A SURVEY ON MACHINE LEARNING CLASSIFIERS IN DECISION SUPPORT SYSTEM FOR DISEASE PREDICTION

Dr.N.Arunachalam[1], Dr.R.Raju[2], R.Agnalla[3], R.Bhakiyalakshmi[4], K.Vigneshwari[5] Dept. of Information Technology,
Sri Manakula Vinayagar Engineering College, Madagadipet, Puducherry

Abstract
The medical sector manages huge amount of data on a regular basis. The findings can be influenced by the processing of large data using conventional methods. Machine Learning algorithms can be used to find out facts in medical research, particularly for predicting diseases. Early disease identification is essential for patient care and expert research. Machine Learning algorithms such as Decision trees, Help Vector Machine, Multilayer Perception, Bayes classification, techniques, K-Nearest Ensemble classification techniques, etc. are used to evaluate various ailments. Machine learning algorithms can lead to rapid prediction of disease with high precision. This research paper that focuses on predicting chronic kidney disease, heart disease, diabetes and breast cancer. The paper also looks at the hybrid approach that improves individual classifier efficiency.

Keywords

Machine Learning, conventional methods, disease prediction, patient care.

1. INTRODUCTION

Machine Learning is a branch of Artificial Intelligence that seeks to provide computer method to collect information about change and upgrade intelligent systems. Artificial Intelligence (AI) allows systems to learn from contexts, implement those functions and increase the likelihood of success in solving problems in the real world [1]. AI proves to be an interesting field with technological advances and scientific development. Hence, it contributes to growing emphasis on ML techniques. Machine learning (ML) is an important data analysis approach that iteratively learns from the available data with the aid of learning algorithms. Machine learning is an artificial intelligence field that uses algorithms to improve performance over time, for example, or to find patterns in data [2]. This work introduces new algorithms to discover useful information for three different problems in health care. Increasingly, decision support systems (DSSs) were designed to help clinicians. These DSSs store and use information when a question enters. The expertise of these DSS methods typically stems, for example, from the Direct Input of laws from human experts. The DSSs suggested gather information automatically, and use method of optimization to give appropriate answers to queries. They are DSS built with the methods of machine learning (ML)[3]. In this proposed system the ability of machine learning to promote personalized healthcare, reduce health resource utilization and improve outcomes. This desertion discusses three problems from various fields of medical research and their solutions to machine learning. Each solution is a separate form of decision- support system. They exhibit three specific properties: customized decision support for health care, decreased use of medical resources and improved outcomes [4]. Disease forecasting plays a key role in machine learning. Different types of diseases can be predicted using ML techniques. However, they are many different types of diseases are predicted using machine learning techniques. This research focused on predicting chronic kidney disease, heart disease, lymphatic system diabetes and breast cancer, and lung disorders. Here it describes a brief description of a few diseases. Several metaheuristic algorithms have been developed in recent decades to address a variety of complex issues influenced by natural phenomena or biological behaviour [5]. This survey introduces available methodologies, techniques and their constraints to the practitioners ' understanding intent in this area.

2. RELATED WORK

There are some classification tools that are used in evaluating medical data. These contain Decision trees, K-Nearest Neighbor (KNN), Bayesian network (Naive Bayes), Neural networks, Fuzzy logic, J48 and Support Vector Machine (SVM), Random Forest. All these classifiers are basically learning methods and adopt sets of rules: Prediction of liver disease can be done using Bayesian Classification using Naïve Bayes and Functional Tree (FT) algorithms [6].

The primary objective is to predict the class type of classes such as Liver Cancer, Cirrhosis Hepatitis and No Diseases. Using data mining techniques, he predicted and analyzed liver disease using a weka method. Finally, the author compared the outputs obtained from Naïve Bayes and FT algorithms and concluded that the Naive Bayes algorithm plays a key role in predicting liver disease. The key imitation of Naive Bayes is the presumption of an independent predictor. Naïve Bayes implicitly assumes that all attributes are mutually independent. In real life, it's almost difficult to get a set of predictors that are completely independent. When the liver is ill, it can cause serious damage to our bodies. There may be a number of things and aspects of health that may unknowingly cause liver damage. Solanki [7] used weka as a data mining technique Prevalent for the classification of sickle cell disease. The author compared J48 and Random tree algorithms and provided a predictive classification model with respect to a person's age of different types of blood groups. It is concluded from the experiment that the Random tree is a stronger algorithm as it makes more in-depth decisions regarding J48 for sickle cell diseases.In the same way, Joshi et al. [8] Diagnosis and prognosis of breast cancer based on classification rules. Breast cancer prediction has long been seen as an important research topic in the scientific and healthcare communities. There are several risk factors for breast cancer, including female sex, obesity,lack of exercise, drinking alcohol, hormone replacement therapy during menopause, ionizing radiation,early pregnancy.By comparing classification rules such as Logistic Model Tree (LMT), Multilayer Perceptron, Stochastic Gradient Descent (SGD), Simple Logistic, Sequential Minimal Optimization (SMO), AdaBoostM1, Attribute Selection,Regression Classification, Filtered Classifier,Multiclass Classifier and J48,the researchers suggested that LMT Classifier would give more accurate diagnosis i.e. 76 percent.Here the constraint falls within the limits of accuracy and performance.Nevertheless, David et al. [9] used classification methods to predict leukemia. A total of 100 types of cancer required special care and diagnosis. One of these types of cancer is leukemia, where it is a cancer of the tissue. The authors compared the performance obtained using K-Nearest Neighbor, Bayesian Network, the randomtree and the J48 tree are based on precision, learning time and error rate. Machine Learning requires massive data sets to learn, which should be inclusive unbiased and of good quality. There may also be occasions when they have to wait for new data to be produced. Bayesian algorithm works well on classification, according to them.In 2013, Vijayaani and S u d h a [10] compared the study of cardiac disease classification feature techniques. Testprocedures are said to be necessary in order to reach the final diagnosis.Nevertheless, too many tests, on the other hand, may confuse the key diagnostic process and lead to difficulties in obtaining the final results, in particular in the case of detecting a disease, several tests should be carried out. This kind of challenge could be overcome by means of machine learning, which could be used directly to obtain the final result with the aid of a few artificially intelligent people. Classification is one of the most important strategies for data mining. Classification was performed using algorithms such as LMT, Multilayer Perceptron and Sequential Minimal Optimization algorithms to predict heart disease. The downside of this is that which cannot solve non-linear logistic regression problems since its decision-making surface is linear. Sugandhi et al. [11] studied a popu in 2011.According to them, Tree is the best classification algorithm for cataract patient disease. This deals with the care of patients with functional impairment due to cataract and improved function as a result of treatment for the disease. Taking into account the prevalence of cataract among men and women, the objective of the analysis is to define the characteristics that assess the existence of C. Naïve Bayes implicitly assumes that all features mutually independent. In real life, it's almost difficult to get a set of predictors that are completely independent. Bin and Yan [12] compared different various classification strategies used by weka for breast cancer is open source software that consists of a series of machine learning algorithms for data mining tasks. The purpose of the research is to examine the efficiency of different classification approaches using WEKA for breast cancer. They used algorithms such as Bayes network classifier, radial base function, nearest neighbor algorithms. The key disadvantages of these algorithms are that they build as many trees on the data subset and aggregate the output of all trees. This reduces over-fitting problems and also eliminates uncertainty and therefore improves precision.

Throughout their analysis, various algorithms were used to predict the effects of each algorithm and to train it. Suchithra, Dr. P. Uma Maheswari [13] has developed a Clinical Decision Support System for Diagnosing Heart Disease. The statistical data on the cause-specific mortality rate sugges that cardiovascular disease plays an important role in leading to mortality. The demographic prediction graph shows a significant rise in cardiovascular disease. The growing population shifts moral and lifestyle levels. The main objective of this paper is to establish an expert diagnostic method for the existence of Ischemic Heart Disease with an integrated automatic classifier. Cost of production Implementation of a process where an automation system is very expensive.Veena and Anjali [14] have introduced Decision Support Systems to Predict Diabetes Mellitus. Diabetes Mellitus is caused by diabetes mellitus. Increased sugar content in the blood. This can cause serious problems such as kidney failure, stroke, cancer, heart disease, and blindness. The Weka is the method under consideration for this analysis.

The dataset was taken from the University of California, Irvine (UCI) machine learning server. The main objective of this analysis is to explore the advantages of various pre- processing strategies fordecision- supporting diabetes prediction systems based on the Support Vector Machine (SVM), the Naive Baye Classifier and the Decision Tree. The main drawback is data shortage and lack of accuracy. In 2016, Ani et al. [15] studied the Decision Support System for the Diagnosis and Prediction of Chronic Renal Failure using Random Subspace Classification. Chronic Renal Failure (CRF) is one of the major diseases affecting human life. The stage of CRF starts with the loss of renal function and eventually progresses to complete failure of all renal function. In this paper, classification techniques such as neural network based back propagation (BPN), probability based Naive Bayes, LDA classifier, lazy learner K- Nearest Neighbor (KNN), tree-based decision tree, and Random subspace classification algorithms are studied first. On the basis of a classification algorithm, the decision support system is used to make CRF predictions. For large datasets, the expense of measuring the distance between the new point and each current point is immense, which degrades the efficiency. The KNN algorithm does not work well with high dimensional data because with a large number of dimensions, it is difficult for the algorithm to measure the distance in each dimension.In 2017, Pradnya Kul et al. [16] used the Information Retrieval Medical Decision Support System for Privacy Protection. The Clinical Decision System (CDS) is a program used to treat patients in different ways.This primarily monitors symptoms of the patient, the test result and the his torical data.They used naive Bayesian to train and identify the patient's data set. But naive Bayesian doesn't make it effective. In the same year, Saima Safdar et al. [17] studied Machine Learning Based Decision Support Systems (DSS) for the diagnosis of heart disease. The current study leads to a comprehensive overview of decision support systems for the treatment of heart disease in clinical settings. The authors independently reviewed and abstracted studies of the cardiac disease-based clinical decision support system. The svm algorithm has been used for classification.

3. RESEARCH ISSUES AND OPPORUNITIES

In this context, the similar approaches that the machine learning based methods to the general classification-based methods, various kinds of the chronic disease prediction methods were studied. The significance of the proposed framework (which is based on machine learning and block chain) over previously designed methods, which further paves a way for its application for chronic disease detection in cloud server. More precisely, the performance of this framework overpasses other models in term of runtime disease detection. In addition to this, when a certain new type of malware or new feature is added the accuracy is increased further. Therefore, our proposed method can handle the multi-features by combing the advantages of machine learning and block chain that effectively detect the chronic disease prediction for cloud platform.

4. CONCLUSION

The main objective of this paper is to predict chronic kidney disease, Diabetes, Heart Disease, Breast Cancer. Using two algorithms i.e. Random Forest and Multilayer Perceptron for our experiments. These algorithms are implemented using WEKA machine learning tool to analyze accuracy which is obtained after running these algorithms in the output window. After running these algorithms, the outputs are compared on the basis of accuracy achieved. These algorithms have been compared with classification accuracy to each other on the basis of correctly classified instances, mean absolute error, Kappa statistics and RMSE metric. The results show that MLP classifier outperforms Random Forest classifier in all sector with respect to the parameters specified. It is concluded that MLP classifier is the best prediction algorithm for chronic kidney disease diagnosis, heart disease, diabetes, Breast Cancer. In future it consists of some more classification algorithms distributed in nature and analyze their performance with the same dataset and also by changing the dataset. However Deep Neural Network will be considered among other classifiers.

5. REFERNCES

[1] J. S. Park, T. Y. Youn, H. B. Kim, K. H. Rhee, and S. U. Shin, ``Smart contract-based review system for an IoT data marketplace," Sensors, vol. 18, no. 10, p. 3577, 2018.

[2] B. L. R. Stojkoska and K. V. Trivodaliev, ``A review of Internet of Things for smart home: Challenges and solutions," J. Cleaner Prod., vol. 140, no. 3, pp. 14541464, 2017.

[3] M. Damshenas, A. Dehghantanha, K.-K. R. Choo, and R. Mahmud, ``M0Droid: An android behavioral-based malware detection model," J. Inf. Privacy Secur., vol. 11, no. 3, pp. 141157, Sep. 2015.

[4] J. Walls and K. K. R. Choo, ``A review of free cloud-based anti-malware apps for android," in Proc. 14th IEEE Int. Conf. Trust, Secur. Privacy Comput. Commun., vol. 1, Aug. 2015, pp. 10531058.

[5] H. Chen et al., ``Malware collusion attack against SVM: Issues and countermeasures," Appl. Sci., vol. 8, no. 10, p. 1718, Sep. 2018.

[6] Dhamodharan S., Liver Disease Prediction Using Bayesian Classification, ISSN:2320-0790, Special Issues,4th National Conference on Advance Computing , Application Technologies, May 2014.

[7] SolankiA.V., Data Mining Techniques using WEKA Classification for Sickle Cell Disease, International Journal of Computer Science and Information Technology, 2014.

[8] Joshi J, Rinal D, Patel J, Diagnosis And Prognosis of Breast Cancer Using Classification Rules, International Journal of Engineering Research and General Science, October 2014.

[9] David S. K., Saeb A. T., Al Rubeaan K., Comparative Analysis of Data Mining Tools and Classification Techniques using WEKA in Medical Bioinformatics, Computer Engineering and Intelligent Systems,2013.

[10] Vijayarani,S., Sudha, S.,Comparative Analysis of Classification Function Techniques for Heart Disease Prediction, International Journal of Innovative Research in Computer and Communication Engineering, 2013.

[11] Dogru and K. Ömer, ``Web-based android malicious software detection and classication system," Appl. Sci., vol. 8, no. 9, p. 1622, Sep. 2018

[12] G.ShanmugasundaramG. Sankarikaarguzhali, "IOT Healthcare Analytics - Machine Learning Approach For Prediction Of Heart Diseases", Journal of Advanced Research in Dynamical and Control Systems, ISSN 1943-023X, Special Issue – 12 Vol. 9/ 2017 (Scopus Indexed)

[13] Strategy Analytics: Android Captures Record 88 Percent Share of Global Smartphone Shipments in Q3 2016, Businesswire, San Francisco, CA, USA, 2016.

[14] A. Demontis et al., ``Yes, machine learning can be more secure! A case study on android malware detection," IEEE Trans. Dependable Secure Comput., to be published.

[15] S. Y. Yerima, S. Sezer, and I. Muttik, ``Android malware detection using parallel machine learning classiers," in Proc. 8th Int. Conf. Next Gener. Mobile Apps, Services Technol., Sep. 2014, pp. 3742.

[16] W. Enck et al., ``TaintDroid: An information- row tracking system for realtime privacy monitoring on smartphones," Commun. ACM, vol. 57, no. 3, pp. 99106, 2014.

[17] Homer Jar DW, Lemeshow S, Sturdivant RX. Applied logistic regression. Wiley; 2013

[18] Shanmugasundaram G , R. Raju , Saravanan R , Balaji S ,Vasumathi B and Hemavarni R, "Heart Disease Prediction Using Artificial Neural Network" , International Journal of Pure and Applied Mathematics, Volume 119, No. 14 2018, 17-27, ISSN: 1314-3395 (Scopus Indexed)

[19] Joachims T. Making large-scale SVM learning practical. SFB 475: Komplex itätsreduktion Multivariaten Datenstrukturen, Univ. Dortmund, Dortmund, 2015.

[20] Joachims T. Making large-scale SVM learning practical.SFB 475: Komplex it ä sre duktionMultivariaten Datenstrukturen, Univ. Dortmund,Dortmund, Tech.2017.

[21] Quinlan JR. Induction of decision trees. Mach Learn. 2016

[22] Breiman L. Random forests. Mach Learn. 2001.

[23] G. Shanmugasundaram, V. M. Selvam, R. Saravanan and S. Balaji, "An Investigation of Heart Disease Prediction Techniques", 2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA)