Improved Heart Disease Prediction Using Deep Neural Network

Mohd Ashraf[1], M. A. Rizvi[2] and Himanshu Sharma[3]
[1]Department of Computer Science & Engineering, Maulana Azad National Urdu University, Hyderabad, Telangana, India
[2&3]Department of Computer Engineering and Application,
National Institute of Technical Teachers Training and Research Institute, Bhopal, Madhya Pradesh, India
E-Mail: ashraf.saifee@gmail.com, marizvi@nitttrbpl.ac.in, hs13867@gmail.com

Abstract -Heart disease is biggest challenge for medical professionals. Modern life style made it an epidemic; according to a survey conducted by WHO heart attack is leading cause of death all over the world especially in the western world. It is surveyed that 23% of the death in US is due to Heart related disease [1]. It has been observed assistance is needed for helping medical professionals in detecting the chance of heart attack in the human. In recent times a lot of work related to providing an automated support system for predicting chance of Heart attack in human has been done. After advancement of computer science, researchers felt that they can help in some of the key interdisciplinary areas like medical science. Machine learning techniques are compared on the single data set which does not reflect true potential of any algorithms. They also suffer from some of the key anomalies such as accuracy and manual data set pre-processing. In this paper, we propose Deep Neural Network methods for creating an automated system for heart attack prediction. It is tested on multiple dataset to find out true potential and providing certainty in the accuracy. Method also promises to remove all the mentioned anomalies from the system like lack of accuracy and automated approach in pre- processing of the data set. In result analysis, it has been observed that prediction is much more efficient and minimum accuracy achieved through this proposed method is 87.64% on any of the data set taken under consideration.
Keywords: Deep Neural Network, Heart Attack, Machine Learning

I. INTRODUCTION

Medical science has to deal with verity of the challenges, related to diagnose of disease. Professional in this field need accurate support system for decision making related to diagnose of disease many approach has been done in this regard[2] , it has been found that most proficient doctor can predicate disease up to an accuracy of 67% leaving behind a lot of question because failure rate is too much and human life is at state. This gives birth to the support system for doctor relevant to the disease, some of the pioneer works in this regard has been discussed in following work.

According to American Heart Association heart attack can occur when blood flow to a part of the heart is blocked by a blood clot. Heart disease is an umbrella term coined for any type of heart disorder that affects the heart. Heart disease means cardiac disease [2].Heart disease is an acute disease which need proper diagnose and analysis of patient data. It is not possible for a human brain to analysis so much of data

obvious solution is support system for doctor working in the field of heart disease. Earlier time decision support system is based predicate logic[1] and first expert system called as Dendril is used to find blood related disease in human. Later with advancement of computer processing power, machine learning came into existence[3] which does not need reprogram itself machine learning put its all focus on dataset and its manipulation. All major algorithms like SVM, Naïve Bayes, Decision Tree neural network is used to building support system for predicting chance of heart attack in human. Each method has its own merits and demerits. It has observed after literature survey there is slope of art of work, regarding much aspect such as lack of accuracy, lack of data independency, better generalized scope and automated pre-processing stage.

Deep learning is an approach where multiple hidden layers are used to reveal the secret inside the data. Deep learning is used for predication of other disease[4]. According to our knowledge it is first time heart attack is predicted using deep neural network .approach use fully automated data pre-processing stage and since it is based on deep neural network , it is obvious accuracy achieved would much higher than any other machine learning technique for evaluation of this method multiple data set has been taken . A broad overview has been done in the result analysis section it is discussed the obtained result and compare them with all other machine learning algorithm methods. In the coming section give overview of heart disease and deep learning. In the proposed methodology section broad discussion is done, where almost all the component along with their inter flowing is explained. Purpose of this paper[3]is to highlight effect of deep neural network on predicting heart attack, So that more deep learning method could be used for heart attack, because it has bright future ahead. It can be relevance with the output obtained with Deep Neural Network which is discussed in result and analysis section.

II. DEEP LEARNING

Deep learning uses artificial neural network having more hidden layer. Neural Network is arranged in the cascading pattern to process non- linear data set, i.e. cascading layers of non-linear processing units.
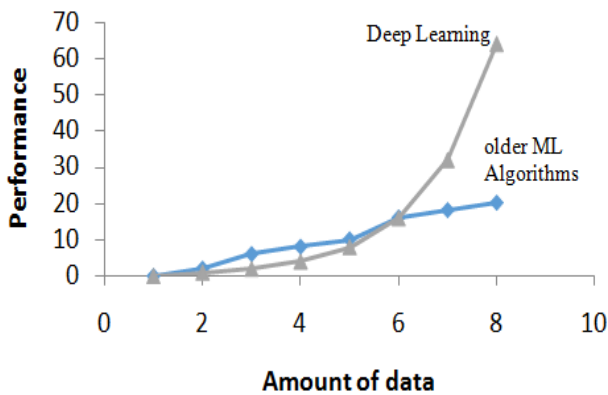
Fig. 1 Comparative graph explaining about nature of Deep Learning algorithms and older Machine Learning Algorithms on the parameter of number of data set

Deep Learning was developed with the idea of enhancing accuracy of the existing machine learning algorithm. It can be seen through the figure 1, in initial phase when data set is small older algorithm underperform the Deep learning techniques but with increase in size of data set , Deep Learning method started gaining momentum and accuracy is get significant edge over all other older machine learning algorithms.

Inception of Deep Learning method starts with origin of recurrent neural network in the year 1973, recurrent network works well with the temporal data where input data set and output dataset is not fixed[5]. It is suitable for the field of language processing and NLP data set. Due to implementation complexities need for the Deep learning method is felt which is fulfilled by the convolutional Network was proposed in year 1978, it works on the fixed data set with fixed output. It is obvious that this method best suitable for the image processing [6].

It is based on the fundamental of breaking complex global weight matrix into change of convolution kernel. Gaussian kernel is used to smooth the data set, edge can be seen with the help of canny kernel and to obtain gradient feature Gabor kernel is used. It is clear that single kernel is not used in the implementation of Convolutional network instead of that it follows the data specific kernel. This model is designed for the very specific task, all major image recognition model are based on it. Other than this application it is hardly applicable for any other application. Auto encoder is simple technique based on the simple fundamental of feedback effect. Model is called encoding because of its architecture which is explained in the fig.2.

Input layer (visible layer) is connected to hidden layer where number of neurons is much less than the visible layer but each neuron is connected backward to its input neuron, due to less neuron model follows encoding mechanism. Finally at the outer layer decoding is done to preserve all the content of the input layer, keeping this goal in this mind number of the neuron is equivalent to the outer layer.
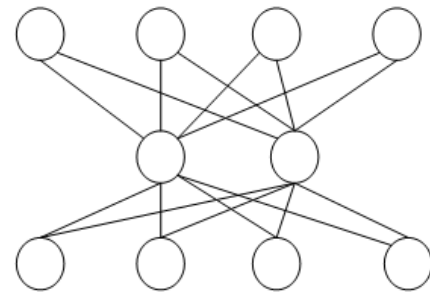

Fig. 2 An overview of the architecture for Auto-Encoder model

Other than these models there exists numerous other variations of the algorithms such as LSTM (Long short term memory) which is consider to be better than Recurrent neural network because it can remember information to long time, this due to the reason LSTM uses special memory unit to remember other than standard memory unit[7]. A set of memory gate is used for controlling entered information, output and forgotten information.

On the other hand RNN loss information due to gradient of the loss function decay exponentially with the time also called as vanishing gradient problem.

LSTM is much more complex architecture due to this a simpler version of LSTM has been proposed called as Gated Neural unit. It uses set of gates to control flow of information and also instead of using separate memory cell, it uses fewer gates.

III. DEEP NEURAL NETWORK

Deep neural network uses stacked neural network. Network with several layers and each layer having several neurons. A node can combine weight with capacity to magnify the value of input provided to it and this sum is passed through activation function. Function divide up to what extent signal will carry further. In deep neural network each layer can switch between on or off where output of one layer act as input for layer in forward direction. Deep neural network varies from other neural network in number of hidden layer whereas artificial neural network consists of one input and one output layer and maximum one hidden layer but deep neural network must have more than one hidden layer [8].

Each neuron in the network got trained on the different set of features which is provided by previous layer output. Increasing the number of level in the neural network directly affect capability of processing feature. It mean it can process recognises more complex features.

This is called feature hierarchy where next layer combine and present most abstract and complex output which possess deep neural network capable-ary to operate on high dimensional non-linear dataset and perform automatic linear feature extraction [9].

IV. PROPOSPED METHODLGY

Basically methodology uses the Deep Neural network fundamental for achieving output for any organisation in deep learning or machine learning data set play an important role.

A. Data Set: Data set used for implementation of proposed work is taken from UCI repository (https://archive.ics.uci. Edu/ml/datasets/Heart Disease). Nature of the data set is defined by its dimensionality. It consists of 13 columns which represents features of the data set. A complete overview has been shown in the figure given below as like other data set it may consist of certain redundancy, noise and missing values for talking case of it proper steps need to be taken. In next subsection pre-processing is explained in brief. Attributes of dataset need to be taken care of and their respective value need to standardize.

Fig. 3 Cleveland dataset
(https://archive.ics.uci.edu/ml/datasets/Heart+Disease)

Attributes are explained in the table I.

TABLE I DATASET DESCRIPTION

Attribute	Data type	Description
age	Continuous	Years
sex	Binary	1:male, 0:female
Chest Pain	1 -4	1:typical angina, 2:atypicalangina,3:non-anginal pain, 4:asymptomatic
trestbps	Continuous	Resting blood pressure
Chol	Binary	Serum cholesterol
Fbs	Binary	fasting blood sugar
Resteeg	0 - 2	Resting electrocardiographic results,0:normal, 1:having ST-T wave abnormality,2:showing probable or definite leftventricular hypertrophy by Estes' criteria
Thalach	Continuous	Maximum heart rate achieved
Exang	Binary	Exercise induced angina
Old peak	Continuous	ST depression induced by exercise relativeto rest
Slop	1-3	1:upsloping, 2:flat, 3:downsloping
Ca	Continuous	Number of Major Vessels
num	0 - 4	Class Target the predicted attribute

B. Data Pre-Processing: Every dataset consist of various types of anomalies such as missing values, redundancy or any other problem for removing this problem there is need of certain step called as processing data [10]. Pre-processing step is needed to overcome from such problem. There are three pre-processing steps:

1. *Formatting:* The data set is used for implementation is taken from UCI repository, it may contain certain attributes whose names are not clear in the (dataset name) also contain certain unrelated attribute which is not useful for the greater performance of proposed work . An attribute name as "Thal" has been removed from dataset by using following command in R, Dataset$Thal<-Null

2. *Cleaning:* This part of pre-processing belongs to remove or fixing of missing out entry in the data frame.

Row containing these incomplete columned to be removed also for removing certain redundant entries in data frame this step is recommend

3. *Sampling:* Sampling is also done on the dataset to enhance the performance of the algorithm on sample data set may lead algorithm to take longer time

C. Transforming Data Set: This is the final step in the processing the data, this step is problem specific related to the data frame .Generally these three data transformation tools are used: Scaling, decomposition, aggregation. For this implementation scaling is used to convert the data into the new vectors because different attribute uses different unit to normalize each attribute on a single uniform unit [11].

This step need to be done for this implementation following code is used for scaling.

```
Set.seed(1)

X<-runif(7)

(x-meanx))/sd(x)

Scale(x)
```

After pre-processing stage, it is time for application of data set on the deep neural network. DNN is deep learning algorithm, it is based on the artificial neural model but number of hidden layer is more than one. At the input layer the data set is feed number of neurons at the outer layer is always equal to the number of feature in the dataset. In according with data set taken number of neuron s would be 13. Each neuron is connected to the inner hidden layer with 6 neuron at the hidden layer1 and 4 neurons at hidden layer 2. Each hidden layer carries extracted feature information to the forward layer for better feature extracted feature information to the forward layer for better feature extraction at the outer layer the output is generated in form of refined feature properties. Purpose of hidden layer is to process non-linear data transformation into feature vector, interpretation of such vector by a human brain is difficult task

D. Architecture: In the building block of propose work is defined in the six major section each of them carries its own functionalities.
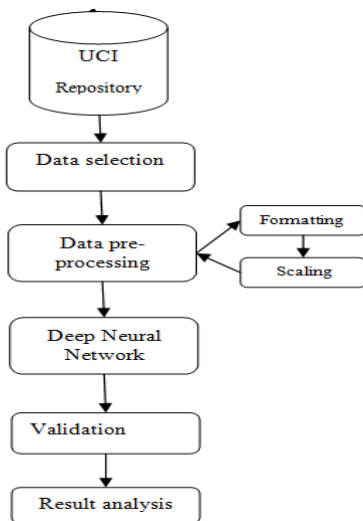

Fig. 4 Architecture

1. *Data Selection:* Data is collected from UCI repository the name of dataset is Cleveland. Discussion on type dataset used for implementation is already being done earlier in this section.

2. *Data Pre-processing:* Data pre-processing for the proposed work is dividing into subsection formatting and scaling. It is also being discussed in previous section.

3. *Deep Neural Network (DNN):* DNN is deep learning mechanism which is used for getting high accuracy for the purpose of predicting heart attack in the patient. Purpose of selecting DNN is to achieve accurate result regarding the predication. Manipulation and classification of dataset can be easily handled by DNN because of its complex nature[12]. Proposed work has used two hidden layer for the purpose of extracting quality feature from the dataset. Working of DNN is explained in great detail in the coming implementation section.

4. *Validation:* Validation is required for every neuron based method seen in earlier phase of epochs there exist a wide gap between actual and predicted result ,for correcting this error continuous effort is needed which can achieved with the help of validation set.

5. *Result:* All the obtained result based on different parameter such as accuracy , execution time and mean squared error is utilize for evaluation of proposed method with the existing work done . The discussion on this topic in coming section of result analysis.

V. IMPLEMENTATION

It is very important to prove the purpose work through the help of some existing coding platform. This work uses R for converting the idea of purposed work into realistic framework. R is a smart platform for implementation for special related to pattern recognition. So it can used for classification of medical dataset, earlier in literature review is examined that quite of few research on medical dataset has been evaluated with this intelligent platform. It is very rich in machine learning and deep learning library with very simple method for coding.

The proposed work has been divided into different module where each module is very independent of one another but output generate by earlier module act as input of later module. Some module such as data pre-processing May consists of more sub module such as formatting and scaling .implementation of each module will be discussed separately with help of suitable codes. Data collection is done manually from the website from UCI repository .dataset may consist of several anomalies which has been already described in the previous sections. To remove this anomalies data type processing has been divided into sub section in the formatting section those attributes has been taking care which is either not useful or may act as performance degrader for the purposed work due to this attribute "thal" has been removed from the dataset by allocating null value to each cell of the column.

Cleveland$Thal<-Null

The next subsection deals with removal of missing value and redundancy from the dataset. These section is called as cleaning of dataset it can be achieved by the help of removal

of those entry from the dataset .finally for transforming the dataset because different attribute use different unit to normalise every attribute on same parameter scaling has been done the following formula is used for normalizing each attributes.

```
Set.seed(1)
X<-runif(7)
(attribute-mean)/sd
Scale(attribute)
```

The core of implementation is based on correct deployment of deep neural network in the R for these purpose different packages in R is responsible which take the purified dataset and produce the correct output several packages needed for correct deployment of the module R MASS, CARET and NEURAL NET, each of them work in coordination to produce the output. Modules also takes the help of two other are R package name as GGPLOT, LATTICE to representing the output in a graphical manner. Functionality of each module is necessary to understand the working of purposed work. First of all CARET(classification and regression training) package is used for data preparation which include function like reducing skewness for removing co related predicator and centralizing data for the purpose of making the dataset ready for the package MASS where the package MASS prepare the logistic regression model of the graph after preparing the logistic regression model of the dataset the output is fade to the neural net package, it is used for training for neural network, NEURALNET package can supported multiple hidden layer with backward and forward propagation. These packages produce the output after proper training of itself by the dataset. For plotting the output produce by the neural network can be plotted by the help of two packages LATTICE and GGPLOT where lattice can display the relationship between multivariate data and GGPLOT can produce the line graph. Validation is one of the key stages of producing the correct output because it retrains the model until the most optimal results could be achieved.

VI. RESULT ANALYSIS

In the deep neural network their great importance of the hidden layers and the weight allocated to them. In accordance to the figure 5 it can be seen that different weight parameters is allocated to each neurons with the different activation function value at each level. Activation functions and weight value both them are adjusted during training of the model to provide most optimal value at the output. A keen analysis on the output has been done for evaluating the proposed work on each parameter especially related to the accuracy and mean square error. Before evaluation of result it is carefully analysed about dataset for that skewness of dataset is evaluated. Evaluation in the dataset is done to know the biasness of the dataset. According to biasness of data further processing of the model has been done like predicating the optimal value for

epochs, weight and activation function value, biasness of the considered data is shown in the fig.6.
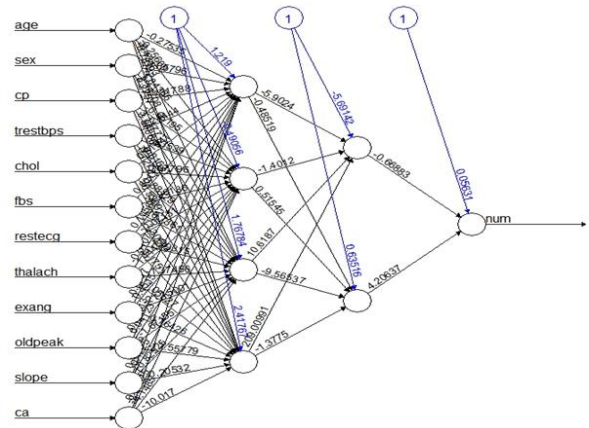

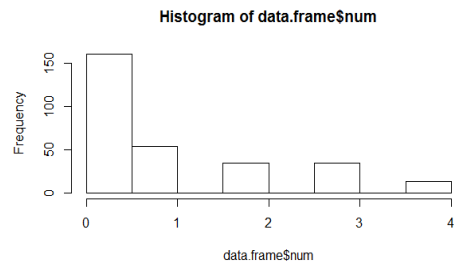Fig. 5 Deep Neural Network with weights allocated to each neuron and Activation Function at each level

Fig. 6 Represents skewness in the data set

A. Evaluation of Mean Square Error and Accuracy: Mean square error is defined as array deviation of present value and estimated value. It is obtained in the present experimental setup that in the initial phase of the output when training is done on less number of instances, mean square error achieved is very low, even on increasing the number of epochs the accuracy did not cross the value of the 0.36.
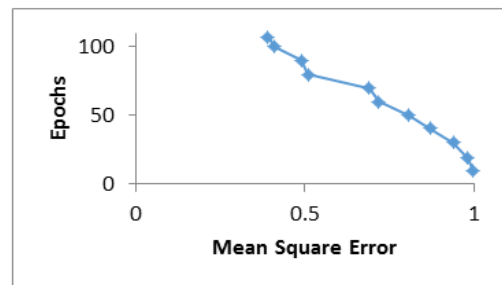

Fig. 7 Fixed error rate after certain epochs keeping number of Instance Constant

In the fig.7,it can be seen that after 107 epochs mean square error remains constant keeping number of instance increased to 298 then the mean square error is reduced to 0.15 with the number of epochs equal to the 1237 after increasing number of epochs, value of error remain the same. In the final evaluation of the experiment entire dataset with 512 rows (instances) is feed to the network after 1717 epochs mean square error reduced to the value of 0.0567

with certain variation in the weight factor of neuron and activation function at each level the error value reduced to the level of 0.0413. Modulation of weight factor is kept around the value 1.217 to -5.87 from input layer to the hidden.
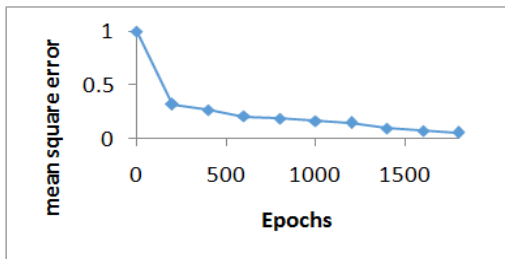

Fig. 8 Mean square error on entire data set

Layer, between two hidden layer the weigh is in the range of -5.804 to -1.379 and the final layer one weight factor is -0.65 and another is 4.207. activation function values at level one different for each neuron which is in the range of 1 to 2.5, at level two value are-5.69142 and 0.635 and final layer it 0.631.

B. Comparison with Other Algorithms: Deep neural network is most feasible algorithm when compared with other algorithms of machine learning. In the fig. 8, it can be easily evaluated the accuracy achieved by the deep neural network is for ahead of its entire contemporary machine learning algorithms. Comparison is done using accuracy as a parameter with both supervised and on supervised learning algorithms. It could be seen from the figure 8 only algorithm whose accuracy is near to the deep neural network is the random forest [13]. Other algorithm such as SVM, Naïve Bayes and ANN does not show promising results. It is interesting to observe that K-mean clustering algorithm also show accuracy better than excepted.
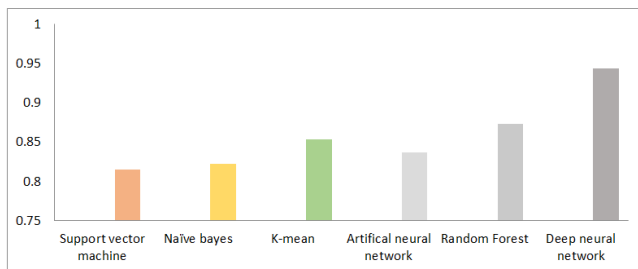

Fig. 9 Comparisons between different algorithms in term of accuracy

VII. CONCLUSION AND FUTURE WORK

Heart attack prediction with help of machine learning and deep learning algorithm has a bright feature ahead recently lot of research work has been diverted to the medical dataset. A lot has been concurred and a lot is need to be done in this area this work is dedicated to improve accuracy with the help of deep learning algorithm keeping in mind very few work has been dedicated to evaluate heart attack prediction using deep learning method this work can opened door of opportunity for further improvement accuracy by

using various other deep learning algorithm. This paper explained every key point related to the work along with basic terminology and concept which is necessary to be discussed. A complete architecture model has been explained along with minute details of each component of architecture. It is very much necessary to know about working of any architecture due to this work explained about inter related working explanation. Implementation detail along with the flow of code and package required for execution has been discussed in this section. result analysis is one major area which need board attention, hence detail of every result obtained based mean square error and accuracy is plotted graphically, to express the achievement obtained by the purposed work a comparison is required. Comparison has been done with various other machine learning algorithm which is earlier used for predicting heart attack which clearly shows super messy of deep neural network over all other algorithm. As discussed earlier there is wide scope for achieving better result in terms of accuracy by using other Deep learning algorithm. In future, this work will extended to other deep learning algorithm for achieving better result regarding accuracy.

REFRENCES

[1] E. O. Olaniyi, O. K. Oyedotun, and K. Adnan, "Heart Diseases Diagnosis Using Neural Networks Arbitration",*Int. J. Intell. Syst. Appl.*, Vol. 7, No. 12, pp. 75–82, 2015.

[2] G. Parthiban, "Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients",*International Journal of Applied Information Systems (IJAIS)*Vol. 3, No. 7, pp. 25–30, 2012.

[3] Yuming Hua, Junhai Guo, and Hua Zhao, "Deep Belief Networks and deep learning",*Proc. 2015 Int. Conf. Intell. Comput. Internet Things*, pp. 1–4.

[4] J. Schmidhuber, "Deep Learning in neural networks: An overview", *Neural Networks*, Vol. 61, pp. 85-117, Jan 2015.

[5] T. A. Lasko, J. C. Denny, and M. A. Levy, "Computational Phenotype Discovery Using Unsupervised Feature Learning over Noisy, Sparse, and Irregular Clinical Data",*PLoS One*, Vol. 8, No. 6, 2013.

[6] P. De, "Modified Random Forest Approach for Resource Allocation in 5G Network",*Int. J. Adv. Comput. Sci. Appl.,*Vol. 7, No. 11, pp. 405–413, 2016.

[7] S. Hochreiter and J. UrgenSchmidhuber, "Long Short-Term Memory",*Neural Compuational.*, Vol. 9, No. 8, pp. 1735–1780, 1997.

[8] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques",*2008 IEEE/ACS Int. Conf. Comput. Syst. Appl.,* pp. 108–115, 2008.

[9] P. N. Druzhkov and V. D. Kustikova, "A survey of deep learning methods and software tools for image classification and object detection",*Pattern Recognit. Image Anal.*, Vol. 26, No. 1, pp. 9–15, 2016.

[10] M. Sultana, A. Haider, and M. S. Uddin, "Analysis of data mining techniques for heart disease prediction", *3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, 2016 *The IEEE website*. [Online] Available: http://www.ieee.org/

[11] D. K. Srivastava and L. Bhambhu, "Data classification using support vector machine",*J. Theor. Appl. Inf. Technol.*, pp. 1-6,2009.

[12] M. A. Jabbar, P. Chandra, and B. L. Deekshatulu, "Prediction of risk score for heart disease using associative classification and hybrid feature subset selection",*Int. Conf. Intell. Syst. Des. Appl. ISDA,* pp. 628–634, 2012.

[13] Bo Pang, Lillian Lee and ShivkumarVaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques", *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, Vol. 10, pp. 79-86, 2002.