International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 04 | Apr-2018 www.irjet.net p-ISSN: 2395-0072

Heart Attack Prediction Using Deep Learning

Abhay Kishore[1], Ajay Kumar[2], Karan Singh[3], Maninder Punia[4], Yogita Hambir[5]

[1,2,3,4] *Department of Computer Engineering, Army Institute of Technology, Pune, Maharashtra*
[5] *Professor, Department of Computer Engineering, Army Institute of Technology, Pune, Maharashtra*

Abstract - *Cardiovascular disease is one of the most heinous disease, especially the silent heart attack, which attacks a person so abruptly that there's no time to get it treated and such disease is very difficult to be diagnosed. The lack of specialist doctors and increase in wrong diagnosed cases has necessitated the need for building an efficient heart disease detection system. Various medical data mining and machine learning techniques are being implemented to extract the valuable information regarding the heart disease prediction. Yet, the accuracy of the desired results are not satisfactory. This paper proposes a heart attack prediction system using Deep learning techniques, specifically Recurrent Neural Network to predict the likely possibilities of heart related diseases of the patient. Recurrent Neural Network is a very powerful classification algorithm that makes use of Deep Learning approach in Artificial Neural Network. . The paper discusses in detail the major modules of the system along with the related theory. The proposed model incorporates deep learning and data mining to provide the accurate results with minimum errors. This paper provides a direction and precedent for the development of a new breed of heart attack prediction platform.*

Keywords: Heart Attack Prediction System, Data Mining, Artificial Neural Network, Recurrent Neural Network, Gated Recurrent Unit.

1. INTRODUCTION

Cardiovascular diseases are one of the highest flying diseases of the modern world. According to a survey, about more than 17.7 million deaths occur all across the world annually due to heart diseases[1]. Of these deaths, an estimated 7.4 million were due to coronary heart disease and 6.7 million were due to stroke[2]. Heart Attacks are one of the most deadly diseases which can knock one down at any point of time without any invitation and silent heart attacks are something which most doctors are not able to predict. The lack of specialists and increasing wrong diagnosed cases have necessitated the need for building an efficient cardiovascular disease prediction system. This has led to research and development of new medical data mining techniques and various machine learning techniques. The main objective of this work is to identify the key patterns and features from the medical data using the classification algorithms and then to select the most relevant attributes for silent heart attack diagnosis. The use of Recurrent Neural Network will further enhance the accuracy of the results. While the implementation of such a system is not unprecedented, the existing systems have drawbacks and do not aim at finding out the possibilities of silent heart attacks. This paper aims to address these and propose implementation of innovative features to develop a more comprehensive system.

2. LITERATURE REVIEW

Author	Purpose	Techniques used	Tool	Accuracy
M.A. Nishara Banu, B Gomathy	This system is used to predict the heart attack and also discussed various uses of various data mining algorithm for disease prediction.	C4.5 MAFIA K-Means clustering	Weka	89%
Aqueel Ahmed Shaikh Abdul Hannan	classification techniques in data mining and performance of classification among them	SVM Decision tree	R tool	91%
Rashedur M Rahman Farhana Afroz	Comparison of differernt data mining for diabetes diagnosis	Neural Network	Weka	85.83%
Nidhi Batla Kiran Jyoti	Analysis of Heart Disease Prediction using Different Data Mining Techniques	Naïve Bayes Neural Network	Weka 3.6.6	96%

Table 1 - Various ML techniques used for heart disease prediction

3. PRESENT SYSTEM

Present systems used for prediction of heart attack are failing to meet the desired accuracy in the results. As seen in the literature survey, the machine learning techniques used are pushing the accuracy till a certain limit. Moreover, the issue with the present heart attack prediction system is the uses of attributes. The attributes to be selected for the prediction of heart attack are the conventional ones and thus the results are generating wrong results many-a-times. The proposed model aims to extract the proper attributes from the datasets which will enhance the precision of the prediction. It will also provide the users with proper diagnosis so that the user understands the problem well without much difficulty.

4. PROPOSED HEART DISEASE PREDICTION SYSTEM

The project sets itself apart by harnessing the powers of both Deep learning and data mining. The paper proposes a system, with a strong prediction algorithm, which implements powerful classification steps with a comprehensive report generation module. The project aims to implement a self-learning protocol such that the past inputs of the disease outcomes determine the future possibilities of the heart disease to a particular user. The proposed model makes use of strong preprocessing tools so that the classification and prediction do not show any errors relating to the dataset. A huge no. of training sets will be used to make the prediction more and more accurate. Not only does the datasets but also the attributes to be used are selected taking into consideration the various important parameters and attributes.

Implementing the afore mentioned model meets the goal of developing a system with increased accuracy of estimating the patient getting heart attack percentage while addressing the drawbacks of the existing system

The model which is proposed for Heart Attack Prediction System is created using Deep learning algorithms and approach. The proposed Flowchart for work is shown in Fig.1

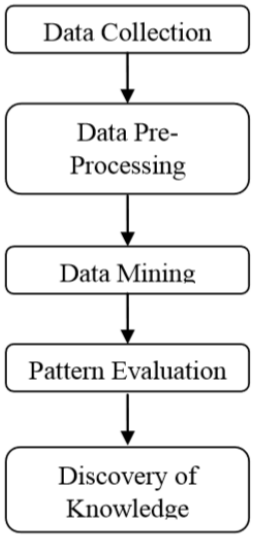

Fig 1 - Flowchart of proposed work

4.1 Data Source

As we know dataset is a collection of data objects which are having various numbers of attributes that deal with the basic characteristics of an object. We obtained 303 records with 75 medical attributes (factors) from the UCI Machine Learning Data Repository [3]. After preprocessing use of 270 records with 13 medical attributes. Out of these 13 attributes, 7 have discrete values whereas 6 have continuous values.

Name	Type	Description
Age	Continuous	Age Age in years
Sex	Discrete	0 = female 1 = male
Cp	Discrete	Chest pain type: 1 = typical angina, 2 = atypical angina, 3 = non-anginal pain 4 =asymptom
Trestbps	Continuous	Resting blood pressure (in mm Hg)
Chol	Continuous	Serum cholesterol in mg/dl
Fbs	Discrete	Fasting blood sugar>120 mg/dl: 1-true 0=False
Exang Continuous Maximum heart rate achieved	Discrete	Exercise induced angina: 1 = Yes 0 = No
Thalach	Continuous	Maximum heart rate achieved
Old peak ST	Continuous	Depression induced by exercise relative to rest
Slope	Discrete	The slope of the peak exercise segment : 1 = up sloping 2 = flat 3 = down sloping
Ca	Continuous	Number of major vessels colored by fluoroscopy that ranged between 0 and 3.
Thal	Discrete	3 = normal 6 = fixed defect 7= reversible defect
Class	Discrete	Diagnosis classes: 0 = No Presence 1=Least likely to have heart disease 2= >1 3= >2 4=More likely have heart disease

Table 2 - Selected Heart Disease Attributes

4.2 Data Mining

4.2.1 Recurrent Neural Network

This deep learning model has a simple structure with a built-in feedback loop allowing it to act as a forecasting engine. RNN in essence is a regular neural network with an additional hidden state where the hidden state influences the neural network output. The hidden state is updated on each input step. It is a model which can not only learn local and temporal dependencies in data but also can accommodate variable sequence lengths.

In RNN the output of a layer is added to the next input layer and fed back into the same layer. Unlike feed-forward networks RNN can receive sequence of values as inputs and it can produce sequence of values as outputs. But in RNN problem of vanishing gradient is worse therefore it leads to exponentially small gradients and decay of information through time.

There are several ways to address this problem the most popular of which is gating. The most popular of gating types are LSTM and GRU with GRU being used in the paper.

4.2.2 Gated Recurrent Unit

It is an improved version of the standard recurrent unit. It has two gates:

i) Update Gate: determines how much of the past date needs to be passed to the future

ii) Reset Gate: determines how much of the past data to forget

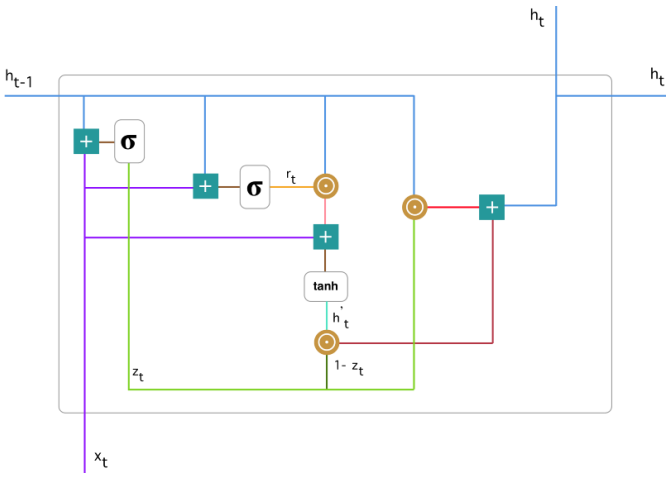

Fig 2 - Gated Recurrent Unit[4]

i) Update Gate

$$z_t = \sigma (W^{(z)}x_t + U^{(z)}h_{t-1}) \quad \ldots (1)$$

ii) Reset Gate

$$r_t = \sigma (W^{(r)}x_t + U^{(r)}h_{t-1}) \quad \ldots (2)$$

iii) Current Memory Content

$$h_t = \tanh (Wx_t + r_t \odot Uh_{t-1}) \quad \ldots (3)$$

iv) Final Memory at Current Time Step

$$h_t = z_t \odot h_{t-1} + (1-z_t) \odot h_t \quad \ldots (4)$$

where,

z_t = update gate

$W^{(z)}$ = weight of x_t

h_{t-1} = holds information of previous states

$U^{(z)}$ = weight of h_{t-1}

σ = sigmoid function

r_t = reset gate

Sigmoid layer is present at forget layer which determines which information to keep and which to discard. A value of 0 means forget everything, value of 1 means retain everything and value of 0.75 means keep exactly 0.75 percent of current data. Since the gates allow the network to forget the states that are no longer needed the computational load of the recurrent network decreases.

4.3 GRU based RNN for heart attack prediction

The preprocessed data is given as input at the input layer. The word embedding layer converts the input data into dense vector representation. With the help of learning GRU layer the vector values are processed. The softmax activation function present at fully connected layer helps in classifying into heart disease or non-heart disease patient.

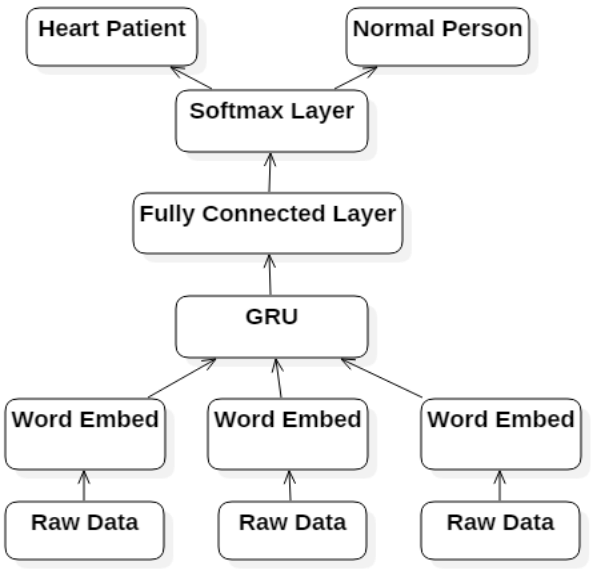

Fig 3 - Design of GRU based RNN for heart attack prediction

4.4 Experimental Result

RNN is implemented in python and the optimization using Theano, a library to perform operations on a GPU. Input dimensions were 13 for each attribute and 100 hidden layers were chosen for the RNN. 150 records were used for training set and 120 records were used for testing set. The model was trained for 100 epochs and batch size as 10 and adadelta was used as the optimization algorithm. At each epoch, "Validation AUC" will be calculated using the validation set, and if it is the best "Validation AUC" so far, the test set will be used to calculate "Test AUC". The model with the best "Test AUC" was saved at the end of the training which came to be 92%.

Accuracy of the model = 92%

5. PERFORMANCE ANALYSIS COMPARISON CHART

TECHNIQUE USED	ACCURACY
CNN & Decision Tree[5]	82%
SVM[6]	84.12%
Naïve Bayes[7]	86.53%
KSOM[8]	88.9%
C4.5 MAFIA K-means Cluster[9]	89%
DBN[10]	90%
RNN	92%

Table 3 - Performance Analysis Comparison Chart

6. CONCLUSION

As we have been through number of projects and their papers and found different algorithms had different accuracy starting from ML algorithms to deep learning algorithms accuracy kept on increasing but we were not able to obtain good results for silent heart attack prediction. Hence after analysis we thought of using RNN and GRU to make the system more accurate and efficient to predict the silent heart attacks and inform the user at the earliest possible. This system has increased the heart attack prediction accuracy to 92% and has proved to be an excellent source in predicting silent heart attacks.

REFERENCES

[1] http://www.who.int/mediacentre/factsheets/fs317/en/

[2] Ms. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte, "A data mining approach for prediction of heart disease using neural networks, international journal of computer engineering and technology",2012

[3] https://archive.ics.uci.edu/ml/datasets/Heart+Disease

[4] https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be

[5] S.Florence, N.G.Bhuvaneswari Amma, G.Annapoorani, K.Malathi "Predicting the Risk of Heart Attacks using Neural Network and Decision Tree", International Journal of Innovative Research in Computer and Communication Engineering, 2014 Nov

[6] Milan Kumari, Sunila Godara, "Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction" , International Journal of Computer Sci ence and Technology, IJCST Vol. 2, Iss ue 2, June 2011

[7] Abhishek Taneja, "Heart Disease Prediction System Using Data Mining Techniques", Oriental Journal Of Computer Science & Technology, ISSN: 0974-6471 Vol. 6, No. (4), Dec 2013

[8] Velu C. M. and Kashwan K. R., IEEE Member , 'Heart Disease Diagnosis Using Multiple Kohenen Self Organizing Maps'

[9] M.A.Nishara Banu and B.Gomathy," Disease Forecasting System Using Data Mining Methods",2014

[10] Dr. T. Karthikeyan1, V.A.Kanimozhi," Deep Learning Approach for Prediction of Heart Disease Using Data mining Classification Algorithm Deep Belief Network",2017