use image inputs, we modified the pre-trained models to accept 1-dimensional input (This is elaborately explained in Section 4.2). After analyzing the performance of multiple pre-trained deep neural networks, the authors tested the pre-trained models with various optimizers and recorded the results.

The authors conducted the experimentation with LeNet-5 and AlexNet, applying all seven optimizers to test. The results can be seen in Tables 9 and 10, respectively. From Table 9, it is clear that the result of Lenet-5 on PTB as well as MIT-BIH database are not up to the mark. AlexNet provided sound results; however, it had a lot of parameters. Thus, the authors experimented with Xception, ResNet, and DenseNet for all seven optimizers on both the MIT-BIH Arrhythmia and the PTB Myocardial datasets, results of which are shown in Tables 11, 12, and 13 respectively.

The MIT-BIH Dataset has five classes: N, SVEB, VEB, F, and Q. During experimentation, our primary interest was to devise a model that was efficient in terms of both accuracy and training time since the MIT-BIH dataset has an enormous sample space. In the MIT-BIH

**Table 8** Results obtained from 10-fold cross-validation experiments on various architectures and the proposed model EfuseNet

	MIT-BIH			PTB		
	AlexNet + LSTM Accuracy	AlexNet + GRU Accuracy	Fused Model Accuracy	AlexNet + LSTM Accuracy	AlexNet + GRU Accuracy	Fused Model Accuracy
Fold 1	98.59%	98.43	98.78	97.84	88.35	98.59
Fold 2	98.52	98.52	98.79	98.55	99.66	99.69
Fold 3	98.49	98.41	98.67	98.63	99.45	99.52
Fold 4	98.57	98.49	98.75	98.63	99.45	99.52
Fold 5	98.45	98.54	98.78	98.44	99.55	99.55
Fold 6	98.76	98.39	98.80	99.59	99.62	99.76
Fold 7	98.54	98.55	98.74	96.05	99.69	99.66
Fold 8	98.66	98.55	98.78	95.84	98.76	99.42
Fold 9	98.53	98.50	98.73	98.42	99.66	99.59
Fold 10	98.61	97.92	98.73	99.11	99.14	99.38
Average	<b>98.57% (+/- 0.08%)</b>	<b>98.43% (+/- 0.18%)</b>	<b>98.76% (+/- 0.04%)</b>	<b>97.81% (+/- 1.96%)</b>	<b>98.34% (+/- 3.34%)</b>	<b>99.48% (+/- 0.32%)</b>

The bold results are the results from the proposed approach and better from past works

**Table 9** Outcomes of the experiments conducted with the LeNet-5 network using various optimizers on the MIT-BIH and PTB datasets

Model Name	Optimizers	MIT-BIH Tr A (%)	MIT-BIH VA (%)	MIT-BIH TA (%)	PTB Tr A (%)	PTB VA (%)	PTB TA (%)
LeNet-5	RMSprop	88.32	90.82	91.06	87.56	92.27	91.12
LeNet-5	Nadam	95.07	96.69	96.53	95.46	98.54	98.32
LeNet-5	SGD	94.82	96.34	96.2	92.69	96.39	96.12
LeNet-5	Adamax	95.9	97.25	96.98	96.16	98.11	97.45
LeNet-5	Adadelta	89.82	91.7	91.6	74.29	76.91	77.30
LeNet-5	Adagrad	91.57	93.19	93.3	81.47	84.72	84.81
LeNet-5	Adam	94.92	96.54	96.46	95.49	98.8	98.52

**Table 10** Outcomes of the experiments conducted with the Alexnet network using various optimizers on the MIT-BIH and PTB datasets

Model Name	Optimizers	MIT-BIH Tr A (%)	MIT-BIH VA (%)	MIT-BIH TA (%)	PTB Tr A (%)	PTB VA (%)	PTB TA (%)
AlexNet	RMSprop	99.81	98.85	98.54	99.97	99.57	99.59
AlexNet	Nadam	99.75	98.87	98.63	99.88	99.66	99.41
AlexNet	SGD	100	99.01	98.73	100	99.4	99.62
AlexNet	Adamax	99.98	98.9	98.73	99.98	99.74	99.59
AlexNet	Adadelta	99.97	98.33	98.3	100	98.71	98.76
AlexNet	Adagrad	100	98.68	98.5	100	99.23	99.34
AlexNet	Adam	99.81	98.58	98.44	100	99.57	99.62

**Table 11** Outcomes of the experiments conducted with the Xception network using various optimizers on the MIT-BIH and PTB datasets

Model Name	Optimizers	MIT-BIH Tr A (%)	MIT-BIH VA (%)	MIT-BIH TA (%)	PTB Tr A (%)	PTB VA (%)	PTB TA (%)
Xception	RMSprop	99.99	98.87	98.73	99.99	99.48	99.48
Xception	Nadam	99.98	98.98	98.81	100	99.57	99.55
Xception	SGD	99.99	98.9	98.61	100	99.48	99.58
Xception	Adamax	99.99	98.85	98.77	99.93	99.4	99.69
Xception	Adadelta	99.88	98.36	98.36	99.97	98.2	98.48
Xception	Adagrad	99.99	98.6	98.51	99.99	98.97	99.27
Xception	Adam	99.96	98.93	98.75	100	99.23	99.55

**Table 12** Outcomes of the experiments conducted with the ResNet50 network using various optimizers on the MIT-BIH and PTB datasets

Model Name	Optimizers	MIT-BIH Tr A (%)	MIT-BIH VA (%)	MIT-BIH TA (%)	PTB Tr A (%)	PTB VA (%)	PTB TA (%)
ResNet	RMSprop	99.79	98.72	98.56	99.92	98.11	97.97
ResNet	Nadam	99.88	98.78	98.52	100	99.48	99.51
ResNet	SGD	100	98.93	98.71	100	99.48	99.45
ResNet	Adamax	99.97	98.98	98.72	100	99.66	99.69
ResNet	Adadelta	99.29	98	97.88	98.81	97.68	97.97
ResNet	Adagrad	99.96	98.48	98.43	99.99	98.71	99.24
ResNet	Adam	99.91	98.97	98.74	99.96	99.57	99.72

**Table 13** Outcomes of the experiments conducted with the DenseNet network using various optimizers on the MIT-BIH and PTB datasets

Model Name	Optimizers	MIT-BIH Tr A (%)	MIT-BIH VA (%)	MIT-BIH TA (%)	PTB Tr A (%)	PTB VA (%)	PTB TA (%)
DenseNet	RMSprop	99.11	98.42	98.2	99.98	98.03	98.62
DenseNet	Nadam	99.67	98.33	98.37	99.98	99.31	99.41
DenseNet	SGD	99.93	98.66	98.35	100	99.23	99.2
DenseNet	Adamax	99.88	98.56	98.45	99.9	98.28	98.21
DenseNet	Adadelta	99.91	97.74	97.64	99.98	94.25	96.28
DenseNet	Adagrad	99.94	98.54	98.34	100	98.71	98.69
DenseNet	Adam	99.64	98.34	98.18	99.9	97.94	98.41

Loss Curve of MIT-BIH

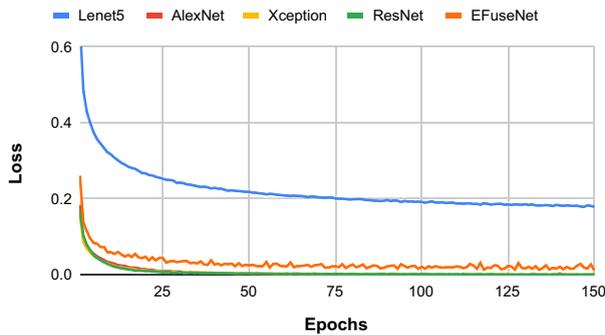


Fig. 12 Loss curve of training procedure of eFuseNet on MIT-BIH dataset

Arrhythmia dataset, Xception gave the best accuracy of 98.81% with Nadam Optimizer with an F-1 score of 0.93. Xception, when used with Adamax optimizer, gave an accuracy of 98.77% with an F-1 score of 0.926. Beyond these, AlexNet gave an accuracy of 98.73% with Adamax and SGD optimizers, respectively. However, since the MIT-BIH dataset has five classes, Xception was slow to classify a sample into five classes, despite its impressive results. Xception proved superior in accuracy, but AlexNet outshined Xception in efficiency matters.

The PTB dataset has nine classes. However, we are performing Binary Classification on the dataset into two classes: Myocardial Infarction and Healthy Controls. Xception gave an accuracy of 99.69%, a sensitivity of 99.26%, a specificity of 99.85, and an F1 score of 0.99. However, even with fewer samples in the PTB dataset, Xception took a long time to train. AlexNet gave impressive results, slightly less than Xception, but much more efficient in terms of training time as compared to Xception. AlexNet, when trained with optimizers RMSprop and Adamax, gave a 99.59% accuracy, which matched Xception’s performance. The loss and accuracy curve of various architectures and EfuseNet on the MIT-BIH dataset is shown in Figs. 12 and 13. At the same time, the loss and accuracy curve of various architectures and EfuseNet on the PTB dataset is shown in Figs. 14 and 15.

Accuracy Curve of MIT-BIH

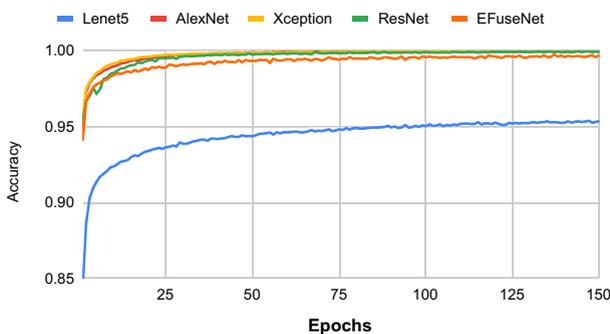
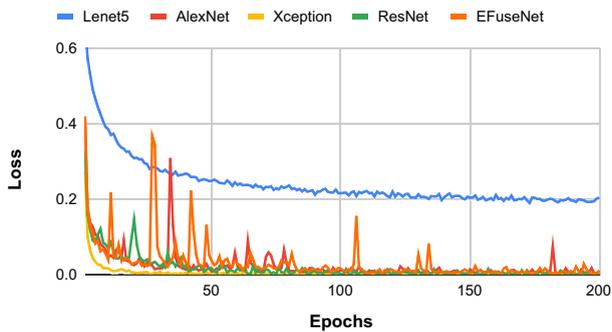


Fig. 13 Accuracy curve of training procedure of eFuseNet on MIT-BIH dataset

Loss Curve of PTB



**Fig. 14** Loss curve of training procedure of eFuseNet on PTB dataset

## 6 Discussion

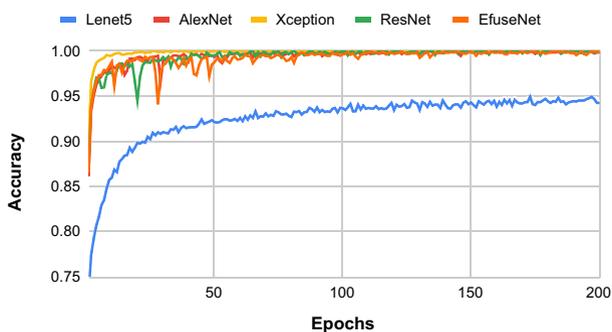
This paper presents a comprehensive study of the performance of various neural networks along with an eFuseNet: a proposed ensemble fusion deep neural network that detects and classifies Arrhythmia and Myocardial Infarction. The results of various pre-trained neural networks are presented in the results section. Furthermore, the authors have tested various neural networks on all seven optimizers - Adam, SGD, RMSprop, Adamax, Adagrad, Adadelata, and Nadam.

Since most of pre-trained models required image inputs, the authors modified the network architecture that enabled us to give 1-Dimensional Electrocardiograms as input. We observed that a lot of research had been done to improve metrics such as accuracy and F1-score.

Thus, the authors proposed eFuseNet model, that gives robust results and does not take much time to train. Furthermore, since previous studies highlighted the bias problem because of unbalanced datasets, the authors used an ensemble fusion technique to counter that bias.

eFuseNet gives an accuracy of 98.76% on the MIT-BIH and a 99.48% accuracy on the PTB dataset, demonstrating competitive performance compared to pre-existing methods to classify Arrhythmia and Myocardial Infarction. Furthermore, eFuseNet is computationally efficient compared to other studies mentioned in Table 2.

Accuracy Curve of PTB



**Fig. 15** Accuracy curve of training procedure of eFuseNet on PTB dataset

## 7 Conclusion and future direction

This study explored the utilization of ECG signals for detecting and classifying arrhythmia and myocardial infarction using the MIT-BIH and PTB datasets. We conducted a thorough investigation of various pre-trained networks and optimizers, offering insights into effective configurations for ECG analysis. Our proposed ensemble model, combining AlexNet, LSTM, and GRU, demonstrated exceptional performance with test accuracies of 98.76% and 99.48% on the MIT-BIH and PTB datasets, respectively. These findings can serve as a valuable reference for researchers in the field.

While our results are promising, it's important to acknowledge potential weaknesses. The reliance on the MIT-BIH and PTB datasets could be expanded in future work to include more diverse and potentially noisy real-world ECG data. Additionally, the computational complexity of our ensemble model could be addressed through optimization techniques, making it more suitable for resource-limited clinical environments. Moreover, integrating explainability methods could enhance the model's interpretability for clinicians. Looking ahead, our approach might be adaptable to diagnose other cardiovascular conditions or applied to different types of medical sensor data. This work lays a foundation for further advancements in automated cardiac diagnosis, with the potential to improve clinical workflows and patient outcomes.

**Data availability** The data used in this study is publicly available at below links:

1. **MIT-BIH:** <https://physionet.org/content/mitdb/1.0.0/>.
2. **PTB:** <https://physionet.org/content/ptbdb/1.0.0/>.

## Declarations

**Conflicts of interest** All the authors do not have any possible conflicts of interest.

## References

1. Cardiovascular diseases. World Health Organization
2. Wacker-Gussmann A, Oberhoffer-Fritz R (2022) Cardiovascular risk factors in childhood and adolescence. MDPI
3. Reed GW, Rossi JE, Cannon CP (2017) Acute myocardial infarction. *The Lancet*. 389(10065):197–210
4. Cenitta D, Arjunan RV, Prema K (2022) Ischemic heart disease multiple imputation technique using machine learning algorithm. *Engineered Science*. 19:262–272
5. Stuart SDF, De Jesus NM, Lindsey ML, Ripplinger CM (2016) The crossroads of inflammation, fibrosis, and arrhythmia following myocardial infarction. *J Mol Cell Cardiol* 91:114–122
6. Fenton FH, Cherry EM, Glass L (2008) Cardiac arrhythmia. *Scholarpedia*. 3(7):1665
7. Paffenbarger RS Jr, Wing AL, Hyde RT (1978) Physical activity as an index of heart attack risk in college alumni. *Am J Epidemiol* 108(3):161–175
8. Singh N, Singh P (2019) Cardiac arrhythmia classification using machine learning techniques. In: *Engineering vibration, communication and information processing*, pp 469–480. Springer. ???
9. Tang X, Ma Z, Hu Q, Tang W (2019) A real-time arrhythmia heartbeats classification algorithm using parallel delta modulations and rotated linear-kernel support vector machines. *IEEE Trans Biomed Eng* 67(4):978–986
10. Alfaras M, Soriano MC, Ortín S (2019) A fast machine learning model for ecg-based heartbeat classification and arrhythmia detection. *Frontiers in Physics*. 7:103
11. Tang DH, Gilligan AM, Romero K (2014) Economic burden and disparities in healthcare resource use among adult patients with cardiac arrhythmia. *Appl Health Econ Health Policy* 12(1):59–71
12. Newman MEJ (2013) Network data. <http://www-personal.umich.edu/~mejn/netdata/>

13. Kropf M, Hayn D, Schreier G (2017) Ecg classification based on time and frequency domain features using random forests. In: 2017 Computing in cardiology (CinC), pp 1–4. IEEE
14. Han C, Wang P, Huang R, Cui L (2022) Hctnet: An experience-guided deep learning network for inter-patient arrhythmia classification on imbalanced dataset. *Biomed Signal Process Control* 78:103910
15. Sinha N, Tripathy RK, Das A (2022) Ecg beat classification based on discriminative multilevel feature analysis and deep learning approach. *Biomed Signal Process Control* 78:103943
16. Acharya UR, Fujita H, Oh SL, Hagiwara Y, Tan JH, Adam M (2017) Application of deep convolutional neural network for automated detection of myocardial infarction using ecg signals. *Inf Sci* 415:190–198
17. Khalifa Y, Mandic D, Sejdić E (2021) A review of hidden markov models and recurrent neural networks for event detection and localization in biomedical signals. *Information Fusion*. 69:52–72
18. Lynn HM, Pan SB, Kim P (2019) A deep bidirectional gru network model for biometric electrocardiogram classification based on recurrent neural networks. *IEEE Access*. 7:145395–145405
19. Yildirim O, Baloglu UB, Tan R-S, Ciaccio EJ, Acharya UR (2019) A new approach for arrhythmia classification using deep coded features and lstm networks. *Comput Methods Programs Biomed* 176:121–133
20. Kim J-K, Jung S, Park J, Han SW (2022) Arrhythmia detection model using modified densenet for comprehensible grad-cam visualization. *Biomed Signal Process Control* 73:103408
21. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *nature*. 521(7553):436–444
22. Li J, Pang S-p, Xu F, Zhou S, Shu M (2022) Two-dimensional ecg-based cardiac arrhythmia classification using dse-resnet
23. Srivastava G, Chauhan A, Jangid M, Chaurasia S (2022) Covixnet: A novel and efficient deep learning model for detection of covid-19 using chest x-ray images. *Biomed Signal Process Control* 103848
24. Srivastava G, Pradhan N, Saini Y (2022) Ensemble of deep neural networks based on condorcet's jury theorem for screening covid-19 and pneumonia from radiograph images. *Comput Biol Med* 149:105979
25. Pal A, Srivastava R, Singh YN (2021) Cardionet: An efficient ecg arrhythmia classification system using transfer learning. *Big Data Research*. 26:100271
26. Hammad M, Ilyyas AM, Subasi A, Ho ES, Abd El-Latif AA (2020) A multitier deep learning model for arrhythmia detection. *IEEE Trans Instrum Meas* 70:1–9
27. Zhang J, Liu A, Gao M, Chen X, Zhang X, Chen X (2020) Ecg-based multi-class arrhythmia detection using spatio-temporal attention-based convolutional recurrent neural network. *Artif Intell Med* 106:101856
28. Hu R, Chen J, Zhou L (2022) A transformer-based deep neural network for arrhythmia detection using continuous ecg signals. *Comput Biol Med* 144:105325
29. Natarajan A, Chang Y, Mariani S, Rahman A, Boverman G, Vij S, Rubin J (2020) A wide and deep transformer neural network for 12-lead ecg classification. In: 2020 Computing in cardiology, pp 1–4 . IEEE
30. Strodthoff N, Strodthoff C (2019) Detecting and interpreting myocardial infarction using fully convolutional neural networks. *Physiol Meas* 40(1):015001
31. Wu J, Bao Y, Chan S-C, Wu H, Zhang L, Wei X-G (2016) Myocardial infarction detection and classification—a new multi-scale deep feature learning approach. In: 2016 IEEE international conference on digital signal processing (DSP), pp 309–313. IEEE
32. Wang H, Zhao W, Jia D, Hu J, Li Z, Yan C, You T (2019) Myocardial infarction detection based on multi-lead ensemble neural network. In: 2019 41st Annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 2614–2617. IEEE
33. Mousavi S, Afghah F, Khadem F, Acharya UR (2021) Ecg language processing (elp): A new technique to analyze ecg signals. *Comput Methods Programs Biomed* 202:105959
34. Teijeiro T, Félix P, Presedo J, Castro D (2016) Heartbeat classification using abstract features from the abductive interpretation of the ecg. *IEEE J Biomed Health Inform* 22(2):409–420
35. Das MK, Ari S (2014) Ecg beats classification using mixture of features. *Int Scholarly Res Notices* 2014
36. Yang W, Si Y, Wang D, Guo B (2018) Automatic recognition of arrhythmia based on principal component analysis network and linear support vector machine. *Comput Biol Med* 101:22–32
37. Arif M, Malagore IA, Afsar FA (2010) Automatic detection and localization of myocardial infarction using back propagation neural networks. In: 2010 4th International conference on bioinformatics and biomedical engineering, pp 1–4. IEEE
38. Liu B, Liu J, Wang G, Huang K, Li F, Zheng Y, Luo Y, Zhou F (2015) A novel electrocardiogram parameterization algorithm and its application in myocardial infarction detection. *Comput Biol Med* 61:178–184
39. Padhy S, Dandapat S (2017) Third-order tensor based analysis of multilead ecg for classification of myocardial infarction. *Biomed Signal Process Control* 31:71–78

40. Banerjee S, Mitra M (2013) Application of cross wavelet transform for ecg pattern analysis and classification. *IEEE Trans Instrum Meas* 63(2):326–333
41. Sharma L, Sunkaria R (2020) Myocardial infarction detection and localization using optimal features based lead specific approach. *Irbm*. 41(1):58–70
42. Ali G, Dastgir A, Iqbal MW, Anwar M, Faheem M (2023) A hybrid convolutional neural network model for automatic diabetic retinopathy classification from fundus images. *IEEE J Transl Eng Health Med*
43. Ahmad I, Rashid J, Faheem M, Akram A, Khan NA, Amin Ru (2024) Autism spectrum disorder detection using facial images: A performance comparison of pretrained convolutional neural networks. *Healthcare Technology Letters*
44. Alarood AA, Faheem M, Al-Khasawneh MA, Alzahrani AI, Alshdadi AA (2023) Secure medical image transmission using deep neural network in e-health applications. *Healthcare Technology Letters*. 10(4):87–98
45. Zeeshan Aslam M, Raza B, Faheem M, Raza A (2024) Aml-net: Attention-based multi-scale lightweight model for brain tumour segmentation in internet of medical things. *CAAI Trans Intell Technol*
46. Zhan Z-Q, Li Y, Han L-H, Nikus KC, Birnbaum Y, Baranchuk A (2020) The de winter ecg pattern: Distribution and morphology of st depression. *Ann Noninvasive Electrocardiol* 25(5):12783
47. Teixeira R, Lourenço C, António N, Monteiro S, Baptista R, Jorge E, Ferreira MJ, Monteiro P, Freitas M, Providência LA (2010) The importance of a normal ecg in non-st elevation acute coronary syndromes. *Arq Bras Cardiol* 94:25–33
48. Kashou AH, Basit H, Malik A (2022) St segment. In: *StatPearls* [Internet]. StatPearls Publishing, ???
49. Moody GB, Mark RG (2001) The impact of the mit-bih arrhythmia database. *IEEE Eng Med Biol Mag* 20(3):45–50
50. Boussejot R, Kreiseler D, Schnabel A (1995) Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet
51. Albawi S, Mohammed TA, Al-Zawi S (2017) Understanding of a convolutional neural network. In: 2017 International conference on engineering and technology (ICET), pp 1–6. Ieee
52. Wang J, Li Z (2018) Research on face recognition based on cnn. In: *IOP Conference Series: Earth and Environmental Science*, vol 170, p 032110. IOP Publishing
53. Mathunjwa BM, Lin Y-T, Lin C-H, Abbod MF, Shieh J-S (2021) Ecg arrhythmia classification by using a recurrence plot and convolutional neural network. *Biomed Signal Process Control* 64:102262
54. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L (2021) Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*. 8(1):1–74
55. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 25:1097–1105
56. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Pereira F, Burges CJ, Bottou L, Weinberger KQ (eds) *Advances in Neural Information Processing Systems*, vol. 25. Curran Associates, Inc., ??? <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
57. Saxena S (2021) Introduction to the architecture of alexnet. *Analytics Vidhya*
58. Wei J (2019) Alexnet: The architecture that challenged cnns. *Towards Data Science* 3
59. Sak H, Senior A, Beaufays F (2014) Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. *arXiv preprint arXiv:1402.1128*
60. Wang Y (2017) A new concept using lstm neural networks for dynamic system identification. In: 2017 American control conference (ACC), pp 5324–5329. IEEE
61. Van Houdt G, Mosquera C, Nápoles G (2020) A review on the long short-term memory model. *Artificial Intelligence Review* 53. <https://doi.org/10.1007/s10462-020-09838-1>
62. Dey R, Salem FM (2017) Gate-variants of gated recurrent unit (gru) neural networks. In: 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS), pp 1597–1600. IEEE
63. Luo S, Johnston P (2010) A review of electrocardiogram filtering. *Journal of Electrocardiology*. 43(6):486–496. <https://doi.org/10.1016/j.jelectrocard.2010.07.007>
64. Qiu X, Liang S, Zhang Y (2020) Simultaneous ecg heartbeat segmentation and classification with feature fusion and long term context dependencies. In: *Pacific-Asia conference on knowledge discovery and data mining*, pp 371–383. Springer
65. DeepAI: Loss function. DeepAI (2019)
66. Parmar R (2018) Common loss functions in machine learning. *Towards Data Science*

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.