Major and Sub-Class Classification of Arrhythmia using Eigen Vectors in ConvNet

Umamaheswari S
Department of Information Technology
Anna University MIT Campus
Chennai, India
uma_sai@mitindia.edu

Sangeetha D
Department of Information Technology
Anna University MIT Campus
Chennai, India
sangeetha@mitindia.edu

Aditya Sriram S
Department of Information Technology
Anna University MIT Campus
Chennai, India
adityasriram34@gmail.com

Nandhinipriya J
Department of Information Technology
Anna University MIT Campus
Chennai, India
nandhinipriya745@gmail.com

Abstract—**Cardiac Arrhythmia is a heart disease that corresponds to abnormal rhythm of heart. It means that the heart is either beating too quickly, too slowly, or sporadically. Arrhythmia is recognized and categorized effectively so as to improve the living conditions of the patients. The Electro Cardiogram (ECG) is a tool for recording electrical activity and determining the electrical impulses in the heart. There are four main classes of arrhythmia which occur due to abnormal heartbeat which are being classified. The main objective of this proposed work is to provide better performance in predicting arrhythmia since even a small error can become dangerous to a person's life. The existing methods uses CNN as the feature extraction model which delays the time of prediction. Here, a novel feature extraction method is introduced based on 1D-Convolutional Neural Networks using the Eigen Vectors functionality. This feature extraction model proves to outperform the existing works in accurately classifying the different classes of arrhythmia. Finally, the ANN model is trained using the K-fold Cross Validation method to achieve this performance and is compared with an ensemble model containing a SVM, ANN and a Decision Tree.**

Keywords— Machine learning approaches; Feature extraction; ECG signals; heart disease Co-variance matrix, Eigen Values, Eigen Vectors, CNN, K-fold Cross Validation, Ensemble Method, ROC

I. INTRODUCTION

Cardiovascular diseases are the largest cause of death in the world, according to the World Health Organization (WHO). They account for 31% of global deaths[7][10]. Some ectopic beats may serve as markers for persistent arrhythmia detection. Detecting abnormal heartbeats[1][12][17] from normal ECG data is therefore critical. The primary focus of this study was to identify ectopic beats and the recognition of persistent arrhythmias. The heart's electrical activity is represented by the ECG signal. As a result, the ECG signal is a useful source of information on the heart's basic functions and diseases, as well as its current state[5][8]. It's critical to accurately predict the different types of arrhythmias because even a slight inaccuracy can be fatal. To accurately predict each of the four classes, the suggested system's positive predictivity and sensitivity are optimized. The goal of this study is to diagnose four different types of chronic arrhythmia using raw ECG signals, including Normal Beats, Supraventricular Ectopic Beats, Ventricular Ectopic Beats, and Fusion Beats. The existing model is improved by implementing a novel feature extraction method based on Convolutional Neural Networks with Eigen Vectors. Further defining the subclasses of arrhythmia will help patients in obtaining effective and timely treatment.

II. DATASET

Between 1975 and 1979, the raw ECG signals were obtained from the MIT-BIH arrhythmia database, the Beth Israel Hospital Arrhythmia laboratory[20]. Sixty percent of the recordings came from inpatients, remaining 40% of recordings were obtained from 47 outpatients.. Each recording is 30 minutes long containing 360 samples per second. with 25 men and 22 women between the ages of 23 and 89. ECG Signals are measured using two lead channels. i.e., MLII and V1. The participants ranged in age from 32 to 89 years old. The MIT-BIH dataset consists of ECG signals of 48 patients. First 23 patients are from normal heartbeats. Remaining 25 are from arrhythmia patients. It also consists of annotators such as reference beat, rhythm and, signal quality.

III. PROPOSED WORK

The architecture diagram of the overall proposed system is shown in Fig 1. It involves pre-processing of the raw signals, which includes reduction of noise and sampling the raw ECG signal, so as improve the classification metrics. This feature extraction technique uses only CNN's forward propagation by altering the functionality of the filters present in conventional CNN algorithm, eliminating the use of back propagation. The extracted features is then given to the model for classification of four classes of arrhythmia.
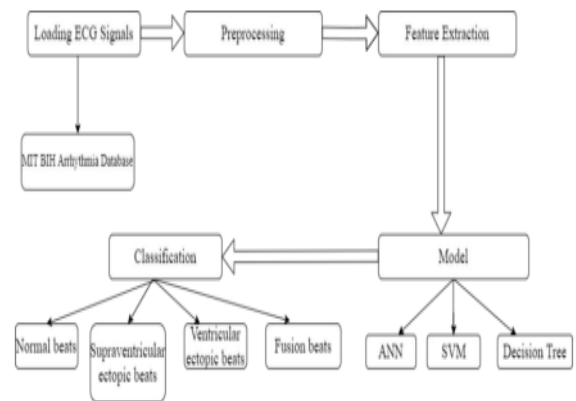
A. Pre-Processing Raw ECG Signals

The raw signals from each file having a format of .hea were extracted to a NumPy array using the WFDB package of python which is known as Wave Form Database. WFDB helps in reading the raw ECG signals and extracts the signals to corresponding NumPy arrays. It returns two values the NumPy arrays and the data annotated to the corresponding ECG signal. Since difference in the range of values of each signal were so high, normalizing the data was helpful as it scaled the signal values between the range   of -1 to 1.

B. Z-Score normalization

It is the process of converting every value in a dataset to a mean of 0 and a standard deviation of 1. The Z-score normalization can be calculated using equation (1)

$$Z = (x - \mu) / \sigma \tag{1}$$

Where x corresponds to score, μ corresponds to mean and σ is standard Deviation.

Then, the recordings were reshaped in such a way that they were to fit 30 minutes. The recording consisted of 360 samples per second, so the new reshaped array of each file consists of a 1800x360x2 matrix, which is our required 2d data.

C. Sampling

Since the dataset is imbalanced i.e., number of samples in each of 4 classes varies in high range. The dataset needs to be balanced for accurate prediction (or) classification. Here the total number of samples in each class are Normal beats (45845), Supraventricular beats (943), Ventricular beats (3788), and Fusion beats (415). The accuracy obtained from training any model with this dataset is not reliable as the model is more biased towards Normal beats. This problem is subdued by sampling techniques. There are two types of sampling which are used, down-sampling the normal beats and up-sampling the rest, which is achieved using different techniques.

a) Up sampling by using Synthetic Minority Over-sampling Technique (SMOTE) and duplication of minority class data.

b) Down sampling the normal beats.

IV. FEATURE EXTRACTION

Figure 2, shows the Single Layer of CNN as [2][6] with Eigen Vectors[11][18], the proposed system is 1D block is made up of a configurable number of filters, each of which has a fixed size; the vector and the filter are convoluted, yielding a new vector with the same number of channels as the number of filters as an output. To introduce nonlinearity, each value in the tensor is fed through an activation function. When executing convolutions, the default is to shift filters of a certain Width by one element at a time; this is known as Horizontal stride, and it can be changed by the user. The output vector will be smaller as the Conv1D Eigen Value stride increases. This can be used to reduce the number of parameters and the amount of memory needed, but it results in data loss.
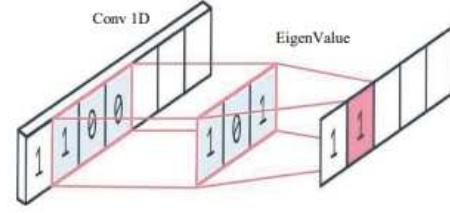

Fig. 2.   Single layer Eigen vector Extraction

A novel feature extraction method is proposed involving CNN's [16] forward propagation eliminating the use of backpropagation which increases the efficiency by reducing the time required to run the traditional CNN algorithm for feature extraction. The pre-processed data is given to this model which extracts the features of the signal and also reduces the number of channels[15] as the effect of Eigen vectors calculation. Three layers of this model are used consecutively to extract the maximum number of dependent features, corresponding to each class. The sample points which contain maximum features are projected using eigenvalues during each iteration of CNN forward propagation. Since the signals are 1 dimensional, convolution 1 d is used in this process of feature extraction. The purpose of filter matrix which is usually learned using backpropagation is eliminated. The filter matrix size is alone used to find the maximum occurring features in each filter range. This is done using the finding of the covariance matrix of the filter size matrix. The covariance matrix helps us to find the relationship between each value in the matrix. Covariance matrix is then used to find the eigenvalues and eigenvectors, which provides us the values of highly projected features. The eigenvalues are now arranged in descending order to find the major features affecting the different classes. Eigen values help in transforming the data into a set of most important features. The resultant data is then sent to the model for training.

Algorithm 1: Feature extraction with eigen vector

Step 1: Start

Step 2: Initialize the hyper parameters - filter size, padding and strides

Step 3: Compute the resultant feature size. Create a matrix to hold the output features

Step 4: Repeat the process for the entire dataset

Step 5: Copy the ith data signal to a variable 'a' to perform computations in that data

Step 6: Loop through n filters:

    Step 6a: Loop through C channels of the data

    Step 6b: Initialize start and end points of the filters

    Step 6c: Copy the sliced contents of the ith data to a variable (a)

    Step 6d: Reshape the variable (a) matrix to calculate the eigen values

    Step 6e: Compute the covariance matrix

    Step 6f: Compute Eigen values and Eigen vectors from the covariance matrix

    Step 6g: Sort the Eigen vectors and values in descending order

    Step 6h: Project the Eigen vectors

    Step 6i: Remove the Features that are less projected

Step 7: Stop

V. Implementaion

A. Ensemble Method

An Ensemble method [9][19] is used to improve the performance of the existing works since the positivity and the sensitivity scores of all the 3 classes other than Normal beats are very poor. To achieve this, three different models are trained and the predictions from each model is considered to classify each signal so that the model stays unbiased. This approach seems to provide better results in terms of sensitivity and positive predictivity than other individual models by itself. The Three models chosen for this method are SVM, ANN +SMOTE and ANN, which proved to be better solitarily.

In algorithm 2, the ensemble model initially creates two lists one to store the predictions of each method (Values list) and other to store the final prediction values (final prediction list). Then it loops over each prediction from all the three methods and save it in the values list, which is then compared to find the maximum occurring or highest voted prediction. Finally, the predictions are stored in the final prediction list and the values list is cleared and kept ready for the next iteration.

Algorithm 2: Ensembling method

- Start
- Initialize an empty list named values to store the predictions from each iteration
- Create an empty list to store the final predictions of the size same as the size of the testing set
- Loop over each prediction:
- Append the values of all the three models to the values matrix
- Find the maximum occurring prediction from the Values list and store it in the final predictions list
- Clear the Values list to store the values of the next iteration
- End

B. K-Fold Cross-Validation

K-fold cross-validation divides the data into k subgroups. The holdout method is now performed k times, with each time one of the k subsets acting as the test or validation set, and the other (k-1) subsets as the training set. This considerably minimizes bias because the majority of the data is utilized for fitting, as well as variance. After all, the majority of the data is also used in the validation set.

C. Subclass Classification Of Arrhythmmia

Classifying the different sub-classes of arrhythmia[4] plays a vital role in understanding the severity of the disease. There are 11 sub classes of arrhythmia which are being classified as shown in Fig 3. The process of classification of these subclasses is similar to that of the main class classification, using the Eigen Vectors in CNN and K-fold cross validation since it yielded the best performance metrics. The Normal class include 3 subclasses namely, Normal beats, Left Bundle Branch Beat and Right Bundle Branch Beat. The Supraventricular class consists of Atrial Premature Beats [3], Aberrated Atrial Premature beats, Nodal Escape beats, Nodal Premature Beat. The premature

ventricular and Ventricular Escape Beats are from Ventricular class of arrhythmia.
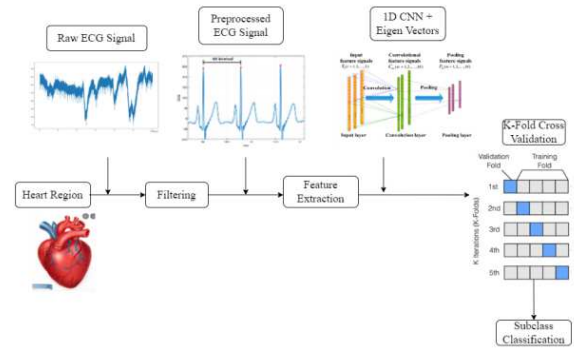

Fig. 3. Architecture Diagram of Sub class Classification

The sub class classification process begins with filtering raw ECG signals using band pass filter to reduce any noise present in the signal. The data is then normalized to minimize the range of peaks and lows of the ECG signals. The normalized data is then up-sampled using duplication since the dataset was more biased to the normal beats. The data is then passed to the feature extraction phase where the model CNN[14][13]-Eigen Vectors learns to differentiate the various signal types of arrhythmias. Finally, the output of feature extracted model is given to the K-fold cross validation method consisting of 10 folds, which uses 4 layers of dense neural networks for classification of the subclasses.

VI. Results And Discussion

A. Results of Major Class Classification Of Arrhythmia

Since the dataset is biased towards the single class (Normal beats), accuracy cannot be considered as a metric even though it achieves an accuracy of 96.75 %. Hence sensitivity and positive predictivity are taken as the metrics which will provide the most accurate evaluation of the trained model.

- Positive Predictivity

Precision is also called as Positive Predictivity. The performance of the classifier is measured by using various metrics such as precision, recall, F1 score and support. Precision measures the ability of a classifier to identify only the correct instances for each class it can be calculated using Equation (2).

$$Positive\ Predictivity = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2)$$

- Sensitivity

Recall is also called as Sensitivity. It measures the ability of a classifier to find all correct instances per class which is calculated using Equation (3).

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3)$$

- F1-score

F1 score combines both positive predictively and sensitivity. It is defined as the harmonic mean of Precision and Recall as shown in Equation (4). Since it gives equal

weightage to both positive predictivity and sensitivity, it is not biased.

$$F1\ score = 2 * \frac{Precision * Recall}{Presicion + Recall} \quad (4)$$

The results obtained from ensemble method using the Eigen vectors in CNN feature extraction method is depicted in Table I. Even though the model performs well on the imbalanced data, the sensitivity of the supraventricular beats seems to be low making the model inutile.

TABLE I.   RESULTS FROM ENSEMBLE METHOD

Class	Precision	Recall	F1-Score
Normal Beats	89.84%	98.83%	94.12%
Supraventricular Ectopic Beats	86.42%	24.73%	38.46%
Ventricular Ectopic beats	88.52%	80.10%	84.10%
Fusion Beats	99.27%	91.28%	95.27%

The precision and recall values obtained 7Th fold of the 10-fold cross validation model using the Eigen Vectors in CNN feature extraction method is better when compared to existing models as shown in Table II.

TABLE II.   RESULTS OBTAINED FROM K-FOLD CROSS VALIDATION

Class	Precision	Recall	F1-Score
Normal Beats	99.51%	99.32%	99.42%
Supraventricular Ectopic Beats	84.15%	89.47%	86.73%
Ventricular Ectopic beats	96.11%	98.14%	97.12%
Fusion Beats	83.78%	73.81%	78.48%

B. Results of Sub Class Classification Of Arrhythmia

- *Confusion Matrix*

The confusion matrix obtained from subclasses classification of Arrhythmia using the novel feature extraction method and K-fold cross validation is shown in figure 4. Precision and sensitivity values of each sub class is presented in Table III.
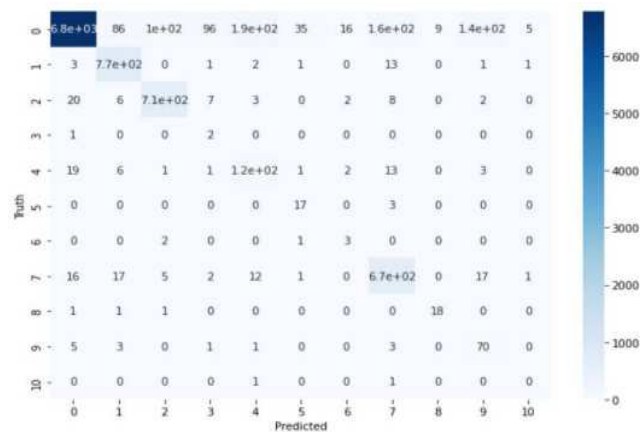

Fig. 4.   Confusion Matrix of subclass classification

TABLE III.   RESULTS OF SENSITIVITY AND POSITIVE PREDICTIVITY OF SUB CLASS CLASSIFICATION OF ARRHYTHMIA

Subclasses	Precision	Sensitivity
Normal Beat	99.05%	88.97%
Left Bundle Branch Beat	86.58%	97.21
Right Bundle Branch Beat	86.34%	86.34%
Atrial premature beat	18.1%	66.66%
Aberrated atrial premature beat	35.47%	71.60%
Nodal (junctional) escape beat	30.35%	85.00%
Nodal (junctional) premature beat	13.04%	50.00%
Premature ventricular contraction	76.81%	90.36%
Ventricular escape beat	66.66%	85.71%
Fusion of ventricular and normal beat	30.30%	84.33%

The ROC curve as shown in Figure 5, proves that the model performs well when compared to a random classifier.
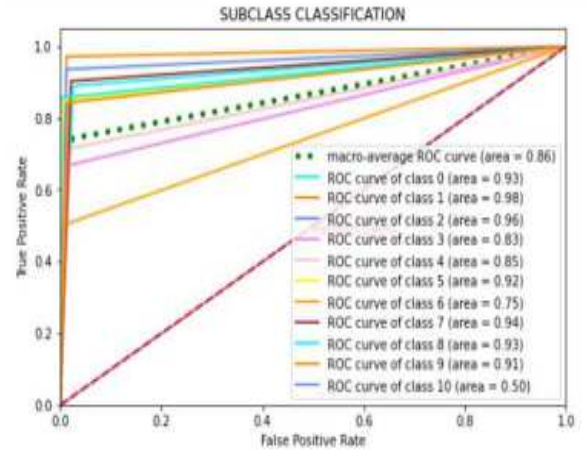

Fig. 5.   ROC Curve for Subclass Classification using K-fold

VII. CONCLUSION AND FUTURE WORK

The primary objective of the proposed work is to predict and classify different types of arrhythmias using raw ECG signals efficiently. To optimize the metrics considered for evaluation and to predict each class with at most precision, a novel feature extraction method is proposed using CNN. For training the model, K-fold cross validation method is used which proved in optimizing the metrics, as every data is being presented to the model in either validation or training set during each fold. The seventh fold of this process gave the highest result in which each class is predicted more efficiently with an average precision of 90.89%, average recall of 90.19% and average F1-Score of 90.44% when compared to that of ensemble method 91.01%, 73.74% and 77.99% respectively. This result also outperforms the existing works in classification of Arrhythmia. Classifying these sub classes might help the doctors even more in finding the problems of the patients accurately and treating them with significant medicines that will cure the patients earlier than expected. There are 11 sub classes predicted in this paper in which some are acute and some are chronic. Thus

classifying and predicting them in their initial states will help the patients to get timely treatment.

The metrics corresponding to the unclassified beats (Q) seems to be low and should be increased in future work. The prediction of Arrhythmia before its occurrence will make a huge impact in the patient's life in curing and preventing any cardio vascular disease well in advance. This prediction of impending arrhythmia might create a huge difference in modern science helping arrhythmia patients to undergo timely and accurate treatment.

REFERENCES

[1] A. Ekhlasi, N. Moieni, H. Ahmadi and A. Ekhlasi, "Analysis of ECG signals to classify abnormal patterns by employing Artificial Neural Network and Discrete Wavelet Coefficients," 2020 6th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), pp. 1-5,2020.

[2] C. Sarvan and N. Özkurt, "ECG Beat Arrhythmia Classification by using 1-D CNN in case of Class Imbalance," 2019 Medical Technologies Congress (TIPTEKNO), pp. 1-4, 2019.

[3] D. Lai, Y. Bu, Y. Su, X. Zhang and C. S. Ma, "Non-standardized patch-based ECG lead together with deep learning-based algorithm for automatic screening of atrial fibrillation," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 6, pp. 1569–1578, 2020.

[4] Sanamdikar, S.T., Hamde, S.T. & Asutkar, V.G. Analysis and classification of cardiac arrhythmia based on general sparsed neural network of ECG signals. SN Appl. Sci. 2, 1244 (2020). https://doi.org/10.1007/s42452-020-3058-8

[5] D. Verma and S. Agarwal, "Cardiac Arrhythmia Detection from Single-lead ECG using CNN and LSTM assisted by Oversampling," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 14-17, 2018.

[6] E. İzci, M. Değirmenci, M. A. Özdemir and A. Akan, "ECG Arrhythmia Detection with Deep Learning," 2020 28th Signal Processing and Communications Applications Conference (SIU), pp. 1-4, 2020.

[7] Fabrice Tueche, Youssoufa Mohamadou, Arsene Djeukam, Landry Cabrel Njike Kouekeu, Rene Seujip, and Merlin Tonka," Embedded Algorithm for QRS Detection Based on Signal Shape", IEEE Transactions on Instrumentation and Measurement, Vol. 70, pp. 1-12.2021.

[8] I. B. Salah, R. D. Rosa, K. Ouni and R. B. Salah, "Automatic diagnosis of valvular heart diseases by impedance cardiography signal processing," Biomedical Signal Processing and Control, vol. 57, pp. 101758, 2020.

[9] J. Xu and J. Zhang, "A Heterogeneous Ensemble Learning based Method for ECG Classification," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), pp. 439-443, 2019

[10] M. Gala, J. Barabbas and J. Krajnak, "Robust QRS complex detector algorithm based on modified pan-tompkins method and wavelet transform," in Proc. Int. Conf. on Telecommunications and Signal Processing, Milan, Italy, pp. 633–636, 2020.

[11] M. Kallas, C. Francis, L. Kanaan, D. Merheb, P. Honeine and H. Amoud, "Multi-class SVM classification combined with kernel PCA feature extraction of ECG signals," 2012 19th International Conference on Telecommunications (ICT), pp. 1-5, 2012.

[12] Mohamed Hammad, Abdullah M. Iliyasu, Abdulhamit Subasi, Edmond S. L. Ho, and Ahmed A. Abd El-Latif," A Multitier Deep Learning Model for Arrhythmia Detection", IEEE Transactions on Instrumentation and Measurement, Vol. 70, pp. 1-9,2021.

[13] M. Thirugnanam and M. S. Pasupuleti, "Cardiomyopathy-induced arrhythmia classification and pre-fall alert generation using convolutional neural network and long short-term memory model," Evolutionary Intelligence, vol. 14, no. 2, pp. 789–799, 2021.

[14] Iyer, T.J., Kishan, B., Nersisson, R. (2021). Prediction and Classification of Cardiac Arrhythmia Using a Machine Learning Approach. In: Komanapalli, V.L.N., Sivakumaran, N., Hampannavar, S. (eds) Advances in Automation, Signal Processing, Instrumentation, and Control. Lecture Notes in Electrical Engineering, vol 700. Springer, Singapore. https://doi.org/10.1007/978-981-15-8221-9_53

[15] Ruxin Wang, Jianping Fan, and Ye Li," Deep Multi-Scale Fusion Neural Network for Multi-Class Arrhythmia Detection", IEEE Journal of Biomedical and Health Informatics, Vol. 24, no. 9, pp. 2461-2472, Sept. 2020

[16] Wu M, Lu Y, Yang W and Wong SY (2021) A Study on Arrhythmia via ECG Signal Classification Using the Convolutional Neural Network. Front. Comput. Neurosci. 14:564015. doi: 10.3389/fncom.2020.564015

[17] Wenliang Zhu, Xiaohe Chen, Yan Wang, Lirong Wang," Arrhythmia Recognition and Classification using ECG Morphology and Segment Feature Analysis", IEEE Transactions on Computational Biology and Bioinformatics, Vol. 16, no. 1, pp. 131-138, 1 Jan.-Feb. 2019.

[18] Y. Yeh, T. Chiang and H. Lin, "Principal Component Analysis Method for Detection and Classification of ECG Beat," 2011 IEEE 11th International Conference on Bioinformatics and Bioengineering, pp. 318-322, 2011.

[19] Z. Sun, C. Wang, Y. Zhao and C. Yan, "Multi-Label ECG Signal Classification Based on Ensemble Classifier," in IEEE Access, vol. 8, pp. 117986-117996, 2020

[20] Dataset - https://physionet.org/content/mitdb/1.0.0/