Proceedings of the Fifth International Conference on Inventive Computation Technologies (ICICT-2020)
IEEE Xplore Part Number:CFP20F70-ART; ISBN:978-1-7281-4685-0

Heart Disease Prediction Using Deep Neural Network

P. Ramprakash[1], R. Sarumathi[2], R. Mowriya[2], S. Nithyavishnupriya[2]

[1]Assistant Professor, Department of Information Technology, Dr. Mahalingam College of Engineering and Technology, Pollachi, India
[2]Student, Bachelor of Information Technology, Dr. Mahalingam College of Engineering and Technology, Pollachi, India

ABSTRACT

Healthcare occupies an indispensable part in human lives. The healthcare industry contain large amount of psychiatric data hence machine learning models were used to provide conclusion effectively in the heart disease prediction. The classification of healthy person and non-healthy person can be done reliably by using machine learning methods. We developed a framework in this exploration that can understand the principles of predicting the risk profile of patients with the clinical data parameters. The proposed model is constructed using Deep Neural Network and χ^2-statistical model. The problem of under fitting and over fitting is eliminated. This model shows better results on both the testing and training data. DNN and ANN were used to analyse the efficiency of the model which accurately predicts the presence or absence of heart disease.

KEY WORDS: Deep Neural Network, Hypothesis, χ^2 statistical model, Heart disease, over fitting and under fitting.

I. INTRODUCTION

Heart disease describes a range of condition that affects the heart. The term cardiac disease is regularly utilized with cardio vascular disease (CVD). The blood to the heart is supplied by coronary supply routes and narrowing of coronary arteries is the major cause for heart failure. Prediction of cardiovascular disease is considered as one of the most important subject in the section of data analytics. The major cause of heart attack in United States is coronary artery disease. Cardiac disorder is widespread in male than that of female. The survey analyzed by World Health Organization (WHO) estimates that 24% of people died in India due to cardiac disorder. Analysts have recorded the different components that increment the possibility of cardiac disorder and coronary artery disease.

The elements are put down into two types that are the elements that could be modified and elements that could not be modified. Age, smoking, heredity, sex, high blood pressure, poor diet, consumption of alcohol, physical inactivity are considered to be the risk factor of heart diseases. Age, sex, heredity are the elements that could not be modified while the elements like smoking, intake of alcohol can be modified. The determinant condition related to lifestyle could be vicissitude with meditation and lifestyle changes.

The most common method used for treating coronary artery disease was angiography. Angiography is the most costliest method and have more reactions. The risk factors make the detection of heart diseases very difficult, these constrains leads to the evolution of modern methods which detects the presence of cardiac diseases. It is hard to identify cardiac disorders manually hence conventional methods are used which is mostly based on examination of patient data.

Machine learning techniques are useful for predicting the diseases. We develop a diagnosis system which is based on χ^2-DNN. The most objective of the determination framework is to dispense with the issue of over fitting and under fitting. The training data is feed in to the neural organizes. The testing dataset is utilized to access the execution of the neural organize. The DNN with multiple hidden layer is used hence the proposed model has high performance than ANN. χ^2 -DNN is developed for improving the classification accuracy of cardiac disorder prognosis.

II. BACKGROUND

Numerous frameworks for coronary disorder forecast utilizing AI calculation was developed early. Algorithm like back vector machine, K-nearest neighbour and Artificial Neural Network was developed early to anticipate the existence or nonappearance of cardiac disorder. Survey

Authorized licensed use limited to: University of Exeter. Downloaded on June 17,2020 at 19:49:49 UTC from IEEE Xplore. Restrictions apply.

indicates that ANN based models are mostly used in heart disease prediction and the accuracy of previous works was less in contrast with the developed model.

S. Radhimeenakshi proposed a classification system for the prediction of cardiac disorder. The work which he carried out focuses on two algorithms namely Support Vector Machine (SVM) and Decision Tree. He carried out his work in weak and python the same predictive model is used for both the training dataset and the testing dataset. The model achieved more accuracy when decision tree classification algorithm is used. The accuracy of the model is evaluated using confusion matrix and the model achieved accuracy of 55% for SVM.

Ⅲ. MATERIALS AND METHODS

A.DATASET DESCRIPTION

The Cleveland coronary disorder dataset is utilized which was taken from onlineUCI AI archive. This dataset is used for research study. The dataset has 303 instance and 14 attribute.

TABLE 1: Description of the elements of the dataset

S.No	Feature Code	Feature Description	Value
1	AGE	Ageof the patient in years	29-77
2	SEX	Gender of patient	1=male 0=female
3	CP	Type of chest pain	0=Atypical angina, 1=typical angina , 2=asymptotic, 3=non angina pain
4	TRESTB PS	Resting Blood pressure	94-200
5	CHOL	Serum cholesterol level	126-564
6	FBS	Fasting blood sugar 1>=120,0<=120	0=false 1=true
7	RESTECG	Resting electrocardiographic results	0=normal 1=ST-T wave abnormalities 2= left ventricular hypertrophy
8	THALACH	Maximum heart rate Achieved.	71-202
9	EXANG	Exercise Induced Angina	0=no 1=yes
10	OLD PEAK	ST depression induced by exercise related to rest	0.0-6.2
11	SLOPE	Slope of the peak exercise ST segment	0= un sloping 1=flat 2=down sloping
12	CA	Count of major vessels coloredBy Fluoroscopy	0-3
13	THAL	Thallium Scan	3=normal 6=fixed 7=reversible effect
14	Target	Class Attribute	0=no 1=yes

The dataset contains 8 categorical attribute and 6 numeric attribute. Table 1 contains the complete information about the dataset. The age attribute in the dataset differs from twenty nine to seventy nine. Periodical studies have shown that people who are greater than 65 are hardly suffering from cardiac disorder.

The male patient has gender value of one and female patient has gender value of zero. Male patient are at the high risk of heart disease than that of female patients. It is found that the female patients with diabetes are more likely to suffer from heart disease than that of male patient with diabetics.

Typical angina, atypical angina, non-angina pain and asymptotic are the different types of chest pain. Angina is the chest pain caused due to the absence of blood that is rich in oxygen which is supplied the heart. Typical angina is caused due to the reduced blood flow to the heart muscle. Atypical angina is caused by emotional or mental stress. Asymptotic is a not symptom of heart disease.

TRESTBPS indicates the resting blood pressure value of an individual the unit of TRESTBPS is mmHg. Serum Cholesterol is the total level of cholesterol accumulated.(LDL) represents Low-thickness lipoprotein which is named as awful cholesterol. Presence of high level of LDL narrows the arteries. (HDL) represents for High-thickness lipoprotein which is named as acceptable cholesterol. The presence of high level of HDL reduces the risk of heart attack.

FBS indicates the fasting blood sugar value of an individual. If the FBS is less than 120mg/dl then the value is 1. If the FBS is more than 120mg/dl then the value assigned to the attribute is 0. Not responding properly to the insulin secreted leads to the increase in the blood sugar level which increases the risk of heart disease. RESTECG displays the resting electrocardiographic result the value is assigned to 0 if the RESTECG is normal, the value is assigned to 1 if the RESTECG have ST-T wave abnormality. The value is assigned to 2 if RESTECG have left ventricular hypertrophy.

Maximum Heart Rate Achieved displays the maximum

Authorized licensed use limited to: University of Exeter. Downloaded on June 17,2020 at 19:49:49 UTC from IEEE Xplore. Restrictions apply.

heartbeat rate achieved by an individual. Increase in heart beat rate by 10% increase the cardiac death by at least 20%. EIA is recorded as 0 if there is no pain and recorded as 1if there is pain. Angina is usually felt in the centre of the chest it may even spread to both of the shoulders.

The important consideration is the duration of ST-segment depression as the recovery after the peak stress results in positive ECG stress test. SLOPE represents the slope of the ST segment. THAL represents duration in exercise test in minutes which displays the thalassemia. ECG stress test is considered as abnormal when there is down sloping ST segment depression> 1mm at 60-80ms. Target is the class attribute it is recorded as zero for non-diseased person and one for the person suffering with heart disease.

B.METHODOLOGY OF PROPOSED SYSTEM

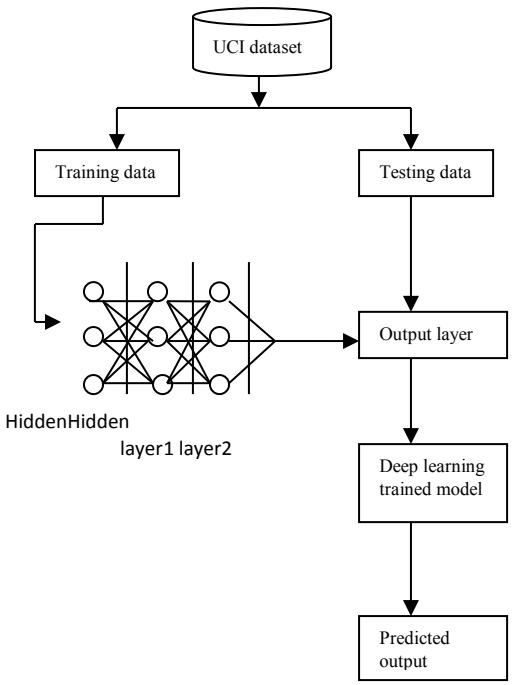

FIGURE1: Block diagram of proposed model

In practical application high accurate results are generated using neural network. The proposed system increases the classification accuracy. The dataset is divided into the testing data and training dataset. The training dataset was given to the neural network. Neural networks are set of algorithms that are used to recognize patterns. The layers in the neural network are made up of activation function. The training features are provided to the network through the input layer. The features are given to hidden layer where actual processing happens with the help of weighted connection. Output layer of the network is attached with the hidden layer. Generation of hypothesis through deep learning models was the aim of predictive model. Hypothesis is the relationship between data which can be tested by collecting data and making observation. We can generate the hypothesis by minimising the error in the training instances.

The performance of the network is dependent on the number of guidelines used which decide the behaviour of the network. Model with less parameter leads to low capacity which results in under fitting. Model with more number of parameters than required leads to high capacity which results in over fitting hence the model should be in such a way that it generates a hypothesis with optimum capacity. The hypothesis is formulated using forward propagation. The input is given to the neurons which performs some operation to generate the output this process is called activation function. The activation function defines the output of a node.

Deep neural network contains more than one hidden layer.

The activation of neurons present at the output layer.

$$f(x) = \frac{1}{1+e^{-x}} \qquad (1)$$

The sigmoid activation function is used in the output layer. Feature selection is used for the filtering of redundant features from the dataset. Feature extraction is dissimilar from feature selection. Feature extraction is getting useful elements from the existing data. Relevant elements are feed to the neural network by eliminating the irrelevant features using feature selection.

Irrelevant elements are removed using χ^2 statistical model. The dependence between the feature and class are measured using χ^2 test. Features are ranked in the first step. From the ranked features optimal features are searched in the second step.

TABLE 2: Confusion matrix

	Predicted heart disease patient	Predicted healthy patient
Actual heart disease patient	TP	FN
Actual healthy patient	FP	TN

Heart disease binary classification contains two classes one is the positive class and the other one is the negative class. Further it also contains t instances.

E_a represent the expected values. The expected value of two independent value based on null hypothesis can calculated as

Authorized licensed use limited to: University of Exeter. Downloaded on June 17,2020 at 19:49:49 UTC from IEEE Xplore. Restrictions apply.

$E_a=(a+b)\ \frac{a+b}{t}$ (2)

$x^2=\frac{1}{d}\sum_{k=1}^{n}\frac{(O_K-E_K)^2}{E_K}$ (3)

Threshold for the number of features must be decided after feature ranking which is denoted by n. The subset of feature with n=1 is taken and the optimum number of features is found through exhaustive search. The subset of feature is applied to DNN. Grid search is used to evaluate the DNN performance. After saving the result of first subset another subset with n=2 is taken and the optimum feature is found then it is applied to DNN and the result is saved. These procedures are continued till all the features are appended with subgroup of the features. The subgroup of feature which gives the best performance result is declared as the optimal result.

IV. VALIDATION SCHEME AND EAVALUTION METRICS

A. VALIDATION SCHEME

In the proposed model approval is performed utilizing train-test-hold-out approval conspire. The 80-20 train-test validation was applied which states the 80% of the information is utilized as preparing information and 20% of the information is utilized as testing information. The testing and training data are portioned using the sklearn library. Out of 303 samples 242 samples or instance is selected and utilized for preparing the model and the rest of 61 instances are utilized as testing information which is used to assess the execution of the developed model.

B. EVALUATION METRICS

Accuracy, sensitivity, specificity and Matthews correlation coefficient (MCC) are the various assessing methods which are used for assessing the execution of the developed predictive model. Exactness is the rate of accurately classified subject in the testing data.

Formula for evaluation metrics.

$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN}$ (6)

$\text{.Sensitivity}=\frac{TP}{TP+FN}$ (7)

$\text{Specificity}=\frac{TN}{TN+FP}$ (8)

Performance of Deep Neural Network with different epochs

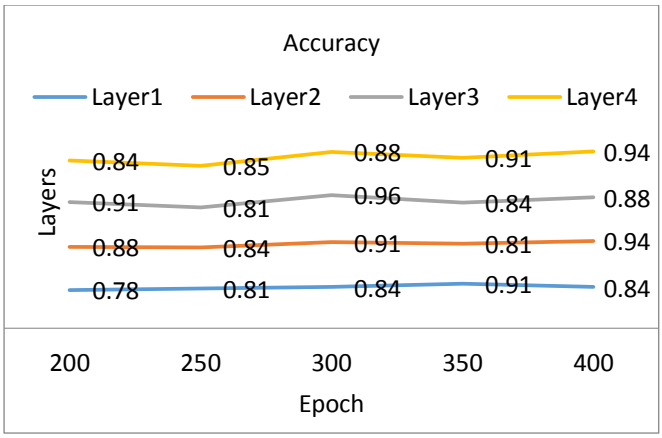

FIGURE2: Accuracy of Neural Network

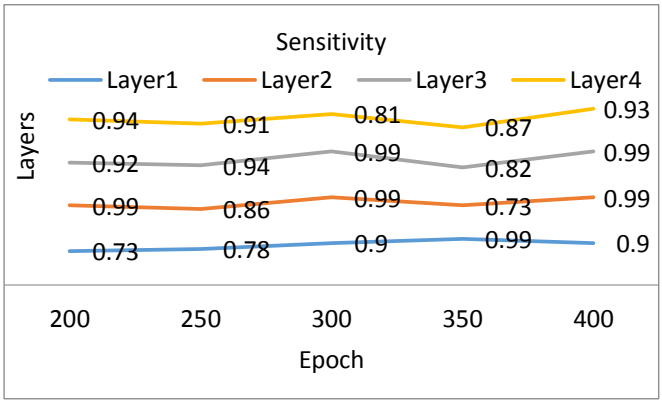

FIGURE3: Sensitivity of Neural Network

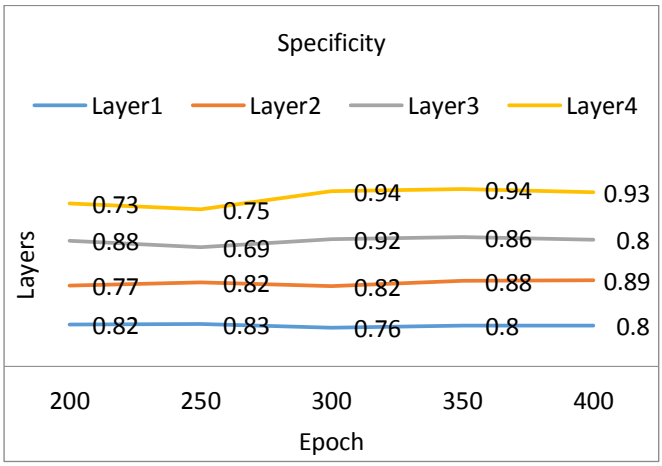

FIGURE4: Specificity of Neural Network

V. EXPERIMENTAL RESULTS AND DISCUSSION

The model is instructed on 242 instances and validated using the balance61 instances. The same elements were chosen for both testing and training data. The grid search algorithm is used to search the optimized configuration. 80-20 holdout validation is used for the optimization of the weights. The algorithm used is IBFGS where α=0.00001.

Authorized licensed use limited to: University of Exeter. Downloaded on June 17,2020 at 19:49:49 UTC from IEEE Xplore. Restrictions apply.

The proposed model contains 2 hidden layers. The conventional DNN is less accurate than that of the proposed χ^2-DNN.

In previous studies ANN is used with one hidden layer. ANN does not contain the idea of depth. In ANN it is not possible to append additional layers during tuning process.χ^2-ANN was designed in previous studies but the accuracy of χ^2 -ANN was less when compared with the χ^2 – DNN.

VI CONCLUSION AND FUTURE WORK

Heart disease is one of the corners for society. In this paper we developed a self-operating diagnosis model for cardiac disorder disease detection using deep neural network. The experimental result concludes that the proposed system improves standard of prediction during prognosis process. This work will be useful for identifying the patients who suffers from heart disease. When a patient is predicted with positive result their reports and data can be closely analyzed. Genetic algorithm can be used in future for more accuracy.

Family history of heart disease is also a reason for developing heart disease; hence this information of the patient can also be included in the dataset which improves the accuracy of the model.

VI REFERENCES

[1] S. Radhimeenakshi, "Classification and prediction of heart disease risk using data mining techniques of Support Vector Machine and Artificial Neural Network," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2016, pp. 3107-3111.

[2]O. Y. Atkov, S. G. Gorokhova, A. G. Sboev, E. V. Generozov, E. V. Muraseyeva, S. Y. Moroshkina, and N. N. Cherniy, "Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters," Journal of Cardiology, Vol. 59, Issue 2, pp. 190-194, 2012.

[3]Sana Bharti,Shailendra Narayan Singh,Amityuniversity,Noida, India Analytical study of heart disease prediction comparing with different algorithms(May 2015)

[4]U. R. Acharya, H. Fujita, O. S. Lih, M. Adam, J. H. Tan, and C. K. Chua, ‘‘Automated detection of coronary artery disease using different durations of ecg segments with convolutional neural network,’’ Knowl.-Based Syst., vol. 132, pp. 62–71, Sep. 2017.

[5] 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, 21 pages.

[6] HimanshuSharma,M A Rizvi Prediction ofHeart Disease using Machine Learning Algorithms:A Survey (August 2017) TulayKaraylan and OzkanKilic, "Prediction of heart disease using neural network," 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, 2017, pp. 719-723

TulayKaraylan and OzkanKilic, "Prediction of heart disease using neural network," 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, 2017, pp. 719-723

[7] J.Vijayashree andN.Ch. SrimanNarayanaIyengar Heart Disease Prediction System Using Data Mining and HybridIntelligent Techniques: A Review (2016)

Asdfghjkl;kjhgfddfl

[8] SellappanPalaniyappan, RafiahAwangIntelligent heart disease prediction using datamining techniques (August 2008)

[9] RamandeepKaur, 2Er. PrabhsharnKaurAReview - Heart Disease Forecasting Pattern usingVarious Data Mining Techniques (June 2016)

[10] K.Polat,S.Sahan,H.Kodaz,andS.Günes,‘‘Anewclassificationmethod to diagnosis heart disease: supervised artificial immune system (AIRS),’’ in Proc. Turkish Symp. Artif.Intell.Neural Netw.(TAINN), 2005.

[11] S. Ozsen, S. Gunes, S. Kara, and F. Latifoglu, ‘‘Use of kernel functions in artificialimmune systems for then on linearclassificationproblems,’’IEEE Trans. Inf. Technol. Biomed., vol. 13, no. 4, pp. 621–628, Jul. 2009.

[12] N. Jankowski and V. Kadirkamanathan, ‘‘Statistical control of RBF-like networks for classification,’’ in Proc. Int. Conf. Artif.Neural Netw. Berlin, Germany: Springer, 1997, pp. 385–390.

[13] J Thomas, R Theresa Princy Human heart disease prediction system using data mining techniques(2016)

[14]Purushottam, KanakSaxena, Richa Sharma Efficient heart disease prediction system usingDecision tree (2015)

[15] V.Krishnaiah, G.Narsimha,N.Subhash Chandra Heart Disease Prediction System using Data Mining Techniques and Techniques and Intelligent Fuzzy Approach:A Review(February 2016)

Authorized licensed use limited to: University of Exeter. Downloaded on June 17,2020 at 19:49:49 UTC from IEEE Xplore. Restrictions apply.