

A review on Machine, Transfer and Deep learning approaches for ECG classification

Mohammed A. Atiea^{1,*}, Hosam E. Refaat²

¹ Computer Science Department, Faculty of Computers and Information, Suez University, Suez, Egypt

² Department of Information System, Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt.

ARTICLE INFO

Article history:

Received 17 November 2022

Received in revised form 1 December 2022

Accepted 2 December 2022

Available online 3 December 2022

Keywords

Electrocardiogram Classification,
Support Vector Machines,
Convolutional Neural Network,
Recurrent Neural Network

ABSTRACT

Cardiovascular Diseases (CVDs) diagnosis requires an expert interpretation of ECG (Electrocardiogram). The ECG is an essential tool that is used to diagnose CVDs for medical treatment to take place. The ECG represents the electrical events of the cardiac cycle which coordinates the contraction and relaxation of the heart chambers to circulate oxygenated and deoxygenated blood. Automation of ECG classification is considered recently to accelerate the diagnoses process and enable continuous monitoring to detect abnormalities in heart functions. ECG classification problem comes with some challenges that need to be considered such as noise, feature extraction, segmentation, and classification. This review article discusses various techniques of classification in a machine, deep, and transfer learning context as well as it considers various denoising methods to enhance the performance of different classifiers. These different classifiers are trained and tested by various and different data sets which may affect their performance as well as the number of classification classes.

1. Introduction

CVDs shape an immense segment of death causes globally. Due to World Health Organization (WHO), in 2019, there are about 17.9 million people passed away because of CVDs; this presents 32% of all death cases world-wide [1].

The electrical events of the cardiac cycle are modeled as a voltage function versus time where the sources of the electrical potentials are the polarization and depolarization processes of contractile cardiac muscle cells (cardiomyocytes).

The typical ECG type that is used most frequently in clinical procedures is the 12-lead surface ECG. To acquire a 12-lead-surface ECG signal, you interface human skin via 10 contact electrodes where six precordial electrodes are positioned on the human chest and four limb electrodes are contacted via human limbs as illustrated in Figure 1.

These ten contact electrodes give different points of view to the heart electrical activity which is the 12-lead signal. The 12-lead signal gives a complete view of heart structure and function that helps detect abnormalities and diagnose different diseases efficiently and accurately. It also gives a precise description for each part of the heart about its function performance which is required to do the correct recognition and classification [2].

ECG signal is constructed from three main components which are the P wave, QRS complex, and T wave as illustrated in Figure 2. These three components are the main parameters that are encountered in any diagnosis or disease detection to perform this diagnosis accurately.

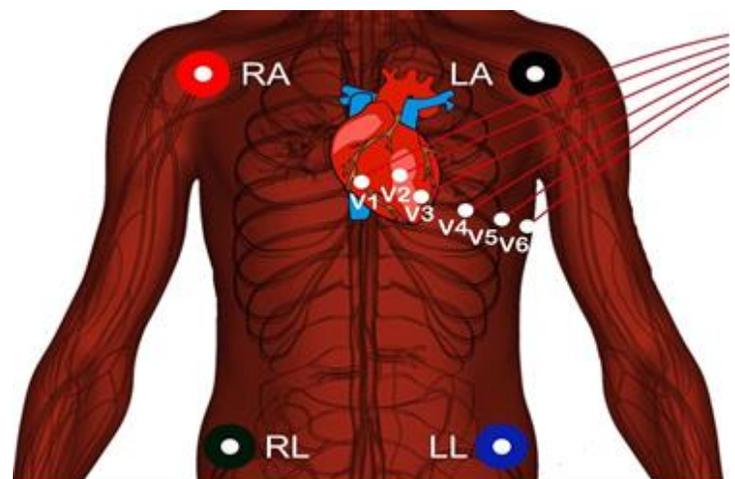


Fig.1: ECG electrodes positions to acquire 12-lead signal

* Corresponding author at Suez University

E-mail addresses: mohammed.a.atiea@fci.suezuni.edu.eg
(Mohammed A. Atiea)

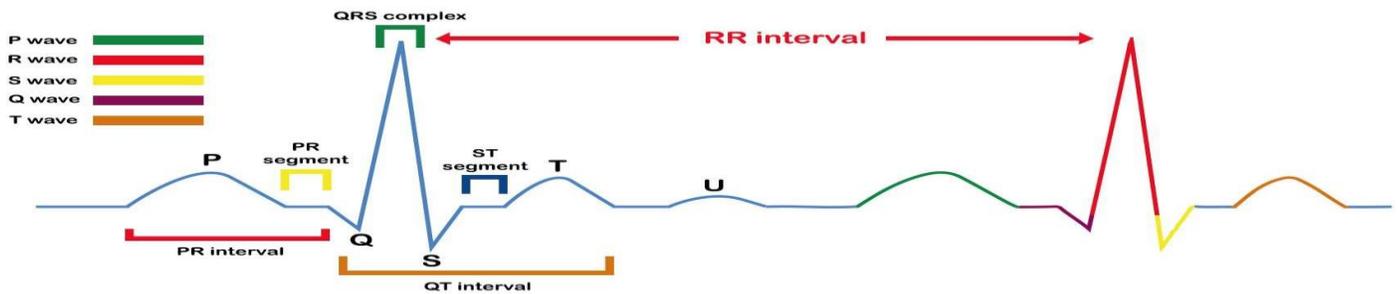


Fig. 2: An ECG signal sinus rhythm

ECG needs a professional doctor to conduct the ECG acquisition process correctly and to investigate the ECG to give the correct diagnosis.

The mentioned problem had exposed the need for automatic classification of ECG signals to facilitate and accelerate the diagnosis process correctly and efficiently.

The classification process of ECG signals can be generally splitted into two phases, the first phase is data preparation and the second is the classification process itself [2].

In the data preparation phase, data must be converted from its raw format into convenient representation to unleash the full potential of the classification algorithm and acquire the desired performance.

In the classification phase, a learning algorithm must be selected as its classification mechanism matches with the nature of the problem, or ECG raw data can be transformed into appropriate representation that can be beneficial to the classifier. This algorithm is fed by training data to predict the disease class of any future signal after its validation by the testing data [3].

There are various techniques to solve ECG classification problems in both data preparation and classification phases. In the data preparation phase, algorithms can be used for denoising signals for low- and high-frequency noise that is represented in the signal such as generative adversarial networks or different function domain transformation techniques where noise components are demolished or can be ignored such as different families of Fourier and Wavelet transforms [2, 3]. Also, there are algorithms to clean the data and reduce their dimensions to expose the principal components for the classification stage, some of these algorithms are principal component analysis (PCA), independent component analysis (ICA), and linear discriminant analysis (LDA).

In the classification phase, various techniques of machine, transfer, and deep learning can be utilized such as support vector machines, random forests, multiple layer perceptron, and different architectures of neural networks.

The organization of this article proceeds as follows, classical machine learning algorithms are discussed first followed by artificial neural networks and some of its layouts then introducing generative adversarial networks which can be used for ECG cleaning and its integration

with convolutional neural networks and the last discussed algorithm is the recurrent neural networks and some of its layouts.

2. Machine Learning Algorithms

2.1 Random Forests

Random Forests (RF) are supervised machine learning algorithms that were introduced in [4] to solve the limitation of classical decision trees in dealing with new data items since the structure of classical decision trees is fixed. RF comes with the flexibility in dealing with new data to overcome this static structure.

2.1.1 Integrating RF with multiple features fusion

In the following section, ECG classification is introduced via RF through four stages which are signal pre-processing, feature extraction and selection, and classification.

The pre-processing stage deploy a discrete Wavelet transform (DWT) was used to remove low- and high-frequencies noise from the raw ECG signal. A Haar wavelet was used to decompose the ECG waveforms in the time-frequency domain. The ECG signal is splitted into eight levels. The detail coefficient band of the first level did not contain the main components of the ECG signal and the approximate coefficient band of the eighth level mainly contained baseline drift noise. Therefore, the first-level detail coefficients and the eighth-level approximation coefficients were all set to zero under the wavelet transform then the filtered ECG was restored by the Wavelet reconstruction method, which can effectively eliminate low- and high-frequency noise.

The heartbeat was segmented in the following approach: 100 sampling stamps were located before the peak point of the R wave and 150 sampling stamp after the peak point of the R wave, totaling 250 sampling points, that is, 0.69 s for each beat cycle.

Feature fusion concept is used to enhance classifier performance, the fusion process was done on medical features, statistical features, and morphological features as follows:

RR interval feature: The RR interval used in this study was the two RR intervals adjacent to the R wave peak point.

Standard deviation (Std): For different types of heart beats, the formula for calculating standard deviation is given by:

Standard deviation (Std): For different types of heart beats, the formula for calculating standard deviation is given by:

$$D(X) = \sigma = \sqrt{\frac{\sum_{n=1}^N [X(n) - \bar{X}]^2}{N}} \quad (1)$$

where $X(n)$ is the ECG signal and N is the signal length.

Kurtosis coefficient (KU) and skewness coefficient (SK): The KU reflects the flatness of the top of the ECG waveform, while the SK is a statistic describing the symmetry of ECG signal distribution. The calculation formula for the coefficients is as follows:

$$KU = \frac{\sum_{i=1}^k (X_i - \bar{X})^4 f_i}{ns^4} \quad (2)$$

$$SK = \frac{\bar{X} - M_o}{\sigma} \quad (3)$$

where M_o represents the mode and σ provides the standard deviation.

Wavelet packet coefficient: Wavelet packet decomposition can divide the frequency band of the ECG signal into several levels. For the 360 Hz sampling frequency of the MIT-BIH database [5], the ECG signal was decomposed into four levels, and the wavelet packet coefficients were taken lies in the range of 0–45 Hertz (main energy of the ECG signal) as features after decomposition.

Discrete Wavelet Transform (DWT) divides a signal into distinct sets, each of which is a time series of coefficients describing the signal's time evolution in the specified frequency band [6]. DWT is used to decompose the ECG signal to extract the main features.

Principal Component Analysis (PCA) is used to apply dimensionality reduction of the wavelet packet coefficients obtained from the previous step and select the main information as the feature. It is a statistical procedure for dimensionality reduction of multidimensional data. It keeps the important information from high dimensional data and converts it to low-dimensional data to eliminate data redundancy, reduce the number of computer operations, and increase the algorithm's performance [7].

The wavelet packet transform to ECG signal was used as a morphological feature, and the RR interval was used as a medical feature in combination with standard deviation, the skewness coefficient, and the kurtosis coefficient.

The performance of the model is obtained through 10-fold cross-validation. The model was evaluated based on accuracy (ACC), sensitivity (SEN), and specificity (SPE) which are given by the following equations:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (4)$$

$$SEN = \frac{TP}{TP + FN} \times 100 \quad (5)$$

$$SPE = \frac{TN}{TN + FP} \times 100 \quad (6)$$

where True positive TP is the positive item correctly recognized by the classifier, True negative TN is the negative item correctly recognized by a classifier, False positive FP is a negative item that is incorrectly marked as a positive item and False negative FN is a positive item that is incorrectly marked as a negative item.

There are two paradigms that can be used in testing the model which are inter-patient paradigm and intra-patient paradigm. The RF classifier achieved the best classification results with features fusion in both intra-patient and inter-patient classification modes, with an accuracy rate of 99.08% and 92.31%, respectively [8]. RF accomplishes a sensitivity of 89.31% and specificity of 99% [8]. RF has some limitations as feature extraction depends on QRS complex wave detection, if the location of the QRS complex is not accurate, it affected the recognition accuracy of the algorithm [8]. And because of the imbalanced database, the method's sensitivity for inter-patient was low.

2.2 Support Vector Machine

Support Vector Machine (SVM) [9] is considered as a supervised machine learning approach [10] that targets to construct $n-1$ hyperplane to separate between two classes of n vectors or entity features that satisfy maximum margin condition. Since ECG classification problem is multi-class classification problem, SVM must be adapted in some schemes to provide multi-class classification model [11, 12].

There are two schemes to enable SVM to handle multi-class classification that are one against all (OAA), and one against one (OAO). Considering N is the total number of target classes:

- 1) In OAA, SVM models are constructed one per class
- 2) In OAO, SVM models are constructed between each pair of classes, resulting in $(N(N-1)/2)$ models.

2.2.1 Integrating SVM with tunable Q-wavelet transform

To distinguish between different classes of ECG rhythms, features can be extracted from these signals in both time and frequency domains as well as features of the cardiac rhythm morphology.

To distinguish among different ECG rhythms, Wavelet-based features can be deployed. The solution can be interpreted as four processes: preprocessing, R-peaks localization and beat segmentation, features extraction, and classification.

MIT-BIH arrhythmia database is used in this classification. MIT-BIH contains ECG signals in its raw form which contains high-frequency noise as well as baseline wandering. There are two filters used to remove such noise

which are Savitzky–Golay (SG) filter to eliminate high-frequency noise, and Butterworth filter with the third order and cutoff frequency 1 Hz to remove baselines.

PanTompkins algorithm was utilized to detect R-peaks in raw ECG which is essential to boost classifier performance. PanTompkins algorithm had shown 99.3% QRS detection accuracy over the MIT-BIH arrhythmia database.

ECG segmentation is done once R-peaks have been detected. Each segment has T_1 and T_2 as boundaries that cover the P wave, QRS complex, and T wave of an ECG signal, T_1 is set 0.25s before the R-peak and T_2 is set 0.45s after it [10], resulting in a 70 seconds of the ECG as a total duration of each segment.

Tunable Q-Wavelet Transform (TQWT) was used to extract features. Low- and high-pass scaling are introduced to preserve low-frequency and high-frequency contents respectively [13]. Two-channel analysis filter banks are provided at each level of signal decomposition for low and high-frequency components. The scaling factors α and β handle low- and high-pass scaling, respectively.

The TQWT's filter-bank provides limited redundancy and flawless reconstruction properties by following the relations: $0 < \alpha < 1$, $0 < \beta \leq 1$ and $\alpha + \beta > 1$.

The low-pass and high-pass filters are designated by $H_0(\omega)$ and $H_1(\omega)$ respectively, where ω represents the angular frequency. $H_0(\omega)$ and $H_1(\omega)$ can be represented as a function of $\Theta(\omega)$ which represents the frequency response of the Dabuchies filter [14, 15]:

$$H_0(\omega) = \theta \left(\frac{\omega + (\beta - 1)\pi}{\alpha + \beta - 1} \right) \quad (7)$$

$$H_1(\omega) = \theta \left(\frac{\alpha\pi - \omega}{\alpha + \beta - 1} \right) \quad (8)$$

Where is $\Theta(\omega)$ represented as:

$$\theta(\omega) = \frac{1}{2} (1 + \cos(\omega)) \sqrt{2 - \cos(\omega)} \quad (9)$$

Q-factor (Q) and oversampling rate or redundancy (r) are related to scaling factors β and α respectively via the following relations:

$$\beta = \frac{2}{Q + 1} \quad (10)$$

$$\alpha = 1 - \frac{\beta}{r} \quad (11)$$

The relationship which determines the maximum number of decomposition levels J_{max} is given by:

$$J_{max} = \left\lceil \frac{\log\left(\frac{\beta N}{8}\right)}{\log\left(\frac{1}{\alpha}\right)} \right\rceil \quad (12)$$

where N is signal length $x(n)$ [14]. Unitary Discrete Fourier Transform (DFT) is complemented with TQWT to perform normalization to represent low-pass and high-pass

spectrum of the ECG signal as J^{th} decomposition level coefficients which will be fed to the classifier as entity features. To adapt SVM to multi-class classification, an OAA scheme is deployed.

This model is able to handle 8 classes where the training and testing data come from the MIT-BIH arrhythmia database and it offers average sensitivity, specificity, and accuracy of 96.22%, 98.22%, and 99.27% respectively [14].

2.2.2 Integrating ensemble SVM with multiple feature extraction

Ensemble Learning is a powerful concept to apply in machine learning classifiers since it aggregates knowledge from different classifiers to take the final decision. An ensemble-based SVM classifier was proposed to classify ECG beats into four classes using the MIT-BIH arrhythmia database.

Four types of features were extracted from raw ECG which are Wavelets, statistical features, R-R intervals, and morphological features.

There are many methods of noise removal methods that can be used with this approach [16, 17] to remove baseline and high-frequency noises. Baselines removal is considered only to preserve the raw signal for the next step of feature extraction. To compute ECG baseline, two sequential median filters of 200 ms and 600 ms are provided. Then, the obtained baseline is subtracted from the raw ECG, which produces ECG signal that is baseline wonder free.

For normalization, Z-score is applied to all items in the data set by the following equation:

$$z = \frac{(x(i) - \mu)}{\sigma} \quad (13)$$

Where (μ) represents the mean, $x(i)$ represents the heartbeat point and σ represents the standard deviation. For heartbeat segmentation, the R-peak point is localized and then the heartbeats are segmented by using a window size of 180 centered at R-peak. Each heartbeat signal includes 90 points on the right of R peaks and 90 points on the left side. In the feature extraction phase, four types of features are extracted from raw ECG signal as follows:

Wavelet transform: it is used to extract information in time and frequency domains both of them [18].

High Order Statistics (HOS): exhibit good performance to describe ECG morphology [18]. each beat is splitted into five segments, creating a 10-dimensional feature, calculating the value of skewness and kurtosis value over each segment as illustrated in equations (2, 3).

Morphological descriptor: can be represented as euclidean distance of separation between the R-peaks and four points of the ECG beat.

R-R intervals: the time between two subsequent beats is provided to establish the R-R interval.

In the **classification phase**, the OAO approach is preferred because there are many samples and it handles

the imbalanced, unorganized data and it requires less time to train OAA.

To enhance the final prediction, assembling is used to combine the decisions of the different classifiers that make up the system. Ensemble of SVMs is applied by combining several SVM models that have been trained with distinct features via employing different kinds of feature extraction methods such as R-R intervals, HOS, wavelets, and amplitude values.

Ensemble of SVMs with an overall accuracy of 94.40%, sensitivity of 65.26%, specificity of 93.25%, precision of 69.11% and F-score of 66.24% [18]. precision and F-score can be obtained from the following equations:

$$precision(P) = \frac{TP}{TP + FP} \times 100 \quad (14)$$

$$F - score = 2 \times \frac{P \times SE}{P + SE} \quad (15)$$

3. Transfer and Deep Learning Algorithms

3.1 Multiple layer Perceptron

The Multiple layer Perceptron (MLP) [19] architecture is presented in figure 3, it consists of main three modules: the input layer, the hidden layer/s, and the output layer. MLP processes information due to the following equation, for each node:

$$y_i = f \left(\sum_j w_{ji}^{(1)} f \left(\sum_k w_{kj}^{(2)} x_k \right) \right) \quad (16)$$

Here, $w_{ji}^{(1)}$ and $w_{kj}^{(2)}$ are the weights of the output and the hidden neurons, respectively, y_i represents the output of the network, and f is the used activation function [20].

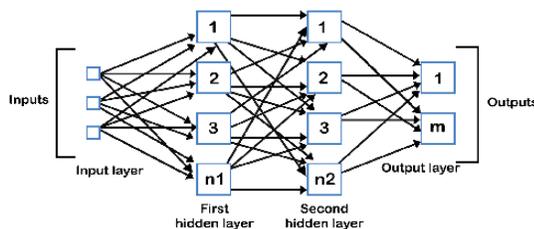


Fig. 3: MLP architecture layout

The activation function introduces the nonlinearity approach to facilitate separation of nonlinear separable classes, it can be represented by various mathematical functions such as *sigmoid* and *tanh* which can be represented by the following equations:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (17)$$

$$\text{sigmoid}(x) := \frac{1}{1 + e^{-x}} \quad (18)$$

The training process of the model can be processed in forward-propagation or back-propagation to update the

weights of the input features to the network. The back-propagation method is more preferred because it shows the effect of the updated weight on the output function, so updating process is more manageable. The effect of neuron wight update can be represented by the following equations, for output layer neurons:

$$\delta_i^{(1)} = y_i(1 - y_i)(t_i - y_i) \quad (19)$$

here, $(t_i - y_i)$ is the difference between the i^{th} output and the corresponding target value [20]. For hidden layer neurons:

$$\delta_i^{(2)} = h_j(1 - h_j) \sum_i \delta_i^{(1)} w_{ji} \quad (20)$$

Via these equations the network weights can be updated due to the following equations, output-layer neurons:

$$W_{ji}^{(1)} := W_{ji}^{(1)} + \eta \delta_i^{(1)} h_i \quad (21)$$

Hidden-layer neurons:

$$W_{kj}^{(2)} := W_{kj}^{(2)} + \eta \delta_j^{(2)} x_k \quad (22)$$

The extent of weight correction is therefore determined by $\eta \delta_i^{(1)} h_i$ or $\eta \delta_j^{(2)} x_k$, where η represents the learning rate such that $\eta \in (0, 1)$ [18].

3.1.1 Integrating MLP with DWT and Independent Components Analysis (ICA)

MLP can be integrated with DWT and ICA to classify ECG signals into three classes using the MIT-BIH arrhythmia database for setting the training and testing processes of the classifier [21].

The system consists of 2 phases, the first one is ECG Signal Processing where the filtration of the ECG signal is done by moving the average filter to eliminate the baselines. This elimination of baselines is done by Low Pass Filter which aids in smoothening the signal.

$$y(i) = \frac{1}{2N + 1} (x(i + N) + x(i + N - 1) + \dots + x(i - N)) \quad (23)$$

Where $y(i)$ denotes the smoothened value of data point in i^{th} , N denotes the total count of data points on both the sides of $y(i)$, $2N+1$ represents the span and x denotes the input vector [22].

In the post-processing phase, the features extraction of the ECG signal is being made by using two seconds of the ECG from its record.

For the dimensionality reduction part, the ICA was used to reduce the wavelet coefficients vector since the signal is decomposed into eight levels via the family of Daubechies Wavelet of 6th order [21].

Mean Square Error (MSE) was used to evaluate the performance of MLP NN [21] and it can be computed by the following equation:

$$MSE = \frac{1}{m} \sum_{i=1}^m (t_i - y_i)^2 \tag{24}$$

The best training performance is obtained at MSE of value 0.18306 with 50000 epochs of the proposed MLP neural network [21]. The performance of MLP NN is recognized by 98.6% [21].

3.2 Convolutional Neural Networks

A Convolutional Neural Network (CNN) is a special deep feedforward network, where the input of each layer is computed as a function of the output of the previous layer. The accuracy will increase by increasing the depth of it, but it will slow down the training. And after a certain depth, the accuracy may decline.

CNN consists of a series of layers such as a convolution layer, a pooling layer, a normalization layer, and a fully connected layer, where the output of any layer is the input of the next layer [23, 24].

An extra layer is added for backward error propagation that learns updated weights for the CNN. The convolution layer applies a set of weights to extract the features from the data. The pooling layer is used to apply dimensionality reduction via calculating the average or maximum convolution features [25]. Loss layer can be used to measure the discrepancy between the CNN prediction and target value.

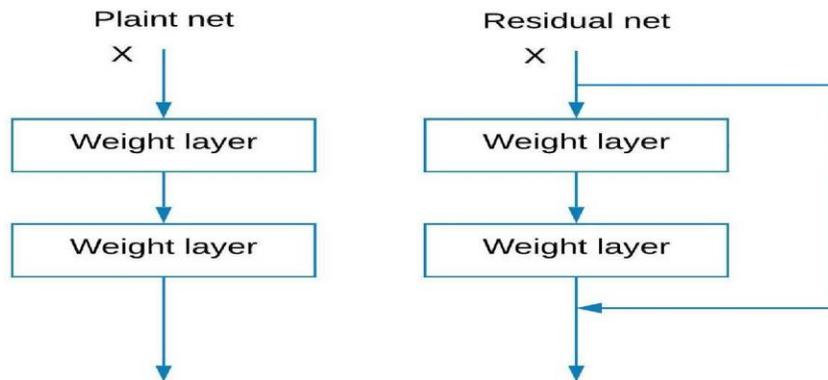


Fig.4: Residual network structure compared to the plain network

Fast Compression Residual convolutional neural Networks (FCResNet) consists of several modules, as shown in Figure 5, in fast down-sampling module mainly consist of two convolution layers where each one is preceded by a dropout layer as well as a batch normalization layer to prevent model over-fitting.

Also, there are three residual convolution modules: each consists of consecutive convolution layers as illustrated in residual network structure, followed by max-pooling layer to reduce the dimension of the feature vectors. The last module is the classification module, which is structured from one flatten layer, two fully connected layers, and a softmax classifier.

A convolution layer is added before the flatten layer to reduce the feature vector dimensions. Also, there is a random dropout layer that is used after the flattening layer

3.2.1 Integrating CNN with maximal overlap wavelet packet transform

Maximal overlap wavelet packet transform (MOWPT) was used for ECG classification. The classification of ECG arrhythmia was done due five classes by using the MIT-BIH arrhythmia database.

Pre-processing: wavelet transform was used to transform the ECG signal from the time-amplitude space to a time-frequency space.

Wavelet packet transform outperforms wavelet transform because it has a higher resolution. It handles the limitations of poor frequency resolution of the wavelet transform in the high-frequency band [26].

The MOWPT was used for preprocessing to decompose the input signal into several levels through low- and high-pass filters. It handles the high-frequency band signal and improve the frequency resolution.

Classification: Residual neural network (ResNet): as illustrated in Figure 4, the network output of the previous network layers does not go through the next layer but directly serves as an input part of the network layer behind. The input data comes from several intersections of the previous network layers and does not only depend on the output of the previous layer.

to prohibit overfitting. The model's achieves 98.79% of accuracy and its average loss value equals 0.0255.

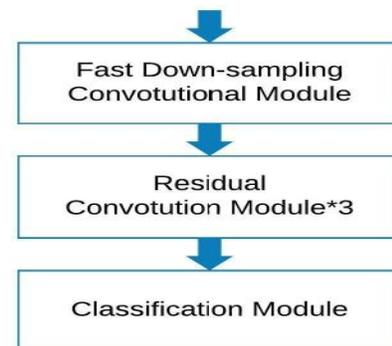


Fig.5: FCResNet Architecture

3.2.2 Integrating CNN with continuous wavelet transform

Continuous Wavelet Transform (CWT) was used to pre-process ECG signals then pass them to a CNN to classify ECG signal over five classes by using MIT-BIH arrhythmia database.

To overcome baseline wandering, two median filters were used to get baseline wandering, then remove it from the original ECG signal. Outliers neatly can also be removed by the median filter without increasing phase distortion.

Wavelet Transform (WT) [27] has highly flexible time and frequency resolutions [25], it can be computed by the following equation:

$$c_a(b) = \frac{1}{\sqrt{a}} \sum_{-\infty}^{\infty} x(t) \cdot \varphi\left(\frac{t-b}{a}\right) dt \quad (25)$$

where a is a scaling factor, b is a translation factor, and $\varphi(t)$ is the wavelet function. The Mexican hat wavelet (mexh) was used as the wavelet function, and it can be computed as

$$\varphi(t) = \frac{2}{\sqrt{3^4 \sqrt{\pi}}} (1 - t^2) e^{-\frac{t^2}{2}} \quad (26)$$

The scale can be converted to frequency by

$$F = \frac{F_c * f_s}{a} \quad (27)$$

where F_c represents the center frequency of the function, f_s is the sampling frequency $x(t)$ [28]. He initialization [29] was used to initialize the convolution layers weights as well as fully connected layers. This method achieved 70.75%, 67.47%, 68.76%, and 98.74% for positive predictive value, sensitivity, F1-score, and accuracy, respectively [25].

3.2.3 Integrating CNN with recurrence plots

Since CNN unleashes its full potential with images, it is needed to find a method to transform 1D ECG signal representation into a sort of special plot or representation which in turn exposes principal features to enhance classification efficiency and facilitates the noise analysis process. Such requirements can be provided in recurrence plotting. Recurrence plots (RPs) can be useful to analyze the bidirectional representation of m-dimensional trajectory. RP and CNN were used to classify ECG signals over six classes by using four different datasets [30]. The target classes were normal (NOR), noise, atrial fibrillation (AF), atrial premature contraction (APC), ventricular fibrillation (VF), and premature ventricular contraction (PVC). The proposed CNN operates at two phases of classification: the first phase classifies noise and VF while the second phase

classifies the rest classes of arrhythmia by using R-peak detection algorithm [31].

There was a two-second ECG signal segment fed to the model, one second before the R-peak and another second after the R-peak, then this segment is converted into a RP. Transfer learning concept can be seen clearly in CNN. The used CNNs in this classification are VGG16 and VGG19 because these two networks has more layers such as filters, padding, and max pooling over the AlexNet and it follows its architecture [30].

The rectified linear unit (ReLU) is preferred over the other activation functions due to its outstanding performance [30,32, 33]. ReLU can be computed by

$$ReLU(x) = \max(0, x) \quad (28)$$

After applying the ReLU activation function and five-fold cross-validation with the mentioned CNNs, the captured average testing accuracy was $95.3\% \pm 1.27\%$ and $98.41\% \pm 0.11\%$ in the first stage and second stage, respectively [30].

3.3 Recurrent Neural Network

Recurrent Neural Network (RNN) is a special variant of neural network architectures that are suited to deal with time series which is appropriate in the case of the ECG classification. It updates the internal state of their nodes based on the previous state of the node which is important to give decision based on the order of input data. This mechanism is applied via a recurrence relation as:

$$h_t = \sigma(W^{hx}x^{(t)} + W^{hh}h^{(t-1)} + b_n) \quad (29)$$

$$\hat{y}^{(t)} = \text{softmax}(W^{yh}h^{(t)} + b_y) \quad (30)$$

Here W^{hx} is the matrix of conventional weights between the input and the hidden layers and W^{hh} is the matrix of recurrent weights between the hidden layer and itself at adjacent time steps. The vectors b_n and b_y are bias parameters that allow each node to learn an offset [34] as introduced in Figure 6 where L_i represents the loss in each cell.

To update the initial conventional via backpropagation such as back-Propagation Through Time (BPTT) [35, 36], gradients tend to either (1) blow up or (2) vanish: the temporal evolution of the back-propagated error exponentially depends on the size of the weights [37]. Case (1) may lead to oscillating weights, while in case (2) learning to bridge long time lags take a prohibitive amount of time or do not work at all [38]. The problem introduces the need for introducing Long-Short Term Memory (LSTM) [38] to solve gradient flowing problems.

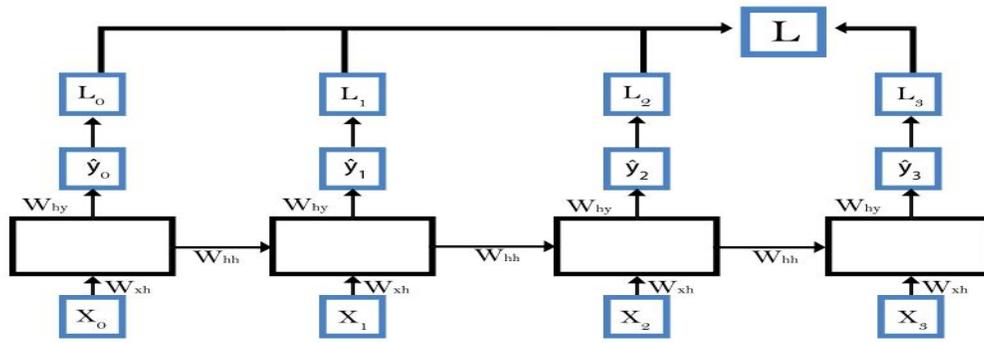


Fig.6: RNN architecture

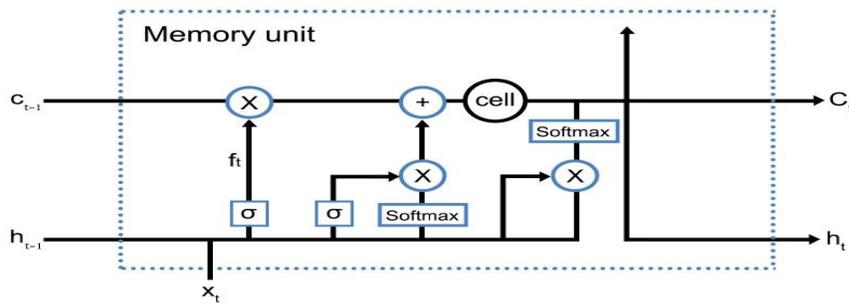


Fig.7: LSTM NN node structure

The LSTM architecture utilizes the concept of gated cells where each cell of computational block can control information flow where information is added or removed by special structures that are called gates as illustrated in Figure 7. LSTM networks operate due to four actions as follows:

1. Forgetting irrelevant data that is coming from the previous state
2. Storing new relevant data into the cell state
3. Updating the cell state values
4. Determining what information is sent to the next time step via the control of output gate

Through this mechanism, BPTT can be processed with uninterrupted gradient flow. Bidirectional Recurrent Neural Network (BRNN) permits to train of the RNN simultaneously in positive and negative time direction to boost the load of input information that is usable by the network.

The BRNN structure outperforms the other ANN structures [39] which can be applied to the LSTM RNN layout, introducing Bidirectional LSTM (BLSTM). In the following part, two methods of ECG classification by LSTM and BLSTM are reviewed.

3.3.1 Integrating time-frequency moments with LSTM

It was proven that the spectral feature extraction using logarithmic transform with LSTM provides satisfying

performance [40]. LSTM input data is fed as two spectral inputs (TF moments):

Instantaneous frequency (IF) and spectral entropy (SE). Both IF and SE were used in feature extraction which was done to boost the training and testing accuracies of the classifier.

In this solution, IF and SE are used to express the information that is contained in a 2D spectrogram of the ECG signal, since LSTM does not behave well with multidimensional data. IF can be calculated via $f_{inst}(t)$ function as:

$$f_{inst}(t) = \frac{\int_0^\infty f P(t, f) df}{\int_0^\infty P(t, f) df} \tag{31}$$

where: $P(t, f)$ represents the power spectrum estimate of an input signal, t is sample vector and f represents spectrum frequencies [41]. The SE measures the spectral power distribution of the ECG signal, and it can be formulated by the following equation:

$$H(t) = - \sum_{m=1}^N P(t, m) \log_2 P(t, m) \tag{32}$$

here $P(t, m)$ is the probability distribution of the spectrogram at time t can be computed by the following equation:

$$P(t, m) = \frac{S(t, m)}{\sum_f S(t, f)} \quad (33)$$

Here $S(t, f)$ is the instantaneous spectral entropy for power spectrogram [40]. After computing TF moments, IF and SE are fed into the LSTM model representing the information that is contained in the signal spectrogram. The model classifies six classes, five classes of cardiac dysfunctions which are AF, bradycardia, tachycardia, Ventricular Tachycardia (VTach) 160 bpm, and PVC; in addition to the NOR class.

The training data for this model is generated using FLUKE "ProSim 4 Vital Sign and ECG Simulator", the

dataset is counted 3121 signals each one in 5s captured at 5 KHz.

This model has an accuracy of 100% with respect to six classes classification and synthetic used signals but there is uncertainty about the behavior of the model at dealing with signals that are coming from ECG devices.

3.3.2 Integrating wavelet sequence with BLSTM

This solution classifies ECG signals into five classes by using the PhysioBank MIT-BIH arrhythmia database for training and testing.

Since ECG signals are represented in digital form, DWT is introduced as an approach to reduce the cost of CWT computation to capture the main components of the input signals which is a crucial part of classifier performance. DWT can be decomposed as serial digital filters representing multi-resolution analysis.

Table 1: Comparison of the state-of-the-art algorithms used for ECG classification with number of classification classes of each.

Literature	Number of classes	Feature	Classifier	ACC	SEN	SPE
P. Pławiak et al [42]	17	Power Spectral Density (PSD), Discrete Fourier Transform (DFT), Hamming Window, Genetic Algorithm	DGEC	99.37	94.62	99.66
W. Yang et al [43]	5	Linear SVM, PCANet	SVM	99.77	-	-
H. Dang et al [44]	5	Gramian Angular Field (GAF), Recurrence Plot (RP) and Markov Transition Field (MTF)	multi-scale fusion CNN	95.48	96.53	87.74
F. Li et al [45]	5	-	CNN	99.00	95.4	99
M. A. Atiea et al [23]	14	-	1D-CNN + Transformer	99.6	99.6	99.6
J. Huang et al [46]	17	Short-Time Fourier Transform	CNN	99.00	-	-
Kumar M. A. [2]	4	Fast Fourier Transform (FFT)	AlexNet	99.7	98.3	99.2
Shuai Ma et al [24]	5	-	1D-CNN + DCGAN	98.7	98.2	98.5
M. A. Ahamed et al [47]	5	-	ANN	98.06	-	-
M. A. Ahamed et al [47]	2	-	Ensemble Approach	97.664	97.062	96.904
X. Xu et al [48]	5	-	1D CNN Bi-LSTM	95.9	-	-
Cao, Minh et al [3]	4	Short-Term Fourier Transform (STFT)	2D CNN (ResNet18)	90.8	-	-
A. M. Shaker et al [49]	15	End to end	Deep CNN	98.3	-	-
A. M. Shaker et al [49]	15	End to end	Two-Stage Deep CNN	98	-	-
J.-S. Huang et al [26]	5	MOWPT	FCResNet	98.79	-	-
T. Wang et al [50]	5	CWT	CNN	98.79	-	-
Ö. Yildirim et al [41]	5	End-to-end	DBLSTM-WS3	99.39	67.47	-
G. Kłosowski et al [39]	6	TF moments: (IF) (SE)	LSTM	100	-	-
M. Ramkumar et al [21]	3	ICA, Daubechies Wavelet	MLP NN	98.6	-	-
Z. Wang et al [8]	5	TF moments: (IF) (SE)	LSTM	100	-	-
C. K. Jha et al [14]	8	TQWT	SVM	99.27	96.22	98.22
S. K. Pandey et al [18]	4	Wavelet, R-R interval, HOS, Morphological	Ensemble of SVMs	94.40	65.26	93.25

These filters consist of High-pass (HP) filters and Low-pass (LP) filters are described by the following equations:

$$H = \sum_{k=-\infty}^{\infty} S[k] \varphi_h[2n - k] \quad (34)$$

$$L = \sum_{k=-\infty}^{\infty} S[k] \varphi_g[2n - k] \quad (35)$$

where S is input signal, H and L are the outputs of filters, and φ_h and φ_g are low-pass and high-pass filters, respectively [41].

There are three levels of DWT to obtain low the lower resolution components. To combine DWT with BLSTM, new layer named Wavelet Sequence was added which is based on the Wavelet transform [41].

This method has shown 99.39% accuracy [41] for five classes classification due to the MIT-BIH arrhythmia database. Table 1 represents the achieved results of ECG classification of the state-of-art proposed solutions, and it shows that ECG classification problem has multiple parameters each one of them has a considerable effect on the overall system performance.

4. Conclusion

Classifier performance is associated with the number of classes, used datasets, different techniques in different stages of the classification system which are signal pre-processing, feature extraction, signal segmentation, and the classification process itself. The key to getting an outstanding performance of a classifier is to get a smooth harmony between all each stage of the system. It is a must to represent data in such a format that is capable to acquire the full potential of the classifier in both manners: first is exposing principal features that carry the critical key of the classification as well as representing it in a convenient way that can be accepted by the classifiers.

Acknowledgment

The Authors gratefully acknowledge the support of Mark A., Mahmoud R., Mennat-Allah M., Ahmed M. and Shimaa A. from faculty of computers and information, Suez university, Suez, Egypt. for their valuable comments and suggestions.

References

- [1] Cardiovascular diseases (cvds) (11 June 2021). URL [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)), Last Accessed Date: 10/2022.
- [2] Kumar M. A, Chakrapani A. "Classification of ECG signal using FFT based improved Alexnet classifier". PLoS ONE (2022) 17(9): e0274225. doi:10.1371/journal.pone.0274225 .
- [3] Cao, Minh, Zhao, Tianqi, Li, Yanxun, Zhang, Wenhao, Benharash, Peyman & Ramezani, Ramin. "ECG Heartbeat classification using deep transfer learning with Convolutional Neural Network and STFT technique". The 4th International Conference on Computing and Data Science (CONF-CDS 2022) doi:10.48550/arXiv.2206.14200.
- [4] Ho, T. K. "Random decision forests". Proceedings of the International Conference on Document Analysis and Recognition, ICDAR 1, 278–282 (1995). doi:10.1109/ICDAR.1995.598994 .

- [5] Mit-bih database and software catalog. URL <http://ecg.mit.edu/dbinfo.html>, Last Accessed Date: 10/2022.
- [6] Hosseinzadeh, M. "Robust control applications in biomedical engineering: Control of depth of hypnosis", Control Applications for Biomedical Engineering Systems, Pages 89-125 (2020) doi:10.1016/B978-0-12-817461-6.00004-4 .
- [7] Ethem Alpaydin, "Introduction to machine learning ", 3ed edition ISBN: 9780262028189, (August - 2014).
- [8] Wang, Z., Li, H., Han, C., Wang, S. & Shi, L. "Arrhythmia classification based on multiple features fusion and random forest using ecg". Journal of Medical Imaging and Health Informatics 9, 1645–1654 (2019). doi:10.1166/jmihi.2019.2798 .
- [9] Cortes, C. & Vapnik, V. "Support-vector networks". Machine Learning 20, 273– 297 (1995). doi:10.1007/BF00994018 .
- [10] Qin, Q., Li, J., Zhang, L., Yue, Y. & Liu, C. "Combining low-dimensional wavelet features and support vector machine for arrhythmia beat classification". Scientific Reports 7, 6067 (2017). doi:10.1038/s41598-017-06596-z .
- [11] Duan, KB., Keerthi, S.S. "Which Is the Best Multiclass SVM Method? An Empirical Study". In: Oza, N.C., Polikar, R., Kittler, J., Roli, F. (eds) Multiple Classifier Systems. MCS 2005. Lecture Notes in Computer Science, vol 3541. Springer, Berlin, Heidelberg (2005). doi:10.1007/11494683_28
- [12] Hsu, C.-W. & Lin, C.-J. "A comparison of methods for multiclass support vector machines". IEEE Transactions on Neural Networks 13, 415–425 (2002). doi:10.1109/72.991427 .
- [13] Selesnick, I. W. "Wavelet transform with tunable Q-factor". IEEE Transactions on Signal Processing 59, 3560–3575 (2011). Doi:10.1109/TSP.2011.2143711 .
- [14] Jha, C. K. & Kolekar, M. H. "Cardiac arrhythmia classification using tunable Q-wavelet transform based features and support vector machine classifier". Biomedical Signal Processing and Control 59, 101875 (2020). doi:10.1016/j.bspc.2020.101875 .
- [15] Cohen, A. "Ten lectures on wavelets, cbms-nsf regional conference series in applied mathematics, vol. 61, i. daubechies, siam, 1992, xix + 357 pp". Journal of Approximation Theory 78, 460–461 (1994). doi:10.1006/jath.1994.1093 .
- [16] Pongpon Sri, S. & Yu, X.-H. "An adaptive filtering approach for electrocardiogram (ecg) signal noise reduction using neural networks". Neurocomputing 117, 206–213 (2013). doi:10.1016/j.neucom.2013.02.010 .
- [17] Blanco-Velasco, M., Weng, B. & Barner, K. E. "Ecg signal denoising and baseline wander correction based on the empirical mode decomposition". Computers in Biology and Medicine 38, 1–13 (2008). doi:10.1016/j.combiomed.2007.06.003 .
- [18] Pandey, S. K., Janghel, R. R. & Vani, V. "Patient specific machine learning models for ecg signal classification". Procedia Computer Science 167, 2181–2190 (2020). doi:10.1016/j.procs.2020.03.269 .
- [19] Orbach, J. "Principles of neurodynamics. perceptrons and the theory of brain mechanisms". Archives of General Psychiatry 7, 218 (1962). doi:10.1001/archpsyc.1962.01720030064010 .
- [20] Kubat, M. "An Introduction to Machine Learning", Springer International Publishing, (2017) doi:10.1007/978-3-319-63913-0 .
- [21] M. Ramkumar, C. Ganesh Babu, K Vinoth Kumar, D Hepsiba, A. Manjunathan & R. Sarath Kumar. "Ecg cardiac arrhythmias classification using dwt, ica and mlp neural networks". Journal of Physics: Conference Series 1831, 012015 (2021). doi:10.1088/1742-6596/1831/1/012015 .

- [22] Ubeyli, E. D. "Wavelet/mixture of experts network structure for ecg signals classification". *Expert Systems with Applications* 34, 1954–1962 (2008). doi:10.1016/j.eswa.2007.02.006 .
- [23] M. A. Atiea & M. Adel, "Transformer-based Neural Network for Electrocardiogram Classification", *International Journal of Advanced Computer Science and Applications (IJACSA)*, 13(11), (2022). doi:10.14569/IJACSA.2022.0131139
- [24] Shuai Ma, Jianfeng Cui, Chin-Ling Chen, Xuhui Chen & Ying Ma, "An effective data enhancement method for classification of ECG arrhythmia", *Measurement*, Volume 203, (2022), doi:10.1016/j.measurement.2022.111978.
- [25] Guo, M.-F., Zeng, X.-D., Chen, D.-Y. & Yang, N.-C. "Deep-learning-based earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems". *IEEE Sensors Journal* 18, 1291–1300 (2018). doi:10.1109/JSEN.2017.2776238 .
- [26] Huang, J. S., Chen, B. Q., Zeng, N. Y., Cao, X. C. & Li, Y. "Accurate classification of ecg arrhythmia using mowpt enhanced fast compression deep learning networks". *Journal of Ambient Intelligence and Humanized Computing* 1, 1–18 (2020). doi:10.1007/S12652-020-02110-Y/TABLES/4 .
- [27] Daubechies, I. "The wavelet transform, time-frequency localization and signal analysis". *IEEE Transactions on Information Theory* 36, 961–1005 (1990). doi:10.1109/18.57199 .
- [28] Wu, Z., Lan, T., Yang, C. & Nie, Z. "A novel method to detect multiple arrhythmias based on time-frequency analysis and convolutional neural networks". *IEEE Access* 7, 170820–170830 (2019). doi:10.1109/ACCESS.2019.2956050 .
- [29] He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing & Sun, Jian. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification". *IEEE International Conference on Computer Vision (ICCV 2015)*. doi:10.1109/ICCV.2015.123 .
- [30] Mathunjwa, B. M., Lin, Y.-T., Lin, C.-H., Abbod, M. F. & Shieh, J.-S. "Ecg arrhythmia classification by using a recurrence plot and convolutional neural network". *Biomedical Signal Processing and Control* 64, 102262 (2021). doi:10.1016/j.bspc.2020.102262 .
- [31] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK & Stanley HE. "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals". *Circulation*. (2000). doi: 10.1161/01.cir.101.23.e215 .
- [32] Simonyan, Karen & Zisserman, Andrew. "Very Deep Convolutional Networks for Large-Scale Image Recognition". arXiv 1409.1556. (2014) doi:10.48550/arXiv.1409.1556 .
- [33] Krizhevsky, A., Sutskever, I. & Hinton, G. E. "ImageNet classification with deep convolutional neural networks". *Communications of the ACM* 60, 84–90 (2017). doi:10.1145/3065386 .
- [34] Lipton, Z. C., Berkowitz, J. & Elkan, C. A "critical review of recurrent neural networks for sequence learning", arXiv:1506.00019 (2015). doi: 10.48550/arXiv.1506.00019 .
- [35] Schmidhuber, J. "A fixed size storage o(n³) time complexity learning algorithm for fully recurrent continually running networks". *Neural Computation* 4, 243–248 (1992). doi:10.1162/neco.1992.4.2.243 .
- [36] Werbos, P. J. "Generalization of backpropagation with application to a recurrent gas market model". *Neural Networks* 1, 339–356 (1988). doi:10.1016/0893-6080(88)90007-X .
- [37] Hochreiter, S. & Schmidhuber, J. "Long short-term memory". *Neural Computation* 9 (8), 1735–1780 (1997). doi:10.1162/neco.1997.9.8.1735 .
- [38] Schuster, M. & Paliwal, K. "Bidirectional recurrent neural networks". *IEEE Transactions on Signal Processing* 45, 2673–2681 (1997). doi:10.1109/78.650093 .
- [39] Kłosowski, G.; Rymarczyk, T.; Wójcik, D.; Skowron, S.; Cieplak, T. & Adamkiewicz, P. "The Use of Time-Frequency Moments as Inputs of LSTM Network for ECG Signal Classification". *Electronics* (2020), 9, 1452. doi:10.3390/electronics9091452
- [40] Zihlmann, M., Perekrestenko, D. & Tschannen, M. "Convolutional recurrent neural networks for electrocardiogram classification", *Computing in Cardiology (CinC)* (2017), arXiv:1710.06122. doi:10.48550/arXiv.1710.06122
- [41] Ozal Yildirim. "A novel wavelet sequence based on deep bidirectional lstm network model for ecg signal classification". *Computers in Biology and Medicine* 96, 189–202 (2018). doi:10.1016/j.compbiomed.2018.03.016 .
- [42] P lawiak, P. & Acharya, U. R. "Novel deep genetic ensemble of classifiers for arrhythmia detection using ecg signals". *Neural Computing and Applications* 32, 11137– 11161 (2020). doi:10.1007/s00521-018-03980-2 .
- [43] Yang, W., Si, Y., Wang, D. & Guo, B. "Automatic recognition of arrhythmia based on principal component analysis network and linear support vector machine". *Computers in Biology and Medicine* 101, 22–32 (2018). doi:10.1016/j.compbiomed.2018.08.003 .
- [44] H. Dang, M. Sun, G. Zhang, X. Zhou, Q. Chang & X. Xu, "A novel deep convolutional neural network for arrhythmia classification," 2019 International Conference on Advanced Mechatronic Systems (ICAMechS), (2019), pp. 7-11, doi: 10.1109/ICAMechS.2019.8861645 .
- [45] F. Li, J. Wu, M. Jia, Z. Chen & Y. Pu, "Automated Heartbeat Classification Exploiting Convolutional Neural Network With Channel-Wise Attention," *IEEE Access*, vol. 7, pp. 122955-122963, (2019), doi: 10.1109/ACCESS.2019.2938617 .
- [46] J. Huang, B. Chen, B. Yao & W. He, "ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network," *IEEE Access*, vol. 7, pp. 92871-92880, (2019), doi: 10.1109/ACCESS.2019.2928017 .
- [47] M. A. Ahamed, K. A. Hasan, K. F. Monowar, N. Mashnoor & M. A. Hossain, "ECG Heartbeat Classification Using Ensemble of Efficient Machine Learning Approaches on Imbalanced Datasets," 2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT), (2020), pp. 140-145, doi: 10.1109/ICAICT51780.2020.9333534 .
- [48] X. Xu, S. Jeong & J. Li, "Interpretation of Electrocardiogram (ECG) Rhythm by Combined CNN and BiLSTM," *IEEE Access*, vol. 8, pp. 125380-125388, (2020), doi: 10.1109/ACCESS.2020.3006707 .
- [49] A. M. Shaker, M. Tantawi, H. A. Shedeed & M. F. Tolba, "Generalization of Convolutional Neural Networks for ECG Classification Using Generative Adversarial Networks," *IEEE Access*, vol. 8, pp. 35592-35605, (2020), doi: 10.1109/ACCESS.2020.2974712 .
- [50] T. Wang; C. Lu; Y. Sun; M. Yang; C. Liu & C. Ou, "Automatic ECG Classification Using Continuous Wavelet Transform and Convolutional Neural Network". *Entropy* (2021), 23, 119. doi:10.3390/e23010119.