JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

An Efficient Machine Learning Technique for Prediction of Cardiac Diseases using ECG Dataset

[1]Sameeksha Rahangdale, [2]Priyanka Singh, [3]Vinita Keer, [4]Taruna Saxena,

[1]Assistant Professor, [2]Assistant Professor, [3]Assistant Professor, [4]Assistant Professor
[1]Department of Electronics and Communication Engineering,
[2,3,4]Department of Computer Science and Engineering,
Lakshmi Narain College of Technology Excellence, Bhopal, India

Abstract : The heart is significant organ of human body part heart disease describes a scope of conditions that influence your heart. Machine learning (ML) is the scientific study of algorithms and statistical models that PC systems use to play out a specific task without using unequivocal instructions, depending on patterns and deduction instead. There are various algorithms, which predict the heart disease. Accuracy is a key parameter to judge the algorithm. This paper proposed the efficient machine learning technique based on random forest. Accuracy, classification error, F-measure, Recall and Precision parameters are calculated. 92% accuracy achieved by proposed algorithm.

Index Terms **– Machine learning, F-measure, Recall and Precision.**

I. INTRODUCTION

Heart disease is perhaps the biggest cause of grimness and mortality among the number of inhabitants on the planet. Prediction of cardiovascular disease is viewed as one of the most significant subjects in the section of clinical information analysis. The measure of information in the healthcare industry is tremendous. Information mining turns the huge assortment of crude healthcare information into data that can assist with making educated decisions and predictions.

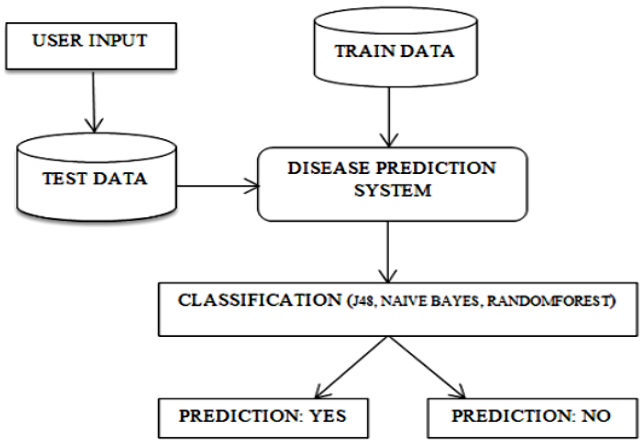

Figure 1: Basic architecture of prediction model

It comprises of planning of dataset and client commitment as the test dataset. Arrangement procedures in information mining are equipped for preparing a lot of information. It tends to be utilized to foresee straight out class marks and orders information dependent on preparing set and class names and it very well may be utilized for grouping recently accessible information. The term could cover any setting wherein some choice or figure is made based on by and by accessible data. Grouping techniques perceived strategy for over and over creation such choices in new circumstances. Here expect that issue is a worry with the development of a system that will be applied to a proceeding with succession of cases in which each new case must be appointed to one of a lot of pre characterized classes based on watched highlights of information.

A few learning calculations target finding better portrayals of the sources of info gave during training.[11] Great models fuse head segments examination and group investigation. Feature learning calculations, likewise called portrayal learning calculations, frequently attempt to save the information in their data yet also change it with the end goal that makes it valuable, regularly as a pre-handling step before performing order or forecasts. This technique permits remaking of the sources of info beginning from the obscure data creating dissemination, while not being essentially committed to arrangements that are unrealistic under that circulation. This replaces manual component structuring, and permits a machine to both get acquainted with the highlights and use them to play out a particular undertaking.

Highlight learning can be either regulated or solo. In managed part learning, highlights are discovered utilizing named input data. Models fuse counterfeit neural systems, multilayer perceptrons, and managed dictionary learning. In solo component learning, highlights are discovered with unlabeled data. Models consolidate word reference learning, independent portion investigation, autoencoders, organize factorization and different types of grouping.

Complex learning calculations attempt to do as such under the requirement that the insightful portrayal is low-dimensional. Meager coding calculations attempt to do as such under the requirement that the academic portrayal is scanty, inferring that the logical model has a colossal number. Multilinear subspace learning calculations expect to increase low-dimensional portrayals authentically from tensor portrayals for multidimensional data, without reshaping them into higher-dimensional vectors. Profound learning calculations find different degrees of portrayal, or a hierarchy of leadership of highlights, with increasingly noteworthy level, dynamically unique highlights described as far as (or delivering) lower-level highlights. It has been fought that a wise machine is one that learns a portrayal that unravels the essential elements of assortment that explain the watched data.

II. BACKGROUND

A. Kumar, et al. heart diagnosis at starting stage of the disease can lead to a successful cure of the disease. The classifications that are used to prevent the diagnosis are: - Naive Byes classification, the Support Vector Machine classification, and the k-NN classification. Hence, some deep learning techniques need to be tested on the dataset which has been taken and needs to be tested on it [1].

D. Dube, et al., presents works utilize various datasets, software tools, performance evaluation measures, prediction techniques, which are analyzed and the performance attained by different techniques are discussed and the research gaps for further contributions are identified in such a way that it enables the researcher to contribute much for the benefit of the society [2].

B. Wang et al., presents Renal brokenness, which is related with horrendous clinical results, is one of the most notable confusions of cardiovascular breakdown (HF). Advantageous expectation of renal brokenness can empower clinical staffs to intercede exactly on schedule to dodge calamitous outcomes. At the present time, is proposed a perform various tasks profound and wide neural system (MT-DWNN) for anticipating deadly intricacies during hospitalization [3].

T. S. Brisimi, et al., presents urban living in present day colossal urban communities has critical unfriendly impacts on wellbeing, expanding the danger of a few relentless infections. it is center around the two driving groups of interminable sicknesses, coronary illness and diabetes, and make data driven techniques to foresee hospitalizations as a result of these conditions. it is base these forecasts on the patients' clinical history, later and progressively far off, as portrayed in their Electronic Wellbeing Records (EHRs) [4].

A. Mdhaffar, et al., presents this work presents a novel wellbeing examination approach for cardiovascular breakdown forecast. It depends on the utilization of complex occasion handling (CEP) development, got together with factual methodologies. A CEP engine forms moving toward wellbeing data by executing edge based examination rules. Rather than having to truly set up edges, our novel factual figuring normally registers and updates edges as showed by recorded chronicled information [5].

J. Zhang et al., presents starting late, the utilization of insightful advancements in clinical dynamic in the telehealth condition have begun to accept a vital activity in improving the idea of patients' lives and decreasing the expenses and remaining task at hand related with their step by step medicinal services. At this moment, incredible clinical recommendation framework that utilizes a quick Fourier change coupled AI outfit model is proposed for transient sickness chance expectation to outfit unending coronary illness patients with legitimate proposals about the need to step through a clinical examination or not on the coming day dependent on analyzing their clinical data [6].

N. Alshurafa, et al., presents Far off wellbeing checking (RHM) frameworks are getting even more broadly got by clinicians and emergency clinics to distantly screen and talk with patients while smoothing out clinician time, diminishing medical clinic costs, and improving nature of care [7].

A. Khan et al., presented discusses some usually used machine learning techniques in Intrusion Detection System and conjointly reviews a number of the prevailing machine learning IDS proposed by researchers at different times. in this paper an experimental analysis is performed to demonstrate the performance analysis of some existing techniques in order that they will be used further in developing Hybrid Classifier for real data packets classification. The given result analysis shows that KNN, RF and SVM performs best for NSL-KDD dataset [8].

D. Tay, et al., presents myocardial dead tissue (MI) is one of the primary drivers of death in many made nations. Consequently, early disclosure of MI occasions is fundamental for incredible protection treatments, potentially lessening avoidable mortality. One philosophy for early ailment forecast is the utilization of hazard expectation models made utilizing AI methods. One huge fragment of these models is to outfit clinicians with the flexibility to tweak (e.g., the expectation range) and utilize the hazard forecast model that they considered generally profitable for their patients [9].

S. Nikolaiev et al., presents the perspective change from conceded interventional to Insightful, Preventive and Customized Drug is a fundamental overall test in the 21st century. Omnipresent association of convenient applications, new savvy sensors and AI strategies make conceivable creation of new age customized modified social insurance checking and pathologies area frameworks [10].

J. S. Sonawane et al., presents in clinical field the illness conclusion is regularly made dependent on the information and experience of the clinical pro. In light of this there are odds of mistakes, unwanted inclinations and moreover takes longer time in

exact analysis of sickness. If there should be an occurrence of coronary illness, its analysis is most inconvenient assignment. It relies upon the mindful examination of different clinical and hypochondriac data of the patient by clinical specialists, which is befuddled procedure. On account of movement in AI, PC and information advancement, the analysts and clinical professionals in gigantic degree are keen on the improvement of automated framework for the forecast of heart disease [11].

D. R. Patil et al., presents in clinical field the determination of coronary illness is most problematic errand. It relies upon the wary investigation of different clinical and over the top data of the patient by clinical specialists, which is puzzled procedure. Due to progress in AI and information development, the analysts and clinical specialists in enormous degree are keen on the improvement of motorized framework for the forecast of coronary illness that is astoundingly precise, fruitful and steady in early finding [12].

III. METHODOLOGY

The proposed research work can be understand by using following flowchart-

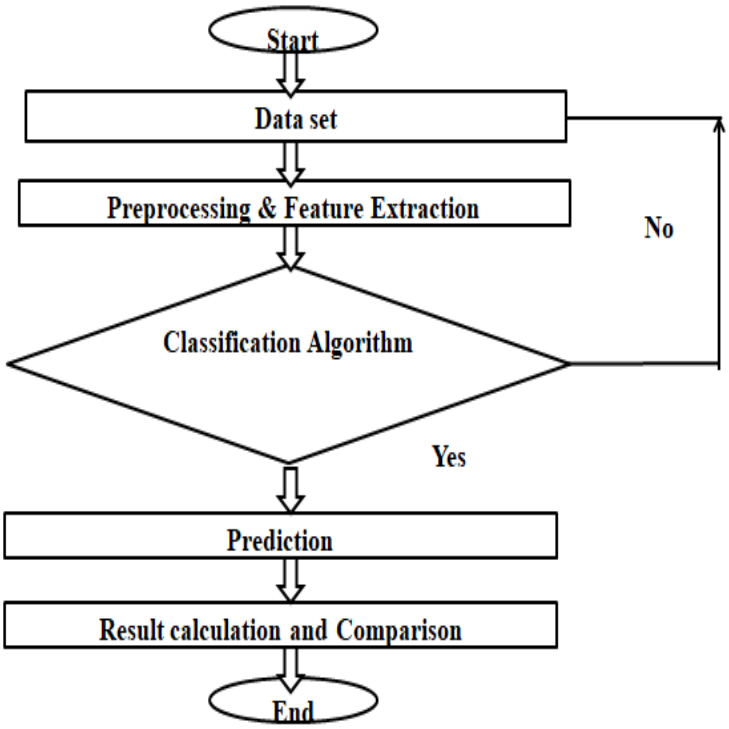

Figure 2: Flow Chart

Steps-

Random Forest is a powerful machine learning technique that can be used for predicting cardiac diseases using ECG (Electrocardiogram) dataset. ECG is a valuable diagnostic tool for assessing the electrical activity of the heart.

Data Collection and Preprocessing:

Obtain a dataset that includes ECG data along with labels indicating the presence or absence of cardiac diseases.

Preprocess the data by removing noise, normalizing the signals, and extracting relevant features from the ECG recordings, such as QRS complex duration, heart rate, ST-segment changes, etc.

Data Splitting:

Divide the dataset into training, validation, and test sets. Typically, you'd use around 70-80% of the data for training, 10-15% for validation, and the remaining 10-15% for testing.

Feature Selection:

Identify the most relevant features that have the most impact on the prediction of cardiac diseases. Random Forest can automatically handle feature selection to some extent, but you might want to perform feature engineering based on domain knowledge.

Random Forest Model Building:

Train a Random Forest classifier using the training dataset. Random Forest is an ensemble method that consists of multiple decision trees, each trained on a random subset of the data. This ensemble approach improves prediction accuracy and generalization.

Hyperparameter Tuning:

Tune the hyperparameters of the Random Forest model to optimize its performance. Important hyperparameters include the number of trees, maximum depth of trees, and the number of features to consider at each split.

Cross-Validation:

Perform cross-validation to ensure the model's robustness. This helps prevent overfitting and provides a more accurate estimate of the model's performance.

Model Evaluation:

Evaluate the Random Forest model using the validation set or through cross-validation. Common evaluation metrics for classification problems include accuracy, precision, recall, F1-score, and the ROC curve.

Model Testing:

Once you are satisfied with the model's performance, use the test dataset to assess its ability to generalize to new, unseen data.

IV. SIMULATION RESULTS

The implementation of the proposed algorithm is done over python spyder 3.7. The sklearn, numpy, pandas, matplotlib, pyplot, seaborn, os library helps us to use the functions available in spyder environment for various methods like decision tree, random forest, naive bayes etc.

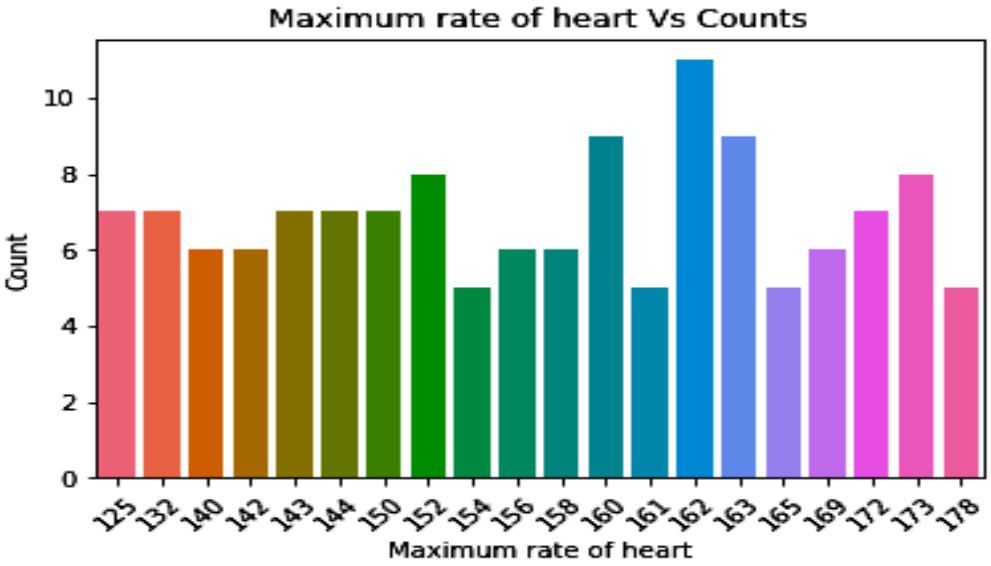

Figure 3: Rate of heart vs count

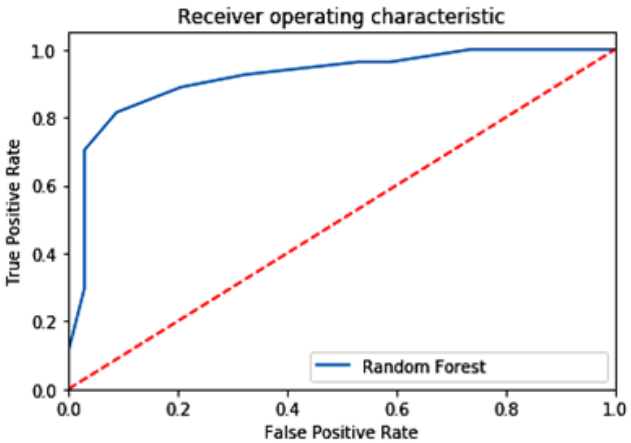

Figure 4: ROC result graph of Random forest

Table 1: Simulation Result

Sr. No.	Parameters	Value (%)
1	Accuracy	92
2	Classification error	8
3	Precision	90
4	Recall	91
5	F-measure	89

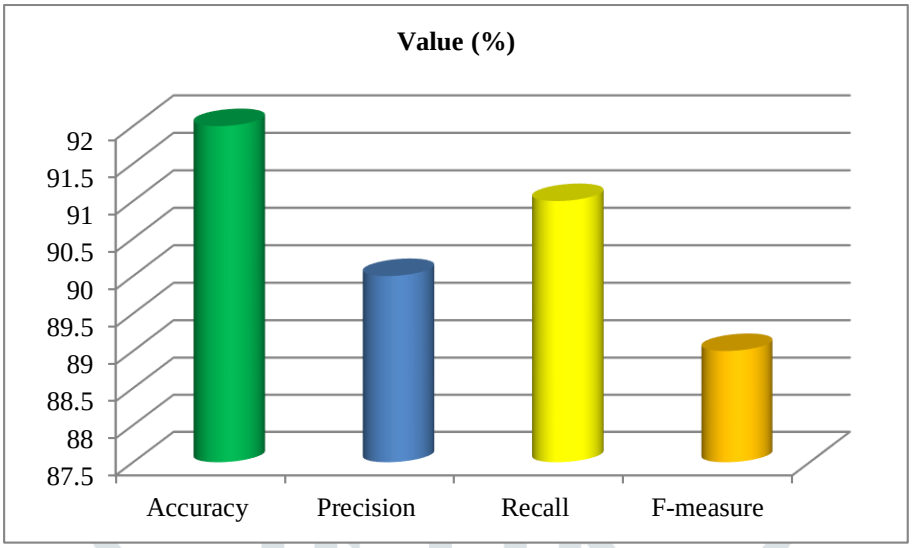

Figure 5: Result graphs of performance parameters

Figure 5 presents the graphical representation of the proposed method result graph in terms of the accuracy, precision, recall and F_Measure.

V. CONCLUSION

Present day medication generates a lot of data stored in the clinical database. For instance, clinical information may contain MRIs, signals like ECG, clinical data like glucose, circulatory strain, cholesterol levels, and so forth. Therefore this paper focused to develop a random forest method for predicting the threat of heart disease to a patient with the medical records got from the patients. Proposed method had the different attributes that it did not use already determined number of hidden units, but the hidden units got summed with one another till the error was decreased using proposed algorithm. The simulation results have proved that the proposed approach has achieved improvement in accuracy and other performance parameters.

REFERENCES

1. A. Kumar, K. Rathor, S. Vaddi, D. Patel, P. Vanjarapu and M. Maddi, "ECG Based Early Heart Attack Prediction Using Neural Networks," 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2022, pp. 1080-1083, doi: 10.1109/ICESC54411.2022.9885448.
2. D. Dube, D. S. Kumar and S. K. Gupta, "An Empirical study of the IoT arrhythmia detection methods: Review and research gaps," 2021 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2021, pp. 1-8, doi: 10.1109/ICCCI50826.2021.9402472.
3. B. Wang *et al.*, "A Multi-Task Neural Network Architecture for Renal Dysfunction Prediction in Heart Failure Patients With Electronic Health Records," in *IEEE Access*, vol. 7, pp. 178392-178400, 2019.
4. T. S. Brisimi, T. Xu, T. Wang, W. Dai, W. G. Adams and I. C. Paschalidis, "Predicting Chronic Disease Hospitalizations from Electronic Health Records: An Interpretable Classification Approach," in *Proceedings of the IEEE*, vol. 106, no. 4, pp. 690-707, April 2018.
5. A. Mdhaffar, I. Bouassida Rodriguez, K. Charfi, L. Abid and B. Freisleben, "CEP4HFP: Complex Event Processing for Heart Failure Prediction," in *IEEE Transactions on NanoBioscience*, vol. 16, no. 8, pp. 708-717, Dec. 2017.
6. J. Zhang *et al.*, "Coupling a Fast Fourier Transformation With a Machine Learning Ensemble Model to Support Recommendations for Heart Disease Patients in a Telehealth Environment," in *IEEE Access*, vol. 5, pp. 10674-10685, 2017.
7. N. Alshurafa, C. Sideris, M. Pourhomayoun, H. Kalantarian, M. Sarrafzadeh and J. Eastwood, "Remote Health Monitoring Outcome Success Prediction Using Baseline and First Month Intervention Data," in *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 2, pp. 507-514, March 2017.
8. A. Khan and A. Nigam, "Analysis of Intrusion Detection and Classification using Machine Learning Approaches", IJOSCIENCE, vol. 3, no. 10, Oct. 2017. DOI: https://doi.org/10.24113/ijoscience.v3i10.13.
9. D. Tay, C. L. Poh, E. Van Reeth and R. I. Kitney, "The Effect of Sample Age and Prediction Resolution on Myocardial Infarction Risk Prediction," in *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 3, pp. 1178-1185, May 2015.

10. S. Nikolaiev and Y. Timoshenko, "Reinvention of the cardiovascular diseases prevention and prediction due to ubiquitous convergence of mobile apps and machine learning," *2015 Information Technologies in Innovation Business Conference (ITIB)*, Kharkiv, 2015, pp. 23-26.
11. J. S. Sonawane and D. R. Patil, "Prediction of heart disease using learning vector quantization algorithm," *2014 Conference on IT in Business, Industry and Government (CSIBIG)*, Indore, 2014, pp. 1-5.
12. D. R. Patil and J. S. Sonawane, "Prediction of heart disease using multilayer perceptron neural network," *International Conference on Information Communication and Embedded Systems (ICICES2014)*, Chennai, 2014, pp. 1-6.