# Multi-module Recurrent Convolutional Neural Network with Transformer Encoder for ECG Arrhythmia Classification


1st Minh Duc Le
*Department of CSCE*
*University of Arkansas*
Fayetteville, AR, USA
le.m.duc86@gmail.com

2nd Vidhiwar Singh Rathour
*Department of CSCE*
*University of Arkansas*
Fayetteville, AR, USA
vsrathou@uark.edu

3rd Quang Sang Truong
*Department of CSCE*
*University of Arkansas*
Fayetteville, AR, USA
tqsang97@gmail.com

4th Quan Mai
*Department of CSCE*
*University of Arkansas*
Fayetteville, AR, USA
quanmai@uark.edu

5th Patel Brijesh
*School of Medicine*
*West Virginia University*
Morgantown, WV, USA
brijesh.patel@wvumedicine.org

6th Ngan Le
*Department of CSCE*
*University of Arkansas*
Fayetteville, AR, USA
thile@uark.edu



***Abstract*—The automatic classification of electrocardiogram (ECG) signals has played an important role in cardiovascular diseases diagnosis and prediction. Deep neural networks (DNNs), particularly Convolutional Neural Networks (CNNs), have excelled in a variety of intelligent tasks including biomedical and health informatics. Most the existing approaches either partition the ECG time series into a set of segments and apply 1D-CNNs or divide the ECG signal into a set of spectrogram images and apply 2D-CNNs. These studies, however, suffer from the limitation that temporal dependencies between 1D segments or 2D spectrograms are not considered during network construction. Furthermore, meta-data including gender and age has not been well studied in these researches. To address those limitations, we propose a multi-module Recurrent Convolutional Neural Networks (RCNNs) consisting of both CNNs to learn spatial representation and Recurrent Neural Networks (RNNs) to model the temporal relationship. Our multi-module RCNNs architecture is designed as an end-to-end deep framework with four modules: (i) time-series module by 1D RCNNs which extracts spatio-temporal information of ECG time series; (ii) spectrogram module by 2D RCNNs which learns visual-temporal representation of ECG spectrogram ; (iii) metadata module which vectorizes age and gender information; (iv) fusion module which semantically fuses the information from three above modules by a transformer encoder. Ten-fold cross validation was used to evaluate the approach on the MIT-BIH arrhythmia database (MIT-BIH) under different network configurations. The experimental results have proved that our proposed multi-module RCNNs with transformer encoder achieves the state-of-the-art with 99.14% $F_1$ score and 98.29% accuracy.**

***Index Terms*—ECG Classification, CNNs, RNNs, LSTM, Encoder, Transformer, MIT-BIH**


## I. Introduction

Cardiovascular diseases are the leading cause of death in the USA [7]. An electrocardiogram (ECG) records the electrical activity of the heart, thereby providing the summative evaluation of the cardiac electrical activity. It has been estimated that up to 300 million ECGs are recorded annually in Europe alone [23], these enormous amounts of ECG data highlights the importance of computer-aid interpretation. A high-accuracy computer-aid interpretation can save expert clinicians considerable time and efforts, as well as reducing the number of misdiagnoses.

Deep neural networks (DNN) [37], inspired by information processing and distributed communication nodes in biological systems, has been receiving massive interest in both academia and industry for a decade. They are computational models comprising of multiples layers, in which output of a layer is the input of the successive layer. The hierarchy of layers enables the network to learn the increasingly abstract, higher-level representations of the input data. DNNs have been showing their dominating performances in various intelligent tasks including biomedical [8] and health informatics [3] [17]. In the last decade, various DNNs-based methods have been employed in ECG-based automatic arrhythmia classification. Convolutional Neural Networks (CNNs) is the most favorable method [4] and could be categorized into 2 main groups: 1D CNNs in time series and 2D CNNs on time-frequency spectrograms. The former uses raw ECGs as the input [2] [25] [15] [16], split each ECG signal into multiple smaller segments which are then classified into labels in prediction step. The second approach focuses on frequency characteristic of the ECG signal, using its time-frequency spectrogram as the input of a 2D CNN for classification [12] [14] [36] [35]. Although the CNNs-based approaches have proven to be effective for arrhythmia classification, they suffer following limitations

*Lack of temporal relationship*: Either 1D CNN on time series or 2D CNNs on spectrogram first partitions an ECG signal into a set of 1D segments or 2D spectrograms at different time. Then, a CNN-based network is applied into each 1D segments or 2D spectrograms. There is no mechanism to model the temporal relations between these segments or spectrograms within the same sample coming from one patient.

*Meta-data is not taken into consideration*: ECG signal is presented in a high dimensional space while meta-data is given in a binary number (i.e. gender) or scale (age). Combining a high dimensional space of ECG signal (either times series or spectrogram) and very low dimensional space of meta-data is challenging. Most existing works do not take meta-data into account.

*Single module*: Most of the existing works is single module, i.e. they target at either time series with 1D CNNs frameworks or spectrogram with 2D CNNs frameworks. None of the previous works explores how to fuse multiple modules to inherit the merits from both time series and spectrogram.

To address the aforementioned limitations, we proposed a multi-module Recurrent Convolution Neural Networks (RCNNs) with transformer encoder. Our network makes use of LSTM [11] as a RNNs and contains four modules as follows. (i) time series module by a 1D RCNNs: In this module, 1D CNNs is first utilized to extract spatio-information from time series segment and LSTM then is used to model the temporal relations between 1D segments. (ii) spectrogram module by a 2D RCNNs: Given an ECG signal, spectrograms at different times are extracted by Short Time Fourier Transform (STFT). A 2D CNNs is used to learn visual representation in spatial domain and a LSTM network is applied to model the tempo-information between spectrograms within an ECG signal; (iii) meta-data module: An autoencoder to featurize/vectorize the metadata to learn semantic information from both sex and gender; (iv) fusion module: the information from three modules is then fused under a transformer encoder. The entire network is illustration in Fig. 1, each module is presented in one colored block.

## II. Related Work

Recently, researchers have made major efforts in using DL-based techniques to outperform specialist cardiologist in ECG interpretation. Various ideas have been proposed and Convolutional Neural Network (CNN) has been widely implemented in automatic arrhythmia diagnosis. Yildirim in [39] proposed a novel approach to classify 10-second ECG signal fragments involving 17 classes. Hannun [2] also proposed an end-to-end DL approach to classify 12 rhythm classes using single-lead ECG recordings. Although the work achieved good results, it raised a question if DNN would be useful in a realistic clinical setting, where 12-lead ECGs are the clinical standard. Ribeiro [25] partially addressed the question by presenting a DNN model using 12-lead ECG recordings to classify 6 types of abnormalities. Recurrent neural network (RNN) is also widely applied for arrhythmia diagnosis due to their highly dynamic behavior. Wang [33] proposed a global and updatable classification scheme named Global Recurrent Neural Network (GRNN). Zhang [40] introduced a patient-specific ECG classification using RNN to learn time correlation among ECG signal points. Long short-term memory (LSTM) and its improved version, gated recurrent unit (GRU) are among best DNN candidates in ECG classification [6], [26] and [31].

The aforementioned studies show that an end-to-end DNN can successfully learn complex representative features of ECG signals with less or without excessive dependencies on manual feature extraction. Although the end-to-end approach extracts the ”deep features” automatically along the network layers, it neglects one important feature of ECG, the frequency response. The importance of ECG frequency content was recognized from the beginning of the $20^{th}$ century [5] [34], and has been studied in various medical research nowadays, such as [28] and [29]. There are various time-frequency transformation methods used for ECG feature extraction, Short-time Fourier Transform (STFT) is extensively used to achieve ECG's spectral content. To exploit frequency characteristic of ECGs, several efforts have been made. Huang [12] used STFT-based spectrogram and 2D CNN for ECG arrhythmia classification. Each ECG signal is transformed into 2D-image of spectrogram to be subsequently fed into 2D CNN for image classification. Xia [36] [35] proposed using STFT and stationary wavelet transform (SWT) transformations to obtain two-dimensional (2-D) matrix input suitable for deep CNNs. Yildirim [38] proposed a novel wavelet sequence based on deep bidirectional LSTM network model.

The work mentioned above merely focused on ECG signal characteristics. Other important characteristics such as patients' physical state (e.g. age, gender) are not considered [4]. Macfarlane [20] showed that ECG interval measurements, including QRS duration, heart rate, QT dispersion, and selected Q-wave durations are highly influenced by patients' gender, age and race. Therefore, age and gender differences in the ECG should be incorporated into a variety of criteria for ECG interpretation [21]. In this paper, we propose an ECG arrhythmia classification method using multimodality - ECG signal, its frequency response and demographic factors (age and gender).

## III. Proposed multi-module RCNNs with Transformer Encoder

Our proposed network consisting of four modules, i.e. Time Series Module, Spectrogram Module, Metadata Module and Fusion Module is detailed as follows:

### A. Time Series Module: 1D RCNNs

This module aims to extract spatio-temporal information given an ECG time series signal. Let $\mathbf{X}$ be a recording ECG signal, $\mathbf{X}$ is partitioned into $n$ segments i.e. $\mathbf{X} = \{x_i\}_{i=1}^{i=n}$. Each segment length is set as $l$. There are two steps in this module. At the first step, the spatio-feature of each segment is extracted by 1D CNNs. We use function $\mathcal{F}$ to represent 1D CNNs, which transforms input segment $\{x_i\}_{i=1}^{i=n}$ into spatial representation vector:

$$\{f_i\}_{i=1}^{i=n} = \mathcal{F}(\{x_i\}_{i=1}^{i=n}) \tag{1}$$

In the second step, a bidirectional LSTM (BLSTM) [9] is applied to model the temporal relations between 1D segments.

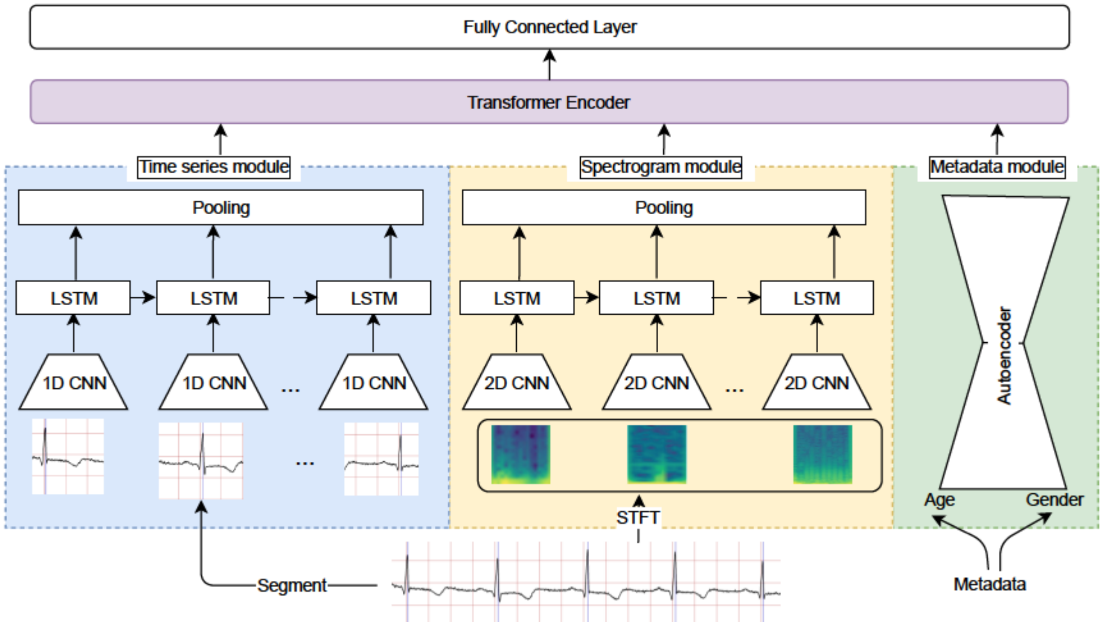


Fig. 1. Network architecture of proposed multi-module RCNNs with four modules: (i) time series module by a 1D RCNNs: 1D CNNs is first utilized to extract spatio-information from time series segments and LSTM is then used to model the temporal relations between 1D segments. (ii) spectrogram module by a 2D RCNNs: spectrograms at different times are extracted by STFT. A 2D CNNs is used to learn visual representation in spatial domain and a 2D LSTM network is applied to model the tempo-information between spectrograms; (iii) meta-data module: An autoencoder to featurize the metadata to learn semantic information from both sex and gender; (iv) fusion module: the information from three modules is then fused under a transformer encoder.

Let denote $\mathcal{R}$ as the temporal modeling function of BLSTM and $h_i$ is the BLSTM's output.

$$\{h_i\}_{i=1}^{i=n} = \mathcal{R}(\{f_i\}_{i=1}^{i=n}) \tag{2}$$

where $f_i \in \mathbf{R}^L$. In our experiments, we set $l = 360$, $n = 10$, Resnet-20 [10] is chosen as backbone network for 1D CNNs. The output of this module $\{h_i\}_{i=1}^{i=n}$ is then passed to transformer encoder module.

### B. Spectrogram Module: 2D RCNNs

This module uses 2D time-frequency responses of ECG as an input. Short Time Fourier Transform (STFT) is utilized to extract time-frequency responses of ECG. The method involves sliding a small window over the signal and then performing discrete Fourier transform for each corresponding window. The equation for STFT is shown in Eq:3 where $\mathcal{S}$ is STFT function and $g(n-m)$ is the window function. Usually a Hann or a Gaussian window is used and the width of window is specified by $m$.

$$\begin{gathered}\{s_i\}_{i=1}^{n} = \mathcal{S}(\{x_i\}_{i=1}^{n}) \\ \mathcal{S}(\{x_i\})(k,m) = \sum_{n=0}^{N-1} x(n)g(n-m)e^{-\frac{j2\pi kn}{N}}\end{gathered} \tag{3}$$

where $k$ and $m$ are known as time index and frequency index. The time-frequency responses $\{s_i\}_{i=1}^{i=n}$ is passed through a 2D RCNNs network. Similar to 1D CNNs, 2D RCNNs is designed with a 2D CNNs followed by a BLSTM network. Let $\{h'_i\}_{i=1}^{i=n} \in \mathbf{R}^L$ be denoted as the output from 2D RCNNs module and it is passed to transformer encoder module.

### C. Metadata Module: Autoencoder

In addition to ECG signal, metadata is studied in our network. Different from ECG signal which is presented in a long time series, metadata is presented by two scales corresponding to gender ($g$) and age $a$. In order to featurize metadata, we first utilize word2vec technique to convert metadata into vectors. We use function $\mathcal{W}$ to present word2vec which transforms an input $x$ into a vector

$$f_a^w = \mathcal{W}(a) \text{ and } f_g^w = \mathcal{W}(g) \tag{4}$$

In order to extract semantic information from metadata, we also apply an autoencoder $\mathcal{A}$ into concatenated vector $[f_a^w f_g^w]$. The output $d \in \mathbf{R}^L$ from the metadata module is: $d = \mathcal{A}([f_a^w f_g^w])$

### D. Fusion Module: Transformer Encoder

In our proposed multi-module, the outputs $\{h_i\}_{i=1}^{i=n}$ from time series module, $\{h'_i\}_{i=1}^{i=n}$ from spectrogram module and $d$ from metadata module are then fused by a transformer encoder. Transformer Encoder [32] is employed to re-weight these features by a proper ratio. This helps the overall model to know which information should be more emphasized to better semantically fuse these features. The final feature $f$ is generally computed as follows

$$f = w_0 d + \{w_i\}\{h_i\}_{i=1}^{i=n} + +\{w'_i\}\{h'_i\}_{i=1}^{i=n} \tag{5}$$

where $w_0$, $\{w_i\}$ and $\{w'_i\}$ are learnt by Transformer Encoder [32].

Finally, we employ a fully connected layer with softmax to convert the outputs $f \in \mathbf{R}^L$ into into categorical probabilities

TABLE I
COMPARISON BETWEEN CNNS AND RCNNS ON MIT-BIH

| Features | | F1_score | Accuracy |
|---|---|---|---|
| Time series | CNNs | 98.76 | 97.86 |
| | **RCNNs** | 98.71 | 98.09 |
| Spectrogram | CNNs | 98.67 | 97.07 |
| | **RCNNs** | 98.66 | 97.22 |

TABLE II
PERFORMANCE OF DIFFERENT MODULES ON MIT-BIH

| Feature | F1-score | Accuracy |
|---|---|---|
| Time Series | 98.71 | 98.09 |
| Spectrogram | 98.66 | 97.22 |
| Time Series - Metadata | 98.84 | 97.91 |
| **Time Series-Spectrogram-Metadata** | **99.14** | **98.29** |

$K$ classes. Let $\theta$ represents the linear mapping to $\mathbf{R}^K$ of the fully connected layer and $\phi$ denotes the softmax function performed on the classes.

$$\phi(\tilde{f}) = \frac{e^{\tilde{f}_i}}{\sum_{k=1}^{K} e^{\tilde{f}_k}}, \text{ where } \tilde{f} = \theta(f) \text{ and } \tilde{f} \in \mathbf{R}^K \quad (6)$$

## IV. EXPERIMENTAL RESULTS

**Datasets** MIT-BIH Arrhythmia dataset consists of 48 thirty minutes long two-lead ECG recordings of 47 subjects. The recordings are digitized using a sampling frequency of 360Hz. The database consists of a total 20 labels. In our experiments, we follow similar experiment setup in [12], i.e. we choose five most common labels i.e. Normal beat (N), Left bundle branch block beat (L), Right bundle branch block beat (R), Premature ventricular contraction (V), Atrial premature beat (A) and Others ( ) as all the other beats. This dataset has 2 leads and lead V5 is used. The split of training:validation is 90:10 and label that occurs most was used as the sample class.
**Metrics** $F_1$*-score* is computed as the harmonic mean of the precision and recall:

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (7)$$

*Accuracy* is the measure of how well the model could perform classification. It is the fraction of correct predictions among the total number of predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

### A. Performance Comparison

In this section, we first examine the effectiveness of RCNNs compared to CNNs as show in Table I. In this table, both time series and spectrogram features conducted on CNNs and RCNNs are investigated. The Table I demonstrated both CNNs and RCNNs obtain quite competitive F1-score while RCNNs outperforms CNNs in accuracy. Take time series as an instance; CNNs perform better than RCNNs in F1-score by a small margin of 0.05% but worse in accuracy by 0.23%.

We then evaluate the effectiveness of our proposed multi-module as shown in Table II. In this experiment, we conduct the performance on each individual module i.e. time series, spectrogram, time series & metadata and the combination of time series & spectrogram & metadata. While time series and spectrogram provide competitive performance, metadata (age and gender) is proven to play importance role to improve the classification performance. Combination of three modules i.e. time series, spectrogram, metadata provides the best performance at both F1-score and accuracy.

TABLE III
PERFORMANCE COMPARISON BETWEEN OUR PROPOSED MULTI-MODULE RCNNS WITH TRANSFORMER ENCODER AND OTHER SOTA APPROACHES ON MIT-BIH DATASET. ACC. IS FOR ACCURACY

| | Module | Framework | Classes | $F_1$score ↑ | Acc. ↑ |
|---|---|---|---|---|---|
| [22] | Time series | SVM | 6 | – | 91.67 |
| [14] | ECG images | 2D CNNs | 8 | 97.00 | 98.81 |
| [13] | ECG images | 2D CNNs | 5 | – | 97.42 |
| [24] | Time series | 1D CNNs | 7 | – | 93.60 |
| [39] | Time series | 1D CNNs | 17 | – | 91.33 |
| [18] | Time series | 1D CNNs | 4 | – | 97.50 |
| [1] | Time series | 1D CNNs | 5 | – | 94.03 |
| [27] | Time series | RNNs | 2 | 94.10 | 95.00 |
| [33] | Time series | RNNs | 2 | – | 97.40 |
| [6] | Time series | RNNs | 1 | – | 98.51 |
| [12] | Spectrogram | 2D CNNs | 5 | – | 99.00 |
| [19] | Spectrogram | 2D CNNs | – | 84.94 | 97.96 |
| [30] | Spectrogram | 2D CNNs | 8 | 98.00 | 98.92 |
| **Ours** | **multi-module** | RCNNs | 6 | **99.14** | **98.29** |

The performance comparison with other SOTA approaches on MIT-BIH dataset is given in Table III. In this Table, F1-score and accuracy are obtained by various modules (e.g. time series, spectrogram, 2D ECG images) with different machine learning techniques e.g Support Vector Machine (SVM), 1D CNNs, 2D CNNs, RNNs. In this experiment, different methods are conducted on different number of classes while our approach is conducted on the most common classes i.e. five most common classes and 1 class other for all the other 15 labels.

## CONCLUSION

In this paper, we proposed an ECG arrhythmia classification method based multi-module Recurrent Convolutional Neural Networks (RCNNs). The experiment has been conducted on six classes (five most common classes and the other classes) from MIT-BIH arrhythmia database. Our network takes all time-series, spectrogram and metadata into consideration. The proposed multi-module RCNNs is able to model both spatial information through CNNs and temporal information through LSTM. Our experiments have shown that metadata plays an important role to improve the classification performance. Our multi-module network outperforms most SOTA approach on the same dataset, with F1-score = 99.14%, and accuracy = 98.29%.

## ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Award No OIA-1946391 and NSF 1920920.

# References


[1] U Rajendra Acharya, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, Muhammad Adam, Arkadiusz Gertych, and Ru San Tan. A deep convolutional neural network model to classify heartbeats. *Computers in biology and medicine*, 89:389–396, 2017.

[2] Hannun A.Y., Rajpurkar P., and Haghpanahi M. et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med*, 25:65–69, 2019.

[3] Sujata Dash, Biswa Ranjan Acharya, Mamta Mittal, Ajith Abraham, and Arpad G Kelemen. *Deep learning techniques for biomedical and health informatics*. Springer, 2020.

[4] Zahra Ebrahimi, Mohammad Loni, Masoud Daneshtalab, and Arash Gharehbaghi. A review on deep learning methods for ecg arrhythmia classification. *Expert Systems with Applications: X*, 2020.

[5] Wilhelm Einthoven. The different forms of the human electrocardiogram and their signification. *The Lancet*, 179(4622):853–861, 1912.

[6] Oliver Faust, Alex Shenfield, Murtadha Kareem, Tan Ru San, Hamido Fujita, and U Rajendra Acharya. Automated detection of atrial fibrillation using long short-term memory network with rr interval signals. *Computers in biology and medicine*, 102:327–335, 2018.

[7] Centers for Disease Control and Prevention. Underlying cause of death, 1999–2018, 2018. CDC WONDER Online Database. Atlanta, GA: Centers for Disease Control and Prevention.

[8] Valeriy Gavrishchaka, Olga Senyukova, and Mark Koepke. Synergy of physics-based reasoning and machine learning in biomedical applications: towards unlimited deep learning with limited data. *Advances in Physics: X*, 4(1):1582361, 2019.

[9] Alex Graves and Jürgen Schmidhuber. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural networks*, 18(5-6):602–610, 2005.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[11] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[12] J. Huang, B. Chen, B. Yao, and W. He. Ecg arrhythmia classification using stft-based spectrogram and convolutional neural network. *IEEE Access*, 7:92871–92880, 2019.

[13] Elif Izci, Mehmet Akif Ozdemir, Murside Degirmenci, and Aydin Akan. Cardiac arrhythmia detection from 2d ecg images by using deep learning technique. In *2019 Medical Technologies Congress (TIPTEKNO)*, pages 1–4. IEEE, 2019.

[14] Tae Joon Jun, Hoang Minh Nguyen, Daeyoun Kang, Dohyeun Kim, Daeyoung Kim, and Young-Hak Kim. Ecg arrhythmia classification using a 2-d convolutional neural network. *arXiv preprint arXiv:1804.06812*, 2018.

[15] S. Kiranyaz, T. Ince, and M. Gabbouj. Real-time patient-specific ecg classification by 1-d convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 63(3):664–675, 2016.

[16] Serkan Kiranyaz, Turker Ince, Ridha Hamila, and Moncef Gabbouj. Convolutional neural networks for patient-specific ecg classification. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2608–2611. IEEE, 2015.

[17] Gloria Hyunjung Kwak and Pan Hui. Deephealth: Review and challenges of artificial intelligence in health informatics. *arXiv preprint arXiv:1909.00384*, 2019.

[18] Dan Li, Jianxin Zhang, Qiang Zhang, and Xiaopeng Wei. Classification of ecg signals based on 1d convolution neural network. In *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pages 1–6. IEEE, 2017.

[19] Qichen Li, Chengyu Liu, Qiao Li, Supreeth P Shashikumar, Shamim Nemati, Zichao Shen, and Gari D Clifford. Ventricular ectopic beat detection using a wavelet transform and a convolutional neural network. *Physiological measurement*, 40(5):055002, 2019.

[20] PW Macfarlane, SC McLaughlin, B Devine, and TF Yang. Effects of age, sex, and race on ecg interval measurements. *Journal of electrocardiology*, 27:14–19, 1994.

[21] Peter W Macfarlane. The influence of age and sex on the electrocardiogram. In *Sex-Specific Analysis of Cardiovascular Function*, pages 93–106. Springer, 2018.

[22] Farid Melgani and Yakoub Bazi. Classification of electrocardiogram signals with support vector machines and particle swarm optimization. *IEEE transactions on information technology in biomedicine*, 12(5):667–677, 2008.

[23] World Health Organisation. Uses of the electrocardiogram, 1981. Regional Office for Europe. Report on a WHO study; Copenhagen EURO Reports and Studies 37 (project ICP/ATH 003).

[24] A Rajkumar, M Ganesan, and R Lavanya. Arrhythmia classification on ecg using deep learning. In *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, pages 365–369. IEEE, 2019.

[25] Antônio H. Ribeiro, Manoel Horta Ribeiro, Gabriela M. M. Paixão, Derick M. Oliveira, Paulo R. Gomes, Jéssica A. Canazart, Milton P. S. Ferreira, Carl R. Andersson, Peter W. Macfarlane, Wagner Meira Jr., Thomas B. Schön, and Antonio Luiz P. Ribeiro. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature Communications*, 11(1):1760, 2020.

[26] Shraddha Singh, Saroj Kumar Pandey, Urja Pawar, and Rekh Ram Janghel. Classification of ecg arrhythmia using recurrent neural networks. *Procedia Computer Science*, 132:1290 – 1297, 2018. International Conference on Computational Intelligence and Data Science.

[27] VG Sujadevi, KP Soman, and R Vinayakumar. Real-time detection of atrial fibrillation from short time single lead ecg traces using recurrent neural networks. In *The International Symposium on Intelligent Systems Technologies and Applications*, pages 212–221. Springer, 2017.

[28] Larisa G Tereshchenko and Mark E Josephson. Frequency content and characteristics of ventricular conduction. *Journal of electrocardiology*, 48(6):933–937, 2015.

[29] Larisa G Tereshchenko, Jonathan W Waks, Muammar Kabir, Elyar Ghafoori, Alexei Shvilkin, and Mark E Josephson. Analysis of speed, curvature, planarity and frequency characteristics of heart vector movement to evaluate the electrophysiological substrate associated with ventricular tachycardia. *Computers in biology and medicine*, 65:150–160, 2015.

[30] Amin Ullah, Syed Muhammad Anwar, Muhammad Bilal, and Raja Majid Mehmood. Classification of arrhythmia by using deep learning with 2-d ecg spectral image representation. *Remote Sensing*, 12(10):1685, 2020.

[31] Sujadevi V G, Soman Kp, and Vinayakumar R. Real-time detection of atrial fibrillation from short time single lead ecg traces using recurrent neural networks. pages 212–221, 01 2018.

[32] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS*, pages 5998–6008. 2017.

[33] Guijin Wang, Chenshuang Zhang, Yongpan Liu, Huazhong Yang, Dapeng Fu, Haiqing Wang, and Ping Zhang. A global and updatable ecg beat classification system based on recurrent neural networks and active learning. *Information Sciences*, 501:523–542, 2019.

[34] Carl Wiggers. Principles and practice of electrocardiography. *Southern Medical Journal*, 22(10), 1929.

[35] Yong Xia, Naren Wulan, Kuanquan Wang, and Henggui Zhang. Atrial fibrillation detection using stationary wavelet transform and deep learning. In *2017 Computing in Cardiology (CinC)*, pages 1–4. IEEE, 2017.

[36] Y. Xia, N. Wulan, K. Wang, and H. Zhang. Detecting atrial fibrillation by deep convolutional neural networks. *Computers in biology and medicine*, 93:84–92, 2018.

[37] LeCun Y., Bengio Y., and Hinton G. Deep learning. *Nature*, 521:436–444, 2015.

[38] Özal Yıldırım. A novel wavelet sequence based on deep bidirectional lstm network model for ecg signal classification. *Computers in Biology and Medicine*, 96, 03 2018.

[39] Özal Yıldırım, Paweł Pławiak, Ru-San Tan, and U Rajendra Acharya. Arrhythmia detection using deep convolutional neural network with long duration ecg signals. *Computers in biology and medicine*, 102:411–420, 2018.

[40] C. Zhang, G. Wang, J. Zhao, P. Gao, J. Lin, and H. Yang. Patient-specific ecg classification based on recurrent neural networks and clustering technique. In *2017 13th IASTED International Conference on Biomedical Engineering (BioMed)*, pages 63–67, 2017.