e-ISSN 2320 –7876 www.ijfans.org
Vol.11,S Iss 1, 2022
 UGC CARE Listed (Group -I) Journal

INDIVIDUAL CARDIAC RISK ASSESSMENT FOR A SMART DEVICE USING DEEP LEARNING

Amar B. Deshmukh[1*], Anjali M. Solanke[1], Sumagna Patnaik[2], C. P. Shirley[3], S.V. Evangelin Sonia[3], N. Jagadeesan[4]

[1]Department of Electronics and Telecommunication Engineering, Marathwada Mitra Mandal's College of Engineering, Pune, Maharashtra, India.

[2]Department of Information Technology, J.B. Institute of Engineering and Technology, Hyderabad, Telangana, India.

[3]Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India.

[4]PG Department of Computer Applications (M.Sc. IT & B.C.A), Dwaraka Doss Goverdhan Doss Vaishnav College, Chennai, Tamil Nadu, India

Corresponding mail: amarbdeshmukh@gmail.com

ABSTRACT

Cardiovascular disease is a leading cause of death worldwide. With nearly one person dying from heart disease every minute in today's ordinary contemporary life, the issue has emerged as one of the most pressing problems. Predicting the start of diseases at an early stage is critical nowadays. When used to healthcare, machine learning can accurately and swiftly identify illness. This study evaluates prospective advancements in the treatment of heart illness. The datasets utilised contain medical parameter characteristics. The samples were analysed in Python by means of the ML procedure, specifically the Forest Algorithm. This approach, which analyses patient data from the past to anticipate future deaths, reduces the incidence of fatalities. The Random Forest technique, a powerful Machine Learning tool, is used in this research to construct a reliable heart disease prediction system. The data is read from a CSV file containing patient records. The procedure is carried out after obtaining the information, and the effective cardiac arrest level is created. The proposed approach has the following advantages: it is very customizable, has a high success rate, and provides excellent reliability and efficiency.

1. Introduction

The heart's ability to pump blood is hampered by heart disease. The World Health Organization estimates that heart disease has killed 10 million individuals. The current

problem facing the healthcare sector is identifying illnesses before a person becomes ill [1]. Despite the fact that data or records on a person's medical history may be extensive, they may be few or inconsistent in the actual world. In these conditions, it may not have been feasible in the past to accurately predict and treat each patient's illness at an early stage [2]. Numerous attempts have been made by scientists to develop a model that can anticipate cardiac illness in its early stages, but none of them have been effective. Each suggested remedy comes with a unique combination of drawbacks. It suggested a method based on self-administered questionnaires for the current system [3]. The user must input each symptom for the algorithm to predict the outcome. The analytical data gathered in SAQ is the foundation of our investigation. A method for forecasting cardiac illness was created in another research. One of the machine learning approaches used for categorization and prediction was the Vector Quantization [4]. Neural networks are trained via back propagation and tested afterwards. On the testing set, around 80% accuracy is attained during the testing phase. Utilizing information gleaned from historical documents takes time. The accuracy rate is low. We employ the Random Forest method, which provides faster and more accurate results, to prevent this. The importance of machine learning in the healthcare industry is growing. One use of machine learning is prediction, and our goal is to foresee heart problem or illness by analysing patient datasets and patient data, or users from whom we need to determine the likelihood that a heart problem arises.

2. Research related to the cardiac vascular disease

Although the current heart disease prediction system uses a variety of algorithms, it has several serious drawbacks. Using algorithms for machine learning [5] Each and every algorithm for machine learning must adhere to these principles. Preprocessing "Not a number," or NaN values, are present in the database or dataset that we utilise. We must convert the non-numerical data to numerical data since the computer we use cannot handle it. The NaN values are swapped out in the method used by the column's mean [6]. Splitting Training sets and testing sets are the two groups into which the database's data is split. Eighty percent of the data in the testing set come from the training set [7], [8]. Tree of Decision Decision trees come in a variety of configurations. Their primary quality that sets them apart is that they previously placed the class characteristic first. The tree root characteristic uses knowledge acquisition to minimise entropy in an entropy system. Calculate the information gain of each attribute in the dataset before selecting a tree root. The feature of knowledge

acquisition will next be discussed [9]. KNN One of the simplest and most effective ways of categorization is this one. Since the customer is ignorant of certain dependable constant controls for probability densities, they might be challenging to comprehend during quality assurance [10]. Therefore, these kinds of computations are performed using the KNN classification approach. To predict the K-nearest neighbor's position, training datasets are utilised. To determine how near the training dataset is to the goal, Euclidean distance is utilised. Give the k rows that are the collection of rows being analyzed's closest neighbours [11]. The incomplete rows in the board set, repeat the process. The largest value of K may be chosen in this application, and an identical parallel model is automatically built on all standards of K up to the highest value provided. The exercise dataset, input variables, and production variables must all be constructed, according to the findings of both the WEKA tool and the KNN algorithm. For all standards of K up to the highest known value, parallel models are constructed using the best value of K. K-means grouping It is an unsupervised learning technique that uses datasets and class labels with unlabeled data. The primary goal of the algorithm is to produce a set of current data. K groups are distributed using this method repeatedly. These divisions are based on shared characteristics. Centroid K makes up each group. K is the total number of categories. According to Benn's theory, the K-means algorithm assigns a new value to a particular group based on its similarities. The new variable is grouped with centroid since centroid is necessary for the group [12]. Boost may be altered (Adaboost) It's a really effective strategy. To speed up the implementation of choice trees, it is used to binary cataloguing problems. Since classification rather than regression is where it is most often utilised, it is also known as separate adaptive improvement. It is feasible to improve the performance of machine learning procedures by using adaptive boost. The models marginally improve categorisation accuracy. The decision tree method, even though it only has one level, is often used with adaptive boost. Small decision trees with a single classification-related choice are called decision stumps. Techniques for Categorizing. Both training and test data sets are fed into the programme. Heart disease is predicted using a data mining environment. To predict heart disease, it uses collections of cardio vascular data (CVD) from several outside sources. There are five historical fiction algorithms used to analyse the datasets. The CVD datasets provide True and False as two separate criteria for predicting heart disease. These CVD data sets were all given to the WEKA data mining method as compressed ARFF files with data labels. Any input the user provides to the

programme will be accepted and tracked by the SQL server in the background. Since the programme only takes ARFF files, the WEKA utility converts the input from an EXCEL file to an ARFF file. The prediction system classifies new instances upon receipt and gives them a class label.

3. Cardiac vascular prediction

Our objective is to develop a system for predicting cardiac disease using the potent machine learning tool Random Forest. A excel file is offered as input. Following the fruitful conclusion of the treatment, the anticipated and proved consequence occurs. The workings of the system are shown in Figure 1. The user's data is associated to the data in the current data set using the Forest Algorithm. [10] The supervised learning subcategory includes the complex machine learning methodology known as the Random Forest method. It may be used to problems involving classification and regression. It uses a strategy that combines a number of classifiers to address a challenging problem, improving the effectiveness and precision of the model.

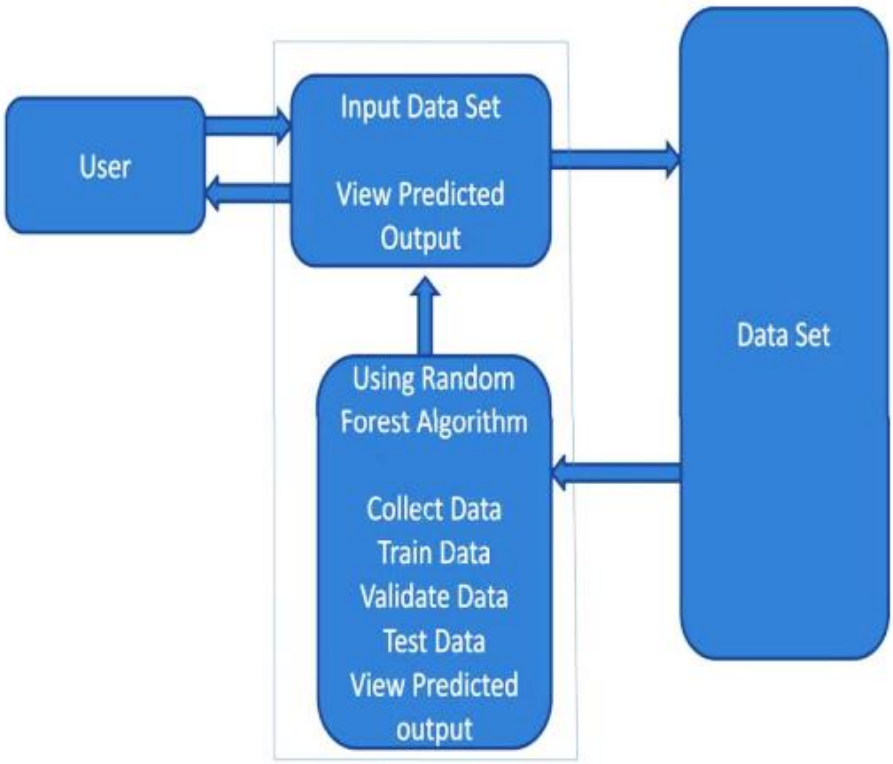

Fig. 1. Research architecture

A classifier called Random Forest uses many choice trees on dissimilar subsets of a given dataset and takes the regular into account to increase the predictability of the dataset. As shown in fig. 2, the RFA technique forecasts the final output using the outcomes of each decision tree rather than relying just on one. The accuracy of the results is influenced by the number of trees utilised; the higher the number of trees used, the higher the accuracy rate. Additionally, overfitting is avoided. The steps that make up the algorithm's functioning are as follows: Picking K data points from the designated training set is the first step. Utilizing the provided data points, construct as many decision trees as seen in fig. 3. The total number of decision trees you want to create is N. Repeat steps 1 and 2 as necessary. Any new data points are predicted by each decision tree and then placed in the category with the highest number of votes. Exams take up 30% of the time, while training takes up 70%. The compensations of the future model are excellent performance and accuracy. It is very successful and extremely versatile. The data or qualities in the data set are organised using the subsequent categories while creating the choice tree:

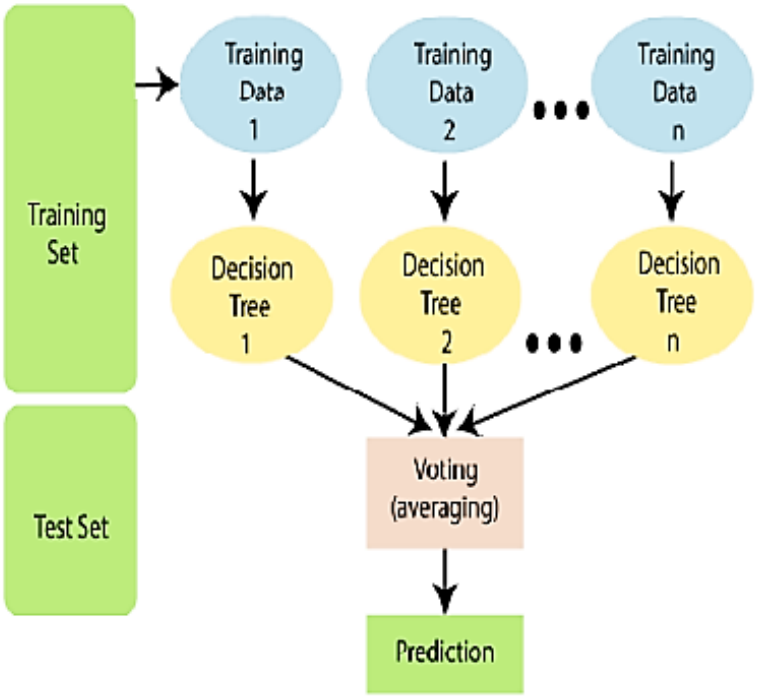

Fig. 2. Forest algorithm

4. Result and Discussion

The primary purpose of our work is to assess whether or not a person has heart disease. and provide suggestions on how to proceed The Random Forest approach has the potential to attain high accuracy rates. The data set that we utilised is shown here (sample). Table 1 has enough information to assess whether or not a person has heart disease.

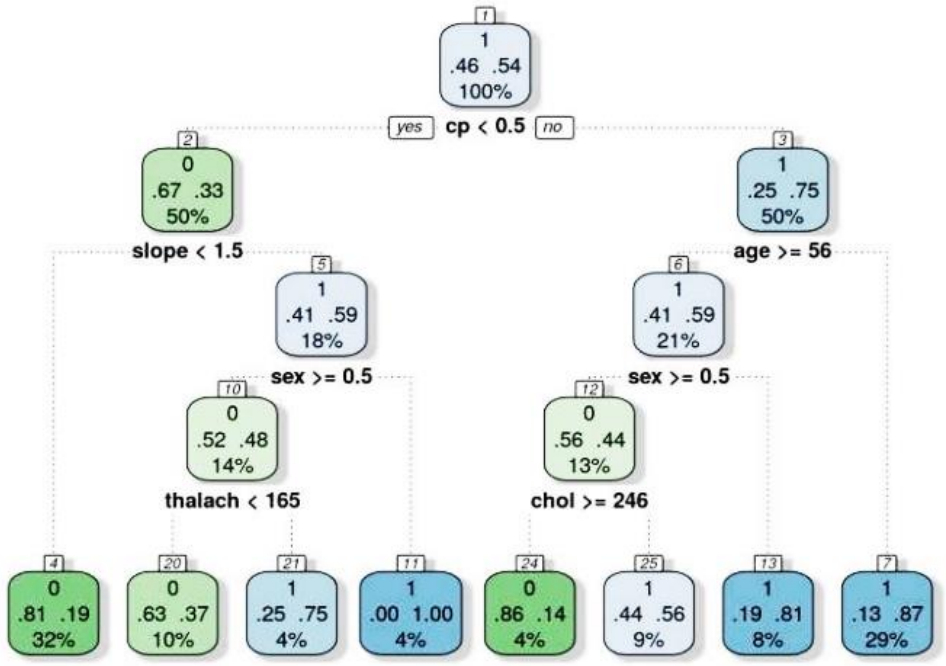

Fig. 3 Decision algorithm

Table 1 Sample data set

Age	62	36	40	55
Cp	4	3	2	2
Trestbps	146	131	131	121
Chol	234	251	205	237
Fbs	2	1	1	1
Thalach	151	188	173	179
Exang	1	1	1	1

Old Peak	3.3	4.5	2.4	1.8
Thal	2	3	3	3
Goal	2	2	2	2

Each data point in the collection was generated by the heart function. Cp-The kind of chest discomfort, for example, is classified into four numbers. (1) Pectoral chest discomfort Gina is unusual. 3. A pain numbing agent 4. There are no indicators The data set's attributes are listed in Table 2.

- Trestbps is an abbreviation for resting blood pressure measurement.
- The units used to measure cholesterol in serum are milligrammes per deciliter.
- Fasting blood sugar levels (if more than 120 mg/dl, indicated as 1, otherwise as 0)
- An electrocardiogram is the consequence of a resting condition.
- Exang-angina is caused by exercising (0-No, 1-Yes)

Table 2 Data set with result

AGE	62	36	16	55
CP	2	2	1	2
TRESTBPS	4	3	1	1
CHOL	234	251	171	201
FBS	2	1	2	2
RESTECG	1	2	1	2
THALCH	151	188	78	80
EXANG	1	1	2	2
OLDPEAK	3.3	4.5	1	2
SLOPE	1	1	3	3
THAL	2	3	4	4
TARGET	2	2	2	2

HEART DISEASE	YES	YES	YES	YES

When compared to a resting situation, previous peak effort generated ST depression. Figure 4 shows that when employing the random forest technique, the software has a greater accuracy rate than when using any other methodology.

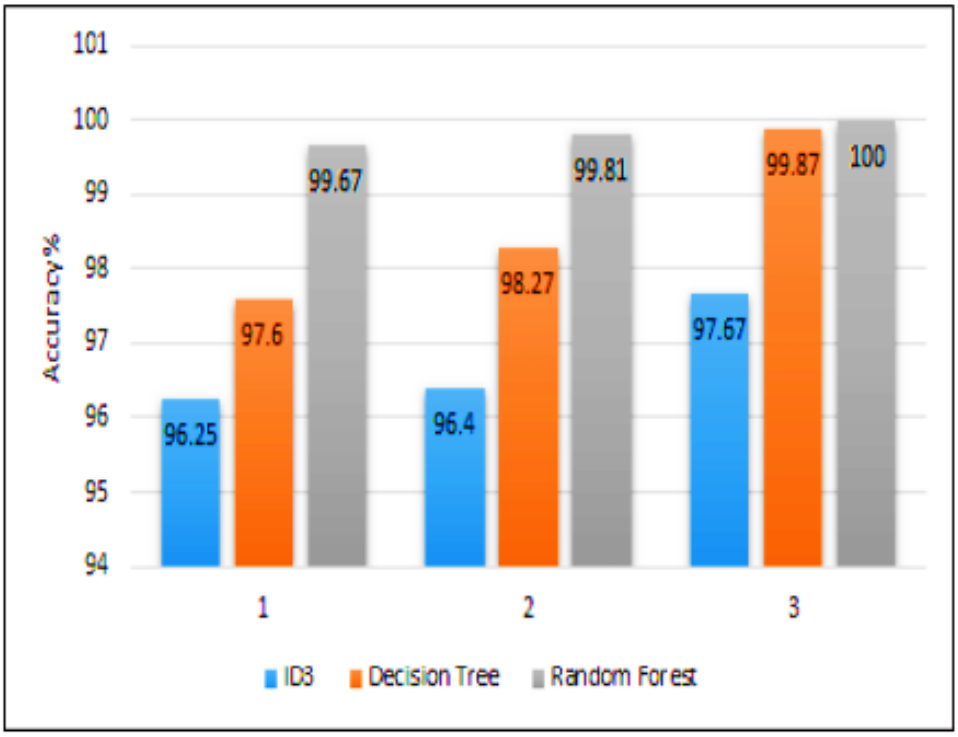

Figure 5. Comparison of various algorithm.

5. Conclusion

For reversing and organisational tasks, the Random Forest algorithm provides a potent collaborative knowledge technique. The function constructs N decision trees and provides a class that represents all decision tree outputs that are regular. As a consequence, it is successfully possible to achieve early-stage forecast accuracy. Processing health data, especially that pertaining to the heart, may help in the early diagnosis of heart disease or other atypical cardiac problems, hence lowering long-term mortality. In today's culture, the prognosis of heart illness is of prime importance. By inputting the report information, a

patient or user who is unable to contact a doctor may use this programme to anticipate illness and choose whether or not to seek medical attention.

Reference

[1] S. Akter, M. Amina, and N. Mansoor, "Early Diagnosis and Comparative Analysis of Different Machine Learning Algorithms for Myocardial Infarction Prediction," *IEEE Region 10 Humanitarian Technology Conference, R10-HTC*, vol. 2021-September, 2021, doi: 10.1109/R10-HTC53172.2021.9641080.

[2] S. Manikandan, "Heart Attack Prediction System," pp. 817–820, 2017.

[3] A. Deb, M. S. Akter Koli, S. B. Akter, and A. A. Chowdhury, "An Outcome Based Analysis on Heart Disease Prediction using Machine Learning Algorithms and Data Mining Approaches," *2022 IEEE World AI IoT Congress, AIIoT 2022*, pp. 418–424, 2022, doi: 10.1109/AIIoT54504.2022.9817194.

[4] R. S, R. Rajkumar, S. P. Shyry, J. A. Mayan, and Y. B. Jinila, "Prediction of Heart attack using Supervised Machine Learning Technique," no. Icesc, pp. 1516–1520, 2022, doi: 10.1109/icesc54411.2022.9885688.

[5] A. Kumar, K. Rathor, S. Vaddi, D. Patel, P. Vanjarapu, and M. Maddi, "ECG Based Early Heart Attack Prediction Using Neural Networks," no. Icesc, pp. 1080–1083, 2022, doi: 10.1109/icesc54411.2022.9885448.

[6] M. S. Raja, M. Anurag, C. P. Reddy, and N. R. Sirisala, "Machine Learning Based Heart Disease Prediction System," *2021 International Conference on Computer Communication and Informatics, ICCCI 2021*, no. Iccci, pp. 0–4, 2021, doi: 10.1109/ICCCI50826.2021.9402653.

[7] D. R. Krithika and K. Rohini, "Ensemble Based Prediction of Cardiovascular Disease Using Bigdata analytics," *Proceedings - 2021 International Conference on Computing Sciences, ICCS 2021*, pp. 42–46, 2021, doi: 10.1109/ICCS54944.2021.00017.

[8] D. Sharathchandra and M. R. Ram, "ML Based Interactive Disease Prediction Model," *2022 IEEE Delhi Section Conference, DELCON 2022*, no. i, 2022, doi: 10.1109/DELCON54057.2022.9752947.

[9] A. Pandiaraj, S. L. Prakash, and P. R. Kanna, "Effective heart disease prediction using

hybridmachine learning," *Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2021*, no. Icicv, pp. 731–738, 2021, doi: 10.1109/ICICV50876.2021.9388635.

[10] N. C. Sattaru, M. R. Baker, D. Umrao, U. K. Pandey, M. Tiwari, and M. K. Chakravarthi, "Heart Attack Anxiety Disorder using Machine Learning and Artificial Neural Networks (ANN) Approaches," *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2022*, pp. 680–683, 2022, doi: 10.1109/ICACITE53722.2022.9823697.

[11] D. Swain, S. K. Pani, and D. Swain, "A Metaphoric Investigation on Prediction of Heart Disease using Machine Learning," *2018 International Conference on Advanced Computation and Telecommunication, ICACAT 2018*, 2018, doi: 10.1109/ICACAT.2018.8933603.

[12] A. Gahane and C. Kotadi, "An Analytical Review of Heart Failure Detection based on IoT and Machine Learning," *Proceedings of the 2nd International Conference on Artificial Intelligence and Smart Energy, ICAIS 2022*, pp. 1308–1314, 2022, doi: 10.1109/ICAIS53314.2022.9742913.