International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1, October, 2019

An Automatic System for Heart Disease Prediction using Perceptron Model and Gradient Descent Algorithm

V. Sahaya Sakila, Akshat Dhiman, Kanyush Mohapatra, Panchal Raj Jagdishkumar

Abstract: ***Many deep learning neural network-based models had been proposed for the prediction of heart diseases but there is no model accessing all 13 features directly from the dataset & feeding them to the 'Perceptron'. The model starts with certain weights and biases to train itself with the given dataset. The error is calculated in each epoch using the popular 'gradient descent' algorithm. Improper arrangement of neurons in a neural network can lead to overfitting or underfitting of the model which will gradually decrease the accuracy. But the perceptron model has only a single neuron performing all the classification and arrangement of data. The weights and biases in the model are updated to fit the model properly in a way that the equation can classify maximum number of data points. The learning algorithm of Perceptron model which calculates the result and performs binary classification is ∑_(i=0)^n wixi >= b. The weights and biases in the equation are changed repeatedly over epochs to decrease the error and fit the dataset with most efficient predicted output. Compared to the old MP Neuron model, the model is more effective with an accuracy of 98.349%.***

Keywords **: Gradient descent, Heart disease, Overfitting, Perceptron model, Underfitting.**

I. INTRODUCTION

The heart is one of your body's most important organs. The disorders that affect the ability of heart to function norm ails are called heart disease. The blocking & narrowing of the coronary arteries, that supply blood to the heart, is a very common cause for failure of heart. The other name for disease related to heart is the cardiovascular disease (CVD). The term cardiovascular disease means any disorder related to heart and blood vessels, that includes hypertension, coronary artery disease (CAD), cardiac dysrhythmias, cerebrovascular disease, valvular heart disease, cardiomyopathies, peripheral vascular disease, and congenital cardiac abnormalities in functioning. Each disorder is characterized epidemiologically; incidence and prevalence rates vary with a wide range for country and different culture. Because hypertension, coronary artery disease, cardiac dysrhythmias, and cerebrovascular disease does most of the cardiovascular rigidity and countries which are developed have mortality. In today's world, common man has no immediate access to the prediction machinery which can tell the chances of a heart disease that can happen to patient. So, people tend to ignore the symptoms rather than going for a proper check-up. Every year several people die due to lack of awareness and immediate access of the treatment. If people use the model as a precaution than they will be able to know that whether they should visit a proper doctor or not. This can decrease the number of deaths due to heart related diseases in future. A blood test is enough to extract the 13 required parameters required to run the model. In future if the model is converted into an app, users will have more easy access to a proper heart companion. Thus, the main purpose is to reduce the increasing number of deaths caused due to heart related issues.

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
V. Sahaya Sakila Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India.
Akshat Dhiman, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India. E-mail: akshatdhiman982@gmail.com
Kanyush Mohapatra, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India. E-mail: rskumarsoro@gmail.com
Panchal Raj Jagdishkumar, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India. Email: roger71011112000@gmail.com

II. RELATED WORKS

2.1 LITERATURE SURVEY

Human heart disease prediction system using data mining technique. The risk of patient is classified with datamining classification techniques such as KNN, Decision Tree Algorithm, Naïve Bayes, Neural Network etc. [1]. The KNN algorithm has the K user defined value to find the values of the factors of heart disease. Decision tree algorithm is used to provide the classified report for the heart disease. The Naïve Bayes method is used to predict the heart disease by using probability [1]. Various techniques used are useful to compare different accuracies.

The heart disease prediction which uses a hybrid technique for classification with data mining. Heart disease is a common reason for deaths happening in India or any other countries of Asia. In the year of 2003 approximate 17.31 million people had died all over the globe and from them, 10 million were due to the coronary heart disease [2]. The paper aims to build a pseudo structure which have the capability to determine and extract not known patterns and heart disease relations from a past heart disease record of database [2].

A system has used Back-Propagation in neural network which is regarded as the best prediction algorithm where we have a non-linear relationship between the data and the output [2].

It has the capability to solve complicated queries for the detection of heart diseases and finally assisting the doctors for making smart clinical decisions [2].

Prediction of Heart Disease Using Machine Learning. Machine learning algorithm neural networks are used for predicting the vulnerability of a heart disease using an application which is developed [3]. It's not practical for a common man to frequently undergo costly tests like the ECG and thus there needs to be a system in place which is handy and at the same time reliable, in predicting whether there is a chance for heart disease or not [3]. Multi-layered Perceptron is used which makes it easy to divide & feed the dataset. The System of prediction for Heart Disease using Machine learning algorithms. MLP provides its users with a prediction result that gives the CAD state of a user [3].

Efficient Heart Disease Prediction System using Decision Tree. An experiment is designed using the Cleveland dataset. In the preprocessing phase All Possible-MV [I] algorithm is used to fill the missing values in the dataset [4]. Results show that a great potential lies in the system for predicting the risk level of heart disease more accurately. A statistical method for analyzing the classifier performance [4]. The database of Cleveland contains 76 raw attributes in total, but in experiments only 14 of them is used. The dataset used in this experiment contains different important parameters like ECR, chest pain, MHR (maximum heart rate), fasting sugar, cholesterol and many more [4].The World Health Organization (WHO) believes that more than 12 million people every year die due to coronary diseases.In 2008 coronary disease took life of 1.73 crores people, and heart disease took 80% of the life of people.[5]. With the use of mining algorithm, we will present rule extraction experiment on heart disease data. We start our examination by cognizance the information source and knowing the quality of the data. Then, we find the first bits of learning into the information [5]. To pre-process the information from harsh dataset to the last is the initial step, and then the data gets arranged [5]. Thus, To identify the risk of heart disease accurately Heart disease prediction system is proposed.

With the use of big data approach, Doing Real-time machine learning for detection of heart disease at the earliest. A real time heart disease prediction system based on apache Spark is proposed by this paper, which act as a strong large-scale distributed computing platform, with the use of which the data event is streaming against machine learning via in-memory computations [6]. The system is composed of two main sub parts, namely streaming processing and data storage and visualization. One uses Spark ML library with Spark streaming and use the application of classification model on data events for prediction of heart disease. Other uses Apache Cassandra for large volume storage of data that is generated [6].

2.2 Existing System:

2.2.1 Disadvantages

- The model has an accuracy of 93.33% in test data set and 84.05% in training data set which is lower than our perceptron model.
- The model uses multiple neutral network layers which means any error in numbers of neuron in hidden layers will result in major differences in the result.
- The model divides the 13 parameters into different blocks of data rather than directly feeding the whole data set to a neuron which decreases the speed of evaluation.

III. PROPOSED SYSTEM

In today's world, common man has no immediate access to the prediction machinery which can tell the chances of a heart disease happening to patient in coming times. So, people tend to ignore the symptoms rather than going for a proper check-up. Every year several people die due to lack of awareness and immediate access of the treatment. If people use the model as a precaution than they will be able to know that whether they should visit a proper doctor or not. This can decrease the number of deaths due to heart related diseases in future. A blood test is enough to extract the 13 required parameters required to run the model. In future if the model is converted into an app, users will have more easy access to a proper heart companion. Thus, the main purpose is to reduce the increasing number of deaths caused due to heart related issues.

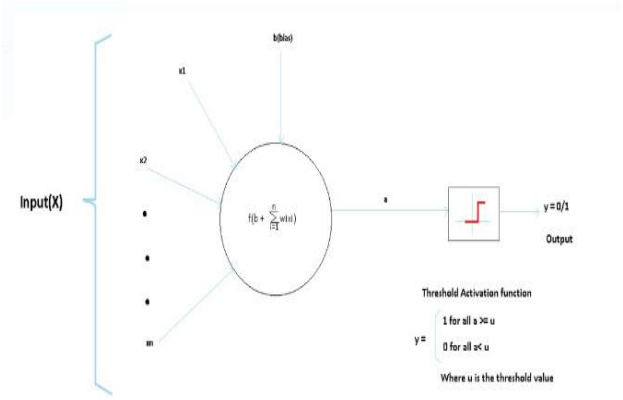


Fig. 1 Perceptron Model

3.1 The Process & Methodologies

3.1.1 Dataset Collection & Uploading

The dataset is collected from online sources like Kaggle, Google Datasets and UCI. The csv format dataset is uploaded into the Google Colab which is the playground where we write the code. The model divides the 13 parameters into different blocks of data rather than directly feeding the whole dataset to a neuron which decreases the speed of evaluation.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Fig. 2 Heart Parameters Dataset

3.1.2 The Perceptron Model

Certain computations to detect features or business intelligence in the input data is done by a neural network unit called Perceptron. It is an algorithm for supervised learning of binary classifiers. This Perceptron algorithm allows the neurons for learning and processing elements in the training set one after the other. This Learning Rule states that the algorithm would learn the optimal weight coefficients on its own. These weights are then multiplied with the input features to discover that whether a neuron is firing or not. Multiple input signals are received by this model, and if a certain threshold is surpassed by the sum of the input signals, then either a signal is outputted or not output is returned. In the context of supervised learning and classification, this can be used for prediction of the class of a sample. It is a function that maps its input "x," multiplied by learned weight coefficient; It generates an output value" f(x)".

$$f(x) = \begin{cases} 1 & \text{if } w \cdot x + b > 0 \\ 0 & \text{otherwise} \end{cases}$$

In the above equation:
"w" is vector denoting weights that are real-valued.
"b" is bias (a component adjusting the borders away from origin without depending on the value that are inputted)
"x" is vector of x values that are fed in.

$$\sum_{i=1}^{m} w_i x_i$$

"m" is number of inputs to the Perceptron
"1" or "0" can be used for representing the output. It can also be represented as "1" or "-1" depending on which activation function is used. The model uses the Sigmoid activation function to train the parameters.
These parameters are then tuned to fit the given dataset perfectly.

```
    if y == 1 and y_pred == 0:
      self.w += lr * x
      self.b += lr * 1
    if y == 0 and y_pred == 1:
      self.w -= lr * x
      self.b -= lr * 1

  accuracy[i] = accuracy_score(self.predict(X), Y)
  if(max_accuracy < accuracy[i]):
    max_accuracy = accuracy[i]
    chkp_w = self.w
    chkp_b = self.b
self.w = chkp_w
self.b = chkp_b
```

Fig. 3 Perceptron Training Code

3.1.3 Gradient Descent Algorithm

How good our model is at making predictions for a given set of parameters is notified by a Loss Functions. Its own curve and gradients are possessed by the cost function. The bias should always be initialized to zero at the starting of the process because any random y-intercept can lead to decrease in accuracy as same number of epochs cannot give same accuracy for any random y-intercept. For each epoch the predictions are made, and error is calculated. This error is calculated using a popular technique known as Gradient Descent Algorithm. In the process, we plot the predicted graph between the data points and check for error from each data point. The calculation and plotting of derivative of the loss function is done with respect to each weight. The partial derivatives of the cost function is calculated with respect to each parameter and the results are stored in a gradient. The graph derives the local minima. It's a pretty simpler approach to plot the differentiation of the original plot and get the local minima to find the lowest possible error.

the cost function as it is given:

$$f(w, b)=1N\sum i=1n(yi-(wxi+b))2f(w, b)=1N\sum i=1n(yi-(wxi+b))2$$

We can calculate the gradient as:

$$f'(w, b)=[\lfloor dfdwdfdb \rfloor]=[1N\sum -2xi(yi-(wxi+b))1N\sum -2(yi-(wxi+b))]\backslash$$

The graph shows an example of how gradient descent works.

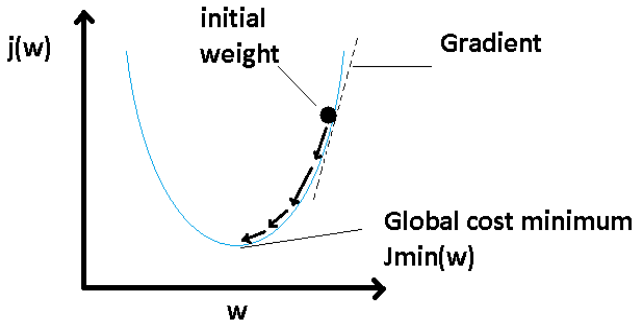
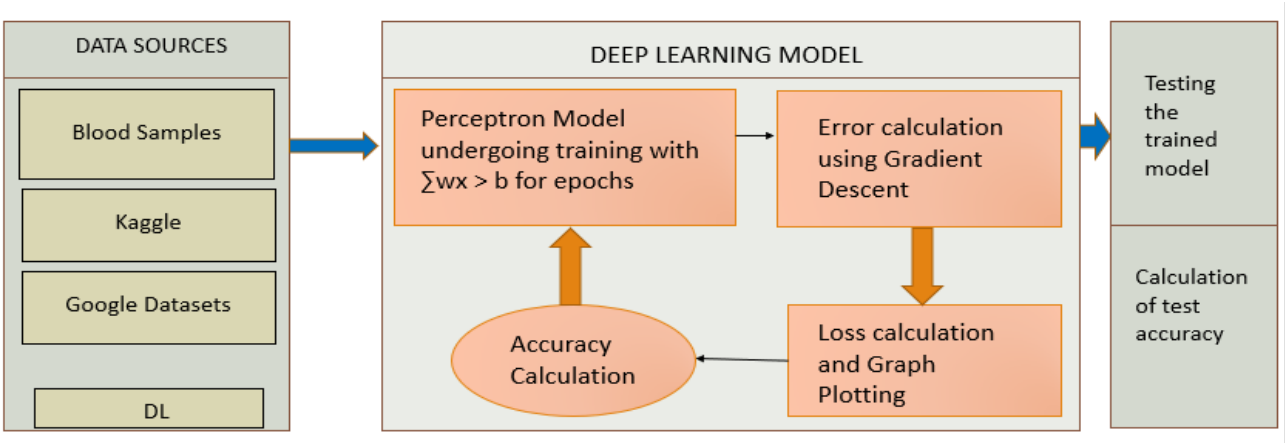


Fig. 4 Gradient Descent Graph

IV. SYSTEM ARCHITECTURE



The system has mainly three different parts. First, the dataset is collected from different data resources available on Google. The dataset from the model is derived from a popular data scientist website called 'Kaggle'. We have taken some references from Google Datasets and UCL also. Secondly, the model is applied. The dataset is trained, and error is calculated. The accuracy graph is plotted, and improvements are done. Thirdly, the test dataset is checked on the model and test accuracy are calculated from the user input.

V. RESULT AND DISCUSSION

After we test the user input data on our trained model, accuracy is calculated and plotted for each epoch. The training accuracy has been found to be more than testing accuracy and the training accuracy of the Perceptron model is more than the accuracy of the previous model.
The accuracy graph is plotted.

5.1 Accuracy Plotting & User Input

After we calculate the error and try to minimize it over our training process, the model get the checkpoint weights and biases where the error is the lowest and then error decrease is plotted over all the epochs. The accuracy is finally calculated using the scikit learn library. The training accuracy is the accuracy received after training the complete dataset to the model after completing all the epochs. Test accuracy is the accuracy received when we test the model using another small portion of the main dataset known as test dataset. In this model, instead of test dataset, user input is taken, and the model is evaluated based on that single input set. The model responds with a proper statement stating whether the patient is healthy or likely to get a disease in future.
For example, a specimen's chemical composition need not be reported if the main purpose of a paper is to introduce a new measurement technique. Authors should expect to be challenged by reviewers if the results are not supported by adequate data and critical details.

VI. CONCLUSION

The main motivation is to provide an insight about detecting heart disease risk rate using data mining techniques. A lot of calculation techniques and different algorithms were used to give different precision depending upon number of parameters.It detected risk rate of heart disease with the help of Perceptron and Gradient Descent algorithm, and accuracy level also provided for 13 attributes that are different and are used from the dataset directly. In future,It could reduce the numbers of attributes, and would increase the accuracy using some other algorithms. We have been planning to develop an android application for patients or users who are feeling a little tension around their hearts. A basic interface of the app was already developed and possibly we can complete the development process in future. The users will need to check their symptoms from a list of options available in the app and app will generate the 13 heart parameters automatically based on the scale of input from the user or user will have an option to input the 13 parameters on their own. The backend connected through java will carry the input parameters to the model, where it will test the data and predict the output and then display in the app.

REFERENCES

1. Theresa Princy. R, J. Thomas, "Human Heart Disease Prediction System using Data Mining Techniques", 2016.
2. Meghna Sharma, Ankita Dewan, "Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification" 2015.
3. Gouthami Kokkula, Isha Pandya, Aditi Gavhane, Prof. Kailas Devadkar (PhD), "Prediction of Heart Disease Using Machine Learning", 2018.
4. Purushottam, Prof. (Dr.) Kanak Saxena, Richa Sharma, "Efficient Heart Disease Prediction System using Decision Tree", 2015.
5. Aditya Jain, Aakash Chauhan, Purushottam Sharma, Vikas Deep, "Heart Disease Prediction using Evolutionary Rule Learning", 2018.
6. Khalil Maalmi , Abderrahmane Ed-daoudy, "Real-time machine learning for early detection of heart disease using big data approach", 2019.

AUTHORS PROFILE

Akshat Dhiman, undergrad at SRM Institute of Technology, Deep Learning geek, Coder, Blogger, AI Enthusiast, In top 60% coders in CoDecode Competition by IIT Techfest, In top 1% in Maharashtra State Board, Member of FnPlus Community.

Kanyush Mohapatra, undergrad at SRM Institute of Technology, Coder, NCC Cadet, Silver Medalist in 15th State Taekwondo Championship.

Panchal *Raj* Jagdishkumar undergrad at SRM Institute of Technology, Coder, Football Player, Java, SQL Database Management, C. PHP Developer, HTML.