JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

USING THE SVM ALGORITHM TO PREDICT AND ANALYZE HEART DISEASE

T.V.Ramana

Department of computer science and engineering, Jain University, Bangalore, India

Abstract

One of the most exciting and difficult tasks is predicting heart disease through data mining. The lack of professionals and the significant number of incorrectly diagnosed cases have prompted the development of a quick and efficient detection method. According to previous systems, combining clinical decision assistance with a computer-based patient record can help to reduce medical errors, improve precision, and hence improve patient safety. We are developing a method that can aid in the prediction of cardiac disease by taking into account risk factors connected with the disease. The system uses a support vector machine algorithm on the patient's historical information/data and gives features such as age, sex, smoking, obesity, alcohol intake, bad cholesterol, blood pressure, and heart rate to make a more accurate prediction of coronary heart disease.

Keywords: prediction, heart disease, medical, mining, cardiovascular

1. Introduction

The efficient functioning of the heart is absolutely necessary for life. The term "heart disease" refers to a problem with the heart's blood vessel system. The heart is a vital organ in the human body [1]. If the body's blood circulation is inadequate, the brain and heart organs stop operating, and death happens in a matter of minutes [2,3]. Since the past 15 years, heart disease has been the top cause of death worldwide [4,5]. Age, family history, sex, stress, high cholesterol, heart rate, smoking, alcohol intake, overweight, physical inactivity, chest pain type, and poor diet are all known risk factors. It is feasible to isolate the record and produce a report on HD if it is positive or negative using information gained by studying the patient's history record.

Heart disease is the leading cause of death worldwide. Patient billing, inventory management, and the generation of simple statistics are all supported by many hospital information systems [6]. Decision support systems are used in some hospitals, however they are mostly limited. [7]Mining is a way of analysing large amounts of data to find hidden patterns and previously unknown links, as well as knowledge detection, to aid in the better comprehension of medical data and the prevention of heart disease [8]. If the classification of coronary heart disease is automated, it can be

beneficial to medical practitioners in terms of finding a speedy and accurate conclusion. The presence of cardiac disease can help individuals live longer. Using Support Vector Machines, the study incorporates the classifications of Heart Disease (SVM) [9,10]. This is a medical decision-making paradigm for identifying coronary artery disease in a rational, purposeful, precise, and timely manner.

2. Review of the Literature

The research is focused on the various strategies for detecting HD. For heart disease prediction models, a variety of technologies and extensive surveys are available.

Naive bayes (NB), Decision tree (DT), Neural network (NN), Genetic algorithm (GA), Artificial intelligence (AI), and Clustering algorithms like KNN and Support vector machine are just a few of the classification approaches available (SVM). The study presents [11] prediction models that use individual strategies as well as combinations of two or more techniques.

Miss. Chaitrali S et al.[12] used a neural network to construct the Heart Disease Prediction System (HDPS). The HDPS system predicts whether or not a patient will develop heart disease. The technology makes predictions based on 13 medical characteristics such as sex, blood pressure, and cholesterol. Obesity and smoking are now incorporated as additional criteria for improved accuracy.

On the basis of an actual dataset, D. Mendes et al. [13] provide a simple and interpretable model. It is made out of a decision tree model structure that employs a smaller number of binary risk factors. The justification is based on a recent dataset provided by the Portuguese Society of Cardiology, which had 77 risk factors originally.

A common feature selection technique for Heart Disease Prediction is presented in this[14] study. The fuzzy measure and the relevant nonlinear integral are used to achieve good results. The fuzzy measure's none additiveness reflects the significance of the feature qualities as well as their interactions. It can forecast the likelihood of individuals developing heart disease based on factors such as age, sex, blood pressure, and blood sugar. As a result, the accuracy improves while the computing time decreases.

Fizar Ahmed et al.[15] provide the architecture for heart rate and other data monitoring techniques, as well as how to apply a machine learning technique such as the kNN classification algorithm to predict heart attacks using a set of heart rate data and other cardiac parameters.

3. Framework that has been proposed

Overview

The biggest cause of death worldwide is cardiovascular disease. A healthy heart rate ranges from 60 to 100 beats per minute. However, a resting heart rate of more than 76 beats per minute has been related to an increased risk of heart attack. An erratic heartbeat may not necessarily indicate a heart attack [16]. However, if it's a new symptom, or if you experience chest pains or difficulty breathing, it could be a sign of a heart attack. The architecture design of the suggested system, which we may use to predict the disease based on the parameters listed, is shown in Figure.

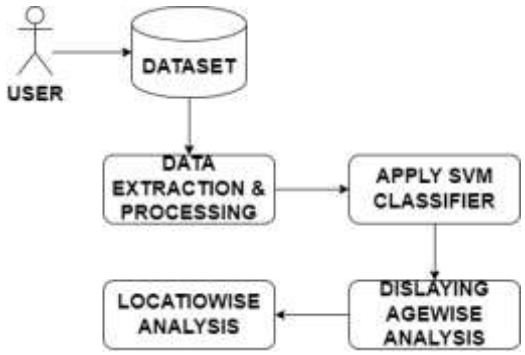

Fig 1. Process flow of the system

Overview of the Project

According to previous systems, combining clinical decision assistance with a computer-based patient record can help to reduce medical errors, improve precision, and hence improve patient safety [17. We are developing a method that can aid in the prediction of cardiac disease by taking into account risk factors connected with the disease. This is where we acquire the patient's historical information/data. Age, sex, smoking, obesity, alcohol intake, bad cholesterol, blood pressure, and heart rate are used to apply the support vector machine algorithm to features like age, sex, smoking, obesity, alcohol intake, bad cholesterol, blood pressure, and heart rate to make a more accurate prediction of coronary heart disease [18].

Patient Database

The Cleveland Clinic Foundation is a non-profit organisation based in Cleveland, Ohio (cleveland.data)

Budapest's Hungarian Institute of Cardiology (hungarian.data)

Long Beach, CA: V.A. Medical Center (long-beach-va.data)

4. Zurich University Hospital, Switzerland (switzerland.data). The instance format is the same for all databases. Only 14 of the 76 raw properties in the databases are actually used. As a result, I've made two copies of each database: one with all of the attributes and one with the 14 attributes that have been used in previous studies [19].

Data Extraction

Determine which factors are prevalent in the same heart disease patient by identifying or studying data from heart disease patients. Experiments were carried out using the Weka 3.6.0 tool [20]. A 1000-record data collection with eight properties is used. This section contains the results of our experimental research for identifying significant patterns for heart attack prediction.

SVM classifier

Many researchers have successfully employed Vapnik and Cortes' support vector machine (SVM) for gender classification tasks [21]. The separation hyper plane in an SVM classifier is chosen to minimise the expected classification error of the unseen test patterns.

SVM is a powerful classifier that can distinguish between two groups. SVM assigns the test picture to the class with the greatest distance to the training's nearest point [22].

The SVM training process created a model that can predict whether a test image belongs to this or another class [23]. Even if we limit ourselves to single pose (frontal) detection, SVM requires a large quantity of training data to identify an affective decision border, and the computational cost is very expensive.

The SVM is a classification learning method. It seeks to find the best separation hyper plane for unseen patterns such that the expected classification error is as low as possible[24]. The input is transferred to a high-dimensional feature space where they can be separated by a hyper plane for linearly non-separable data. Kernels are used to efficiently project data into a high-dimensional feature space. The SVM seeks to find the optimal separating hyper plane defined by the equation WTx+b that maximises the distance between the two classes given a set of training samples and the accompanying decision values -1, 1.

4. Upshot

Analyze the sensor data on the server using the SVM data mining technique, which is appropriate for our situation. Conclusions on the most effective model, the usefulness of conjoint models, and the overall model's accuracy can be inferred from them.

5. Conclusion

By its very nature, heart disease is a lethal disease. This condition can lead to life-threatening complications like heart attacks and death. The value of data mining in the medical domain has been recognised, and measures have been done to implement applicable approaches in disease prediction. We're putting in place a system that will help anticipate heart disease based on a patient's clinical data related to a risk factor for heart disease. We can predict whether individuals will get heart disease or not by analysing medical datasets such as age, sex, blood pressure, obesity, and blood sugar and applying SVM classifier. Furthermore, the SVM's classification accuracy, sensitivity, and specificity have all been proven to be excellent, giving it a superior option for diagnosis. We're also analysing the data to see when it occurs most frequently and which regions are affected by the condition. As a result, precautions can be taken to avert death from heart disease.

References

1) Krishna, N. M., Sekaran, K., Vamsi, A. V. N., Ghantasala, G. P., Chandana, P., Kadry, S., ... & Damaševičius, R. (2019). An efficient mixture model approach in brain-machine interface systems for extracting the psychological status of mentally impaired persons using EEG signals. *IEEE Access*, *7*, 77905-77914.
2) Patan, R., Ghantasala, G. P., Sekaran, R., Gupta, D., & Ramachandran, M. (2020). Smart healthcare and quality of service in IoT using grey filter convolutional based cyber physical system. *Sustainable Cities and Society*, *59*, 102141.

3) Reddy, A. R., Ghantasala, G. P., Patan, R., Manikandan, R., & Kallam, S. Smart Assistance of Elderly Individuals in Emergency Situations at Home. *Internet of Medical Things: Remote Healthcare Systems and Applications*, 95.

4) G S Pradeep Ghantasala, D. Nageswara Rao, Mandal K (2021) MACHINE LEARNING ALGORITHMS BASED BREAST CANCER PREDICTION MODEL. *Journal of Cardiovascular Disease Research*, 12 (4), 50-56. doi:10.31838/jcdr.2021.12.04.04

5) Ghantasala, G. P., Reddy, A. R., & Arvindhan, M. Prediction of Coronavirus (COVID-19) Disease Health Monitoring with Clinical Support System and Its Objectives. In *Machine Learning and Analytics in Healthcare Systems* (pp. 237-260). CRC Press.

6) Chandana, P., Ghantasala, G. P., Jeny, J. R. V., Sekaran, K., Deepika, N., Nam, Y., & Kadry, S. (2020). An effective identification of crop diseases using faster region based convolutional neural network and expert systems. *International Journal of Electrical and Computer Engineering (IJECE)*, *10*(6), 6531-6540.

7) Bhowmik, C., Ghantasala, G. P., & AnuRadha, R. (2021). A Comparison of Various Data Mining Algorithms to Distinguish Mammogram Calcification Using Computer-Aided Testing Tools. In *Proceedings of the Second International Conference on Information Management and Machine Intelligence* (pp. 537-546). Springer, Singapore.

8) Ghantasala, G. P., Kumari, N. V., & Patan, R. (2021). Cancer prediction and diagnosis hinged on HCML in IOMT environment. In *Machine Learning and the Internet of Medical Things in Healthcare* (pp. 179-207). Academic Press.

9) Kumari, N. V., & Ghantasala, G. P. (2020). Support Vector Machine Based Supervised Machine Learning Algorithm for Finding ROC and LDA Region. *Journal of Operating Systems Development & Trends*, *7*(1), 26-33.

10) G. S. Pradeep Ghantasala, Nalli Vinaya Kumari. Mammographic CADe and CADx for Identifying Microcalcification Using Support Vector Machine. Journal of Communication Engineering & Systems. 2020; 10(2): 9–16p.

11) Salma Banu N.K,Suma Swamy"Prediction of Heart Disease at early stage using Data Mining and Big Data Analytics: A Survey",2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT),2016 IEEE.

12) Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte,"A DATA MINING APPROACH FOR PREDICTION OF HEART DISEASE USING NEURAL NETWORKS",INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING and TECHNOLOGY,2012.

13) D. Mendes, S. Paredes, T. Rocha, P. Carvalho "Assessment of Cardiovascular Risk based on a Data-driven Knowledge Discovery Approach",2015 IEEE.

14) SunitaSoni, JyotiSoni,UjmaAnsari,Dipesh Sharma, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", International J ournal of Computer Application (IJCA, 0975 8887) Volume 17 No.8, March 2011.

15) Fizar Ahmed,"An Internet of Things (IoT) Application for Predicting the Quantity of Future Heart Attack Patients " International Journal of Computer Applications (0975 8887) Volume 164 No6, April 2017.

16) Kishore, D. R., Syeda, N., Suneetha, D., Kumari, C. S., & Ghantasala, G. P. (2021). Multi Scale Image Fusion through Laplacian Pyramid and Deep Learning on Thermal Images. *Annals of the Romanian Society for Cell Biology*, 3728-3734.

17) "A survey on Microcalcification identification and classification using CAD System", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.2, Issue 5, page no.186-190, MAY-2015, Available :http://www.jetir.org/papers/JETIR1805783.pdf

18) Ghantasala, G. P., Tanuja, B., Teja, G. S., & Abhilash, A. S. (2020). Feature Extraction and Evaluation of Colon Cancer using PCA, LDA and Gene Expression. *Forest*, *10*(98), 99.

19) Ghantasala, G. P., Reddy, A., Peyyala, S., & Rao, D. N. (2021). Breast Cancer Prediction In Virtue Of Big Data Analytics. *INTERNATIONAL JOURNAL OF EDUCATION, SOCIAL SCIENCES AND LINGUISTICS*, *1*(1), 130-136.

20) Ghantasala, G. P., & Kumari, N. V. (2021). Breast Cancer Treatment Using Automated Robot Support Technology For Mri Breast Biopsy. *INTERNATIONAL JOURNAL OF EDUCATION, SOCIAL SCIENCES AND LINGUISTICS*, *1*(2), 235-242.

21) Ghantasala, G. P., Kallam, S., Kumari, N. V., & Patan, R. (2020, March). Texture Recognization and Image Smoothing for Microcalcification and Mass Detection in Abnormal Region. In *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)* (pp. 1-6). IEEE.

22) Ghantasala, G. P., & Kumari, N. V. (2021). Identification of Normal and Abnormal Mammographic Images Using Deep Neural Network. *Asian Journal For Convergence In Technology (AJCT)*, *7*(1), 71-74.

23) MANDAL, K., GHANTASALA, G. P., KHAN, F., SATHIYARAJ, R., & BALAMURUGAN, B. (2020). Futurity of Translation Algorithms for Neural Machine Translation (NMT) and Its Vision. *Natural Language Processing in Artificial Intelligence*, 53.

24) Sreehari, E., & Ghantasala, P. G. (2019). Climate Changes Prediction Using Simple Linear Regression. *Journal of Computational and Theoretical Nanoscience*, *16*(2), 655-658.