

Graphical User Interface Based Heart Disease Analysis In Machine Learning Algorithms

D.Manojkumar¹,P.Rajeswari²,C.Nandhakumar³,J.Ramprasath⁴

¹Assistant Professor, manojkmr314@gmail.com,

²Assistant Professor, radhap2886@gmail.com

³Assistant Professor, nandhakrish1997@gmail.com

⁴Assistant Professor (SS), jrprasath@gmail.com

^{1,2,4}Dept.of IT, Dr. Mahalingam College of Engineering and Technology, Pollachi

³Dept of CSE, M .Kumarasamy College of Engineering, Karur

Abstract-As we all know; heart disease affects the majority of individuals in the world today. About one person per minute in this age passed away from heart-related illnesses. As a result, heart disease-related mortality has recently become a serious problem. Roughly one person jumps into the pool every minute as a result of heart illness. We are all aware that enormous volumes of unethical data are produced in our healthcare systems. We need a system that is well-structured and user-friendly for handling this massive type of data. We can therefore introduce machine learning algorithms to the outline as a result of this. We have included a GUI-based machine-learning method for heart disease prediction in our future system. We have gathered the 14-field dataset for our system and pre-processed it utilizing data analysis algorithms. Then, using Python Jupiter, we compared with the fundamental classification methods such as SVM, DT, RF, and LR. Our analysis of this technique led us to the conclusion that logistic regression ensured a precision and warmth of 98% and 89%, correspondingly.

Keywords-ML, SVM, heart disease prediction, Decision Tree, Random Forest

1. INTRODUCTION

One of the most common and dangerous diseases that threatens public health around the world, heart disease frequently results in heart attacks and strokes. Heart disorders are becoming a major factor in human lifespan reduction. Blood pressure, stress, blood sugar, work overload, and a number of other factors all contribute to heart diseases, which in turn cause deaths. Healthcare professionals have long had a crucial and difficult task: predicting and identifying cardiac disease. Therefore, being able to detect cardiac disease in its early stages will be helpful to individuals everywhere so they can take the required precautions before it becomes severe. Data mining techniques were applied in

medical diagnosis, particularly in the prediction of heart disease. There are many algorithms for classifying and predicting the outcomes of heart disease, for example. Machine learning techniques, for example, are applied to medical science and are rising rapidly due to their propensity for providing high accuracy in classification and calculating outcomes, promoting patients' health, and improving the value and quality of healthcare. Heart attack diseases are thought to be the most common of all illnesses that cause death. Medical professionals undertake several surveys on heart conditions and collect data on heart patients, their signs, and the course of the disease. Gradually, reports of nearby individuals with common diseases and characteristic indications.

People today want to live very expensive lives, so they work very hard similar machines to make a lot of money and be comfortable. As a result, they neglect to mind of themselves, and as a result, their eating ways and way of life change. They become more tense, develop blood gravity and sugar issues at a young age, don't get plenty sleep, and eat whatever they can grow.

In a report out in 2020, the World Health Organization WHO specified that heart disease is the top cause of passing global. A change of applications, integrating advertising, client relationship management, engineering, health analysis, web mining, and mobile computing, have employed data mining. Healthcare fraud and abuse incidents have been effectively identified via data mining. Through statistical treatment and data mining, medical complexity is decreased. To improve the experience and ends that have been given, we have used some data mining techniques. Health care businesses gather enormous quantities of data that may contain some unknown information, which is valuable for making decisions based on data and for making choices that would produce the best results. A major contributing factor to developing heart disease is having high blood pressure. Determining the presence of heart disease is therefore a serious challenge. Knowing if someone has heart disease or not is tough to expect.

Therefore, machine learning may be a useful option for diagnosing any human heart condition in order to avoid this complication. The act was compared in terms of detection accuracy, prediction time, care, and specificity. The supervised machine learning techniques used in this prediction of heart disease are LR, DT RF, SVM, and Nave Bayes Therefore, using search algorithms rather than machine learning approaches for prediction will provide better outcomes in the prediction of heart disease in the future.

2. LITERATURE REVIEW

Early detection is important because heart disease is among the main causes of mortality worldwide. The computer-based computer - aided diagnostic system aids doctors in the diagnosis of heart disease. Early detection

of cardiac disease can therefore help in high-risk patients decision-making regarding lifestyle modifications, reducing consequences, which can be an important milestone in the area of medicine. [1] An effective desktop approach and predictive modelling that can help with the early identification or prediction of cardiac disease was developed using machine learning techniques. According to a number of factors, this model predicts if a patient will develop heart disease or not. Heart disease prediction relies on the use of data gathering, processing, feature selection, and classification methods including K - means, Random Forest, Logistic Regression, SVM, or Decision Tree. Filtering, wrapping, and integrating are the three techniques for selecting features. In this instance, filter-based feature selection methods like chi-square, data fusion, and correlation were applied. Of the 13 qualities, only 10 aspects are determined to be crucial for model development. It will lower the computational expense of modelling or, in some situations, be utilized to raise the effectiveness of the model. The accuracy of each algorithm will be compared at the end, and the algorithm that yields the most reliable data will be chosen. Users of this GUI programmer can examine their health at their convenience while sitting down. Anyone can use the GUI to conduct a check-up to see if a person has heart disease because of how it is made to be user-friendly.

Globally, heart disease kills most (CVD). Clinical data shows that an ML system may detect CVD early and reduce mortality. Several studies have used machine learning to diagnose CVD and assess patient severity. [2] Despite good results, none of these studies optimized Models for CVD detection and severe degree categorization. This study addresses imbalance distribution using SMOTE. Six ML classifiers and Super Parameter Optimization (HPO) are utilized to determine the patient's status. All structures were built and tested using two accessible datasets. [3] SMOTE and Additional Tree (ET) optimized using hyper band outperformed previous models and state-of-the-art CVD identification trials by 99.2% and 98.52%, respectively. The Cleveland dataset-related algorithm also converges to 95.73%. Physicians can assess cardiac disease using the indicated method. Hence, early therapy prevents heart disease-related death.

Cardiovascular disease is a complex condition that affects many people all over the world. In treatment, and notably in the discipline of cardiology, the timely and precise identification of heart disease is essential. In this study, we proposed a ML-based method for quickly and precisely identifying cardiac disease. The system is based on classification algorithms like SVM, LR, artificial neural networks, KNN, Naive bays, and DT. Standard feature selection algorithms like Reliever, Negligible backup maximum possible significance, least squares wastage anywhere from, and Local learning have been used to eliminate redundant and irrelevant features. In addition, we proposed a special fast conditionally information gain feature selection technique to address the feature selection issue. [4] In order to increase classification accuracy and reduce classification system execution time, the approaches for feature selection are applied. In order to determine the optimum methods for model evaluation and hyper parameter tweaking, the leaving one case out cross-validation methodology was also used. Classifier performance is assessed using performance measures. The performances of the classifiers have been assessed using the features chosen utilizing feature selection methods. According to the experimental findings, an elevated intelligent system for detecting heart illness might be made by combining the suggested feature selection approach (FCMIM) with a classifiers support vector machine. The suggested diagnosis system (FCMIM-SVM), when compared to previously described methods, achieved good accuracy. The suggested method is also user-friendly in the healthcare industry for the identification of cardiac illness.

Medicine's hardest problem is predicting heart disease. Medical and other specialists take a long time to figure out the cause. This work uses LR, KNN, SVM, GBC, and Grid Search to forecast cardiac illness. 5-fold

cross-validation validates the system. Contrast these four methods. The Extreme Patterns Boosting Encoder using Mesh Refinement CV had the best and virtually identical test and training accuracy for both datasets at 100% and 99.03%. Further analysis shows that the XG Boost Classifier without Grid Search CV has the best testing and training accuracies for both datasets at 98.05% and 100%. (Hungary, Switzerland & Long Beach V and UCI Kaggle). [5] The proposed method's systematic results are compared to heart disease prediction studies. This Extreme Conjugate Gradient Boosting Encoder with Grid Search CV has the highest max parameter for measuring accuracy. This paper proposes a new model-creation technique for practical challenges.

Modern heart disease is a leading cause of death. Clinical data analysis has trouble predicting cardiovascular disease. Machine learning (ML) helps healthcare providers predict and make decisions from their massive data sets. Current Internet of Things breakthroughs have used machine learning in several sectors (IoT). [6] ML-based cardiac disease prediction study offers little insight. This work proposes a novel machine learning model for identifying significant variables to improve heart disease prediction. The prediction model is introduced using several feature mixtures and many well-known classification algorithms. To predict cardiovascular disease, we use hybrid randomized forests with a linear model to improve performance with 88.7% accuracy (HRFLM).

3. PROPOSED SYSTEM

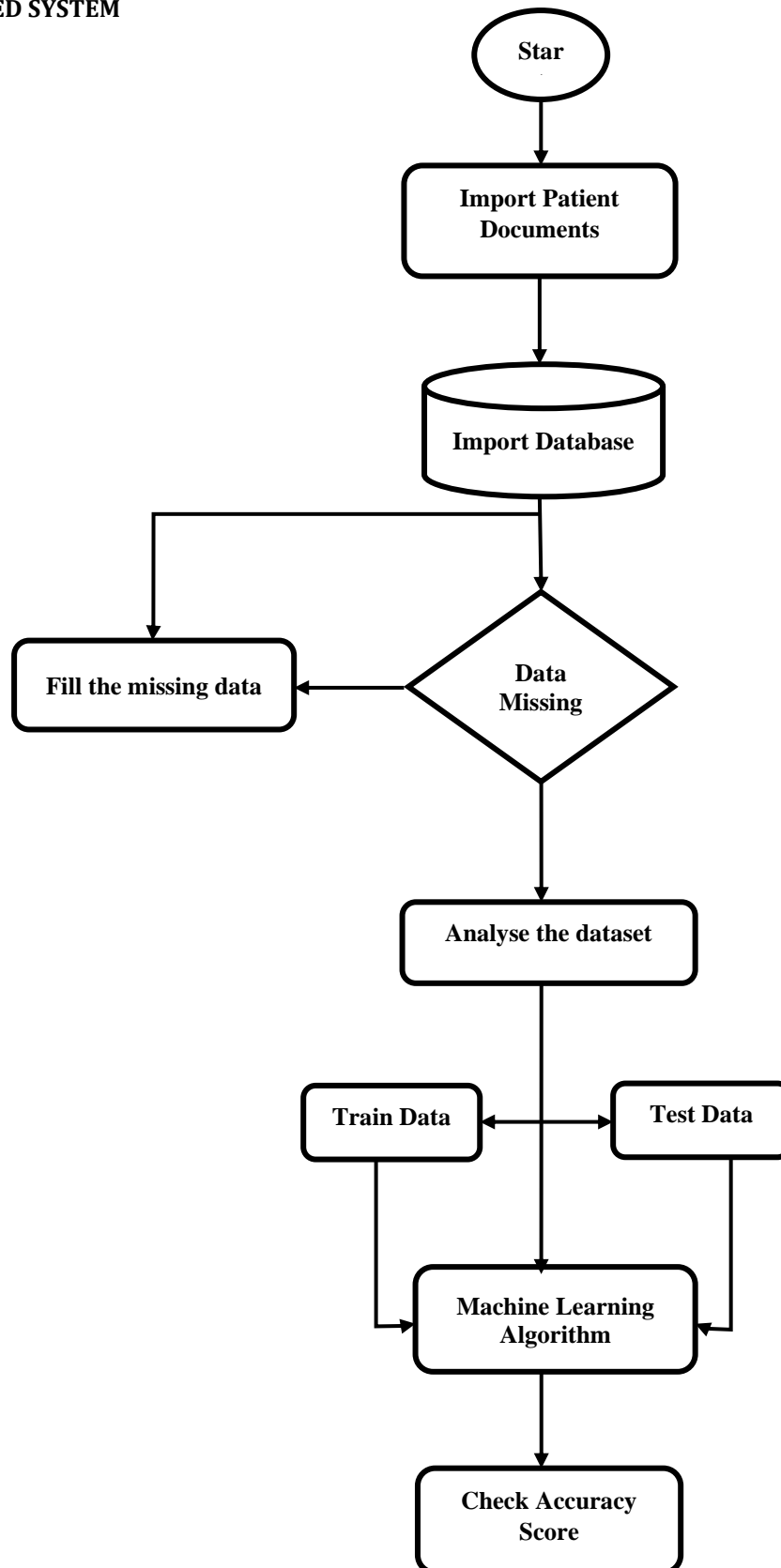


Figure 1. Proposed Data Flow

Figure 1 illustrated the proposed data flow for UCI Machine Learning repository. It includes an actual dataset of 300 samples of data with 14 different attributes (13 predictors; 1 class), such as blood pressure, the type of chest discomfort, and the ECG result. To construct a model with the highest level of accuracy achievable for this system, we applied four algorithms to determine the causes of heart disease.

3.1 Classification Algorithm

Given their matches, each object in the collection is categorized. The most popular and widely utilized DM strategy is classification. Accurately predicting the target class of objects for which the class label is unknown is the goal of classification methodology. There are nine groups of the categorization algorithms offered. These algorithms were selected for use in this system.

3.1 Random Forest

A Random Forest classifier uses numerous decision trees on unlike subsets of the input data to increase the dataset's expected precision. When using the Random Forest technique, you use the mean squared error to solve regression difficulties (MSE).

ALGORITHM

1. **Input: Importing the libraries**
2. **Output: best configuration**
3. # importing the dataset
4. `a = ds.iloc[:, :-1].values`
5. `b = ds.iloc[:, 13].values`
6. #splitting dataset into training set and test set
7. `from sklearn.model_selection import train_test_split`
8. `A_train,A_test,b_train,b_test=train_test_split(A,b,test_size=0.25)`
9. #feature scaling
10. `from sklearn.preprocessing import StandardScaler`
11. #exploring the dataset
12. #VISUALIZATIONS---relationship between attributes
13. End
14. Return best configuration

4. RESULT AND EXPERIMENTS

For simulation, we use the system windows 10 OS with Intel i5 CPU and 4G RAM and 15 GB hard disk. Data-science programs appropriate for Windows, Linux, and macros are included in the bundle. Anaconda, Inc. created and maintains it.

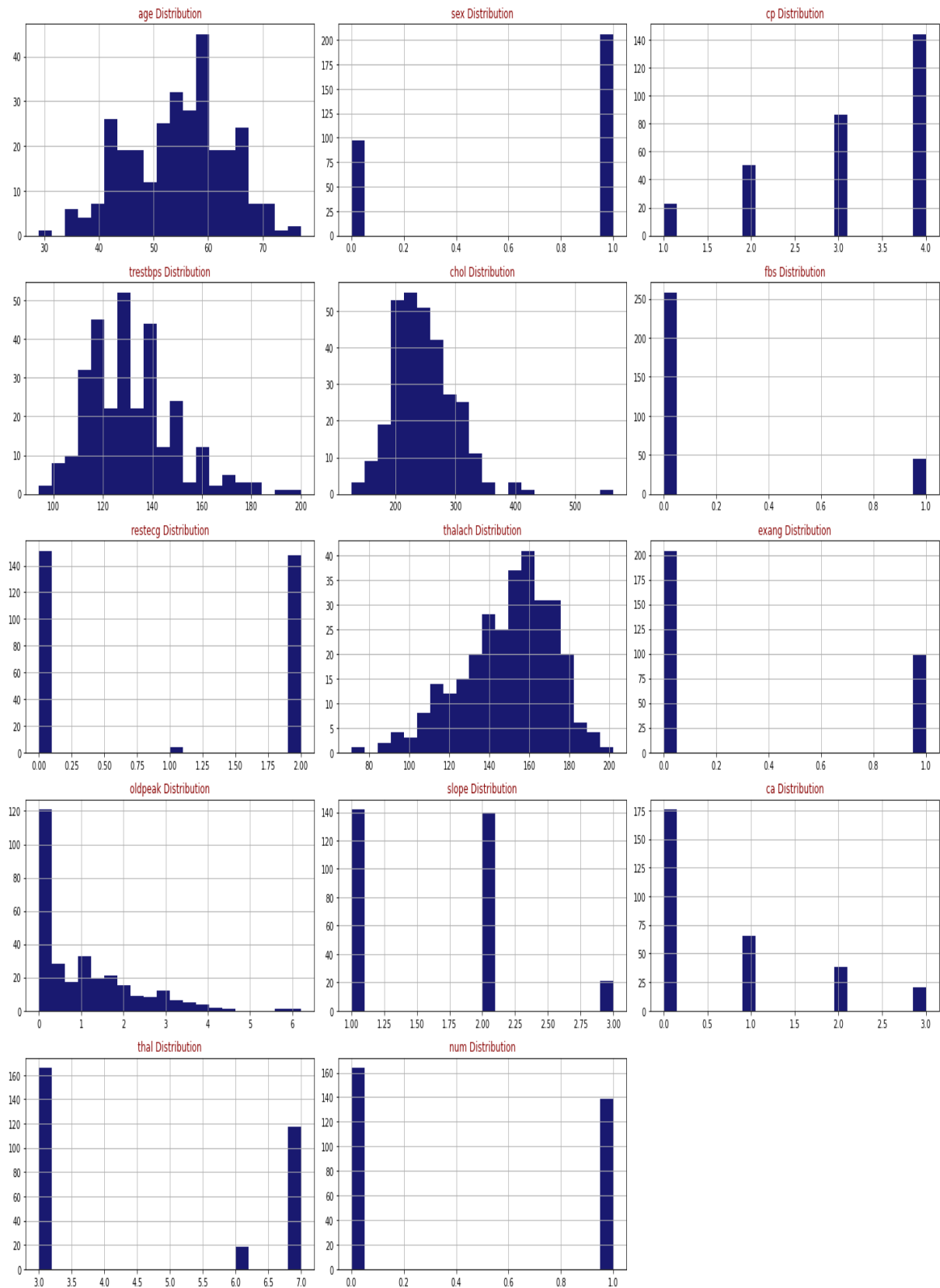


Figure 2. Output of Random Forest Algorithm

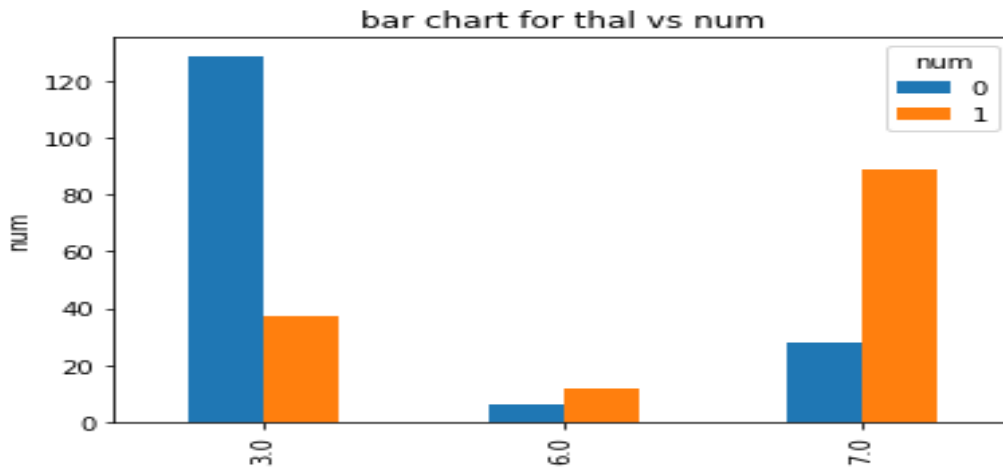


Figure 3. Random Forest Algorithm Graphical Representation

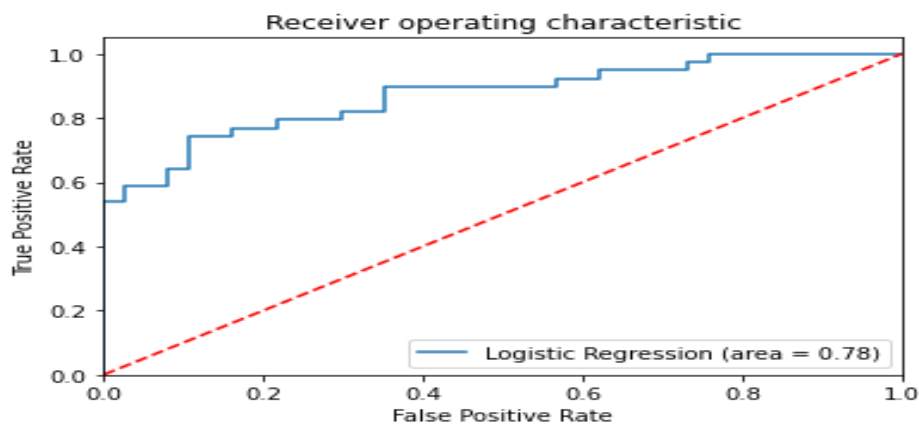


Figure 4. Bar Chart for True Positive vs False Positive

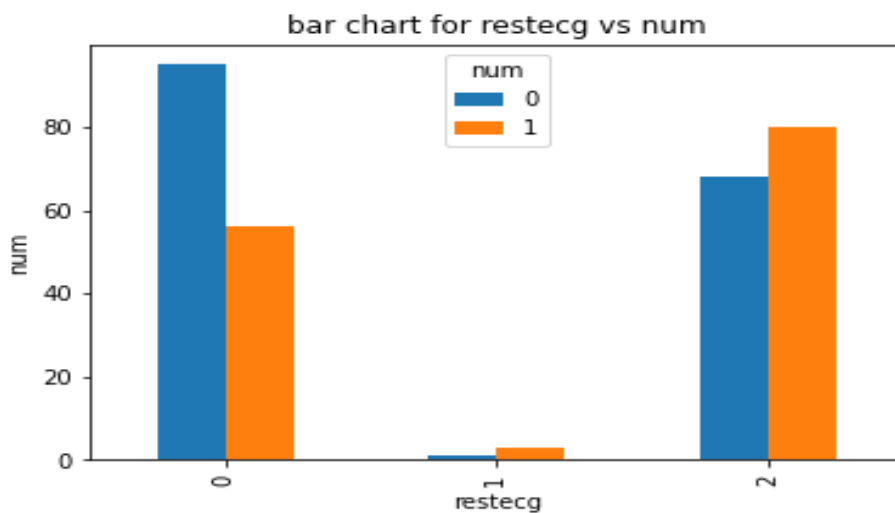


Figure 5. Bar Chart for Restecg vs Num

Table 1 Contains comparison of the future method's accuracy performance with that of existing solutions. A comparison of the suggested approach with other recent developments in

machine learning for heart disease. Proposed system provide better accuracy then naïve bayes and SVM. Random forest algorithm output are project in Figure 2. Random Forest Algorithm

Graphical Representation are these projected in Figure 3. True Positive vs False Positive accuracy

are projected in Figure 4. Restecg vs Num details are illustrate in Figure 5.

Table 1 Proposed Performance Comparison

Algorithm	Accuracy	Predicative	FP Rate
Proposed System	91	87	3
Naive bayes	87	81	7
SVM	85	81	8

5. CONCLUSION AND FUTURE WORK

There are numerous machine learning and data mining approaches available. It is challenging to design and build precise and scalable classifications for medical applications using predictable data mining approaches due to the volume of medical data generated. By incorporating the most modern IT of open-source and big data processing engines into the healthcare system, the issues will be considerably decreased. In order to determine which categorization was the best, we tried to assess the efficiency of various algorithms in terms of constructing time, testing the models' sensitivity, specificity, and accuracy. LG performs best with a precision rate of 97.50%, followed by SVM and RF. In conclusion, LG has proven to be successful in foretelling heart disease and performs best in terms of accuracy. However, this work shows that Python Anaconda is a quick information processing framework, mainly when used with plenty of data. Using this classification algorithm, we evaluated the precision of the data generated by mentioned algorithm SVM, LR, DT, and RF. We found that, among these four, Logistic Regression generated the most sensitive and accurate data from the pre-processing.

References

[1] CHEN Shao-Hua, "A Novel Approach for Building Open Library Student Credit Evaluation System" – 2012
 [2] Dimitri N. Kopaliani "Digital Asset Management System for Libraries" – 2007
 [3] Heru Supriyono" Developing a QR Code-based Library Management System with

Case Study of Private School in Surakarta City Indonesia" – 2018
 [4] Maneesh Kumar Bajpai "Researching through QR Codes in Libraries" – 2015
 [5] Nyoman Karna "Self Service System for Library Automation"- 2019
 [6] Rachna Patnaik "Role of Content Management Software (CMS) in Libraries for Information Dissemination" – 2015
 [7] Sparse Support Vector Machine for Intrapartum Fetal Heart Rate Classification., Jordan Frecon , Roberto Leonarduzzi , Nelly Pustelnik, Patrice Abry , Muriel Doret.
 [8] Changes in Daily Measures of Blood Pressure and Heart Rate Improve Weight-based Detection of Heart Failure Deterioration in Patients on Telemonitoring., Rohan Joshi and Illapha Cuba Gyllensten
 [9] Heart Disease Prediction using Evolutionary Rule Learning., Aakash Chauhan, Aditya Jain, Purushottam Sharma, Vikas Deep
 [10] Ramakrishnan S, Muthanantha Murugavel A S, Sathiyamurthi P, Ramprasath J, Seizure Detection with Local Binary Pattern and CNN Classifier, Journal of Physics: Conference Series, Vol. 1767, Issue 012029, 2021
 [11] Balasamy K., Krishnaraj N, Ramprasath J, Ramprakash P, "A secure framework for protecting clinical data in medical IoT environment", Smart Healthcare System Design: Security and Privacy Aspects, 2021.
 [12] Vaishnavi P, Shebin Sharief, J. Ramprasath, "Second Wave Covid-19: Survey on Attitudes and Acceptance of Vaccines and Psychological Impact of Second Wave", International Journal of Current Research and Review, Vol 14(5), March 2022

- [13] M Balakrishnan, AB Christopher, AS Murugavel, J Ramprasath, "Prediction of Data Analysis Using Machine Learning Techniques", *Int. J. of Aquatic Science*, Volume 12, Issue 3, pp. 2755-2762, 2021
- M Balakrishnan, AB Christopher, AS Murugavel, J Ramprasath, "Biological system Administrations for Compelling Utilize of Information Driven Modeling", *Journal of Education: Rabindra bharati University*, Vol.23, No.12, pp: 251-259, 2021.
- [14] Heru Supriyono "Developing a QR Code-based Library Management System with Case Study of Private School in Surakarta City Indonesia" - 2018
- [15] Attia, Z. A., & Khedr, A. E. (2020). Heart disease diagnosis using ensemble learning algorithms. *Neural Computing and Applications*, 32(23), 17271-17280