



MULTIPLE CARDIO VASCULAR DISEASE PREDICTION USING NEURAL NETWORKS

ASWIN K¹, KANIMOZHI N², JEEVITHA M³, MATHUMITHA M⁴, DURGA K⁵

¹Department of CSE, Assistant Professor, Dhanalakshmi Srinivasan Engineering College ,Perambalur

²Department of CSE , UG Student, Dhanalakshmi Srinivasan Engineering College , Perambalur

³ Department of CSE , UG Student, Dhanalakshmi Srinivasan Engineering College , Perambalur

⁴Department of CSE , UG Student, Dhanalakshmi Srinivasan Engineering College , Perambalur

⁵Department of CSE , UG Student, Dhanalakshmi Srinivasan Engineering College , Perambalur

ABSTRACT

Machine learning is wide accustomed build the machine to be told and to predict once exposed to new knowledge. because of several advancements in machine learning, there square measure numerous ways which will be adopted to predict the center malady of a private. heart condition is one amongst the foremost diseases moving the individual round the world. There square measure many risk factors that ends up in heart condition. the mix of provision multivariate analysis and neural network provides a completely unique approach in predicting the center malady. at first provision regression is applied to pick out the foremost risk factors for predicting the malady. It produces the many risk factors that square measure helpful in predicting the center malady supported applied mathematics p-value. the danger factors that aren't having the many impact square measure known and removed. The resultant vital factors square measure provided as input to the neural network. And neural network is trained for the danger factors that's obtained from provision regression and accustomed check whether or not the person has the center malady or not. Thus, the combination of provision regression and neural network is applied in predicting the center malady.

Keywords: —Machine learning, Logistic regression model, neural networks, prediction, heart disease

INTRODUCTION

Cardiovascular disease (CVD) is that the leading reason behind death worldwide and a serious public health concern. Heart Disease is one amongst the leading sickness round the world. The effective functioning of the center plays a significant role within the body. There square measure many sorts of heart diseases like MI, cardiac muscle anaemia, noninheritable heart condition, Coronary heart condition, asystole, Peripheral heart condition etc., There square measure numerous methodologies accessible in predicting the center sickness. Machine learning may be a form of computing that produces the machines to find out from coaching knowledge and makes predictions on the check knowledge supported the learned knowledge. the essential plan behind the machine learning is to search out the patterns among the info and build the predictions. There square measure varied applications in machine learning like in recommender systems, diagnosis, bioinformatics etc., Basically, there square measure 3 forms of learning in machine learning like supervised learning, unsupervised learning and reinforcement learning. prognostic analytics embody numerous applied mathematics techniques from prognostic modeling, machine learning (ML) and data processing to form predictions supported the present or historical knowledge. the utilization of prognostic analytics square measure within the client relationship management, tending business and

in several different fields. Deep learning features a vital impact on the prognostic analytics. There square measure several models in the predictive modeling like Naive Thomas Bayes, logistical regression, Neural networks, Support Vector Machine, Classification and Regression trees etc., Artificial neural network (ANN) is one amongst the mathematical or recursive approach. it's just like the human brain neurons. the factitious neural network has connections, propagation direction and separate layers. every layer is created from nodes with the arrows that represents the interconnections between them. within the neural network, there's several nodes within the input layer. These input layer nodes square measure connected to the hidden layer nodes. every input is appointed with the weights. The input nodes within the network passes {the knowledge|the info|the information} to the nodes within the hidden layer that performs some tasks or computations and send the processed data to the output node. The output layer has the node that yields the ultimate result. this is often an outline of the method of neural network.

OVERVIEW OF WORK

The risk factors for Coronary sickness|heartcondition|cardiovascular disease} (CHD) or arteria coronaria disease, victimization the in-built imputation algorithm and particle swarm improvement. it's obtained that physical inactivity in addition forms the possibility issue of CHD. the selection rules unit utilized in predicting the possibility factors of cardiovascular disease. among the danger assessment in patients ill with symptom failure is finished by automatic classifier. victimization the live referred to as long sign variability, this classifier classifies the lower risk patients from higher risk patients automatically. Classification and Regression tree (CART) is helpful in classifying the higher risk patients and thus the lower risk patients. the middle rate variability is in addition major very important think about finding the possibility assessment of symptom failure. the possibility assessment of heart targeting the reduction of Coronary cardiovascular disease (CHD) events unit classified into before the event and once the event. The event before the CHD and once the event that unit non-modifiable and thus the modifiable unit best-known. The events unit animal tissue coronary intervention (PCI), infarct (MI) and blood vessel bypass graft surgery (CABG). The C4.5 decision tree algorithm is utilized for these three events of coronary cardiovascular disease. used association rules for prediction of cardiovascular disease. These association rules unit applied on the medical dataset and it generates many rules that unit orthogonal. thus on spot the foundations that unit truly essential for predicting the middle unhealthiness unit best-known by victimization search constraints that searches the association rules in coaching job dataset and eventually validates on the check set. The hybrid system is utilized with the planet improvement of genetic algorithm and this technique is utilized to initialize the neural network weights. A multilayered feed-forward network is utilized. The input nodes, hidden nodes and output nodes unit twelve, 10 and a combine of severally. The input nodes unit primarily the possibility factors that unit utilized in predicting the middle unhealthiness. the possibility factors of heart condition otherwise referred to as infarct is thought supported the selection trees and apriori algorithm supported these ways in which, the possibility factors that unit referred to as economical among the detection of heart condition unit pain, diabetes, smoking, gender and physical inactivity, age, lipids, sterol, glyceride, vital sign.

MULTI-LAYERED PERCEPTRON

In a multi-layered perceptron (MLP), perceptrons are organized in interconnected layers. The input layer collects input patterns. The output layer has classifications or output signals to that input patterns could map. as an example, the patterns could comprise a listing of quantities for technical indicators a few security; potential outputs might be "buy," "hold" or "sell." Hidden layers fine-tune the input weightings till the neural network's margin of error is borderline. it's hypothesized that hidden layers extrapolate salient options within the input file that have prophetic power relating to the outputs. This describes feature extraction, that accomplishes a utility the same as applied math techniques like principal element analysis. A neural network evaluates value knowledge and finds opportunities for creating trade choices supported the information analysis. The networks will distinguish delicate nonlinear interdependencies and patterns different ways of technical analysis cannot. in line with analysis, the accuracy of neural networks in creating value predictions for stocks differs. Some

models predict the right stock costs fifty to sixty p.c of the time, whereas others are correct in seventy p.c of all instances. Some have posited that a ten p.c improvement in potency is all associate capitalist will fire from a neural network.

PROPOSED METHOD

The ultimate goal is to mix the logistical regression model and neural network based mostly approach within the prediction of cardiovascular disease. the guts illness dataset has 303 observations of people out of that 297 observations are taken for thought. The planned system in the main consists of 2 elements. the primary half is to search out the necessary risk factors in predicting the guts illness from the accessible risk factors within the dataset supported the p-value. This p-value yields the numerous codes for every attribute. and therefore the second half is to divide the dataset into coaching and testing dataset. The neural network is build for the coaching dataset and therefore the learned neural network is in a position to predict the testing dataset. Data assortment the information employed in this project is obtained from the Cleveland cardiovascular disease info. a complete of 297 records with fourteen medical attributes that is employed to predict the guts illness. logistical Regression Model logistical regression model is one among the simple regression model and it's the capability to live the connection between the explicit variable and one or a lot of variable quantity. Here the freelance variables are age, sex, pain sort, resting force per unit area, humor sterol, abstinence blood glucose, resting electrographic results, most pulse rate, exercise elicited angina, oldpeak-slope of the height, slope of the height exercise, blood vessels affected, thal defect. The variable is that the category that is to be expected as healthy or having cardiovascular disease. within the binary logistical multivariate analysis, the variable is coded as "0" or "1" indicates that the individual having the guts illness or not severally. The logistical regression model computes the chance of the haert illness as a operate of the chance factors. we are able to reason the contingent probability $p(y=1 | X)$, wherever $X = (x_1, x_2 \dots x_n)$ represents n risk factors related to the illness. As a result we tend to may calculate the probability of a private suffering the illness. The cutoff price is set to zero.5. If the cutoff price is larger than zero.5, we are able to infer that the individual suffers from the guts disease; otherwise, the individual is free from the guts illness. except for this, the logistical regression model has the capability to pick out factors that have vital impact on the guts illness supported the applied mathematics significance p-value. The p-value is that the short kind for chance price and it's the chance that's given by the outline of the logistical regression model. applied mathematics hypothesis testing create use of p-values and it's employed in several fields of analysis like social science, social science etc., The p-value is outlined because the marginal significance to represent the chance of the prevalence of a given event. The logistical regression model is in a position to pick out the necessary risk factors or attributes that are employed in predicting the guts illness from the fourteen attributes of the dataset. By applying the logistical regression model, the chance factors or attributes that are having the pvalue but zero.05 ($p < 0.1$). The larger pvalue indicates the changes within the variable quantity aren't related to the changes within the variable. {the vital|themany|the numerous} risk factors that are obtained from applied mathematics significant p-value of logistical regression model are sex, pain sort, resting force per unit area, abstinence blood glucose, exercise elicited angina, slope of the height exercise, blood vessels affected and thal defect. coaching and Testing Dataset The dataset of 297 records are divided into coaching and testing dataset. The coaching dataset is employed to make a prognostic relationship and it's the set of examples that's used for learning and to suit weights of the classifier. The check set is ready of examples that is employed to judge the performance of a fully-specified classifier. The coaching and testing dataset is split into seventy fifth and twenty fifth severally. Building Neural Network The neural network may be a procedure model supported biological neural networks. Artificial neural networks (ANN) is predicated on observation of a person's brain. Human brain is extremely sophisticated internet of neurons . Analogically ANN is Associate in Nursing interconnected set of 3 units like input, hidden and output unit. In diagnosing, the patient's risk factors or attributes is employed as Associate in Nursing input. The effectiveness of artificial neural network was verified in medication. ANN are employed in predicting coronary cardiovascular disease. Here the input layer consisting of eight neurons corresponds to eight vital attributes. there's one output category variable that takes the worth either zero or one. {the price|the worth} zero represents that the individual isn't stricken by cardiovascular disease and therefore the value one represents that the individual suffers from cardiovascular disease. the amount of nodes employed in the hidden layer are three. the most advantage of neural networks is high accuracy. The applications of neural

network ar accounting, medicine , fraud detection etc. supported the learned network or coaching dataset, the neural network is in a position to predict the presence or absence of cardiovascular disease for the testing dataset.

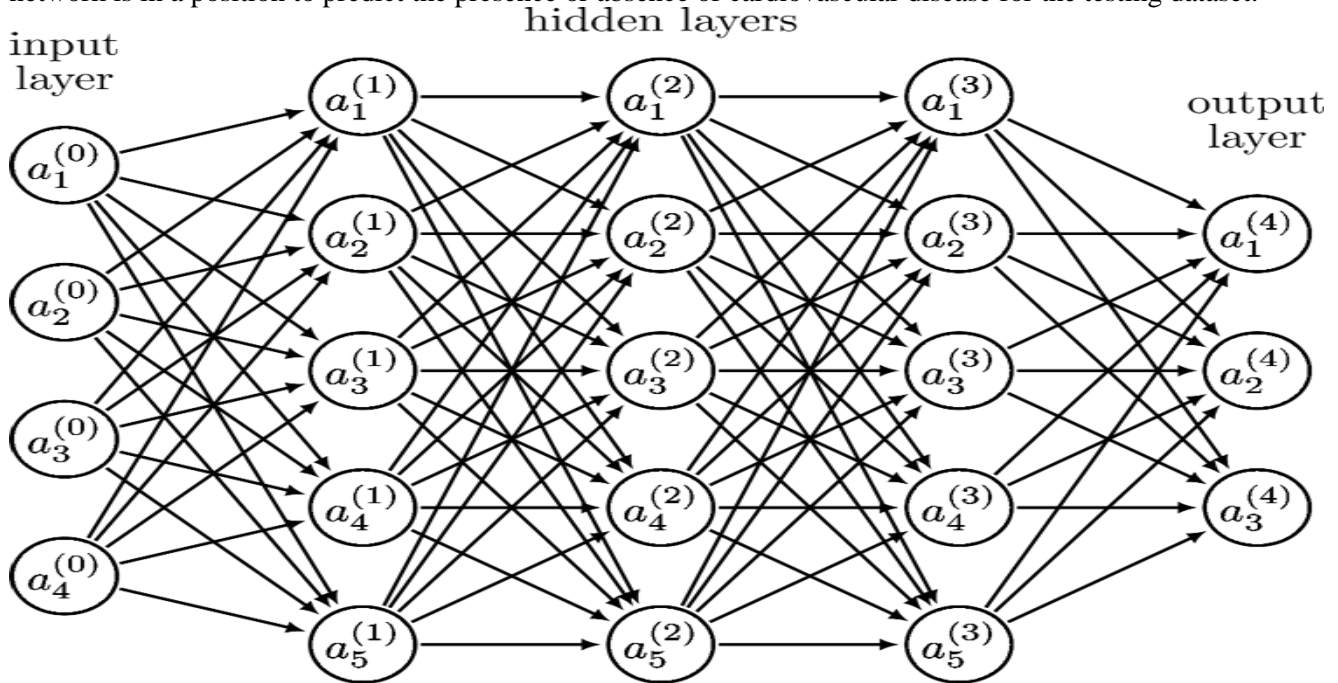


Fig : 1Neural Network

PERFORMANCE MEASURES

The performance measures of neural network are calculated mistreatment numerous measures like accuracy, specificity and sensitivity. The accuracy obtained by the neural network is eighty four. and also the sensitivity and specificity obtained are ninety one.4% and 78.5%.

Accuracy = $(TP+TN) / (TP+TN+FP+FN)$

Specificity = $TN / (TN+FP)$

Sensitivity = $TP / (TP+FN)$

where, TP = True Positive; that's the amount of samples that ar classified as having cardiopathy whereas they were even have cardiopathy. TN = True Negative; that's the amount of samples that are classified as not having cardiopathy whereas they were truly not. FN = False Negative; that's the amount of samples that ar classified as not having cardiopathy whereas they were even have cardiopathy FP = False Positive; that's the amount of samples that are classified as having cardiopathy whereas they were truly not. The accuracy, sensitivity and specificity of Neural Network is tabulated in Table.

Table 1. Accuracy, Sensitivity and Specificity of Neural Network

Classifier	Accuracy	Sensitivity	Specificity
Neural Network	87%	93.4%	76.5%

CONCLUSION

Machine learning is active analysis space during which several researchers square measure operating in aid domain for illness risk identification. The advantage of supplying regression is that the interpretability of model parameters and easy use. The advantage of neural network is it needs less formal applied math coaching to develop and might implicitly notice advanced non-linear relationships between dependent and freelance variables. the combination of supplying regression and neural network provides the novel approach in predicting the center illness of a personal. the long run work are often extended for longitudinal studies of the patients and to enhance the accuracy in prediction of heart condition.

REFERENCES

- [1] V. Sree Hari Rao, M. Naresh Kumar, "Novel Approaches for Predicting Risk Factors of Atherosclerosis," IEEE Journal of Biomedical and Health Informatics ., vol. 17, No. 1, Jan 2013.
- [2] Paolo Melillo, Nicola De Luca, Marcello Bracale and Leandro Pecchia , "Classification Tree for Risk Assessment in Patients Suffering From Congestive Heart Failure via Long-Term Heart Rate Variability", IEEE Journal of Biomedical and Health Informatics., Vol. 17, No. 3, May 2013.
- [3] Minas A. Karaolis, Joseph A. Moutiris, Demetra Hadjipanayi, Constantinos S. Pattichis, "Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees," IEEE Transactions on Information Technology in Biomedicine, Vol. 14, No. 3, May 2010.
- [4] Carlos Ordonez , "Association Rule Discovery With the Train and Test Approach for Heart Disease Prediction", IEEE Transactions on Information Technology in Biomedicine, Vol. 10, No. 2, April 2006.
- [5] Syed Umar Amin, Kavita Agarwal, Dr. Rizwan Beg, "Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors", Proceedings of IEEE Conference on Information & Communication Technologies, 2013.
- [6] Sikander Singh Khurl, Gurpreet Singh, "Ranking Early Signs of Coronary Heart Disease Among Indian Patients", IEEE International Conference on Computing for Sustainable Global Development, 2015.
- [7] The UCI Machine Learning Repository[online] <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>
- [8] Stephenie C. Lemon, Jason Roy, and Melissa A. Clark, Peter D. Friedmann, William Rakowski, "Classification and Regression Tree Analysis in Public Health: Methodological Review and Comparison With Logistic Regression", The Society of Behavioral Medicine, Vol. 26, No. 3, 2003, pp. 172-181.
- [9] David W. Hosmer, Jr., Stanley Lemeshow, Rodney X. Sturdivant, "Applied Logistic Regression", 2013.
- [10] Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction ", International Journal of Computer Applications, Vol. 17, No.8, March 2011, pp. 43-48.
- [11] HarleenKaur , Siri Krishan Wasan and Vasudha Bhatnagar, "The Impact of Data Mining Techniques on Medical Diagnostics", Data Science Journal, Vol. 5, October 2006, pp. 119-126.
- [12] K.Srinivas, B.Kavihta Rani , A.Govrdhan , "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks", International Journal on Computer Science and Engineering (IJCSE), Vol. 02, No. 02, 2010, pp. 250-255.

[13] Asha Rajkumar, G.Sophia Reena, "Diagnosis Of Heart Disease Using Datamining Algorithm", Global Journal of Computer Science and Technology, Vol. 10 , September 2010.

