Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-10150-x

A novel approach for early prediction of sudden cardiac death (SCD) using hybrid deep learning

Rabin Kaspal[1] · Abeer Alsadoon[1] · P. W. C. Prasad[1] · Nedhal A. Al-Saiyd[2] · Tran Quoc Vinh Nguyen[3] · Duong Thu Hang Pham[3]

Received: 30 May 2020 / Revised: 1 October 2020 / Accepted: 23 October 2020
Published online: 31 October 2020

Abstract
Importance of early prediction of Sudden Cardiac Deaths (SCD) has been rising as a large percentage of mortality of patients with cardiovascular diseases. Various deep learning methodologies has been developed to predict the onset of SCDs, Their key limitation is either classification accuracy or the processing time. This research tries to improve the classification accuracy and decrease the processing time. A Convolutional Neural Network (CNN) is combined with a Recurrence Complex Network (RCN) along with Dropout Regularization to enhance the accuracy of SCD classification. Initially, the synchronization feature of individual heartbeat of the electrocardiogram (ECG) signal is constructed by RCN. The recurrence matrix from the (RCN) will generate Eigen values. Then, CNN will be employed to extract features and detect SCD by analysing the Eigen values. Finally, the performance of the classification is improved by the developing a voting algorithm for the SCD detection. MIT-BIH SCD database is used to evaluate the proposed system. The average accuracy and processing time for MIT-BIH Arrhythmia dataset is 93.24% and 21 epochs, MIT-BIH SCD Holter dataset is 90.60% and 11.5 epochs, and Apnoea-ECG dataset is 92.13% and 13.5 epochs. The average processing time has also been reduced to 20.77 milliseconds against the current processing time of 32.96 milliseconds. The proposed system enhances the classification accuracy and the processing time of the prediction system. The study eradicates the issue of gradient saturation during the training of the CNN by proposing a new activation function as well as eliminates the risk of overfitting by implementing dropout regularization in CNN.

Keywords Sudden cardiac death · Convolution neural network · Recurrence complex network deep learning · Dropout regularization · Electrocardiogram (ECG) signals

✉ Abeer Alsadoon
alsadoon.abeer@gmail.com

Extended author information available on the last page of the article

1 Introduction

Earlier methods in the field of mortality prevention due to Sudden Cardiac Deaths (SCD) involved development of effective targeted therapeutic interventions; such as implantable cardioverter defibrillators (ICDs) [21]. The drawback of these methods is the limited cost effectiveness because of a small number of people receiving inappropriate ICD shocks in follow-up clinics as well as most SCDs seen in patients not having high risk profile [4]. Moreover, the traditionally developed computer-aided diagnosis systems were proven to work relatively well only when the illness is already existing in the patient [3]. Currently, deep learning methods based on convolution neural networks (CNNS) are gaining popularity to overcome the limitations of traditional methods. Hence, the avoidance of complex mathematical abstractions or manual interventions can improve computational efficiency in the early prediction [23].

Deep learning technology has gained increasing attention in the recent years, and has enhanced greatly the ways of medical diagnosis at earlier stages [23]. Deep learning algorithms are used in many domains such as in; the image processing, text recognition, natural language processing, etc. and has been used to derive other algorithms. CNN shows high performance in the classification of 2D and 3D medical image, brain tumor [M. A. Khan, I. Ashraf, M. Alhaisoni, R. Damaševiˇcius, R. Scherer, A. Rehman, and S. A. Chan Bukhari, "Multimodal Brain Tumor Classification Using Deep Learning and Robust Feature Selection: A Machine Learning Application for Radiologists," Diagnostics, Aug 2020, vol. 10, no. 8, pp.:565., DOI: https://doi.org/10.3390/diagnostics10080565.]. The convolution neural network (CNN) is applied to extract features and reduce computation time in classification [NEW], [16] of patients having heart conditions using their ECG signals. CNNs are well-known in pattern recognition technique that combines feature extraction, feature reduction and classification techniques with the use of multiple convolution layers, pooling layers and a fully connected layer. Convoluted optimal features are extracted and classified using feed-forward artificial neural networks that employs a fully connected layer with back propagation learning mechanism [23]. However, the CNN has a low convergence speed, the back-propagation learning is slow and requires large number of iterations [6]. Other concerns related to network performance degradation when applying different datasets to construct a general feature extraction and classification mechanism to fit all datasets, is quite challenging [12].

Currently, a variety of combination of the feature extraction and classification techniques has been developed in order to enhance the performance of the system in terms of accuracy and speed rate when detecting SCD. The model proposed in the state of art method presents an accuracy of 94.59% [27]. The 11-layer deep CNN architecture proposed by [1] has a classification accuracy of 95.22%, while [9]'s system has an accuracy of 84.28%, ~10% less than the highest record. However, these systems are still unreliable to be implemented in the real-world scenario, where the consistency in the result is not prevalent since various factors such as; the population size, training and learning techniques, feature maps, activation functions, etc. The back-propagation algorithm used in the previous studies has not considered the risk of overfitting during training and produced misclassification due to inconsistency and inaccuracy of selected features due to noises introduced. Therefore, there are still some areas in the current studies that need improvements.

This study aims to improve the classification accuracy and the processing time of the proposed model by implementing two modified features. A modified loss function is combined with ridge regression in the back-propagation algorithm in order to avoid overfitting and a modified activation function to reduce the learning time of the network by quickly allowing the model to converge. The aim is to increase the classification accuracy of the system by combining regularization in the training in order to preserve the important features from noises as well as maintaining consistency of the selected features. Therefore, the accuracy and processing time of the system are improved in the prediction of sudden cardiac death.

In this study, a new activation functions and a modified learning technique are proposed to overcome the drawbacks in the current systems. The new proposed activation function is the modified Rectified Linear Unit (ReLU), which can effectively eliminate the problem of gradient saturation. Thus, classification accuracy and relatively faster convergence speed can be observed. To overcome loss of features or inconsistencies, regularization is used. Regularization, while preserving and maintaining consistency of selected features, also significantly reduces the variance of the model without increasing in its bias. The lesser the variance, the more different types of data the model can classify.

2 Literature review

There have been a lot of works done on the prediction of SCD. In this section, we summarize prior studies and their applied algorithms and how the problems related to the prediction of SCD are solved taking into consideration the function used, prediction performance, prediction time, and prediction accuracy.

[10] improved the prediction time of SCD to 13 min with an accuracty of 84% by proposing an optimum strategy to systematize the extracted features and decide on the appropriate processing method, which improve the prediction time. Since it has been known to be ineffective to apply the same set of features in all time intervals, this study applied feature selection using reinforcement learning that allows various features to be selected from different parts of the sample space. The obtained accuracy, sensitivity, specificity and precision were 84.28%, 85.71%, 82.85% and 83.33% respectively. However, it ignored the fact that other features may be present in the processing methods and discussed forming optimal combination in addition to other features extracted. Non-linear features obtained from ECG signal could provide newer information that were left before. Additionally, the chaotic and non-linear features grow drastically significant and relatively dominant as getting further away from the incident of SCD. Thus, it might pose more importance to examine a greater number of features while trying to optimize the prediction time.

[24] enhanced the signal processing times to around 4 to 5 min. Implantable cardioverter defibrillators (ICDs) have been used to reduce SCD. This study did a comparative study on the prediction performance of the features that address heart rate changes before SCD, using a common database and a support vector machine classifier. From the comparison of the features, the time domain features are found to have high performance and also high AUROC with 85.7% in sensitivity, specificity and accuracy and an analysis window of 5 min. However, the drawback of this solution is that the processing times is marginally increased when different features are selected in each minute. This will take more time and slow down the system.

[21] have tried to automatically identify a developing SCD in patients using machine learning approaches on the arrhythmic risk markers using the intensively clinically demonstrated electrocardiographic markers.The study introduced more effiecient and practical approach based on calculation of SCDI, an integration of the informative markers, simplified the

prediction method. The automated strategy employing the five classifiers was able to predict SCD in less than one second with average accuracy of 98.91% (KNN), 98.70%(SVM), 98.99% (DT), 97.46% (NB) and 99.49% (RF) for up to thirty minutes prior to the occurrence of SCD. However, the study was performed on a relatively smaller data from a public database, which is why the results cannot be applied to all the real cases. The arrhythmias risk markers are currently in epidemiological studies and not routinely. Additionally, there is a problem in overfitting (failing to fit additional data) during the training process, which reduces the accuracy of the recognition of risk markers of SCD.

[3] improved the prediction time and the accuracy of the prediction of SCD by proposing a methodology based on ECG signals applying Wavelet Packet Transform as a signal processing technique, Homogeneity Index (HI) as a nonlinear measurement for time series signals and an Enhanced Probabilistic Neural Network (EPNN) as the classification algorithm. The effectiveness and usefulness is evaluated using a database of measured ECG data for patients with and without SCD. The proposed solution is capable of predicting the risk of developing an SCD event up to 20 min prior to the incident with a high accuracy of 95.8% and decreasing processing times for signals as ECG signal are directly used. The drawbacks to this model is that; the test data is relatively small compared to the number of real SCD cases and the variations in each of them. The HI didn't consider various linear and non-linear features of the signal that might contribute to lowering or increasing the prediction time.

[8] presented a system that is designed based on a novel multi-class approach for the prediction of SCD 10 min before its the occurance. The approach includes the HF patients and normal people, people as the comparison group, while previous studies only included normal patients over the same database. A hybrid approach of unsupervised and ensemble learning algorithms has been constructed for dimensional space reduction and optimal feature selection of different features of HRV based on clustering the data into different groups. Among the ensemble learning algorithms, boosting, bagging, and random subspace are popular algorithms to construct a set of classifiers based on weighted votes. The bagging algorithm that builds deep trees is time and memory consuming in comparison to the boosting and subspace methods. Moreover, among the multi-class predictions, the recommended algorithms are subspace and RUSBoost, in which the latter is extremely efficient in the case of random sampling of test cases. Therefore, the hybrid approach of the SFS-RUSBoost algorithm produced a minimum classification error of 20% with eight selected subset of features. The solution presented by the study performs prediction of SCD 10 min earlier to occurrence of SCD episode, with a novel multi-class classification approach, and average classification accuracy of 83.33 for SCD prediction.

[2] enhanced the classification accuracy by using two artificial intelligence classifiers; namely KNN and LDA.The SCD and normal patients were classified using 8 different time and frequency domain features. The ECG signals from the patients are filtered, the amplitude is normalized, and features are normalized to improve the performanace and accuracy of the classifiers. Ten-fold cross validation was performed on both of classifiers to test the robustness of the classifiers. When the features are normalized, the difference in the accuracy of the KNN classifier was found, which is increased significantlymore than when it is not normalized. The risk of occurrence of SCD was predicted within 10 min before the onset with an impressive accuracy of 97% for the classifier with highest performance (KNN). Although an impressive prediction time has been calculated and a higher accuracy is gained, but various other considerations should be made before implementing the solution. A larger dataset of a diverse population should be considered so as to obtain a generalized result. Also, various other features of the time-frequency domain should be considered.

[9] enhanced the prediction time and accruacy by providing an optimal strategy to organize the extracted features and decide on the appropriate processing method to improve the prediction time. Since applying the same set of features in all time intervals has been known to be ineffective. This study implemented non-linear, time-frequency feature selection for local decision makers using reinforcement learning, which selected different features from different parts of the sample space. Prediction time; before the onset of SCD, with acceptable accuracy rate is obtained at 12 min. The obtained accuracy was 83% for a Multi-layer perception classification.

[18] significantly improved the prediction time and accuracy by proposing a deep-learning early warning system (DEWS) that used a total of 36,190 patients. The study data are decomposed into a set of 28,045 subjects for derivation data and a set of 8145 subjects for validation, these patients are taken from two hospitals. Only the derivation data is used to develop DCAPS, using a multilayer perceptron (MLP). In the training dataset for predicting neurologic recovery, the AUROC of DCAPS, LR, RF, SVM, and conventional model was 0.976. 0.955, 0.965, 0.951, and 0.821. In training dataset for predicting survival to discharge, the AUROC of DCAPS, LR, RF, SVM, and conventional model was 0.928, 0.891, 0.886, 0.856, and 0.734 respectively. Attempt to fix it by adjusting event/non-event data will lead to high sensitivity which is not preferred. Because data from one hospital is used for model derivation, and data from other hospital for testing purposes, the result is not guaranteed on other hospitals.

[19] proposed and enhanced the performance of a deep learning model based on echocardiography results predicted in-hospital mortality among HD patients more accurately than the existing prediction models and other machine learning models. It developed and compared the result of the Deep Learning (DL) model with a Logistic Regression and Random Forest machine learning model. Furthermore, subgroup analyses were performed for coronary heart disease (CHD) and heart failure (HF) patients. The DL model outperformed all the risk scores as well as the machine learning model in terms of AUROC and AUPRC. The proposed method comprised of 25,776 patients among which 1026 were deceased. The DL model calculated a risk score of 22% and the areas under the receiver operating characteristic curve (AUROC) were 0.912, 0.898, 0.958, and 0.913 for internal validation, external validation, CHD, and HF, respectively. They have shown outperforming results over other considered algorithms. However, since the deep learning and machine learning are derived from the relationship between given data and results and not from medical knowledge, the model is strongly based on the memory of the derivation data. Therefore, the performance of the model in other situations is not guaranteed without external validation.

[25] extraordinarily enhanced the prediction times and accuracy is done for the automatic classification of abnormal ECG beats differentiated from normal ones, based on a Deep Neural Network (DNN). The DNN for the heartbeat classification was developed by using the Tensor Flow framework, the deep learning library from Google. The DNN classifier is constructed using an input layer for the raw descriptors, seven hidden layers, with 5, 10, 30, 50, 30, 10 and 5 neurons respectively, and an input layer for the raw descriptors, based on the ReLU (Rectified Linear Unit) activation function. To confirm the quality of correct classification for the approach in terms of accuracy, the authors have performed a comparison of eleven well-known classification techniques using WEKA tool. The result for the accuracy of the DNN model showed more efficient over all the datasets with values greater than 99%, and concerning sensitivity and specificity the model achieves good value in comparison with the other algorithms. A well-known MIT–BIH Arrhythmia Database is used in the experiments. This approach can be implemented to develop a model to predict SCD.

[1] enhanced the accuracy of the prediction by proposing a 11-layer deep CNN that uses a dataset obtained from the Physikalisch-Technische Bundesanstalt Diagnostic (PTBD) ECG database. ECG data of 200 subjects were taken along with 12-lead signal from each subject. Each signal is sampled at 1000 samples per second. A total of 10,546 normal ECG beats and 40,182 MI ECG beats have been used in this study. Each ECG beat consists of 651 samples (250 and 400 samples for before and after R-peaks detection respectively) comprising of one P-QRS-T wave. The study carries out these steps twice one for signal with noise and on without noise. The proposed system performed better when the noise was removed. The signal with noise reduced the performance of the system. However, the model has realized comparable results for ECG with or without noise and the ability to undersatand the underlying structure of a noisy ECG beat. An average accuracy, sensitivity, and specificity are achieved of 93.53%, 93.71%, and 92.83% respectively for ECG beats with noise. Furthermore, the highest average accuracy of 95.22% sensitivity of 95.49% and specificity of 94.19% is obtained for ECG beat without noise. On the contrary, the system required computationally intensive time to learn the features and big data is required to train the system for better performance.

[13] proposed three ANN models; multilayer perceptron (MLP), long-short-term memory (LSTM), and hybrid for the prediction of cardiac arrest are trained and then compared to other classifiers including; the modified early warning score (MEWS), and non-ANN models (logistic regression, and random forest). AUROC, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) are used for the comparison. However the degree of actual numeric difference was relatively small. They clarify that this statistically minor difference is not because the ANN is insignificantly different from other models, but because even the MEWS (with relative low accuracy compared to other classification techniques) seemed to have high performance in the emergency department. A total of 374,605 ED visits and 2,910,321 patient status updates. The ANN models (MLP, LSTM, and hybrid) obtained better results based on AUROC (AUROC: 0.929, 0.933, and 0.936; 95% confidential interval: 0.926–0.932, 0.930–0.936, and 0.933–0.939, respectively) than the best achieved results of non-ANN models, and the hybrid ANN model that utilizes baseline and sequence information showed the best performance. The only limitation is not including the temporal adequacy in the clinical study; such as: sensitivity and PPV, to evaluate the early warning system. This model cannot accurately measure whether the information about a deteriorating patient is given in timely manner. Since the reason for prediction is not reported, the action to be taken may not be clear. Using of the model is not sufficient in identifying the number of prevented cardiac arrest.

[17] enhanced a prognostic prediction model that uses a deep learning on data from a large national registry for the clinical outcomes of patients after OHCA. 52,131 patients from 2 hospitals were included during the study period. The hospital data were split by date into a derivation set and a validation set. A recurrent neural network was trained. Through validation, this study demonstrated that the accurate performance of the deep-learning model, DCAPS, was excellent for predicting neurologic recovery and survival to discharge. When DCAPS was applied to electronic health records (EHR) in a hospital and Emergency medical service (EMS), the possibility of neurologic recovery could be calculated in real time. As comparative measures, we used the area under the receiver operating characteristic curve (AUROC), the area under the precision–recall curve (AUPRC), and the net reclassification index. For 352 input vectors labelled cardiac arrest in hospital, the deep learning–based early warning system (DEWS) that shows (AUROC: 0.850; AUPRC: 0.044), which is significantly performed better than a modified early warning score (MEWS) that shows (AUROC: 0.603; AUPRC: 0.003), a random forest algorithm

(AUROC: 0.780; AUPRC: 0.014), and logistic regression (AUROC: 0.613; AUPRC: 0.007). The area under the receiver operating characteristic curve, and the area under the precision–recall curve (AUROC) are used to measure the model.

[20] enhanced the accuracy of the prediction of heart failure (HF) patients. They found a DL predictive model based on echocardiography, for predicting the in-hospital patients mortality with AHF using two hospital datasets, and validated DAHF using separated AHF registry data. The DL model consisted of 3 hidden neural network layers with 362 nodes, batch normalization and dropout layers using TensorFlow. The model has achieved high performance in predicting the mortalities and is more accurately than the existing prediction models and other machine learning models. For predicting in-hospital mortality, the area under the receiver operating characteristic curve of the DAHF were 0.880 (95% confidence interval, 0.876–0.884). These results significantly outperformed those of the GWTG-HF (0.728 [0.720–0.737]) and other machine learning models. For predicting 12 and 36-month endpoints, DAHF (0.782 and 0.813) significantly outperformed MAGGIC score (0.718 and 0.729).

[23] proposed an improved CNN-SVM model that enhanced the recurrence classification performance of AF patients. While traditional CNN model extract feature maps through convolution layer, reducing these maps from the pooling layer, implementing a MLP classifier to classify the characteristic information of the objects, then integrating CNN with non-linear SVM classifier to classify the feature maps. The developed system is tested with signals obtained from body surface potential mapping (BSPM) to find out that the performance of the modified system (CNN-SVM) outperforms that of a CNN. The constructed model ultimately achieved an accuracy of 96%, a sensitivity of 88%, and a specificity of 96%. For this process, 14 patients of AF are followed up for one year after their first treatment of which 10 remained in normal sinus rhythm while other four returned to AF. The ECG data for these patients are obtained through the 128-Lead BSPM system. However, the BPSM signals and label from patients are limited and a lot of human and material resources are required to continuously follow up patients. Moreover, there is no public database that records the required data [7, 14, 15].

2.1 State of art

Figure 1 shows the block diagram that explains the features of the state-of-art model proposed by [27] for the prediction of atrial fibrillation, which is one of the major causes for sudden cardiac deaths. The area enclosed in a blue dotted box depicts the good features, while the limitations are enclosed in a box with red dotted lines. [27] proposed a deep CNN that utilizes a RCN for generation and analysis of the synchronization features of the electrocardiogram signal to detect atrial fibrillation. This is followed by a voting algorithm to increase the classification performance of the beat-wise atrial fibrillation detection algorithm. The use of RCN for the analysis of the signal is accounted for the ECG signal being a non-stationary time series [2]. In the state of art model, the RCN is used for extracting low level atrial fibrillation features. It uses the time delay method of the phase space reconstruction method [8] for the generation of the recurrence matrix, which will be used for feature extraction. The use of convolutional layers, pooling layers and fully connected layers in the network architecture have improved the signal classification accuracy [11]. The model uses a Softmax classifier and a sigmoid function as an activation function to classify ECG signals. With this model, the obtained sensitivity, specificity, and accuracy of the state of art algorithm are 91.28%, 92.91%, and 89.59%, respectively. This model consists of four stages as shown in Fig. 1. i) Data Preprocessing ii) Low level feature extraction with RCN iii) Classification using CNN and iv) Majority Voting.

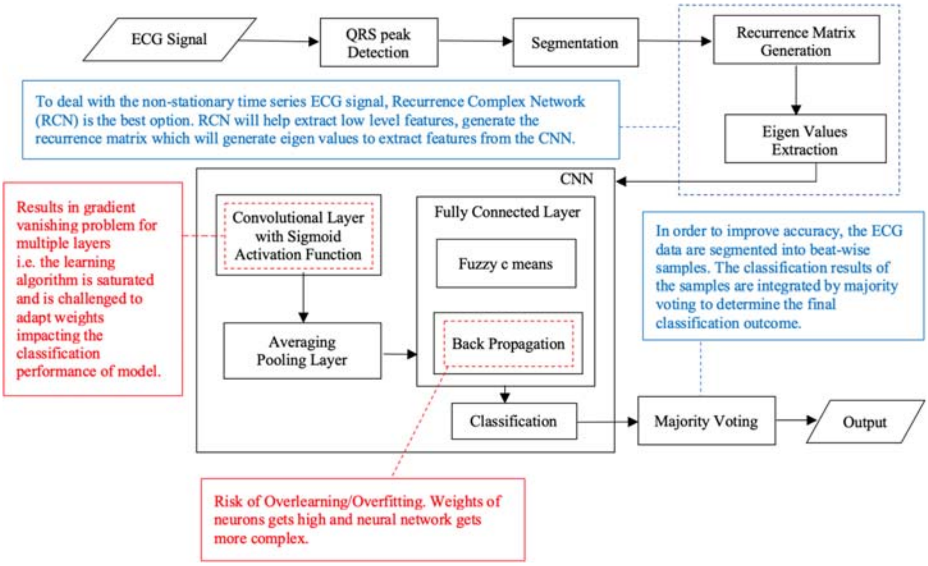


Fig. 1 Block Diagram of the State of Art System [27]. The blue dotted box represents the good features while the red dotted box shows the limitation of the state of art solution for Sudden Cardiac Death prediction using Deep Convolution Neural Networks

2.1.1 Data preprocessing

The real data provided by the MIT-BIH AF includes 25 long-term ECG recordings of humans with AF and contains 299 AF episodes. Each recording contains ECG1 and ECG2 signals, which are sampled at 250HZ and 12-bit resolution. Only ECG1 signals are used to evaluate the AF detection methods. To reduce the baseline wander and noise components of ECG signal, each single data recording is filtered with a seven-order Butterworth bandpass filter with poles at 0.5 Hz and 49 Hz. It can correct the baseline and reduce the noise effects. The local maximums of the convolution is calculated between ECG recording and a set of predefined QRS model to detect the onset of the QRS wave. At each onset point of the QRS, the QRS wave and noise are removed, based on the most matched model that find the local maximums of the convolution between the ECG recording and a set of predefined QRS models [27]. The rest of the signals are split into segments; each segment is approximately represented the output AA signal of a heartbeat, where almost the ventricular signals are removed. The signals are interpolated into 128-bit data samples using Fourier transform interpolation. Based on the samples, an ECG classification algorithm is developed next. To evaluate each method of denoising, the signal-to-noise ratios (SNR) of the original and filtered signals were calculated. The maximum value of SNR depending on the length of the signal.

2.1.2 Low level feature extraction with recurrence complex network

The RCN is a popular tool for processing non-stationary time series [27]. Since ECG data is a non-stationary time series, thus, it is analysed by the RCN. The RCN is used to generate the recurrence matrix using the time delay method of the phase space reconstruction method. Traditionally, the recurrence matrix is binarized to extract some numerical features manually which are then used to classify the input samples using fuzzy c-means (FCM). However, due to the complexity involved in the manual extraction of the appropriate features of ECG data, the features are extracted from the recurrence matrix automatically using CNN. The Eigen values of the recurrence matrix is calculated first and fed into the CNN. The CNN extracts the features and classifies the data. Each Eigen values of the data sample is a 92-byte feature vector.

2.1.3 Classification using CNN

Using CNN for ECG signal classification is composed of alternating layers of convolution and pooling layers. CNNs are able of automatically generating high-level features (i.e., weights and thresholds) by training. The CNN proposed in this study for the classification of heartbeats consists of two convolutional layers, two pooling layers, one flattening layer, and an output layer [27]. Each convolutional layer in the network can be considered to be a fuzzy filter to enhance the characteristics of the original signal, reduce noise and generate feature maps [17]. The convolution kernel size of each feature map is 5×11 and the stride is one. The sub-pooling layer subsamples the data using the principle of local correlation and retains useful information while reducing the number of parameters [25]. The pooling size is 2×2 and the stride is two. The data in the pooling layer is transformed into feature vectors by the flattening layer.

2.1.4 Majority voting

The classification accuracy obtained from the beat-wise AF detection algorithm is relatively low. So, in order to improve the classification accuracy, majority voting methodology is used. For this, before the classification, the ECG data is segmented into beat-wise data samples. Each adjacent sample is used as a collective candidate for the classification of signal. The above method is used to classify the samples of one candidate and the results are then integrated by majority voting to determine the class of the signal. The number of adjacent samples to be considered in calculated experimentally. As an activation function sigmoid function is used in the convolution layer. This activation function is prone to encountering the gradient saturation problem. Therefore, the network is unable to adapt weights during training and this will impact the classification performance [18]. Also, the used back-propagation algorithm is incapable of updating weights of the neurons throughout the learning, which increases the complexity of the network and will in turn increase the bias of the network. This makes the network learning very slow and unable to classify different types of data [13]. The problem with sigmoid function is saturated, i.e. large values snap to 1 and small values snap to 0. The function is only sensitive to changes around the mid-point of its input. Once it is saturated, it becomes a challenging issue for the learning algorithm to adapt the weights. The layers that are deep in the network do not receive useful gradient information. Errors are propagated through the network to update weights, are decreased with additional layers through which it is propagated [24]. This is called gradient vanishing problem, which will prevent the network from learning effectively and thus influences the convergence time and overall classification accuracy. Further, the loss function for the back-propagation algorithm is incapable of updating weights and/or decreasing variance of the model [19]. Therefore, the model is not able of generalizing datasets that are different from the training data. Also, since the datasets are limited, they are amplified for the training purposes. In this process, when the data is not enough or the model overtrains, then there is a risk of overfitting as the weights of the neurons get high and the network becomes more complex [26]. As a result, effective classification cannot be done, and the model may produce inaccurate outcomes.

The proposed model demonstrated higher performance with the algorithm achieving sensitivity, specificity, and accuracy of 91.28%, 92.91%, and 89.59%, respectively. The proposed method proved more effective to the problem of individual variation in the atrial fibrillation detection compared to traditional algorithm. The training algorithm is presented in Table 1 and the flowchart is presented in Fig. 2.

Activation functions in neural networks compute the weight of input and bias that will help to decide whether a neuron can be fired or not. In the presented solution, sigmoid function is used as an activation function to calculate the output of the neurons in the convolution layer as expressed in eq. (1). The input to previous layer, weight and bias will generate the input for the sigmoid function. However, the sigmoid function will suffer from gradient vanishing problem affecting the classification accuracy and efficiency. This problem can be mitigated by replacing sigmoid function with Rectified Linear Unit (ReLU) function as activation function [23].

The output from the output layer of the CNN is obtained as

$$o_n = \frac{\exp\left(\sum_{m=1}^{N_F} P_{j,m} W_{m,n}\right)}{\sum_{n'=1}^{2} \exp\left(\sum_{m=1}^{N_F} P_{j,m} W_{m,n}{}'\right)} \tag{1}$$

Where,

o_n	the units in the output layer, $n = 1, 2$.
N_F	total number of pooling layers.
m	1, 2… N_F.
$W_{m,n}$	the weight between $P_{F,m}$ and o_n.
$P_{j,m}$	final output unit in the final pooling layer

Table 1: Pseudocode for Beat-Wise AF Detection (BWAD) algorithm

```
Algorithm: BWAD method with RCN for ECG classification
Input: ECG samples E
Output: ECG Signal Classification
BEGIN
Step 1: For each ECG sample in E, create a LxL recurrence matrix
        R(i,j) = || X(t_i) – X(t_j) ||,    i , j = 1, 2 ……. L
        Where, L = N – (m-1)τ

Step2: From each of the recurrence matrix, extract Eigen values to be fed to the CNN

Step 3: Calculate convolution feature maps as
        C_j = σ(Σ_i^{N_I} O_i * W_{i,j})

        Perform averaging pooling on the feature maps to get pooling feature maps as:
        P_{j,m} = r Σ_{n=1}^{G} C_{j,(m-1)*s+n}

        Then, the output layer will generate the output as:
        o_n = exp(Σ_{m=1}^{N_F} P_{F,m} W_{m,n}) / Σ_{n'=1}^{2} exp(Σ_{m=1}^{N_F} P_{F,m} W_{m,n}')

Step 4: Perform majority voting on the collective candidate of adjacent samples.
Step5: The integrated result from majority voting is the final output.
END
```

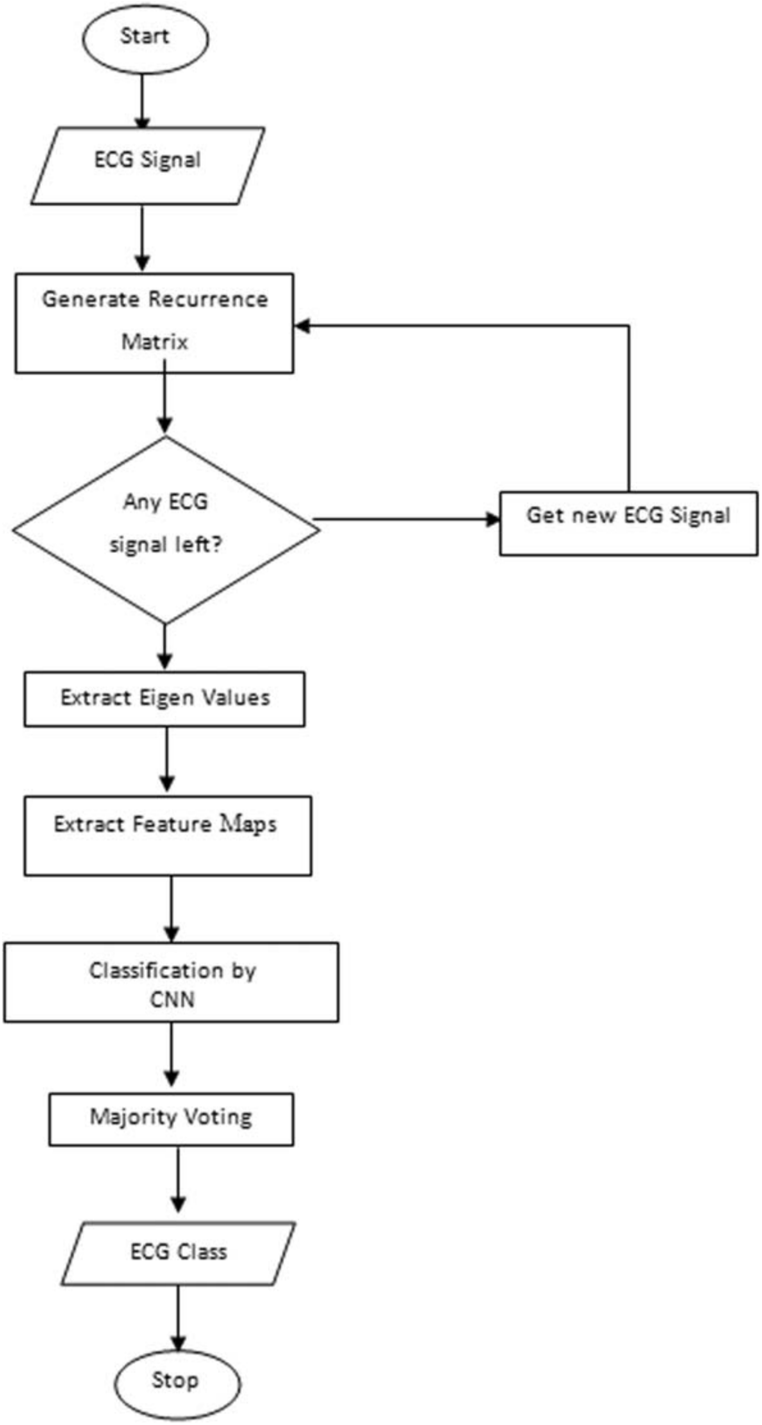


Fig. 2 Flow Diagram of The State of Art Algorithm

$$\mathrm{P}_{\mathrm{j,m}} = \ \mathrm{r} \sum_{n=1}^{G} C_{j,(m-1)*s+n} \tag{2}$$

Where,

$C_{j,\,(m-1)*s+n}$	the ((m-1) * s + n)' th unit of convolution layer feature map and is given by eq. (3)
$\mathrm{P}_{\mathrm{j,\,m}}$	m'th unit of P_j, which is the pooling layer feature map
G	pooling size
s	shift size that determines the overlap of adjacent pooling windows
r	scaling factor, selected as one in G

$$C_j = \sigma\left(\sum_i^{N_I} O_i * W_{i,j}\right) \tag{3}$$

Where,

C_j	fully connected feature map of convolution layer
j	1, 2... NO; NO = number of fully connected feature maps
N_I	total number of input feature maps
O_i	input feature maps
$W_{i,\,j}$	weight vectors
σ	the activation function as given by eq. (4)

$$\sigma \ = \frac{1}{1+e^{-x}} \tag{4}$$

Where,

e	is the natural logarithm
σ	is the activation function
x	is the input to the activation function

The CNN is trained with a backpropagation algorithm using loss function that is given by eq. (5). During machine learning, when there is a little training data, when dataset has small sample size, when dataset has noise or random fluctuations, or overtraining. It negatively impacts the performance of the model on new data. Amplifying the datasets generally will solve it. For limited data, ridge regression method is chosen, also known as L2 regularization to reduce weights of neurons and decrease complexity [23].

$$E_{x_i}(\theta) = -\sum_{n=1}^{2} y_{i,n} O_n \tag{5}$$

Where,

$E_x(\theta)$	is the loss function
θ	is all the weights of the CNN
x_i	is input sample
$y_{i,\,n}$	$[y_{i,\,1}, y_{i,\,2}]$ is binary encoding vector target for x_i
O_n	is the output unit given by eq. (1):

3 Proposed solution

From the analysis of feature extraction and classification of ECG signals using deep learning, it is found that activation function, loss function and weights are significant factors that needs to be considered for the performance of convolutional neural networks. We adopt the solution proposed by [27] because it is used CNN with majority voting to eliminate the overfitting problem. If there is overfitting, the neural network will have difficulty to generalise unknown data and lead to result in under-performance. Using dropout will select participating parameters randomly in the training. This allows the network to be switched to multiple combinations to enhance the generalization ability and steer clear of overfitting, thereby improving the accuracy of classification [27].

However, there are few issues with this solution. It uses sigmoid function in the convolution layer as the activation function, which is highly likely to suffer gradient saturation problem that affects the classification accuracy of the model and also reduces the convergence speed [10]. The other limitation is that the loss function for the back-propagation algorithm is incapable of updating weights and/or decreasing variance of the model [19]. Therefore, the model is not able of generalizing datasets that are different from the training data. This will lead to inaccurate outcomes and degrades the performance of the classifier. To fix the gradient saturation problem, the proposed solution will employ a Modified Rectified Linear Unit (MReLU) as activation function, which is inspired by [23]. Another feature of the proposed solution is the adaptation of work from [23] to deal with loss in the back-propagation algorithm. These two features will improve the classification accuracy and convergence speed. The proposed system consists of two stages as shown in Fig. 3. I) Preprocessing ii) CNN for Feature Extraction and Classification.

3.1 Preprocessing

To reduce the baseline wandering and noise of the signal, each single data recording is filtered with a seven-order Butterworth bandpass filter with poles at 0.5 Hz and 49 Hz. The local maximums of the convolution between ECG recording and a set of predefined QRS model is calculated to detect the onset of the QRS wave. At each onset point of the QRS, the QRS wave is cancelled based on the most matched model. The rest of the signals are departed into segments; each approximately the AA of a heartbeat segment. The signals are interpolated into 128-bit data samples using Fourier transform interpolation. Based on the samples, an ECG classification algorithm is developed.

3.2 CNN for feature extraction and classification

CNNs are able of automatically generating high-level features (i.e., weights and thresholds) by training. The CNN proposed in this study consists of two convolutional layers, one pooling layer, one flattening layer, and two fully connected layers. Each convolutional layer in the network can be considered to be a fuzzy filter enhancing the characteristics of the original signal and reducing noise and generate feature maps. The proposed solution uses Modified ReLU (MReLU) as activation function as shown in Fig. 3. The MReLU is simpler in comparison to sigmoid function and can solve the problem of gradient saturation. The neurons are allowed to update effectively and accelerate the convergence of the model. The sub-pooling layer subsamples the data using the principle of local correlation and retains useful

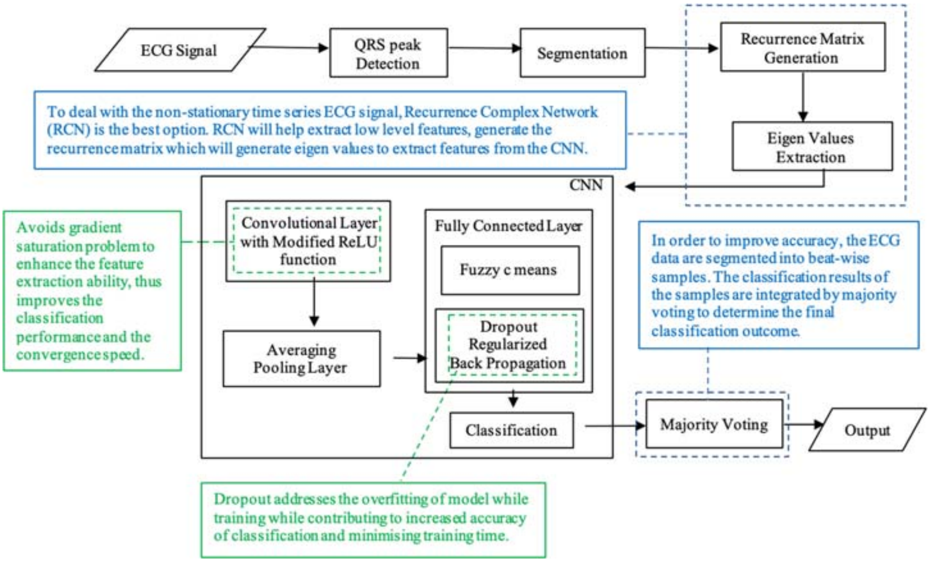


Fig. 3 Workflow of the proposed model for ECG signal classification using Modified Rectified Linear Unit (ReLU) activation function and dropout regularized back propagation learning. The green dotted box represents the new modified feature while the blue dotted box represents the existing good feature of the system

information while reducing the number of parameters (dropping out parameters). The data in the pooling layer is transformed into feature vectors by the flattening layer. The two fully connected layers employ the biased dropout method [26]. The loss is calculated by the modified loss function as in Fig. 3. The new loss function takes weight of each class into account to reduce the bias towards frequent classes [6]. The calculated loss is back propagated through the network to update weights and bias for optimization.

The training strategy namely Beat-Wise AF Detection (BWAD) algorithm with L2 Regularization is shown in Table 2 and the flowchart of the algorithm is presented in Fig. 4.

3.2.1 Proposed equations

The Rectified Linear Unit (ReLU) activation function forces the input values less than zero to zero and thus eliminates the vanishing gradient problem. The function is given as [23]:

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \tag{6}$$

Where,

x is the input to the activation function
f(x) is the activation function

The ReLU is sometimes fragile during training and allow some gradients to die, which leads to the death of some of the neurons. This will hinder in the weight update process of the training phase.

The Exponential Linear Unit (ELU) solves the problem of the ReLU by using identity for positive values, while negative values will be pushed close to zero reducing computation

Table 2: Beat-Wise AF Detection (BWAD) algorithm with L2 Regularization

Algorithm: L2 Regularized modified loss function Input: Training ECG samples, regularization parameter λ Output: Regularized loss function

BEGIN
Step 1: Initialize weights of neurons 'w'
Step 2: For each training ECG samples,
Check if w > 0. If yes,
Calculate square of weights of neuron.
Reduce the weight of the neuron by the penalty factor as in equation (12).
Repeat the above process for all neurons for all iterations.
End For
Step 3: For each ECG sample in E, create a LxL recurrence matrix
$R(i,j) = \lVert X(t_i) - X(t_j) \rVert, \quad i, j = 1, 2 \ldots\ldots L$
Where, $L = N - (m-1)\tau$
Step 4: From each of the recurrence matrix, extract eigen values to be fed to the CNN
Step 5: Calculate convolution feature maps as
$C_j = M\sigma\left(\sum_i^{N_I} O_i * W_{i,j}\right)$
Perform averaging pooling on the feature maps to get pooling feature maps as:
$P_{j,m} = r\sum_{n=1}^{G} C_{j,(m-1)*s+n}$
Then, the output layer will generate the output as:
$o_n = \frac{\exp\left(\sum_{m=1}^{N_F} p_{F,m} w_{m,n}\right)}{\sum_{n'=1}^{2} \exp\left(\sum_{m=1}^{N_F} p_{F,m} w_{m,n'}\right)}$
Step 6: Perform majority voting on the collective candidate of adjacent samples.
Step 7: The integrated result from majority voting is the final output.
END

complexity and improving learning speed. The Exponential Linear Unit (ELU) activation function with $0 < \alpha$, is given as [5]:

$$
\begin{aligned}
f(x) &= \begin{cases} x, & x > 0 \\ \alpha\exp(x) - 1, & x \le 0 \end{cases} \\
f'(x) &= \begin{cases} 1 & x > 0 \\ f(x) + \alpha & x \le 0 \end{cases}
\end{aligned}
\tag{7}
$$

Where,

f(x) is the activation function

α ELU hyperparameter that controls the saturation point for negative net inputs

The eq. (7) has some part that will eliminate the limitation of eq. (6). ELU reduces the gap between the normal gradient and the unit natural gradient because of a reduced bias shift effect for units in next layer. Therefore, ELU improves and accelerate learning characteristics in DNN compared to a ReLU network with the same architecture. ELUs have a clear saturation plateau in its negative values with smaller arguments, allowing them to learn a more robust to noise and stable representation. Saturation means a small derivative, which decreases the variation [27]. We will use only that part in green colour in eq. (7) to modify the old ReLU function. Therefore, eq. (6) will now be modified by us using eq. 7 to create an eq. (8).

$$
Mf(x) = \begin{cases} x, & x > 0 \\ \exp(x), & x \le 0 \end{cases} \tag{8}
$$

Equation (8) is the required modified activation function that eliminates the gradient vanishing problem in the state of art system. Here, we modified the state of art activation function which is eq. (4) by our modified eq. 8. Therefore, eq. (4) [27] will be modified to eq. (9) as:

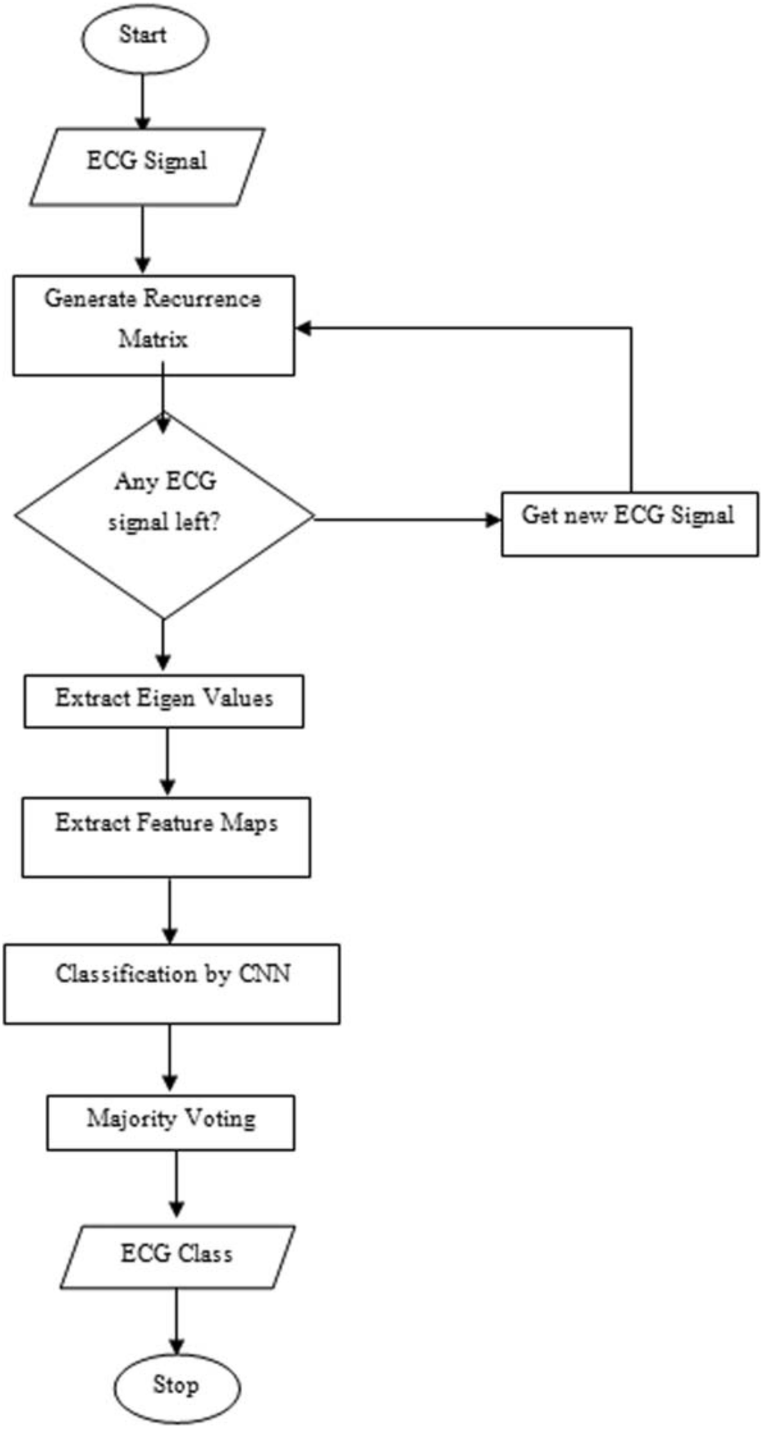


Fig. 4 Flow diagram of the proposed MReLU activation function and modified loss function for ECG classification

$$M\sigma = Mf(x) \tag{9}$$

Where,

$M\sigma$	is the modified version of the activation function
x	is the input

The solution provided by (Li et al., 2019) uses L2 regularization method to compensate for the limited data in their study. This will prevent the model from overfitting. The ridge regression adds L2 norm with as penalty factor λ thereby reducing weight and lowering the complexity of the network. This aids in proper data fitting. The loss function with L2 Regularization given by (Li et al., 2019) is expressed as eq. 10:

$$l(x) = L(X) + \lambda \sum_{i=1}^{3} \| \theta_i \|^2 \tag{10}$$

Where,

$l(x)$	is the loss function
θ	the feature weights of ith layer of the fully connected layer of the CNN
$L(X)$	is cross entropy loss
λ	is penalty factor after cost function
$\lambda\sum_{i=1}^{3}\| \theta_i \|^2$	is the regularization parameter.

Taking only the regularization parameter into account, eq. (10) should be modified by us to be eq. (11). This modified equation will then be used in eq. (5).

$$Ml(x) = \lambda \sum_{i=1}^{3} \| \theta_i \|^2 \tag{11}$$

Where,

$Ml(x)$	is the modified loss function
θ	the feature weights of ith layer of the fully connected layer of the CNN
λ	is penalty factor after cost function
$\lambda\sum_{i=1}^{3}\| \theta_i \|^2$	is the regularization parameter.

By using the regularization parameter in eq. (10), we obtain the modified loss function for the state of art system. The modified loss function is expressed as:

$$ME_{x_i}(\theta) = -\sum_{n=1}^{2} y_{i,n} O_n + Ml(x) \tag{12}$$

Where,

$ME_{x_i}(\theta)$	the enhanced loss function
θ	all the weights of the CNN
x_i	input sample
$M\ l(x)$	the modified loss function
$y_{i,n} = [y_{(i,1)}, y_{(i,2)}]$	binary encoding vector target for x_i
O_n	is the output unit given by eq. (1)

3.2.2 Area of improvement

An equation for a modified loss function was proposed in eq. (12) to improve the performance of the network model. The modified loss function is the embellishment of the state of art loss function with the addition of regularization, which reduces the risk of overfitting. Regularization significantly reduces the variance of the model without increase in its bias. The lesser the variance, the more different types of data the model can classify [28]. Regularization also decreases size of weights in the network, which helps to reduce the loss making the network simple and efficient in generalization [23]. Not implementing the regularization increases the weights over time, makes the network more complex and decreases the learning rate of the network. Another area that could be improved on the state of art solution is the activation function. By proposing a new enhanced Rectified Linear Unit activation function, the problem of gradient saturation is overcome and also the learning time of the network is reduced by allowing the model to converge. Therefore, the classification accuracy and processing time are improved.

3.2.3 Why modified loss function ($ME_{x_i}(\theta)$)?

Recently, deep learning algorithms like Convolution Neural Networks (CNN) have been widely used to eliminate the traditional explicit feature extraction process by introducing an automatic feature extraction process, which consists of convolution layers and subsampling layers. The CNN can learn to implicitly and automatically extract characteristic features from training data [22]. The complexity and parameters of CNN model is greatly reduced due to local receptive fields, shared weights and subsampling. However, because of large number of parameters involved in CNN, the training data are often vulnerable to over-fitting, which exacerbates when there is not enough training data or due to overtraining [16]. To avoid risk of potential over-fitting, early stopping and regularization are used, and the training is terminated when the accuracy of the verification set reaches certain limits.

Overfitting restricts the CNN model from generalizing newer data failing to produce accurate classification and decrease the performance of the network. Therefore, the overfitting must be prevented to train the network more effectively [29]. The possibility of overfitting can be reduced by increasing dataset size, increasing network size or regressions. Regularization significantly reduces the variance of the model without increase in its bias. The lesser the variance, the more different types of data the model can classify [28]. Regularization also decreases size of weights in the network, which helps to reduce the loss making the network simple and efficient in generalization [23]. Not implementing regularization increases the weights over time, makes the network more complex and decreases the learning rate of the network.

4 Results and discussion

For the implementation of this research, we used a software called PyCharm 2018.2 (Professional Edition) with Python 3.7.0 along with the libraries such as; Keras, NumPy, Tensorflow, sklearn, NLTK, and matplotlib. We implemented the state of art and our proposed solution, and reviewed them using three different datasets available freely. The datasets include MIT-BIH Arrythmia dataset [21], MIT-BIH SCD Holter dataset [10] and Apnoea-ECG datasets for

Table 3 Statistics of datasets used for the evaluation of State of Art model and the proposed model

Dataset Name	Total Number ECG recordings	Total Number of samples per recording	Total Number of subjects
MIT-BIH Arrhythmia Database	48	65,000	47
SCD Holter Database	23	45,000	23
Apnoea-ECG Database	70	30,000	70

a more comprehensive review of the solution proposed. Each dataset varies in the total number of samples and data balance degree. Six-fold cross-validation is used, where the training data is randomly divided into six equal sized sets. For testing purposes, one out of the six data sets is used while the rest of the data sets are used for training [29]. The model is trained using stochastic gradient descent with the dropout rate of 0.3, the learning rate of 0.18, the filter size of 4, 5, 6 respectively, batch size of 90, hidden unit of 180 (Table 3).

The system used for the experiment was configured with Intel® Core™ i5-3337U CPU@ 2.40 GHz and 4GB RAM. The experiment is divided into two parts. In the first part, the experiment covers three different scenarios, a small dataset with relatively small number of samples (Apnoea-ECG dataset), a medium sample from a small number of recordings (MIT-BIH SCD Holter dataset) and a large dataset with large number of samples per recording of ECG (MIT-BIH Arrhythmia dataset). 20% of the data from each dataset were used as validation data, while the other 80% were used for training purposes. The results for the training and validation are shown in Table 4. The second part of the experiment includes testing the CNN model for all three datasets and the results are shown in Tables 5, 6 and 7.

The accuracy and processing time for the three datasets are computed by the metric function from python Keras library. The mean method of the python NumPy library is used to calculate the average of the processing times and the accuracy. The results for different datasets are shown in Fig. 7 and 8. For the review of each datasets, the predict function from python Keras library is used to compute the classification accuracy. The now function of the datatime library is used to calculate processing time. The averages of the metrics are calculated to construct a bar graph for comparison as shown in Figs. 9, 10 and 11.

The convolutional neural network model is created to extract features from the labelled training data automatically during the feature extraction stage. The extracted features from the feature maps are fed into the max pooling layer the output of which are used by the fully

Table 4 Accuracy and Processing Time Results of State of Art solution and proposed solution for three different datasets

Dataset Name	Stage	State of Art Solution		Proposed Solution	
		Accuracy (%)	Processing Time (epoch)	Accuracy (%)	Processing Time (epoch)
MIT BIH Arrythmia Dataset	Training	89.24%	27 epochs	97.93%	23 epochs
	Validation	82.73%	23 epochs	88.54%	19 epochs
SCD HolterDataset	Training	90.05%	17 epochs	92.93%	14 epochs
	Validation	88.32%	11 epochs	91.26%	9 epochs
Apnoea-ECG Dataset	Training	92.48%	19 epochs	96%	16 epochs
	Validation	80.75%	13 epochs	88.25%	11 epochs

Table 5 Accuracy and Processing time Results of State of Art solution and proposed solution for ECG classification for MIT-BIH Arrhythmia Dataset

Sample No.	State of Art Solution		Proposed Solution	
	Accuracy (%)	Processing Time (ms)	Accuracy (%)	Processing Time (ms)
1	92.23%	32.35 ms	94.82%	22.46 ms
2	90.06%	29.80 ms	96.71%	19.54 ms
3	89.25%	31.75 ms	94.22%	22.45 ms
4	92.81%	29.55 ms	95.74%	20.47 ms
5	94.72%	36.28 ms	97.51%	21.88 ms
6	88.34%	33.45 ms	98.13%	20.92 ms
7	89.21%	31.28 ms	93.62%	21.80 ms
8	89.72%	30.35 ms	91.92%	23.38 ms
9	91.32%	33.21 ms	93.04%	23.20 ms
10	90.04%	38.45 ms	92.24%	24.21 ms

connected layer to obtain the ECG classification by applying softmax classifier in the classification stage. The trained model is then used to evaluate the validation data.

The accuracy performance of the MIT-BIH Arrhythmia dataset during the training stage is shown in Fig. 5 for both the state-of-art solution and our proposed solution. In the training stage, both the solution achieves similar accuracy. However, the proposed solution reaches maximum accuracy within fewer epochs. This provides the evidence that the proposed solution reduces the processing time for a larger dataset during the training of the model.

The accuracy performance of the MIT-BIH Arrhythmia dataset during the training validation stage is shown in Fig. 6 for both the state-of-art solution and our proposed solution. The accuracy of the proposed solution has dramatically increased in even fewer epoch compared to the state of art solution. The accuracy and processing times of the three different datasets based on the training and validation stages is presented in Table 4.

The data report and bar graphs generated from the results obtained are used for the comparison of the state of art solution and our proposed solution. The comparison is presented in following tables and figures. The results are obtained in two ways. One in terms of training and validating the model on differing dataset sizes and the other for comprehensive testing of each models with the three datasets.

Table 6 Accuracy and Processing time Results of State of Art solution and proposed solution for ECG classification for MIT-BIH SCD Holter Dataset

Sample No.	State of Art Solution		Proposed Solution	
	Accuracy (%)	Processing Time (ms)	Accuracy (%)	Processing Time (ms)
1	93.28%	31.02 ms	96.58%	19.45 ms
2	89.88%	33.54 ms	93.32%	18.29 ms
3	90.01%	37.25 ms	95.43%	21.28 ms
4	87.09%	29.82 ms	93.35%	18.93 ms
5	88.21%	30.23 ms	94.25%	21.02 ms
6	89.92%	35.82 ms	92.27%	21.25 ms
7	90.58%	33.21 ms	91.65%	20.93 ms
8	91.82%	36.81 ms	95.45%	21.84 ms
9	93.07%	30.53 ms	96.22%	19.21 ms
10	88.25%	32.88 ms	92.55%	20.44 ms

Table 7 Accuracy and Processing time Results of State of Art solution and proposed solution for ECG classification for Apnoea-ECG Dataset

Sample No.	State of Art Solution		Proposed Solution	
	Accuracy (%)	Processing Time (ms)	Accuracy (%)	Processing Time (ms)
1	89.21%	32.45 ms	94.46%	18.78 ms
2	90.5%	33.65 ms	96.58%	19.25 ms
3	91.22%	33.94 ms	94.21%	20.15 ms
4	82.21%	36.10 ms	95.63%	19.35 ms
5	85.32%	32.46 ms	92.28%	20.32 ms
6	87.25%	33.68 ms	94.50%	22.78 ms
7	89.21%	30.58 ms	97.29%	21.98 ms
8	90.23%	29.89 ms	95.34%	18.55 ms
9	91.24%	32.56 ms	93.29%	19.56 ms
10	89.28%	35.98 ms	94.76%	19.48 ms

For the first part, the results are shown based on the stages of training and validation. The result for each dataset is shown in terms of accuracy and processing times. The accuracy is calculated in terms of percentage of correctly classified samples against the total labelled samples, whereas the processing time is calculated in epochs required for the model to converge. This test was performed on the 20% of the samples in three datasets (i.e. MIT-BIH Arrhythmia, MIT-BIH SCD Holter and Apnoea-ECG datasets). The overall average accuracy is calculated by averaging the accuracy result of the training and validation stages and the overall average processing time is computed by averaging the processing times of the result of training and validation stages. The result thus obtained is presented in Table 4 and visualized in Fig. 7 and 8. For state-of-art, the average accuracy and processing time for MIT-

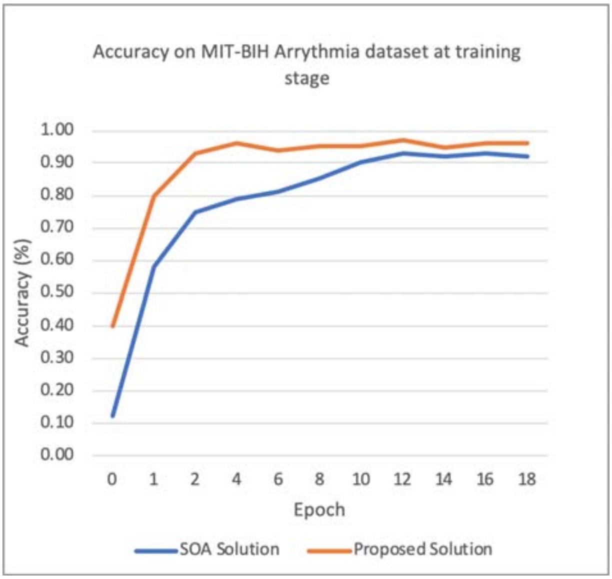


Fig. 5 The classification accuracy performance on MIT-BIH Arrhythmia dataset for state-of-art solution and proposed solution at the training stage. a) The blue line shows the classification accuracy versus epoch of the state-of-art solution. b) The orange line shows the classification accuracy versus epoch of our proposed solution

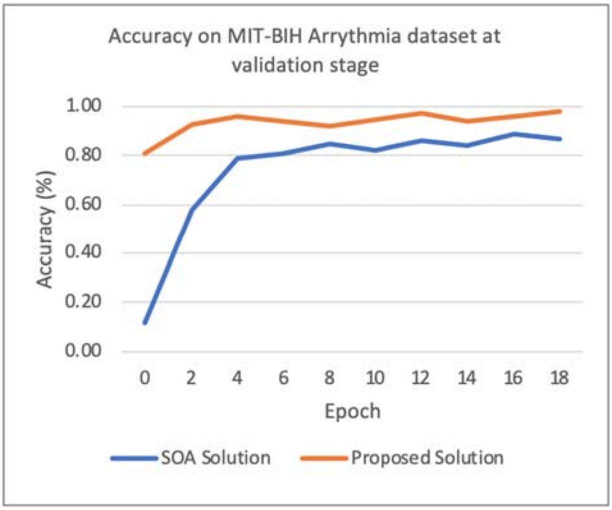


Fig. 6 The classification accuracy performance on MIT-BIH Arrythmia dataset for state-of-art solution and proposed solution at the validation stage. a) The blue line shows the classification accuracy versus epoch of the state-of-art solution. b) The orange line shows the classification accuracy versus epoch of our proposed solution

BIH Arrhythmia dataset is 85.99% and 25 epochs, MIT-BIH SCD Holter dataset is 89.91% and 14 epochs, and Apnoea-ECG dataset is 86.62% and 16 epochs. For the proposed solution, the average accuracy and processing time for MIT-BIH Arrhythmia dataset is 93.24% and 21 epochs, MIT-BIH SCD Holter dataset is 90.60% and 11.5 epochs, and Apnoea-ECG dataset is 92.13% and 13.5 epochs.

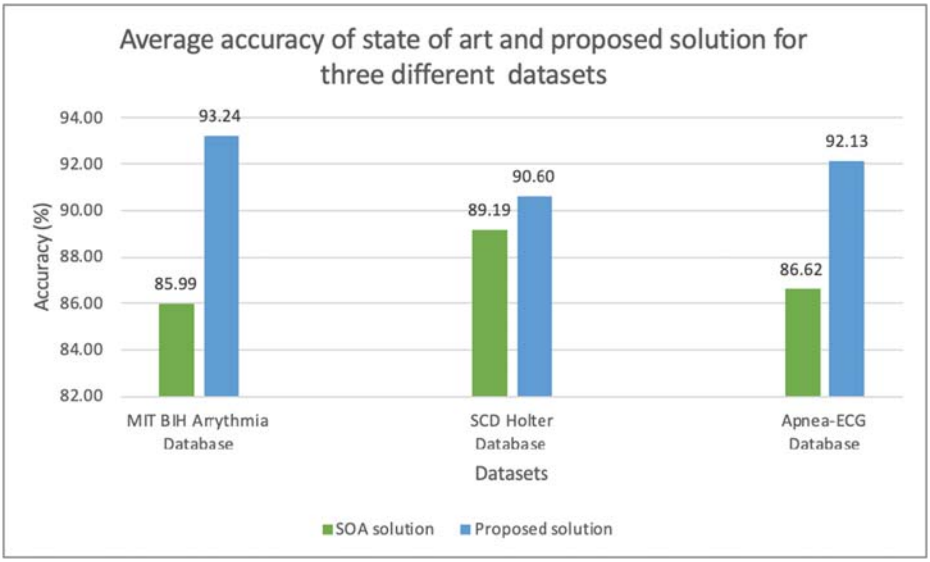


Fig. 7 Average classification accuracy in percentage for the three different datasets. The green colour indicates the accuracy of the state-of-art solution while the blue colour indicates the accuracy of the proposed solution. a) The first couple of bar graphs indicate the average accuracy for the MIT-BIH Arrhythmia dataset. b) The second couple of bar graphs indicate the average accuracy for MIT-BIH SCD Holter dataset. c) The third couple of bar graphs indicate the average accuracy for Apnoea-ECG dataset.

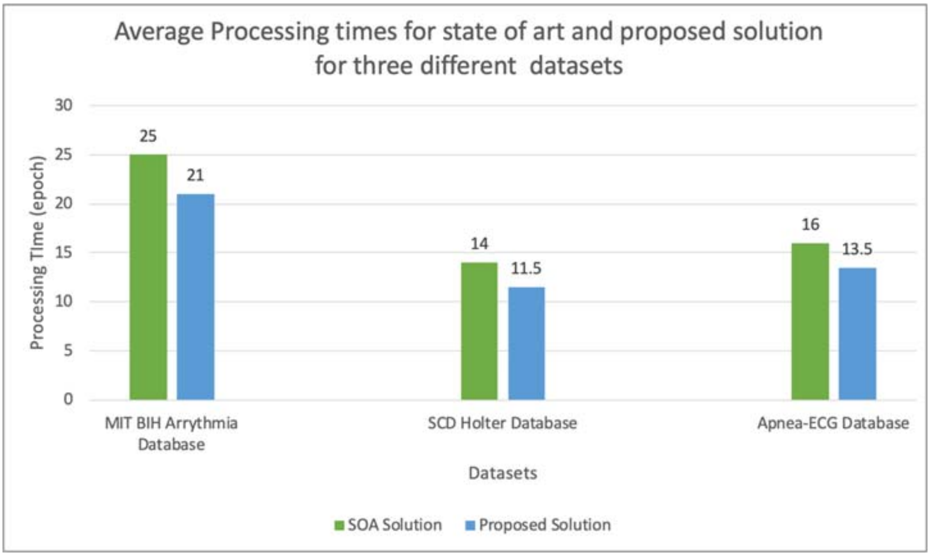
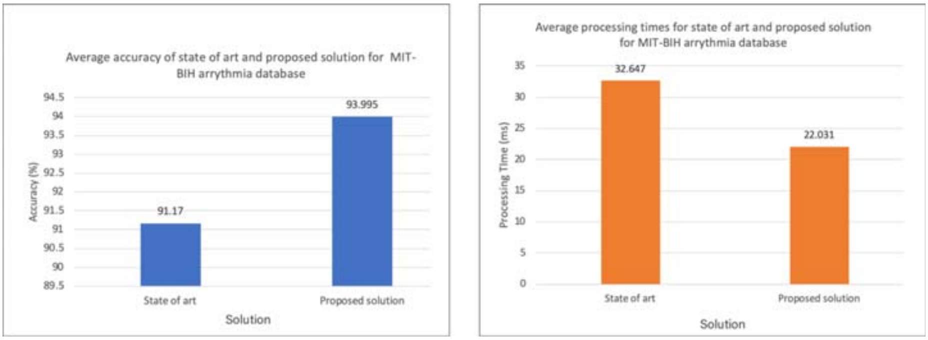


Fig. 8 Average processing times in epochs for the three different datasets. The green colour indicates the accuracy of the state-of-art solution while the blue colour indicates the accuracy of the proposed solution. a) The first couple of bar graphs indicate the average processing times for the MIT-BIH Arrhythmia dataset. b) The second couple of bar graphs indicate the average processing times for MIT-BIH SCD Holter dataset. c) The third couple of bar graphs indicate the average processing times for Apnoea-ECG dataset

For the second part, the samples from each dataset have been tested and the results for each sample are shown in Tables 5, 6, and 7 and visualised in Figs. 9, 10 and 11. The test has been conducted for ten samples from each dataset to obtain the classification accuracy and processing time. The classification accuracy is expressed as the percentage probability of the labelled data for each class whereas the processing time is the execution time for the test sample to be classified.

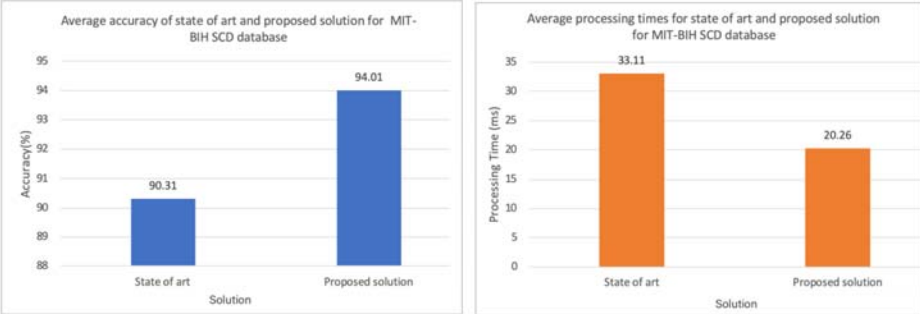
The accuracy and processing times are the metrics that are calculated in order to evaluate the state of art and the proposed solution. These results were obtained during the classification stage in the convolutional neural network. This involved the use of tools such as Python with



(a) Average accuracy in percentage

(b) Average processing times in ms

Fig. 9 Average accuracy and processing times results for State of Art Solution and Proposed solution for MIT-BIH Arrhythmia dataset. (a) The average accuracy in percentage for the state of art solution and proposed solution. (b) The average processing times in milliseconds for state of art and proposed solution



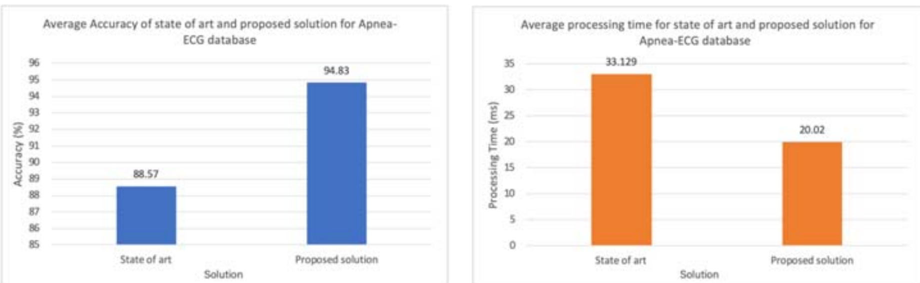
(a) Average accuracy in percentage

(b) Average processing times in ms

Fig. 10 Average accuracy and processing times results for State of Art Solution and Proposed solution for MIT-BIH SCD Holter dataset. (a) The average accuracy in percentage for the state of art solution and proposed solution. (b) The average processing times in milliseconds for state of art and proposed solution

Tensorflow backend and Keras library. The analysis of the results is done in two phases: training and validating the model for each solution and then review the model for each of the three different datasets. The classification accuracy and processing time has been improved in the proposed solution by using eq. (9) to eliminate the risk of gradient saturation and eq. (12) to prevent overfitting of the network therefore minimizing the processing times.

The results demonstrate the improvement in classification accuracy and processing time of the proposed solution comparing to the state-of-art solution of the classification of ECG signals for SCD prediction. For the modified activation function and modified loss function, the average classification accuracy of 94.7% which is 4.72% higher than the current solution [27]. Also, the convergence speed of the model is decreased with 15.3 epochs to obtain the optimization which is 5 epochs less that the current solution. Additionally, the average processing time has been decreased by 12.19 ms in the proposed solution at 20.77 ms compared to the current solution. For each of the datasets, the predict function from python Keras library computes the classification accuracy using true positives and true negatives. The now function of the datatime library calculates processing time using the start and stop time. The averages of the metrics are calculated using AVERAGE method of MS Excel, which will help in the comparison of the two solutions.



(a) Average accuracy in percentage

(b) Average processing times in ms

Fig. 11 Average accuracy and processing times results for State of Art Solution and Proposed solution for Apnoea-ECG dataset. (a) The average accuracy in percentage for the state of art solution and proposed solution. (b) The average processing times in milliseconds for state of art and proposed solution

Table 8 Comparison table of state of art and proposed solution

	Proposed Solution	State of art Solution
Name of solution	Enhanced Rectified Linear Unit activation and regularized loss algorithm	Beat wise Sudden Cardiac Death Detection Algorithm
Applied Area	Classification of Sudden Cardiac Death	Detection of Sudden Cardiac Death
Activation Function	Modified Rectified Linear Unit	Sigmoid Function
Equation	$M\sigma = Mf(x) = \begin{cases} x, & x > 0 \\ \exp(x), & x \leq 0 \end{cases}$	$\sigma = \frac{1}{1+e^{-x}}$
Loss Function	Modified Loss function	Back propagation Loss Function
Proposed equation	$ME_{x_i}(\theta) = -\sum_{n=1}^{2} y_{i,n} O_n + \lambda \sum_{i=1}^{3} \parallel \theta_i \parallel^2$	$E_{x_i}(\theta) = -\sum_{n=1}^{2} y_{i,n} O_n$
Contribution 1	The activation function is the modified Rectified Linear Unit (ReLU) which can effectively eliminate the problem of gradient saturation. Thus, classification accuracy and faster convergence speed can be observed.	Using sigmoid function as an activation function is prone to gradient saturation. Thus, the classification accuracy is reduced, and processing time is increased.
Contribution 2	Regularization significantly reduces the variance of the model without increase in its bias. The lesser the variance, the more different types of data the model can classify.	Doesn't consider the regularization resulting is possible overfitting which restricts the CNN model from generalizing newer data and thus failing to produce accurate classification and decrease the performance of the network.

The accuracy is calculated using eq. (13)

$$Accuracy = \frac{True\ Positives + True\ Negatives}{All\ Samples} \tag{13}$$

Where,

True Positives	the number of correctly identified positive samples
True Negatives	the number of correctly identified negative samples

Using modified ReLU (MReLU) function; as the activation function in the convolutional layer of the networks, eliminates the problem of gradient saturation and improves the performance of the proposed system. The main idea of the MReLU function is its gradient that will never saturate, which allows efficient updating of the parameters during training therefore, the model optimization is achieved. Using regularized loss function is another feature. Regularization significantly reduces the variance of the model without increase in its bias. The lesser the variance, the more different types of data the model can classify [28]. Regularization also decreases size of weights in the network which helps to reduce the loss making the network simple and efficient in generalization [23]. Using a modified loss function combined with ridge regression in the back-propagation algorithm in order to avoid overfitting and a modified activation function that reduces the learning time of the network by quickly allowing the model to converge. The aim is to increase the classification accuracy of the system by combining regularization in the training in order to preserve the important features from noises as well as maintaining consistency of the selected features.

The primary objectives of this solution have consistently been to achieve higher classification accuracy while using lower processing time. The limitation of the current solution has been successfully solved in this research with an average accuracy of 94.7% against the current accuracy of 91.17%. The improvement in these metrics are accounted by the use of modified activation to minimize the risk of gradient saturation and the regularized loss function to reduce the bias in the data. Hence, the proposed solution exhibits better performance in different data scenarios. .

Table 8 shows the comparison between state of art and proposed solution.

5 Conclusion

In this paper, a novel deep learning SCD detection algorithm has been presented. A CNN structure is employed leveraging multilayer structures and presenting highly abstract representation of the input. Optimizing the signals to obtain high-level features and then classify the input. A majority voting algorithm has been utilized to improve the performance of the algorithm. However, there still exist limitations in accuracy and processing time. The aim of this research is to increase the classification accuracy and reduce the processing time using deep learning methods. The modified Rectified Linear Unit activation function has been developed. This improvement reduces the risk of gradient saturation when the model is trained in order to improve the accuracy and convergence speed. Combining the loss function with ridge regression will allow preserving and maintaining consistency of selected features that will improve the classification accuracy. It also helps minimize the negative effect of bias in the imbalance dataset and further enhances the classification accuracy. It also decreases size of weights in the network which helps to reduce the loss making the network simple and efficient in generalization. Therefore, the proposed solution has reduced the convergence time by 5 epochs on average. Apart from this, the proposed solution has improved the accuracy by 4.72% on average and decreased the processing time by 12.19 milliseconds on average

Appendix

Table 9 Annotations for abbreviations used

SCD	Sudden Cardiac Death
ECG	Electrocardiogram
CNN	Convolution Neural Network
ReLU	Rectified Linear Unit
MReLU	Modified Rectified Linear Unit
RCN	Recurrence Complex Network

References

1. Acharya UR, Fujita H, Oh SL, Hagiwara Y, Tan JH, Adam M (2017) Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Information Sciences* 415–416:190–198. https://doi.org/10.1016/j.ins.2017.06.027
2. K. A. Alfarhan, M. Y. Mashor, A. Zakaria, and M. I. Omar, "Automated Electrocardiogram Signals Based Risk Marker for Early Sudden Cardiac Death Prediction," *Journal of Medical Imaging and Health Informatics,* vol. 8, no. 9, pp. 1769–1775, 2019, doi: https://doi.org/10.1166/jmihi.2018.25311769.
3. Amezquita-Sanchez JP, Valtierra-Rodriguez M, Adeli H, Perez-Ramirez CA (2019) A Novel Wavelet Transform-Homogeneity Model for Sudden Cardiac Death Prediction Using ECG Signals. *Journal of Medical Systems* 42(10):176. https://doi.org/10.1007/s10916-018-1031-5
4. Chugh SS, Kelly KL, Titus JL (2000) Sudden cardiac death with apparently Normal heart. Circulation 102(6):649–654. https://doi.org/10.1161/01.CIR.102.6.649
5. Clevert D-A, Unterthiner T, Hochreiter S (2015) Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs), *CoRR,* vol. abs/1511.07289.
6. Dang H, Sun M, Zhang G, Qi X, Zhou X, Chang Q (2019) A novel deep arrhythmia-diagnosis network for atrial fibrillation classification using electrocardiogram signals. IEEE Access 7:75577–75590. https://doi.org/10.1109/ACCESS.2019.2918792
7. Delakis M, Garcia C (2008) *Text Detection with Convolutional Neural Networks*. pp. 290–294
8. Devi R, Tyagi HK, Kumar D (2019) A novel multi-class approach for early-stage prediction of sudden cardiac death. *Biocybernetics and Biomedical Engineering* 39(3):586–598. https://doi.org/10.1016/j.bbe.2019.05.011
9. Ebrahimzadeh E, Manuchehri MS, Amoozegar S, Araabi BN, Soltanian-Zadeh H (2019) A time local subset feature selection for prediction of sudden cardiac death from ECG signal. *Medical & Biological Engineering & Computing* 56(7):1253–1270. https://doi.org/10.1007/s11517-017-1764-1
10. Ebrahimzadeh E et al (2019) An optimal strategy for prediction of sudden cardiac death through a pioneering feature-selection approach from HRV signal. *Computer Methods and Programs in Biomedicine* 169:19–36. https://doi.org/10.1016/j.cmpb.2018.12.001
11. Fujita H, Acharya UR, Sudarshan VK, Ghista DN, Sree SV, Eugene LWJ, Koh JEW (2016) Sudden cardiac death (SCD) prediction based on nonlinear heart rate variability features and SCD index. Appl Soft Comput 43:510–519. https://doi.org/10.1016/j.asoc.2016.02.049
12. Greenspan H, Ginneken BV, Summers RM (2016) Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. IEEE Trans Med Imaging 35(5):1153–1159. https://doi.org/10.1109/TMI.2016.2553401
13. Jang D et al (2019) Developing neural network models for early detection of cardiac arrest in emergency department. Am J Emerg Med 38:43–49. https://doi.org/10.1016/j.ajem.2019.04.006
14. Ji S, Xu W, Yang M, Yu K (2013) 3D convolutional neural networks for human action recognition. IEEE Trans Pattern Anal Mach Intell 35(1):221–231. https://doi.org/10.1109/TPAMI.2012.59

15. Korpusik M, Collins Z, Glass J (2017) Semantic mapping of natural language input to database entries via Convolutional neural networks, *Proceedings of* 2017 *IEEE international conference on acoustics, speech and signal processing,* pp. 5685–5689
16. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Neural Information Processing Systems 25:01/01–01/90. https://doi.org/10.1145/3065386
17. Kwon J, Lee Y, Lee Y, Lee S, Park J (2018) An Algorithm Based on Deep Learning for Predicting In-Hospital Cardiac Arrest. *Journal of the American Heart Association* 7(13):e008678. https://doi.org/10.1161/JAHA.118.008678
18. Kwon J, Jeon KH, Kim HM, Kim MJ, Lim S, Kim KH, Song PS, Park J, Choi RK, Oh BH (2019) Deep-learning-based out-of-hospital cardiac arrest prognostic system to predict clinical outcomes. Resuscitation 139:84–91. https://doi.org/10.1016/j.resuscitation.2019.04.007
19. Kwon J et al (2019) Artificial intelligence algorithm for predicting mortality of patients with acute heart failure. *PLOS ONE* 14(7):e0219302. https://doi.org/10.1371/journal.pone.0219302
20. Kwon J, Kim K, Jeon K, Park J (2019) Deep learning for predicting in-hospital mortality among heart disease patients based on echocardiography. Echocardiography 36(2):213–218. https://doi.org/10.1111/echo.14220
21. Lai D, Zhang Y, Zhang X, Su Y, Heyat MBB (2019) An automated strategy for early risk identification of sudden cardiac death by using machine learning approach on measurable arrhythmic risk markers. IEEE Access 7:94701–94716. https://doi.org/10.1109/ACCESS.2019.2925847
22. Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324. https://doi.org/10.1109/5.726791
23. Li Z, Feng X, Wu Z, Yang C, Bai B, Yang Q (2019) Classification of atrial fibrillation recurrence based on a convolution neural network with SVM architecture. IEEE Access 7:77849–77856. https://doi.org/10.1109/ACCESS.2019.2920900
24. Parsi A, Loughlin DO, Glavin M, Jones E (2019) Prediction of sudden cardiac death in implantable Cardioverter defibrillators: a review and comparative study of heart rate variability features. IEEE Rev Biomed Eng:1–1. https://doi.org/10.1109/RBME.2019.2912313
25. Sannino G, De Pietro G (2018) A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. *Future Generation Computer Systems* 86:446–455. https://doi.org/10.1016/j.future.2018.03.057
26. Si Y, Xu T, Jiang S (2018, 2018) Deep Convolutional Neural Network Based ECG Classification System Using Information Fusion and One-Hot Encoding Techniques. *Mathematical Problems in Engineering* (7354081):10. https://doi.org/10.1155/2018/7354081
27. Wei X, Li J, Zhang C, Liu M, Xiong P, Yuan X, Li Y, Lin F, Liu X (2019, 8057820) Atrial Fibrillation Detection by the Combination of Recurrence Complex Network and Convolution Neural Network. *Journal of Probability and Statistics* 2019:9. https://doi.org/10.1155/2019/8057820
28. Zhai X, Tin C (2018) Automated ECG classification using dual heartbeat coupling based on convolutional neural network. IEEE Access 6:27465–27472. https://doi.org/10.1109/ACCESS.2018.2833841
29. Zhang M, Diao M, Guo L (2017) Convolutional neural networks for automatic cognitive radio waveform recognition. IEEE Access 5:11074–11082. https://doi.org/10.1109/ACCESS.2017.2716191

Affiliations

Rabin Kaspal[1] • Abeer Alsadoon[1] • P. W. C. Prasad[1] • Nedhal A. Al-Saiyd[2] • Tran Quoc Vinh Nguyen[3] • Duong Thu Hang Pham[3]

[1] School of Computing and Mathematics, Charles Sturt University, Sydney, Australia

[2] Faculty of Information Technology, Applied Science Private University, Amman, Jordan

[3] Faculty of Information Technology, The University of Da Nang – University of Science and Education, Da Nang, Vietnam