



This is a self-archived – parallel-published version of an original article. This version may differ from the original in pagination and typographic details. When using please cite the original.

©2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

AUTHOR	Hammad Mohamed, Iliyasu Abdullah M, Subasi Abdulhamit, Ho Edmond S. L., El-Latif Ahmed A. Abd
TITLE	A Multi-tier Deep Learning Model for Arrhythmia Detection
YEAR	2021
DOI	10.1109/TIM.2020.3033072
VERSION	Author's accepted manuscript
CITATION	M. Hammad, A. M. Iliyasu, A. Subasi, E. S. L. Ho and A. A. A. El-Latif, "A Multitier Deep Learning Model for Arrhythmia Detection," in <i>IEEE Transactions on Instrumentation and Measurement</i> , vol. 70, pp. 1-9, 2021, Art no. 2502809, doi: 10.1109/TIM.2020.3033072.

# A Multi-tier Deep Learning Model for Arrhythmia Detection

Mohamed Hammad, Abdullah M Iliyasu, Abdulhamit Subasi, Edmond S. L. Ho, and Ahmed A. Abd El-Latif

**Abstract**— An electrocardiograph (ECG) is employed as a primary tool for diagnosing cardiovascular diseases (CVD). ECG signals provide a framework to probe the underlying properties and enhance the initial diagnosis obtained via traditional tools and patient-doctor dialogues. Notwithstanding its proven utility, deciphering large datasets to determine appropriate information remains a challenge in ECG-based CVD diagnosis and treatment. Our study presents a deep neural network (DNN) strategy to ameliorate the aforementioned difficulties. Our strategy consists of a learning stage where classification accuracy is improved via a robust feature extraction protocol. This is followed by using a genetic algorithm (GA) process to aggregate the best combination of feature extraction and classification. Comparison of the performance recorded for the proposed technique alongside state-of-the-art methods reported the area shows an increase of 0.94 and 0.953 in terms of average accuracy and F1 score, respectively. The outcomes suggest that the proposed model could serve as an analytic module to alert users and/or medical experts when anomalies are detected.

**Index Terms**— AAMI standard, Arrhythmia Detection, Cardiovascular Diseases, Deep Neural Network, E-healthcare devices, Electrocardiogram (ECG), Genetic Algorithm.

## I. INTRODUCTION

ACCORDING to the World Health Organization (WHO), as of 2016, cardiovascular diseases (CVDs) are the main cause of deaths globally [1]. An estimated 17.9 million succumbed to CVD-related ailments, representing 31% of all deaths worldwide in the same year. Two-thirds of these deaths occurred in low-income countries [1] where medical facilities for early detection and supports for patients of CVD is deemed generally poor.

Arrhythmia is a situation that arises when a person's heart pacemaker does not work properly, or its functions are impaired by ectopic focuses [2]. There are several diagnostic methods available to detect CVDs. Electrocardiography (ECG) is the most well-known and widely applied method for the detection of various heart diseases [3]. ECG is an affirmed method for the detection of acute coronary syndromes, intra-ventricular conduction disturbances, and arrhythmias [4]. ECG lead refers to an imaginary line between two ECG electrodes [5]. The

standard ECG has 12 leads, of which six leads are placed on the arms and/or legs of the individual, and called *limb leads*, while the remaining leads are placed on the torso and are referred to as *precordial leads*. The limb leads are labelled as I, II, III, aVL, aVR and aVF while the precordial leads are similarly labelled as V1, V2, V3, V4, V5 and V6.

Overall, ECG has valuable potential to reveal hidden disease features. However, the manual analysis of the long-term ECG signals is a tiresome task even for an expert. Hence, developing a Computer-Aided Cardiac Diagnosis (CACD) system helps clinicians in obtaining a second opinion regarding cardiac diagnosis, they serve as tools supporting the decline diagnostic errors in cardiology [6].

Several CACD systems have been developed to support different stages of CVD diagnosis and treatment based on Machine Learning (ML) strategies. Most classical ML methods are based on the extraction of features from different signal types, typically ECG, which are subsequently processed using different algorithms such as random forest [7], support vector machines [8], etc. Similarly, employing convolutional neural networks (CNN), unsupervised feature extraction mechanisms have been used in the classification of signals. While effective in its use to classify such signals, little has been invested in understanding the intrigues in CNN's automated classification of the extracted features.

Meanwhile, advances in terms of available computational tools and algorithms have revealed their use in automated early detection and diagnosis of cardiac abnormalities are on the rise. Recently, focus on ECG rhythm (ECGr) classification has similarly been on the increase. ECGr classification can be grouped into areas that focus on finding effective extraction methods [9], improving classification outcomes [10-14], and utilization of deep learning methods to enhance the performance of classification [15-18].

Akhmani *et al.* [11] presented a method based on Gaussian mixture modeling (GMM) with enhanced expectation maximization (EM) for arrhythmia detection. They obtained an accuracy of 99.7% for arrhythmia classification. In [12], Hammad *et al.* reported the use of a classifier based on characteristics of ECG signals to detect abnormal heart conditions. As reported therein, their strategy yielded an

Mohamed Hammad is with Faculty of Computers and Information, Menoufia University, Menoufia, Egypt (e-mail: mohammed.adel@ci.menoufia.edu.eg).

Abdullah M Iliyasu is with Electrical Engineering Department, College of Engineering, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia and with School of Computing, Tokyo Institute of Technology, Yokohama 226-8502, Japan. Also, he with School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130022, China (e-mail: a.iliyasu@psau.edu.sa; abdul-m-elias@ieee.org)

Abdulhamit Subasi is with Faculty of Medicine, Institute of Biomedicine, University of Turku, Turku, Finland (e-mail: abdulhamit.subasi@utu.fi).

Edmond S. L. Ho with the Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom (e-mail: e.ho@northumbria.ac.uk)

Ahmed A. Abd El-Latif is with the Mathematics and Computer Science Department, Faculty of Science, Menoufia University, Shebin El-Koom 32511, Egypt, and is with School of Information Technology and Computer Science, Nile University, Egypt (e-mail: a.rahiem@gmail.com; aabdellatif@nu.edu.eg).

average classification accuracy of 99%. In their contribution, Tuncer *et al.* [14], presented an automated method for arrhythmia detection using hexadecimal local patterns (HLP) and discrete wavelet transform (DWT). For classification, they used 1 nearest neighborhood (1NN) classifier and reported an accuracy of 99.7%.

Despite excelling in suppressing classification errors, in the face of big data of ECG records, the reported ML methods (i.e., [10-14]) might be less effective. This is attributed to the inherent shortcoming associated with training ML models on limited or small data sets. Among others, important differences between our study and those enumerated earlier include:

- Unlike the focus of most of the highlighted studies (i.e., [10-14]) on designing preprocessing of ECG signals, feature extraction, feature selection and classification protocols, this work proposes a DNN model optimized using genetic algorithm (GA) to aggregate the best combination of feature extraction and classification.
- As widely established, data inadequacy is crucial for learnability of machine and deep learning models [19]. Unlike the reported use of average-sized datasets in [12-14], which manifests in their validation on big datasets, the proposed study exhibits the same performance on both small and big datasets.
- Although with its own merits, some of the highlighted methods employed multiple ECG recordings which increases the complexity of the respective methods [11, 12]. To overcome this, the proposed employs only one lead ECG.
- Unlike [10,12,13], where classification was confined to one or two types of arrhythmia, our proposed model is employed to classify five arrhythmia categories.

To overcome of the shortcomings attributed to the highlighted ML techniques, deep learning models (DLMs) are used. In fact, today DLMs are so ubiquitous that their applications traverse different domains [19], including in applications in auspicious medical areas [20]. Specific to arrhythmia detection, numerous methods have been reported [15-18]. In addition to the studies highlighted earlier (i.e., [10-14]), some of the shortcomings in the studies on arrhythmia detection will also be ameliorated as part of contributions of this study.

To accentuate, the work presented in this study is tailored towards the automated detection of CVD ailments based on ECG signals. Specifically, a deep learning model is suffused with other ML and optimization (i.e., ResNet-LSTM+GA) techniques for efficient detection of *five* types of arrhythmia based on the association for the advancement of medical instrumentation (AAMI) standards using the most standard ECG classification dataset available, i.e., the MIT-BIH Arrhythmia [21]. Consequently, the main contributions of this study are summarized as follows:

1. Design of an optimized DNN model for feature learning required to distinguish different cardiac rhythms using a single lead ECG.
2. Integrating LSTM-based features with a  $k$ -nearest neighbor ( $k$ -NN) classifier to improve the classification efficacy.

3. Design of a new DNN method based on the cross-validation combined with genetic algorithm for optimization of features and parameters.

To the best our knowledge, this is the first attempt at arrhythmia detection in the manner enumerated above. Outcomes show promising results in terms of detection accuracy, sensitivity, specificity, positive predictivity and F1-Score.

The flow of the remainder of the study is outlined as follows. The details of the proposed techniques as well as datasets employed throughout its execution and validation are discussed in Section II. The presentation of the performance analysis and discussion of outcomes in Section III. Conclusions are drawn in Section IV.

## II. PROPOSED METHODOLOGY

The proposed technique fuses the adaptability and flexibility in input-output relationships of deep neural networks (DNN) models with the “learnability” of classical ML methods as well as repeatability inherent to the mutation, crossover and other properties of GA and other optimization techniques to realize a prodigious yet efficient strategy for arrhythmia detection.

Execution of the proposed strategy entails feeding the ECG signal into the DNN model where deep features of each patient are extracted. Next, these features are fed into a genetic algorithm that determines their optimum combination. Following that, several classifiers, including  $k$ -NN, support vector machine (SVM), multilayer perception (MLP) are utilized to classify the features following which final CVD detection is realized. Figure 1 presents the general outline of the proposed technique as explained. The remainder of this Section presents details of the *three* (i.e., the deep learning, genetic algorithm, and traditional machine learning) tiers that make up the proposed technique.

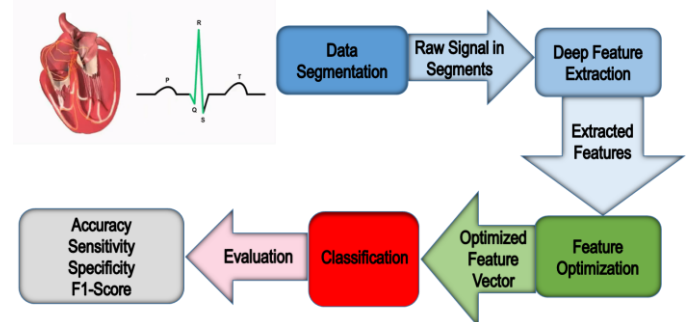


Figure 1. Outline of proposed multi-tier deep learning model (DLM)

### A. Proposed deep learning model (DLM)

DNN is a class of artificial neural networks ingrained with mathematical manipulation to turn an input state into an output via unit-wise calculation of the probability of each output [22]. Each mathematical manipulation is considered a layer of the neural network and so, in the case of complex architectures, the network requires a deep layer of computations; hence, the name DNN.

As the features for ECG classification are usually manifested in the morphology of the ECG waveforms, algorithms for feature extraction should be crafted to characterize ECG waveforms by a set of parameters. The convolution operation

in a CNN model has been proven effective in the extraction of features from 1D or 2D data that exhibit good robustness to noise [23]. The residual convolutional neural network (ResNet), which is an improved version of CNN, augment a known drawback regarding the problem of degradation associated with DNNs by adding shortcut links between its layers. Therefore, a ResNet is potentially viable for extraction the features from the raw ECG waveforms.

However, an ECG recording used to diagnose CVD ailments could extend to *tens* of seconds in duration. As a result, morphological features of ECG recordings can be very complex and hard to characterize. Moreover, an average segment of ECG recording is considered enough to make a diagnosis since parts of it are either irrelevant or redundant. Consequently, it is neither effective nor expedient to rely solely on ResNet. In view of this, the proposed DLM utilizes ResNet for local feature extraction from the raw ECG signals while using other network components, such as long short-term memory (LSTM) layer to summarize the local feature series. Furthermore, the proposed DLM ResNet-LSTM network, whose main function is to extract ECG features is structurally composed of three parts: namely, the local features learning, global features learning and the classification units. Figure 3 presents the detailed structure of the proposed ResNet-LSTM deep learning model and details of its *three* units are presented in the sequel.

#### 1. Local feature learning (LFL) unit

This part of the proposed DLM is focused on the raw ECG signals. A local feature vector manifests the morphological information of a short period in an ECG recording. The resulting feature map is composed of a sequence of local features ordered by time. As discussed earlier, in this study, the LFL unit is implemented using ResNet. It consists of a few initial layers while its main body is made up of repeating substructures. As presented in Figure 2 each substructure consists of *one* max-pooling layer whose size is present at 2 and a residual module. The length of the feature map will be split through each of the substructures whose number depends in part on the input length such that a longer input requires more pooling layers to compress the feature map to a certain length. Each residual module contains two convolutional layers, each preceded by two layers: a rectified linear unit (ReLU) activation layer and a dropout layer. Through a shortcut connection, the input of a residual module is merged by combining it with the output of its second convolutional layer. As required by the merge operation, and to ensure equal input and output length, feature maps are padded before input into each convolutional layer. The kernel size of the first convolutional layer in the network is set at 32, and through every 4 substructures, the kernel size is reduced by half. The LFL unit terminates with a ReLU activation layer where the feature maps are fed as input to the global features learning part.

#### 2. Global feature learning (GFL) unit

The global features learning (GFL) unit transforms the feature map from the preceding LFL unit into a global vector that is subsequently used in the classification unit. The LSTM, which is a kind of recurrent neural network (RNN), is used to characterize the properties of the global vector that emanates from the final layer of the LSTM. This vector's length is determined by the hidden units in LSTM layers (preset at 64 in

this study) at which stage the test samples achieved the minimum classification error using this number.

#### 3. Classification unit

The classification process is the final step of the proposed DLM whence classification process is done based on the global feature vector extracted from preceding units. Classification is accomplished using two fully connected (i.e. dense) layers that are each followed by a ReLU activation layer and a SoftMax activation layer respectively. In the case of multi-class classification, the number of cells in the last dense layer is the same as the number of classes. However, in the two-class classification, the last dense layer usually has just one cell. The outputs of the last layer are the predicted probabilities showing the class that each ECG recording belongs to. Although traditional DNN models can make predictions in an end-to-end fashion, to improve classification performance, an optimization layer is integrated to the proposed DLM, whose details are discussed in the next section.

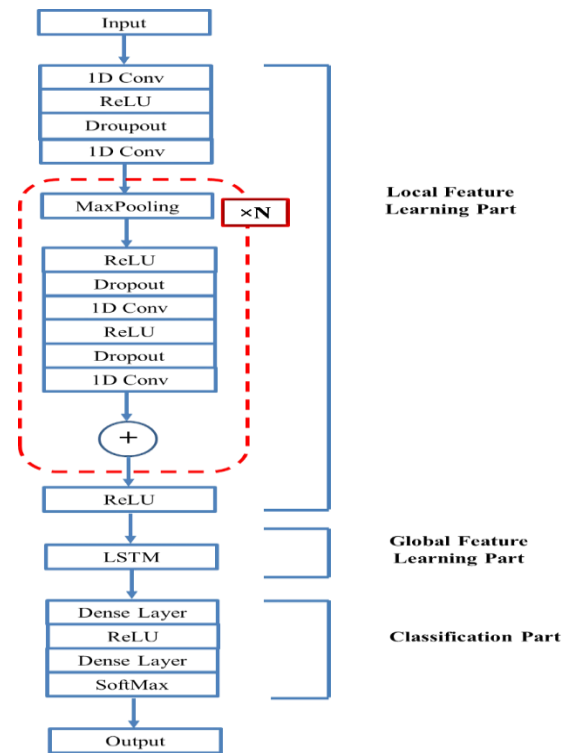


Figure 2. Architecture of proposed DLM

#### B. GA-based optimization tier

A high dimensional feature set perturbs classification accuracy while imposing temporal constraints on available resources. A wide range of approaches such as GA [24], and particle swarm optimization (PSO) [25] are employed to mitigate these encumbrances. In this study, a GA is infused into the DLM discussed earlier with the objective of minimizing the number of features and classification errors associated with the variability of signal characteristics and random permutation of the signals.

Moreover, recently GA has been found very adaptive and efficient for feature selection [26]. Typically, GA has *four* steps

and its use as the optimization tier of the proposed technique is outlined in the following steps.

- Initialize a population that consists of  $N$  chromosomes, each of length  $L$ .
- Select optimum features from the feature space comprising of 36 different features, which can be defined as the initial population in the GA optimization procedure. From the feature space, the features that have higher fitness values (as measured by a fitness function) will be eliminated, whereas those with lower fitness values will be retained.
- In each generation, select *two* parents and using the classifiers to categorize the feature combination of each of these parents separately.
- Calculate the fitness value of each classification and terminate the optimization when the generation passes the 100th iteration.
- Send the selected parents to the crossover stage and continue the optimization until the generation meets the defined number of iterations. A default crossover probability of 0.8 used to choose the parents with lower fitness values to generate the next population: Table I presents the default parameter settings for the proposed GA-based optimization procedure.

Table I. Default choice of best parameters for GA-based optimization

GA parameter	Value
Fitness Function	Accuracy or F1-score
Generations Numbers	100
Crossover	Arithmetic Crossover
Crossover Probability	0.8
Mutation	Uniform Mutation
Mutation Probability	0.1
Selection scheme	Tournament of size 2
EliteCount	2

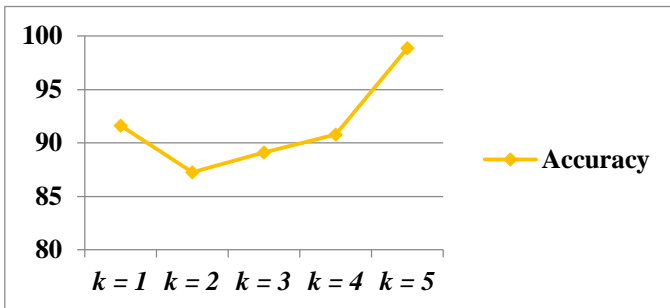


Figure 3. Impact of changing parameter  $k$  on the system accuracy

### C. Classification tier of proposed model

In this study, classification is realized using several classifiers such as  $k$ -NN, SVM and MLP. These classifiers were implemented and evaluated to choose a suitable classifier for our model as outlined in the remainder of this subsection.

#### 1. $k$ -NN classifier

The  $k$  nearest neighbor ( $k$ -NN) is a supervised machine learning algorithm that is credited ease of deployment and low temporal demands. The  $k$ -NN algorithm performs the test point classification based on the surrounding training points, i.e. the neighbors closest to the test point (hence, the name nearest neighbors). When the distance of some of these neighboring points is closer than the rest, the test point will be considered as

belonging to the category of those points. As a drawback, however, the  $k$ -NN is sensitive to errors during the training sequence [27]. In this study, the  $k$ -NN classifier with  $k = 5$  is employed. Figure 3 depicts the  $k$  selection criterion where preference is given to a value that yields highest accuracy.

#### 2. SVM classifier

The support vector machine (SVM) is another type of supervised learning algorithm that is employed for regression and classification problems by finding a hyperplane that splits the features into diverse domains [28]. When employed in classification tasks, the SVM is called Support Vector Classification (SVC). Similarly, when SVM is used in regression problems the SVM is called Support Vector Regression (SVR). Irrespective of where it is used, SVM can be broadly classified into two main types, namely, Linear SVM (L-SVM) for two-class problems and Non-Linear SVM (N-SVM) for multi-class problems. This study is built on the N-SVM, where the classification is accomplished via a kernel function to map the data into a different space with a hyper-plane. A Kernel function can be chosen from the different types available, such as linear, polynomial and radial basis function (RBF). However, the choice is guided by the kind of classification problem. Constrained by limitations of dataset available, RBF kernel, where the number of observations is larger than the number of features is chosen is used in this study. Formally, the RBF (Gaussian) kernel is defined as follows [29]:

$$K(x, x_i) = \exp\left(-\frac{\|x-x_i\|^2}{2\sigma^2}\right) \quad (1)$$

where:  $\sigma$  is a real value standard variance of Gaussian distribution and each  $x_i \in \mathbb{R}^p$  is a  $p$ -dimensional real vector. In this study, a default value of  $\sigma = 1$  is used.

#### 3. MLP classifier

Multi-layer perceptron (MLP) is a class of feed-forward artificial neural networks that consists of three layers, the input layer, a hidden layer, and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function (e.g. *sigmoid* function). In this study, MLP classifier with *one* hidden layer and *five* output nodes is used for arrhythmia detection, is used. Layers in the network are completely connected to each other and trained in a way that each input vector results in an output vector with one non-zero element corresponding to the estimated class label. For training, Levenberg-Marquardt back-propagation algorithm [30] is used. The learning procedure is continued until either 30 validation failures are obtained or when the maximum number of iterations (100 epochs) is reached.

## III. EXPERIMENTS AND VALIDATION

As mentioned earlier, a *five*-fold cross-validation method [31] is employed in this study and the performance of the system was evaluated in each fold. The average result of all the *five* folds was calculated as the final performance of the system.

The performance of the proposed DLM was assessed in three steps. First, the impact of using five-fold cross-validation on an end-to-end model was evaluated. Second, the role of proposed GA-based optimization on the classification is assessed. Finally, outcomes of the classification were compared alongside standard techniques in the area of automated



detection of cardiac abnormalities, which is an important step in CVD diagnosis and treatment.

The dataset employed and metrics used in validating the proposed model are introduced in subsequent subsections.

#### A. MIT-BIH Arrhythmia Dataset

The MIT-BIH database [21] contains 48 ECG recordings obtained from 47 subjects composed of 47% female and 53% male participants whose age range varied between 23 and 89 years. Each subject is represented by one ECG recording that is obtained using two leads placed on the chest. The lead II (MLII), which is at the top and lead V1, which is the lower one produced the signals at a sampling frequency of 360 Hz. As

explained in [26], in the upper signal, the normal QRS complexes are usually prominent while in some cases (such as in recordings 102 and 104), surgical dressings on the patients prevented the use of lead II, so, in this case, lead V5 was used. According to the AAMI specifications, the MIT-BIH database can be grouped into 5 five arrhythmia groups namely: normal (N), supraventricular ectopic (S), ventricular ectopic (V), fusion (F), and unknown (Q). The distribution of the data over the training and testing sets for each of the five types of arrhythmia are presented in Table II, while Figure 4 presents samples of three recordings (i.e., normal, and other abnormal cases) from the MIT-BIH database.

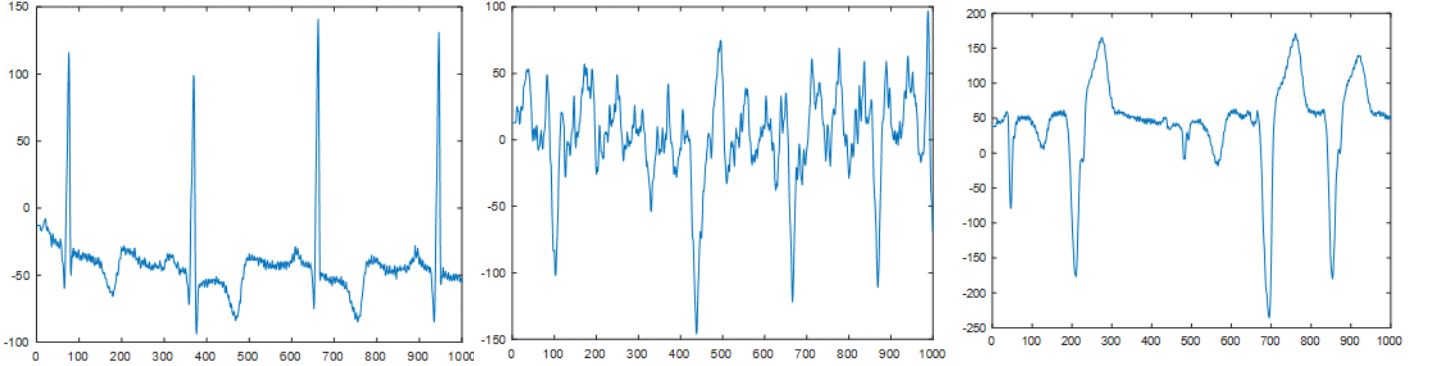


Figure 4. Samples of normal and abnormal classes from MIT-BIH dataset [21]

Table II. Distribution of data samples for each of the five types of Arrhythmia

Class	Training	Testing	All data
N	59777	14945	74722
S	2223	556	2779
V	5782	1446	7228
F	642	160	802
Q	208	52	260
<b>Total</b>	<b>68632</b>	<b>17159</b>	<b>85791</b>

#### B. Performance Metrics

The classification results were evaluated using standard metrics including the Accuracy (Acc) sensitivity (SEN), positive predictivity (PPV) and specificity (SPE). These parameters are defined in the equation matrix in Table III.

In addition, F1-score defined in Eq. (2) [32] is used to evaluate the classification.

$$F1 - score = \frac{1}{M} \left( \sum_i^M \frac{\left( \frac{\sum_i^N \frac{2TP}{2TP + FN + FP}}{N} \right)}{2} \right) \quad (2)$$

where  $N$  and  $M$  represent the number of sets applied in the five-fold and the number of classes, respectively. TP and TN are true positives and true negatives, FP and FN are false positives and false negatives.

$$\text{Positive Predictivity (PPV)} = \left( \sum_i^N \frac{TP}{TP + FP} \right) / N \quad (3)$$

$$\text{Sensitivity (SEN)} = \left( \sum_i^N \frac{TP}{TP + FN} \right) / N \quad (4)$$

$$\text{Specificity (SPE)} = \left( \sum_i^N \frac{TN}{TN + FP} \right) / N \quad (5)$$

$$\text{Accuracy (Acc)} = \left( \sum_i^N \frac{TP + TN}{TP + TN + FP + FN} \right) / N \quad (6)$$

#### C. Preprocessing and implementation details

Since ECG recordings in the selected dataset are disproportionate in length, the first preprocessing operation is that to convert them to the same length by padding or truncating for the convenience of DNN training. However, the choice of padding method can increase the computing complexity of the model, while the truncation method can damage the recording integrity. Moreover, 90% of recordings in the selected database are no longer than 30 seconds in length. In view of this, as a trade-off between the computing efficiency and recording integrity, 30 seconds long recordings are chosen as the target length. Furthermore, a zero-padding value was used because it is the baseline value of the ECG signals. The padding was positioned at the head of each recording because LSTM is considered more sensitive to values in the recording tail. Meanwhile, for the truncation, the model uses the same method that truncates a segment randomly from the original signal. Additionally, other fixed length recordings (such as 10 and 20 seconds) were considered but the 30 seconds length recordings remained the optimal choice.

The randomness involved in the padding/truncation methods, augment and balance the dataset during this process. However, considering the additional memory requirements imposed by the augmentation, the data augmentation and balancing are done in batches. For each model training, a batch is generated by randomly selecting a certain number of recordings from each class and then padding or truncating these recordings as the case may demand. For example, for an adopted batch size of 63, and a class numbers of 5, a batch contains 7 recordings from each class. The randomness in the padding (except that for the

ResNet-LSTM) and truncation ensure that, with the exception of the ResNet-LSTM, no two training samples are the same. Further, no additional memory is required for the augmented recordings because they are generated just prior to preparing a batch and discarded immediately after processing of the batch.

As stated earlier, the recordings are padded or truncated to 30 seconds, which adds up to 15000 sampling points. The local features learning pipeline of the proposed ResNet-LSTM has 7 residual substructures. As each substructure has a max-pooling layer with a pool size of 2, the resulting output length of the local features learning pipeline is set at 117. Since a recording may have more than one label, then the resulting task is a multi-label classification problem. Consequently, as outlined in Section III, SoftMax is used as the final activation function which generates the predicted probability for each class independently.

The models are implemented on the Keras framework with a Tensorflow backend. The hardware environment is a workstation equipped with an Intel Core i7-6800K CPU, a Nvidia GTX 1080ti GPU and 16 GB memory, while the training time of each batch for ResNet-LSTM is 307ms.

#### D. Results and performance analysis

In this section, the results of experiments and performance of the proposed model are presented and discussed. Specifically, the proposed models are validated via three experiments. In the first experiment, the proposed end-to-end model is employed without any ML stages. In the second experiment, the proposed deep model is used for feature extraction and an external classifier ( $k$ -NN classifier) is employed for classification. In the last experiment, a comparison of the two proposed models (i.e., end-to-end and DLM with external classifier) alongside state-of-the-art approaches reported in the literature.

As discussed earlier, based on the AAMI specification EC 57 [33], using the MIT-BIH database [21], arrhythmia is considered as beat level that can be classified as N, S, V, F and Q classes. Each beat contains 300 sampling points: 100 before and 200 after the R peak (with  $F_s = 360\text{Hz}$ ). The lead II data is used for the classification, so the channel number for the input is 1. Four residual substructures in the DLM (i.e. ResNet-LSTM) model produce a local feature map of length 18. Since this is a multi-class and single-label classification problem, SoftMax is used as the final activation function. Furthermore, both intra- and inter-patient models are trained for this problem. For the training of intra-patient models, the beats from all recordings are first shuffled and divided into 5 segments for the cross-validation, whereas for the training of intra-patient models, the beats from 22 recordings are used for training and another 22 recordings for validation, as stipulated in the AAMI standards. However, since the beat numbers of classes F and Q are small, only the other three classes (i.e., N, S and V) are involved in the inter-patient classifications. The results of the intra- and inter-patient models are presented in Table III and Table IV, respectively.

From Tables III and IV, it can be deduced that GA-optimized the features of the proposed DLM better than other methods. Similarly, it can be inferred that  $k$ -NN is better suited for GA optimization since it produced better accuracy relative to other classifiers. Figures 5 and 6 present graphs showing the validation of the proposed  $k$ -NN ML model with and without using GA-based optimization. A reading of the two plots confirms the choice of the GA-optimized strategy as more efficient.

Table III. Comparison of intra-patient's classification performance of using MIT-BIH database

Method	F1- score	Acc	N			S			V			F			Q		
			SEN	PPV	SPE	SEN	PPV	SPE	SEN	PPV	SPE	SEN	PPV	SPE	SEN	PPV	SPE
End-to-End	0.884	0.977	0.980	0.996	0.965	0.926	0.787	0.992	0.958	0.892	0.991	0.902	0.695	0.997	0.2	0.088	0.999
SVM + GA	0.888	0.978	0.984	0.994	0.953	0.900	0.791	0.992	0.951	0.918	0.993	0.888	0.711	0.997	0.333	0.250	0.999
KNN + GA	0.897	0.980	0.984	0.995	0.963	0.931	0.789	0.992	0.957	0.917	0.993	0.892	0.761	0.998	0.267	0.274	0.999
MLP + GA	0.894	0.973	0.996	0.971	0.929	0.811	0.993	0.977	0.962	0.853	0.986	0.918	0.706	0.996	0.333	0.062	0.999
SVM	0.755	0.830	0.818	0.985	0.634	0.919	0.758	0.945	0.903	0.871	0.922	0.847	0.871	0.983	0.333	0.250	0.999
KNN	0.814	0.859	0.899	0.958	0.934	0.741	0.794	0.876	0.972	0.887	0.991	0.925	0.658	0.935	0.267	0.274	0.999
MLP	0.630	0.781	0.711	0.941	0.932	0.648	0.957	0.868	0.924	0.868	0.704	0.957	0.560	0.737	0.333	0.062	0.999

Table IV. Comparison of inter-patient's classification performance of using MIT-BIH database

Method	F1- score	Acc	N			S			V		
			SEN	PPV	SPE	SEN	PPV	SPE	SEN	PPV	SPE
End-to-End	0.609	0.779	0.780	0.977	0.845	0.705	0.537	0.975	0.810	0.226	0.807
SVM + GA	0.562	0.781	0.785	0.980	0.867	0.727	0.447	0.963	0.770	0.229	0.819
KNN + GA	0.884	0.872	0.813	0.970	0.872	0.802	0.960	0.931	0.822	0.965	0.941
MLP + GA	0.735	0.821	0.832	0.711	0.672	0.717	0.583	0.922	0.782	0.813	0.864
SVM	0.546	0.589	0.528	0.957	0.605	0.246	0.436	0.897	0.664	0.212	0.597
KNN	0.805	0.815	0.719	0.946	0.865	0.711	0.921	0.932	0.796	0.918	0.868
MLP	0.675	0.751	0.650	0.675	0.868	0.636	0.556	0.749	0.721	0.813	0.881

#### E. Comparison with similar state-of-the-art methods

In this section, the performance of the proposed techniques is evaluated alongside longstanding methods in the area. For a level playing ground, outcomes with the best performance are used for the assessment. For this study, this is the DLM suffused with  $k$ -NN and GA-based optimization.

From the results the comparison presented in Table V, it can be concluded that the proposed method matches or betters the other methods in terms of robustness for the dataset reported. As noted earlier, as a motivation for embarking on this study, while it was observed that there are a good number of the machine or deep learning approaches for detecting heart disorders (such as arrhythmia) most of them are encumbered by

the computational overhead associated with the complex frameworks utilized in their detection models [10,11,13,14,15,16,17,18]. Many of these frameworks are slow [15,16,17,18] while others are overburdened by memory-related constraints [10,11,12,13]. Furthermore, most of the developed models are especially primed to detect only one type of signal. Unlike these previous studies, a multi-tier DLM capable of detecting more than one type of signal is proposed. Its use in detecting heart disorders such as arrhythmia is demonstrated.

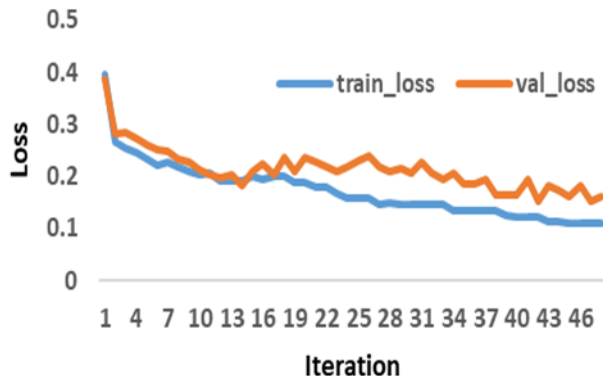


Figure 5. Plot of proposed ResNet-LSTM with GA when using  $k$ -NN as a classifier (the best classifier)

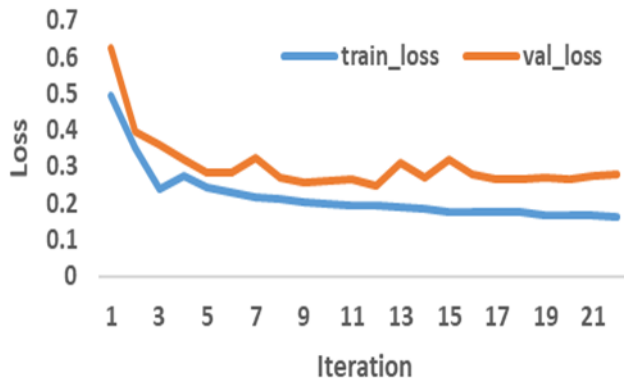


Figure 6. Plot of proposed ResNet-LSTM without GA when using  $k$ -NN as a classifier (the best classifier)

In addition, the bidirectional long-short term memory network (BiLSTM) method is reported to perform creditably for arrhythmia detection [35, 36]. Compared to our proposed model (Table V), the BiLSTM method in [35] offers marginally better performance in detection accuracy. However, [35] reports detecting RR intervals, which increases computational overhead. Furthermore, the method is more susceptible to noisy signals and its accuracy is data intensive since large amounts of ECG data is required. Additionally, it requires double LSTM cells, which makes it costlier. Consequently, the time complexity of [35] potentially prohibitive because of the complexity of BiLSTM.

Meanwhile, the intuition to suffuse GA-based optimization into the proposed DLM ensures that (the best) lower features are extracted from the ECG signals of more than one lead, despite the use of the lead II ECG signals. This suggests that the

proposed method is less complex than other competing techniques. The proposed model requires four hours for training with a computational complexity (defined in terms of the decision time for each recording) of 10ms.

Finally, as presented in Table V, for the datasets employed, the proposed technique outperforms similar ones in terms of classification accuracy, which is an important precursor for arrhythmia detection and subsequent diagnosis.

Method	Feature extraction method	Classifier	# of classes	Performance reported (%)
[11]	Gaussian mixture modeling (GMM) + enhanced expectation maximization (EM)	Ensemble of decision trees	5	Acc = 96.15 SEN = 93.30 PPV = 88.90
[12]	Modified Pan-Tompkins algorithm [34]	Characteristics of ECG signals	2	Acc = 99.00
[13]	Temporal vectorcardiogram	SVM	3	Acc = 91.39
[14]	5-levels DWT + 1-dimensional hexadecimal local pattern (1D-HLP) + neighborhood component analysis (NCA)	1 Nearest neighborhood (1NN)	5	Acc = 99.70
[16]	Deep Auto-Encoders (DAEs) pre-training	DNN	10	Acc = 99.73 SEN = 91.20 SPE = 99.80 F1-score = 91.80
[17]	Pan-Tompkins algorithm [34]	CNN	5	Acc = 94.03 SEN: 96.71 SPE: 91.54
[18]	Very deep convolutional neural network (VDCNN) + multi-canonical correlation analysis (MCCA)	QG-MSVM	4	Acc = 97.37 SEN = 98.75 SPE = 99.22
[35]	ECG morphology + 21 RR intervals	BiLSTM-Attention	5	Acc = 99.49 SEN = 96.80 SPE = 99.40 PPV = 98.28
[36]	R-peak and T episode detection based on threshold	1D CNN + Bi-LSTM	5	Acc = 95.90 SEN = 95.90 PPV = 96.34 F1-score = 95.92
<b>Proposed</b>	<b>Deep learning model + GA</b>	$k$ -NN	5	<b>Acc = 98.00</b> <b>SEN = 99.70</b> <b>SPE = 98.90</b> <b>PPV = 95.80</b> <b>F1-score=89.7</b>

#### F. Complexity analysis

To evaluate the complexity of our proposed model, the CPU time required to execute each phase is computed as presented in Table VI. Therefore, we note that, in the worst case, the cost of the LSTM amounts to ( $O(n^2)$ ).



Meanwhile, this cost is half that expected for BiLSTM method (as reported in the rightmost column of Table V). Consequently, it is imperative that the marginal improvement in terms of prediction accuracy for the BiLSTM model (as reported earlier in Table V) over the proposed LSTM DLM with GA classifier, comes at the expense of increases in computational overhead. We believe this tradeoff between cost and marginal improvement in detection accuracy must be carefully considered depending on the application.

Table VI. Comparison between computational complexity of LSTM and BiLSTM deep learning models in ECG-based arrhythmia detection

Parameter	LSTM	BiLSTM [35,36]
Conv	$O(n)$	$2O(n)$
ReLU	$O(n)$	$2O(n)$
Max. Pooling	$O(n)$	$2O(n)$
Dropout	$O(n)$	$2O(n)$
Addition operation	$O(c)$	$2O(c)$
<b>Total running time</b>	$O(n^2)$	$2O(n^2)$

#### IV. CONCLUDING REMARKS

Efficient and accurate classification of ECG signals is the first important step in the detection and diagnosis of many CVD ailments. This study presents a multi-tier DLM suffused with ML and GA-based optimization for effective discrimination of cardiac abnormalities in the ECG signal. The proposed techniques fuse the adaptability and flexibility in input-output relationships of DNN models with the “learnability” of classical ML methods as well as repeatability inherent to the mutation, crossover and other properties of GA and other optimization techniques to realize a prodigious yet efficient strategy for early detection of different CVD ailments. These properties ensure that the proposed approach alleviates many of the shortcomings that encumber traditional approaches. Outcomes show that the proposed model matches and outperforms many of the competing approaches in either or both the datasets used. Moreover, with average accuracy values of 98%, the proposed technique is efficient in arrhythmia detection on the MIT-BIH dataset. Furthermore, the performance was reported in terms of specificity (98.9%) sensitivity (99.7%) and positive predictivity (95.8%) for the five-fold cross-validation. The metrics validate the potentials for incorporating the proposed model in state-of-the-art applications for CVD detection as well as its subsequent diagnosis, treatment and management. In ongoing work, we are exploring effective techniques to fuse the proposed method into other aspects of e-health framework. In future work, we intend to exploit the potency of the BiLSTM method to enhance other aspects of CVD detection.

#### ACKNOWLEDGMENT

This study is sponsored by the Prince Sattam Bin Abdulaziz University, Saudi Arabia via the Deanship for Scientific Research funding for the Advanced Computational Intelligence & Intelligent Systems Engineering (ACIISE) Research Group Project Number 2019/01/9862.

#### REFERENCES

[1] Sudarshan, V. K. Acharya, U. R., Oh, S. L., et al. (2017). Automated diagnosis of congestive heart failure using dual tree complex wavelet transform and statistical features extracted from 2, s of ECG signals. *Computers in Biology and Medicine*, 83(Complete), 48-58.

[2] N. Singh and P. Singh, “Cardiac arrhythmia classification using machine learning techniques,” in *Lecture Notes in Electrical Engineering*, vol. 478, Springer Verlag, 2019, pp. 469–480.

[3] M. Liu and Y. Kim, “Classification of Heart Diseases Based on ECG Signals Using Long Short-Term Memory,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2018, vol. 2018-July, pp. 2707–2710.

[4] P. Kligfield et al., “Recommendations for the standardization and interpretation of the electrocardiogram: Part I: The electrocardiogram and its technology: A scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society,” *Circulation*, vol. 115, no. 10. pp. 1306–1324, Mar-2007.

[5] Lewis Potter, *Understanding an ECG*. March 5,2011. <https://geekymedics.com/understanding-an-ecg/>. Accessed [13-07-2020].

[6] N. Khosravan, H. Celik, B. Turkbey, E. C. Jones, B. Wood, and U. Bagci, “A collaborative computer aided diagnosis (C-CAD) system with eye-tracking, sparse attentional model, and deep learning,” *Med. Image Anal.*, vol. 51, pp. 101–115, Jan. 2019.

[7] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[8] Y. Freund and R. E. Schapire, “Experiments with a new boosting algorithm,” in *icml*, 1996, vol. 96, pp. 148–156.

[9] Hammad, M., Zhang, S., and Wang, K. (2019). A novel two-dimensional ECG feature extraction and classification algorithm based on convolution neural network for human authentication. *Future Generation Computer Systems*.

[10] Sadhukhan, D., Pal, S., and Mitra, M. (2018). Automated identification of myocardial infarction using harmonic phase distribution pattern of ECG data. *IEEE Transactions on Instrumentation and Measurement*, 67(10), 2303-2313.

[11] Afkhami, R. G., Azarnia, G., and Tinati, M. A. (2016). Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals. *Pattern Recognition Letters*, 70, 45-51.

[12] Hammad, M., Maher, A., Wang, K., Jiang, F., and Amrani, M. (2018). Detection of abnormal heart conditions based on characteristics of ECG signals. *Measurement*, 125, 634-644.

[13] Garcia, G., Moreira, G., Luz, E., and Menotti, D. (July 2016). Improving automatic cardiac arrhythmia classification: Joining temporal-vcg, complex networks and svm classifier. In *2016 International Joint Conference on Neural Networks (IJCNN)* (pp. 3896-3900). IEEE.

[14] Tuncer, T., Dogan, S., Pławiak, P., & Acharya, U. R. (2019). Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals. *Knowledge-Based Systems*, 186, 104923.

[15] Zhang, J., Chen, X., Liu, A., Chen, X., Zhang, X., & Gao, M. (2020). ECG-based Multi-Class Arrhythmia Detection Using Spatio-Temporal Attention-based Convolutional Recurrent Neural Network. *Artificial Intelligence in Medicine*, 101856.

[16] Nurmaini, S., Umi Partan, R., Caesarendra, W., Dewi, T., Naufal Rahmatullah, M., Darmawahyuni, A., ... & Firdaus, F. (2019). An automated ECG beat classification system using deep neural networks with an unsupervised feature extraction technique. *Applied Sciences*, 9(14), 2921.

[17] Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., Adam, M., Gertych, A., and San Tan, R. (2017). A deep convolutional neural network model to classify heartbeats. *Computers in biology and medicine*, 89, 389-396.

[18] Amrani, M., Hammad, M., Jiang, F., Wang, K., and Amrani, A. (2018). Very deep feature extraction and fusion for arrhythmias detection. *Neural Computing and Applications*, 30(7), 2047-2057.

[19] Sedik, A.; Iliyasa, A.M.; Abd El-Rahiem, B.; Abdel Samea, M.E.; Abdel-Raheem, A.; Hammad, M.; Peng, J.; Abd El-Samie, F.E.; El-Latif, A.A.A. (2020). Deploying Machine and Deep Learning Models for Efficient Data Augmented Detection of COVID-19 Infections. *Viruses* 2020, 12(7), 769.

[20] Alghamdi, A., Hammad, M., Ugail, H., Abdel-Raheem, A., Muhammad, K., Khalifa, H. S., & Abd El-Latif, A. A. (2020). Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities. *Multimedia Tools and Applications*, 1-22.

- [21] Goldberger AL, Amaral LA, Glass L, et al. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*. 2000; 101(23): E215.
- [22] Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, 25(1), 65.
- [23] Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., and Inman, D. J. (2019). 1D Convolutional Neural Networks and Applications: A Survey. arXiv preprint arXiv:1905.03554.
- [24] Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and computing*, 4(2), 65-85.
- [25] Rini, D. P., Shamsuddin, S. M., Yuhaniz, S. S. (2011). Particle swarm optimization: technique, system and challenges. *International journal of computer applications*, 14(1), 19-26.
- [26] Huang, C. L., & Wang, C. J. (2006). A GA-based feature selection and parameters optimization for support vector machines. *Expert Systems with applications*, 31(2), 231-240.
- [27] Berkaya, S. K., Uysal, A. K., Gunal, E. S., Ergin, S., Gunal, S., and Gulmezoglu, M. B. (2018). A survey on ECG analysis. *Biomedical Signal Processing and Control*, 43, 216-235.
- [28] Hammad, M., Plawiak, P., Wang, K., & Acharya, U. R. (2020). ResNet-Attention model for human authentication using ECG signals. *Expert Systems*, e12547.
- [29] Hammad, M., and Wang, K. (2017, April). Fingerprint classification based on a Q-Gaussian multiclass support vector machine. In *Proceedings of the 2017 International Conference on Biometrics Engineering and Application* (pp. 39-44). ACM.
- [30] Hagan, M. T., & Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE transactions on Neural Networks*, 5(6), 989-993.
- [31] Duda RO, Hart PE, Stork DG. *Pattern Classification 2nd Edition*. Pattern classification. New York, John Wiley and Sons, 2001:55-88.
- [32] Liu F, Liu C, Zhao L, et al. An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. *J Med Imaging Health Inform* 2018;8(7):1368-1373.
- [33] Association for the Advancement of Medical Instrumentation. (2013). ANSI/AAMI EC57: 2012—Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms. American National Standard.
- [34] Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE transactions on biomedical engineering*, (3), 230-236.
- [35] Li, R., Zhang, X., Dai, H., Zhou, B., & Wang, Z. (2019). Interpretability analysis of heartbeat classification based on heartbeat activity's global sequence features and BiLSTM-attention neural network. *IEEE Access*, 7, 109870-109883.
- [36] Xu, X., Jeong, S., & Li, J. (2020). Interpretation of Electrocardiogram (ECG) Rhythm by Combined CNN and BiLSTM. *IEEE Access*, 8, 125380-125388.

**Mohamed Hammad** received his M.Sc. degree in 2015, Information Technology Department, Faculty of Computers and Information, Menoufia University, Egypt and the Ph.D. degree in 2019, School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China. He is currently an Assistant Professor in Faculty of Computers and Information, Menoufia University, Egypt. His research interests include Computer Vision, Machine Learning, Pattern Recognition, and Biometrics.

**Abdullah M. Ilyasu** (aka Abdul M. Elias) (Senior Member, IEEE) received the M.E., Ph.D., and Dr. Eng. degrees in computational intelligence and intelligent systems engineering from the Tokyo Institute of Technology (Tokyo Tech.), Japan,

where he is currently a Research Faculty with the School of Computing. He is also a Professor with the School of Computer Science and Technology, Changchun University of Science and Technology, China. In addition to being among the pioneers of research in the emerging quantum image processing (QIP) subdiscipline, he has to his credit more than 100 publications traversing the areas of computational intelligence, quantum cybernetics, quantum image processing, quantum machine learning, cyber and information security, hybrid intelligent systems, the Internet of Things, 4IR, health informatics, and electronics systems reliability. Prof. Ilyasu is an Associate Editor in many other journals, including IEEE ACCESS and Information Sciences.

**Abdulhamit Subasi** is specialized in Machine Learning, Data mining and Biomedical Signal Processing. Concerning application of machine learning to different fields, he wrote seven book chapters and more than 170 published journal and conference papers. He is also author of the books, “Practical Guide for Biomedical Signals Analysis Using Machine Learning Techniques” and “Practical Machine Learning for Data Analysis Using Python”. Since 2020, he has been working as a Professor of medical engineering at University of Turku, Turku, Finland. He has worked on several projects related to biomedical signal processing and data analysis.

**Edmond. S. L. Ho** received the B.Sc. degree in Computer Science from the Hong Kong Baptist University, in 2003, the M.Phil. degree in Computer Science from the University of Edinburgh, Scotland, in 2006 and the Ph.D. degree in Informatics from the University of Edinburgh, Scotland, in 2011. He is currently a Senior Lecturer with the Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom. His current research interests include computer graphics and animation, computer vision, machine learning, and robotics.

**Ahmed A. Abd El-Latif** received the B.Sc. degree with honor rank in Mathematics and Computer Science in 2005 and M.Sc. degree in Computer Science in 2010, all from Menoufia University, Egypt. He is an associate professor of Computer Science at Menoufia University, Egypt. He is author and co-author of more than 130 ISI papers. His areas of interests are multimedia content encryption, secure wireless communication, IoT, applied cryptanalysis, perceptual cryptography, secret media sharing, information hiding, biometrics, forensic analysis in digital images, and quantum information processing. Dr. Abd El-Latif is an associate editor of Journal of Cyber Security and Mobility.