ABSTRACT
The rapidly growing field of data analysis plays a significant role in healthcare. The healthcare industry has become big business. The healthcare sector produces enormous amounts of data every day. This data helps to extract the hidden information, which is useful to predict disease at the earlier. In medical field, predicting heart disease is treated as one of the intricate tasks. Therefore, there is a necessity to develop a decision support system to forecast the cardio vascular disease in a patient. Machine learning plays a vital part in disease prediction. In this paper, various machine learning methods were used to predict the heart disease and their performances were compared. The results obtained show the superiority of the Random forest algorithm.

KEY WORDS: CLASSIFICATION ACCURACY, HEART DISEASE, MACHINE LEARNING.

INTRODUCTION
Data mining is used to examines and unearths important information from a massive collection of data. This can be further helpful in exploratory and illustration out patterns for making intelligent business-related decisions. One of the most threatening in medical domain is heart disease, which occurs instantly when its limitations are reached. Machine learning plays a vital role in disease prediction Rajathi N et al., Cardiovascular disease generally refers to narrowed or blocked blood vessels, which can also lead to heart attack, chest pain or stroke. In general, blood pressure, cholesterol and pulse rate are the main reasons for a heart attack. Heart attack is the main heart disease.

Cardiovascular diseases (CVDs) are the most common explanation for global death. It is estimated that 17.9 million die annually. Heart attacks- once characterized as a part of “the old man’s disease” but in this era it can causes for more young people. The heart attack occurs when the coronary arteries become blocked. It causes a serious attack when one or more coronary arteries become blocked.

Bad clinical results would be the doorway in the death of a patient. A computer-based support system can be developed to make a good decision in order to achieve correct and cost-effective treatment. Most of the hospitals maintains their patient data in the form of images, texts and numbers using database systems. This data contains much of the hidden information that has not yet explored, which are useful to make right decisions. Therefore, there is a need to develop an excellent system to help the practitioners to predict heart disease before it occurs. This paper mainly concentrates on the prediction of heart disease considering the past heart disease database records.
MATERIAL AND METHODS

Various studies with respect to diagnosis of heart disease are discussed in this section. Feixiang Huang et. al., used a data mining process to foresee hypertension from patient medical histories and concluded that J-48 classifier produces better results. M. Amiri et. al., developed diagnosis systems heart sounds. They used 116 heart sound signals to classification and regression trees. M.A. Nishara Banu et. al., used clustering and classification algorithm to forecast the hazard level of the patients. The authors Theresa Princy et. al., discussed about classification methods including Naïve Bayes, neural network, KNN, decision tree for predicting the risk level of a patient they consider age, gender, pulse rate, blood pressure, cholesterol of each patient.

The various the machine learning algorithms are used by Min Chen et. al., for effective prediction of chronic disease. A multimodal disease risk prediction method was adopted for structured and unstructured data. The prediction accuracy the algorithm is better than other with a convergence speed. Tikotikar A et.al., data mining technique are used in the medical field for clinical diagnosis. It is inferred that an exhaustive survey of medical data help to make well informed diagnosis and decisions.

Cincy Raju et. al., proved that the SVM technique is an efficient method for predicting heart disease. Praveen Kumar Reddy. M, et. al., used decision tree algorithm to prove the better prediction by comparing its performance with SVM. The authors Akash et. al., applied structured data and the text data of the patient to the k-mean algorithm and archived better accuracy. Reddy et. al., employed machine learning methods for heart disease prediction. All these created an interest to employ machine learning to prediction of heart disease.

Proposed Methodology: In this paper, various machine learning methods including Naïve Bayes classifier, logistic regression, random forest, support vector machine, decision tree classifier and KNN are employed to forecast heart disease. The Python language is used for implementation. The working of the model proposed is pictorially depicted in Figure 1. The dataset is pre-processed in order to remove irrelevant data which helps to achieve better accuracy.

Dataset: The heart disease dataset available in UCI repository taken for this study. The dataset consisting of the parameters including age, sex, chest pain type, serum cholesterol, resting blood pressure etc. After preprocessing the dataset was separated into training (70%) and testing (30%). The models used logistic regression, K-Nearest Neighbor, Support Vector Machine, Naïve Bayes, Decision tree and Random forest are trained using the training data and finally tested with the testing set.

RESULTS AND DISCUSSION

The overall objective of this paper is to forecast more accurately the occurrence of heart disease. Simulation based experiments were conducted using six methodologies named Naïve Bayes Classifier, Logistic Regression, Random Forest, SVM, Decision Tree Classifier and KNN. From the result it’s been seen that the random forest gives more accuracy as compared as other five techniques. The data set used is decomposed into a training set and testing set. Here, 70% of the dataset is taken for training and the remaining is considered for testing. From the dataset, it is identified that there are more people suffering from heart disease in the 50-60 age group. This is pictorially represented in figure 2.

From the dataset, it is inferred clearly that a greater number of men are suffering from heart diseases as compared to women. While the range of men suffering from heart disease lies between 80-100, the number of women suffering from heart disease lies between 60-80. This is shown in figure 3. The performance of the classification models on the test data was represented using confusion matrix, per class accuracy and classification accuracy and is given in table 1.
The classification accuracy of various algorithms is graphically represented in figure 4 and the results are presented in table 2. From the results achieved it is inferred that random forest algorithm gives best prediction accuracy than other algorithms.

**CONCLUSION**

In the field of disease prediction, machine learning plays a significant role. In this paper, various machine learning approaches are used for heart disease forecast. The experimental results illustrate that the Random Forest algorithm achieves the highest accuracy of 91.8% and thus successfully achieving the objective of improving the prediction accuracy. The future work is towards more investigation on evolutionary computation techniques for the problem undertaken and study their performances.

**REFERENCES**


Praveen Kumar Reddy, M., Sunil Kumar Reddy, T.,
Malavika G et al.,


Tikotikar, A., & Kodabagi, M., 2017. A survey on technique for prediction of disease in medical data. In 2017 International Conference on Smart Technologies for Smart Nation (Smart Tech Con) (pp. 550-555). IEEE.