Listen to Your Heart: Arrhythmia Detection using Machine Learning

Presish Bhattachan
Department of Systems Design Engineering
University of Waterloo
Waterloo, Canada
p4bhatta@uwaterloo.ca

Ilir Lazoja
Department of Systems Design Engineering
University of Waterloo
Waterloo, Canada
ilazoja@uwaterloo.ca

Kevin Qiu
Department of Mechanical and Mechatronics Engineering
University of Waterloo
Waterloo, Canada
longlai.qiu@uwaterloo.ca

Abstract— **Electrocardiogram (ECG) signals are commonly used in the medical field to obtain information on patient heart health and allow doctors to make diagnosis of cardiovascular diseases. Arrhythmia, a manifestation of such diseases, describes an irregular heartbeat and often leads to more serious conditions. With the aging population, heart related death tolls are expected to rise and early detection of arrhythmia can mitigate such concern. However, accurately detecting arrhythmia is often a difficult task amongst existing machines and health care practitioners. To address this problem, the work in this paper applies various machine learning techniques to ECG signals from four different classes. Four different classifiers including SVM, random forest, MLP and CNN are evaluated based on their precision, recall, and f1-score at classifying normal heartbeats and 3 different arrhythmia types according to the AAMI EC57 standard. The data was acquired through Physionet's MIT-BIH Arrhythmia Dataset and used two feature extraction methods as inputs to SVM and random forest. The results from this study indicate that convolutional neural networks are the best method in classifying the heart arrythmias, providing high scores in precision and recall, ultimately resulting in the highest f1-score as well.**

I. INTRODUCTION

Electrocardiography is the most common method in diagnosing cardiac disease and provides critical information about heart health and pathology. Carefully analyzing electrocardiogram (ECG) signals, allows doctors to diagnose patients with heart diseases such as myocardial infarction, ventricular tachycardia and atrial fibrillation [1]. The implications of such diseases pose a serious societal issue due to its high prevalence, mortality rates and cost of treatment. Each year, 17.3 million people die from cardiovascular diseases which accounts for 37% of all global deaths [2]. The aging population will be of the highest risk and as a result, death toll may rise significantly in the coming decade.

Heart arrhythmias, caused by irregular electrical signals in cardiac tissue, are important manifestations for such diseases. In fact, early detection of arrhythmia is critical in preventing more severe heart related problems and allows for appropriate treatment sooner. Detecting arrhythmia was often a challenging task for computers to interpolate and required expert cardiologists for secondary review. Previous literature found that only 50% of computer predictions were accurate in diagnosing for non-sinus rhythms [3] and only 1 in 7 atrioventricular block signals were correctly identified [4].

Traditional algorithms for arrhythmia recognition are based on morphological features of few QRS complexes. These methods encompass the use of noise filtering of ECG signals [5], signal segmentation [6] and feature extraction [7]. While such methods are commonly popularized in literature, these algorithms must recognize distinct relationships between various wave types. This is more error prone due to heart beat variability amongst individuals and the presence of noise [8].

To address such limitations in detecting arrhythmia accurately, the use of recent machine learning techniques such as deep learning has shown promising results. Specifically, image processing was the first field where deep learning was introduced which surpassed existing approaches by a large margin. Since then, deep learning has been incorporated into several other similar architectures such as the use of detecting arrhythmia from ECG signals.

In this paper, we apply different machine learning techniques used in literature on the MIT-BIH arrhythmia dataset. The motivation behind our work is to compare and contrast different machine learning methods to one another and evaluate the most effective method to use for identifying individuals with heart arrhythmia. The proposed scheme will analyze four different classifiers; MLP, CNN, SVM and random forest.

II. RELATED WORK

The recent rise of machine learning approaches in detecting arrhythmia from ECG signals has been quite successful. The proposed techniques range from simple classifiers such as Linear Discriminant Analysis (LDA) or decision trees to more sophisticated ones such as deep learning, Support Vector Machines (SVM) and Multi-Layer Perceptrons (MLP). The results have been so impressive that Hannun et al. in [1] proposes an approach that was able to achieve cardiologist level detection using a convolutional neural network (CNN). The 34-layer network learns to classify twelve different arrhythmia types in approximately 65,000 unique ECG recordings from

30,000 patients. The network relies on methods discussed in related literature such as the use of residual connections found in Residual Network architectures [9] and the use of batch normalization [10] between each convolution layer. The evaluation criteria compares two different metrics against the cardiologist committee observations denoted as the ground truth. This benchmark assesses the accuracy of the model using the F1 score for each sequence (roughly once per second) and for each set (every 30 seconds). The model with the best results outperforms a committee of cardiologists in both recall and precision for a given test set.

Another method commonly found in literature used to classify arrhythmia are SVMs. Nanjundegowda et al. in [11] introduces a method using a binary SVM classifier to detect normal and abnormal arrhythmia based on feature extraction in a dataset. The paper utilizes multiple extraction techniques to obtain subsets of both linear and non-linear features. This is accomplished using Principal Component Analysis (PCA), Kernel Independent Component Analysis (KICA) and Higher Order Spectrum (HOS). The linear extracted features allow for more optimal results under noise free conditions, while the non-linear subset contains hidden information that is able to achieve greater performance with noisy data. The proposed method evaluates performance based on three criteria; specificity, sensitivity and accuracy. The experimental results indicate improvements up to 1% over existing methods.

Similarly, Osowski et al. in [12] uses both HOS and Hermite characterization for feature extraction and employs a weighted voting principle to combine the two preprocessing methods into a SVM classifier. The weights of each neural classifier are then adjusted based on the Least Mean Squares (LMS) method. The combined classifier demonstrates an improved 1-2% average testing error over its individual respective method.

Park et al. in [13] applies a random forest classifier on features segmented from extracted QRS-complex. This was carried out using a Pan-Tompkins algorithm which performs a series of filtering, differentiation, squaring and integration over a moving window. The extracted features contain differences in amplitude between PQ, QR and RS signals. This feature set is then inputted into a random forest decision tree that is desirable to noisy data. The paper compares the use of five different classifiers and assesses performance based on accuracy, sensitivity and specificity. The results indicate that of the five classifiers evaluated, random forest achieved the best results for a one hundred random tree in all three metrics.

III. DATASET DESCRIPTION

The dataset used in this paper is from Physionet's MIT-BIH Arrhythmia Dataset which is deemed as a reputable source found amongst scientific literature [14]. The dataset has been used extensively for heartbeat classification and is also found in aforementioned papers in [11]-[13]. The dataset stemmed from a joint collaboration between Boston's Beth Israel Hospital (now Beth Israel Deaconess Medical Center) and Massachusetts Institute of Technology in 1980 [14]. The collection from the collaboration was the first readily available source of testing material for arrhythmia detection and has soon expanded to numerous other cardiovascular related research at over 500 sites worldwide [14].

The dataset contains 48 half-hour excerpts of two channel ECG recordings, taken from 47 different patients over a span of 4 years [14]. The heartbeats themselves were annotated by doctors and physicians, and recorded at 360 samples per second.

There are two versions available and will be used in this project. One version is the raw ECG data series for each patient. The second dataset is derived from the first dataset and contains preprocessed ECG signals segmented into heartbeats [15][16]. Each heartbeat has been downsampled to 125 Hz, cropped and padded to the fixed dimension of 188 [16]. Each heartbeat is labeled according to the Association for the Advancement of Medical Instrumentation (AAMI) EC57 standard [15]. Each heartbeat is grouped into these 5 classes.

['N': 0, 'S': 1, 'V': 2, 'F': 3, 'Q': 4]

N: Normal beat
S: Supraventricular premature beat
V: Premature ventricular contraction
F: Fusion of ventricular and normal beat
Q: Unclassifiable beat

The Q classification will be excluded from the dataset. Tables 1 and 2 demonstrates the breakdown of the original and preprocessed datasets.

TABLE I.   ORGINIAL DATASET

	N	S	V	F
Training	54,988	1171	4545	496
Testing	44,033	2050	3220	388
Total	99,021	3221	7765	884

TABLE II.   PREPROCESSED DATASET

	N	S	V	F
Training	72,471	2223	5788	641
Testing	18,118	556	1448	162
Total	90,589	2779	7236	803

IV. METHODOLGY

The proposed scheme evaluates the performance of four different classifiers; SVM, random forest, MLP, and CNN. The inputs of the SVM and random forest are features extracted from the raw patients' ECG signal from the MIT-BIH dataset for each heartbeat. Two feature extractors were used to segment features from the signal of each heartbeat and were evaluated individually.

An ECG signal is an asymmetric periodic signal that results from the electrical nodes in the heart. It results in a signal seen in Figure 1. The first feature set (FS1) consists of peak locations, peak heights, peak widths, peak prominence, and local max and min locations from the preprocessed segmented dataset. It finds the local maxima and minima of the heartbeat, which would correspond to the P, Q, R, S, T components of a heartbeat. Each heartbeat sample feature vector was zero padded to match the heartbeat sample feature vector with the most amount of peaks and local maxima and minima.
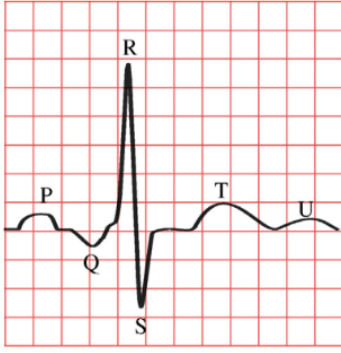
Fig. 1. Generic ECG Signal [19]

The second feature set (FS2) is more complex, and the use of features specific to the ECG signals were employed. The features were extracted from the original dataset, after being processed and segmented, using [17] and consists of wavelets, higher order statistics, local binary patterns, morphological descriptors and R-R intervals (104 features in total). A wavelet transform is used to extract features in both temporal and spectral frequencies. These transformations represent a signal as a combination of finite wavelets, rather than infinite sine and cosine waves like the Fourier transform, of varying width and amplitudes [18]. Higher order statistics, such as kurtosis and skew, are also advanced methods in describing ECG signals [19]. R-R intervals are the distance between two QRS peaks, and is another feature that is more exclusive to the ECG signal [19]. Finally, the distance from the R peak at specific fiduciary points, such as the location of the P wave, Q wave, S wave and T wave are calculated by finding the minimum and maximum values between specific time intervals in which those specific features of the ECG wave are found [19].

A. SVM

SVMs are hyperplanes that separate training data by a maximal margin tolerance. All vectors that lie on one side of the hyperplane are labelled as "1" or "0" and all instances closest to the hyperplane are referred to as support vectors. SVMs allow the original training data to be projected onto higher dimensionality through kernel selection.

The scikit-learn library was used to implement the SVM and random forest classifier. The SVM classifier decision function used was 'one-against-rest'. Different regularization parameters, C, were used (0.001, 0.01, 0.1, 1, 10, 100). The C which produced the best results on the validation set was kept. All other parameters were kept default from the scikit-learn library.

B. Random Forest

Random forest is a learning method that works by constructing multiple decision trees on a random subset of features and aggregating the unit vote of each tree to determine the most popular class. Random forest is advantageous to data that is prone to overfitting as the classifier averages multiple deep decision trees. The random forest parameters used were default from scikit-learn.

C. MLP

MLP is a class of feedforward artificial neural networks. An MLP architecture consists of at least three layers of nodes; an input, output and hidden layers. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. Each neuron is assigned an importance in the form of weights and biases. The layers of an MLP consist of multiple fully connected layers where each neuron in the current layer is connected to the one prior. The parameters of each neuron are independent from one another and possess a unique set of weights. The output of the network provides a class score.

The MLP architecture consists of 2 hidden layers each containing 50 neurons, with two Dropout layers (rate = 0.2) in between for regularization. The MLP was trained for 200 epochs.

D. CNN

CNN is a deep learning algorithm that follows a similar architecture to that of MLP. The main difference is that CNN have layer have hidden layers consisting of convolution and pooling used to extract features from an input. Convolution layers allow for the number of parameters to be minimized by having neurons share weights. Pooling layers can scale down on the number of parameters even further, this is more effective and easier to train than MLP. Finally, CNN layers are not required to be fully connected in which every node does not connect to every other node.

The proposed architecture for the CNN is shown in Figure 2 and includes a 1D convolution layer, followed by 5 blocks, each containing a convolutional layer with 32 kernels of size 5, a batch normalization function (which provides some regularization) with momentum 0.1 and epsilon 0.00001, an activation function (ReLU), and a max pooling layer of size 5 and stride 2. These blocks behave as a convolutional encoder. Once these blocks are traversed, a fully connected layer with 32 neurons, and an activation function in between each of the fully connected layers, followed by a softmax layer to predict the output class. The CNN was trained for 75 epochs.

Both the inputs to the CNN and MLP are the preprocessed segmented heartbeats and implemented using TensorFlow. The loss function was cross entropy loss from the softmax outputs. An Adam optimizer with default parameters was used for training and a model checkpoint was implemented which saves the model at the point with lowest validation loss.

Given that the dataset classes shown in Table 1 are imbalanced, it was decided that the use of accuracy as a metric was disregarded. Hence precision, recall (sensitivity), and f1-score were used as evaluation metrics for each approach. Furthermore, false positives and negatives are crucial when identifying arrhythmia, so these metrics are a better indication of how well each model performs as opposed to accuracy.
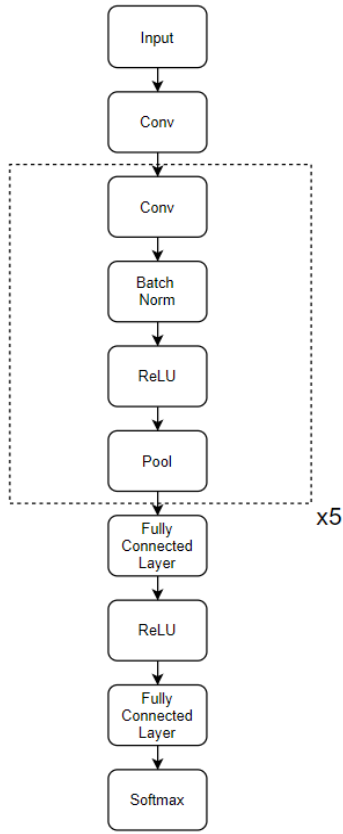
Fig. 2. CNN Architecture

V. RESULTS

Tables 3-8 display the confusion matrix and the results of precision, recall and f1-score for each proposed method. For both SVM and random forest, the results of two different feature sets are shown.

TABLE III. FS1 SVM (C=100)

	N	S	V	F	
N	17,927	79	104	8	
S	262	271	23	0	
V	299	9	1121	19	
F	63	0	17	82	
					Average
Precision	0.97	0.75	0.89	0.75	0.84
Recall	0.99	0.49	0.77	0.51	0.69
F1-Score	0.98	0.59	0.83	0.61	0.75

TABLE IV. FS2 SVM (C=100)

	N	S	V	F	
N	43,304	282	203	244	
S	1701	219	128	2	
V	246	15	2958	1	
F	317	0	68	3	
					Average
Precision	0.95	0.42	0.88	0.01	0.57
Recall	0.98	0.11	0.92	0.01	0.50
F1-Score	0.97	0.17	0.90	0.01	0.51

TABLE V. FS1 RANDOM FOREST

	N	S	V	F	
N	18,064	15	39	0	
S	347	190	19	0	
V	424	1	1014	9	
F	100	0	15	47	
					Average
Precision	0.95	0.92	0.93	0.84	0.91
Recall	1.00	0.34	0.70	0.29	0.58
F1-Score	0.98	0.50	0.80	0.43	0.68

TABLE VI. FS2 RANDOM FOREST

	N	S	V	F	
N	43,670	20	343	0	
S	1988	38	24	0	
V	503	2	2715	0	
F	352	0	36	0	
					Average
Precision	0.94	0.63	0.87	NaN	NaN
Recall	0.99	0.02	0.84	0	0.46
F1-Score	0.96	0.04	0.86	NaN	NaN

TABLE VII. MLP

	N	S	V	F	
N	18,069	17	28	4	
S	206	341	9	0	
V	104	3	1317	24	
F	37	0	12	113	
					Average
Precision	0.98	0.94	0.96	0.80	0.92
Recall	1.00	0.61	0.91	0.70	0.80
F1-Score	0.99	0.74	0.94	0.75	0.85

TABLE VIII. CNN

	N	S	V	F	
N	18,019	53	30	16	
S	95	450	9	2	
V	29	3	1402	14	
F	12	0	15	135	
					Average
Precision	0.99	0.89	0.96	0.81	0.91
Recall	0.99	0.81	0.97	0.83	0.90
F1-Score	0.99	0.85	0.97	0.82	0.91

The results of each method indicate that the CNN was most effective, with results of 91%, 90% and 91% for precision, recall and f1-score, respectively. Random forest using FS2 produced the worst results, as none of the signals were classified within the "F" class, resulting in zero true positives for that classification. This renders the calculation for precision, recall and f1-score as unattainable.

Almost all of the methods employed classified normal heart rhythms well. This is reasonable given that a regular ECG signal has many fiduciary points that are common amongst all members of the population. Precision for "S" class was always

higher than both the recall and f1-score. The "F" classification, which was a combination between ventricular ("V" class) and normal beat ("N" class), was challenging for SVM and random forest, but not for MLP and CNN. Feature set 1 and 2 had stark differences in terms of the effectiveness, as FS1 had greater precision and recall in both random forest and SVM, than compared to feature set 2. It may have been the larger complexity coming with FS2 that had resulted in the reduced precision and sensitivity.

The method that proves to be the most effective is CNN, due to the fact that the model is able to create and learn about temporal features of the ECG signal that are most appropriate for each classification. The training and validation curve for the MLP and CNN can be found in Figure 3 below. The CNN was able to learn the data better as it had a higher model capacity, however, it did slightly overfit the training data.
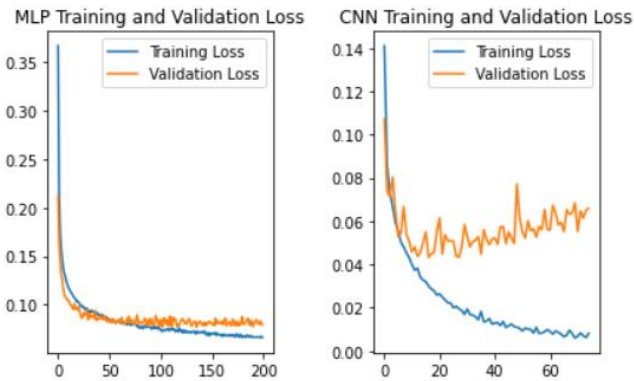

Fig. 3. Training and Validation Loss Curve for MLP and CNN

VI. CONCLUSION

In this paper, we compare various machine learning techniques to classify both normal and arrhythmia related heartbeats. Four different classifiers; SVM, random forest, MLP and CNN were applied towards the MIT-BIH dataset. It was discovered that CNN achieved the best results based on the evaluation metrics. The classifier achieved an average of 91%, 90% and 91% for precision, recall and f1-score, respectively. The implementation of such machine learning methods can extend past the use of arrhythmia detection. Within the healthcare industry, machine learning can also be used for drug discovery, medical imaging and even outbreak predictions.

REFERENCES

[1] A. Hannun et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network", Nature Medicine, vol. 25, no. 1, pp. 65-69, 2019. Available: 10.1038/s41591-018-0268-3.

[2] Heart Disease, Stroke and Research Statistics At-a-Glance; AHA; 2016

[3] A.P. Shah, S.A. Rubin, "Errors in the computerized electrocardiogram interpretation of cardiac rhythm", Journal of electrocardiology, 40(5):385–390, 2007.

[4] M.E. Guglin, D. Thatai, "Common errors in computer electrocardiogram interpretation", International journal of cardiology, 106(2):232–237, 2006.

[5] P. deChazal, M. O'Dwyer and R. Reilly, "Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features", IEEE Transactions on Biomedical Engineering, vol. 51, no. 7, pp. 1196-1206, 2004. Available: 10.1109/tbme.2004.827359.

[6] Y. Hu, S. Palreddy and W. Tompkins, "A patient-adaptable ECG beat classifier using a mixture of experts approach", IEEE Transactions on Biomedical Engineering, vol. 44, no. 9, pp. 891-900, 1997. Available: 10.1109/10.623058.

[7] C. Ye, B.V. Kumar, M.T. Coimbra, "Heartbeat classification using morphological and dynamic features of ECG signals", IEEE Trans. Biomed. Eng. 2012, 59, 2930–2941.

[8] E. Luz, W. Schwartz, G. Cámara-Chávez and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey", Computer Methods and Programs in Biomedicine, vol. 127, pp. 144-164, 2016. Available: 10.1016/j.cmpb.2015.12.008.

[9] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. Available: 10.1109/cvpr.2016.90

[10] S. Ioffe, C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", arXiv preprint arXiv:1502.03167, 2015.

[11] N, Raghu, Meshram, Vaibhav, "Arrhythmia recognition and classification using kernel ICA and higher order spectra", International Journal of Engineering and Technology. 7. 10.14419/ijet.v7i2.9535.

[12] S. Osowski, L. Tran, T. Markiewicz, "Support Vector Machine-Based Expert System for Reliable Heartbeat Recognition", IEEE Transactions on Biomedical Engineering. 51. 582-9. 10.1109/TBME.2004.824138.

[13] J. Park, S. Lee, K. Kang, "Arrhythmia detection using amplitude difference features based on random forest", 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). doi:10.1109/embc.2015.7319561

[14] G. Moody and R. Mark, "The impact of the MIT-BIH Arrhythmia Database", IEEE Engineering in Medicine and Biology Magazine, vol. 20, no. 3, pp. 45-50, 2001. Available: 10.1109/51.932724.

[15] M. Kachuee, S. Fazeli and M. Sarrafzadeh, "ECG Heartbeat Classification: A Deep Transferable Representation", IEEE International Conference on Healthcare Informatics (ICHI), 2018. Available: https://arxiv.org/pdf/1805.00794.pdf.

[16] S. Fazeli, "ECG Heartbeat Categorization Dataset", Kaggle, 2018. [Online]. Available: https://www.kaggle.com/shayanfazeli/heartbeat. [Accessed: 20- Apr- 2020].

[17] V. Mondejar, "mondejar/ecg-classification", GitHub, 2020. [Online]. Available: https://github.com/mondejar/ecg-classification/blob/master/python/features_ECG.py. [Accessed: 20- Apr- 2020].

[18] S. Sahoo, B. Kanungo, S. Behera and S. Sabut, "Multiresolution wavelet transform based feature extraction and ECG classification to detect cardiac abnormalities", Measurement, vol. 108, pp. 55-66, 2017. Available: 10.1016/j.measurement.2017.05.022.

[19] V. Mondéjar-Guerra, J. Novo, J. Rouco, M. Penedo and M. Ortega, "Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers", Biomedical Signal Processing and Control, vol. 47, pp. 41-48, 2019. Available: 10.1016/j.bspc.2018.08.007