Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Deep Learning: Its Empirical Study and Comparison in Heart Disease Prediction

[1]Abhishek Saxena, [2]Dr.Harish Kumar Taluja, [3]Dr.Neeta Verma

[1]Doctor of Philosophy (Pursuing), School of Engineering and Technology, Noida International University, Greater Noida, Uttar Pradesh.

[2]Professor, School of Engineering & Technology, Noida International University, Greater Noida, Uttar Pradesh.

[3]Professor, Department of Computer Science and Engineering, Inderprastha Engineering College, Ghaziabad, Uttar Pradesh.

Article Info
Page Number: 10963 - 10981
Publication Issue:
Vol 71 No. 4 (2022)

Article History
Article Received: *15 September 2022*
Revised: *25 October 2022*
Accepted: *14 November 2022*
Publication: *21 December 2022*

Abstract

Due to the impact of fast technological advancement and adoption of advanc lifestyle, the life of humans has become very comfortable and had completely chang which ultimately is the reason for the rise in many cardiovascular diseases. During past decades, the diseases caused by heart emerged to be the biggest reason of sudc and untimely death for many persons across the world. Since, its early diagnosis is s a challenge for a medical expert, therefore many researchers are applying various ot techniques for its prediction. Since the application of Deep Learning has been wid applied for various other complex tasks, and had proved with its ability and potentia extract more appropriate and high level features, it gives a ray of hope to ma researchers for helping the health care professionals. This paper conducts the empiri study of Deep Learning, and also compares it with various other classificat algorithms of machine learning, to embark the early detection of the occurrences heart disease. On the basis of the results obtained from this study, it is concluded t the Deep Learning emerges to be a better choice, in comparison to others, and have capability to be used for the prediction of heart disease in a patient.

Keywords: Deep Learning, Machine Learning, Classification Algorithms, Accuracy

1. INTRODUCTION

The Humans are the best creature on earth, as they have the capability to create changes of any kind, but during the entire life-span, they are perhaps, mostly infected by various diseases and health problems. These diseases, of any type and in any form, if not identified early and treated properly, with utmost guidance, can be hazardous for the life of a person.

The World Health Organization [13] in its previous reports has predicted that diseases occurred by cardio-vascular, are alone the major cause of deaths across the countries. Since heart is the most vital organ of human anatomy, on which the entire life span of individual depends, and is deeply associated in providing smooth functionality to other organs of body, its fitness is highly recommended. In the last year, a report had estimated that many million peoples have lost their precious life from CVDs worldwide, in which three quarters solely belongs to the countries of middle and low incomes. It is therefore, necessary to detect CVDs early, as its major cause and early prediction is still a challenge.

The heart disease comprises of many symptoms which includes acute pain in chest, arms, left shoulder, elbows, jaws or in back. It also includes breathlessness, discomfort in the center of chest, nausea, faintness, vomiting, mild headache, a sweating with cold body and turning pale yellow. As in the healthcare system, the diagnosis for any disease is the most crucial task, hence for CVDs many cardiologists depend on traditional clinical tests, such as blood tests and electro-cardiogram, while in many cases it has been observed that besides these, some other complex combination of both clinical and pathological examinations comprising of blood pressure, glucose, elevated levels of sugar etc. is also preferred.

Although these previous techniques, along with physical examination and detailed analysis of previous medical record, gives an idea for its evaluation by medical expert, which if unclear and unreliable, often causes inaccurate diagnosis and delays in treatment. Besides time consuming and lethargic in progress, they are also expensive and computational intensive.

As compared to previously, when no specific approach was there to find the feasible treatment of heart disease, and is entirely depended on clinical investigations and on the intuition and experience of an expert, today its early detection is not a big task, as many clinical professionals and researchers have shown interest to study it for its accurate prediction.

With the vast and rapid technological advancement by artificial intelligence, its subset, the Machine Learning [22], have paved the way, for getting absolute analysis about its occurrence, and had shown complete insight about its early detection. While the scope of Machine Learning is found in various applications of medical field, it can also be mainly used to make the early forecast of heart disease.

As the machine learning algorithms [14] have paved the doors for elaborating new opportunities to get the timely and accurate diagnosis of heart disease. The deep learning is an emerging area, and a more advanced technique based on biological nervous system having multiple layers, processing units for extraction of features. The Deep Learning enlarges the perspective and scope of finding the more accurate analysis of patterns to the medical expert for speedy diagnosis. Therefore, [15] it has been observed that the developed models of deep learning are more reliable, and hence, have the ability for reduction of error up to a great extent.

This paper conducts the empirical study of Deep Learning technique for the chances of occurrence of heart disease in a patient. It is sub-divided into following parts:

Part 2 focusses on literature review being done by various researchers being progressed in this area, *Part 3* implements the architecture of proposed research, *Part 4* defines methodology used, *Part 5* depicts the experimental results obtained between the various classification algorithms, *Part 6* describes the conclusion of paper, and *Part 7* defines the result and future scope of it.

2. LITERATURE REVIEW

The broad category of affecting the heart is the cardiovascular diseases, and the early methods used to forecast it had helped in making decisions about the changes. Since heart plays a very major role in the entire life of living organisms therefore, the prediction and diagnosis of heart diseases, requires more knowledge and accuracy, since a negligible error can only be the cause of fatigue, or death of a person [22]. Also, according to the various world reports, it is estimated that there are extreme amount of deaths related with heart, and its percentage is growing exponentially by every year.

To deal with such problems many researchers have found that there is an urgent need for a prediction system [26] to spread its preliminary awareness and detection. Machine learning, a subset of AI, provides a complete support and platform for the prediction of any kind of disease, which takes data from natural ways. Since the machine learning [23], [24] algorithms, have also been used previously for the prediction of heart disease, hence it plays an essential role and emerges as a better option for delivering accurate results.

But, since the splitting of dataset is one of the easy tasks in machine learning, and is also a required feature in operating dataset, it is highly needed to extract the information present in the corporate databases of large dataset. Therefore, [1],[31] analyzed that the use of artificial neural network, k-nearest neighbors, naive bayes, and logistic regression, are some algorithms of machine learning, that can efficiently diagnose the occurrence of heart disease.

As it is quite uncommon to identify beforehand about an individual being affected by heart disease, since it requires patient's data and previous records, hence identification of parameters is a difficult task, and therefore, requires additional tools to make the decisions.

To make the chances for improving the efficiency of the aforesaid approach, another researcher [17] therefore, utilized the Kaggle dataset with 303 instances for training the model, and 14 attributes for labelling class, as the primary cause for identifying heart disease. In this work he concluded that, although data is applied on logistic regression of classification algorithm, but the proposed approach gives better accuracy, in comparison to the other algorithms used in machine learning.

Similarly, another researcher [3] also used some other algorithms of machine learning, for finding the true classification of patients with, or without heart disease. In their work, they used the dataset which comprises the data of chest pain, level of sugar, blood pressure and other, including the records of previous medical history of patient. On the basis of these, they concluded that the result achieved by using these have good accuracy, in comparison to the previously used techniques of classification.

Although the designed model can be able to achieve higher rate of accuracy by the use of large quantity of training data, it also gives an idea to decide whether there exists a chances for the existence of heart disease or not. Due to this, many researchers [4], [25], [5], [6], [7], [8], [9] have time-to-time tested the algorithms on UCI dataset [16], and found good results. Besides this, there are still many algorithms that are not capable to forecast the true and accurate prediction of disease.

Therefore, the technique of Deep Learning emerged to be used as an approach to predict the disease efficiently, which is caused due to the blockage in heart.

Since the recent and widespread popularity gained by the Deep Learning approach, due to its capability and acceptance to predict the true accuracy, another researcher Harshit Jindal, Agrawal etc. [11] & [19] conducted an investigation for the amount of work being done by using it, in the field of heart disease. In their work, they concluded that the use of Deep Learning can widely be used for it as it gives more accurate results in comparison to the previously applied techniques.

Another researcher Sajja & Kalluri [10] in his work proposed the application of deep learning by using the convolutional neural network to make an early stage prediction for the spread of cardiovascular disease. In this work, for conducting the experiment, they used dataset of UCI machine repository, and made a model by using two convolutional layers, two dropout layers and a single output layer. On the basis of this setup, they made a comparison between it and machine learning, and concluded that the deep learning achieves a very high rate of accuracy, in comparison to the other algorithms of machine learning.

Similarly, another researcher Solanki & Sharma [12] introduced the concept of defining enhanced deep neural network made up from thirteen variables of predictor and single target variable, obtained by applying the dataset of machine repository on it. The model is trained by using the keras package of Deep Neural Network, to validate it. On the basis of this work, they concluded that using this approach has increased the speed of prediction, and increased the chances of getting the enhancement in accuracy.

3. PROPOSED ARCHITECTURE

The main emphasis of this work is to utilize the selected dataset, to propose a model, for the design of an algorithm, to elaborate the concept of deep learning, so that the true and accurate detection of disease is made. For making this work possible, it undergoes by using Python, due to its support for vast libraries and packages, including *Tensorflow, Keras, Numpy, Pandas, Sklearn.* It also uses Anaconda, an open source distribution [18] containing an integrated development environment Spyder, an open cross platform used for scientific work.

This work also implicit the empirical analysis of deep learning, using the proposed model, and compares it with the different classification algorithm of machine learning at different levels of evaluations.

For this purpose, the following algorithms have been proposed:

Table 1: Proposed Algorithm for Deep Learning

ALGORITHM

Step 1: *Choose the dataset of Heart disease.*
Step 2: *Perform the preprocessing of data.*
Step 3*: Prepare the data, by removing noise and missing values.*
Step 4*: Make proper feature selection in it.*
Step 5*: Split the data into two parts: 80% for training & 20% for testing.*
Step 6*: Build the model of deep neural network by using four hidden layer.*
Step 7: *Apply ReLu activation function in first input layer,* ***Step 8:*** *Apply ReLu activation function in four hidden layers and use sigmoid function in output layer.*
Step 9: *Train the model using 80% of data.*
Step10*: Apply cross validation in the training data.*
Step 11*: Perform testing on 20% of data.*
Step 12*: Obtain the predicted result*
Step 13*: Repeat the steps 9-12, if not obtained*
Step 14*: End*

able 2: Algorithm for Machine Learning

ALGORITHM

Step 1: *Choose the dataset of heart disease.*
Step 2: *Perform the preprocessing of data.*
Step 3*: Prepare the data by removing noise and missing values.*
Step 4*: Make proper feature selection in it.*
Step 5*: Split the dataset into two parts: 80% for training & 20% for testing.*
Step 6*: Apply the selected algorithm.*
Step 7*: Train the model.*
Step 8*: Apply cross validation in the training data.*
Step 9*: Perform the testing by using 20% of data.*
Step 10*: Obtain the predicted result.*
Step 11*: Repeat step 6, for various algorithms.*
Step 12*: Repeat the steps 7-10, for better results*
Step 13*: End*

4. METHODOLOGY

4.1. Dataset Collection:

In this proposed work, the dataset is selected through [17] which had originally been obtained from the UCI repository of machine learning [16] specifying Cleveland database. The data consists of

303 values having 76 attributes as total, but it is referred by all the previous experiments of published work to use only 14. In general, the database of Cleveland is used in various researches of machine learning up till now, as it is the only available dataset for it. Therefore, it is also being used here with the hope to detect the correct and accurate prediction of it, irrelevant of its type.

The detailed descriptions of the various attributes used in this database are as follows:

Table 3: Description & Values of dataset

No	Features	Type	Description	Values
1	Age	Real	Age	In years
2	Sex	Categorical	Sex	Male = 1, Female = 0
3	CP	Categorical	Chest Pain type	Value 1: typical angina Value 2: atypical angina Value 3: non-anginal pain Value 4: asymptomatic
4	Trestbps	Real	Resting Blood Pressure	Measured in mm Hg.
5	Chol	Real	Serum Cholesterol	Measured in mm/dL
6	FBS	Real	Fasting blood sugar > 120 mg/dl)	(1 = true; 0 = false)
7	RestECG	Categorical	Resting electrocardiographic results	Value 0: normal Value 1: having ST-T wave abnormality Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
8	Thalach		Maximum Heart Rate achieved	60-200 bph
9	Exang	Categorical	Exercise Induced Angina	(yes = 1; no = 0)
10	Oldpeak	Real	ST depression induced by exercise relative to rest	
11	Slope	Categorical	slope of the peak exercise ST	Value 1: upsloping Value 2: flat Value 3: down-sloping
12	Ca	Real	number of major vessels colored by flouro-scopy	Values between (0-3)

13	Thal	Categorical	Thalassemia(defect type)	normal =3, fixed defect = 6, reversible defect =7
14	Num	Real	Diagnosis of Heart Disease	Value 1: Yes Value 0: No

***4.2 Data Preprocessing*:**

As the data of real-life contains many redundant values, missing information, and noise therefore, it is desired to make it clean, by removing the noise and filling it with missing values. This step is most important, and utmost care must be taken to make the data perfect, as later the data needs to be transformed efficiently in making the model. The present work undergoes the transformation of data, by applying it to the dataset of UCI [16] machine repository, to reduce it into a straightforward form, for predicting the more accurate results of the model.

***4.3 Data Preparation*:**

In this step, a brief description of analysis of database is performed, for predicting the variables nature by finding the mean, standard deviation etc. It also gives other values for quantitative variables by the use of absolute and relative frequencies. The data is prepared by removing the missing values in each variable, as it is necessary to eliminate them, for the smooth and efficient working of some algorithms.

***4.4 Feature Selection*:**

The extraction and selection of feature is the most relevant steep, as it helps to impart the goal more nearly in reaching the accurate results. Since the selection of best features is a critical task, as many irrelevant features can be extracted out which may affect the efficiency of the classifier therefore, proper steps and measures must be imparted for it.

4.5 Data Splitting:

The data splitting is an essential step needed for evaluating the performance of model. This step is being used to split the chosen dataset into two parts: training and testing. The 80% data is used for training the model, while the 20% data is used for testing the performance of model. The dataset needed for training is also used for nine-fold cross-validation.

4.6. Training Algorithms*:*

The proposed research work uses the algorithm of Deep Learning, and the various classification algorithms of machine learning, for training the model and classifying the highest rate of accuracy for the chances of heart disease:

[A] Deep Learning

(i) Multi-Layer Perceptron (MLP): In this work, the major emphasis is to implicate the use of deep learning in the application of occurrence of heart disease. The multi-layer perceptron or MLP is the algorithm being applied in deep learning. The MLP works on the same principle of artificial neural network, which mimics the functionality and working of human brain. Like ANN, the MLP is also inspired by the working of biological nervous system and contains input layer, weight and output layer. The Deep Neural Network only differs with ANN, in having multiple hidden layers of complex interconnecting neurons between the two layers. The proposed algorithm of deep learning in Table 1, imparts a practical way for learning the real and vector-valued functions, over continuous and discrete-valued attributes, in a manner to perform the robustness of training data.

[B] Machine Learning

(i) **Logistic Regression**: The first classification algorithm of machine learning used in this work to train a model for detecting the higher chances of the occurrence of heart disease is the logistic regression. It is based on the concept of probability, and is used to assign the observations of a discrete class. The algorithm in its simple form is represented by binary or binomial form, which takes the dependent variable or target variable in only two possible values of 1 and 0. The hypothesis used in logistic regression ceases the cost function between 0 and 1, and transforms the output by sigmoid or logistic function, by allowing the model to define the relation between target and multiple predictor variables. The linear function in this algorithm is simply used to provide input to another function, such as:

$$h\theta(x) = g(\theta Tx) \text{ where } 0 \le h\theta \le 1 h\theta(x) = g(\theta Tx) \text{ where } 0 \le h\theta \le 1$$

where, g is the logistic or sigmoid function used. It can be defined as follows −

$$g(z) = 11 + e - z \text{ where } z = \theta Tx$$

(ii) **K-Nearest Neighbor (KNN)**: Another training algorithm of classification, to specify the instance based, nonparametric and lazy learning, for making the prediction of new samples of class, on the basis of its neighbor is called the k-nearest neighbor or KNN. The KNN is used in several groups to forecast the problems related to regression and classification. But, it is mostly applied in the industrial problems related to classification algorithm of machine learning, as it performs fairly well, across all the examined criteria being performed. It is mostly preferred, when assessing the functionality of a technique, due to its easily understanding and lower time of computation.

(iii) **Support Vector Machine (SVM)**: The Support Vector Machine or, SVM in short, is used to classify the linear and non-linear data. Although it can be used for both the classification and regression algorithms, it is primary used in classification algorithms. The goal of SVM is to find the best decision boundary or hyperplane, out of various other present, of data item between the two classes, in an n-dimensional feature space, where n is the number of features. The SVM algorithm actually works by first plotting each data point in the hyperplane, irrespective of its value in a specific coordinate, and then performs the classification of data point, to find the best hyperplane of the two classes on the basis of maximum margin. The marginal distance is defined, as the distance of a class

between its nearest instance, and the decision of a data item on the hyperplane for the respective member of class. Due to this, the SVM gives the best accuracy in low computation time.

(iv) **Naive Baye's**: The Naive Bayes algorithm is another algorithm of machine learning used for classification. This algorithm is based on the concept of Bayes Theorem, and is represented by family of probabilistic classifier. The Naive Bayes algorithm makes use of statistical independence, and the main objective to use this classifier is to find the posterior probabilities i.e. the probability of a class label L, of the given observed features, P ($L \mid features$). The algorithm works to find these probabilities by the use of Bayes theorem. On the basis of which the probability can be expressed in the quantitative form as:

P(L|features)=P(L)P(features|L)P(features) P(L|features)=P(L)P(features|L)$P(features)$

where, P ($L \mid features$) is used to define the posterior probability of class. The (L) is the prior probability of class, ($features \mid L$) is the occurrence of probability of predictor of a given class, and ($features$) is the prior probability of a predictor. The algorithm of Naive Baye's assumes to have the sturdy independence in between the various features, and is therefore, essential to make predictions [21]. Also, since this classifier is quite easy to build and generally performs well therefore, it is an appropriate choice to be used in medical field for the diagnosis of diseases.

(v) **Decision Tree**: Another algorithm of machine algorithm that is represented by a tree like structure is called the decision tree. The decision tree is used to show the instances of a prediction in a tree, by sorting it from the root to the leaf node, for making the classification and measuring the homogeneous target variable within these subsets, to provide a measure for making a good quality split. The decision algorithm is the best, accurate, and yet powerful algorithm for the predictive modeling, with having branches, internal and terminal nodes. The algorithm uses these nodes present in the internal tree, for performing the "test" on the features, and also uses branches for observing the "conclusion" of test, with class label representing each leaf node.

4.7 Evaluating Methods:

The proposed model is evaluated by describing the various criteria of metrics, for ex- the percentage of accuracy, recall value, the precision value, the F-score, the ROC, and the AUC.

(i) The Accuracy: The accuracy is defined, and considered to be one of the most important metric, for evaluating the correctness of the model being trained by algorithm. It is evaluated by finding the ratio between the total number of correct values and the total values. Mathematically, it is shown as:

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

where, TP = True Positive

TN = True Negative
FP = False Positive
FN = False Negative

(ii) Recall: The recall is defined as another metric of evaluation and used to measure the accuracy of model on the given data. It is used to predict the ability of the trained model, by the applied algorithm, in predicting the positive values. The recall is defined by the following:

$$\text{Recall or Sensitivity} = \frac{TP}{(TP + FN)} \quad (2)$$

(iii)Precision: The precision is used to analyze and predict about how often a model's positive value is correct? It is another criterion for evaluating the performance of model, and is calculated by measuring the ratio between the true positives values with the sum of all positive values. The precision is represented in the following way:

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (3)$$

(iii) F1 Score: The F1 score, or F score is the metric that is used to evaluate the performance of the model by application of particular algorithm. It is referred as the harmonic mean of precession and recall, and is considered a better metric of evaluation than accuracy and precision. The F1 score is represented by:

$$\text{F1 Score} = \frac{2 \times TP}{2 \times TP + FN + FP} \quad (4)$$

(iv) Receiver Operating Curve (ROC): The receiver operating curve is another criterion for evaluating the efficiency of a model, at all the levels of thresholds of classification algorithms. This roc curve is used to show the two parameters, true positive and false positive values on the plot respectively. It is obtained by indicating the area of the curve which is used to indicate the ability of classifier for differentiating among the various classes. It is therefore, predicted that if the area under the curve is greater, the efficiency of model is also increased greater, in differentiating between the two values.

The following figure - 1 depicts the complete flowchart of different processing steps of methodology of the proposed system.

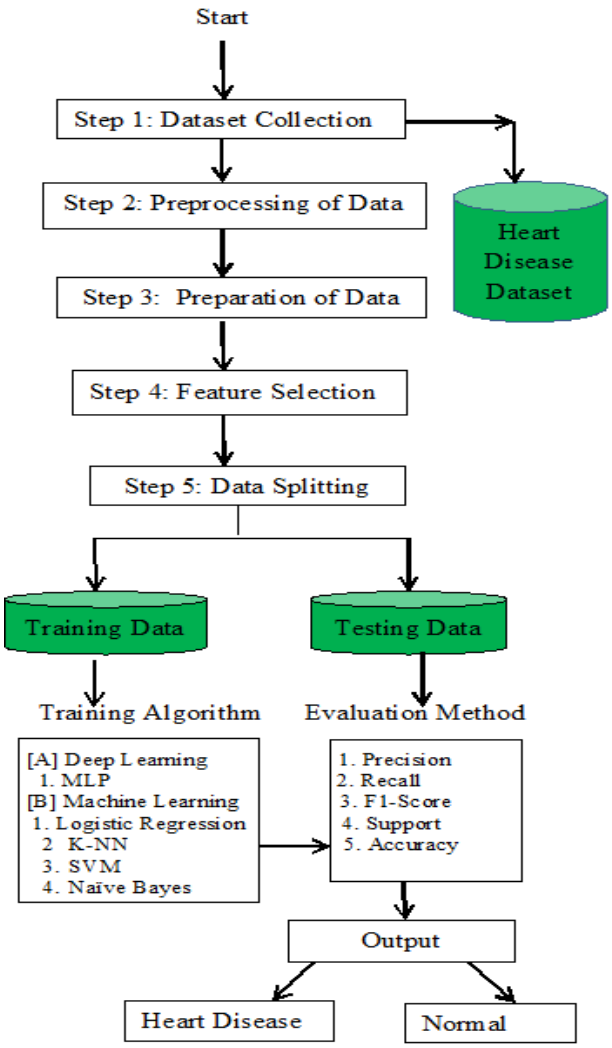

Figure 1: Flowchart for proposed system

5. THE RESULTS

5.1. Result of Data Preprocessing:

Table 3: The Summary Statistics

	Age	Sex	CP	---------	Ca	Thal	Output
Count	303.000	303.000	303.000	--------	303.000	303.000	303.000
Mean	54.366337	0.683168	0.966997	--------	0.729373	2.313531	0.544554
Std	9.082101	0.466011	1.032052	---------	1.022606	0.612277	0.498835
Min	29.000000	0.000000	0.0000000	---------	0.0000000	0.000000	0.000000
25%	29.000000	0.000000	0.000000	---------	0.0000000	2.000000	0.000000
50%	55.000000	1.000000	1.000000	---------	0.000000	2.000000	1.000000
75%	61.000000	1.000000	2.000000	---------	1.000000	3.000000	1.000000
max	77.000000	1.000000	3.000000	---------	4.000000	3.000000	1.000000

5.2. Deep Learning Results: By the application of deep learning, the following results were obtained:

I. Comparison with ANN

Table 4: Comparison between ANN & Deep Neural Network

S. No	Prediction Technique	Algorithm Used	Hidden Layers	Epoch	Accuracy
1.	Artificial Neural Network	BPN	1	100	83.667%
2.	Deep Learning	Multilayer Perceptron	5	100	88.524%

II. Model's Performance

Table 5: Classification Report for DNN

	Precision %	Recall %	F1-score %	Support %
0	0.94	0.76	0.84	38
1	0.85	0.96	0.90	53
Accuracy			0.88	91
Macro Average	0.89	0.86	0.87	91
Weighted Average	0.89	0.88	0.88	91
Accuracy Score: 87.91%				

5.3. Machine Learning Results:

A. Result for Logistic Regression:

Table 6: Classification Report for Logistic Regression

	Precision %	Recall %	F1-score %	Support %
0	0.94	0.76	0.84	38
1	0.85	0.96	0.90	53
Accuracy			0.88	91
Macro Average	0.89	0.86	0.87	91
Weighted Average	0.89	0.88	0.88	91

Accuracy Score: 87.91%

B. Result for K-Nearest Neighbor

Table 7: Classification Report for K-NN

	Precision %	**Recall %**	**F1-score %**	**Support %**
0	0.87	0.68	0.76	38
1	0.80	0.92	0.86	53
Accuracy			0.82	91
Macro Average	0.83	0.80	0.81	91
Weighted Average	0.83	0.82	0.82	91
Accuracy Score is : 82.41%				

C. Result for Support Vector Machine

Table 8: Classification Report for SVM

	Precision %	**Recall %**	**F1-score %**	**Support %**
0	0.90	0.71	0.79	38
1	0.82	0.94	0.88	53
Accuracy			0.85	91
Macro Average	0.86	0.83	0.84	91
Weighted Average	0.85	0.85	0.84	91
Accuracy Score: 84.61%				

D. Result for Naïve Bayes

Table 9: Classification Report for Naïve Bayes Classifier

	Precision %	**Recall %**	**f1-score %**	**Support %**
0	0.88	0.74	0.80	38
1	0.83	0.92	0.88	53
Accuracy			0.85	91
Macro Average	0.86	0.83	0.84	91
Weighted Average	0.85	0.83	0.84	91
Accuracy Score: 84.61%				

E. Result of Decision Tree

Table 10: Classification Report for Decision Tree

	Precision %	**Recall %**	**f1-score %**	**Support %**
0	0.70	0.79	0.74	38
1	0.83	0.75	0.79	53
Accuracy			0.77	91
Macro Average	0.77	0.77	0.77	91
Weighted Average	0.85	0.83	0.84	91
Accuracy Score: 76.92%				

5.8. The Cluster Plot:

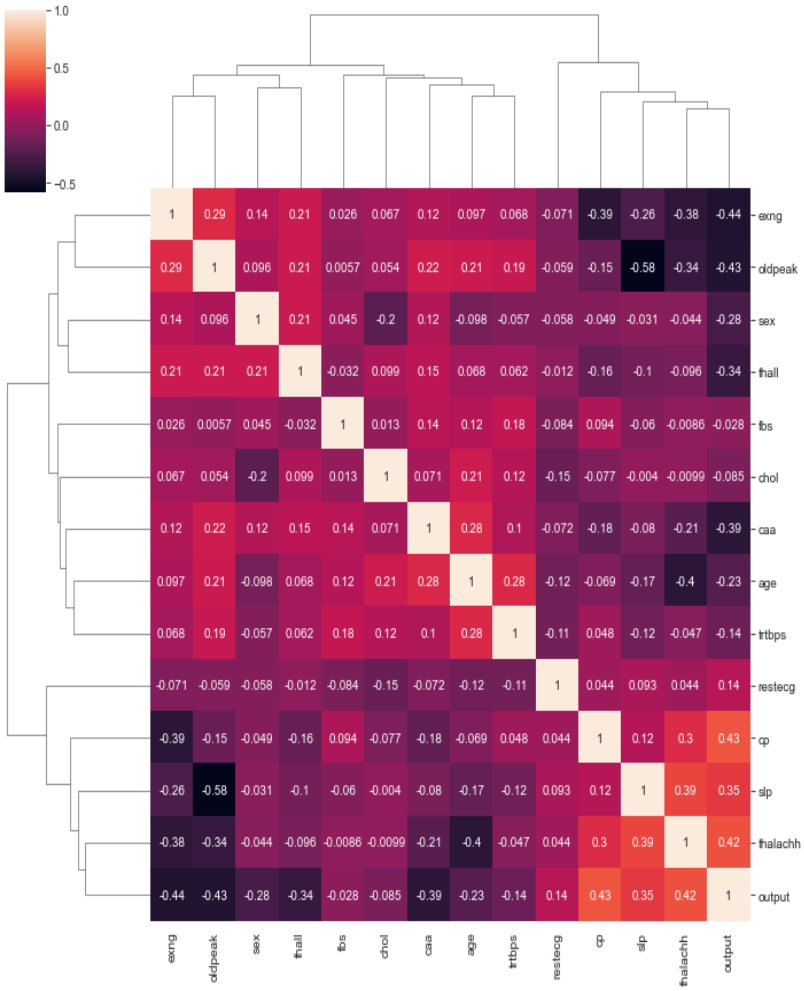

Figure 2: The Cluster Plot

5.5. The ROC Curve

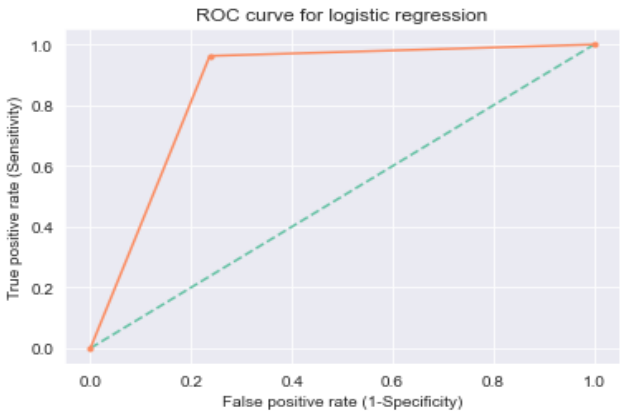

Figure 3: The ROC Curve for logistic regression

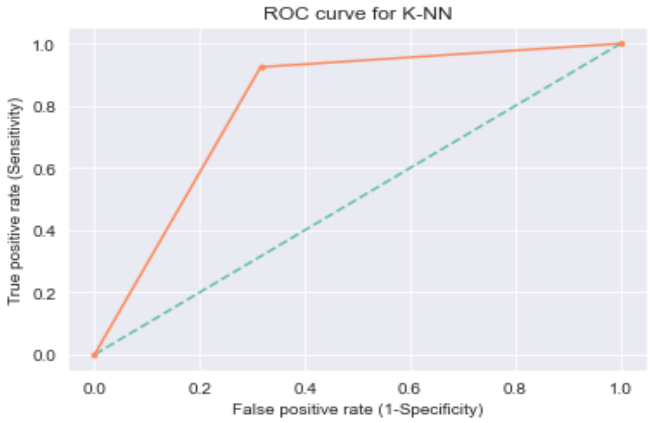

Figure 4: ROC Curve for K-NN

Figure 5: ROC Curve for SVM

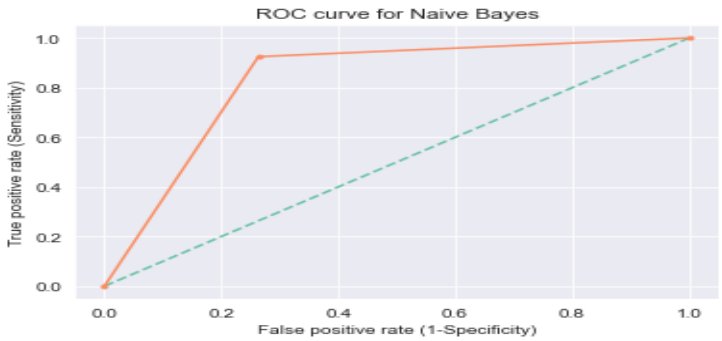

Figure 6: ROC Curve for Naïve Bayes

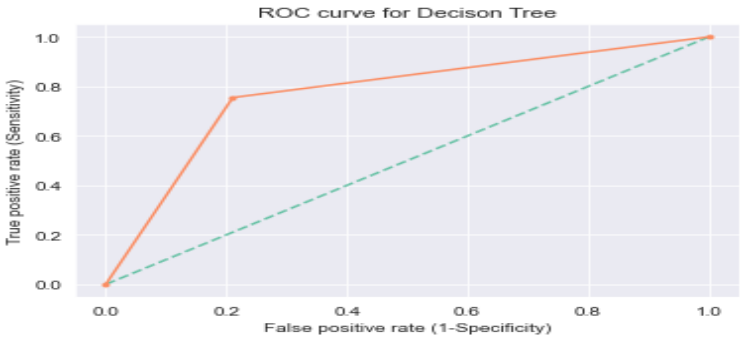

Figure 7: ROC Curve for Decision Tree

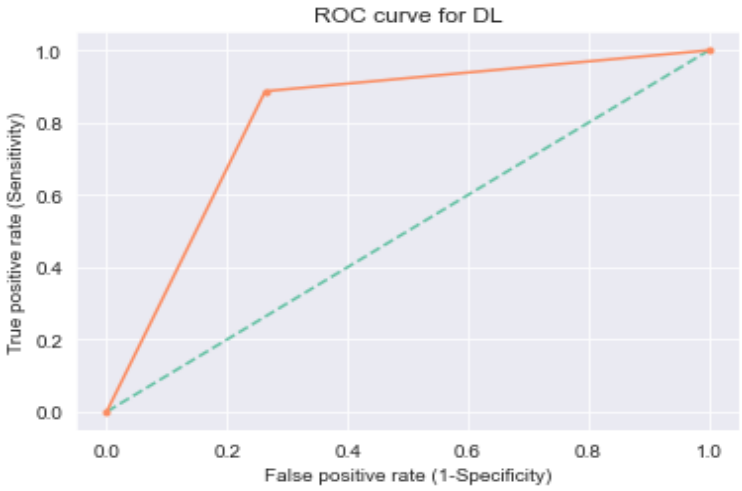

Figure 8: ROC Curve for Decision Tree

5.6. Accuracy Comparison Results

Table 11 Accuracy Comparison Chart

S.No	Algorithm	Accuracy
1	Logistic Regression	87.9%
2	K-Nearest Neighbor	82.41%

3	Support Vector Machine	84.61%
4	Naïve Bayes	84.61%
5	Decision Forest	76.92%
6.	Deep Neural Net	87.91%

6. CONCLUSION

It has been observed by this work that the use of deep learning technique of machine learning, proved to be a better option for predicting the chances of occurrence of heart disease, in comparison to the various other classification algorithm. It is also verified by this work that the concept of deep learning emerged to be a good choice, as it gives good results in comparison to the simple Artificial Neural Network. Hence, it is concluded, on the basis of the results obtained by the comparison of deep learning with other classification algorithms of machine learning that deep learning proves to be a better choice, and can be used efficiently by medical professionals for its accurate detection.

7. RESULT & FUTURE SCOPE

In this work it is found that the concept of Deep Learning emerges as a powerful technique, and is a better option for investigating the accurate chances of occurrences of heart disease by a medical professional, in comparison to the other machine learning techniques. In reference to that it is also expected that this research work, using the concept of deep learning can also be extended in future, to enhance and analyze it further with some hybrid approaches, for getting better investigations techniques with the timely prediction and high percentage of accuracy.

8. REFERENCES

[1] S. S. Yadav, S. M. Jadhav, S. Nagrale, and N. Patil, “Application of machine learning for the detection of heart disease,” 2nd International Conference on Innovative Mechanisms for Industry Applications, (ICIMIA), vol. 165–172, Bangalore, India, March 2020

[2]S.Adithya Varun, G.Mounika, Dr. P.K. Sahoo, K. Eswaran, “Efficient system for Heart disease prediction by applying Logistic regression” International Journal of Computer Science & Technology, Vol 10, issue 1, march 2019

[3] Heart disease prediction using machine learning algorithms to cite this article: Harshit Jindal et al 2021 IOP Conference Series: Mater. Sci. Eng. 1022 012072

[4] M. Aljanabi, H. Qutqut, and M. Hijjawi, “Machine learning classification techniques for heart disease prediction: a review”, International Journal of Engineering and Technology, Vol. 7, pp. 5373–5379, 2018.

[5] Devansh Shah, · Samir Patel, Santosh Kumar Bharti, “Heart Disease Prediction using Machine Learning Techniques”, Springer Nature Computer Science 2020 1:345 https://doi.org/10.1007/s42979-020-00365-y

[6] P.Santhi, R.Ajay, D.Harshini and S.S.Jamuna Sri, “A Survey on Heart Attack Prediction Using Machine Learning”, Turkish Journal of Computer and Mathematics Education, Vol.12 No.2, pp 2303 – 2308, 2021

[7] Rubini PE, Dr. C. A. Subasini, Dr. A. Vanitha Katharine, V. Kumaresan, S. Gowdham Kumar, T.M. Nithya, “A Cardiovascular Disease Prediction using Machine Learning Algorithms”, Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 2, 2021, Pages. 904 - 912 Received 20 January 2021; Accepted 08 February 2021

[8] Alexandra Norman, James Harding, Daria Zhukova, "Machine Learning in the Health Industry: Predicting Congestive Heart Failure and Impactors," SMU Data Science Review: Vol. 5: No. 1, Article 7, 2021

[9] Chetan Sharma , Shankar Shambhu, Prasenjit Das, Shaily Jain, Sakshid, “Features Contributing Towards Heart Disease Prediction Using Machine Learning”, ACI’21: Workshop on Advances in Computational Intelligence at ISIC 2021, February 25-27, 2021, Delhi, India

[10] Tulasi Krishna Sajja, Hemantha Kumar Kalluri, “A Deep Learning Method for Prediction of Cardiovascular Disease Using Convolutional Neural Network”, International Information and Engineering Technology Association, Vol. 34, No. 5, October, 2020, pp. 601-606

[11] Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain and Preeti Nagrath, “Heart disease prediction using machine learning algorithms”, IOP Conf. Series: Materials Science and Engineering, 1022 (2021) 012072 IOP Publishing doi:10.1088/1757-899X/1022/1/012072

[12] Yogita Solanki, Sanjiv Sharma, “Analysis and Prediction of Heart Health using Deep Learning Approach”, International Journal of Computer Sciences and Engineering Open Access Research, Paper Vol.-7, Issue-8, Aug 2019 E-ISSN: 2347-2693,pp 309-315

[13] W. H. O. (WHO), "Cardiovascular Diseases." https://www.who.int/healthtopics/cardiovascular-diseases#tab=tab_1 (11 June, 2021)

[14] Himanshu Sharma and M A Rizvi, "Prediction of Heart Disease using Machine Learning Algorithms: A Survey", International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 5, no. 8, August 2017.

[15] K.H. Miao, H.J. Miao, “Coronary heart disease diagnosis using deep neural networks”, International Journal of Advance Computer Science Application, Vol 9,Issue 10, 2018.

[16] UCI Machine Learning Repository. Available from: https://archive.ics. uci.edu/ml/index.php. Accessed November 01, 2018.

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

[18] https://docs.anaconda.com/anaconda/

[19] S. Pouriyeh, S. Vahid, G. Sannino, G.D. Pietro, H. Arabnia, et al., “A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease”, IEEE Symposium on Computers and Communications (ISCC), pp. 204–207,2017

[20] C. Sowmiya and P. Sumitra, "Analytical study of heart disease diagnosis using classification techniques," 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Srivilliputhur, pp. 1-5, 2017 doi: 10.1109/ITCOSP.2017.8303115.

[21] A. N. Repaka, S. D. Ravikanti and R. G. Franklin, "Design And Implementing Heart Disease Prediction Using Naives Bayesian," 2019 3rd International Conference on Trends in

Electronics and Informatics (ICOEI), Tirunelveli, India, pp. 292- 297, 2019 doi:10.1109/ICOEI.2019.8862604

[22] Awan, S. E., Sohel, F., Sanfilippo, F. M., Bennamoun, M., & Dwivedi. G. "Machine learning in heart failure", Current Opinion in Cardiology, Vol 33(2), pp 190-195,2018,doi:10.1097/hco.0000000000000491

[23] Dinesh K.G, Arumugaraj K, Santhosh K.D, Mareeswari V, "Prediction of Cardiovascular Disease Using Machine Learning Algorithms", Published in IEEE, 2018 DOI: 10.1109/ICCTCT.2018.8550857

[24] Archana Singh; Rakesh Kumar, "Heart Disease Prediction Using Machine Learning Algorithms", 2020 International Conference on Electrical and Electronics Engineering (ICE3), 2020 DOI**:** 10.1109/ICE348803.2020.9122958,

[25] Khaled Mohamad Almustafa, "Prediction of heart disease and classifiers' sensitivity analysis", BMC Bioinformatics, Volume 21, Article number: 278, 2020

[26] Lafta R, YanLi, Tseng VS. "An Intelligent Recommender System based on Short Term Risk Prediction for Heart Disease patients", IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). Vol 3,6-9 Dec, 2015