ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 6, June 2021

Early Warning Heart Disease Prediction System Using Machine Learning

Ms. MS. Surya[1], L. Karthik[2], K. Santhiya[3], S. Pavithra[4], E. Asika[5]

[1]Assistant Professor, [2,3,4,5] Scholar, Department of Electronics and Communication Engineering,
Jansons Institute of Technology, Coimbatore, Tamil Nadu, India.

surya.m@jit.ac.in[1], karthiksurya611@gmail.com[2], santhiyakaliappan24@gmail.com[3],
pavithcharu@gmail.com[4], asikaelangovan@gmail.com[5]

Abstract: Cardiovascular infections are the most widely recognized reason for death worldwide throughout the recent years in the developed, developing and also underdeveloped nations. It is expanding at a quick rate in both elder and younger age people these days. In every one minute, nearly one individual dies of Heart Disease in India alone. To bring down the quantity of passing from heart sicknesses, there must be a quick proficient recognition strategy. The aim of our project is to protect human resources in clinical focuses and improve precision of determination. The early forecast of cardiovascular sicknesses can help in settling on choices on way of life changes in high danger patients and thusly decrease the confusions, which can be an extraordinary achievement in the field of medication. In our undertaking, we utilized the SVM idea for the model handling with highlights utilizing python and shown that from the basic information and pre-processed the highlights and expanded the exactness of the model.

Keywords: Heart Disease, ML, SVM, Data-visualization, Classification

I. INTRODUCTION

Heart Disease or Coronary illness is the main cause of death in all over the world. It is otherwise called as coronary artery disease. The main causes of Heart disease are diet, ageing, genetics, diabetes, smoking, gender, lifestyle and so on. The symptoms are pain in chest, trouble in breathing, feeling weak, palpitations, cyanosis, swelling of heart and Shortness of breath. It can be prevented by avoiding or reducing the stress, avoid smoking, by eating balanced meals which are low in cholesterol, doing regular exercises, keeping blood pressure normal and under control. This paper means to foresee future Coronary illness by investigating information of patients which groups if they have coronary illness utilizing ML algorithms.

II. DATA ANALYSIS IN CLINICAL FIELD

The important test in the present medical care is arrangement of best quality administrations and powerful precise determination. Regardless of whether heart illnesses are found as the great wellspring of death on the earth. The entire exactness in administration of an illness lies on the appropriate season of location of that infection. Records of enormous arrangement of clinical information made by clinical specialists are accessible for examining and separating important information from it. Information mining methods are the methods for removing important and concealed data from the huge measure of information accessible. For the most part the clinical data set comprises of discrete data. Thus, dynamic utilizing discrete information becomes unpredictable and intense assignment. ML handles huge scope very much organized informational index productively. In the clinical field, AI can be utilized for conclusion, location and expectation of different infections. The primary objective of our undertaking is to give an apparatus to specialists to identify coronary illness as beginning phase. This thus will assist with giving compelling treatment to patients and stay away from extreme outcomes. ML assumes a vital part to identify the secret discrete examples and along these lines examine the given information. After examination of information ML strategies help in coronary illness expectation and early conclusion.

III. DATABASE DESCRIPTION

This data set contains 76 characteristics, however completely distributed trials allude to utilizing a subset of 14 of them. Specifically, the Cleveland information base is the one in particular that has been utilized by ML specialists to this date. The “objective” field alludes to the presence of coronary illness in the patient. It is number esteemed from 0 to 4. Examinations with the Cleveland data set have focused

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 6, June 2021

on just endeavouring to recognize presence (values 1,2,3,4) from nonattendance (esteem 0). The names and federal retirement aide quantities of the patients were as of late eliminated from the data set, supplanted with faker qualities.

IV. METHODOLOGY

The Support Vector Machine is an extremely incredible and flexible AI model, equipped for performing straight or nonlinear characterization, relapse, and even exception identification. It is quite possibly the most famous models in AI, and anybody intrigued by AI ought to have it in their tool kit. SVMs are especially appropriate for arrangement of complex yet little or medium-sized Datasets. Straight SVM Characterization is the principal thought behind SVMs and is best clarified for certain photos. The two classes can unmistakably be isolated effectively with a straight line.

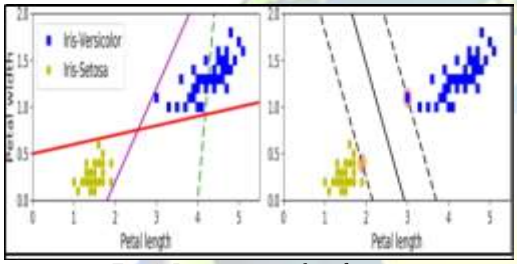

Fig. 1. Large margin classification

The left graph shows the choice limits of three potential direct classifiers. The model whose choice limit is addressed by the ran line is terrible to such an extent that it doesn't separate the classes appropriately. The other two models work completely on this preparation set, yet their choice limits come so near the examples that these models will presumably not proceed also on new occasions. Interestingly, the strong line in the plot on the privilege addresses the choice limit of a SVM classifier; this line isolates the two classes as well as stays as distant from the nearest preparing occurrences as could be expected. SVM classifier as fitting the largest conceivable road between the classes. This is called huge edge grouping. The adding of additional preparation examples "off the road" won't influence the choice limit by any stretch of the imagination: it is completely decided by the occasions situated on the edge of the road. These occasions are known as the help vectors. SVMs are delicate to the component scales, on the left plot, the upward scale is a lot bigger than the even scale, so the broadest conceivable road is near even. After highlight scaling, the choice limit looks much better.

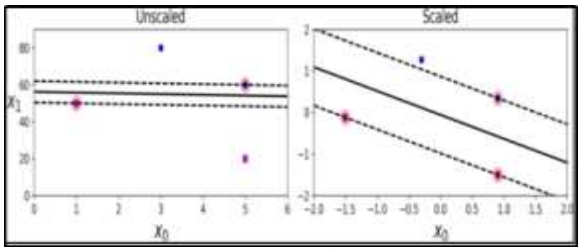

Fig .2. Sensitivity to feature scales

V. PROPOSED WORK

The primary goal of this paper is to viably anticipate if the patient influenced from coronary illness or not. To anticipate the likelihood of having coronary illness when the wellbeing proficient enters the info esteems from the patient's wellbeing report and the information is taken care of into model. To improve the exhibition of the order model to anticipate the patients yield precisely.

A. *Design Flow:*

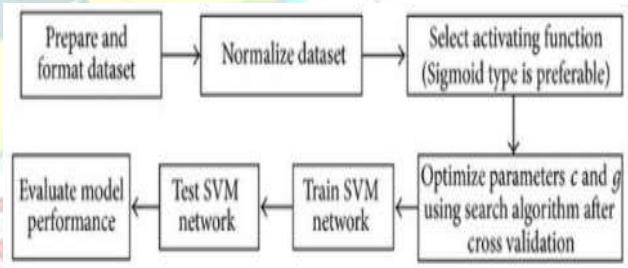

Fig .3. Work Flow

B. *Performance Measures:*

Assessing a classifier is frequently altogether trickier than assessing a regressor. The Stratified Fold class performs delineated examining to deliver folds that contain a delegate proportion of each class. At every emphasis the code makes a clone of the classifier, prepares that clone on the preparation creases, and makes expectations on the test overlap. At that point it tallies the quantity of right forecasts and yields the proportion of right expectations.

C. *Confusion Matrix:*

A greatly improved approach to assess the exhibition of a classifier is to take a gander at the disarray network. The overall thought is to tally the occasions examples of

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 6, June 2021

influenced are delegated not influenced. For instance, to know the occasions the classifier mistook for yes as no, FP and FN are thought of.

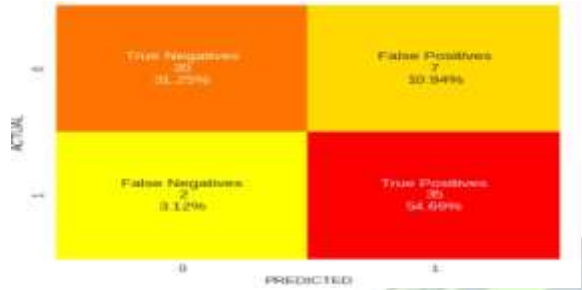

Fig .4. Confusion Matrix

D. Precision and Recall:

Scikit-Learn gives a few capacities to register classifier measurements, including exactness and review: It is regularly advantageous to join exactness and review into a solitary measurement called the F1 score, specifically on the off chance that you need a straightforward method to look at two classifiers. The F1 score is the consonant mean of accuracy and review though the normal mean treats all qualities similarly, the symphonises mean gives substantially more weight to low qualities. Subsequently, the classifier will possibly get a high F1 score if both review and accuracy are high.

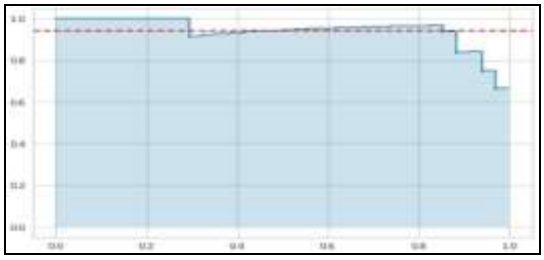

Fig .5. Precision Recall Curve

E. The ROC Curve:

The receiver operating characteristic (ROC) bend is another normal instrument utilized with double classifiers. It is basically the same as the accuracy/review bend, yet as opposed to plotting exactness versus review, the ROC bend plots the genuine positive rate against the bogus positive rate. The FPR is the proportion of negative occurrences that are mistakenly delegated positive. It is equivalent to one less the genuine negative rate, which is the proportion of negative occasions that are effectively delegated negative. The TNR is likewise called particularity. Consequently, the ROC bend plots affectability versus 1 – explicitness. To plot the ROC bend, to figure the TPR and FPR for different edge esteems, utilizing the ROC bend () work is required. By and by there is a trade-off: the higher the review (TPR), the falser positives (FPR) the classifier produces. The spotted line addresses the ROC bend of a simply irregular classifier; a decent classifier stays as distant from that line as could really be expected.

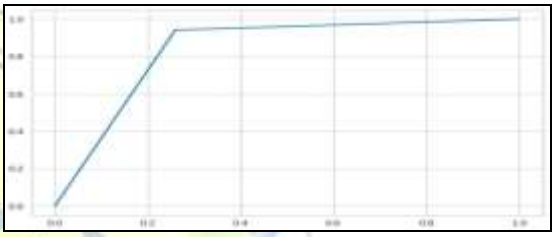

Fig .6. ROC curve

VI. Result And Discussion

The result shows that the SVM algorithm has a better accuracy of 85.93% as compared to other classification technique.

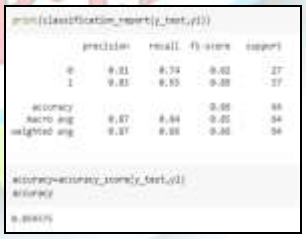

Fig .7. Accuracy

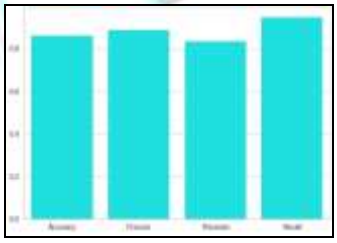

Fig .8. SVM Results

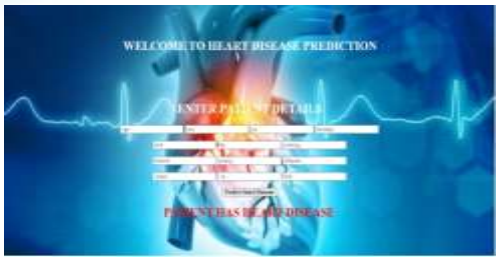

Fig .9. Predicting Patient who has Heart disease

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 6, June 2021

Fig .10. Predicting Patient who has no Heart disease

VII. CONCLUSION

The early guess of cardiovascular illnesses can help in settling on choices on way of life changes in high danger patients and thus diminish the confusions, which can be an extraordinary achievement in the field of medication. Coronary illness is one of the significant worries for society today. It is hard to physically decide the chances of getting coronary illness dependent on hazard factors. Nonetheless, AI procedures are valuable to anticipate the yield from existing information. This paper resolved the parameter tuning behind the models and successfully predict the heart disease, with 85.93% accuracy. The goal was successfully achieved by developing a Support Vector Machine algorithm maximizing the classification performance. Further for future enhancement, we can train on models and predict the types of cardiovascular diseases providing recommendations to the users, and also use more enhanced Model.

ACKNOWLEDGEMENT

Our sincere thanks to our project guide Assistant Professor Ms. MS. Surya for the valuable comments which greatly helpful to improve our project and the manuscript.

REFERENCES

[1]. Radhanath Patra, Bonomali Khuntia “Predictive Analysis of Rapid Spread of Heart Disease with Data Mining”, 978-1-5386-8158-9/19/$31.00, 2019IEEE.

[2]. M. A. Karaolis, J. A. Moutiris, D. Hadjipanayi and C. 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, pp. 559-566, May 2010.

[3]. Jesmin Nahara, Tasadduq Imama, Kevin S. Ticklea and Yi-Ping Phoebe Chen, "Computational intelligence for heart disease diagnosis: A medical knowledge driven approach", Expert Systems with Applications, vol. 40, no. 1, pp. 96-104, 2013.

[4]. A. Rairikar, V. Kulkarni, V. Sabale, H. Kale and A. Lamgunde, "Heart disease prediction using data mining techniques", 2017 International Conference on Intelligent Computing and Control (I2C2), pp. 1-8, 2017.

[5]. S. Babu, E.M. Vivek, K.P. Famina, K. Fida, P. Aswathi, M. Shanid, et al., "Heart disease diagnosis using data mining technique", 2017 International Conference of Electronics Communication and Aerospace Technology (ICECA), pp. 750-753, 2017.

[6]. E. O. Olaniyi, O. K. Oyedotun, A. Helwan and K. Adnan, "Neural network diagnosis of heart disease", *2015 International Conference on Advances in Biomedical Engineering (ICABME)*, pp. 21-24, 2015.

[7]. C. Sowmiya and P. Sumitra, "Analytical study of heart disease diagnosis using classification techniques", *2017 IEEE International Conference on Intelligent Techniques in Control Optimization and Signal Processing (INCOS)*, pp. 1-5, 2017.

[8]. M. A. Jabbar and S. Samreen, ”Heart disease prediction system based on hidden nave bayes classifier,” 2016 International Conference on Circuits, Controls, Communications and Computing (I4C), Bangalore, 2016, pp. 1-5.

[9]. Nidhi Bhatla and Kiran Jyoti, “An Analysis of Heart Disease Prediction Dierent Data Mining Techniques”, International Journal of Engineering Research & Technology (IJERT), 2012, 2278-0181.

[10]. Chitra R and Seenivasagam V, “Review of Heart Disease Prediction System Using Data Mining Technique and Hybrid Intelligent Techniques”, ICTACT journal on Soft Computing, 2013, 2229-6956(ONLINE), p. 605-608.