Comparative Analysis of Machine Learning Models for Cardiovascular Disease Prediction

Dr. Neena Nanda

Mr. Ayaskanta Mohapatra

ABSTRACT

In most sectors Machine Learning algorithms have reinforced the diagnostic and prognostic capacity of traditional regression techniques. As with many of those sectors, clinical environment must also make use of machine learning for predicting coronary heart diseases or heart attack. There are various data mining techniques that can be used for the identification and anticipation of heart disease among patients. The focus of the present study is purely methodological, given our overarching goal of exploring the applicability of Machine Learning (ML) models in predicting heart attack. Specifically, the primary objective was to compare internal validity and accuracy of several supervised ML algorithms to the prediction of heart attack. Part of the data used came from an open source on Towards Data Science available on a platform for data scientists that published ML-based prediction projects using publicly available databases. Additional 40 responses were collected from Cardiologists, pathological labs and from people who had suffered with heart attack in and around Odisha. The methodology of machine learning with python is proposed. Python scripting is implemented using the Anaconda jupyter notebook. Six supervised machine learning techniques KNN, SVM, Decision Tree, Random Forest, Logistic Regression, and Naïve Bayes classifier were analysed and tested on this data. Results showed KNN provided best outcomes for heart attack prediction when metrics related to accuracy and error rate were considered.

Keywords: Machine learning, Supervised Machine Learning, KNN, SVM, Decision Tree, Random Forest, Logistic Regression, Naïve Bayes classifier, heart attack prediction.

1. Introduction

In Heath Care industry heart attack is one of the most pressing problems. Cardiac disease has become the initiator of primary health issues and the major cause of death globally as heart is one of the most vital organs in an individual. Psychological problem, unhealthy foodstuff, lack of exercise, environmental factors etc. are some reasons for a heart attack. Doctors use ECG, Stress Test, and Heart MRI etc. for predicting heart disease (Weng, Reps, Kai, Garibaldi, & Qureshi, 2017). According to survey of WHO, seventeen million total global deaths are due to heart attacks and strokes. Work overload, mental stress, and a variety of other issues contribute to heart disease deaths in many countries. In general, it is discovered to be the leading cause of death in adults. Diagnosis is a difficult and crucial process that must be completed correctly and swiftly. Diagnosis is frequently done based on the doctor's expertise and knowledge. This leads to unfavourable outcomes and high medical expenses for therapies given to patients. (S.Dangare & S. Apte, 2012). A patient's analysis is usually based on information including signs, symptoms, and a physical examination. Predicting cardiac disease is one of the most difficult and time-consuming tasks in medicine. If the diagnosis of cardiac disease based on multiple syndromes is inaccurate, it might lead to mistaken assumptions and unexpected outcomes.

Nowadays, the healthcare industry creates a vast amount of data about patients, hospital resources, disease diagnosis, medical devices, and other topics. These data can be extremely crucial in insight extraction, cost-cutting support, and decision-making by timely predictions. The amount of information stored in Healthcare Industry is not feasible to handle manually. For adequate analysis, human intelligence is insufficient. In data mining field these large amounts of data can be used to get useful information and to generate relationships within the attributes (Beunza et al., 2019).The studies of the past are mainly based on a 13-feature dataset. In addition, classification is common in most past studies to predict if a patient has heart disease or not (Cleveland, 2020). A dataset of 340 medical records and 14 attributes (13 inputs and 1 output) are used in this study to test and justify the differences between algorithms. Finally, the optimum algorithm KNN with the relatively highest accuracy is introduced.

2. Background and Literature Review

The provision of high-quality services at a reasonable cost is a significant challenge for healthcare organisations. Early diagnosis and effective treatment procedures for patients are examples of quality services. A proper diagnosis should always be carried out, as a misdiagnosis can have disastrous consequences. Clinical examinations and procedures should also be kept to a minimum. Implementing computer-based information and/or decision support systems can help achieve these goals (Singh, Singh, & Pandi-Jain, 2018). For the management of their healthcare or patient data, many hospitals today use some form of hospital information system. These systems produce a lot of data in the form of numbers, texts, graphs, and images. These data, on the other hand, are rarely used to aid clinical decision-making. Here arises an important aspect; 'How can the data be converted into useful information that can enable healthcare personnel to make intelligent clinical decisions?' This becomes the main motive of the research.

Using a heart illness dataset, Pandey et al. in 2013 proposed the performance of a clustering technique. They looked at how well different clustering algorithms performed and how accurate their predictions were. The manner of classes to clusters evaluation will be used to calculate cluster performance. Finally, they recommended Make Density Based Cluster as the most adaptable algorithm for heart disease diagnosis, with a prediction accuracy of 85.8086 percent.

Comparative analysis of classification tree algorithm like Decision Stump, Random Forest using WEKA, an open source software consisting of a collection of machine learning algorithm for data mining tasks has proposed that bagging improved higher accuracy as compared to boosting. When the weak classifiers are ensembled with majority voting, the accuracy was improved by a maximum of 7.26% and stacking improved the accuracy by a maximum of 6.93%.(Vijayarani & Sudha, 2013).

A neural network is used to construct a heart disease prediction system in a h study and the suggested method incorporated 13 medical parameters. The results of the trials done proved that the suggested algorithm outperforms similar state-of-the-art alternatives (Al-Milli, 2013).

With the help of a large number of input features, Chaitrali S. D. in 2012 investigated a computation framework for cardiac syndrome. A few medical terms were recycled, such as blood pressure, sex, cholesterol, and 13 other qualities, to forecast heart disease in a certain person or patient. He also used two separate characteristics, such as smoking and obesity. For evaluating the heart disease database, different data mining techniques were utilised, such as Decision trees, neural networks, and nave baye's. The accuracy supplied by the system is critical to the coordination of these procedures. Decision tree accuracy is 99.62 percent, neural network accuracy is 100 percent, and nave bayes accuracy is 90.74 percent.

By obtaining HRV and ECG signals, Kiyong Noh et al. (2006) used a classification technique to remove multi-parametric structures. As the cornerstone of this associative technique, Kiyong used the FP-growth algorithm. A degree of rule consistency was achieved, allowing for a strong press on trimming designs in the technique of producing designs.

Alp Aslandogan, et al. (2004) used Dempsters' rule to combine three different classifiers: K-nearest Neighbour (KNN), Decision Tree, and Nave Bayesian. This classification, which is based on a merged idea, has shown to be more accurate.

3. Methodology

In contrast to traditional logistic regression, comparative methodological research was conducted and six supervised machine learning techniques KNN, SVM, Decision Tree, Random Forest, Logistic Regression, and Naïve Bayes classifier were analysed and tested. Part of the data (300 responses) used came from an open source on Towards data science available on a platform for data scientists that published ML-based prediction projects using publicly available databases (Rawat,2021). Additional 40 responses were collected from Cardiologists, pathological labs and from people who had suffered with heart attack in and around Odisha.

In all fourteen clinical attributes were utilised .The importance of these attributes was confirmed after consulting doctors in Odisha. The literature review related to reasons for cardiac problems were also indicative of these parameters.

Description of dataset : The attributes which are used in this research and for what they are used or resemble are described as follows: (i)Age—age of patient in years, sex—(1 = male; 0 = female).(ii)Cp—chest pain type.(iii)Trestbps—resting blood pressure (in mm Hg on admission to the hospital). The normal range is 120/80 (if one has a normal blood pressure reading, it is fine, but if it is a little higher than it should be, then changes in lifestyle are recommended .(iv)Chol—serum cholesterol shows the amount of triglycerides present. Triglycerides are another lipid that can be measured in the blood. It should be less than 170 mg/dL (may differ in different Labs).(v)Fbs—fasting blood sugar larger than 120 mg/dl (1 true). Less than 100 mg/dL (5.6 mmol/L) is normal, and 100 to 125 mg/dL (5.6 to 6.9 mmol/L) is considered prediabetes.(vi)Restecg—resting electrocardiographic results.(vii)Thalach—maximum heart rate achieved. The maximum heart rate is 220 minus your age.(viii)Exang—exercise-induced angina (1 yes). Angina is a type of chest pain caused by reduced blood flow to the heart. Angina is a symptom of coronary artery disease.(ix)Oldpeak—ST depression induced by exercise relative to rest.(x)Slope—the slope of the peak exercise ST segment.(xi)Ca—number of major vessels (0–3) colored by fluoroscopy.(xii)Thal—no explanation provided, but probably thalassemia (3 normal; 6 fixed defects; 7 reversible defects).(xiii) thal - 3 = normal; 6 = fixed defect; 7 = reversable defect. (xiv) target - have disease or not (1=yes, 0=no).

Before writing code, all of the necessary Python libraries like NumPy, Pandas, Scikit-Learn (Sklearn), Matplotlib, Seaborn were imported.

The data of 340 responses was pre-processed in order to use it in the models. In this dataset of 13 features and 1 target variable no missing data was identified. Certain categorical variables which were found not to be useful were changed into dummy columns. The dataset was also scaled so that it is balanced for machine learning algorithms. The data was then divided into training (80%) and test data (20%).The training data was then applied to the classification models, followed by the test data to compare the predicted and true values. The accuracy as well as other characteristics were then calculated. Machine learning algorithms like K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree, Logistic Regression, Navi Bayes, and Random Forest algorithms for classification were carried out.

4. Data Analysis

With the help of rcParams, a feature correlation matrix for data analysis was generated. As shown in Figure-1, pyplot displayed the correlation matrix. Then xtick and ytick were used to add the names to the correlation matrix. Colorbar() was used to display the matrix's color bar. We can see from the graph that some of the features have a negative connection with the goal value, and none of them have a high association with it. To draw the charts, histograms were plotted for scale purposes in Figure.2. It also shows that each feature and label is dispersed throughout several ranges, emphasizing the necessity for scaling. Discrete bars were used to check for categorical variables since its necessary to first take care of the category variables before implementing machine learning

methods. To get a better outcome from machine learning models, the dataset must be generally balanced as an irregular dataset will be of no help. A bar chart was plotted to check it out. Unique values for the X-axis from the target column were considered, then their names were changed with X-ticks. The value count for the Y-axis was determined as in Figure.3. The plots show how each feature and label is distributed along different ranges, which further confirms the need for scaling. Discrete bars, mean that each of these is a categorical variable. Our target labels have two classes, 0 for no disease and 1 for disease.

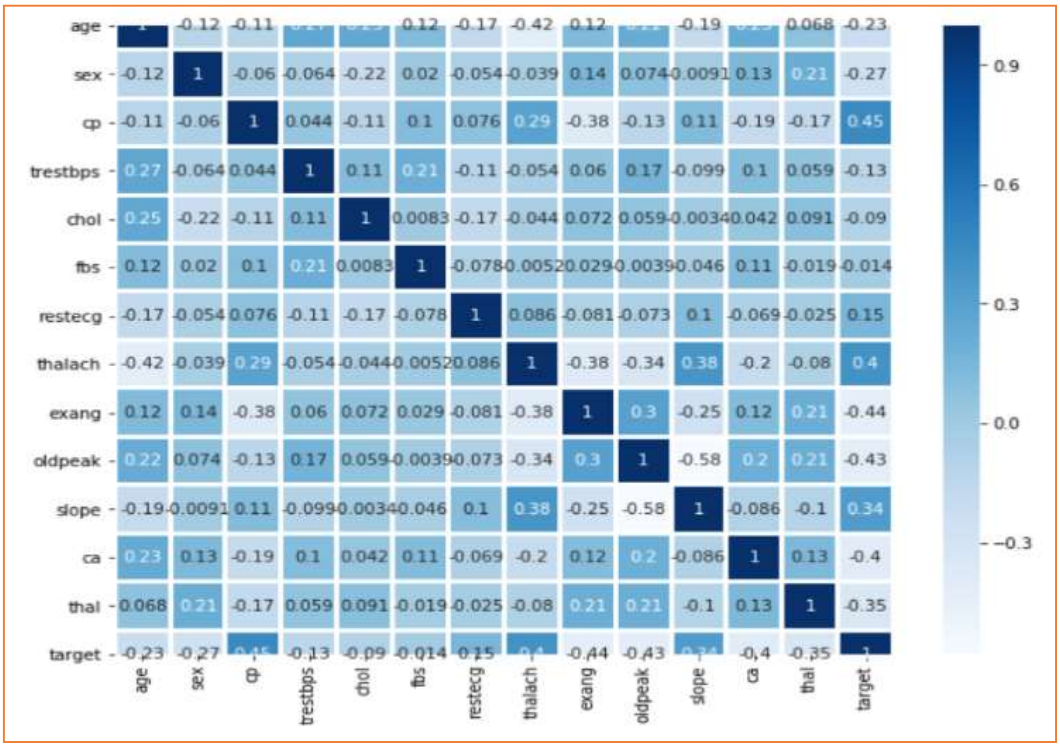

Figure 1: **Correlation Matrix of given dataset**

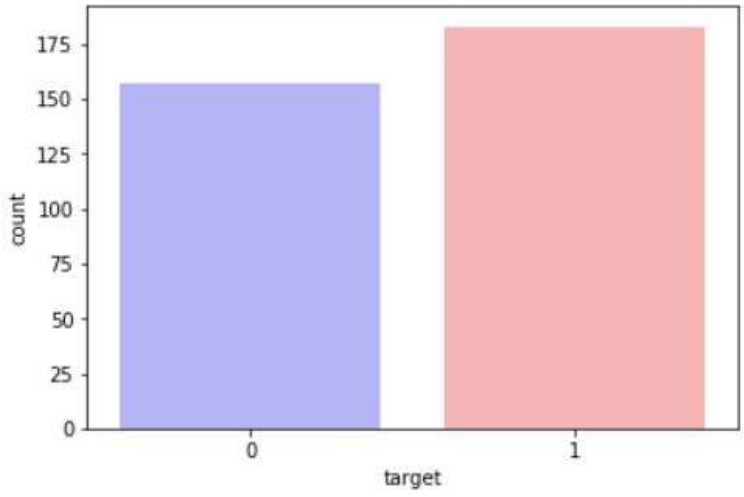

Figure 2: **Number of each target class**

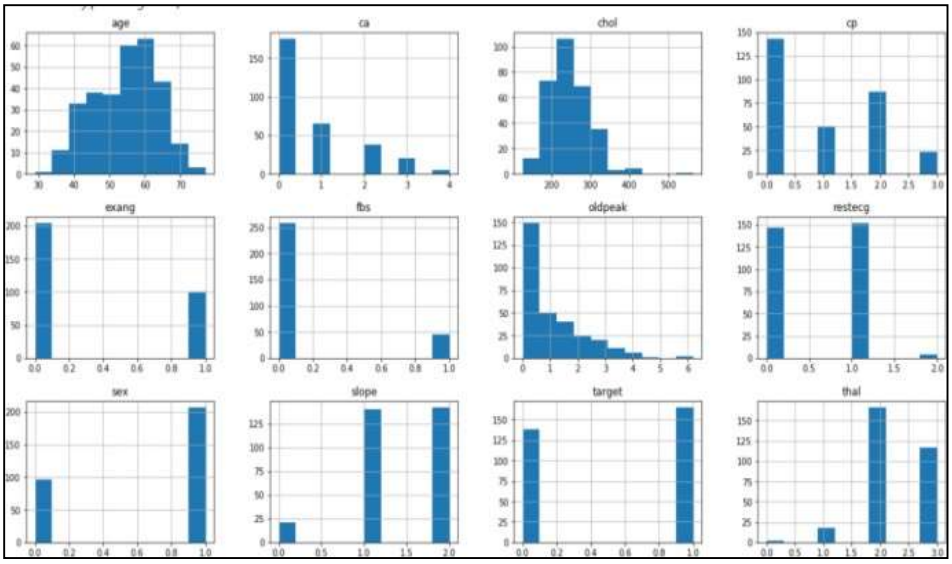

Figure 3: **Feature versus Label**

To initiate with the analysis the study used different types of, machine learning techniques to compare the accuracy rate. Fitting into the appropriate algorithms resulted in the confusion matrix. An N x N matrix is used to evaluate the performance of a classification model, where N is the number of target classes. The matrix compares the actual goal values to the machine learning model's predictions. This provides us with a comprehensive picture of how well our classification model is working and the types of errors it makes. The four values are produced using the confusion matrix.

The number of times the classifier correctly predicts the positive class as positive is referred to as True Positive (TP).

The amount of times the classifier correctly predicts the negative class as negative is referred to as True Negative (TN).

False Positive (FP): This term refers to the amount of times a classifier wrongly predicts a negative class as a positive.

False Negative (FN): This is the number of times the classifier predicts the positive class as negative.

Accuracy is calculated using formula: **(TP+TN)/(TP+TN+FP+FN)**.

To arrive at a conclusion, the performance measurements of various algorithms are compared, and the algorithm with the greatest rates is the one with the best performance metrics.

The relevance and results obtained of applied machine learning algorithms are analyzed here.

4.1 KNN

Both classification and regression predicting issues can be solved with KNN. However, in the industry, it is more commonly employed in categorization difficulties. We normally look at three crucial criteria while evaluating any technique: The output is simple to interpret. Time to calculate, Predictive Ability. KNN is one of the most fundamental yet essential algorithms in machine learning, with applications in pattern recognition, data mining, and intrusion detection, as well as in the supervised learning arena.It makes no assumptions about the data and is typically used for classification jobs when little or no prior knowledge of the data distribution is available. This approach involves locating the k data points in the training set that are the closest to the data point for which a target value has been determined (Masethe & Masethe, 2014).On the basis of majority class, it generally looks for the given data points for the classes of K nearest neighbours. It was found to give an accuracy of 88.24%.

TRUE CLASS

PREDICTED CLASS		POSITIVE	NEGATIVE
	POSITIVE	TP (20)	FP (7)
	NEGATIVE	FN (1)	TN (40)

Table 1: **Confusion matrix for KNN Classifier**

4.2 SVC

Clustering is the process of dividing a data collection into groups based on some criterion in order to organise data into a more comprehensible format. This goal can be accomplished in a variety of ways. Clustering can be done using a parametric model or by

grouping points based on a distance or similarity metric, as in hierarchical clustering. Cluster boundaries are naturally placed in regions of data space where there is little data, i.e. in "valleys" in the data's probability distribution. Support vector clustering (SVC), which is based on the support vector technique, follows this pattern. A kernel function is used in SVC to map data points from data space to a high-dimensional feature space. Using the Support Vector Domain Description algorithm, the algorithm searches the kernel's feature space for the smallest sphere that encloses the image of the data. When projected back to data space, this sphere creates a collection of outlines that encircle the data points. The contours are then read as cluster boundaries, and SVC associates the points encompassed by each contour to the same cluster (Walker, Khan, Katic, Maassen, & Zeiler, 2020) .

We utilise a support vectors classifier to get the optimum hyper plane for the independent classes (Beunza et al., 2019). The hyper plane is determined by a few kernels. For the dataset in question, the top score was 85.29 percent.

TRUE CLASS

PREDICTED CLASS		POSITIVE	NEGATIVE
	POSITIVE	TP (18)	FP (9)
	NEGATIVE	FN (1)	TN (40)

Table 2 : **Confusion Matrix for SVC Classifier**

4.3 Decision Tree Algorithm

Decision Tree is a supervised learning technique that may be used to solve both classification and regression problems, however it is most commonly employed to solve classification issues. Internal nodes represent dataset attributes, branches represent decision rules, and each leaf node provides the conclusion in this tree-structured classifier. The Decision Node and the Leaf Node are the two nodes of a Decision tree. Leaf nodes are the output of those decisions and do not contain any more branches, whereas Decision nodes are used to make any decision and have several branches. The decisions or tests are made based on the characteristics of the given dataset. It's a graphical depiction for obtaining all feasible solutions to a problem/decision depending on certain parameters. It's termed a decision tree because, like a tree, it starts with the root node and grows into a tree-like structure with additional branches. We utilise the CART algorithm, which stands for Classification and Regression Tree algorithm, to form a tree.

The many types of decision trees we have are determined by the type of target variable we have. There are two types of it:

• Categorical Variable Decision Tree: A categorical variable decision tree is a decision tree with a categorical target variable.

• Continuous Variable Decision Tree: A Continuous Variable Decision Tree is one that has a continuous target variable.

A supervised machine learning algorithm where the data is frequently split according to certain parameters. This classifier creates a decision tree that assigns class values to each data point in order to assign class values to each data point. Most of the aspects of the model can be changed while it is being created. After reviewing the results, we can see that the maximum score is 76.47 percent, indicating that the task has been completed (Vijayarani & Sudha, 2013).

TRUE CLASS

PREDICTED CLASS		POSITIVE	NEGATIVE
	POSITIVE	TP (19)	FP (8)
	NEGATIVE	FN (8)	TN (33)

Table 3: **Confusion Matrix for Decision Tree Classifier**

4.4 Random Forest

Random forest is a supervised learning technique that can be used to classify and predict data. However, it is mostly employed to solve categorization issues. A forest, as we all know, is made up of trees, and more trees equals a more healthy forest. Similarly, the random forest method constructs decision trees from data samples, extracts predictions from each, and then votes on the best option. It's an ensemble method that's superior than a single decision tree because it averages the results to reduce over-fitting. This is essentially an ensemble of decision trees that employs decision tree concepts. The total features are used to generate a cluster of trees using this model. The highest score of 80% as accuracy result was found with this algorithm (Decker & Hutchison, 2013; Virtanen et al., 2013; Weng et al., 2017).

TRUE CLASS

PREDICTED CLASS		POSITIVE	NEGATIVE
	POSITIVE	TP (19)	FP (8)
	NEGATIVE	FN (4)	TN (37)

Table 4**: Confusion matrix for Random Forest Classifier**

4.5 Logistic Regression

The relationship between a categorical dependent variable and a group of independent (explanatory) factors is investigated using logistic regression analysis. When the dependent variable has just two values, such as 0 and 1 or Yes and No, logistic regression is utilised. Iother words, as a function of X, the logistic regression model predicts P(Y=1).

As a method for analysing categorical response variables, logistic regression competes with discriminant analysis.Considering the fact that many statisticians believe that logistic regression is more adaptable than discriminant analysis and is better suited to modelling most situations (Collins & Stevens, 2002; Damen et al., 2016; McClelland et al., 2015). As it does not require the independent variables to be regularly distributed this research also carried out logistic regression on the dataset. The accuracy of logistic regression was 82.35 percent.

TRUE CLASS

PREDICTED CLASS		POSITIVE	NEGATIVE
	POSITIVE	TP (17)	FP (10)
	NEGATIVE	FN (2)	TN (39)

Table 5: **Confusion Matrix for Logistic Regression**

4.6 Naive Bayes Classifiers

The Bayes' Theorem is used to create a family of classification algorithms known as Naive Bayes classifiers. This is not a single algorithm, but a collection of algorithms that all follow the same basic principle: each pair of criteria being classed is unrelated to the others. The accuracy of the Naive Bayes classifiers was 69.12 percent.

TRUE CLASS

PREDICTED CLASS		POSITIVE	NEGATIVE
	POSITIVE	TP (23)	FP (4)
	NEGATIVE	FN (17)	TN (24)

Table 6: **Confusion matrix for Naïve Bayes Classifier**

5. RESULT AND ANALYSIS

The accuracy of different algorithms as presented in the table show that highest accuracy score, recorded were 88.24%, 85.29%, 82.35%, 82.35%, 76.47% and 69.12% from the KNN, SVC, Decision Tree , Random Forest, Logistic Regression ,and Naive Bayes respectively.

Sr. No	Classifier	Accuracy (In %)
1	KNN	88.24
2	SVC	85.29
3	Decision Tree	76.47
4	Random Forest	82.35
5	Logistic Regression	85.35
6	Naive Bayes	69.12

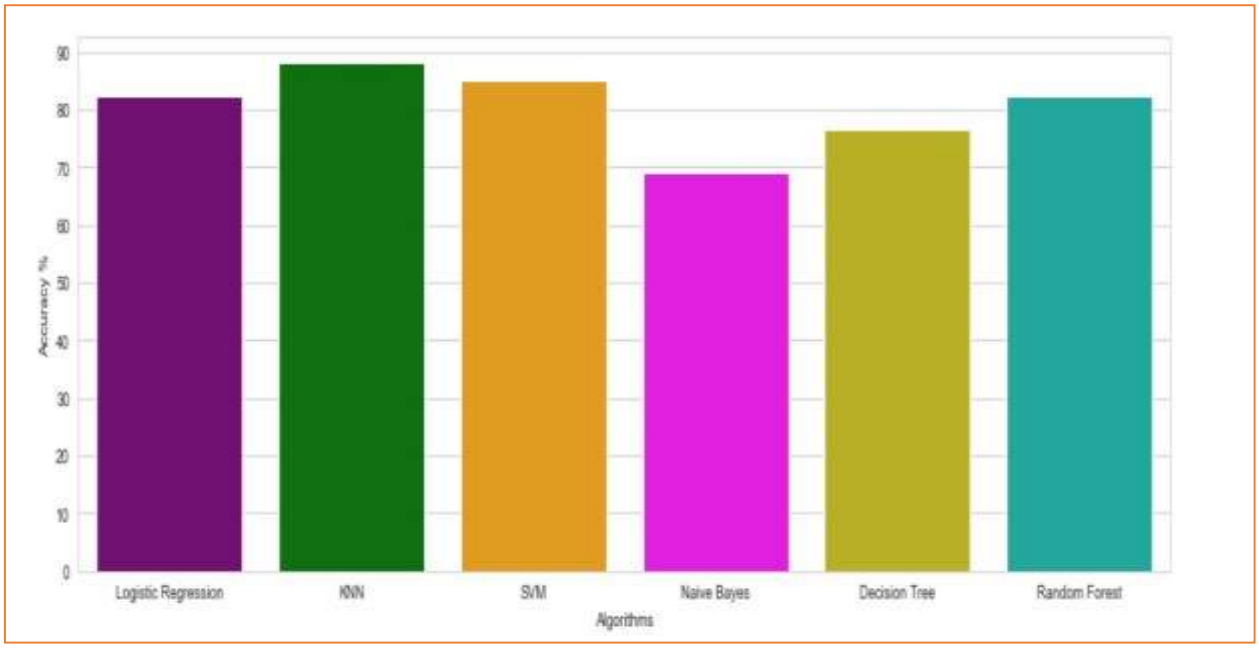

Figure 4 **: Comparative Accuracy of the six classifiers**

6. Conclusion and Future Work

Approximately 17.9 million people every year die due to cardiovascular according to the recent status from WHO. Hence the demand for an efficient and accurate prediction of heart disease is there. This research deals with various techniques involving the Machine learning model and Deep learning models. In all 6 models are trained and tested with maximum scores as Decision tree 76.47%, KNN 88.24%, Random Forest 82.35%, Navi Bayes 69.12, logistic regression 82.35% and SVM 85.29%. KNN gave the best accuracy.

The current study has a number of advantages. The most important is that, in addition to data from the public domain, additional data was acquired from Cardiologists, pathological labs and from people who had suffered with heart attack in and around Odisha. There was also a lot of sensitivity analysis (different models of variable management, feature engineering), as well as model and hyperparameter creation. The fact that the code was made public allowed to analyse and reproduce our findings, as well as use the current work as a springboard into the field of health machine learning in general and programming in particular.The fundamental flaw in this study is that when using data supplied as part of a Machine Learning project, neither clinical reliability nor the quality of the results can be guaranteed. Furthermore, as previously stated, the restricted number of observations certainly reduces the prediction capacity of trained models, which are built to cope with considerably greater numbers of observations.

7. Future Work

In this research KNN models gave the best accuracy, but KNN is only suitable for smaller-scale data. hence as a part of the future work, RNN and some other deep learning models can be applied to check on any improvement in the accuracy of the model.

The next stage in the development of Machine Learning algorithms in healthcare is to apply them to a broader range of situations and populations. Although huge numbers of health data are difficult to get today, we have found that applying Machine Learning to lower data volumes (small data) without compromising validity or application is doable, especially when the data quality is excellent. The benefits to the health-care system might be substantial. Our clinical research teams will be able to become comfortable with these methodologies by working with smaller datasets, allowing them to be prepared for when big data encompassing clinical and healthcare information becomes a reality. However, it is critical to emphasise that each method has its own benefits and may be better to others in certain circumstances.

As a result, there is no way to compare them in broad terms. The data, context, user, imputation, variable selection, parameter tuning, re-balancing, and data partitioning operations will all influence the results. Even if one approach outperforms others on some criteria, different scenarios will most likely alter comparisons.

References

1. Al-Milli, N. (2013). Backpropogation neural network for prediction of heart disease. *Journal of Theoretical and Applied Information Technology*, *56*(1), 131–135.
2. A. K. Pandey, P. Pandey, K. L. Jaiswal, and A. K. Sen, "Data Mining Clustering Techniques in the Prediction of Heart Disease using Attribute Selection Method," International Journal of Science, Engineering and Technology Research (IJSETR), ISSN: 2277798, Vol 2, Issue10, October 2013.
3. Beunza, J. J., Puertas, E., García-Ovejero, E., Villalba, G., Condes, E., Koleva, G., … Landecho, M. F. (2019). Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease). *Journal of Biomedical Informatics*, *97*(July), 103257. https://doi.org/10.1016/j.jbi.2019.103257
4. Chaitrali S. Dangare, Sulabha S. Apte, Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications (0975 888)Volume 47No.10, June 2012.
5. Collins, C. J., & Stevens, C. K. (2002). The relationship between early recruitment-related activities and the application decisions of new labor-market entrants: a brand equity approach to recruitment. *The Journal of Applied Psychology*, *87*(6), 1121–1133. https://doi.org/10.1037/0021-9010.87.6.1121
6. Damen, J. A. A. G., Hooft, L., Schuit, E., Debray, T. P. A., Collins, G. S., Tzoulaki, I., … Moons, K. G. M. (2016). Prediction models for cardiovascular disease risk in the general population: Systematic review. *BMJ (Online)*, *353*. https://doi.org/10.1136/bmj.i2416
7. Decker, H., & Hutchison, D. (2013). *Database and Expert*.
8. Kiyong Noh, HeonGyu Lee, Ho-Sun Shon, Bum Ju Lee, and Keun Ho Ryu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", Springer 2006,Vol:345, page no. 721- 727
9. Masethe, H. D., & Masethe, M. A. (2014). Prediction of heart disease using classification algorithms. *Lecture Notes in Engineering and Computer Science*, *2*, 809–812.
10. McClelland, R. L., Jorgensen, N. W., Budoff, M., Blaha, M. J., Post, W. S., Kronmal, R. A., … Burke, G. L. (2015). 10-Year Coronary Heart Disease Risk Prediction Using Coronary Artery Calcium and Traditional Risk Factors. *Journal of the American College of Cardiology*, *66*(15), 1643–1653. https://doi.org/10.1016/j.jacc.2015.08.035
11. S.Dangare, C., & S. Apte, S. (2012). Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques. *International Journal of Computer Applications*, *47*(10), 44–48. https://doi.org/10.5120/7228-0076
12. Singh, P., Singh, S., & Pandi-Jain, G. S. (2018). Effective heart disease prediction system using data mining techniques. *International Journal of Nanomedicine*, *13*, 121–124. https://doi.org/10.2147/IJN.S124998
13. Vijayarani, S., & Sudha, S. (2013). An Effective Classification Rule Technique for Heart Disease Prediction, *1*(4), 81–85.
14. Vijayarani, S., & Sudha, S. (2015). An Efficient Clustering Algorithm for Predicting Diseases from Hemogram Blood Test Samples , Indian Journal of Science and Technology, *1*(8), 1–8. https://indjst.org/articles/an-efficient-clustering-algorithm-for-predicting-diseases-from-hemogram-blood-test-samples
15. Virtanen, M., Nyberg, S. T., Batty, G. D., Jokela, M., Heikkilä, K., Fransson, E. I., … Kivimäki, M. (2013). Perceived job insecurity as a risk factor for incident coronary heart disease: Systematic review and meta-analysis. *BMJ (Online)*, *347*(7921), 1–15. https://doi.org/10.1136/bmj.f4746
16. Walker, S., Khan, W., Katic, K., Maassen, W., & Zeiler, W. (2020). Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings. *Energy and Buildings*, *209*, 109705. https://doi.org/10.1016/j.enbuild.2019.109705
17. Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can Machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS ONE*, *12*(4), 1–14. https://doi.org/10.1371/journal.pone.0174944
18. Y. Alp Aslandoganet. al.," Evidence Combination in Medical Data Mining", Proceedings of the international conference on Information Technology: Coding and Computing (ITCC'04) 0-7695-2108-8/04©2004 IEEE.

Author Details

1. Dr Neena Nanda

Assistant Director

Institute for Technology and Management

Plot No. 25/26, Institutional Area

Sector 4, Kharghar East

Navi Mumbai - 410210
2. Ayaskanta Mohapatra
Student: PDGM Business Analytics 2019-21
Institute for Technology and Management
Plot No. 25/26, Institutional Area
Sector 4, Kharghar East
Navi Mumbai - 410210