IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

HealthyHeart: ML based Analysis and Predictionof Cardiovascluar Diseases

Karishma Gowda *Information TechnologyVESIT*
Mumbai, India

Nikita Makhija *Information TechnologyVESIT*
Mumbai, India

Hitesh Ochani *Information TechnologyVESIT*
Mumbai, India

Prof. Shanta Sondur
Information Technology VESIT
Mumbai, India

***Abstract*—Electrocardiogram or ECG waveform is one of the most common names that comes up in the healthcare sector. It is often the foremost step that is taken by the healthcare professionals when it comes to dealing with any cardiovascular issues. The signals from the ECG waveform are a reflection of how the heart functions. Any abnormalities encountered in the waveform are a reflection of underlying problems that a person may be suffering from. Using technology, especially Machine Learning and Artificial Intelligence in the healthcare industry can open up new doors and provide valuable assistance in the treatment of diseases. HealthyHeart is our proposed way that classifies CVD by using Random Forest Algorithm on the data obtained from the ECG signals. An accuracy of 97.7% was achieved by training the developed model over 979273 data entries. The accuracy rate indicates that HealthyHeart is a good fit for producing accurate results.**

***Index Terms*—ECG, Machine Learning, Healthcare industry, Cardiovascular diseases (CVD), Random Forest (RF)**

I. Introduction

Early detection and proper medication are the two pillars that can help a person tackle any health complications. There have been many unfortunate incidents where late detection of a disease cost a person their life.

As the technology is growing, its benefits can be seen in the healthcare industry too. The most promising breakthrough in recent times has been the levels that ML and AI have reached. Some work about the same has already been carried out and researchers and scientists are optimistic about ML and AIbeing the next most promising thing in the healthcare industry.When it comes to CVD or Cardiovascular diseases, usingML and AI based solutions for early detection and diagnosiscan be helpful. As per the surveys conducted globally, CVDshave the highest death rates; contributing to 31% of the deathsacross the globe. Around 85% of CVD affected patients die ofheart attacks and strokes that block their arteries and prevent proper blood flow throughout their body.

Categories that the CVDs can be classified into depend on each patient's condition and the symptoms that they show. Some of the categories of the CVDs include the following;

- Coronary heart disease
- Congenital heart disease
- Cerebrovascular disease
- Rheumatic heart disease
- Peripheral Arterial disease, etc.

Multiple methods in the medical field can help to detect CVDs by carrying out expensive and complicated tests. How- ever, irregular heartbeat is the most easily detectable and alarming symptom for it. ECG is known for recording the pattern in which the heart beats. It keeps a track of the rhythm and the rate at which the heart is beating. It is crucial to determine if there is any abnormality in the functioning of a person's heart. It is the fastest and the easiest way of detecting irregular heartbeats.

Any difference encountered in the ECG data can be a sign of an underlying cause. It is necessary to analyze and determine the cause of the encountered abnormality before the situation goes out of hand. Analyzing the ECG signals can help to understand if there are any narrowed or blocked arteries that can lead to chest pains or a heart attack.

Using ML and AI algorithms for this process can help to not only speed up the detection process but also to increase the precision of CVD diagnosis. By using data from the ECG wave forms with an efficient Machine Learning algorithm, time-efficient results that have an increased performance can be successfully determined.

The aim of the proposed system is to be of assistance in the early detection of CVD. The objective of choosing this project is to develop a ML based model that uses ECG data, analyse it, and gives accurate prediction about whether a person is at a risk of CVD.

II. Background

The proposed system aims to work for benefiting the healthcare industry by using ML based algorithms to deliver a system that can make the CVD detection process faster. ECG and the data that can be analysed from it is the most importantfactor in this system.

A. ECG Waveform

ECG signal records the rhythmic beating of the heart. This signal is divided into different wave forms or sections; each one having a particular role to play in determining the heart's health. These wave forms are used to determine the electrical intervals recorded during one heartbeat. The intervals are named as P, Q, R, S, T, and U.

As we know, the heart is made up of 4 chambers or sections; 2 out of those contribute to being the right and left atrium whereas the other 2 contribute to being the right and left ventricles. The divided sections of the waveform represent the functioning of the atria and the ventricles.

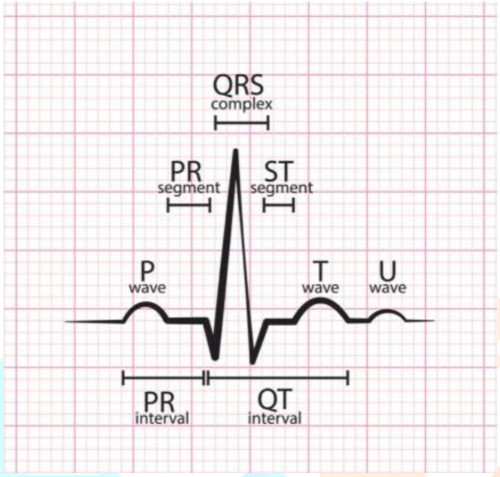

Fig. 1. An ECG Wave

The first short upward wave, labeled as the P wave indicates that the atria are contracting and pumping blood into the ventricles.

The next wave encountered is a complex that consists of the Q, R, and S intervals. Q interval begins with a downward deflection and then leads to the R interval, which is a peak. The peak then changes to the S interval which is the next encountered downward wave. Collectively, the QRS complex represents ventricular depolarization and contraction.

The PR segment of the waveform also contributes to representing the condition of the heart. It indicates the transit time that was needed for the electrical signal to move from the sinus node to the ventricles.

The T interval is next in line. It is an upwards waveform that represents ventricular re-polarization. A small wave may follow the T interval. This is labeled as the U interval and represents the small remnants of ventricular re-polarization.

The ST segment of the wave that was encountered after the QRS complex represents both the ventricles being completely depolarized.

Two intervals, PR and QT are measured. They both are of significant importance. The PR interval represents the time from the beginning of the atrial depolarization to the beginning of the ventricular depolarization.

The ideal values of all of these intervals should be as mentioned below.

- PR: 0.12 - 0.20 sec (3-5 small squares)
- QRS: 0.08 - 0.12 sec (2-3 small squares)
- QT: 0.35 - 0.43 sec

TABLE I
Normal Range of ECG Wave Intervals

Component	Characteristics
Heart Rate	60-100 bpm
PR Interval	0.12-0.20 sec
QRS Interval	0.06-0.10 sec
QT Interval	Less than half of R-R interval
ST Interval	0.08 sec

Table 1 shows the ranges of normal ECG wave segments.

The proposed system is built to classify diseases based on a person's heart rate. BPM or beats per minute is calculated from the ECG signal by using the P,Q,R,S,T, and U intervals. The specific range for this classification is as given below;

- Ventricular Tachycardia (150-250 bpm)
- Atrial Flutter (100-175 bpm)
- Sinus Bradycardia (less than 60 bpm)
- Atrial Fibrillation (150-200 bpm)
- Atrioventricular Block [AVB] (80-90 bpm)

These classified CVDs fall under Tachycardia, Bradycardia, and Arrhythmia.

B. Machine Learning

Machine Learning (ML) is the process by which a system can be trained to work and perform in a particular way. It is a subsection of Artificial Intelligence (AI).

Machine Learning is important for developing systems that can constantly learn and improve themselves to provide better results. By using Machine Learning and its different algorithms, a system can learn and improve itself without the need for any additional programs. Depending on the need, there are numerous algorithms and learning methods by which ML can be incorporated in a system.

As concerned with the healthcare industry, researchers are optimistic that ML based solutions can help to lay the foundation of better, more accurate, and efficient diagnosis of diseases. They are also inclined towards the idea of such technological solutions playing a major role in precise treatment of diseases.

III. Related Work

A data set having 8 columns that indicate the 8 param- eters of the ECG signal and 979273 rows having different values was used to build the model. An accuracy of 97.7% was achieved by applying Random Forest algorithm on the collected data. The achieved accuracy rate is an indication that the developed model is a good fit and will provide accurate results.

Various methods were thoroughly studied and analysed before the idea of HealthyHeart was put forth. During this study, one particular work that became the foundation for proposing this system is ’Analysis of Electrocardiograph (ECG) Signal for the Detection of Abnormalities Using MAT- LAB’ by Durgesh Kumar Ojha, Monica Subashini [16]. In the mentioned work, thorough analysis of the ECG signals was carried out to detect any underlying abnormalities in the functioning of the heart. The abnormalities in that system were detected by using MATLAB tool to plot the ECG wave from the given data. The plotted wave form was then compared to the ideal ECG wave form and the variations and abnormalitiesin it was detected.

’ECG arrhythmia classification using a 2-D convolutional neural network’ by Tae Joon Jun, Hoang Minh Nguyen, Daeyoun Kang, Dohyeun Kim, Daeyoung Kim, Young-Hak Kim [17] is another work that was an inspiration for coming up with the proposed system. In this work, 2-dimensional CNN were used for developing a system that recognised patterns.

IV. Proposed System

When it comes to CVD, early and accurate detection can help a patient in remarkable ways. It can help to speed up the medication process and make it easier for the healthcare professionals and the patient to tackle the disease effectively and in time.

A lot of necessary and crucial information about a person’s heart condition can be obtained from an ECG signal. This information when provided to an ML algorithm as input can result in the algorithm being an efficient CVD predictor.

HealthyHeart is one such system that makes use of the services that ML has to offer. Study of various algorithms was conducted and the one that proved to be the best fit for this initiative was chosen. A lot of training and testing has been carried out for the algorithm to give the most accurate results.

The proposed system is chosen after thorough research and an in depth literature survey that had shown some light on the methods that already exist to improvise and be of use in the healthcare industry.

The flow in which the system would work is as shown in Fig. 4.

1) The collected ECG data will undergo data pre-processing techniques that would refine the data at hand.
2) Required parameters and entities will be extracted/ computed from the refined data.
3) ML algorithms will be applied on this extracted data.
4) Predictions will be made by the ML algorithm.
5) CVD Classification will be carried out.

A. *Random Forest Algorithm*

Different ML algorithms were studied and implemented on a model to find out the one that would be the most efficient.

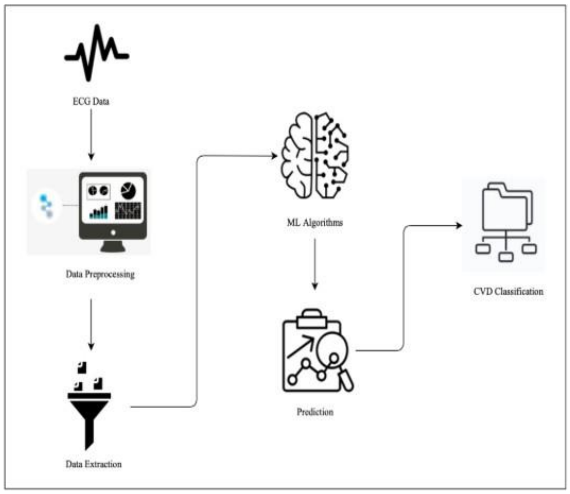

Fig. 2. Work Flow of the Proposed System

Random Forest was found to give the most accurate re- sults amongst them. RF algorithm belongs to the decision tree family itself. It is based on the concept of integrating multiple decision trees at a time to give an output that can be compartmentalised into a class.

Based on the class that a decision tree is trained to predict, multiple decision trees working together as Random Forest algorithm come up with multiple output. The output that is most accurate and similar to the condition of the input is then chosen as the final output of the algorithm. All of these steps take place within the algorithm itself.

When it comes to HealthyHeart and CVD classification, Random Forest is the algorithm that can give the best results.

B. *HealthyHeart : Phases*

The system works on the principle of thorough analysis and categorization. The work that was required to put the system together was distributed across 4 phases; each of them having an equal importance in developing the system. The phases of HealthyHeart include;

- Phase 1: Data Preparation
- Phase 2: Data Visualization
- Phase 3: Model Implementation
- Phase 4: Testing and Deployment

1) Phase 1: Data Preparation: Preparing the data that will be used to test and train the model is an integral step when it comes to ML. The model to be developed will be dependent on the way in which it is trained and tested. It is thus necessary to provide concise and abundant data to the model in its training phase.

For HealthyHeart, the data set was divided into a 70:30 ratio; 70% of this data set was used for the training phase of the model whereas the remaining 30% was used for the testingphase.

The intervals or the sections incorporated from the ECG wave form were analyzed and compared to the normal range of the those sections provided to the algorithm. Any abnormality encountered was used to predict the CVD and its class.

2) *Phase 2: Data Visualization:* When working with a data set, it is necessary to visualise and relate the parameters and entities in it with one another. Data visualization does just that. It is the graphical way of representing the data and the information that you are working with.

A lot can be understood and concluded by visualizing the data. Outliers, Patterns, Trends can be easily detected and studied with the help of this process.

For carrying out data visualization, the parameters obtained from the data set are plotted against each other in the form of graphs, maps, or charts and the co-relation or trend in them is found out.

For the HealthyHeart project, values obtained from the ECG data were plotted against each other. It made it easier to understand and work with the relationship that the parameters have with each other.

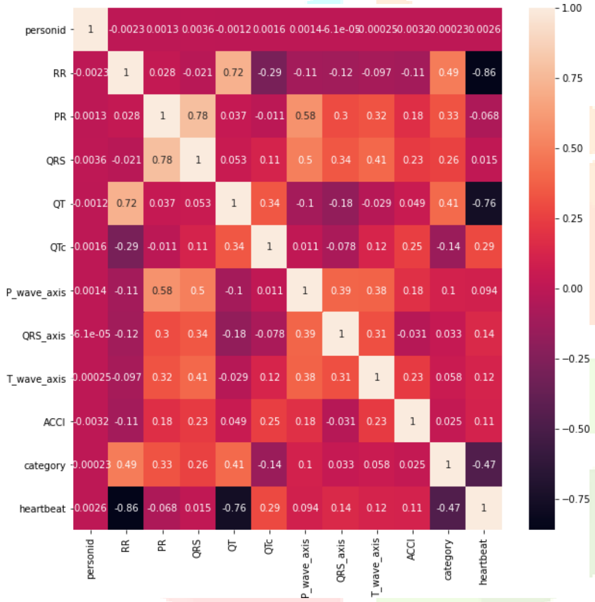

Fig. 3. Heat map for ECG data

The Heat map plotted for the ECG data shows the co- relation between its parameters. The segments of the ECG waveform PR, QRS, and QT are in co-relation. The distance between two consecutive RR segments is used to calculate beats per minute (bpm).

Visualization of data made it easier to compare the param- eters against the expected values of a normal ECG. Under- standing the implementation of the developed ML model also became easier with the help of data visualization.

3) *Phase 3: Model Implementation:* Once the data is pre-processed, cleaned, and visualised, the next step is to provide that data to the developed model. Training a model is a crucial step and should be given a lot of thought. Variations of data should be provided so that the model is well versed with all kinds of values that it can incorporate.

The Random Forest based model for HealthyHeart was implemented and trained with the 70 percent of data that was segregated for this process in the Data Preparation phase itself. Before going forward with Random Forest algorithm, vari- ous other algorithms were tried as well. The one that offered the most accurate results was then chosen for this Project.

Table 2 shows the Confusion Matrix of the ML algorithms that were implemented before choosing Random Forest for further integration in the system.

TABLE II
Confusion Matrix for ML algorithms

Parameters	SVM	AdaBoost	Random Forest
Accuracy	97.38%	79.83%	97.77%
Precision	95.27%	49.46%	97.01%
Sensitivity	91.50%	47.22%	91.69%
Specificity	98.86%	87.96%	99.29%
Error rate	2.61%	20.16%	2.21%
F1-score	93.34%	48.31%	94.27%
FPR	1.14%	12.03%	0.70%

As per the obtained numbers, the Random Forest model gives the most promising results. It offers an Error rate of 2.21% which is the least value in comparison to SVM model’s 2.61% and AdaBoost model’s 20.16%. The accuracy acquired for Random Forest model was around 97.77%; which is considered to be a good fit for a ML model. An accuracy range of less than 75% would be an under fit whereas a range of more than 98% would be an over fit.

4) *Phase 4: Testing and Deployment:* Testing a model after its implementation is necessary to make sure that the model is serving the purpose for which it was developed. The testing phase requires the model to be given data values that are not redundant to its training data. The parameters and the entitiesof the data will be the same but the range of values will vary.The remaining 30% of the data that was kept aside at the start of this project’s Data Preparation phase was used for testing the HealthyHeart model.

An Accuracy of 97.77% was achieved by implementing the Random Forest model. Besides that, Precision and Specificity of 97.01% and 91.69% respectively were achieved by the Random Forest model. Prior to this, tests were carried out on the SVM and the AdaBoost models as well. They gave an accuracy of 97.38% and 79.83% respectively. Both of these models were not suitable for the HealthyHeart system as the SVM model takes a lot of time to generate results whereas the AdaBoost model under-fits the system in terms of Precision and Sensitivity.

When compared, Random Forest out performs the other two models in every aspect. The most important feature being its time-efficiency. Thus, Random Forest was chosen as the best suitable option for the proposed system.

V. RESULTS

An accuracy of 97.77%, precision of 97.01% and sensitivity of 91.69% was achieved for this system. After carrying out in-depth research and analysis, it was confirmed that this accuracy is the best for the proposed system. The said accuracy will provide accurate results for the CVD classification. An accuracy less than the said percentage would have been an under-fit. On the other hand, aiming to achieve an accuracy greater than this percentage could over-fit the model and not provide precise results.

VI. CONCLUSION

A range of CVDs can be detected in-time and appropriate medication can be provided to the patients by using the proposed system.

With an accuracy of 97.77%, precision of 97.01% and sensitivity of 91.69%, this system that will provide accurate CVD detection results. This has been possible due to the study and integration of ML algorithms into this system.

Improvising and modifying this system as per the future needs can change the face of the Healthcare industry by leaps and bounds.

ACKNOWLEDGMENT

We would like to acknowledge the faculty of the Informa- tion Technology Department, Vivekanand Education Society's Institute of Technology [VESIT], Mumbai for their support and guidance.

REFERENCES

[1] M.Vijayavanan, V.Rathikaram, "Automatic classification of ECG signal for Heart disease diagnosis using morphological Features", International Journal of Computer Science and Engg.Technology, Vol 5 No.4, 2014, pp 449-455.

[2] Serkan Kiranyaz, Turker Ince, Jenni Pulkkinen, Moncef Gabbouj, "Per- sonalized long-term ECG classification" A Systematic approach in Expert Systems with Applications, 38 (2011) 3220–3226

[3] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks," ar X iv preprint ar X iv:1707.01836, 2017.

[4] Artis, S.G.; Mark, R.G.; Moody, G.B. Detection of atrial fibrillation using artificial neural networks. In Proceedings of the Computers in Cardiology, Venice, Italy, 23–26 September 1991; IEEE: Piscataway, NJ, USA, 1991; pp. 173–176.

[5] Salam, A.K.; Srilakshmi, G. An algorithm for ECG analysis of arrhyth- mia detection. In Proceedings of the IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 5–7 March 2015; IEEE: Piscataway, NJ, USA, 2015;pp. 1–6.

[6] Perez, R.R.; Marques, A.; Mohammadi, F. The application of supervised learning through feed-forward neural networks for ECG signal classifi- cation. In Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Vancouver, BC, Canada, 15–18 May 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–4.

[7] Gautam, M.K.; Giri, V.K. A neural network approach and wavelet analysis for ECG classification. In Proceedings of the 2016 IEEE International Conference on Engineering and Technology (ICE TECH), Coimbatore, India, 17–18 March 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1136–1141.

[8] Nilanon, T.; Yao, J.; Hao, J.; Purushotam, S.; Liu, Y. Normal/abnormal heart recordings classification by using convolutional neural networks. In Proceedings of the IEEE Conference on Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 11–14 September 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 585–588.

[9] Kiranyaz, S.; Ince, T.; Gabbouj, M. Real-time patient-specific ECG classification by 1-D convolutional neural network. IEEE Trans. Biomed. Eng. 2016, 63, 664–675. [CrossRef] [PubMed]

[10] Hian Chye Koh and Gerald Tan, "Data Mining Applications in Healthcare", Journal of Healthcare Information Management — Vol. 19, No.2

[11] Nilakshi P. Waghulde1, Nilima P. Patil, " Genetic Neural Approach for Heart Disease Prediction" International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume- 4 Number-3 Issue-16 September-2014.

[12] Jun TJ, Park HJ, Minh NH et al (2016). Premature ventricular contraction beat detection with deep neural networks. IEEE International Conference on Machine Learning and Applications pp 859-864

[13] Maas AL, Hannun AY, Ng AY (2013). Rectifier nonlinearities improve neural network acoustic models. International Conference on Machine Learning 30(1):3

[14] Rajni, I. Kaur, Electrocardiogram signal analysis-an overview. Int. J. Comput. Appl. 84(7), 22–25 (2013)

[15] M. Llamedo, J.P. Martinez, Heartbeat classification using feature selection driven by database generalization criteria. IEEE Trans. Biomed. Eng. 58(3), 616–625 (2011)

[16] Durgesh Kumar Ojha, Monica Subashini, Analysis of Electrocardiograph (ECG) Signal for the Detection of Abnormalities Using MATLAB. World Academy of Science, Engineering and Technology International Journal of Biomedical and Biological Engineering Vol:8, No:2, 2014

[17] Tae Joon Jun, Hoang Minh Nguyen, Daeyoun Kang, Dohyeun Kim, Daeyoung Kim, Young-Hak Kim, ECG arrhythmia classification using a 2-D convolutional neural network, Division of Cardiology, University of Ulsan College of Medicine, Asan Medical Center,Seoul, Republic of Korea