Grenze International Journal of Engineering and Technology, Jan Issue

Heart Disease Prediction and Diagnostic Analysis

N. Malarvizhi[1], Sakshi Kumari[2], Kunal Nayan[3] and Vamsi Kalyan Jupudi[4]
[1-4]VelTech Rangarajan Dr. Sagunthala R&D Institue of Science and Technology/Department of CSE, Chennai, India
Email: drnmalarvizhi@gmail.com, {vtu11831, vtu11856, vtu15955}@veltech.edu.in

***Abstract*— A heart disease is a significant health concern worldwide and is responsible for a considerable number of deaths every year. Early detection of heart disease can significantly reduce the risk of morbidity and mortality associated with it. This has led to the development of various machine learning techniques for predicting and analyzing heart disease. This study investigates the pertinence of machine learning techniques to predict long term outcomes in heart disease patient. The study aims to evaluate the performance of different machine learning algorithms such as Decision tree, KNN, and SVM for predicting the occurrence of heart disease in patients. The data used in this project is collected from Kaggle Platform. The performance of each algorithm is evaluated using metrics which is accuracy and the results of this study will provide valuable insights into the effectiveness of different Machine learning algorithms for heart disease prediction and analysis, which can be used for early detection and management of heart disease. Varieties of data points that could lead to heart disease are examined. The diagnostic analysis additionally demonstrates how some factors, like Age, Sex, Cholestrol, CP, Restbps etc. used in data-sheet are likely to cause heart disease. Tableau has been used for Data Visualization. Django has been used as web framework. User have to enter required input value and then prediction analysis will result whether patient having heart disease or not.**

***Index Terms*— Artificial Intelligence, face recognition, machine learning, facial features.**

I. INTRODUCTION

Heart illnesses refer to a broad spectrum of problems that limit the heart's ability to operate correctly. The most common types of cardiac disease include heart attacks, heart failure, arrhythmias, congenital heart disease, and coronary artery disease. Coronary artery disease occurs when the blood vessels that supply the heart with oxygen and nutrients become narrow or blocked, leading to chest pain, shortness of breath, and, in severe cases, heart attack. Heart failure is a condition in which the heart is unable to pump enough blood to fulfill the demands of the body. This causes weakness, breathlessness, and swelling in the legs and ankles. Arrhythmias are irregular heart rhythms that can cause dizziness, fainting, or even sudden cardiac arrest. Valve disease occurs when the heart's valves do not function properly, leading to shortness of breath, fatigue, and chest pain. Early detection and treatment of heart disease can greatly enhance a patient's prognosis and quality of life. Blood tests, stress tests, and other diagnostic methods are used to identify persons who are at risk of developing heart disease and to identify the specific form of heart disease a person may have.

The use of diagnostic analysis and heart disease prediction has the potential to greatly improve heart disease early detection and treatment, increasing patient outcomes and lowering the overall burden of heart disease on society. This research provides an early heart disease prediction by combining the occurrence of resting blood pressure, Chest Pain, fasting blood sugar, resting electro-cardio graphic result, exercise induced angina, the slope of the

Grenze ID: 01.GIJET.10.1.323

peak exercise segment, that to make more accurate predictions about a person's risk of developing heart disease. The data-sheet contains 1025 rows and 14 columns. Machine Learning algorithm like KNN, DT and SVM has been used for Prediction Analysis. Tableau has been used for Data Visualization

II. Literature Survey

The study of numerous categories of heart malady such as coronary heart illness, coronary course illness, heart disappointment, ischemic heart infection, cardiovascular infection, innate heart illness, valvular heart malady and hypo-plastic cleared out heart disorder are discussed in [1]. From the overview the finding is that neural arrange based procedures contribute more adequacy and some methods have gotten more than 90 percent precision. At last, a few of the research issue is additionally addressed to go before the assist inquire about on the same course. The authors in [2]. proposed a framework that has utilized 14 noteworthy parameters for testing. The dataset comprises of 1026 different persistent data. Eight classification methods are performed to foresee heart hazard conditions. All of them Decision Tree work superior and give amazing accuracy. The framework has connected execution measurements such as exactness, accuracy, sensitivity, specificity, precision, F-measure, and classification error to measure the system performance. This study has developed a website that people can use to simply and swiftly check their heart health at home. Data from 1026 distinct patients were used to train and test the system. It performs better in a variety of methods, including DT (99 percentage), RF (99 percentage), XgBoost (95 percentage), KNN (89 percentage), SVM (85 percentage), LR (85 percentage), Ada Boost (83 percentage) and Naive Bayes (82 percentage). The experiment's findings give a target value of 0 or 1 that indicates whether the patient has heart disease or not.

A model utilizing machine learning techniques to forecast cardiac disorders has been created in [3]. Classification techniques were used to construct this model because they are crucial for prediction. Different classification techniques, such as Logistic Regression, Random Forest, Support Vector Machine, Gaussian Naive Bayes, Gradient boosting, K-nearest Neighbors, Multinomial Naive Bayes, and Decision Trees are used to create the model. The classifiers are tested and trained using the Cleveland data source. Additionally, a feature selection approach called chi square is utilized to pick the most important characteristics from the input data set, which reduces execution time and improves classifier performance. Random forest has the best accuracy of all the classifiers tested using performance measures. As a result, the model created using Random Forest is an effective and workable solution for diagnosing cardiac disorders, and it can be applied in the healthcare industry, which is crucial to the field of cardiology.

The author in [4] suggested the Heart Disease Prediction using Exploratory Data Analysis. Four different types of chest discomfort are taken into account while predicting the likelihood of heart disease. Analysis was performed using cardiac data. The database contains 209 records and records of 8 characteristics, including age, chest pain type, blood pressure, blood glucose level, resting ECG and heart rate, and four types of chest pain. The Cleveland heart disease raw data-set, which contains 76 features of 303 patients, is used to define the suggested algorithm. To eliminate error brought on by inconsistent data, some samples are eliminated during the pre-processing phase. 209 samples with seven independent characteristics, including age, the type of chest pain, blood pressure, blood glucose level, ECG at rest, heart rate, the frequency of physical activity, and ECG, are used to predict heart disease. Age is considered to be the biggest risk factor for heart problems because fatty streaks in the coronary arteries form during adolescence. Only men are included in this data collection since they are more prone than women to get cardiac diseases. Angina is the pain that results from tight muscles. Data is analyzed using visualization tools.K means clustering method is used in big data environment.

A model developed in [5] is used to predict cardiac disease using machine learning algorithms. This model will help doctors diagnose the condition early on with fewer tests and provide the necessary care, perhaps saving many lives. In hospitals, a significant amount of data about patients with heart disease and other illnesses is produced every day. Without data mining techniques, clinicians would struggle to use or handle the patient's data effectively to make judgements. facts mining is extremely advisable for heart disease prediction since it extracts more accurate and usable facts from enormous amounts of data, making prediction simple. Because of its fundamental design, machine learning is able to manage vast amounts of data, process it quickly, and generate predictions early on. Heart disease can be diagnosed using a variety of data mining approaches, including classification, prediction, and pattern recognition. In this study, machine learning classification models are utilized to identify cardiac vascular disorders. Data input is used by classification algorithms to predict and classify the data to which class or category it belongs. Logistic regression, Decision Trees, Random Forest, Gradient Boosting, Support Vector Machine, Naive Bayes, and K-Nearest Neighbor are a few of the classification algorithms. All classification algorithms are trained to forecast heart disease in this study, and performance comparisons are made.

A system has been created to forecast coronary disease [6]. Boosting and bagging ensemble techniques are combined with PCA and LDA feature extraction algorithms to enhance the performance of heart disease prediction. The Cleveland heart disease dataset is utilized to extract key features using feature extraction techniques. Selected features are used to compare the performance of ensemble methods (boosting and bagging) and five classifiers (KNN, SVM, NB, DT, and RF). The bagging ensemble learning algorithm using DT and PCA feature extraction approach had the best performance, according to the experimental data. the DT, KNN, and RF have the highest performance of accuracy, recall, precision, and F-measure which were 98.4 percentage, 98.5 percentage, 98 percentage, and 98 percentage respectively. The authors in [7] defined several data mining methods that can be effectively used to forecast cardiac disease. With fewer attributes and tests, this work seeks to develop efficient and accurate prediction. The model made use of previously edited and pre-processed data. The most effective algorithms are XGBoost with 78.69 percentage and Random Forest with 86.89percentage. The least accurate algorithm, K-Nearest Neighbor, has a performance of 57.83 percentage. Additional data mining techniques, such as time series, clustering and association algorithms, and other ensemble approaches, can be used to advance this research. The limitations of this study need the adoption of more intricate and coupled models in order to increase the precision of heart disease early prediction.
The experimental findings of several current classifiers were described in [8]. First, the effectiveness of each classifier used, including K-Nearest Neighbors (KNN), Decision Trees (DT), Extra-Tree Classifiers (ETC), Random Forests (RF), Logistic Regressions (LR), Naive Bayes (NB), Artificial Neural Networks (ANN), Support Vector Machines (SVM), Adaboosts (AB), and Gradient Boosting (GB), are assessed. The conspicuous and highly variant features are then chosen from the feature space using four feature selection algorithms (FSAs): Fast Correlation-Based Filter (FCBF), Minimal Redundancy Maximal Relevance (mRMR), Least Absolute Shrinkage and Selection Operator (LASSO), and Relief. Additionally, the chosen feature spaces are fed into classification algorithms as input so that the significance may be examined.
Some of the machine learning methods such as K closest neighbors (KNN), logistic regression, and random forest classifiers were investigated in [9]. These algorithms can help practitioners or medical analysts identify heart disease accurately. This essay's research involves looking through journals, papers that have been published, and current cardiovascular disease statistics. The technique is a set of actions that converts the incoming data into patterns of data that are familiar to the users. The proposed approach consists of three stages, the first of which is the data collection, the second of which is the extraction of relevant values, and the third of which is the pre-processing stages where we can study the data. Filling in missing numbers and cleaning up the data are the primary functions of data pre-processing. Following pre-processing, a classifier is utilized to categorize the pre-processed data. K Nearest Neighbor is the classifier employed in the proposed model.

III. PROPOSED SYSTEM

The suggested system of heart disease prediction uses a variety of approaches, algorithms, and other tools to create a system that diagnoses a patient's heart illness based on their symptoms. The patient's symptoms are then compared with the system's previously available data-set. By taking those data-sets and comparing them with the patient's disease, the accurate percentage of disease can be predicted. Web-framework Heart Disease Prediction Application will be used. User have to enter the required field. Then, with the aid of different algorithms and techniques like Decision Tree, KNN, and SVM, those data are classified. The output will be predicted whether the patient is having heart disease or not. Data Visualization is done by Using Tableau.
Advantages of the proposed system

- This system will help in the speedy detection of the heart disease, which can save the patient's lives who dies due to unavailability of tests at the initial stage.
- The accuracy level of the disease's prediction will be far better since the conventional testing.
- The financial cost of the testing will be much cheaper in the long run.

The architecture diagram of the system heart disease prediction using machine learning shown in Figure 1 consist of textual data-sets through which we will compare the symptoms of the user and predicts it, Data sets are transformed into the smaller sets and from there it gets classified based on the classification algorithm. Later on, the classified data is processed into the machine learning technologies through which the data gets processed and goes in to the heart disease prediction model using all the inputs from the user. Then after user entering the above information and overall processed data combines and compares in the prediction model of the system and finally predicts the heart disease.
The proposed system is divided into several modules and each of the module is explained as below:

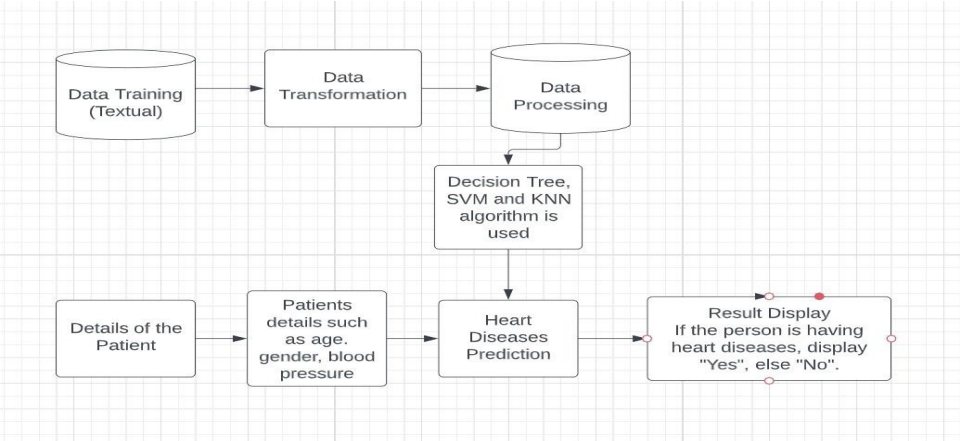

Figure 1. Block diagram of the proposed system

A. Data Collection Pre-Processing

The data-set of heart disease prediction is collected from Kaggle. The obtained data-set is stored in file and uploaded on Jupyter platform. Then data is pre-processed to find any missing values. So, after they are gathered, they are pre-processed into a model-compatible format for the machine learning algorithm. Pre-processing involves a variety of methods and procedures, including: cleansing up data. These manual and automatic methods eliminate erroneously added or categorized data.

B. Training and Modelling

The data-set is fragmented into training and testing ration. Testing and training datasets are trained using different machine learning techniques like Decision Tree, KNN and SVM. The first phase in machine learning is model training, which produces a functioning model that can subsequently be validated, tested, and deployed. How well a model performs during training ultimately determines how well it will perform when it is implemented in an application for end users.

C. Development

The process of integrating a machine learning model into an already-existing production environment is known as deployment, and it allows you to use data to make useful business decisions. It can be one of the most challenging stages of the machine learning life cycle and is one of the final ones. Then Jupyter notebook of stroke prediction file is downloaded and deployed into DJANGO web-framework which is python based, front-end and back-end are connected in this step.

D. Data Visualization

The technique of presenting data in a visual or graphical form to facilitate understanding, analysis, and communication is known as data visualization. It involves using charts, graphs, maps, and other visual elements to convey information. Tableau has been used for data insight.

IV. Results And Discussions

By utilizing the proposed model, we can recognize the danger level of heart disease alongside the order of sickness levels with a precision of 97.7 percent. Decision tree shows efficiency with accuracy score of 97.7 percentage. KNN shows efficiency with accuracy score of 95.5 percentage. SVM shows efficiency with accuracy score of 94.4 percentage. The heart disease prediction model initialized with primary attributes given by the user to find the risk levels. The proposed model is efficient in predicting the heart disease accurate results.

This system will help in the speedy detection of the heart disease, which can save the patient's lives who dies due to unavailability of tests at the initial stage. The accuracy level of the disease's prediction will be far better since the conventional testing. Machine learning algorithm like KNN, SVM and DT as shown in Figure 2. Decision tree shows efficiency with accuracy score of 97.7 percentage. KNN shows efficiency with accuracy score of 95.5 percentage. SVM shows efficiency with accuracy score of 94.4 percentage. The user-provided primary attributes

were used to initialize the heart disease prediction model and determine the risk levels. The proposed model is efficient in predicting the stroke, with accurate results.

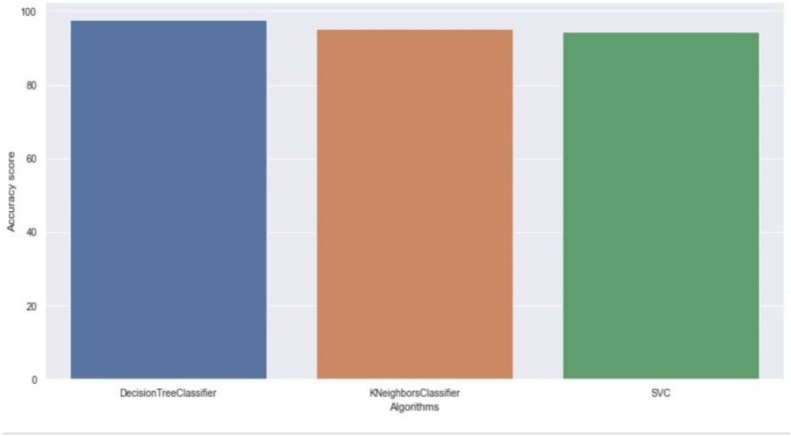

Figure 2: Prediction Accuracy Score of ML algorithm

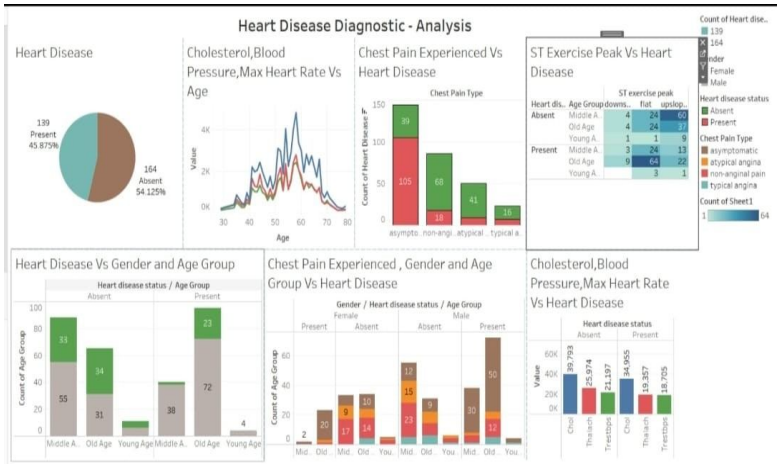

Figure 3: Tableau Dashboard

The Dashboard of Heart Disease Diagnostic Analysis is done by using Tableau as depicted in Figure 3. Finally, the result is predicted whether the patient is having heart disease or not. If the predicted value is 0 the patientis not having heart disease. If it is 1 the patient is having heart disease.

V. Conclusion

Since heart disease is a leading cause of death in India and around the world, the application of cutting-edge technology, such as machine learning, to the early detection of heart disease, would have a significant influence on society. An important development in the world of medicine is the ability of high-risk individuals to make decisions about lifestyle changes that will reduce problems with early diagnosis of cardiac illness. More people are given heart disease diagnoses each year. This necessitates a quick diagnosis and plan of action. The adoption of suitable technology support in this area may have significant positive effects on both the medical community and patients. SVM, Decision Tree, and KNN are some of the machine learning methods utilized in this work to evaluate the performance and predicts the heart disease.

References

[1] Bohdan Koval ,Iulia Khlevan,etc.al,"Development of Heart Disease Diagnosis Concept Using Machine Learning", Information Technology and Implementation (IT&I-2021), December 01–03, 2021.

[2] Bohdan Koval ,Iulia Khlevan,etc.al,"Development of Heart Disease Diagnosis Concept Using Machine Learning", Information Technology and Implementation (IT&I-2021), December 01–03, 2021.

[3] B Padmaja,Chintala Srinidhietc.,"Early and Accurate Prediction of Heart Disease Using Machine Learning Model" Turkish Journal of Computer and Mathematics Education,Vol.12 No.6,pp 4516-452850, 2021.
[4] R. Indrakumari, T. Poongodi,Sowmya Ranjan Jena "Heart Disease Prediction using Exploraory Data Analysis",(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 12, No. 6, pp 2528-30, 2021.
[5] Narender Kumar,Dharmender Kumar "Machine Learning based Heart Disease Diagnosis using Non-Invasive Methods" Journal of Physics:ConferenceSeries.Vol.10,No.5,pp-3450, 2021.
[6] Xiao-Yan Gao,Abdelmegeid Amin Ali etc.al,(2021) "Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method",Volume 2021, Article ID 6663455, 10 pages, 2021.
[7] Pooja Anbuselvan,"Heart Disease Prediction using Machine Learning Techniques",International Journal of Engineering Research Technology (IJERT) Vol 9 Issue 11, 2020.
[8] Muhammad, Y., Tahir, M., Hayat, M. et al. Early and accurate detection and diagnosis of heart disease using intelligent computational model. Sci Rep 10, 19747, 2020.
[9] J.Jeyaganesan,"Diagnosis And Prediction of Heart Disease Using Machine Learning Techniques" Vol 19, Issues 2:pp.1817-1827, 2020.