Performance analysis of Machine Learning Models for Heart Disease

~ Tanu Goswami

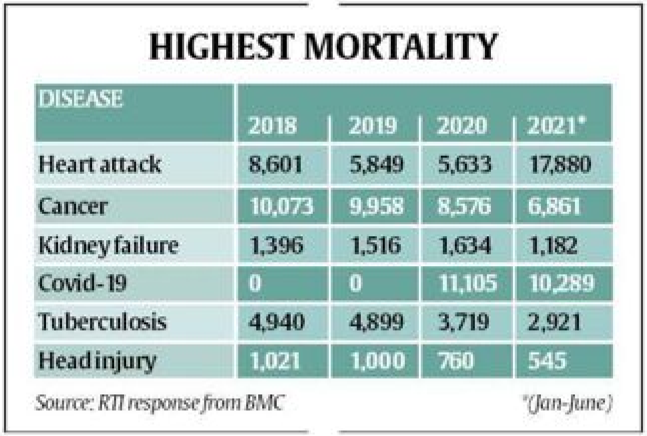
Abstract :

Studies have shown that heart diseases have emerged as the number one cause of death. Heart disease is accountable for deaths in all age groups and is common among males and females. A good solution to this problem is to be able to predict what a patient's health status will be like in the future so the doctors can start treatment much sooner which will yield better results. It's a lot better than acting at the last minute when the patient is already at risk; hence, the prediction of heart disease is widely researched. A lot of research and technological advancement has been recorded in similar fields. This paper aims to report about taking advantage of the various models and develop prediction models for heart disease survivability.

Keywords- machine learning, Heart disease, performance analysis, Prediction, Classification.

Introduction :

Chronic diseases are one of the curses mankind is facing today. Heart disease, Diabetes, and Stroke are some examples of chronic diseases. According to the statistics of the World Health Organization, death due to heart disease is top of the top 10 leading causes of death. This happens mainly because of two reasons; lack of proper medical resources in the hospital and failure of giving proper medical care to the patients at the correct time. Failure to identify the initial stages of heart disease is one of the main reasons why medical professionals are not able to give proper medical care to the patient. Constant monitoring and early detection of chronic diseases are important to avoid the risk related to this kind of disease. Heart attack has the highest mortality in 2021. Heart disease in people increasing day by day.



HIGHEST MORTALITY

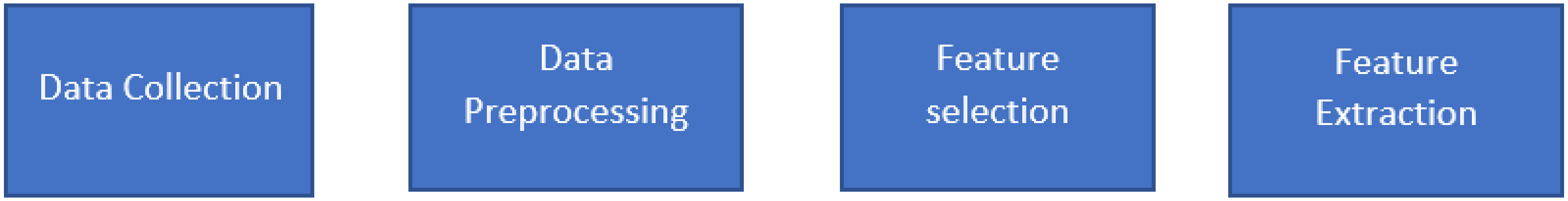
DISEASE	2018	2019	2020	2021*
Heart attack	8,601	5,849	5,633	17,880
Cancer	10,073	9,958	8,576	6,861
Kidney failure	1,396	1,516	1,634	1,182
Covid-19	0	0	11,105	10,289
Tuberculosis	4,940	4,899	3,719	2,921
Head injury	1,021	1,000	760	545

Source: RTI response from BMC

*(Jan-June)

Thus for monitoring purposes, we trained our machine to predict the presence and absence of Heart Disease using some models like SVC, Random Forest, Decision Tree, etc.

Proposed methodology



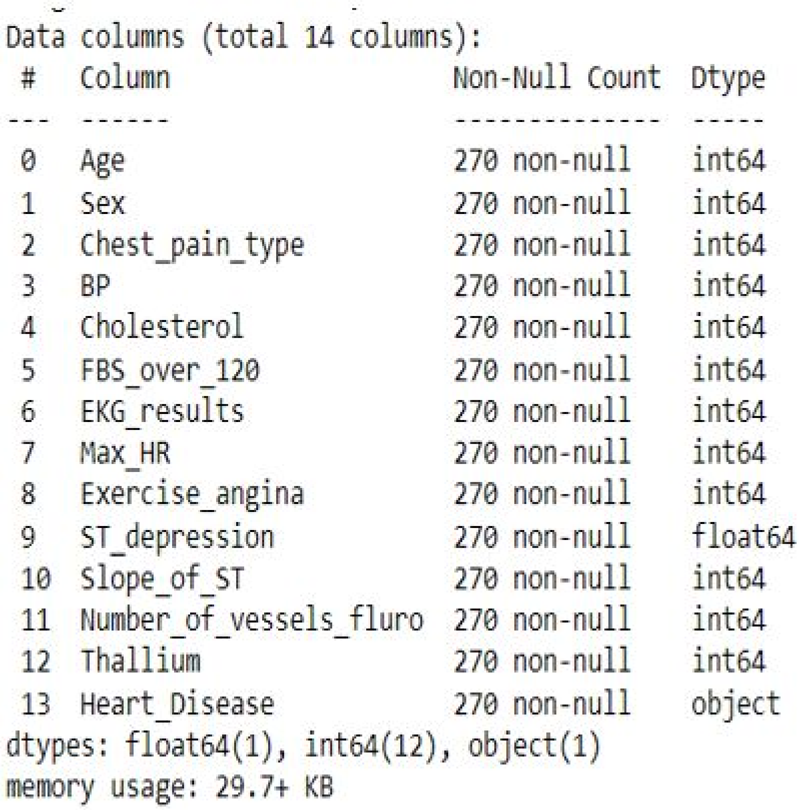
Data Collection

For this project, the data has been downloaded from Kaggle. It consists of 14 classes namely Age, Sex, Chest pain type, FBS over 120, EKG results, Max HR, Exercise Angina, ST depression, Slope of ST, Number of vessels Fluro, thallium, and heart disease. There are a total of 270 samples divided into 2 classes as shown in the figure

Presence (1) =151
Absence (0) =119
Name: Heart Disease

Data Preprocessing

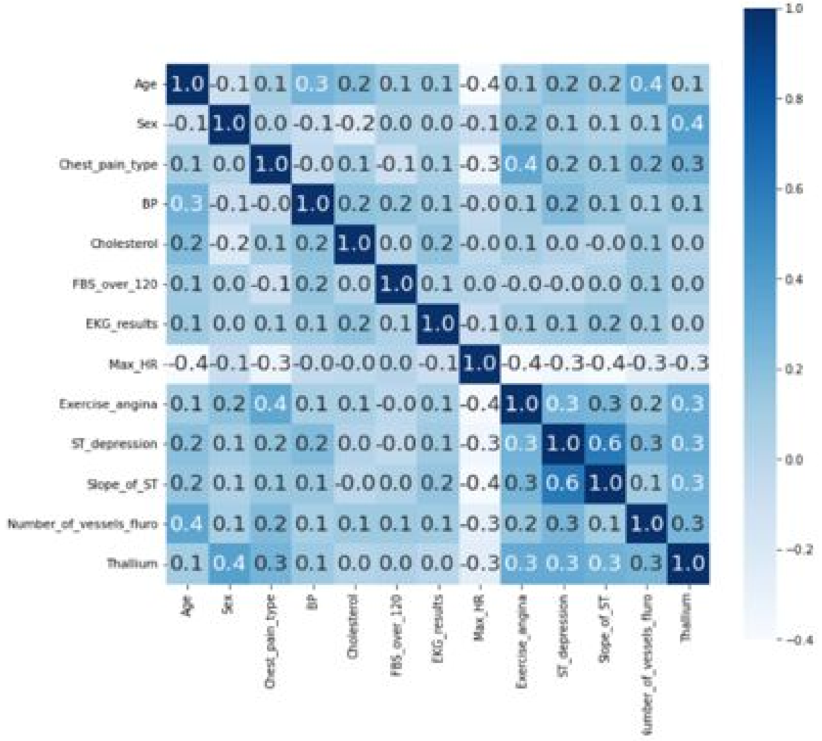
The data is preprocessed it is required to make one change in the dataset i.e. Heart Disease. We have taken presence as 1 and absence as 0.



```
Data columns (total 14 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   Age                      270 non-null    int64
 1   Sex                      270 non-null    int64
 2   Chest_pain_type          270 non-null    int64
 3   BP                       270 non-null    int64
 4   Cholesterol              270 non-null    int64
 5   FBS_over_120             270 non-null    int64
 6   EKG_results              270 non-null    int64
 7   Max_HR                   270 non-null    int64
 8   Exercise_angina          270 non-null    int64
 9   ST_depression            270 non-null    float64
 10  Slope_of_ST              270 non-null    int64
 11  Number_of_vessels_fluro  270 non-null    int64
 12  Thallium                 270 non-null    int64
 13  Heart_Disease            270 non-null    object
dtypes: float64(1), int64(12), object(1)
memory usage: 29.7+ KB
```

Feature selection

For feature selection, we draw a heat map between these parameters of the dataset.



Feature Extraction

PCA (Principle Component Analysis) is used to extract features by reducing the numbers of input variables, we use Random forest with PCA and got an 87.03% accuracy score, Feature importance is also used to find the importance of every parameter.

Result & Discussion

All experiment has been carried out in widows 10 with Intel(R) Core(TM) i5-1135G7 @ 2.40GHz   2.42 GHz, RAM 8 GB. In this analysis, we worked on five Machine Learning Models providing accuracy of 52.06% by Linear Regression,74.07% by SVM, 87% by Naïve Bayes, 72.2%by Decision Tree, and the highest 87.03% by Random forest.

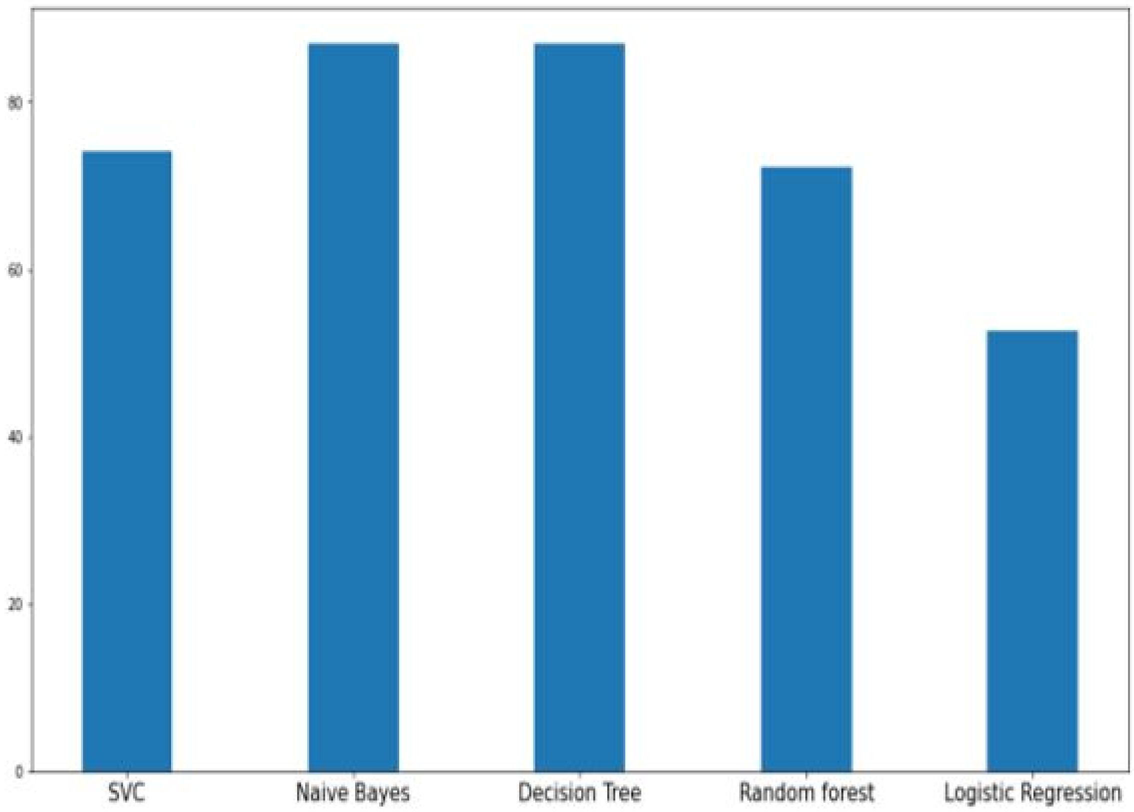


Table 1. Shows the result comparison of various ML techniques

Model	Accuracy	Recall	Precision
SVM	0.74	0.69	0.93
Naïve Bayes	0.87	0.87	0.90
Decision Tree	0.72	0.82	0.62
Random Forest	0.85	0.82	0.93
Linear Regression	0.52	0.45	0.65

Conclusion & Future work

In this study, we illustrate five machine learning models in which Random Forest provides the best accuracy score of 87.03%.

Future work or scope for this project is that it can be further implemented in the form of smart devices which can be deployed in hospitals or other health care centers for early stage Heart Disease detection and these models need to be more trained by large data sets to depict the problem more precisely.

References:

- S. Sreejith, S. Rahul & R. C. Jisha “A Real-Time Patient Monitoring System for Heart Disease Prediction Using Random Forest Algorithm” in 2015
- Sun, S., Huang, R.: An adaptive K-nearest Neighbor algorithm. In: Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), vol. 1 (2010)
- N. Alharbe, A.S. Atkins and A.S. Akbari, "Application of ZigBee and RFID Technologies in Healthcare in conjunction with the Internet of Things", *International Conference on Advances in Mobile Computing & Multimedia*, pp. 191-195, 2013.
- R. Sethukkarasi, S. Ganapathy, P. Yogesh, and A. Kannan, "An intelligent neuro fuzzy temporal knowledge representation model for mining temporal patterns", *Journal of Intelligent & Fuzzy Systems*, vol. 26, no. 3, pp. 1167-1178, 2014.
- S.H. Almotiri, M.A. Khan and M.A. Alghamdi, "Mobile Health (m-Health) System in the Context of IoT", *IEEE International Conference on Future Internet of Things and Cloud Workshops*, pp. 39-42, 2016.