Sequential and Spatial Learning for ECG-Based Arrhythmia Detection Using CNN-LSTM

Anerudh R
Vellore Institute of Technology
Chennai, India
anerudh.r2022a@vitstudent.ac.in

V Varun
Vellore Institute of Technology
Chennai, India
varun.v2022a@vitstudent.ac.in

Harisri Sanjivan
Vellore Institute of Technology
Chennai, India
harisri.sanjivan@vitstudent.ac.in

Abstract—Cardiac arrhythmias are irregularities in the heartbeat that can lead to severe health complications if not diagnosed promptly. Early and accurate detection plays a vital role in preventing potentially life-threatening conditions such as stroke and heart failure. This paper presents a hybrid deep learning architecture that combines a 14-layer Convolutional Neural Network (CNN), dense layer, and Long Short-Term Memory (LSTM) network for the robust classification of arrhythmias using ECG signals. The raw ECG data were first subjected to preprocessing using the Daubechies 4 (db4) wavelet transform, which effectively denoises the signal while preserving the critical morphological features of the heartbeats. The CNN layers were employed to extract high-level spatial features from the ECG waveforms, followed by a dense layer to integrate and refine these features. The LSTM component then models the temporal dependencies between beats, enabling the model to learn sequential patterns that are crucial for accurate arrhythmia classification. Finally, a SoftMax classifier supported by auxiliary classifiers was used for the final decision-making. The model was trained and evaluated on benchmark ECG datasets, and demonstrated high accuracy, sensitivity, and specificity across five arrhythmia classes. Owing to its low latency and high performance, this architecture shows promise for real-time implementation in wearable health-monitoring systems, aiding continuous and automated cardiac care.

Keywords: electrocardiogram (ECG), Arrhythmia Detection, Wavelet Transform, Long Short-Term Memory (LSTM), Convolutional Neural Network, Signal Denoising, Biomedical Signal Processing, Deep Learning.

I. Introduction

Heart disease continues to be one of the world's leading health concerns, and among its many forms, arrhythmias are particularly tricky—they can be sudden, irregular, and difficult to detect. The detection of these irregular heartbeats has improved dramatically in recent years. With the help of deep learning, especially ECG signals, arrhythmia detection can be automated far more efficiently than before [1], [2]. Older methods require manually selecting features from ECG signals, which are both tedious and error-prone. But now, techniques like wavelet transforms help clean signals and make important features stand out [3], [4]. At the same time, convolutional neural networks (CNNs) have become a go-to tool for identifying patterns in ECG data, making classifications more accurate and reliable [5], [6]. As wearable ECG devices become more common, there is also a growing need for models that are not only accurate but also sufficiently light to work in real time. Researchers have begun training CNNs specifically for this type of applications [7]. Among the preprocessing methods, the Daubechies wavelet (db4) stands out for how well it cleans and segments ECG signals [8], [11]. By combining CNNs with LSTM layers, models can now understand both the shape of the signal and how it changes over time—something hybrid models have performed well with in recent studies [9], [10]. Inspired by this progress, our work introduces a deep, 14-layer CNN-LSTM model that includes auxiliary classifiers and a SoftMax output. We pre-processed the ECG signals using db4 wavelet denoising and trained the model to classify them into five types of arrhythmias. The design is lightweight, accurate, future-ready, and ideal for real-time analysis and wearable healthcare devices [12].

II. Literature Review

Wavelet transforms play a key role in ECG signal preprocessing. They can pinpoint short-lived features and reduce noise. Gupta and Mitra [2] showed how well the discrete wavelet transform (DWT) works to clean up ECG signals and spot QRS complexes. This leads to reliable arrhythmia detection. Barros et al. [8] combined wavelet transforms with envelope analysis. This combination made it easier to find QRS complexes in both ECG and photoplethysmogram (PPG) signals, which improved their timing accuracy. Zhang and his team [10] introduced new ways to break down complex ECG signals, like tunable Q-factor wavelet transform (TQWT) and complete ensemble empirical mode decomposition (CEEMD). These methods help obtain detailed information for deep learning models. Similarly, Mehta et al. [11] and Talo et al. [4] used DWT with synthetic minority oversampling (SMOTE) to balance different types of heart rhythm problems, ensuring that they could obtain good information from all kinds of arrhythmias.

.

CNN Architectures for Classification

CNNs are the top choice for sorting out different heart rhythms because they can learn features step-by-step. Shukla et al. [1] set a high bar by mixing wavelet transforms with a 12-layer CNN with an accuracy of 98.7% on the MIT-BIH dataset. Sharma and Gupta [6] took this even further with a 15-layer CNN that used wavelet-based peak detection. They reached 99.42% accuracy, showing that using more layers can make a significant difference.

Murat et al. [5] investigated designs that boost productivity. They made CNNs work better for real-time processing without losing their accuracy. Liu et al. [7] created a small 1D CNN for wearable devices. It showed 97.3% accuracy when spotting three types of arrhythmias. These studies demonstrate the balance between the complexity of a model and its ease of use in real life.

Hybrid and Ensemble Models

Models that combine CNNs with other algorithms have an impact on the boosting performance. Nguyen et al. [3] paired CNNs with extreme learning machines (ELMs) to speed up classification cutting computational costs by 30% while keeping 98.1% accuracy. Al Rahhal et al. [9] came up with a 2D CNN-SVM hybrid using SVMs to fuse decisions to boost specificity in minority classes. Zhang et al. [10] mixed CNNs with long short-term memory (LSTM) networks picking up on time-based patterns in ECG signals to achieve 99.1% accuracy on the PhysioNet dataset.

Tackling Data Imbalance

Uneven class distribution in ECG datasets remains a significant problem. Talo et al. [4] and Mehta et al. [11] used SMOTE to create more samples of rare arrhythmia classes boosting recall rates by 12–15%. Sharma and Gupta [6] got high accuracy without balancing the data hinting that deep architectures might lessen imbalance through layered learning of representations.

Real-Time and Wearable Applications

The move toward on-the-go monitoring has sparked new ideas in lightweight models. Liu and his team [7] fine-tuned a seven-layer CNN for ECGs you can wear making it possible to spot irregular heartbeats non-stop with very little delay. Barros et al. [8] took this further to cover PPG signals, paving the way for building them into smart watches. Murat's group [5] put the spotlight on ways to shrink models, like cutting out parts and using fewer bits, to make them work better on small devices.

.

III. Methodology

The proposed framework for arrhythmia detection involves four main stages: (1) ECG signal acquisition and segmentation, (2) noise elimination through wavelet denoising and moving average filtering, (3) feature extraction by employing a 12-layer 1D CNN with residual blocks, and (4) multiclass classification using SoftMax. Figure 1 shows the pipeline. Each MIT-BIH database ECG record was pre-processed into 180-sample R-peak-Centered segments, normalized, and directly input into the CNN to avoid manual feature engineering.

Input Data Used

PhysioNet (http://www.physionet.org) [21] contains the MIT-BIH database from which we extracted the data for this study. There were 48 recordings of two-channel dynamic ECGs signals in this database. Each recording had a maximum duration of 30 min and a 360HZ sample frequency. MT-BIH annotates every beat with the class to which it belongs. Figure I shows the number of beats per class in the MIT-BIH database. To train and test the validity of our method, 44 ECG recordings from lead II (MLII) in the database were selected. The four beats were overlooked because of their low signal quality for post-processing according to the AAMI standard (102, 104, 107, 217). The data were split into a training set (50%) and a test set (50%). There were 13200 non-duplicate occurrences in each collection. The total number of data selected for every dataset is indicated in Table I.

Data Processing

Generally, several interference noises are easily blended during the collection process because of the poor ECG signal and the acquisition equipment effect. Nevertheless, this noise makes it extremely challenging to analyse the ECG readings. Thus, before the classification of the ECG, adequate preprocessing of the ECG signals is essential. Power frequency interference, baseline drift, and electromyographic interference were the three most prevalent ECG signal interference sounds. The Daubechies 4 wavelet transform and moving average filter are both employed to denoise the ECG signal. In practice, useful signals usually manifest as low-frequency or smoother signals, and noisy signals usually manifest as high-frequency signals during signal processing. High-frequency wavelet coefficients are derived from noisy signals when the noisy signals are divided by the wavelet transform. Then, the high-frequency wavelet coefficients were threshold-processed to eliminate power line interference and electromyography interference. Finally, an inverse wavelet transformation was employed to reconstruct the signals. The baseline drift noise was removed while traveling using the average filter.

CNN-LSTM Model Architecture

Deep learning is a technique that involves automated feature extraction, wherein raw input data are progressively converted into high-level representations. Convolutional Neural Networks (CNNs) are among the most common deep learning architectures, distinguished by their capability to capture spatial hierarchies in data. Unlike traditional machine learning methods, CNNs are resourceful in learning and gaining knowledge about patterns on their own by running them through various convolution layers.

A CNN can be considered as a course of action that is structurally similar to a multilayer perceptron (MLP); however, the difference lies in the use of local connections and weight sharing instead of fully connected layers, which are utilized to process the information efficiently. This further leads to a reduction in the parameters without harming spatial relationships. The basic CNN model consists of three layers: a convolutional layer, a pooling layer, and a fully connected layer. It also had two components.

Feature Extractor: It has convolutional and pooling layers and performs the task of the energy conversion process, whereas the classifier is responsible for the classification of various arrhythmia types based on the features available in the former.

Feature Extractor: Comprising the convolutional and pooling layers and automatically returning the features in the

ECG signals that are most relevant to the task of arriving at a diagnosis.

Classifier: A fully connected multi-layer perceptron (MLP) that classifies the extracted features and thus establishes arrhythmia classification.

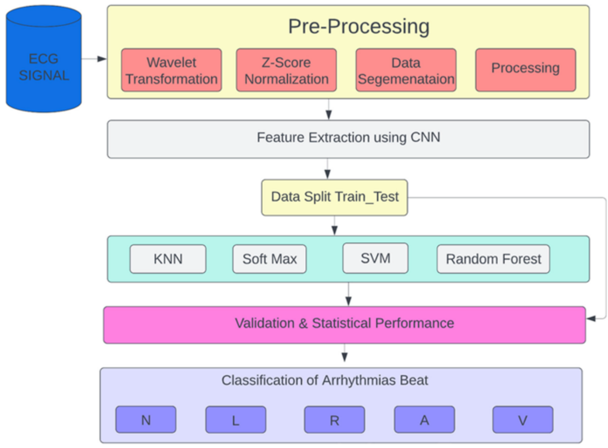

Fig.1 Model Architecture

Feature Extraction using CNN

The feature extraction process is initiated by the employment of convolutional layers, which traverse ECG signals with filters (kernels) to obtain the most valuable features. These filters spot the various features of the ECG signals, such as the P-QS-T components, which correspond to the arrhythmic patterns that are the most critical. The feature map resulting from the convolution operation is represented as:

$$cl,j = \sigma\left(b - \sum_{j=1}^{M} w_j x_{0j}\right)$$

i. w_j is the weight of the j-th feature map,
ii. x_{0j} is the input vector (ECG segment),
iii. b is the bias term,
iv. σ represents the activation function,
v. M is the kernel size.

Performance Metrics

The success of the CNN model was measured in terms of % accuracy and loss value, both of which are useful in evaluating the performance of multiclass classification problems.

Accuracy

The overall accuracy of the model was the proportion of samples that were correctly labelled by the classifier. Equation for determining accuracy as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Here, TP is True positive, TN is True negative, FP is False positive, and FN is False negative. The number of correct predictions divided by the total number of input data is the ratio.

Loss

A loss is a numerical value that reflects the extent to which the model's prediction is off in a certain case. If the model's forecast is perfect, the loss is zero; otherwise, it is greater. The goal of training a model is to find a set of weights and biases that, on average, result in minimal loss across all cases.

Precision

It informs you what percentage of optimistic forecasts were truly positive

$$Precision = \frac{TP}{TP + FP}$$

Recall

It tells you what proportion of all positive samples the classifier correctly predicted. Other synonyms include True-Positive Rate (TPR), Sensitivity, and Probability of Detection.

$$Recall = \frac{TP}{TP + FN}$$

3.4.5 F -Score

It is a measure that combines precision and recall. It is the harmonic mean of precision and recall in mathematics.

$$F - Score = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN}$$

IV. Results and Discussion

The MIT-BIH database was used to evaluate the performance of the proposed technique. The training of the model comprises a total of 10 epochs with a batch size of 32. The model accuracy curves and loss curves with respect to the number of epochs are shown in the figures below. In these figures, the x-axis represents the number of epochs used by the SoftMax layer, and the y-axis represents the accuracy of the model obtained during training and testing on data samples. Similarly, the y-axis represents the training and test loss of the model using the Adam optimizer in the SoftMax layer. As the accuracy and loss become saturated, these graphs suggest that the model has been correctly trained after the ideal value of network weights has been obtained. Furthermore, the test loss looks to be virtually identical to the training loss, indicating that the model has been fine-tuned to minimize loss

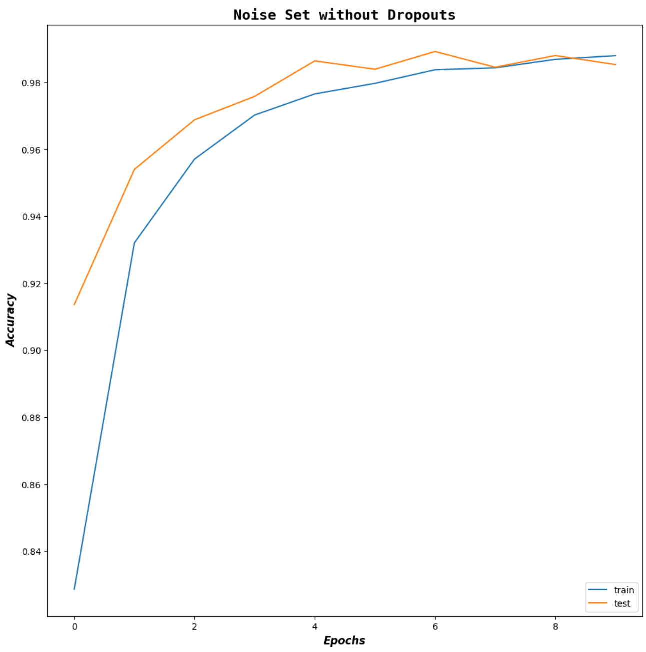

Fig.2 Accuracy graph

Model Architecture
The architecture used is a 14-layer CNN-LSTM model. The layers are as follows:

Layer (type)	Output Shape	Param #
Conv_Layer_1 (Conv1D)	(None, 187, 32)	192
Conv_Layer_2 (Conv1D)	(None, 187, 64)	10,304
Conv_Layer_3 (Conv1D)	(None, 187, 64)	20,544
Conv_Layer_4 (Conv1D)	(None, 187, 128)	41,088
Conv_Layer_5 (Conv1D)	(None, 187, 128)	82,048
Max_Pool_Layer_1 (MaxPooling1D)	(None, 94, 128)	0
Dropout_Layer_1 (Dropout)	(None, 94, 128)	0
LSTM_1 (LSTM)	(None, 94, 210)	284,760
Max_Pool_Layer_2 (MaxPooling1D)	(None, 47, 210)	0
LSTM_2 (LSTM)	(None, 47, 190)	304,760
Max_Pool_Layer_3 (MaxPooling1D)	(None, 24, 190)	0
Flatten Layer (Flatten)	(None, 4560)	0
FC_Layer_1 (Dense)	(None, 5)	22,805
FC_Layer_2 (Dense)	(None, 5)	30

Table.1 14-layer CNN-LSTM architecture

The overall best testing accuracy for 5-class classification was 99.00%, which was found by utilizing an 80:20 training-test ratio. The confusion matrix associated with the presented CNN-LSTM model is obtained and illustrated below. The confusion matrix provides a succinct summary of the categorization of the various individual classes.

Classification Report
The classification report for the model is as follows:

	precision	recall	f1-score	support
0	0.98	0.96	0.97	1821
1	0.98	0.99	0.99	2006
2	0.97	0.99	0.98	2064
3	1.00	0.98	0.99	2057
4	0.99	1.00	1.00	2052
accuracy			0.99	10000
macro avg	0.99	0.98	0.99	10000
weighted avg	0.99	0.99	0.99	10000

Fig.3 Classification Report

V. Confusion Matrix
The confusion matrix for the model is as follows:

$$\begin{bmatrix} 1756 & 34 & 17 & 4 & 10 \\ 18 & 1985 & 3 & 0 & 0 \\ 5 & 1 & 2051 & 6 & 1 \\ 3 & 0 & 43 & 2011 & 0 \\ 1 & 0 & 1 & 0 & 2050 \end{bmatrix}$$

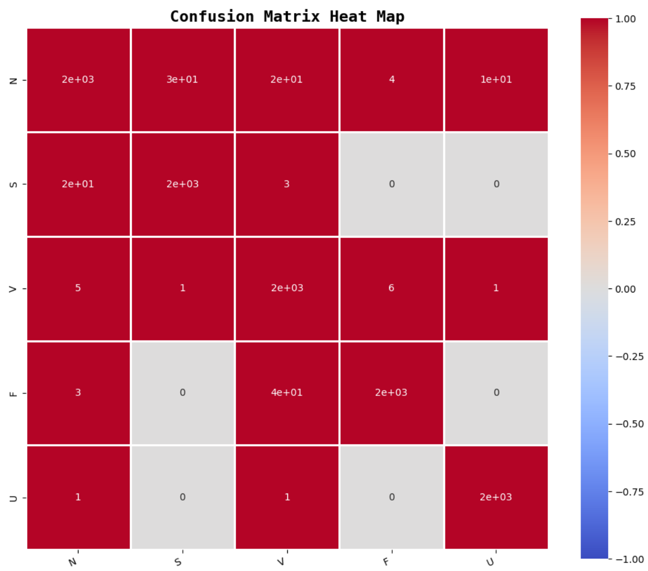

Fig.4 Confusion Matrix

We modified the SoftMax layer to support vector machine classifiers, random forest classifiers, and k-nearest neighbor classifiers to compare the performance of each technique to SoftMax. The comparison between SoftMax, SVM, random forest, and k-nearest neighbor classifiers based on overall accuracy, typical recall, typical precision, and typical F1 score is shown in the table below.

Classifier	Accuracy	Recall	Precision	F-1 Score

SoftMax	0.99	0.98	0.99	0.99
SVM	0.97	0.96	0.97	0.96
Random Forest	0.95	0.94	0.95	0.94
K – Nearest Neighbour	0.93	0.92	0.93	0.92

Table.2 Comparison of models

VI. Conclusion

This research presents a deep learning model to get better at finding cardiac arrhythmias using ECG data. We used a special technique called the db4 wavelet transform to make the ECG signals much clearer and reduce noise. This is really important for spotting tiny changes in heartbeats that we might otherwise miss. The model, which has a complex structure (14 layers of CNN, Dense, and LSTM), is designed to learn both the shape of individual heartbeats and the rhythm of the heart over time. This helps it accurately tell the difference between many kinds of irregular heartbeats, making it a reliable tool for doctors. We added extra parts to the model to make it learn more steadily and quickly and thoroughly tested the model to measure its performance, including confusion matrices, which showed that it's very accurate at identifying different types of irregular heartbeats. The findings of this study emphasize the efficacy of merging advanced signal processing methods, such as wavelet transform, with deep learning models to elevate the accuracy and dependability of arrhythmia detection. Furthermore, the model exhibits significant promise for incorporation into real-time wearable health monitoring devices, where consistent and precise arrhythmia detection could substantially influence patient care by enabling prompt interventions and mitigating the likelihood of negative health outcomes. The success of this research underscores the increasing significance of AI-powered solutions in revolutionizing healthcare technologies and enhancing patient well-being.

VII. Future Work

The present model exhibits robust capabilities in arrhythmia classification. Future research endeavors can prioritize optimizing the model for seamless integration into real-time wearable healthcare systems. This optimization can be achieved by diminishing the model's computational demands via techniques such as pruning, quantization, or the adoption of lightweight architectures, including MobileNet or TinyML frameworks. Furthermore, the integration of attention mechanisms or transformer-based modules holds the potential to deepen the model's comprehension of the temporal context inherent in electrocardiogram (ECG) signals. Personalizing models through transfer learning to accommodate the unique cardiac characteristics of individual patients may also lead to enhanced diagnostic accuracy. Future studies could also look at using more than just ECG data, for example, combining it with signals like PPG or blood pressure might help the model better detect complex heart rhythm problems and also training the model with information from a wider range of people and less common heart issues could make it work better for everyone in a real clinical setting..

References

[1] A. Shukla et al., "Detection of Arrhythmia Heartbeats from ECG Signal Using Wavelet Transform-Based CNN Model," Int. J. Comput. Intell. Syst., May 2023.

[2] R. Gupta and S. Mitra, "ECG Signal Analysis and Arrhythmia Detection using Wavelet Transform," J. Inst. Eng. India Ser. B, Dec. 2016.

[3] H. M. Nguyen et al., "ECG Signal Classification to Detect Heart Arrhythmia Using Convolutional Neural Networks," Multimedia Tools Appl., Nov. 2022.

[4] M. Talo et al., "Arrhythmia Detection in Single-Lead Heartbeat Using ECG and Residual CNN," Artif. Intell. Soft Comput., Dec. 2023.

[5] F. Murat et al., "CNN-Based Cardiac Arrhythmia Classification from ECG Signals," Adv. Intell. Syst. Comput., 2024.

[6] S. Sharma and P. Gupta, "An ECG Based CNN Model for Detection of Different Classes of Arrhythmia," SN Comput. Sci., Jun. 2023.

[7] Y. Liu et al., "A CNN Model for Cardiac Arrhythmias Classification Based on Wearable ECG Data," Cardiovasc. Eng. Technol., Dec. 2021.

[8] A. K. Barros et al., "Heartbeat Detector from ECG and PPG Signals Based on Wavelet Transform and Envelopes," Australas. Phys. Eng. Sci. Med., Jan. 2023.

[9] M. A. Al Rahhal et al., "Cardiac Arrhythmias Detection Using a Hybrid 2D CNN-SVM Model," Artif. Intell. Soft Comput., Dec. 2023.

[10] Y. Zhang et al., "Arrhythmia Detection Using TQWT, CEEMD and Deep CNN-LSTM Network," Multimedia Tools Appl., Nov. 2022.

[11] S. S. Mehta et al., "Automatic ECG Classification Using Discrete Wavelet Transform and CNN," Computing, Feb. 2023.

[12] J. Park et al., "ECG-Based Heartbeat Classification for Arrhythmia Detection Using Continuous Wavelet Transform and ANN," SpringerLink.

[13] M. Acharya, U. Rajendra, H. Fujita, S. L. Oh, and Y. Hagiwara, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Med.*, vol. 100, pp. 270–278, Feb. 2018.

[14] X. Zhang, Y. Xu, and W. Li, "A novel hybrid deep learning model for arrhythmia detection using multimodal ECG data," *Biomed. Signal Process. Control*, vol. 68, p. 102767, Aug. 2021.

[15] K. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.

[16] H. Yildirim, A. Talo, and U. R. Acharya, "A new 1D-CNN with deep learning architecture for robust classification of arrhythmia," *Expert Syst. Appl.*, vol. 145, p. 113122, Jun. 2020.

[17] L. Yao, C. Zhu, and X. Wang, "Multiscale CNN-based deep learning for ECG heartbeat classification," *J. Healthc. Eng.*, vol. 2021, pp. 1–10, Jan. 2021.

[18] P. D. Pham and Y. H. Kim, "Automatic arrhythmia detection based on ensemble learning using hybrid CNN-LSTM architecture," *IEEE Access*, vol. 8, pp. 127778–127785, Jul. 2020.

[19] J. Hannun et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nat. Med.*, vol. 25, pp. 65–69, Jan. 2019.

[20] M. R. Asl, A. Setarehdan, and M. Mohebbi, "Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal," *Artif. Intell. Med.*, vol. 44, no. 1, pp. 51–64, Sep. 2008.

[21] A. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," *Comput. Biol. Med.*, vol. 96, pp. 189–202, May 2018.

[22] M. R. Alfaras, J. Vallejo, and S. Larriba-Pey, "ECG arrhythmia classification using convolutional recurrent neural networks," *Comput. Biol. Med.*, vol. 103, pp. 1–10, May 2018.

[23] S. Hong, M. Zhou, Q. He, Z. Zhang, and Z. Wang, "Opportunistic and context-aware affect sensing on smartphones: A survey," *IEEE Access*, vol. 7, pp. 46896–46933, Apr. 2019.

[24] T. Pławiak, and A. Acharya, "Novel deep genetic ensemble of classifiers for arrhythmia detection using ECG signals," *Neural Comput. Appl.*, vol. 32, pp. 11137–11161, May 2020.

[25] B. Rajpurkar et al., "Deep learning for ECG classification: CardioNet – A new hybrid CNN-RNN architecture," *J. Electrocardiol.*, vol. 59, pp. S64–S70, Sep. 2020.