

Heart Failure Prediction using classification Techniques

Vishal Naidu

Department of Electronics and Telecommunication
Ramrao Adik Institute of Technology
Mumbai, India

Abstract—Heart Diseases are considered to be life-threatening and should be recognized at an early stage to make it less fatal. The most common disease is heart failure, and it is the most fatal of all and needs to be taken care of. There are many methods of treatments available for heart failures, and now machine learning and deep learning have also taken it a step forward. But sometimes due to unnecessary circumstances the prediction can go wrong and can be very fatal. To avoid that thing the authors have taken the dataset which consist of 13 main attributes/features used to predict the failure and the models which have been used to predict it are support vector machine, decision tree, knn, random forest classifier and Logistic Regression. The paper aims to provide the best model out of all the classification models with the help of the final accuracy.

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

I. INTRODUCTION

Heart is the major component of the human survival system and the main issue faced by a human is a failure of it. It is also known as congestive heart failure, and generally the thing occurs when the muscle doesn't pump the blood up to the mark. And when such a thing occurs the blood backs up and the fluid can't build up in the lungs causing shortness of breath. The main symptoms of heart failure can be Fatigue, nausea, irregular heartbeat, weakness etc. Treatment can definitely create an impact on the reduction of the failure by changing the daily sedentary lifestyle such as losing weight, reducing salt and exercising. But as time is passing, a lot of research data and patient records of hospitals are available. There are many open sources for accessing the patient's records and researches can be conducted so that various computer technologies could be used for doing the correct diagnosis of the patients and detect this disease to stop it from becoming fatal.[1]Hence the system proposed would help in the prediction of failures at an early stage and stop it from being more fatal.

II. LITERATURE OVERVIEW

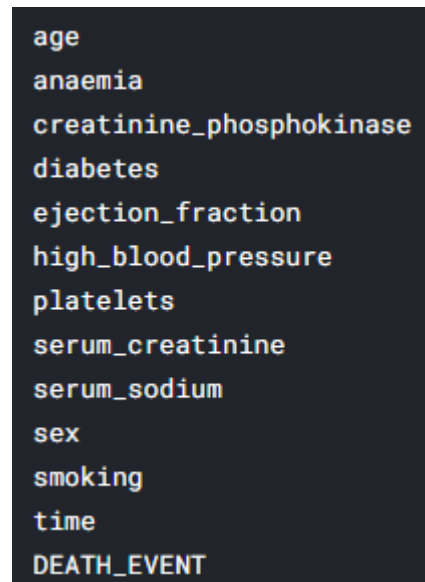
Due to the usage of Artificial Intelligence, it becomes easier for users to predict Heart Failure. By using machine models such as Neural Network, SVM and KNN predicting a Heart failure or disease has become more accurate and easy to catch at its initial stage[2]. According to [3], the accuracy level > 80% of each model. Which is SVM (99.3%), Neural Networks (91.1%) and KNN (87.2%). But while testing they concluded that, Neural Network showed 93% accuracy, SVM showed 90% and KNN showed 85.5% accuracy. Here we can notice that the Neural Network shows maximum accuracy. The whole idea for this discussion is to put stress on how helpful machine learning

techniques are to predict heart failure using the medical data or data set. One of the most important parts of any Machine Learning model is the data set in it, which is where the patient's previous medical history like BMI data, smoking consumption, health status etc. [4]Using this data the algorithms will make predictions. Therefore, this technique can help users to predict any Heart Failure Disease at its initial stage and similar ways can be used for predictions like lung diseases, and Stroke.

III. METHODOLOGY

A. DATA PREPARATION

The data was gathered from the heart failure prediction dataset, which is well-known on the internet. Because data is scarce, the only way to run the model and produce a forecast was to obtain data from a trustworthy source. Various characteristics are included in the dataset, including age, anaemia, diabetes, ejection_fraction, high_blood_pressure and others can be seen in the figure. The dataset contains 300 images with 13 features in it.



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

Figure 1: Features/Attributes Overview

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine
0	75.0	0	582	0	20	1	265000.00	1.9
1	55.0	0	7861	0	38	0	263358.03	1.1
2	65.0	0	146	0	20	0	162000.00	1.3
3	50.0	1	111	0	20	0	210000.00	1.9
4	65.0	1	160	1	20	0	327000.00	2.7

Figure 2: Dataset Overview

B. DATA PREPROCESSING

Because both the real world and our data include faults, preprocessing is an excellent way to enhance it. Whether the data has been preprocessed determines the method's speed. The more preprocessing that is done, the better the model that is used will be. The author first checks for all null values before removing the id column, which will not affect the results.

C. FEATURE SELECTION

Features are critical for obtaining accurate results from the algorithm. Visualization allows us to observe the many characteristics and how they affect the outcomes. Figure 3 shows different features and its visualized results, the first pie chart shows the smoker patients who died and survived followed by High Blood pressure, anemic patients, and diabetic patients. After that, a correlation is necessary to have between all the attributes/features to get an idea about which feature has the most impact on the results. From the heatmap in figure 4, we can see that the 3 features most correlated (both positively and negatively) with a patient's survival outcome (DEATH_EVENT) are ejection fraction, serum_creatinine and time. Therefore, when splitting the data into a train and test set, we can explicitly select the three features mentioned instead of the complete dataset.

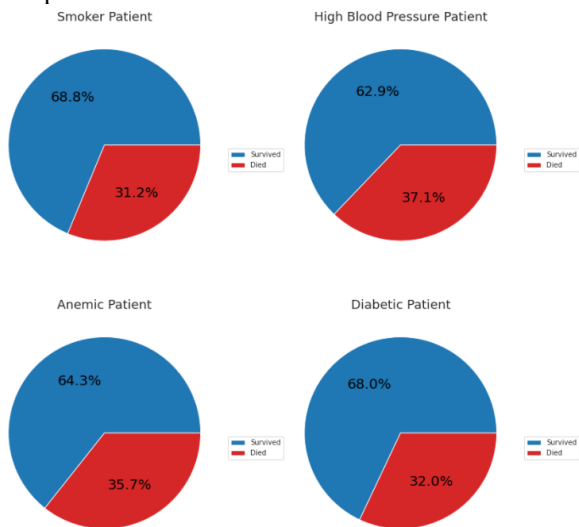


Figure 3: Attributes Pie chart

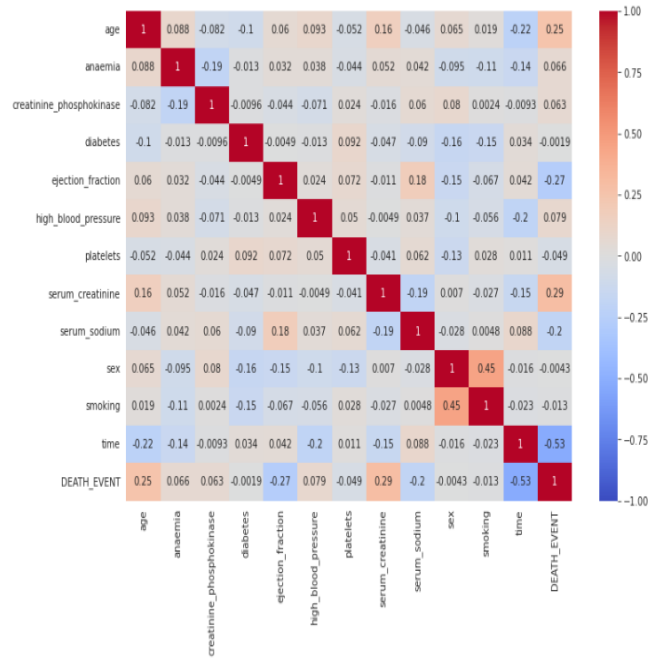


Figure 4: Correlation Matrix

D. MODEL ARCHITECTURE

1) RANDOM FOREST CLASSIFIER

It's a technique that falls under the umbrella of ensemble modelling. It may be used to develop a good prediction model by combining classification and regression approaches. In this work, decision trees are used as the foundation estimators. On their own, decision trees are a poor predictor, but when combined with other decision trees, they improve. Decision trees vote on how to categorize a single instance of input data in classification tasks, and they output the class that is the mode of the classes or the mean of forecasts in regression tasks. This manner, we may prevent parameter tinkering and reduce overfitting.

2) SUPPORT VECTOR MACHINE

To construct nonlinear class borders, Support Vector Machines use a linear model. To identify the target classes, support vectors (lines or hyperplanes) are employed. In order to deal with nonlinear situations, the model uses a mapping function to apply numerous transformations to the input before training a linear SVM model to classify the data in a higher-dimensional feature space..

3) NAIVE BAYES CLASSIFIER

A suite of algorithms based on the Bayes theorem is used in this classification strategy. It's a classifier, therefore it's used to distinguish between items based on particular characteristics. With the use of the Bayes theorem, Bayes' principal duty is classification.

4) DECISION TREE CLASSIFIER

A decision tree is a decision-making tool that employs a tree-like representation of alternatives and their potential outcomes, such as chance event outcomes, resource costs, and utility. It's one approach to show an algorithm composed entirely of conditional control statements.

5) KNN

The number of clusters k would be chosen first, followed by the cluster's centroid. The starting centroid might be any random item or the first k objects in a sequence. As a result, the technique is split down into three steps: first, we discover the centroid's coordinates, then we compute how far each item is from the centroid, and finally, we group the items based on the minimum distance. Following the procedure, we may obtain a centroid.

6) Logistic Regression Classifier

A logistic regression classifier is a technique for calculating the probability of a discrete output given an input variable. The most popular logistic regression models yield a binary outcome, such as true or false, yes or no, etc. Multinomial logistic regression can be used to model events having more than two discrete outcomes. Logistic regression is a valuable approach of analysis.[11]

IV. EXPERIMENTAL RESULTS

The procedure for selecting a dataset and preprocessing it in order to improve the accuracy of the models chosen. As seen in the table below, the models chosen have great accuracy. Starting with the Random Forest classifier, which has a 91 percent accuracy, naive bayes has an 88 percent accuracy, decision tree has an 83 percent accuracy, support vector machine has a 91 percent accuracy, k nearest neighbour has an 89 percent accuracy, and logistic regression has an 88 percent accuracy. As a result of our observations, we may conclude that random forest and SVC perform as well for the given definition.

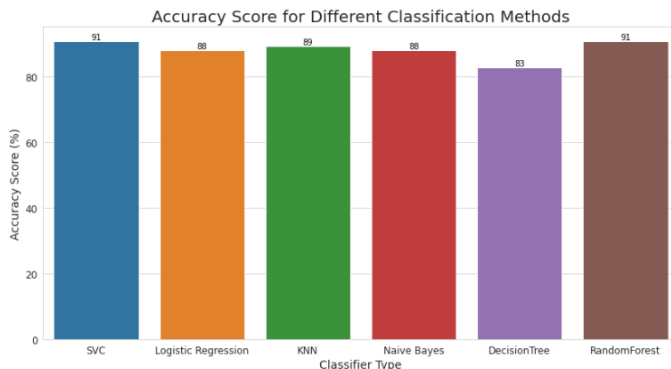


Figure 4: Accuracy chart

V. CONCLUSION

After testing all the models, the authors conclude that SVC(Support Vector Classifier) and Random Forest classifier has the greatest accuracy and performs the best compared to support vector machine, decision tree, knn, random forest classifier and Logistic Regression. Other researchers would utilise the accuracy to choose the best model among the others, and it would undoubtedly aid in the treatment of Heart failure in the majority of cases.

VI. 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] D. E. Salhi, A. T. and M.-T. Kechadi, "Heart Failure Prediction".
[3] F. S. Alotaibi and S. A. Jeddah, "Implementation of Machine Learning Model to Predict Heart Failure Disease".
[4] M.M.S.D. Saqib Ejaz Awan, "Machine learning-based prediction of heart failure readmission or death: implications of choosing the right model and the right metrics".
[5] 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.
[6] 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.
[7] 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.
[8] A. 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 (TAECE), 2015, pp. 101-106, doi: 10.1109/TAECE.2015.7113608.
[9] 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 Innovation in Technology (ASIANCON), 2021, pp. 1-6, doi: 10.1109/ASIANCON51346.2021.9544787.
[10] G. Valenza et al., "Mortality Prediction in Severe Congestive Heart Failure Patients With Multifractal Point-Process Modeling of Heartbeat Dynamics," in IEEE Transactions on Biomedical Engineering, vol. 65, no. 10, pp. 2345-2354, Oct. 2018, doi: 10.1109/TBME.2018.2797158.
[11] 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
[12] P. -Y. Liang, L. -J. Wang, Y. -S. Wu, T. -W. Pai, C. -H. Wang and M. -H. Liu, "Prediction of patients with heart failure after myocardial infarction," 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2020, pp. 2009-2014, doi: 10.1109/BIBM49941.2020.9313253.
[13] 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 System (ICRIS), 2020, pp. 636-639, doi: 10.1109/ICRIS52159.2020.00160.
[14] 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
[15] 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, doi: 10.1109/ICCC47050.2019.9064070.