Heart Diseases Prediction using Deep Learning Neural Network Model

Sumit Sharma, Mahesh Parmar

Abstract:- Deep learning plays an important role in the field of medical science in solving health issues and diagnosing various diseases. So in this paper, we will discuss heart disease.

We proposed a model for heart disease prediction. Heart Disease is on of key area where Deep Neural Network can be used so we can improve the overall quality of the classification of heart disease. The classification can be performed on the various ways like KNN, SVM, Naïve Bayes, Random Forest. Heart Disease UCI dataset will be used to demonstrate Talos Hyper-parameter optimization is more efficient than others.

Keyword:- Deep Learning, Neural Network (NN), CNN, RNN, KNN, SVM, Heart disease dataset.

I. INTRODUCTION

In our day to day life,Today generation is very busy in our daily routine schedule which feels to nervousness, restlessness and stress.Each individual have distinctive pulse rate and blood pressure which ranges from 60 to 100 BPM for pulse rate and 120/80 to 140/90 for blood pressure.

In world wide, heart disease is the major issue in human life. Heart means “Cardio”. Heart disease category is called Cardiologist disease.

The different types of heart disease are following :-

- Congenital heart disease.
- Arrhythmia.
- Coronary artery disease.
- Dilated cardiomyopathy.
- Myocardial infarction.
- Heart failure.
- Hypertrophic cardiomyopathy.
- Mitral regurgitation

Deep learning is a more popular machine learning method. It is not only when applying it in image classification tasks but also uses normal tabular data. In this model, we create a deep learning neural network model using Talos.

Talos is a hyperparameter optimization techniques. In this model, we can use the Talos optimizer with Keras library. Keras is a deep learning neural network library. Keras creates a high-level neural networks model. It developed for easy and fast experimentation. Keras supports both convolutional neural networks (CNN) and recurrent neural networks (RNN), as well as combinations of the two. It runs perfectly on CPU and GPU. Talos is a fully automated POD (Prepare, Optimize, Deploy) pipeline that stability yields state by step prediction results in a wide range of prediction related problems. In this field, large amount of researches has been done by using various algorithms and techniques.

This paper aims to achieve better accuracy and to make the system more efficient so that it can predict the chances of a heart attack.

II. RELATED WORK

Nowadays many researchers used machine learning, deep learning and data mining in healthcare for the predict to the disease but each research gives your opinion and your prediction accuracy according to your research.

Tan et al. [1] suggested a hybrid method using two machine learning algorithms, one is SVM (Support Vector Machine) and another is GA (Genetic Algorithm), both are effectively combined with in this approach. Data mining tools like LIBSVM and WEKA are used for this analysis here we collected 5 different dataset from the IUC repository. When we applied the hybrid model it reaches to an accuracy of 84.07% for heart disease, 78.26 for diabties and 76.20% for breast cancer and 86.12% for Hepatitis

Chaurasia et al.[2] recommended data mining perspectives to detected heart diseases in human body. In this data mining approaches, WEKA tool is used for machine learning purposes which is used multiple algorithms for data mining like - J48, Naïve bayes and bagging. The UCI laboratory is a part of machine learning. In the heart disease dataset 313 attributes and 13 attributes for prediction. Naïve Bayes gives 82.31% accuracy, J48 offers 84.35% and Bagging offers 85.35% accuracy in classification.

As suggested by Vembandasamy et al. [3] using the Naive Bayes algorithm which uses Bayes approach. Naïve Bayes algorithm has a robust principle of independence dataset. Here we use one of the leading diabetes research dataset which consist of 500 patients. WEKA tools is data mining tools and performs classification using 70% for training and 30% for testing dataset. Accuracy of Naive Bayes 86.419%.

SahayaArthyet.al[4]analyses the existing works on heart disease prediction which uses data mining. The data mining techniques are commonly used in heart disease prediction. They also discuss the databases used such as the heart disease data set from UCI repository, tools used such as Weka, Rapid Miner, Data melt, Apache Mahout, Rattle, KEEL, R data mining and soon. They conclude that the use of single algorithm results in better accuracy in prediction. But the use of hybridization of two or more algorithms can enhance and improve the heart disease prediction with good accuracy.

Maratea et al, [5] evaluated a sequential feature selection approach using a neuro-fuzzy classifier. Its obtaining the Cleveland set accuracy of

Revised Manuscript Received on January 5, 2020

***** Correspondence Author

SUMIT SHARMA, M.tech. Scholar,Dept. of CSE & IT, MITS Gwalior,Gwalior, Madhya Pradesh, India.Email: s.s.sumit1996@gmail.com

MAHESH PARMAR, Assistant Professor MITS,Dept. of CSE & IT, MITS Gwalior,Gwalior, Madhya Pradesh, India.

Email: maheshparmarcse@mitsgwalior.in

88.2%. in this paper, training data is using 50% of dataset and testing data is using 50% of the dataset.

Lakshmi et. al. [6] maintaining 10 methods, Its using heart disease UCI repository in the 4 sets, and found that a Partial Least Square Discriminant Analysis (PLS-DA) method accuracy of 86.13%.

R. Sharmila et al, [7] as suggested to enhanced the prediction of heart diseases dataset using the data techniques. SVM provided the better and efficient accuracy 85%. In SVM, parallel fashion gives better accuracy than sequential SVM.

Ashwini Shetty A et al, [9] Different Data Mining Approaches for Predicting Heart Disease. WEKA tool, MATLAB. Accuracy of Neural Network 84% and Accuracy of Hybrid Systems 89%.

Chala Bayen et al, [10] as suggested to Prediction and Analysis for Heart Diseases using data mining techniques. J48, Naïve Bayes (NB), Support Vector Machine (SVM). It gives the better result which helps to improved the quality of services and reduce the cost to individuals.

III. PROPOSED METHODOLOGY

In this paper, We deploy a model "Optimized DNN using Talos" and compare the method to others it is more efficient to others. This model provided a high accuracy compared to others. In this model, we are following some steps.

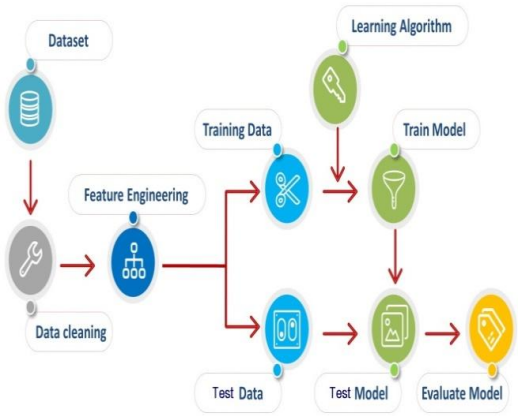

Fig.1- Flow diagram

A. Dataset:- In this paper, we are using the Heart Diseases datasets. In Heart Diseases, 14 attributes are given shown in fig.2 and 303 columns who represents the patient's data. In this database, we apply the DNN using Talos and predict the diseases.

Table 1 – Dataset description

No.	Attributes	Descriptions
1.	Age	Patients Age (in Year)
2.	Sex	0 : female and 1 : male
3.	Cp	Type of Chest pain Type 0: typical Angina Type 1: atypical angina Type 2: non-anginal pain Type 3: asymptomatic
4.	Trestbps	Resting Blood sugar (in mm Hg on admission to the hospital).
5.	Chol	serum cholesterols in mg/dl
6.	Fbs	Fasting blood suger > 120 mg/dl. (1= true; 0= false)
7.	Restecg	Resting ECG result
8.	Thalach	Maximum heart rates Achieved.
9.	Exang	Exercise induced angina.
		(1 = yes; 0 = no)
10.	Oldpeak	ST depression induced by exercise relative to rest.
11.	Slope	Slope or peak exercise ST Segment. Value 1:upsloping Value 2: flat Value 3: downsloping
12.	Ca	number of major vessels (0-3) colored by flourosop
13.	Thal	3 = normal; 6 = fixed defect; 7 = reversable defect
14.	Num	The predicted attribute. 0 : Yes; 1 : No.

B. Data cleaning: - Data cleaning is the first and necessary step for any project processes and data models. It means filtering and modifying your data such that it is easier to explore and understand.

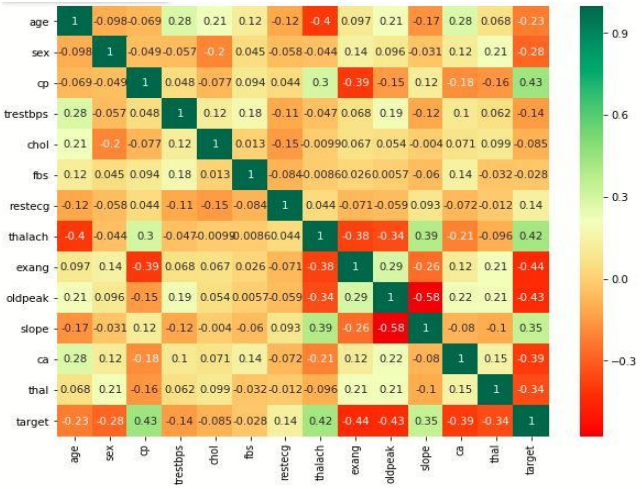

Fig.2- Heat-map

C. Feature engineering:- Features engineering is the most popular part of the deep learning. Features engineering is used for extracting some features data set. Features engineering is the process of transferring the raw data to features data and this features data improve the quality of the model and provide better accuracy.

D. Learning algorithms:- We are using the many types of learning algorithms –

a) Logistic regression :- Logistic regression is a classification algorithm for categorical variables. Sigmoid function is the most part of logistic regression.

Algorithms :- Logistic regression algo. following some steps.

1. Initialize Θ.
2. Calculate $Y = \sigma(\Theta^T X)$ for a customer.
3. Compare the output of Y with actual output of customer y, record it as error.
4. Calculate the cost of all customers.
5. Change the Θ to reduce the cost.
6. Go To Step 2.
7. Stop.

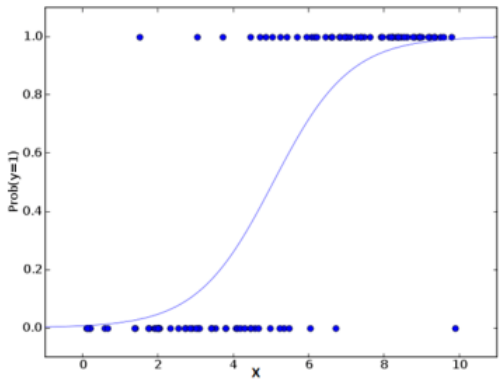

Fig.3- Logistic regression

b) K-NN:- KNN is a non- parametric machine learning algorithm. It is a supervised learning algorithm. It means to predict the output from the input data.

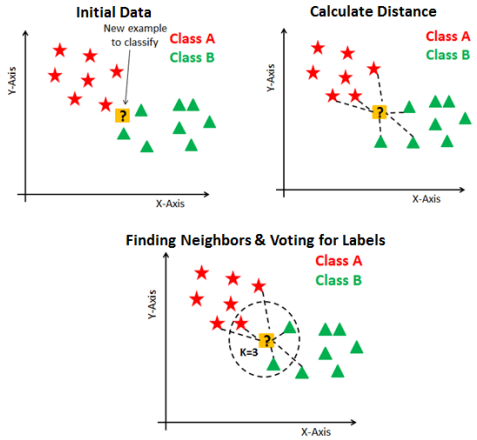

Fig.4- KNN

K-NN Algorithms:- K-NN algo following some steps.

1. Pick a value for K.
2. Calculate the distance of unknown case from all cases.
3. Select the K- observations in the training data that nearest to the unknown data point.
4. Predict the response of unknown data point using the most popular response value from the KNN
5. Stop.

In this algorithms data is divided into training and test data sets. The training dataset is used for model building and training. K- value is decided which is often the square root of the number of observations. Now the test data is predicated on the model built.

c) Support Vector Machine(SVM):- SVM is supervised deep learning algorithm which can be used both classification and regression.

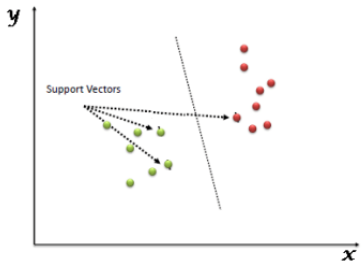

Fig.5- Support vector

The SVM algorithm is used to predict this disease by plotting the training dataset where a hyperplane classifies in two – presence and absence of heart disease.

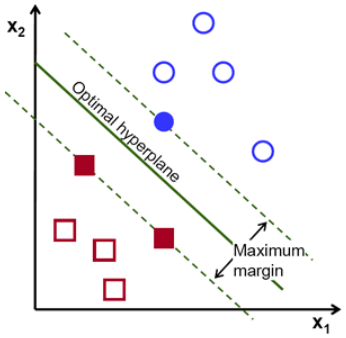

Fig.6- SVM

SVM is used to handle class imbalance. Class imbalance is a problem in machine learning when the total number of positive and negative is not the same and the classifier will not perform well

d) Naïve Bayes:- Naive Bayes is a probabilistic machine learning classification algorithm based on the Bayes Theorem. It is used in a wide variety of classification tasks.

Bayes Rule is a way of going from P(X|Y), known from the training dataset, to find P(Y|X), known from the test data.

$$\rho(X|Y) = \frac{\rho(X \cap Y)}{\rho(Y)} \quad \rho(Y|X) = \frac{\rho(Y \cap X)}{\rho(X)}$$

P(X|Y) is known data means P(Evidence|Outcome) for testing and P(Y|X) is unknown data means P(Outcome|Evidence) for testing.

$$\rho(Y|X) = \frac{\rho(X|Y) * \rho(Y)}{\rho(X)}$$

e) Hyper-parameter optimization(Talos) :- Talos follows POD (Prepare, optimize, Deploy), process workflow and additional functionality for evaluation, reporting, including plots for visual analysis.

Prepare (P): - this is the first process for preparation of defining the hyperparameter space for the experiments and it is the setting oof experiment options such as choosing for the optimization strategy.

Optimize (O): - This is the second process for optimization,. Its automated process of finding an optimal hyperparameter combination for a well generalizing model for a given prediction task.

Deploy (D): - This is the third process for deployment. Its automated the process of sorting locally the required assets for local and remote deployment of a model for production purpose.

Reporting and evaluation is the last process after all 3 POD workflow. Its provides several facility for analysis and evaluation of experiments, including the all plots for epoch-by-epoch visual analysis for experimental progress.

f) Random forest classifier :- Random forest classifier makes a set of decision trees from randomly selected subset of training dataset. It aggregates the votes from different decision trees to decide the final class of the test object.

Random Forest Algorithms :- This algo follows the some steps.
1. Choose an attributes from your dataset.
2. Calculate the significance of attribute in splitting of data.
3. Split data based on the value of the best attribute.
4. Go to Step 1.

E. Train & Test model:-Data is split into test and train dataset where The train set contain a output field on which the model learns.

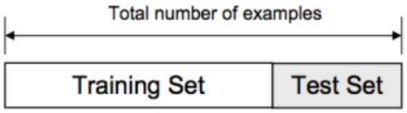

Fig.7- Data splitting

Here we applied cross-validation so our data remains randomly across the train set.

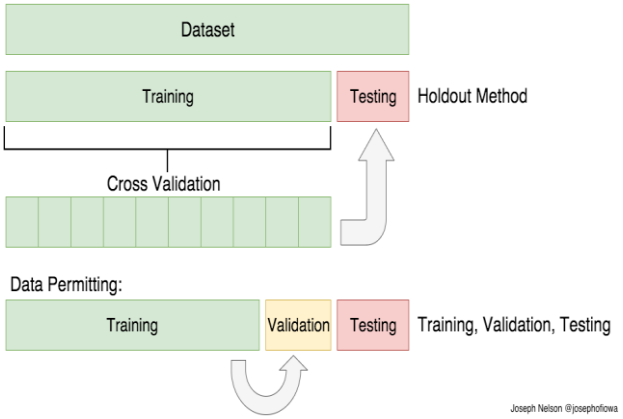

Fig.8 - Training to testing process

In the training/testing model, two words are very popular first is Under-fitting and second is Over-fitting.
If the model is under-fitting then our model doesnot fill well and perfor poorly
And if it Over-fits then it performs poorly on test set.

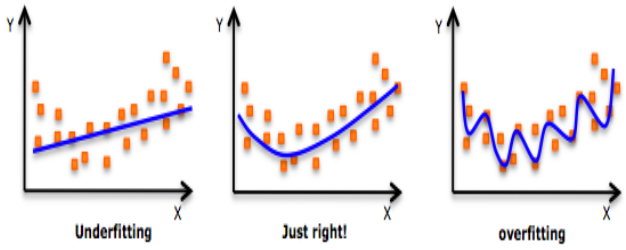

Fig.9 –Under-fitting and over-fitting

F. Evaluate model:-We deploy the neural networks model. In this model, we are using the input layer, output layer, hidden layer, and activation function. In this model, we deploy the deep neural network model using Talos optimization. The purpose of Talos optimization is allowing to use to continue working with Keras models.

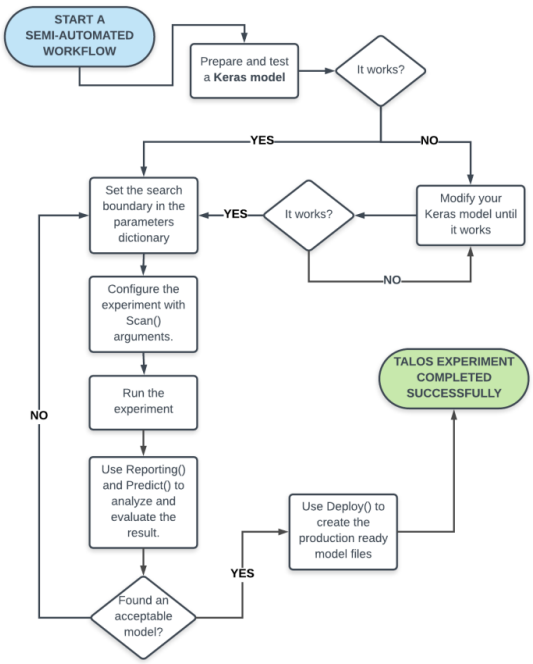

Fig.10- Hyper-parameter optimization technique using Talos

IV. RESULTS

In this paper we applied some classification algorithms (like – K-NN, SVM, Hyper-parameter optimization) on Heart diseases data set and measure the all classification accuracy is available on below mention table.

Table 2.- Algorithms accuracy comparison

S.No.	Classification Algorithms	Accuracy
1.	Logistic Regression	85.25%
2.	K-NN	90.16%
3.	SVM	81.97%
4.	Naïve Bayes	85.25%
5.	Hyper-parameter optimization (Talos)	90.78%
6.	Random forest	85.15%

In this paper, we finally found out the best classification algorithms is Hyper-parameter optimization using Talos for Heart diseases dataset.

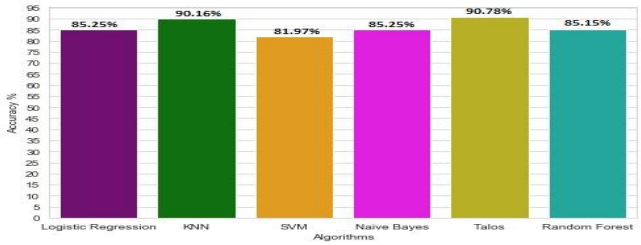

Fig.11- Algorithms Accuracy comparison

V. CONCLUSIONS

A summary of this paper arranged in alogical sequence that generally follows your methodology section. Compare to other algorithms and optimization, it is proved good results for prediction. In this paper, We deploy a deep learning neural networks (DNN) using Talos optimization. Talos optimization is newly

optimization techniques in DNN. Talos provide better accuracy (90.76%) to other optimizations. It is applied on the Heart disease datasets and find out the good prediction. Using the Talos optimization we create a Keras model and deploy it.

REFERENCES

1. K. C. Tan, E. J. Teoh, Q. Yu, and K. C. Goh, "A hybrid evolutionary algorithm for attribute selection in data mining", Expert Systems with Applications, Vol.36, No.4, pp.8616-8630, 2009.
2. V. Chaurasia and S. Pal, "Data mining approach to detect heart diseases", International Journal of Advanced Computer Science and Information Technology, Vol.2, No.4, pp.56-66, 2014.
3. K. Vembandasamy, R. Sasipriya, and E. Deepa, "Heart Diseases Detection Using Naive Bayes Algorithm", *IJISET-International Journal of Innovative Science, Engineering & Technology*, Vol.2, pp.441-444, 2015.
4. A.Sahaya Arthy, G. Murugeshwari "A survey on heart disease prediction using data mining techniques" (April 2018).
5. Hamid Reza Marateb and Sobhan Goudarzi, "A noninvasive method for coronary artery diseases diagnosis using a clinically interpretable fuzzy rule-based system," Journal of Research in Medical Sciences, Vol. 20, Issue 3, pp.214-223, March 2015.
6. K.R. Lakshmi, M.Veera Krishna, and S.Prem Kumar, "Performance Comparison of Data Mining Techniques for Predicting of Heart Disease Survivability," International Journal of Scientific and Research Publications, Vol.3, Issue 6, pp.1-10, June 2013.
7. R. Sharmila, S. Chellammal, "A conceptual method to enhance the prediction of heart diseases using the data techniques", International Journal of Computer Science and Engineering, May 2018.
8. Purushottam, Prof. (Dr.) Kanak Saxena, Richa Sharma, "Efficient Heart Disease Prediction System", 2016, pp.962-969.
9. Ashwini Shetty A, Chandra Naik, "Different Data Mining Approaches for Predicting Heart Disease", International Journal of Innovative in Science Engineering and Technology, Vol.5, May 2016, pp.277-281.
10. Mr. Chala Beyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, 2018.

AUTHORS PROFILE

SUMIT SHARMA, (M.tech.Scholar), Dept. of CSE & IT, MITS Gwalior, Madhya Pradesh, India. Area of Interest: Data Mining, Image Processing, Machine Learning and Deep Learning. I am an student of CSE&IT Department in MITS Gwalior. I have received B.E. degree in information technology from MITS Gwalior. I am three times Gate qualified 2017, 2018 and 2019.

Mr. Mahesh Parmar, as an Assistant Professor in CSE&IT Department in MITS Gwalior and having 10 years of Academic and Professional experience. He received M.E. degree in Computer Engineering from SGSITS Indore. He has guided several students at Master and Under Graduate level. His areas of current research include Data mining and Image Processing. He has published more than 30 research papers in the journals and conferences of international repute. He has also published 02 book chapters. He is having the memberships of various Academic/ Scientific societies including IETE, CSI, and IET etc.