International Journal on Emerging Technologies 11(3): 531-538(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Heart Disease Detection using Core Machine Learning and Deep Learning Techniques: A Comparative Study

Sujay Das[1], Ritesh Sharma[2], Mahendra Kumar Gourisaria[3], Siddharth Swarup Rautaray[4] and Manjusha Pandey[4]

[1]*B.Tech. Student, School of Electronic Engineering, KIIT Deemed to be University, Bhubaneswar (Odisha), India.*
[2]*B.Tech. Student, School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar (Odisha), India.*
[3]*Assistant Professor, School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar (Odisha), India.*
[4]*Associate Professor, School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar (Odisha), India.*

(Corresponding author: Mahendra Kumar Gourisaria)
(Received 04 March 2020, Revised 22 April 2020, Accepted 25 April 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: At the present time the load on an individual person has increased significantly due to an increase in one's work. Because of the dire circumstance which cannot be avoided, there is a high probability that the person is going to suffer from heart disease. In India mostly people die due to heart disease so it is a great matter of concern. Since the ancient times humans have significantly improved in the domain of machine and health care. In the modern times after the entry of machine and AI in the field of medicine and health care there has been a significant development and improvement in the domain of medicine and health care. New diseases are now being discovered frequently and are being eliminated by the help of certain evolved technologies. Heart disease being a matter of primary concern this paper involves a predictive model which discovers whether a person is having heart disease or not by the use of various classification algorithms like logistic regression, decision trees, K-Nearest neighbors, SVM, Deep learning algorithm which is ANN etc. In this paper the authors has compared all the above algorithms and have found that ANN performs best out of all these algorithms.

Keywords: Decision Trees, Logistic Regression, SVM, Machine Learning, Deep learning, Fuzzy neural network, Heart disease detection, ANN, K-NN

Abbreviations: ANN, Artificial Neural Network. K-NN, K-Nearest Neighbor, SVM, Support Vector Machines.

I. INTRODUCTION

Heart is one of the import muscular organs just of the size of a fist of a human being. Heart is responsible for pumping blood into the blood vessels of the circulatory system. A human being heart has a total of four chambers which are the right atrium, the right ventricle, the left atrium and the left ventricle. The function of right atrium is it pumps the blood into the right ventricle which it receives from the veins. The function of right ventricle is it pumps the blood into the lungs which is receives from the right atrium. The function of left atrium is it collects the blood full of oxygen which it receives from lungs and pumps it into left ventricle. The function of left ventricle is to pump the blood which is full of rich oxygen into the rest of the body part [15]. While pumping the blood the heart beats with a rhythm which is recognized because of the pace making cells present in the senatorial node. The average heartbeat of a resting person is 72 bpm. The heartbeat of a person increases which the person is doing exercise but in long term the resting heart rate lowers which in turn improves the health of the heart. The range of conditions which affect your heart is described by heart disease. There are numerous disease that the heart faces which can be proved fatal if not diagnosed properly. The disease which affect a heart are Congenital heart disease, Arrhythmia, Coronary artery disease, Dilated cardiomyopathy, Myocardial infarction, Heart failure, Hypertrophic cardiomyopathy, Mitral valve prolapsed, and Pulmonary stenosis. Congenital heart disease is a general term for some disorientation of heart which was there since the birth. The deformities can be of numerous types like spectral defects, Obstruction defects and cyanotic heart disease. Arrhythmia can also be termed as a irregular heart beat. We can say that a person is suffering from arrhythmia if he/she is suffering from arrhythmia. There are numerous ways in which a heart loses its regular rhythm which are tachycardia, brachicardia, premature and fibrillation. This disease can be fatal when a heart rhythm changes too much or fluctuates too much. This is because of a damaged heart. They need to treat as soon as possible because it can be fatal for a person.

Coronary arteries are the type of artery that supply the muscle of the heart by circulating oxygen and nutrients by circulating blood. Sometimes the coronary arteries become diseased and damaged. This is because due to the deposited of plaque which consists of cholesterol. The plaque which gets build up blocks the artery and which results in not providing the heart with sufficient nutrients and oxygen. Heart failure is a very fatal condition where a person's heart does not pump blood around the body anymore. This is a very serious condition in which a person's life is at risk. In this disease the person has to do consult a doctor as soon as possible [2]. The evolution of AI and ML has developed and improved the field of medicine and health care beyond any person's imagination.

Now a day the machines perform task way more efficiently than a human being. Machine trained by Artificial Intelligence perform task with great ease with much precision. Now days a machine can do all type of work a doctor is capable of doing like interacting with the patients, diagnosing them, and writing down the details of a patient in a notepad. Machine learning not only helps in treating humans but also helps in treating plants and animals [3].

The proposed model takes a set of values like age, sex, chest pain type, cholesterol level, fasting blood sugar level and all and predicts whether a person is suffering from heart disease of not. The model uses some complex classification algorithms of machine learning like Logistic regression, decision trees, Random forest, SVM. The dataset contains age, sex, the type of pain in chest, a person's level of cholesterol, level of fasting blood sugar and all. After the classification algorithm is applied we get the predicted values which are 0 and 1. 0 stating that the person is healthy and 1 stating that person is having some kind of heart disease.

For coding the proposed model we have used python programming language as it is a best choice for machine learning. The sections in this paper are as follows: Section I is the Introduction, section II is the discussion about the related work, section III gives knowledge about our proposed model and the implementation, Section IV compares the model used in previous implementations, section V discusses about the result got after the successful implementation of the model and section VI concludes this paper.

II. RELATED WORKS

Dai *et al.,* (2015) proposes a model whose main aim is to precisely measure and predict very efficiently the heart related hospitalization which depends upon the patients data which is available and his or her medical history. They have used a dataset in which the patients are tagged and are randomly separated into training and testing dataset. They have used a number of machine learning algorithms which are SVM, Naïve Bayes etc. Their model show the result that under 30% false alarm rate, the detection could be as high as 82% [4].

Abdar *et al.*, (2015) proposes a model which uses data mining technique to predict the odds of patients suffering from heart disease. The authors of the paper have used data mining techniques to predict the people who are suffering from heart disease. They have used 5 algorithms which are logistic regression, SVM, K-Nearest Neighbor, Decision Trees and C5.0 neural network. Their model has achieved 92.03% for C5.0, KNN SVM and neural network has achieved an accuracy of 88.37%, 86.05%, and 80.23% respectively [5].

Nayak *et al.*, (2019) has presented a paper which diagnosis heart disease using diverse data mining classification methods like Support Vector Machine, K-NN classification algorithm, Machine classification. These techniques are used to predict the disease at a early stage. They have detected to it in an early stage so that it can be curable and preventive. They have achieved the following accuracy 84.91% for Decision trees, 88.68% for SVM, 96.23% for Naïve bayes, 58.49% for KNN (roc), 62.26% for KNN (Acc). They have used R data analytical tool [6].

Latha and Jeeva (2019) has implemented a model which uses works like ensemble model to increase the accuracy of weak classifiers. It not only increases the accuracy of the weak classifiers but also improves the model so that the disease can be detected at an early stage. It also implements the model to a medical dataset to show the early prediction of the disease. The study shows the use of bagging and boosting. The authors have achieved an increase in accuracy of 7% after using the ensemble model. They have further improved the result by the use of feature selection implementation; the results which they have obtained have showed a significant improvement [7].

Haq *et al.*, (2018) has implemented a model which uses several ML algorithms which are not only popular but they are also effective. They have use three feature selection algorithm, the cross validation model and classifier based evaluation matrices in their paper. They have discussed all the classifiers, data preprocessing methods, validation method, and classification evaluation metric. Their proposed machine learning model will help and assist the doctors to diagnose the disease [8].

Ramalingam *et al.*, (2018) proposes a model various machine learning algorithms to predict heart disease. Their model uses complex machine learning algorithms like Support Vector Machine (SVM), K-Nearest Neighbor, Naïve Bayes, Decision Trees, Random Forest and ensemble models which are found very popular among researchers [9].

Santhana and Geetha (2019) proposes a model which can detect the heart disease using machine learning. They have used the popular machine learning language python to code the model. The algorithms which they have used are also quite popular namely Decision trees and Naïve Bayes. They have also compared which algorithm shows better result in the term of accuracy level of heart disease. They have used Data Mining [10].

III. PROPOSED MODEL

Machine learning is a domain which gives the computers enough power. The term machine learning is quite fascinating and it excited anyone that comes across it. Machine learning is used quite often in today's world, in such places that one individual would expect. There is a huge difference between machine learning and traditional programming [42]. In the traditional programming method we feed in the data and the program to get the output by running it on the machine. But in machine learning we feed in the data and output during the training phase and the machine creates its own programming logic which helps us to predict the result. A machine learning model is splits the dataset into two halves which are training part and testing part. In the training part the model trains on the data by the help of a specific algorithm to find a pattern. After the training phase comes testing phase. In this part we find out how a machine learning model is trained. In the testing phase we find out the accuracy of the model which tells whether a model is under fit, good or over fit [11].

A machine learning model can be of three types which are model which are fine or in other words it produces

accuracy as expected. Another one is under fitting model. A model is defined as under fitting when it performs very bad and gives very less accuracy at the time of testing and training. Such type of model can be termed as under fitting model. A model can be said over fitting if the model performs exceptionally well in the training phase and performs badly and provides less accuracy in the testing phase. Such type of model can be said as over fitting model. An idle machine learning model is a model which performs well in training phase as well as testing phase. To build a machine learning model, we have to divide or segment the steps into important clusters which are gathering past data in any suitable format for preprocessing the data which we have gathered, splitting the data into training and testing sets, building the model with suitable algorithms and testing our model and evaluating it using an evaluation metric such as F1 score, precision and recall [12, 43]. As evident from the name machine learning provides the ability to learn to the machines. Machine learning is used in all day to day applications such as Web search application, Photo tagging application, Spam detector, and all. The algorithms which are used to create a machine learning model are divided into various categories which are supervised learning, unsupervised learning, reinforcement learning, Natural language processing and Dimensionality Reduction [40].

We have proposed a model which trains on a dataset which contains 76 features but we have taken 14 features into our consideration which we found more relatable. We have divided the dataset into training and testing set and have trained them on various machines learning algorithms and also trained them in Artificial Neural Network which is a deep learning algorithm, which have immensely helped us to find out the result and have helped us to obtain high accuracy.

The dataset which we got was not upto the mark. There were lot of anomalies in the data like missing values, duplicate data values and there some attributes which were highly correlated with each other. We have cleaned the dataset by dropping the correlated columns and by implementing appropriate techniques on our dataset.

A. Dataset Description

We have used a heart disease dataset from UCI machine learning repository. It contains 76 attributes but we are using some important 14 attributes out of those. Those attributes used in our model are sex, age, blood pressure, pain in chest, pain location, chest pain type, cholesterol level, resting blood pressure etc. The more information about the dataset is mentioned below [41].

Table 1: Mean of all the features in the dataset.

target	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
0	56.601	0.826	0.478	134.398	251.086	0.159	0.449	139.101	0.550	1.585	1.166	1.166	2.543
1	52.496	0.563	1.375	129.303	242.230	0.139	0.593	158.466	0.139	0.583	1.593	0.363	2.121

B. Data Exploration

Data Exploration is a key part of building model using machine learning. Data Exploration helps in determining which data feature or data point is relevant for building our model, whether the data is left skewed or right skewed, is there any missing data or not and etc. In our dataset we found that there were some missing continuous variables which were replaced by the mean of all the data points and the categorical data points were replaced by the mode of the categorical data points. We also found that our data was perfectly skewed.

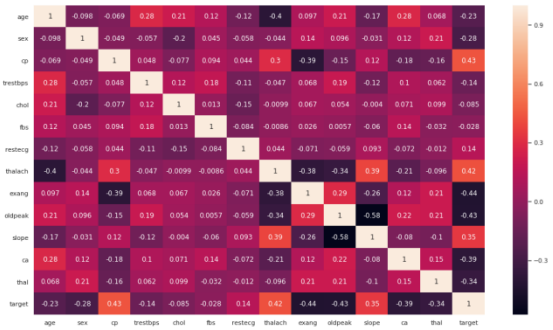

Fig. 1. Correlation matrix.

C. Splitting the Data

Splitting the data turns out to be one of the important criteria for building a model. We have to split the data in order to divide the dataset into two levels that are training part and testing part. For the working of our model we have split the dataset into 80% of data into training set and 10% into testing set and 10% into validation set. The data was also divided randomly to find the precise and accurate result.

D. Logistic Regression

The complex and popular machine learning algorithm which is also known as logistic regression was popularly used in science of biology during the beginning of twentieth century. After that the logistic regression was used in numerous number of social science applications. We can use logistic regression when the dependent variable is categorical [13]. In our problem statement that is predicting heart disease, a threshold value is need to be set if we use linear regression for classification. If the correct outcome comes as the person is suffering from disease and our model predicts there is 0.4 probabilities that the person is suffering from disease and our threshold value of 0.5 and our model classifies it as the person not suffering from any kind of disease then there will be some serious consequence afterwards [14]. So we get to know that linear regression is not bounded and that is the sole reason why logistic regression comes into picture. The value of logistic regression strictly varies between 0 to 1 [1]. There are various types of logistic regression [16, 17] which are namely Binary logistic regression, Multinomial Logistic Regression and Ordinal Logistic Regression. The basis of logistic regression is from sigmoid activation function represented by

$$\sigma(z) = p = \frac{1}{1 + e^{-z}}$$

After running our model using logistic regression we get an accuracy of 86% in our own designed algorithm of logistic regression and 88% accuracy using sklearn logistic regression library.

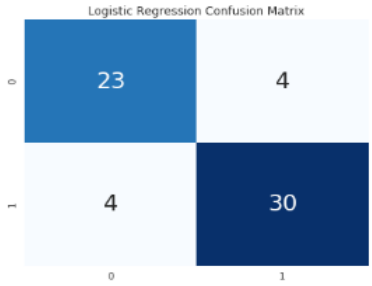

Fig. 2. Confusion Matrix for logistic regression.

E. Decision Trees

A complex ML algorithm can be divided into three types which are regression, classification and clustering [18]. Decision trees is a type of machine learning algorithm that comes under supervised learning as it requires label to get operated. Decision trees algorithm is used where data is regularly split according to a particular parameter [19]. Decision trees can be used for both regression model as well as classification model. They key difference between regression model and classification model is regression model is used to predict a continuous value while classification model predicts the probability or percentage. In our model we have implemented decision tree classifier or a categorical variable decision tree because our model is a binary classifier model and we have predicted whether the person is suffering from heart disease or not [20].

The important expressions which are related to decision trees are root node, splitting, Decision Node, Leaf/Terminal Node, Pruning, Branch, Parent and child node [21].

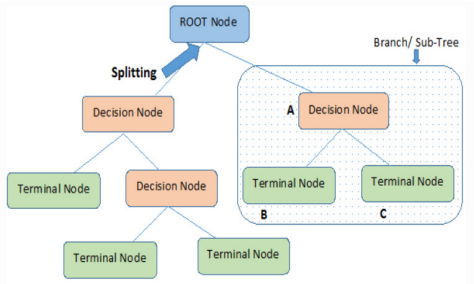

Fig. 3. Overview of decision tree.

After training our model with decision trees we get our accuracy of 79.33%

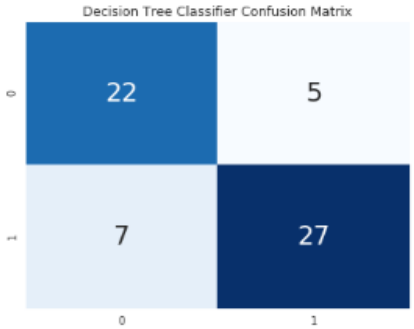

Fig. 4. Confusion matrix for decision tree.

F. Naïve Bayes

In the supervised learning category Naïve Bayes come into play. Naïve Bayes can be stated as a most common Machine Learning technique which is often used for classification purpose. The popular ML technique Naïve Bayes is a complicated classification technique to work with. It is also popularly known as probabilistic classification technique as its core concept is to use probability to make predictions to classify [22]. These algorithms are a collection of famous algorithms which are based in the popular Bayes theorem [23].

We have divided our dataset into a total of two parts that are a matrix which contains features and response vector. The matrix that is feature matrix consists of all the rows of the dataset in which each and every vector contains the effectiveness of dependent features [24]. The response vector consists of the effectiveness of class variable which in this case are yes and no for each. In our dataset the class is person suffering from disease or not. The crux of Naïve Bayes if that each feature makes an independent or equal contribution to the outcome [25].

The classifier Naïve Bayes is completely dependent upon the theorem Bayes theorem. The Bayes theorem is a probability theorem that relates conditional probability. In simple words it discovers the probability of an event occurring give the probability of another already occurred event [26]. It can be mathematically stated as

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

In the above calculation we are trying to find the probability of the given event which is y, given that the event X is true. In this equation event X can be termed evidence [27]. P(y) is the priori of y. P (y|X) can be stated as posterior probability of X.

The equation of Naïve Bayes after implementing it in a classifier model is:

$$y = argmax_y P(y) \prod_{i=1}^{n} P(x_i|y)$$

After implementing our model with Naïve Bayes we get the accuracy as 87.89%.

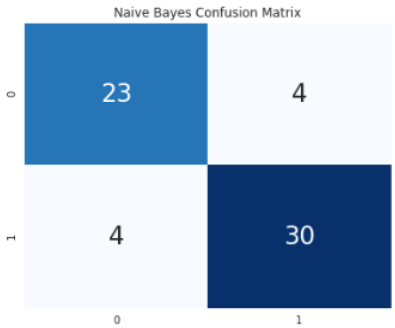

Fig. 5. Confusion matrix for Naive Bayes.

G. Random Forest

The popular classification ML algorithm Random forest is a type of supervised learning algorithm which is very popular when it comes to classification techniques. A decision tree can be considered as a very basic foundation of random forest algorithm [28]. Similarly like

a decision tree random forest algorithm can be used with regression and classification. Just like a forest is comprised of trees, a random forest is comprised of a numerous amount of decision trees. A random forest creates decision trees on data samples selected randomly. It usually gets its prediction from each tree and the best solution is selected by the means of voting [29]. A random forest can be used in variety of applications like search engines, image classification and selecting numerous numbers of features. Random forest follows a classic algorithm which is Boruta algorithm. This algorithm selects the important features of a dataset.

The algorithm random forest works in a very unique yet elegant way. It generally works in four steps. It selects random samples from a specified dataset. Then the output from each decision tree is got by constructing the decision tree from each sample [30]. A vote is performed for each output. The output which gets the most votes wins as the final prediction. After implementing our model with the random forest classifier we get an accuracy of 89.82%.

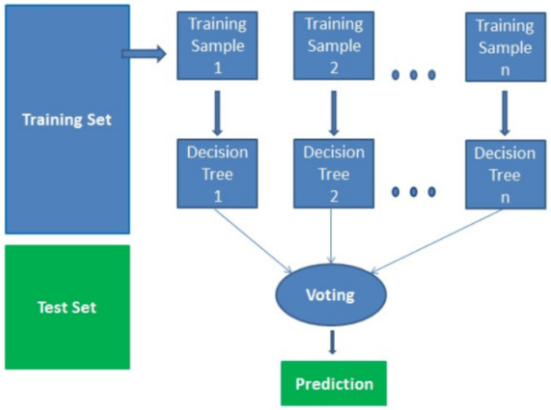

Fig. 6. Working of a random forest model.

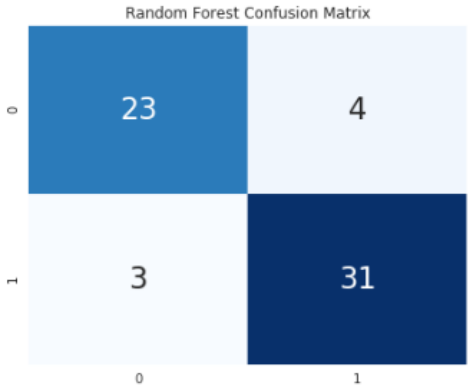

Fig. 7. Random forest confusion matrix.

H. Support Vector Machine

SVM can be considered as a linear model for doing classification and regression tasks [31]. SVM is known for solving linear as well as nonlinear problems and it also does a good job for working with many real world problems. The main idea behind Support Vector Machine algorithm is very elegant yet complex [32]. It creates a hyper plane the separates the data according to their respective classes. SVM finds the best hyper plane or line to separate the classes in a very effective way. It finds the points which are quite aloes to the line from both the classes. The points are popularly known as support vectors. The distance is often called as a margin [33]. The hyper plane for which the margin is maximum is known as optimal hyper plane. The complexity arrives when the data is not linearly separable [34]. In this case the SVM tries to add another axis and then separates the data. After training our model with the SVM algorithm we found out that the accuracy of the model is 89.59%.

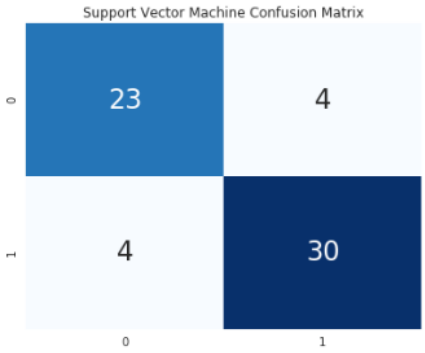

Fig. 8. Support Vector machine confusion matrix.

I. K-Nearest Neighbor

An algorithm can be called supervised if it relies input data which is labeled, to learn and then produce appropriate result when given a new data which is unlabeled. A problem can be called classification problem if it has discrete output values [35]. The algorithm K Nearest Neighbors thinks that the things which are similar exist in close proximity. If we simplify these we find that things which are similar lie near to each other.

We have implemented the KNN algorithm right from the scratch. We have first initialized K to choose a right amount of neighbors. For each data, the distance between the query neighbors and current neighbors were calculated, and then after that the index of an ordered collection is added with the distance [36]. After adding we have sorted the collection indices from minimum to maximum by the distance. After sorting we have picked the first K entries from the sorted collection. Our problem statement is a classification one so we have returned the mode of K labels.

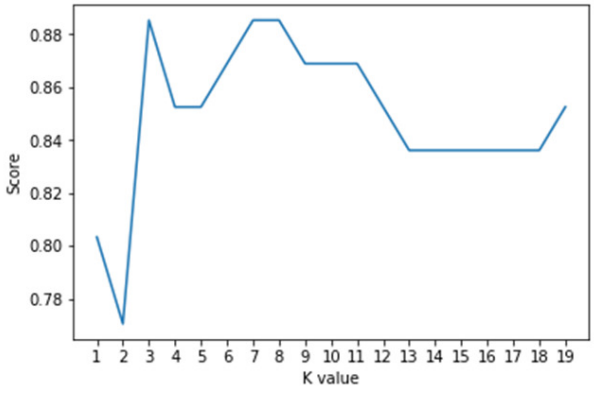

Fig. 9. K-value graph.

After implementing our KNN algorithm we found the accuracy to be 89%.

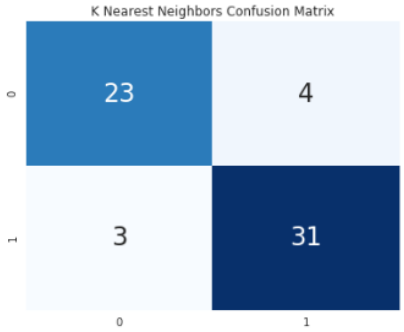

Fig. 10. K-Nearest Neighbor confusion matrix.

J. Artificial Neural Network

Neural network are a set of code that are inspired from biological neurons and have decision making capability just like them [37]. When a system or model is designed using the neural networks then they learn by the help of data. The connections are modeled as weights [44]. A connection which is positive can be represented as excitant connection and a connection which is negative can be stated as inhibitory connection [38]. A simple neural network consists of three items which are the input layer, the hidden layer and the result layer. The inputs that are given in a neural network are given as weights and are summed up. In the hidden layer activation functions are applied, in our model we have applied the ReLU activation function and sigmoid activation function [39]. We have used the binary cross entropy function as the loss function and Adam optimizer.

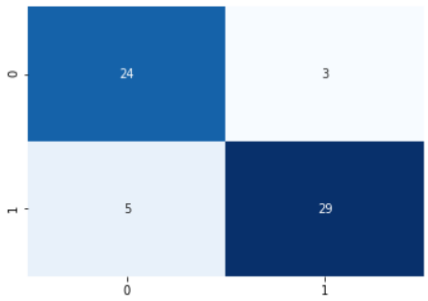

Fig. 11. ANN Confusion Matrix.

After using the ANN model we find out that our model gives an accuracy of 92% when ran for 100 epochs. For this we have used Google's cloud platform that is Colab.

IV. COMPARISION AND RESULT

After implementing our model in variety of machine learning algorithms and implementing our model in ANN we found out the following result. For Logistic regression we get an accuracy of 88%, for Decision Trees we get an accuracy of 79.33%, for Naïve Bayes we get an accuracy of 87.89%, for SVM we get an accuracy of 89.59%, for Random forest we get an accuracy of 89.82%, for K Nearest Neighbor we get an accuracy of 89% and for Artificial Neural Network we get an accuracy of 92%. Dai *et al.*, (2015) has got a false alarm rate of 30% and accuracy of 82% [4]. They have used different machine learning algorithms like SVM and Naïve Bayes. Abdar *et al.,* (2015) proposed a model which shows an accuracy of 80.23% for the neural networks. In our study we found that the highest accuracy in core machine learning model is given by random forest algorithm i.e. 89.82% and our proposed ANN model is giving an accuracy of 92% [5].

Table 2: Comparison of various algorithms used.

Algorithm Name	Accuracy
Logistic Regression	88%
Decision Trees	79.33%
Naïve Bayes	87.89%
SVM	89.59%
Random Forest	89.82%
KNN	89%
ANN	92%

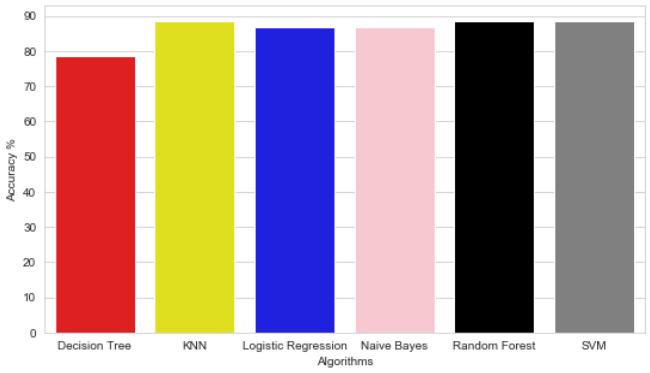

Fig. 12. Comparison graph for various algorithms used.

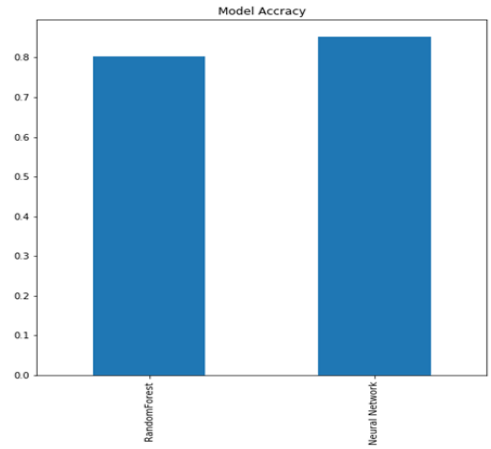

Fig. 13. Comparison of Random Forest and ANN.

V. CONCLUSION AND FUTURE WORK

As we know in modern time's heart diseases are quite common in the world. If not treated within a specific time or detected early heart diseases can prove to be fatal. Heart diseases can occur for a lot of reasons like smoking, internal damage of heart, pressure and stress etc. If not treated properly it can take life of a human being. But during the recent time artificial intelligence, machine learning and deep learning have helped the world by detecting the heart disease with much ease.

They not only perform the work efficiently but also take quite a less time in detecting the heart disease.

Our proposed model takes various features like age, sex, BP, resting blood pressure, chest pain location, cigarettes per day etc. We have used six types of classification algorithms like logistic regression, decision trees, SVM, random forest, naïve bayes, and KNN. We have also ANN (popularly known as Artificial Neural Network) to classify whether a person is suffering from heart disease or not. We have got the highest accuracy as 92% for ANN hence proving that deep learning models are better at classifying the heart diseases.

Our proposed model can be implemented in a website in which a patient has to register and will have to fill the required fills like age, sex, blood pressure etc and the website could predict if he or she is suffering from heart diseases or not. This can also be implemented in IOT devices as well as embedded systems. A mobile can also be made using machine learning and native app development software to detect and predict heart diseases.

The following research can be helpful in many real-world medical and healthcare domain. The model proposed in this study can be used over the bio-technical devices for automated and faster detection of heart diseases which will help the medical industry to grow and take next major step towards automation.

ACKNOWLEDGEMENTS

We would like to thank our co-authors who has given their constant support throughout this research. We would also like to thank KIIT Deemed to be University for their continuous support.

Conflict of Interest. There is no conflict of interest.

REFERENCES

[1]. Mythili, T., Mukerji, D., Padalia, N., & Naidu, A. (2013). A Heart Disease Prediction Model using SVM Decision Trees-Logistic Regression (SDL). *International Journal of Computer Applications*, *68*, 11-15.

[2]. Chaurasia, V., & Pal, S. (2013). Early Prediction of Heart Diseases Using Data Mining Techniques. *Caribbean Journal of Science and Technology, 1,* 208-217.

[3]. Rautaray, S.S., Pandey, M., Gourisaria, M.K., Sharma, R. and Das, S., (2020). Paddy Crop Disease Prediction – A Transfer Learning Technique. *International Journal of Recent Technology and Engineering (IJRTE), 8*(6), 1490-1495.

[4]. Dai, W., Brisimi, T. S., Adams, W. G., Mela, T., Saligrama, V., & Paschalidis, I. C. (2015). Prediction of hospitalization due to heart diseases by supervised learning methods. *International journal of medical informatics*, *84*(3), 189-197.

[5]. Abdar, M., Kalhori, S. R., Sutikno, T., Subroto, I. M., & Arji, G. (2015). Comparing Performance of Data Mining Algorithms in Prediction of Heart Diseases. *International Journal of Electrical and Computer Engineering (IJECE), 5*(6), 1569-1576.

[6]. Nayak, S., Gourisaria, M. K., Pandey, M., & Rautaray, S. S. (2019). Heart Disease Prediction using Frequent Item Set Mining and Classification Technique. *International Journal of Information Engineering and Electronic Business (IJIEEB), 11*(6), 9-15.

[7]. Latha, C. B. C., & Jeeva, S. C. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Informatics in Medicine Unlocked*, *16*, 100203.

[8]. Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R. (2018). A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms. Mobile Information Systems. *Journal of Machine Learning Research, 11,* 849-872

[9]. Ramalingam, V. V., Dandapath, A., & Raja, M. K. (2018). Heart disease prediction using machine learning techniques: a survey. *International Journal of Engineering & Technology, 7,* 684-687.

[10]. Santhana, K. J., & Geetha, S. (2019). Prediction of Heart Disease using Machine Learning Algorithms. *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*, 1-5.

[11]. Hazra, A., Mandal, S. K., Gupta, A., Mukherjee, A., & Mukherjee, A. (2017). Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review. *Advances in Computational Sciences and Technology, 10*(7), 2137-2159.

[12]. Das, K., & Behera, R. N. (2017). A Survey on Machine Learning: Concept, Algorithms and Applications. *International Journal of Innovative Research in Computer and Communication Engineering, 5*(2), 1301-1309.

[13]. Khemphila, A., & Boonjing, V. (2010). Comparing performances of logistic regression, decision trees, and neural networks for classifying heart disease patients. *International Conference on Computer Information Systems and Industrial Management Applications (CISIM)*, 193-198.

[14]. Manogaran, G., & Lopez, D. (2018). Health data analytics using scalable logistic regression with stochastic gradient descent. *International Journal of Advanced Intelligence Paradigms*, *10*(1-2), 118-132.

[15]. WebMD,Picture of the Heart. Retrieved April 10, 2020from https://www.webmd.com/heart/picture-of-the-heart#1.

[16]. Austin, P. C., & Tu, J. V. (2004). Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *Journal of Clinical Epidemiology, 57*(11), 1138-1146.

[17]. Cheung, A. K., Sarnak, M.J., Yan, G., Dwyer, J. T. Heyka, R. J, Rocco, M. V., Teehan, B. P. & Levey, A. S. (2000). Atherosclerotic cardiovascular disease risks in chronic hemodialysis patients. *Kidney International, 58*(1), 353-362.

[18]. Ben-Haim, Y., & Tom-Tov, E. (2010). A Streaming Parallel Decision Tree Algorithm. *Journal of Machine Learning Research, 11*, 849-872.

[19]. Krishnaiah, V., Narsimha, G., & Chandra, N. S. (2013). Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques. *International Journal of Computer Science and Information Technology*, *4*, 39-45.

[20]. Bayat, S., Cuggia, M., Rossille, D., Kessler, M., & Frimat, L. (2009). Comparison of Bayesian Network and Decision Tree Methods for Predicting Access to the Renal Transplant Waiting List. *Medical Informatics in a United and Healthy Europe, 150*, 600-604

[21]. Brijain, M., Patel, R., Kushik, M., & Rana, K. classification. *International Journal of Engineering Development and Research, 2*(1).
[22]. Vijararani, S., & Dhayanand, S. (2015). Data Mining Classification Algorithms for Kidney Disease Prediction. *International Journal on Cybernetics & Informatics, 4*(4), 13-25.
[23]. Subbalakshmi, G., Ramesh, K., & Rao, C. (2011). Decision Support in Heart Disease Prediction System using Naïve Bayes. *Indian Journal of Computer Science and Engineering, 2*(2), 170-176.
[24]. Saboji, R. G. (2017). A Scalable solution for heart disease prediction using classification mining technique, International Conference on Energy, Communication. *Data Analytics and Soft Computing,* 1780-1785.
[25]. Shinde, R., Arju, S., Patil, P., & Waghmare, J. (2015). An Intellient Heart Disease Prediction System using K-Means Clustering and Naïve Bayes Algorithm. *International Journal of Computer Science and Information Technologies, 6*(1), 637-639.
[26]. Zhang, H. (2004). The Optimality of Naïve Bayes. Proceedings of the *Seventeenth International Florida Artificial Intelligence Research Society Conference*.
[27]. Saritas, M. M., & Yasar, A. (2019). Performance Analysis of ANN and Naïve Bayes Classification Algorithm for Data Classification. *International Journal of Intelligent Systems and Applications in Engineering, 7*(2), 88-91.
[28]. Casanova, R., Saldana, S., Chew, E. Y., Danis, R. P., Greven, C. M., & Ambrosius, W. T. (2014). Application of random forests methods to diabetic retinopathy classification analyses. *PLOS one*.
[29]. Masetic, Z., & Subasi, A. (2016). Congestive heart failure detection using random forest classifier. *Computer Methods and Programs in Biomedicine, 130*, 54-64.
[30]. Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M., (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines, *71*, 804-818.
[31]. Yu, W., Liu, T., Valdez, R., Gwinn, M., & Khoury, M. J. (2010). Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC medical informatics and decision making*, *10*(1), 1-7.
[32]. Polat, K., Gunes, S., & Arslan, A. (2008). A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine, Expert Systems with Applications, *34*(1), 482-487.
[33]. Ghumbre, S., Patil, C., & Ghatol, A. (2011). Heart Disease Diagnosis using Support Vector Machine. (2014). A survey on decision tree algorithm for *International Conference on Computer Science and Information Technology,* 84-88.
[34]. Moavenian, M., & Khorrami, H. (2010). A qualitative comparison of Artificial Neural Networks and Support Vector Machines in ECG arrhythmias classification. *Expert Systems with Applications, 37*(4), 3088-3093
[35]. Jabbar, M. A. Deekshatulu, B. L., & Chandra P. (2013). Classification of Heart Disease Using K-Nearest Neighbor and Genetic Algorithm. *International Conference of Computational Intelligence: Modeling Techniques and Applications*, 85-94.
[36]. Saini, I., Singh, D., & Khosla, A. (2013). QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases. *Journal of Advanced Research, 4*(4), 331-344.
[37]. Odom, M. D. & Sharda, R. (1990). A neural network model for bankruptcy prediction. *International Joint Conference on Neural Networks, 2*, 163-168.
[38]. Sharma, R., Das, S., Gourisaria, M. K. Rautaray, S. S., & Pandey, M. (2020). A Model for Prediction of Paddy Crop Disease Using CNN, Progress in Computing, Analytics and Networking. *Advances in Intelligent Systems and Computing*, 533-543.
[39]. Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction. *International Journal of Computer Applications, 17*(8), 43-48.
[40]. Chandna, D. (2014). Diagnosis of Heart Disease Using Data Mining Algorithm. *International Journal of Computer Science and Information Technologies, 5*(2), 1678-1680.
[41]. UCI Machine Learning Repository, Heart Disease DataSet, Retrieved March 27, 2020 from http://archive.ics.uci.edu/ml/datasets/Heart+Disease.
[42]. Nayak, S., Gourisaria, M. K., Pandey, M., & Rautaray, S. S. (2019). Comparative Analysis of Heart Disease Classification Algorithms Using Big Data Analytical Tool. *In International Conference on Computer Networks and Inventive Communication Technologies,* 582-588.
[43]. Geeitha, S. & Thangamani, M. (2020). A Cognizant Study of Machine Learning in Predicting Cervical Cancer at various Levels - A Data mining concept. *International Journal on Emerging Technologies (IJET), 11*(1), 23-28.
[44]. Gourisaria, M. K., Das, S., Sharma, R., Rautaray, S. S., & Pandey, M. (2020). A Deep Learning Model for Malaria Disease Detection and Analysis using Deep Convolutional Neural Networks. *International Journal on Emerging Technologies (IJET), 11*(2), 699-704.

How to cite this article: Das, S., Sharma, R., Gourisaria, M. K., Rautaray, S. S. and Pandey, M. (2020). Heart Disease Detection using Core Machine Learning and Deep Learning Techniques: A Comparative Study. *International Journal on Emerging Technologies*, *11*(3): 531–538.