e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

COMPARATIVE ANALYSIS OF DIFFERENT MACHINE LEARNING TECHNIQUES FOR PREDICTING HEART DISEASE

N. Saivivek*1, P. Gnaneswar*2, Dr. T. Dhilipan Rajkumar*3

*1,2Student, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Anand Nagar, Krishnankoil, India.

*3Assistant Professor, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Anand Nagar, Krishnankoil, India.

ABSTRACT

Heart Attack is a term that provides a large number of heart-related medical conditions. The key to cardiovascular disease (Cardiovascular disease) is to evaluate large amounts of data sets, compare information that can be used to predict, Prevention, Management as a heart attack. Heart disease is mainly caused by stress, family background, High Blood Pressure, etc.... Data analysis is used to classify the earth for use in a useful way to control, corrupt and manage large data sets. It can be used with great success in predicting, preventing, and controlling heart disease. To address this we aim to use SVM-based Data Analytics and Genetic Algorithm to diagnose heart disease. This result reveals the Genetic Algorithm as better predictive models.

Keywords: Random Forest, LSTM, Regression

I. INTRODUCTION

It is difficult to diagnose heart disease because of several risk factors such as diabetes, high blood pressure, high cholesterol, abnormal heart rate and many other factors. Various data mining methods and sensory networks have been used to determine the severity of heart disease among people. The severity of the disease is categorized based on various methods such as the K-Nearest Neighbor Algorithm (KNN), Decision Trees (DT), Logistic Regression (LR), and Vector Support Machine (SVM). The nature of heart disease is complex and therefore, this disease should be treated with care. Failure to do so would cause a heart attack or death. The concept of medical science and data mining is used to diagnose different types of metabolic syndromes. Data mining by classification plays an important role in cardiovascular prediction and data analysis.

II. LITERATURE REVIEW

According to Ordonez [1] heart disease can be predicted by other key factors taken from a patient and in their work we have introduced a system that includes individual factors based entirely on 13 basic factors such as sex, blood pressure, cholesterol and others to predict a patient's risk of heart disease. They added two other attributes namely fat and smoking and expanded the research database. Data classification algorithms such as Decision Tree, Naive Bayes, and Neural Network are used to predict and results are analyzed on a cardiovascular site.

Yılmaz, [2] proposed a method that uses a small square vector support device (LS-SVM) that uses a binary decision tree to separate a cardiotocogram to determine a patient's condition.

Duff, et al. [3] conducted a research project involving 533 patients who had suffered from cardiac arrest and were included in the analysis of risk factors for heart disease. They do archeological analysis and data mining analysis using many Bayesia networks.

Frawley, et al. [4] have performed the task of predicting the survival of Coronary heart disease (CHD) which is a challenging research challenge for the medical community. They also used 10-fold verification methods to determine the neutral value of the three speculative models for performance comparison purposes. Lee, et al. proposed a novel approach to augmenting and learning a multifaceted feature as well as direct and indirect aspects of Heart Range Diversity diagnosing heart disease. They performed various tests on specific and non-linear features to evaluate several class dividers, e.g., Bayesian categories, CMAR, C4.5 and SVM.

Based on their experiments, SVM performed much better than other class dividers. No, et al. propose a divisive approach which is an integrated system built on the effective method of FP growth. Because the volume of patterns can vary and be large, they have provided a law for measuring coherence and allowed for difficult selection of pruning patterns in the pattern production process.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

Parthiban, et al.[5] proposed a new project in which heart disease is diagnosed and predicted using the proposed Coactive Neuro-Fuzzy Inference System (CANFIS). Their model operates based on a combination of neural network flexibility capabilities and is based on a genetic and cognitive algorithm to detect the occurrence of a disease. The effectiveness of the proposed CANFIS model was evaluated according to the training program and phase accuracy. Finally, their results indicate that the proposed CANFIS model has a better chance of predicting heart disease.

Singh, et al.[6] they did the work using, partition clustering algorithm (K-Means) and hierarchical clustering algorithm (agglomerative). The K-means algorithm has high performance and scalability and integrates quickly when produced with large data sets. Consecutive combinations form a series of combinations by usually combining two smaller groups into larger ones or dividing a larger group into smaller ones. Using the WEKA data mining tool, they calculated k-means performance and hierarchical clustering algorithm on the basis of accuracy and duration.

Guru, et al.[7] they have proposed a three-layer multilayer perceptron computational model to be used to expand the decision-making process for the diagnosis of five major cardiovascular diseases. The proposed decision support system is trained using a back-to-back broadcast algorithm with dynamic, flexible reading level and memory tools. Palaniappan, et al. have done research work and developed a model known as the Intelligent Heart Disease Prediction System (IHDPS) using several data mining techniques such as Decision Trees, Naïve Bayes and Neural Network. Shantakumar, et al. have done research work when an intelligent and effective heart attack predictor system is developed using the Multi-Layer Perceptron with Back-Propagation. Consistent with this, frequent patterns of heart disease are generated by the data-based MAFIA algorithm.

III. EXISTING SYSTEM

- The World Health Organization (WHO) estimates that 12 million people worldwide die each year from heart disease.
- About 25% of deaths in people aged 25-69 occur due to heart disease. In urban areas, 32.8%.
- Deaths are due to heart disease, with a percentage of rural areas at 22.9.
- More than 80% of people worldwide die from heart disease. The WHO estimates that by 2030, there will be about 23.6 million.
- People will die of heart disease.
- Diagnosis is an important and tedious task in medicine.
- Treatment for this disease is very high and cannot be affordable for many patients especially in India.

IV. PROPOSED SYSTEM

- Predicting heart disease.
- It helps to reduce the cost of treatment by providing effective treatment.
- Finding the boundary values in forecasts such as accuracy, timing and power consumption.

V. METHODOLOGY

SVM Classifiers:

Early predictors of cardiovascular disease can help make lifestyle changes for high-risk patients and reduce their risk. Research has attempted to identify the most influential features of heart disease and to accurately predict overall risk using homogeneous data mining techniques. Recent research has focused on combining these methods using methods such as hybrid data mining algorithms. This paper proposes a rule-based model to compare the accuracy of the application of the rules in each of the results of the vector support system, decision trees, and retranslation in the Cleveland Heart Disease Database to show an accurate model of predicting heart disease.

Knn:

Medical data mining to explore a hidden pattern in data sets. Monitored algorithms are used to predict early heart disease. Nearby Neighborhood (KNN) is the most widely used isolation laziness algorithm. KNN is the most popular, efficient and effective algorithm used for pattern detection.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

Decision Tree:

The user prepares the procedures by looking at specific details and symptoms of heart disease. Decision tree (ID3) and navie Bayes methods for data mining are used to retrieve information associated with each patient. Based on accurate results predictions, system performance is analyzed.

VI. TECHNOLOGY USED

AN INTRODUCTION TO PYTHON

Python is a standard, translated, interactive, object-oriented language, and advanced programming language. In interpreted language, Python has a design philosophy that emphasizes code readiness (especially using white spacing to limit code blocks rather than curved brackets or keywords), as well as syntax that allows programmers to express concepts in a few lines of code than can be used in languages. like C ++ or Java. Provides structures that allow for clear planning on both small and large scales. Python translators are available for most operating systems. CPython, a Python reference reference implementation, is open source software and has a community-based development model, as it makes almost all of its unique applications. CPython is owned by the nonprofit Python Software Foundation. Python installs a system of variable type and default memory management. It supports a wide range of program paradigms, including object-focused, critical, practical and processed, and has a large and complete general library.

DJANGO

Django is a high-quality Python web framework that promotes rapid development and clean, functional design. Designed for experienced engineers, it takes great care of web development, so you can focus on writing your app without having to re-engineer. It's free and it's open source.

Django's main goal is to facilitate the creation of complex, web-driven websites. Django emphasizes the reusability and “connectivity” of the components, the rapid development, and the goal of never repeating them. Python is used throughout, even settings files and data models.

INTRODUCTION TO MYSQL

MySQL (http://www.mysql.com) is a robust SQL data server developed and maintained by T.c.X DataKonsultAB in Stockholm, Sweden. Available publicly since 1995, MySQL has risen to become one of the world's most popular web servers, thanks to part of the server speed, durability, and flexible licensing policy. (See note for more information on the MySQL licensing strategy.) With the provision of MySQL embedding capabilities, combined with a large and very simple set of pre-defined integration functions, MySQL has undoubtedly become the most popular PHP partner partner.

VII. EXPERIMENTAL RESULTS AND DISCUSSION

1. Upload Training Data:

The law-making process proceeds in two stages. During the first phase, the SVM model is built using training data During each wrap, this model is used to predict class labels Rules are tested on the remaining 10% of the test data to determine accuracy, precision, recall and F rating. . Additionally, the size of the rule set and the length of the intermediate system are also calculated for each round of opposite verification.

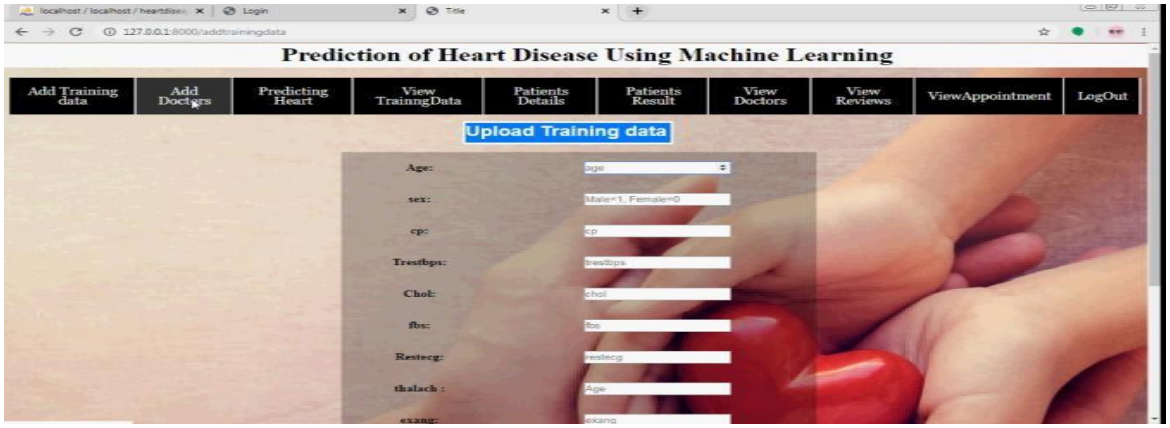

Figure 1: Upload Training Data

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

2. Data Pre- Processing:

Heart disease data is first processed after collection of various records. The database contains a total of 303 patient records, of which 6 records have some missing values. Those 6 records have been removed from the database and the remaining 297 patient records are used in preliminary processing. Multi-category variations and binary classification are presented in the given database attributes.

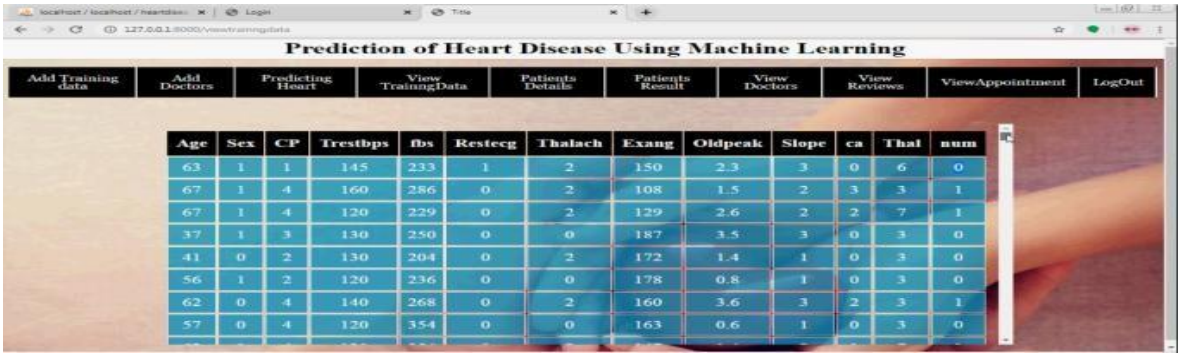

Figure 2: Data Pre- Processing

3. Predicting Heart Disease:

The training set is different than the test set. In this study, we used this method to ensure the universal application of the methods. In a way to ensure k-fold contrast, all databases are used to train and test the partition in Heart Stoke.

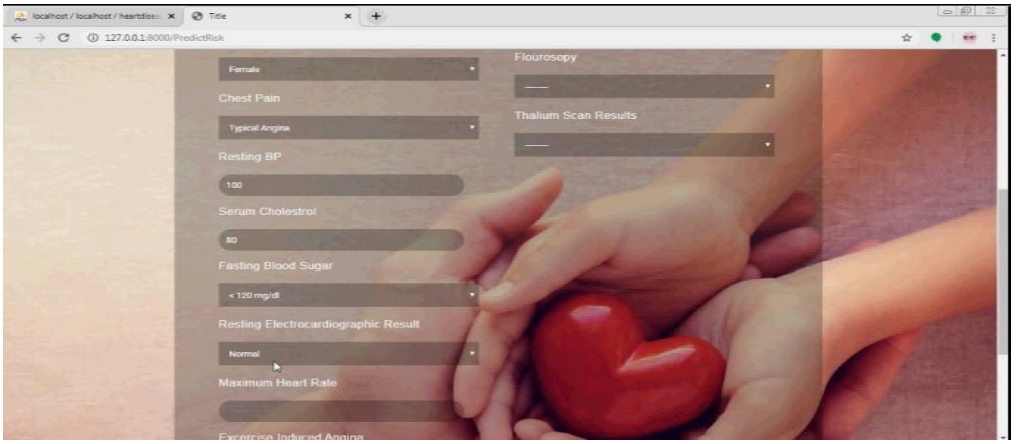

Figure 3: Predicting Heart Disease

4. Graphical Representations:

Analysis of proposed systems is calculated based on authorizations and approvals. This can be measured with the help of graphic texts such as pie chart, bar chart and line chart. Data can be assigned to dynamical data.

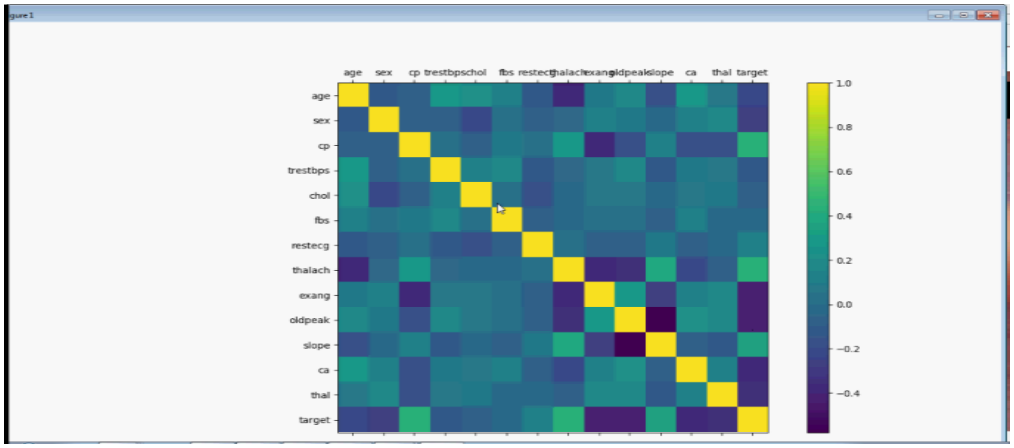

Figure 4: Graphical Representations

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

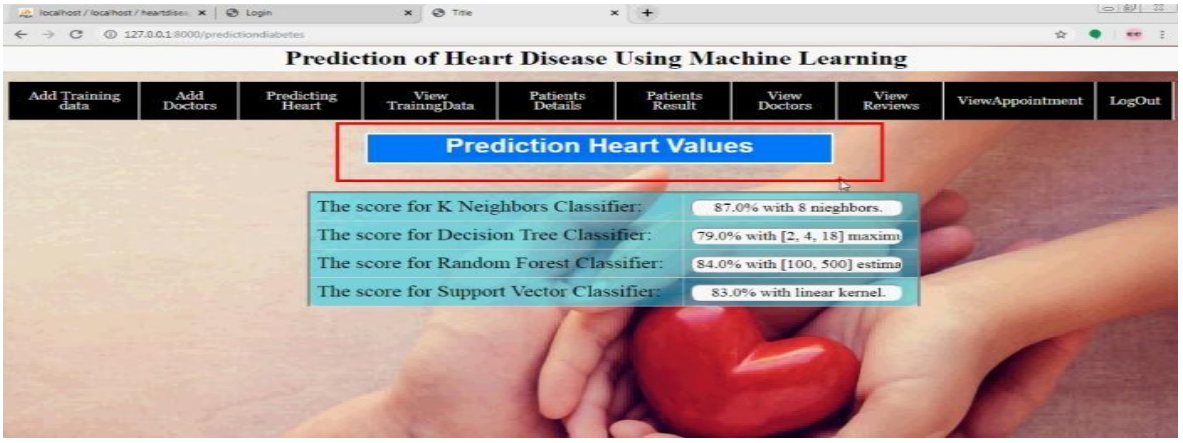

Figure 5: Accuracy

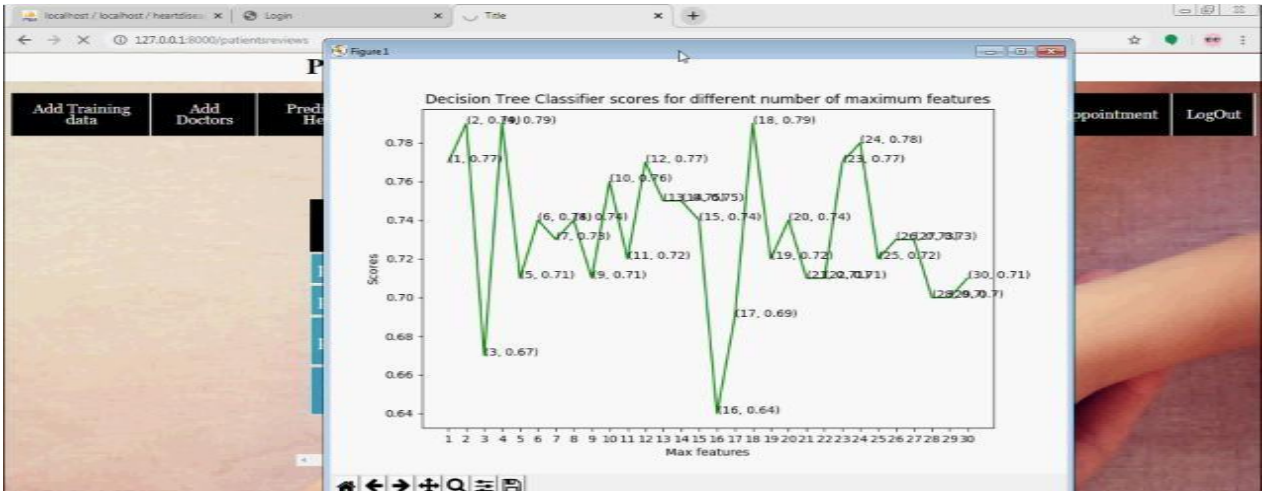

Figure 6: Classifier

VIII. CONCLUSION

Identifying the analysis of raw heart health care data will help save long-term human lives and early detection of heart conditions. Machine learning techniques have been used in this project to process raw data and provide new and unique insights into heart disease. Predicting heart disease is challenging and extremely important in medicine. However, the mortality rate can be very manageable if the disease is diagnosed in the early stages and preventive measures are taken as soon as possible. Further expansion of this study is highly desirable in order to direct research into real-world databases instead of theoretical methods and simulations. The proposed hybrid HRFLM method is used to integrate the features of Random Forest (RF) and Linear Method (LM). HRLMM proved to be very accurate in predicting heart disease.

IX. FUTURE ENHANCEMENT

Future study of this study can be done with a variety of mechanical learning techniques in better guessing techniques. In addition, new traits to select traits may be developed to obtain a broader perspective on key factors to increase the prognosis for heart disease.

X. REFERENCES

[1] Carlos Ordonez, "Improving Cardiovascular Prognosis using the Rules of Prohibited Organization", Introduction to Technical Seminar, University of Tokyo, 2019.

[2] Franck Le Duff, ChristianMunteanb, Marc Cuggiaa and Philippe Mabob, “Predicting the Causes of Survival After a Cardiac Cardiac Surgery Using Data Mining Method”, Health Technology and Informatics Studies, Vol. 107, No.2, pages 1256-1259, 2019.

[3] W.J. Frawley and G. Piatetsky-Shapiro, “Information Access to Information Sources: Overview”, AI Magazine, Vol. 13, No. 3, pages 57-70, 2019.

[4] Heon Gyu Lee, Ki Yong Noh and Keun Ho Ryu, "Bio Signal Data for Mining: Diagnosis of Coronary Artery Diseases Using Online and Indirect HRV Features", Proceedings of International Conference on Emerging Technologies in Knowledge Discovery and Data Mining, pp. 56-66, 2020.

[5] Kiyong Noh, HeonGyu Lee, Ho-Sun Shon, Bum Ju Lee and Keun Ho Ryu, "Integrated Cardiovascular Diagnosis Method", Intelligent Computing in Symptom Analysis and Pattern Recognition, Vol. 345, pages 721-727, 2019.

[6] Latha Parthiban and R. Subramanian, "Intelligent Heart Disease Prediction System uses CANFIS and Genetic Algorithm", International Journal of Biological, Biomedical and Medical Sciences, Vol. 3, No. 3, pages 1-8, 2019.

[7] Niti Guru, Anil Dahiya and Navin Rajpal, "Decision Support Program for Diagnosis of Heart Disease using the Neural Network", Delhi Business Review, Vol. 8, No. 1, pages 1-6, 2019.