International Journal of Research in Engineering, Science and Management
Volume-2, Issue-4, April-2019
www.ijresm.com | ISSN (Online): 2581-5792

Heart Disease Prediction and Medicine Prescription using SVM

D. Harini[1], S. Akash[2], R. S. Vishnu Durai[3], J. Archana[4]
[1,2,4]*Student, Department of Computer Science and Engg., Sri Ramakrishna Engg. College, Coimbatore, India*
[3]*Assistant Professor, Dept. of Computer Science and Engg., Sri Ramakrishna Engg. College, Coimbatore, India*

***Abstract*: Data mining is the most popular knowledge extraction method for knowledge discovery. Machine learning is used to enable a program to analyze data, understand correlations and make use of insights to solve problems and enrich data and for prediction. Data mining techniques and machine learning algorithms play a very important role in medical area. The health care industry contains a huge amount of data. But most of it is not effectively used. Heart disease is one of the main reason for death of people in the world. Nearly 47% of all deaths are caused by heart diseases. We use KNN, Support Vector Machine to predict the heart diseases. Accuracy of the prediction level is high when using more number of attributes. Our aim is to perform predictive analysis using these data mining, machine learning algorithms on heart diseases and analyse the various mining, Machine Learning algorithms used and conclude which techniques are effective and efficient.**

***Keywords*: classification and prediction, convolutional neural network, support vector machine learning.**

1. Introduction

The heart is the hardest working muscle in the body. The average heart beats 100,000 times a day, day and night, to supply oxygen and nutrients throughout the body. Blood pumped by the heart also shuttles waste products such as carbon dioxide to the lungs so it can be eliminated from the body. Proper heart function is essential to support life. Coronary artery disease (CAD), commonly known as heart disease, is a condition in which cholesterol, calcium, and other fats accumulate in the arteries that supply blood to the heart. This material hardens forming a plaque that blocks blood flow to the heart. When a coronary artery narrows due to plaque build up or some other cause, the heart muscle is starved for oxygen and a person experiences chest pain known as angina. The usage of information technology in health care industry is increasing day by day to aid doctors in decision making activities. It helps doctors and physicians in disease management, medications and discovery of patterns and relationships among diagnosis data. Current approaches to predict cardiovascular risk fail to identify many people who would benefit from preventive treatment, while others receive unnecessary intervention. Machine-learning offers opportunity to improve accuracy by exploiting complex interactions between risk factors. We assessed whether machine-learning can improve cardiovascular risk prediction. “Cardiovascular disease is the leading cause of illness and death worldwide,” said Dr. Stephen Weng, of Nottingham University’s National Institute for Health Research School. Our study shows that artificial intelligence could significantly help in the fight against it by improving the number of patients accurately identified as being at high risk and allowing for early intervention by doctors to prevent serious events like cardiac arrest and stroke.” Based on their results, it is clear that artificial intelligence and machine learning techniques have a key role in fine-tuning risk management strategies for individual patients. This paper is organized as follows: Section 2 gives the classification of machine learning prediction techniques. Section 3 describes the work in the literature regarding classification algorithms for medical diagnosis of cardiovascular and heart diseases. Section 4 concludes the survey. We thus propose to collect relevant data pertaining all elements related to our field of study, train the data as per the proposed algorithm of machine learning and predict how strong is there a possibility for a patient to contract a heart disease. For the purpose of patients entering data, we suggest to make use of the easily available sensors in watches and cell phones to measure the simple factors. In the section II, we discuss what all sensors are prevalent in the market and what all symptoms do they measure. Our main aim behind developing the system is to make it user friendly so that regular monitoring of the patient is made possible. Thus it is of utmost importance that the factors required in the input are most accurate and easily available.

2. Related works

To initiate with the work, we have started collecting data in each and every aspect towards the goal of the system. In the first place, the research was in the direction of the main causes or the factors which have strong influence on the heart health. Some factors are unmodifiable like age, sex and family background but there are some parameters like blood pressure, heart rate etc. which can be kept in control by following certain measures [4]. Many doctors suggest healthy diet and regular exercise to keep the heart healthy. Following are the parameters which are considered for the study in designing the system which have major risk percentage with respect to CAD [5]. The research on the classification and prediction of heart disease

have increased to a greater extent during the past decade. The related works vary with various datasets used for the respective classification approaches. Wisconsin's dataset is a popular dataset used to predict if the disease is benign or malignant using various features like clump thickness, size uniformity, etc. Similar kind of classification is done through the Open ML Heart disease dataset. The dataset is in the form of CSV files. It uses various algorithms such as k-means clustering, linear regression, decision tree and artificial neural networks. The features in the Wisconsin's dataset is computed from digitized image of a fine needle aspirate (FNA) of a breast mass. They describe the cell nuclei process in the image. This is a process called cytology. This considers classification done among the cells obtained using chemical methods.

In the other hand, the study of histopathology includes the entire tissue for classification. In former classification techniques using these data's obtained from the histopathological images only two classes were classified that is benign or malignant. Later, the Break his dataset provided the multiclass classification of the heart disease. This helps in accurate diagnosis of the disease. The research works related in the binary classification always provide more accurate results than the multiclass classification. The multiclass classification methods included convolutional neural networks including deep learning mechanisms. It is performed through various accuracy were obtained when applied to the augmented datasets.

Once after the multi disease classification of heart disease has become a leading topic in the research field, the classification of the various stages of disease came into place, But the datasets are kept confidential by the ICIAR 2018- BACH Grand challenge organisation. The datasets can be used only with prior permission from the organisation. This dataset histopathological images belonging to four stages of disease that is benign, In situ, Invasive and normal. The methods used for classification of these datasets including the stain normalization and data pre-processing by obtaining patches from each image. This dataset has a high resolution of 2040*1536 pixels, hence its clarity is much higher. A transfer learning based approach for classification of the H&E stained images is provided and the network uses architectures such as inception V3, Resnet50 which is pre trained on Image Net for classification. These approaches fail to identify the possibility that the disease cells are present in only a small part of the image and the rest of the image appears to be normal or benign.

3. Proposed system

The contribution of the proposed research is to design a machine-learning-based medical intelligent decision support system for the diagnosis of heart disease. In the present study, various machines learning predictive models such as logistic regression, k-nearest neighbor, ANN, SVM, decision tree, Naive Bayes, and random forest have been used for classification of people with heart disease and healthy people. Three feature selection algorithms, Relief, minimal-redundancy-maximal-relevance (mRMR), Shrinkage and Selection Operator (LASSO), were also used to select the most important and highly correlated features that great influence on target predicted value. Cross-validation methods like k-fold were also used. In order to evaluate the performance of classifier, various performance evaluation metrics such as classification accuracy, classification error, specificity, sensitivity, Matthews' correlation coefficient (MCC), and receiver optimistic curves (ROC) were used. Additionally, model execution time has also been computed. Moreover, data preprocessing techniques were applied to the heart disease dataset. The proposed system has been trained and tested on Cleveland heart disease dataset, 2016.

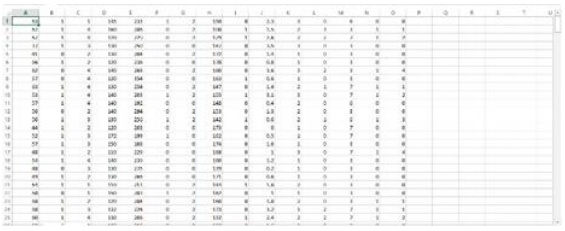

Fig. 1. Heart disease dataset

UCI data-mining repository the dataset of Cleveland heart disease is available online. All the computations were performed in Python on an Intel(R) Core™ i5-2400CPU @3.10 GHz PC. The major contributions of the proposed research work are as follows:(a)All classifiers' performances have been checked on full features in terms of classification accuracy and execution time.(b)The classifiers' performances have been checked on selected features as selected by feature selection (FS) algorithms Relief, mRMR, and LASSO with k-fold cross-validation.(c)The study suggests which feature algorithm is feasible with which classifier for designing high-level intelligent system for heart disease that accurately classifies heart disease and healthy people. We use 80% images for training and 20% images for validation.

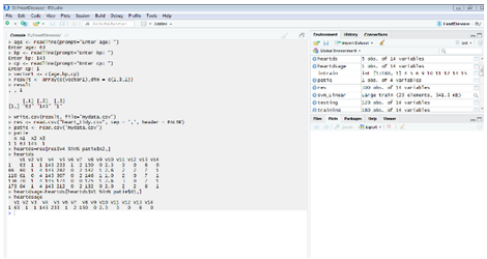

Fig. 2. Basic inputs

Machine Learning (ML) delivers methodologies, approaches, and apparatuses that can help resolving analytic and predictive hitches in a miscellany of medicinal areas. ML is being used for the inquiry of the wild of controlled edges and their mixtures for forecast, e.g. forecast of illness development, removal of medicinal information for consequence

investigation, treatment guidance and provision, and for the overall enduring organization. ML is also being used for statistics examination, such as discovery of proportions in the data by rightly commerce with flawed data, clarification of incessant data used in the Strenuous Care Unit, and brainy troubling subsequent in real and ordered nursing. It is contended that the successful presentation of ML attitudes can help the tally of computer-based structures in the healthcare setting providing chances to ease and enhance the exertion of medical boffins and eventually to recover the competence and excellence of medicinal repair. Below, it précis some main ML requests in medicine. Machine Learning learns the data and produces the result.

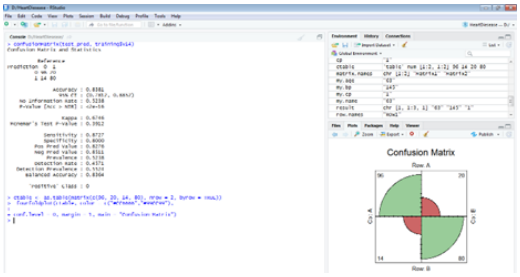

Fig. 3. Confusion matrix

Considering the staging of disease the dataset contains only less images for training and the dimensions are large in size (2040*1536 pixels). There is a complexity requiring more data to generalise well and to avoid over fitting due to only less number of images available. The training data is done using the same process as described before for the classification of Break His dataset. The convolutional layers are increased in case of the classification of BACH dataset in order to increase accuracy. But in case of Break his dataset, increase in layers lead to decrease in accuracy.

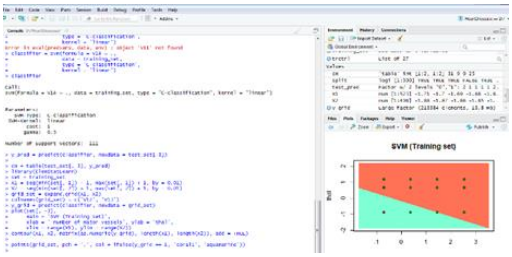

Fig. 4. SVM training set

4. Experiments and results

It has presented an article on an intelligent system for automated breast cancer diagnosis & prognosis using SVM diagnosis of breast cancer disease. Wisconsin diagnostic breast cancer datasets were used to implement SVM model to provide distinction between the malignant & benign breast masses. The article provides the implementation details along with the corresponding results for all the assessed classifiers. Several comparative studies have been carried out concerning both the prognosis and diagnosis problem demonstrating the superiority of the proposed SVM algorithm in terms of sensitivity, specificity and accuracy. It has presented an article on early detection of breast cancer using SVM classifier technique. We have discussed, how to detect tumour from mammograms. In this article the authors have specified an algorithm for tumour detection and have proposed the method that includes the mammograms image, which were filtered with Gaussian filter based on standard deviation and matrix dimensions such as rows and columns. Then the filtered image was used for contrast stretching. The reconstructed image is used for segmentation. Thresholding method was used for segmentation and then the features were extracted from the tumours area. This method can be summarized as the initial step based on grey level information of image enhancement.

SVM is a powerful method for building a classifier. It aims to create a decision boundary between two classes that enables the prediction of labels from one or more feature vectors. This decision boundary, known as the hyperplane, is orientated in such a way that it is as far as possible from the closest data points from each of the classes. These closest points are called support vectors.

The SVM algorithm was originally proposed to construct a linear classifier in 1963 by Vapnik. An alternative use for SVM is the kernel method, which enables us to model higher dimensional, non-linear models. In a non-linear problem, a kernel function could be used to add additional dimensions to the raw data and thus make it a linear problem in the resulting higher dimensional space. Briefly, a function could help do certain calculations faster which would otherwise would need computations in high dimensional space.

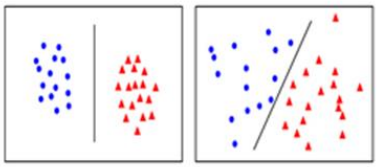

Fig 5. Confusion matrix

5. Conclusion

Heart attack is crucial health problem in human society. This paper has summarized state of art techniques and available methods for predication of this disease. Deep learning an emerging area of artificial intelligence showed some promising result in other field of medical diagnose with high accuracy. It is still an open domain waiting to get implemented in heart disease predication. Some methods of deep learning have been discussed which can be implemented for heart disease predication, along with pioneer machine learning algorithms. An analytical comparison has been done for finding out best available algorithm for medical dataset. In future our aim is to carry forward the work of temporal medical dataset, where dataset varies with time and retraining of dataset is required.

References

[1] Jyoti Soni, Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction, International Journal of Computer Applications, Volume 17– No.8, March 2017.
[2] P.K. Anooj, Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules, Journal of King Saud University – Computer and Information Sciences (2012) 24, 27–40.
[3] Nidhi Bhatla and Kiran Jyoti, An Analysis of Heart Disease Prediction using Different Data Mining Techniques, International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 8, October – 2012.
[4] M. Anbarasi, E. Anupriya, N.Ch.S.N.Iyengar, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm", International Journal of Engineering Science and Technology Vol. 2(10), 2016, 5370-5376.
[5] M. Akhil Jabbar, Priti Chandra, B. L Deekshatulu "Heart Disease Prediction System using Associative Classification and Genetic Algorithm", ICECIT, 2015.
[6] Hegner, Barbara R. dan Esther Caldwell, Asisten Keperawatan: Suatu Pendekatan Proses Keperawatan (Terjemahan). Jakarta: Penerbit Buku Kedokteran EGC, 2003.
[7] Hjalmarson, A. (2007). Heart rate: an independent risk factor in. European Heart Journal Supplements.
[8] Jan A. Kors, Cees A. Swenne, Karin H. Greiser. (2007). Cardiovascular disease, risk factors, and heart rate variability in the general population. Journal of Electro cardiology.
[9] R. Sethukkarasi and Kannan, "An Intelegent System for Mining Temporal rules in Clinical database using Fuzzy neural network", European Journal of Scientific Research, Vol. 70(3), pp 386-395, 2012.
[10] Yahya, A. F, Menaklukan Pembunuh No. 1: Mencegah dan Mengatasi Penyakit Jantung Koroner secara Tepat dan Cepat. Bandung, 2010.