2020 IEEE International Conference for Innovation in Technology (INOCON)
Bengaluru, India. Nov 6-8, 2020

Performance Evaluation of Supervised Machine Learning Algorithms in Prediction of Heart Disease

P. Sujatha
Professor, Department of Information Technology,
Vels Institute of Science, Technology & Advanced Studies,
Pallavaram, Chennai, India.
suja.research@gmail.com

K. Mahalakshmi
Research Scholar, School of Computing Sciences,
Vels Institute of Science, Technology & Advanced Studies,
Pallavaram, Chennai, India.
rkmahalakshmi2016@gmail.com

***Abstract* - Big challenge in health care industry is to record and analyze the massive amount of information about patients. Innovations in technologies made revolution in the healthcare industries. In recent years the data analytics developed as promising tool for problem solving and decision making in healthcare professions. Data analytics process the data automatically to make healthcare system more dynamic and robust. It systematically uses and analyses the data of health care for better treatment with low costs. The chief applications of Machine learning in healthcare are the detection and diagnosis of diseases. The heart is the chief organ of human body. Heart disease increases the mortality rate in the world. Around 90% of heart diseases are preventable. Machine learning plays a remarkable role in the health care industry in prediction of heart disease. In this research paper, the presence of heart disease is predicted by employing Decision Tree, Naïve Bayes, Random Forest, Support Vector Machine, K-Nearest Neighbor and logistic Regression algorithms. The performance of the algorithms was analyzed using parameters such as Accuracy, Precision, AUC and F1-score. From the experimental result, it is found that the Random Forest is more accurate for predicting the heart disease with accuracy of 83.52% compared with other supervised machine learning algorithms. The F1- Score, AUC and precision score of Random forest classifiers are 84.21%, 88.24% and 88.89% respectively.**

Keywords- Heart disease, data analytics, machine learning, decision tree, naïve bayes, random forest, support vector machine, k-nearest neighbor, logistic regression, health care.

I. INTRODUCTION

Major causes of increasing mortality rate are heart disease. The unhealthy life style, stress, obese and health history of patients are the risk factors of heart disease. Heart disease leads to complications like heart attack, heart failure and strokes etc. Most of the heart diseases are preventable with proper diagnosing system and simple lifestyle modification. Usage of Machine Learning techniques has been increased to develop screening tools with pattern recognition and classification. Such tools provide more accuracy as compare to other traditional approaches. Machine learning is used to extract the hidden facts from medical data. Machine learning is a multi-disciplinary filed. It consists of statistics, algebra, data processing and knowledge analytics etc., Machine learning aims to make machine capable of learning. Machine learning is classified into three categories: Supervised, Unsupervised Machine Learning and Reinforcement Learning. Fig.1. shows the classification of machine learning techniques.

A. Supervised Machine Learning

A supervised ML algorithm analyzes the known set of input data with known output value and produces a function. The functions used to find the output of new input. In supervised machine learning, there is a presence of supervisor as teacher during the training process to the machine. The learning algorithm in supervised learning makes prediction on the training data and is corrected by the supervisor, and this learning process continues until the learning algorithm achieves a desired accuracy. Supervised learning algorithms are classified into two categories,

- Classification
- Regression

Some of the popular supervised machine learning Algorithms is,

- Linear Regression
- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine
- Naïve Bayes
- K Nearest Neighbor

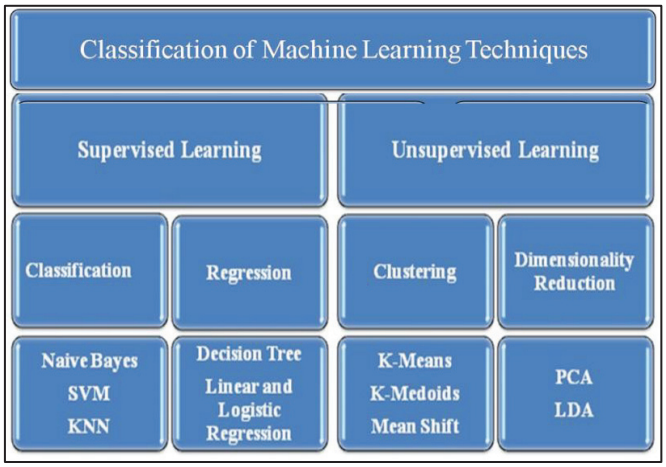

Fig. 1. Classification of Machine Learning Techniques

Machine learning is an effective tool in predicting from large amount of healthcare data. In this research paper, we have implemented only the supervised machine learning algorithms such as Decision Tree, Naïve Bayes, Random Forest, Support Vector Machine, K-Nearest Neighbor and

978-1-7281-9744-9/20/$31.00 ©2020 IEEE

logistic Regression using python for diagnosing the heart related problems. Further, we analyzed the performance of each algorithm using the evaluation metrics such as Accuracy, Precision, AUC and F1-score.

II. REVIEW OF LITERATURE

Various methods have been utilized for early prediction of heart disease using Machine learning techniques. Existing research works related to heart disease prediction using Machine learning techniques are described in this section:

Fahd Saleh Alotaibi [1] utilized most common Machine Learning Approaches viz., Decision Tree, Naïve Bayes, Random Forest, Support Vector Machine and Logistic Regression. 14 attributes of Heart Disease Dataset from Kaggle platform was used. 10-fold cross-validation techniques were applied to increase the accuracy and to decrease the chances of duplication in record selection using Rapid Miner tool. From the result it is inferred that decision tree classifier obtained higher performance with 93.19% of accuracy and Naïve Bayes obtained minimum among all. The author concluded that Rapid miner tool give high accuracy than Matlab and Weka tool.

S.Nandhini et al. [2] proposed a sensor node and machine learning algorithm to predict the heart disease. The implemented model performs monitoring and predicting process of the heart disease. AMPED sensor with Bluetooth 4.0 was used in monitoring System. Monitoring process will send email and an SMS to the nearest hospital and relative of the patient during emergency situation. The diagnosing system experimented with KNN, Decision Tree, Random Forest, Naïve Bayes, SVM and Logistic Regression on 13 features of Cleveland heart dataset. Diagnosing system with Random Forest obtained maximum accuracy of 89% and monitoring system achieved 100% of accuracy

Sanjay Kumar Sen [3] applied common classifiers viz., Naïve Bayes, SVM, Decision Tree, and K-Nearest Neighbor in predicting and diagnosing of heart disease. The experiments were done using 14 input features of UCI Machine Learning Repository dataset via weka tool. The author concluded that naïve bayes algorithms performed better than other algorithms

III. METHODOLOGY

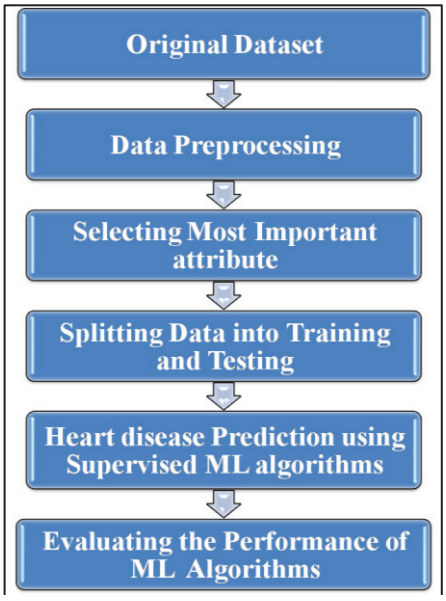

Fig. 2. Schematic Diagram

In this research, we have implemented the aforementioned supervised machine learning algorithms for prediction of heart disease using Heart disease data set from kaggle database with python 3.7. We also evaluated the performance of such algorithms based on the metrics such as accuracy, precision, F1 Score and AUC. The overall architecture of this research work is as given in the Fig.2.

A. Data Set Description

The data are retrieved from the Heart disease data set from kaggle database [4]. The dataset provides the patients' information that consists of 303 records with 14 attributes. Each attribute is a potential risk factor and the descriptions of them are shown in Table I.

TABLE I. LIST OF ATTRIBUTES OF SAMPLE DATA SET

S.No	Attribute	Description
1.	Age in years	Continuous value
2.	Gender	1 represents male 0 represents female
3.	CP – Chest Pain	Chest pain type 0 represents typical angina 1 represents atypical angina 2 represents non-anginal pain 3 represents asymptomatic
4.	trestbps	resting blood pressure
5.	chol	serum cholesterol in mg/dl
6.	Fasting Blood Sugar	fasting blood sugar > 120 mg/dl 1 represents true 0 represents false
7.	Restecg- -Resting electrocardiographic (ECG)	resting electrocardiographic results 0 represents normal 1 represents having ST-T wave abnormality 2 represents Possible or definite left ventricular hypertrophy
8.	thalach	maximum heart rate achieved
9.	Exang - exercise induced angina	Blood supply when you exercise 1 represents yes 0 represents no
10.	oldpeak	ST depression induced by exercise relative to rest
11.	Sl: slope	the slope of the peak exercise ST segment 0 represents upsloping 1 represents flat 2 represents downsloping
12.	ca	number of vessels (0-3) colored by flourosopy
13.	thal	Thallium stress test 1 represents normal; 2 represents fixed defect; 3 represents reversible defect

14.	target	the predicted attribute 0 represents no chances of heart failure 1 represents chances of heart failure

B. *Data Preprocessing*

Data is often incomplete and inconsistent. Inconsistent and incomplete dataset significantly influence the performance of machine learning algorithms. The heart disease data set has been checked for missing values and null values as null value considerably effect on the conclusion that drawn from the data . There were no missing values in our dataset. As the percentage of missing value in our dataset is zero, we further moved to the feature selection process. We also scaled the features to speed up the training of the classifiers. Feature scaling is a process during the data preprocessing step. It is a method to normalize the independent variable or features of data within a particular range.

C. *Feature Selection*

The major purpose of selecting feature is to eliminate unrelated and redundant attributes. An irrelevant feature reduces the accuracy of the algorithm. Our research work used Boruta Feature selection technique to select the most important features. The Boruta algorithm is used to find all the significant attributes from the dataset with respect to an outcome variable [5]. From 13 input attributes, 6 attributes ('age', 'cp', 'thalach', 'oldpeak', 'ca', 'thal') are selected as most important attribute. We have split the data into 70% as training and 30% as testing. The important input attribute selected using Boruta Feature selection are shown with description in Table II.

TABLE II. LIST OF SELECTED ATTRIBUTES USING BORUTO FEATURE SELECTION

S.No	Attribute	Description
1.	Age in years	Continuous value
2.	CP – Chest Pain	Chest pain type 0 represents typical angina 1 represents atypical angina 2 represents non-anginal pain 3 represents asymptomatic
3.	thalach	maximum heart rate achieved
4.	oldpeak	ST depression induced by exercise relative to rest
5.	ca	number of major vessels (0-3) colored by flourosopy
6.	thal	Thallium stress test 1 represents normal; 2 represents fixed defect; 3 represents reversible defect

D. *Implementation of Supervised Machine Learning Algorithms*

We have implemented popular supervised machine learning algorithms on heart disease dataset from kaggle website using the tool Python 3.7. The data set contains 303 records. It includes both positive and negative observations. We have divided these 303 records of the data into training and testing. 30% of records that is 91 records have been used for testing and remaining 212 records for training. Evaluating model is one of the core mechanisms while building machine learning model to check the efficacy of the model. We have also evaluated and compared the performance of the supervised machine learning algorithms using the metrics such as accuracy, precision, F1-score and AUC. One of the popular methods of estimating the performance of the algorithms is to use confusion matrix that contains information pertaining to predicted (columns) and actual class (rows) which can be visualized easily. It is a matrix representation of the prediction results. It is used to describe the performance of the classifier.Confusion matrix gives the counts of test records correctly and incorrectly predicted by the model. The representation of confusion matrix is shown in Fig. 3.

Each prediction can be one of the following four outcomes.

True Negatives (TN): Correctly predicted as negative and are actually negative

True Positives (TP): Correctly predicted as positive and are actually positive

False Negatives (FN): Incorrectly predicted as negative and are actually positive.

False Positives (FP): Incorrectly predicted as positive and are actually negative

	Predicted 0	Predicted 1
Actual 0	**TN**	**FP**
Actual 1	**FN**	**TP**

Fig. 3. Confusion Matrix

The confusion matrix gives us a holistic view of how well our classification model is performing.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Confusion matrices and evaluation results obtained for the methods SVM, NB, KNN, LR, DT and RF are presented in this section. In this research the above mentioned supervised machine learning classifiers made a total of 91 predictions that is 91 records were being tested for the presence of the heart disease. Out of those 91 records in actual, 50 records in the sample have the heart disease that is "Yes" and 41 patients do not have heart disease that is "No". We have compared this actual target values with those predicted by the machine learning model using confusion matrix.

The following Table III shows the confusion matrices obtained by the supervised machine learning algorithms.

TABLE III. CONFUSION MATRICES OF SUPERVISED MACHINE LEARNING METHODS

S.No	Classifiers	Confusion Matrix	Description:
1.	Random Forest	n=91 / Predicted 0 (No) / Predicted 1 (Yes) Actual 0 (No): TN= 36 / FP = 5 / 41 Actual 1 (Yes): FN = 10 / TP = 40 / 50 46 / 45 Confusion Matrix of Random Forest	Out of those 91 records, the random forest classifier predicted "Yes" 45 times and "No" 46 times **True Negatives (TN):** 36 people were correctly predicted "No" and also in actual they do not have the heart disease. **True Positives (TP):** 40 people were correctly predicted "Yes" and also in actual they have the heart disease. **False Negatives (FN):** 10 people were incorrectly predicted "No" while they actually have heart problem. **False Positives (FP):** 5 people were incorrectly predicted "Yes" while in actual they do not have heart problem.
2.	Support Vector Machine	n=91 / Predicted 0 (No) / Predicted 1 (Yes) Actual 0 (No): TN= 34 / FP= 7 / 41 Actual 1 (Yes): FN= 9 / TP= 41 / 50 43 / 48 Confusion Matrix of SVM	Out of those 91 records, the SVM predicted "Yes" 48 times and "No" 43 times **True Negatives (TN):** 34 people were correctly predicted "No" and also in actual they do not have the heart disease. **True Positives (TP):** 41 people were correctly predicted "Yes" and also in actual they have the heart disease. **False Negatives (FN):** 9 people were incorrectly predicted "No" while they actually have heart problem. **False Positives (FP):** 7 people were incorrectly predicted "Yes" while in actual they do not have heart problem.
3.	Naïve Bayes	n=91 / Predicted 0 (No) / Predicted 1 (Yes) Actual 0 (No): TN= 35 / FP= 6 / 41 Actual 1 (Yes): FN= 10 / TP= 40 / 50 45 / 46 Confusion Matrix of Naïve bayes	Out of those 91 records, the Naïve bayes predicted " Yes" 46 times and "No" 45 times **True Negatives (TN):** 35 people were correctly predicted "No" and also in actual they do not have the heart disease. **True Positives (TP):** 40 people were correctly predicted "Yes" and also in actual they have the heart disease. **False Negatives (FN):** 10 people were incorrectly predicted "No" while they actually have heart problem. **False Positives (FP):** 6 people were incorrectly predicted "Yes" while in actual they do not have heart problem.
4.	Logistic Regression	n=91 / Predicted 0 (No) / Predicted 1 (Yes) Actual 0 (No): TN= 32 / FP= 9 / 41 Actual 1 (Yes): FN= 9 / TP=41 / 50 41 / 50 Confusion Matrix of Logistic Regression	Out of those 91 records, the Logistic Regression predicted "Yes" 50 times and "No" 41 times. **True Negatives (TN):** 32 people were correctly predicted "No" and also in actual they do not have the heart disease. **True Positives (TP):** 41 people were correctly predicted "Yes" and also in actual they have the heart disease. **False Negatives (FN):** 9 people were incorrectly predicted "No" while they actually have heart problem. **False Positives (FP):** 9 people were incorrectly predicted "Yes" while in actual they do not have heart problem.

5.	Decision Tree	n=91 / Predicted 0 (No) / Predicted 1 (Yes) Actual 0 (No): **TN=33** / **FP=8** / **41** Actual 1 (Yes): **FN=11** / **TP=39** / **50** **44** / **47** Confusion Matrix of Decision Tree	Out of those 91 records, the Decision Tree predicted "Yes" 47 times and "NO" 44 times **True Negatives (TN):** 33 people were correctly predicted "No" and also in actual they do not have the heart disease. **True Positives (TP):** 39 people were correctly predicted "Yes" and also in actual they have the heart disease. **False Negatives (FN):** 11 people were incorrectly predicted "No" while they actually have heart problem. **False Positives (FP):** 8 people were incorrectly predicted "Yes" while in actual they do not have heart problem.
6.	K Nearest Neighbors	n=91 / Predicted 0 (No) / Predicted 1 (Yes) Actual 0 (No): **TN=29** / **FP=12** / **41** Actual 1 (Yes): **FN=13** / **TP=37** / **50** **42** / **49** Confusion Matrix of KNN	Out of those 91 records, the KNN predicted "Yes" 49 times and "No" 42 times **True Negatives (TN):** 29 people were correctly predicted "No" and also in actual they do not have the heart disease. **True Positives (TP):** 37 people were correctly predicted "Yes" and also in actual they have the heart disease. **False Negatives (FN):** 13 people were incorrectly predicted "No" while they actually have heart problem. **False Positives (FP):** 12 people were incorrectly predicted "Yes" while in actual they do not have heart problem.

A. Performance Evaluation

We have analyzed the performance of various machine learning algorithms using the metrics such as accuracy, F1 Score, AUC and Precision for heart disease dataset [6].

The following metrics were used to evaluate the performance of the algorithms:

1) Area Under The Roc Curve (AUC):

Area under the ROC Curve (AUC) is a performance metrics for classification problem. AUC created by plotting True Positive Rate (TRP) i.e. sensitivity or recall vs. False Positive Rate.

Higher the AUC, better the model is at predicting the heart disease. From the Fig. 10, we found that the random forest got highest AUC of 88.24% and KNN got the least AUC of 76.85%.

2) Classification Accuracy:

Classification Accuracy is the common evaluation metric for classification algorithms. The accuracy is the ratio of the total number of prediction that is correct to the total number of samples.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

Accuracy is the most important metric for evaluating the performance of the supervised machine learning algorithms. The Fig. 11 shows accuracy achieved by Supervised Machine learning algorithms. It is clear that the Random Forest achieved highest accuracy of 83.52%. The other algorithm which is closed to the random forest accuracy is Naïve Bayes and SVM. Naïve Bayes and SVM obtained the same accuracy of 82.42%. KNN obtained least accuracy of 72.53%. The accuracy of logistic regression is 80.22% and the accuracy of decision tree is 79.12%.

3) F1 Score:

The F1- Score is the combination of precision and recall. It is calculated by taking harmonic mean of precision and recall.

$$F1 = 2* (\text{Precision} * \text{Recall})/ (\text{Precision} + \text{Recall})$$

It is clear from the Fig. 12 that the F1 score of random forest classifier is higher in comparison of other models. The f1 score of Random forest classifier is 84.21%. In our study we observed that KNN obtained least F1 score of 74.75%

4) Precision:

Precision is the ratio of true positives to all the positives predicted by the algorithms.

$$\text{Precision} = \frac{TP}{TP+FP}$$

The precision shows the percentage of the algorithms results which are relevant. From the experimental result and Fig.13. we found that the Random forest achieved 88.89% of precision rate which is higher than the other supervised machine learning algorithms. KNN has least precision rate 75.51%.

The following Table IV shows the values of accuracy, F1 Score, AUC and precision score obtained by supervised machine learning algorithms.

TABLE IV. PERFORMANCE EVALUATION OF SUPERVISED LEARNING ALGORITHMS

Classifiers	AUC	Accuracy	F1 score	Precision
Random Forest	0.882439	0.835165	0.842105	0.888889
Support Vector Machine	0.862439	0.824176	0.836735	0.854167
Naïve Bayes	0.860000	0.824176	0.833333	0.869565
Logistic Regression	0.855610	0.802198	0.820000	0.820000
Decision Tree	0.789024	0.791209	0.804124	0.829787
K Nearest Neighbors	0.768537	0.725275	0.747475	0.755102

From the above Table IV, the values of AUC, Accuracy, F1 score and Precision metrics of different supervised Machine Learning algorithms are graphically represented in the following Fig. 4, 5, 6, 7 respectively.

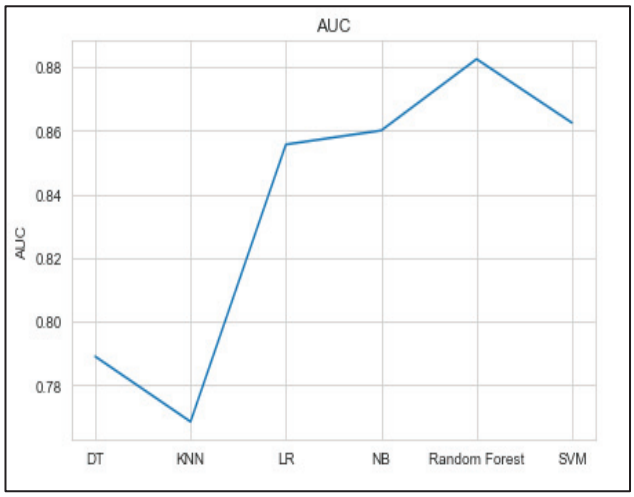

Fig. 4. AUC of Supervised Machine Learning Algorithms

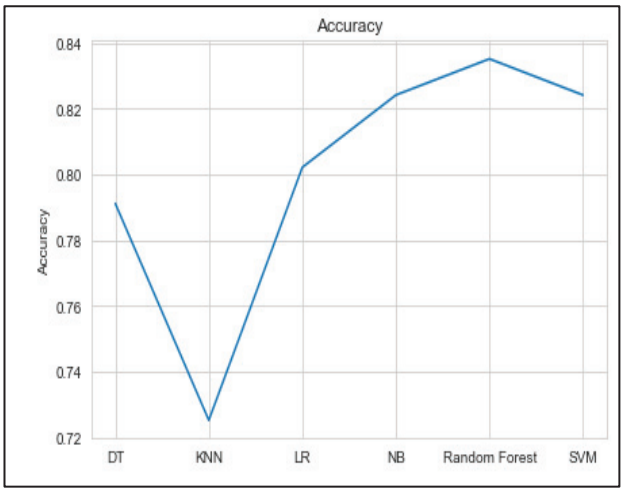

Fig. 5. Accuracy Graph of Supervised Machine Learning Algorithms

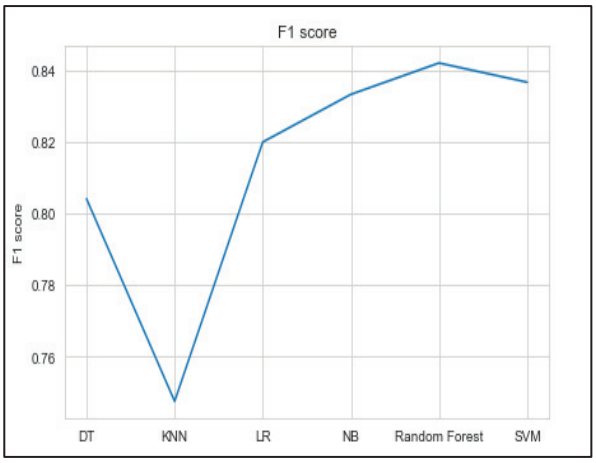

Fig. 6. F1 Score of Supervised Machine Learning Algorithms

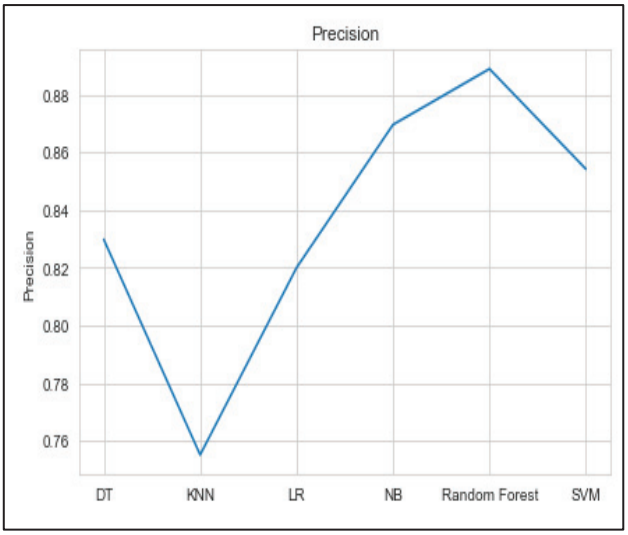

Fig. 7. Precision Score of Supervised Machine Learning Algorithms

The following Fig. 8 shows the Overall Performance Evaluation of supervised machine learning algorithms in terms of AUC, Accuracy, F1 Score, Precision.

From the experimental results we found that the Random Forest classifier achieved better results in all the evaluation metrics such as AUC, Accuracy, F1 Score and Precision in comparison of all other supervised Machine learning algorithm. The KNN showed least performance. The performance of SVM and NB were similar in most of the model evaluation parameter.

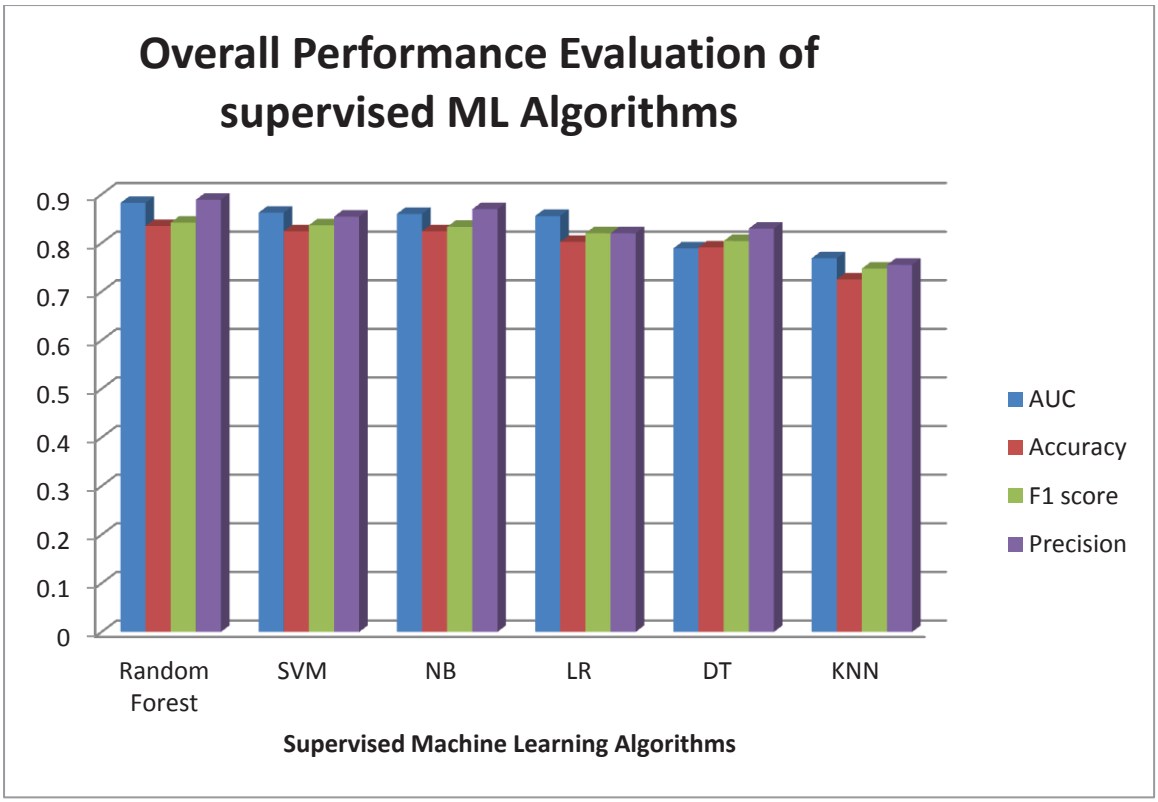

Fig. 8. Overall performance Evaluation of supervised machine learning algorithms.

V. CONCLUSION AND FUTURE WORK

The objective of this research work is to compare the performance of different supervised machine learning algorithms that can be employed in automated heart disease prediction systems. In this research work, heart disease is predicted using decision tree, naïve bayes, SVM, KNN, Logistic Regression and random forest. The classifiers are implemented using Python 3.7 on Heart disease dataset form Kaggle website. From the experimental results it is found that the random forest classifier is more accurate compared to Decision Tree, SVM, Naïve Bayes, Logistic regression and KNN.

The work is done with small size dataset. With more data better models can be built. The future work of this research is to use different Machine learning techniques for heart disease prediction with large size dataset and also to combine some of the machine learning techniques to build reliable and accurate prediction model for heart disease in real time.

REFERENCES

[1] Fahd Saleh Alotaibi, "Implementation of Machine Learning Model to predict Heart Failure Disease," International Journal of Advanced Computer Science and Applications (IJACSA), vol.10, no.6, pp.261-268, 2019.

[2] S.Nandhini, Monojit Debnath, Anurag Sharma and Pushkar, "Heart Disease Prediction using Machine Learning," International Journal of Recent Engineering Research and Development (IJRERD), ISSN: 2455-8761, vol.3, pp.39-46, 2018.

[3] Sanjay Kumar Sen, "Predicting and Diagnosing of Heart Disease using Machine Learning Algorithms," International Journal of Engineering and Computer Science (IJECS), vol.6, pp.21623-21631, 2017.

[4] https://www.kaggle.com/ronitf/heart-disease-uci?select=heart.csv

[5] https://www.datacamp.com/community/tutorials/feature-selection-R-boruta

[6] https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_algorithms_performance_metrics.htm

[7] Amit Juyal, Chetan Pandey, Janmejay Pant, Ankur Dumka and Vikas Tomar, "Performance Analysis of Supervised Machine Learning Algorithms on Medical Dataset," International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, vol.8 Issue-6, March 2020

[8] M. Chandralekha and N. Shenbagavadivu, "Performance Analysis Of Various Machine Learning Techniques To Predict Cardiovascular Disease: An Emprical Study," Applied Mathematics & Information Sciences An International Journal, No. 1, pp.217-226, 2018

[9] Beulah Christalin Latha. C and Carolin Jeeva. S, " Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," Informatics in Medicine Unlocked, Elsevier, pp. 1-9, 2019

[10] Rajesh. N, Maneesha.T, Shaik Hafeez and Hari Krishna, "Prediction of Heart Disease Using Machine Learning Algorithms," International Journal of Engineering & Technology, pp.363-366, 2018