IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105160 |

Heart Disease Prediction using Machine Learning Model

Dr. Shivamurthy R C[*1], Dhanush M S [*2], Shaz Ahammed[*3], Kushi P Kogilvadi[*4], Roopali Singh R L[*5]

Head of the Department, Dept. ofComputer Science & Engineering, Maharaja Institute of Technology, Mysore, Karnataka, India[*1]

Undergraduate Student, Dept. ofComputer Science & Engineering, Maharaja Institute of Technology, Mysore, Karnataka, India[*2.3.4.5]

ABSTRACT: A common tool for cardiovascular diagnostics is the electrocardiogram (ECG). It takes substantial expertise to interpret ECG charts, and physically going through paper records can be time-consuming. These records can be automatically diagnosed and analyzed thanks to digitization. This project uses machine learning to transform paper ECG records into a 1-D signal. Techniques like wave extraction, lead division, smoothing, and binary picture conversion are used. The model is finalized using Principal Component Analysis for dimension reduction and a variety of classifiers, including KNN, Logistic Regression, SVM, and Voting Based Ensemble. This model helps with the diagnosis of cardiac conditions like myocardial infarction and irregular heartbeats as well as the assessment of general heart health

KEYWORDS: ECG, Voting based classifier, KNN, SVM,Logistic Regression

I. INTRODUCTION

According to the World Health Organization, heart disease is the first leading cause of death in the high and second leading cause of death in low-income countries. It has remained the leading cause of death at the global level for the last 20 years. Heart disease is a general phrase that covers a variety of ailments that have an impact on the heart and blood vessels. Heart failure, arrhythmias, coronary artery disease, and valvular heart disease are a few of the disorders that can range in severity. Heart disease is a prominent cause of death globally and is frequently brought on by lifestyle choices including smoking, eating poorly, exercising seldom, and being overweight. However, the onset of cardiac disease can also be influenced by genetics, ageing, and current medical problems including diabetes and hypertension.

Heart disease symptoms can vary, but common ones include chest pain, shortness of breath, exhaustion, disorientation, and swelling in the legs or feet. Early detection and diagnosis are essential for effective therapy, and physical examination, review of medical history, and diagnostic procedures such as ECG, echocardiogram, stress test, and cardiac catheterization are used to determine the presence of heart disease. Treatment options for heart disease include medication, lifestyle modifications, surgical treatments, and implantable devices. Risk reduction for heart disease depends heavily on prevention. A healthy lifestyle can help lower the risk of heart disease by promoting regular exercise, a balanced diet, abstaining from smoking and excessive alcohol use, and managing stress.

By learning to represent the world as a nested hierarchy of concepts, each concept is defined in relation to simpler concepts, and more abstract representations are computed in terms of less concrete conceptions, deep learning is a specific type of machine learning that reaches considerable power and flexibility. With the help of our model, we can help in quick diagnosis of Cardio-vascular diseases.

II. METHODOLOGY

Although there are numerous patients suffering from Cardio-vascular diseases worldwide, the number of digitized ECG's images publicly available online are small and scattered. That is why it is important that we select a validated dataset. which includes the patient's Scanned/Digitized ECG images and parameters required for the project. We propose a Data mining-based model to detect Cardio-vascular disease by making use of digitized ECG images. We pre-process these images and make the necessary adjustment by converting them into Gray-scale images and then dividing

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105160 |

them in separate 12 individual graph leads. We apply supervised classification models on these images extract their features and classify the images as either Normal, Myocardial infarction or Abnormal Heart beat. The chosen model is then deployed to a browser-based solution. User can upload the patient ECG image using the interface. This data is then fed to deep learning model working in the backend and then the prediction results are given with accuracy.

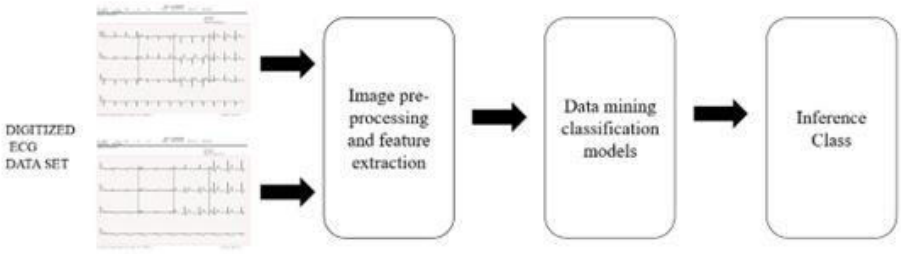

Figure 1 Workflow

Objective

- Acquisition of images, locating the ideal dataset for our purpose. Utilizing various image processing methods to prepare images
- Study and combine the performance of different models using performance metrics.
- To study various approaches used to detect Cardio-vascular diseases.
- To classify digitized ECG images into its sub-types viz: Normal, Myocardial infarction and Abnormal Heart beat using ensemble machine learning algorithms

III. MODELING AND ANALYSIS

The user uploads ECG images to our web app. From there, image conversion techniques such as rgb2gray conversion, denoising, Gaussian Filtering, thresholding and contouring are implemented to extract signals without the grid lines. The signal is then dimensionally reduced, and the necessary waves (P, QRS, T) are extracted using segmentation and fed into our pre-trained model from the analysis. Once the model finishes the analysis, it returns the results back to the user based on the findings

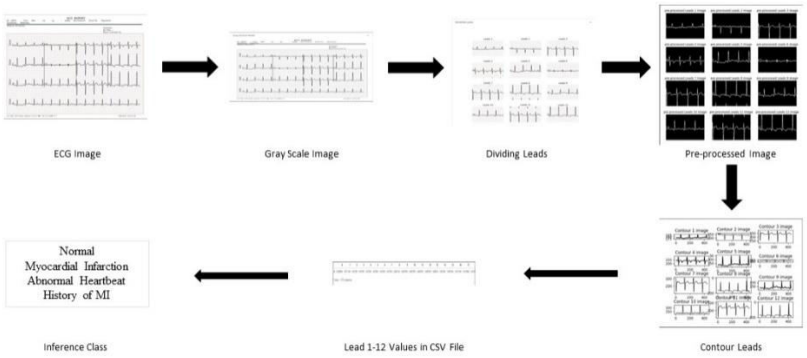

Figure 2 Steps in Classification

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105160 |

To prepare Leads (1-13) for further processing, each individual lead image is transformed by removing Gridlines, converting to Grayscale, applying Gaussian filtering, and performing Thresholding to convert to binary image.

The transformed image is traced to extract only the signals from the image using the contour technique, and the values are scaled using the MinMax Scalar. The normalized output is saved in CSV format as a 2D signal.

The 1D rows are converted into columns using transpose. With both 1D and 2D CSV files and cropped 1 to 13 lead images, we perform different Supervised classification algorithms: k-nearest neighbors (KNN), Logistic Regression, Support Vector Machine (SVM), and Voting Based Ensemble Classifier on based on CSV DATA.

Once we have extracted all of the image's 12 lead 1D values, we combine them into a single csv for further analysis.

Post Dimension reduction, following data mining techniques, are applied on 12 leads combined:

K-nearest neighbors (KNN):

K-nearest neighbors (KNN) algorithm is a type of supervised ML algorithm that can be used for both classifications as well as regression predictive problems. However, it is mainly used for the classification of predictive problems in the industry.

Logistic Regression:

Logistic regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes).

Support Vector Machine (SVM):

Support vector machines (SVMs) are powerful yet flexible supervised machine learning algorithms that are used both for classification and regression. But generally, they are used in classification problems. SVMs have their unique way of implementation as compared to other machine learning algorithms.

XGBoost:

XGBoost is an implementation of gradient boosted decision trees designed for speed and performance that is dominative competitive machine learning

Voting Based Ensemble Classifier with GridSearchCV:

Under Voting-based Ensemble classification, three Machine learning models like k-nearest neighbors (KNN), Support Vector Machine (SVM) and Random Forest Classifier are stacked and voted to pick one model which gives the highest accuracy. For tuning the hyperparameters we have used GridSearchCV

IV. RESULTS AND DISCUSSION

In this project, we perform different Supervised classification algorithms: k-nearest neighbors (KNN), Logistic Regression, Support Vector Machine (SVM), and Voting Based Ensemble Classifier based on CSV DATA. Before performing data modeling, ECG images categorically belonging to four categories of patients i.e., patients with Myocardial Infarction, Abnormal Heartbeat, Myocardial Infarction History, and good Health are combined on the lead level (from 1 to 12) and then convert target column with array (['No', 'HB', 'MI', 'PM'] into numeric using groups encoder. Post dimension reduction technique like Principal component Analysis is applied to understand the data and validate the variance explained is under acceptable limit.

Table 1 Comparison of different ML Model

SN.	Model	Accuracy
1	KNN	0.79
2	SVM	0,91
3	XGBoost	0.88
4	Logistic Regression	0,77
5	VotingBased	0.95

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105160 |

V. CONCLUSION

The empirical results show that we can produce faster and accurate predictions for heart patients by applying the given predictive model to the ECG images of new patients. This study can also be extended to include multiple different heart diseases if the feature extraction from images is done correctly and optimally along with increased accuracy of our model

REFERENCES

[1] Heart disease prediction using machine learning Techniques" by Apurva Garg1, Bhartendu Sharma2 and Rijwan Khan3 1,2Scholar, ABES Institute of Technology, Ghaziabad, Uttar Pradesh – 201009, India3 Professor, Department of Computer Science and Engineering, ABES Institute of Technology, Ghaziabad, Uttar Pradesh – 201009, India was published in the year 2021.
[2] "A Heart Disease Prediction Model using SVM-Decision Trees-Logistic Regression (SDL) by Mythili T., Dev Mukherji, Nikita Padalia, and Abhiram Naidu was published in the year 2013.
[3] "Heart Attack Analysis and Prediction using SVM" by Madhu H.K. and D. Ramesh was published by the year 2021.
[4] "A Heart Disease Prediction Model Based on Feature Optimization and Smote-Xgboost Algorithm" by Jian Yang and Jinhan Guan was published in the year of 2022.
[5] "Prediction of Heart Diseases using Support Vector Machine" by YamalaSandhya was published in the year 2020.
[6] "ECG Prediction Based on Classification via Neural Networks and Linguistic Fuzzy Logic Forecaster by Eva Volna, Martin Kotyrba, and Hashim Habiballa was published by the tear 2022.
[7] "Heart disease prediction using machine learning Algorithms" by Harshit Jindal1, Sarthak Agrawal1, Rishabh Khera1, Rachna Jain2 and Preeti Nagrath2 was published in the year 2021.
[8] "Analysis and classifcation of heart diseases using heartbeat features and machine learning algorithms" by FajrIbrahemAlarsan and Mamoon Younes2 was published in the year 2019.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details