PREDICTION OF RISK DEVELOPING CHRONIC DISEASE SYSTEM USING MACHINE LEARNING

[1]Dr.K.Murali Babu, [2]Dr.V.Venu Gopal, [3]A Benargy, [4]Bandla Venkatasiva Ayyappa
[1,2] Professor, [3]Assistant Professor, [4]Student, Dept. of Computer Science Engineering, Newton's Institute of Engineering, Macherla, Andhra Pradesh, India.

ABSTRACT

In most cases, it is hard to foresee chronic diseases before they impact a person. As a consequence, we employ Machine Learning to create a model that allows us to calculate the likelihood that an individual would have a condition like diabetes, chronic kidney cancer, or a heart attack based on information about their health like their haemoglobin and glucose levels. You or a loved one may have thought on multiple occasions that you or they need emergency medical help, but for a number of reasons, they are unable to provide it.

A project called the Chronic Disease Prediction application aims to offer online counseling and end-user assistance. In this article, we propose a web tool that enables users to predict their risk of developing a disease, whether it be typical or extraordinary and to obtain fast details of an expert for help. The programmed is provided with information about chronic diseases such as heart disease. The user can input health parameters for prediction in this programmed. At this point, certain subtle elements are formed for the client to look for various ailmentsthat may be related to him. Here we use some clever information mining techniques to find the first precise disease that can be related to the patient's interests. Depending on the result, they cansimilarly contact a specialist for further treatment. The frame allows the client to see the doctor's points of interest as well.

INTRODUCTION

The objectives of "Recognize chronic diseases through machine learning" are:

a. Chronic diseases generally cannot be predicted before they affect humans. So we are using machine learning to create a model that will help us predict the probability that will contracta disease such as a heart attack.

b. To reduce the risk of sudden health attacks from a chronic disease.

c. Sometimes you or someone needs immediate medical attention, but not the right guide to counseling. Therefore, it also suggests the specialist along with the details of the hospital.

Developer Responsibilities

1. Install the software on the client hardware and configure the software as required by the client.
2. Undertake any user training that may be necessary in the use of this system.
3. System maintenance over a period of time.

.

Existing System

Although the current system is manual and file-based. We understand that the system thatwe are going to build must provide solutions to the loss of time and space that affects the efficiency of day-to-day activities in the hospital. We are all patients at some point and we all want good health care. Assume that doctors are medical experts and that there is good research behind everydecision. But it predicts diseases before the person is affected by the disease. The physician's needfor in-depth research that goes beyond medical practice.

The current hands system has a lot of work on paper and does not take care of the predictionof the data.

Proposed System

We have developed an expert system called the Online Health Prediction System that is used to facilitate the prediction of chronic diseases. In this system, we have combined structured data and unstructured data in health care to estimate the risk of disease.

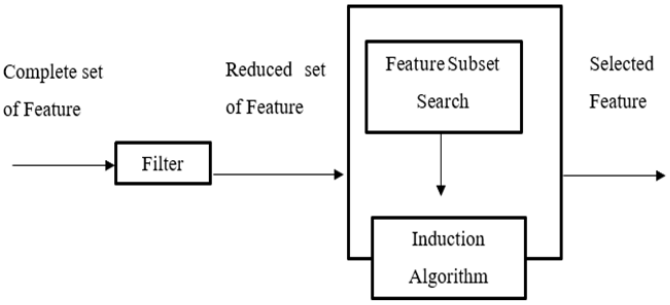
- To deal with structured and unstructured data, we use the kmeans algorithm and linear regression to select the characteristics that will be used by the machine learning model to predict disease.
- Users can seek medical help at any time.
- Doctors are attracting more clients through the Internet.

LITERATURE REVIEW

Machine learning methods for theoretical exploration of life are generally applicable. Numerous studies have been published based on the value of medical diagnostic technologies. Such experiments applied various solutions to problem and obtained a precise classification. theycapture the actual survival of . And al. [6] Even for breast cancer patients, a naive bay, a decision tree, and a backpropagation neural network were used. Although the results obtained were robust(around 90 percent accuracy), were not necessary because the data were divided into two groups:one for survival greater than five years and the other for the who died within the five years. The findings were not relevant. [7] An application selection approach that focuses on evaluating the usability of function selectors. This looks for a simple and consistent set of functions without losingthe dimension of predictive precision. A scoring algorithm is used to assign a confidence ofto thecharacteristics. [8] By using fuzzy reasoning in , you reduce the extent of the original problem, theproposed GA / SVM hybrid solution. A subset of healthy genes, , will be identified and then evaluated by SVM. [9] This study focused on and compares the performance of ArtificialNeural Network (ANN) and Vector Machine Support (SVM) for the classification of liver cancer.For precision, flexibility, specificity and efficiency , both models were contrasted and validated with the liver disorders dataset BUPA. [10] Used in combination with prediction of heart disease in Association Mining and Genetic Algorithm. The proposed methodology used the statistics of the Gini index mutation genetic algorithm for the interaction and crossover method. Used a qualityenhanced feature collection methodology. [17] Designed the Alzheimer's Risk Prediction Systemusing patient EHR data. Here they used the context of active learning to solve a real problem for the patient. In this, the risk model of the active patient was configured. The risk of Alzheimer's disease is used for this active risk prediction algorithm. N.H. Barakat suggested understandable support.

FEATURE SELECTION

The increased use of computers from all angles leads to the collection of large amounts ofdata. This data is large and systematically linked to define the appropriate trends, making data mining a critical field for analysis, estimation, and other data activities. You have entered a dynamic field of research in order to solve theoretical problems in real time. Big data mining is used in various places where data processing operations are required. These development and datamining strategies have been widely used at various levels such as pattern recognition etc., and thecollection of applications in practically all areas plays an important role. The purpose of the selection is to determine the smallest possible subset of characteristics. The framework selects the base of the original characteristics by removing obsolete and unnecessary data dimensionality characteristics without losing any useful details. Before introducing the data mine tasks, this is the critical pre-processing phase. The accuracy of the mines, the calculation time and the understanding of the test are improved. Three types of selection methods consist of filters, wrappers, and built-in approaches.



As discussed in [16], the filter selects the function regardless of the type of classifier used. The advantage of this method is that the characteristics only need to be selected once, it is simple and independent of the type of classifier used, but this method has no relation to the classifier, each characteristic is interpreted separately from the functional dependencies. The Wrapper approach depends on the classification used. The results of the classifier are used to evaluate the quality of the specified attribute or characteristic. Another approach has the advantagethat the filtering cycle eliminates the disadvantage, which is simpler than the filtering system as it still handles all dependencies. The next integrated approach is to combine a filtering algorithm with a wrapper approach to find an optimal subset of features that is integrated into the classification structure. The benefit of this approach is less expensive and less brittle than the wrapapproach. Various uses of feature selection in the last two decades:

- Text mining
- Image processing and computer vision
- Industrial applications
- Bioinformatics

MATERIAL AND METHODS

Weka is a data mining platform that uses various algorithms. These algorithms can be explicitlyadded to the data or from the Java code name. Weka is a set of resources for:

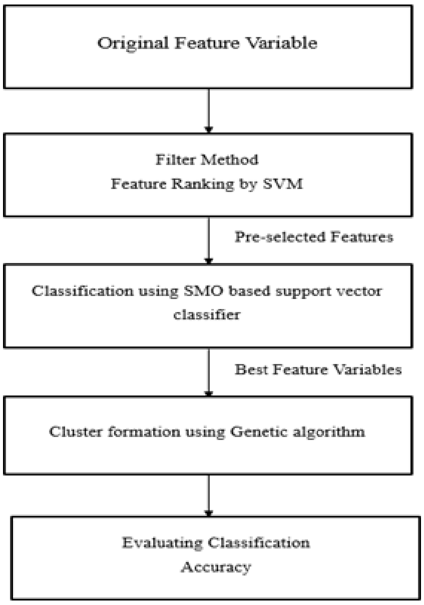
- Regression.
- Cluster.
- Association.
- Data preprocessing.
- Classification evaluation.
- Visualization. ”[13]

We have classifiers in Weka to estimate quantitative or numerical quantities. Decision making and lists are available for learning systems, vector support computers, case-dependent classifiers, logistic regression, and Bayesian networks. All tabs are activated as soon as the data is loaded.

Depending on the specifications, trial and error, the suitable algorithm can be found for a simple data representation. The Clusters tab helps the person classify groups or classes of incidents in the data collection. Clustering provides data to the consumer for review. The specified training set, percentage division, test set, and classes are used for grouping, where users can ignore otherattributes of the data set as needed. "KMeans, EM, Cobweb, Xmeans, and FarthestFirst are included in Weka."

FRAME WORK

An illustrative description of SVM and the gene-dependent clustering model is shown in FIG. Below is an algorithm for the improved SVM method for genetic clustering.

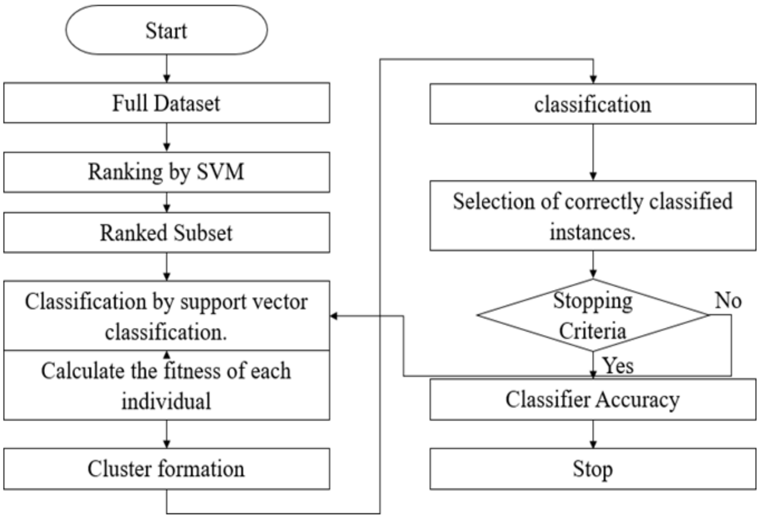


Proposed Framework

The initial variables that are filtered by the SVM filter are selected. Performance is basedon sensitivity, specificity, measurement, memory, precision. In the data mining industry, these metric measures played a crucial role in evaluating the results of various classifications and servedas a guide for algorithms, as shown in Table 3.2. TP = true positive: number of positively predicted examples that are actually positiveFP = false positive: number of positively predicted examples that are truly negative TN = true negative: number of examples that are negatively predicted and are actually negative FN = false negatives: The following metrics are measured in terms of the number of negative predictions that represent good ranking results. Sorting efficiency is measured in the following indices.

The proposed hybrid feature selection algorithm:

In this section we define a basic genetic clustering SMO algorithm as suggested by ourexperiments. Before we can use the genetic algorithm, it is necessary to calculate several operations. It is the first population, fitness, selection and crossing.



CONCLUSION

This article presents a modern approach to feature selection. This selection method is based on logistics algorithm. Our method was tested with severalother classifiers of different types in thebreast cancer, lung cancer, and diabetes data sets. The results show the robustness of the proposedapproach.

REFERENCES

1. Soni, Vishal Dineshkumar, Information Technologies: Shaping the World under the Pandemic COVID-19 (June 23, 2020). Journal of Engineering Sciences, Vol 11, Issue 6,June/2020, ISSN NO:0377-9254; DOI:10.15433.JES.2020.V11I06.43P.112 , = Available atSSRN: https://ssrn.com/abstract=3634361
2. Soni, Vishal Dineshkumar, Challenges and Solution for Artificial Intelligence in Cybersecurity of the USA (June 10, 2020). Available at SSRN: https://ssrn.com/abstract=3624487 or http://dx.doi.org/10.2139/ssrn.3624487
3. American college of radiology, Reston VA, Illustrated Breast imaging Reporting and Data system (BI-RADSTM) , third edition, 1998.
4. Urmaliya, Ajay & Jyoti, Singhai. (2013). Sequential minimal optimization for support vector machine with feature selection in breast cancer diagnosis. 2013 IEEE 2nd International Conference on Image Information Processing, IEEE ICIIP 2013. 481-486. 10.1109/ICIIP.2013.6707638.
5. Bittern, R., Dolgobrodov, D., Marshall, R., Moore, P., Steele, R., AND Cuschieri, A. artificial neural networks in cancer management. e-Science All Hands Meeting 19 (2007),pp.251 – 263.
6. Bellaachia, A., AND Guven, E. Predicting breast cancer survivability using data miningtechniques.
7. Afef Ben Brahim, Mohamed Limam, 2017, "Ensemble feature selection for high dimensional data: a new method and a comparative study", International Federation of Classification Societies (IFCS), vol. 12(4), pp 937-952
8. Huerta, E.B., 2006, ―A Hybrid GA/SVM approach for gene selection and classification ofmicroarray data‖, springer, vol.3907 pp. 34–44
9. Sallehuddin, R.N.H.aidillah, S.H., Mustaffa, N.H., 2014, ―Classification of liver cancer using artificial neural network and support vector machine‖ In: Proceedings of International Conference on Advance in Communication Network, and Computing, Elsevier Science, pp 487-493.
10. Jabbar, M.A., Deekshatulu, B.L., Chandra, 2012, ―Heart disease prediction system using associative classification and Genetic Algorithm‖.
11. S. S. Nikam, ―A Comparative Study of Classification Techniques in Data Mining Algorithms‖, Oriental journal of computer science & Technology, vol.8, pp.13-19, April 2015.
12. S. Neelamegam, E.Ramaraj, ―Classification algorithm in Data mining: An Overview‖,International Journal of P2P Network Trends and Technology (IJPTT), vol.4, pp.369-374, September 2013.
13. Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.
14. ―Platt, J.C.: Sequential minimal optimization: a fast algorithm for training support vector machines. Technical Report MSR-TR-98- 14, Microsoft Research, 1998.‖
15. ‖Islam, M. Z., Estivill-Castro, V., Rahman, M. A., Bossomaier, T. (2018). Combining K-Means and a Genetic Algorithm through a Novel Arrangement of Genetic Operators for High Quality Clustering. Expert Systems with Applications. 91:402-417.
16. Sangaiah, Ilangovan & Vincent Antony Kumar, A.. (2019). Improving medical diagnosis performance using hybrid feature selection via relieff and entropy based genetic search (RF- EGA) approach: application to breast cancer prediction. Cluster Computing. 22. 10.1007/s10586-018-1702-5.
17. B. Qian, X. Wang, N. Cao, H. Li, and Y.-G. Jiang, ―A relative similarity based method forinteractive patient risk prediction,‖ Springer Data Mining Knowl. Discovery, vol. 29, no. 4, pp. 1070–1093, 2015.

18.N.H. Barakat, A.P. Bradley, M.N.H. Barakat. (2010). Intelligible SupportVector Machinesfor Diagnosis