IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research Paper , Journal Volume 10, Iss 08, 2021

PREDICTING HEART FAILURE USING CLASSIFICATION METHODS

#1**Mr.VANGAPALLI RAVITEJA,** *Assistant Professor*

#2**Dr.NALLA SRINIVAS,** *Associate Professor*

Department of Computer Science and Engineering,

SREE CHAITANYA INSTITUTE OF TECHNOLOGICAL SCIENCES, KARIMNAGAR, TS.

Abstract— Many cardiovascular diseases are fatal, thus early detection and treatment are crucial. The most frequent disease, heart failure, has a high fatality rate and requires meticulous monitoring and treatment. Recent advances in machine intelligence and deep learning have expanded heart failure treatment options. However, unexpected variables may cause estimates to be inaccurate, with catastrophic consequences. To fix the problem, the scientists used a dataset with thirteen crucial failure prediction variables. In this study, prediction models include SVM, decision tree, k-nearest neighbors, random forest classifier, and logistic regression. This study seeks the most accurate categorization model.

Keywords—Comparative analysis, machine learning, Heart disease prediction, random forest, Classification

1. INTRODUCTION

Because the heart is necessary for life, a significant number of people suffer from heart failure. The condition known as congestive heart failure is brought on by inadequate blood circulation in the cardiac muscle. Dyspnea is caused by blood flowing in an anticlockwise direction because it prevents lung fluid from building up. Weakness, nausea, fatigue, and an irregular heartbeat are some of the symptoms associated with heart failure. It is possible to lower the likelihood of treatment failure by engaging in regular exercise, reducing sodium intake, and losing weight. The clinical and research information collected from patients at various healthcare facilities is now readily available. Because patient records are easily accessible, researchers are able to conduct studies. The use of computer technology enables early disease detection, which can help prevent fatalities. The proposed method might be able to help detect and avoid problems early on, which would reduce the severity of the faults.

2. LITERATURE OVERVIEW

The ability to forecast cardiac failure is made easier by artificial intelligence. Recent developments in machine learning, such as neural networks, support vector machines (SVM), and k-nearest neighbors, allow for improved early disease diagnosis, including heart failure. According to the statistics, each model has an accuracy of more than 80%. The accuracy of SVMs was 99.3%, while that of neural networks was 91.1%, and that of K-Nearest Neighbors was 87.7%. The K-Nearest Neighbors (KNN) approach had an accuracy of 85.5%, the Support Vector Machine (SVM) method had an accuracy of 90%, and the Neural Network method had an accuracy of 93% during testing. The Neural Network provides the most precise results. The purpose of this workshop is to show how machine learning may be used to predict heart failure based on medical data. It is essential because machine learning models take into account a patient's medical history, including their BMI, smoking habits, and overall health state. Following the acquisition of data, algorithms produce predictions. This method can aid in the early detection of heart failure, as well as the prediction of pulmonary difficulties and strokes.

3. METHODOLOGY

DATA PREPARATION

The data for this study came from the heart failure prediction dataset, which is a well-known resource available online. Because there was a limited amount of data, the model could only be applied

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research Paper Journal Volume 10, Iss 08, 2021

to legitimate data when creating a forecast. The dataset was broken down into categories such as age, anemia, diabetes, ejection fraction, and high blood pressure, as seen in the graph. There is a connection between each of the 300 photographs and one of the thirteen criteria.

```
age
anaemia
creatinine_phosphokinase
diabetes
ejection_fraction
high_blood_pressure
platelets
serum_creatinine
serum_sodium
sex
smoking
time
DEATH_EVENT
```

Figure 1: Brief Features and Qualities

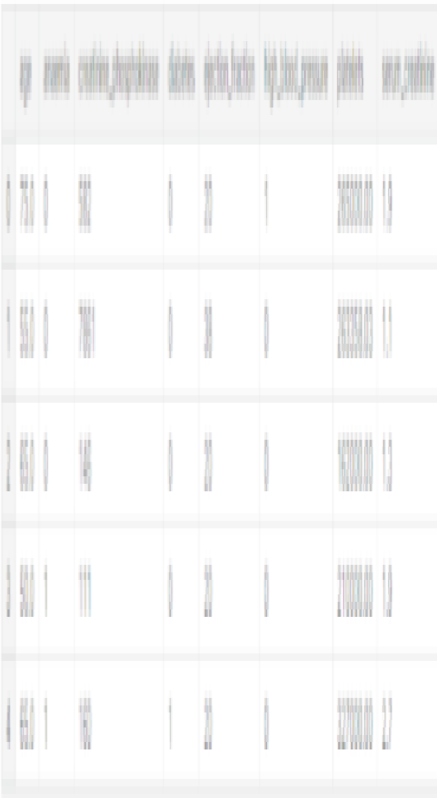

Figure 2: Dataset summary.

DATA PREPROCESSING

Preprocessing improves real-world and our data by fixing errors in both datasets. The operation will be faster with preprocessed data. Additional preprocessing improves model performance. The author checks for null values before removing the id column, however this does not affect the result.

FEATURE SELECTION

The algorithm must have properties to produce accurate results. Visualization lets you analyze various attributes and their effects on results. Figure 3 shows traits and their effects. The first pie graph shows smoking patients who lived or died. Then, diabetes, anemia, and high blood pressure patients are shown. To discover how much each trait or feature affects outcomes, all traits or features must be linked. Ejection fraction, serum creatinine, and time are the three elements most strongly correlated with a patient's survival outcome (DEATH_EVENT), both positively and negatively, according to Figure 4. Thus, by splitting the data into a training set and a testing set, the three needed attributes can be consciously selected.

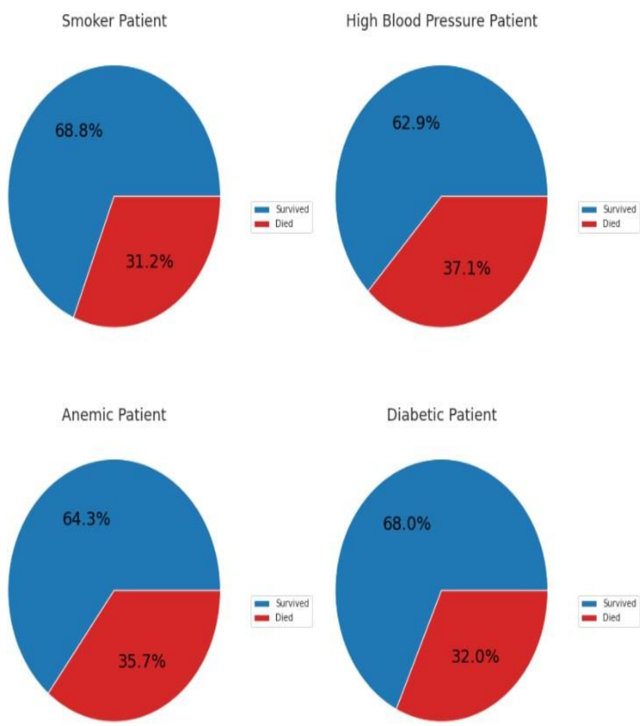

Figure 3: Visualizing qualities with pie charts is the topic.

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research Paper , Journal Volume 10, Iss 08, 2021

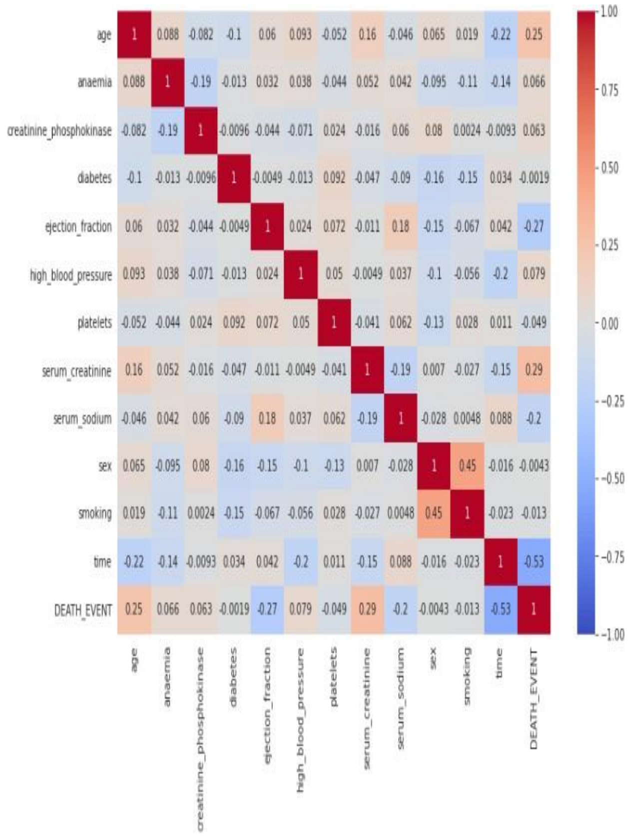

Figure 4: Statistical tools like correlation matrices investigate dataset variables' correlations. The pairwise correlations between each possible combination are displayed.

MODEL ARCHITECTURE

RANDOM FOREST CLASSIFIER

This is called modeling with an ensemble. The use of classification and regression can lead to the development of a reliable prediction model. The estimation in this work is carried out using decision trees. The predictive power of decision trees is increased when they are coupled with one another. In challenges involving categorization, decision trees are used to classify the input data. The output of the decision tree is determined by either the average forecasts of the regression assignments or the class that occurs most frequently. This method minimizes the need for adjusting the parameters and overfitting the data.

SUPPORT VECTOR MACHINE

This type of modeling is known as ensemble modeling. The creation of a reliable prediction model can be accomplished through the application of classification and regression techniques. In this study, estimation is accomplished by the utilization of decision trees. When linked with one another, decision trees have a greater capacity for accurate prediction than they do on their own. Decision trees are utilized in competitions involving categorization to classify the incoming data in the appropriate categories. The output of the decision tree is either the average forecasts of the regression assignments or the class that happens most frequently. Both of these factors are taken into consideration. The necessity of modifying the parameters and overfitting the data is reduced to a minimum when using this strategy.

NAIVE BAYES CLASSIFIER

This method of categorization makes use of a group of algorithms, all of which are derived from the Bayes theorem. Because it's a classifier, it's purpose is to differentiate between different things based on particular qualities that they share. The task of classification was Bayes' primary responsibility when employing the Bayes theorem.

DECISION TREE CLASSIFIER

Many Bayes theorem-based algorithms are used in this classification technique. The classifier uses unique qualities to distinguish things. The Bayes theorem is crucial for categorization.

KNN

A decision tree employs a tree-like form to show multiple options and their outcomes. Chance, resource, and utility effects are examples. Example: Show an algorithm that employs only conditional control statements.

Logistic Regression Classifier

After determining the total number of clusters, or k, the next step is to compute the centers, or centroids, of each cluster. It is possible to find the initial centroid by selecting an object at random from among all the objects in the set of the first k objects. As a result, the method is broken down into three stages: locating the coordinates of the centroid, computing the distance that separates each item from the centroid, and placing the items in descending order according to the shortest

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research Paper , **Journal Volume 10, Iss 08, 2021**

distance. To locate a centroid, proceed through these procedures.

4. EXPERIMENTAL RESULTS

A logistic regression classifier uses one or more input variables to predict a category output. Logistic regression models, used in academic research, produce binary results like true or false, yes or no, etc. Instead of binary classification, multinomial logistic regression characterizes events with several discrete outcomes. Logistic regression is a powerful analysis method.

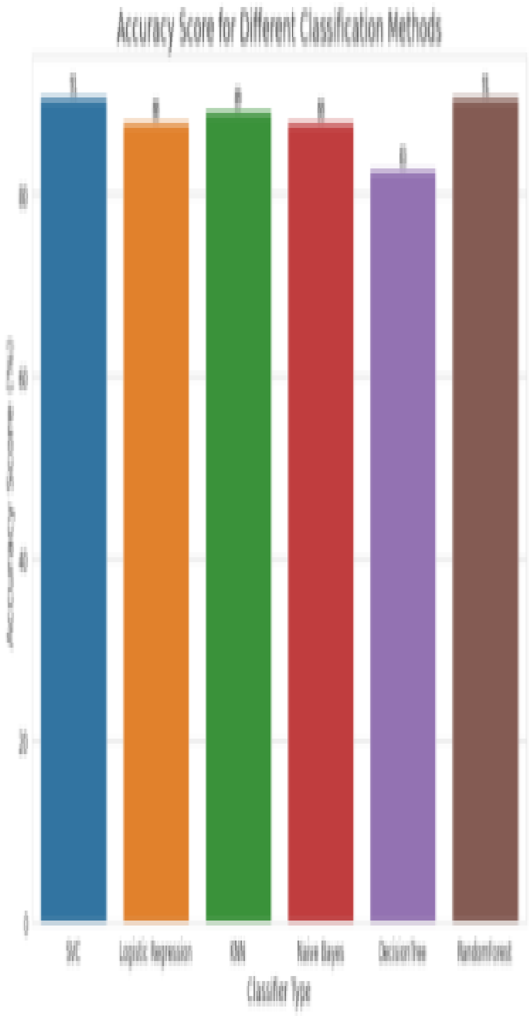

Figure 4: Accuracy chart

5. CONCLUSION

Choosing and preparing a dataset to improve model accuracy. The summary below shows that the selected models are precise. Random Forest is most accurate (91%), followed by Naive Bayes (88%), Decision Tree (83%), Support Vector Machine (91%), K Nearest Neighbor (89%), and Logistic Regression (88%). Our data suggests that random forest and support vector classifier (SVC) meet the criteria.

REFERENCES

1. Rohit Bharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, Parneet Singh, Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning, Computational Intelligence and Neuroscience, vol. 2021, Article ID 8387680, 11 pages, 2021. https://doi.org/10.1155/2021/8387680.
2. S. Sarkar and J. Koehler, A Dynamic Risk Score to Identify Increased Risk for Heart Failure Decompensation, in IEEE Transactions on Biomedical Engineering, vol. 60, no. 1, pp. 147-150, Jan. 2013, doi: 10.1109/TBME.2012.2209646.
3. B. Wang et al., A Multi-Task Neural Network Architecture for Renal Dysfunction Prediction in Heart Failure Patients With Electronic Health Records, in IEEE Access, vol. 7, pp. 178392-178400, 2019, doi: 10.1109/ACCESS.2019.2956859.
4. B. Gnaneswar and M. R. E. Jebarani, A review on prediction and diagnosis of heart failure, 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 2017, pp. 1-3, doi: 10.1109/ICIIECS.2017.8276033.
5. J. Aljaaf, D. Al-Jumeily, A. J. Hussain, T. Dawson, P. Fergus and M. Al-Jumaily, Predicting the likelihood of heart failure with a multi level risk assessment using decision tree, 2015 Third International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE), 2015, pp. 101-106, doi: 10.1109/TAEECE.2015.7113608.
6. K. Arunaggiri Pandian, T. S. Sai Kumar, S. P. Dhandare and S. Thabasum Aara, Development and Deployment of a Machine Learning Model for Automatic Heart Failure Prediction, 2021 Asian Conference on

ISSN PRINT 2319 1775 Online 2320 7876
, **Journal Volume 10, Iss 08, 2021**

Innovation in Technology (ASIANCON), 2021, pp. 1-6, doi: 10.1109/ASIANCON51346.2021.9544787.

7. S. Modi and M. H. Bohara, Facial Emotion Recognition using Convolution Neural Network, 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1339- 1344, doi: 10.1109/ICICCS51141.2021.9432156. 19.9064070.
8. X. Sang, Q. Z. Yao, L. Ma, H. W. Cai and P. Luo, Study on survival prediction of patients with heart failure based on support vector machine algorithm, 2020 International Conference on Robots & Intelligent Syste(ICRIS), 2020, pp. 636-639, doi:10.1109/ICRIS52159.2020.00160.
9. C. B. Rjeily, G. Badr, A. H. A. Hassani and E. Andres, Predicting heart failure class using a sequence prediction algorithm, 2017 Fourth International Conference on Advances in Biomedical Engineering (ICABME), 2017, pp. 1-4, doi: 10.1109/ICABME.2017.8167546.
10. M. G. Asogbon, O. W. Samuel, S. Chen, P. Feng and G. Li, A Hybrid Approach Based on Non-parametric Attribute Learning Technique and Multi-layer Networks for Congestive Heart Failure Risk Prediction, 2019 IEEE 5th International Conference on Computer and Communications (ICCC), 2019, pp. 257-261, doi10.1109/ICCC47050.20