International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

Survey on Asthma Prediction Using Classification Technique

Mrs. J. Cathrin Princy [1], Mrs. K. Sivaranjani [2]

[1]Research Scholar, Department of Computer Science, Bishop Heber College Tiruchirappalli, Tamilnadu, India
[2]Asst. Professor, Department of Information Technology, Bishop Heber College, Tiruchirappalli, Tamilnadu, India

Abstract— Asthma is a common disease characterized by redness and hyper-reactivity in the airways which causes reversible airflow limitation. In this paper, how to simply predict asthma affected people using data mining classification techniques. Asthma prediction is how to find out the people whether asthma affected or not. Prediction will be done certain types of breathing tests there are FEV1/FEC, FEF. Predictions are only using yes or no condition. Mostly machine learning algorithms are to predict asthma there are support vector machine(SVM), artificial neural networks(ANN), k-nearest neighborhood algorithms(KNN), adaboost, random forest algorithms. It comprises the analysis of various classification techniques an asthma prediction. It compares classification techniques and produce the result based on accuracy level.

Keywords— Data mining, Classification, Asthma, SVM, ANN, KNN, breathing tests, Machine learning.

I. INTRODUCTION

Data Mining is a knowledge mining process. It is an interdisciplinary subfield of computer science. It is the computational process of discover patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems.

Classification is a data mining function that assigns items in a collection to target categories or classes.The goal of classification is to accurately predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as low, medium, or high credit risks. Classification models predict categorical class labels. Classification techniques are using classifiers. Here the test data is used to estimate the accuracy of classification rules. Many classification techniques are available in data mining. There are Decision tree, K-nearest neighbor, Bayesian classification, Artificial neural networks(ANN), Support vector machine(SVM) and so on.

Asthma is a lung disease caused by the inflammation and narrowing of the airways that causes recurrent attacks of breathlessness and wheezing, and often can be life-threatening. The common reason for asthma prediction is smoking, dust allergy, chest tightness, cough, breathlessness. Asthma is a chronic inflammatory disorder of the airways. Airway inflammation

is the result of interactions between various cells, cellular elements and cytokines. Asthma develop their first symptoms before the 5th year of age. However, asthma diagnosis in children younger than five years old remains a challenge for the clinical doctors. In this paper, a brief literature survey has been performed in order to give an algorithm with more accuracy.

II. RELATED WORK

In [4], the authors used the combination of machine learning algorithms. There are Context sensitive auto-associative memory neural network model(CSAMM) which has 84.32% accuracy, Backpropogation model which has 82.21% accuracy, C4.5 algorithm which has 83.83% accuracy, Bayesian network which has 81.17% accuracy, Particle Swarm Optimization(PSO) which has 84.16% accuracy. Two types of methods are used here. The first one is questionnaire and the second one is clinical diagnosis. CSAM using inputs of two vectors: one representing the set of possible diseases up to the moment and the other vector corresponding to a new sign, symptom or laboratory result. Then Kronecker product is also used here. Backpropogation is the best-known training algorithm for multi-layer neural networks. Backpropogation can also be considered as a generalization of the delta rule for non-linear activation functions and multilayer networks. C4.5 builds decision trees from a set of training data, using the concept of information entropy. PSO is a population based stochastic optimizations. PSO shares many similarities with evolutionary computational techniques such as Genetic Algorithms. PSO is initialized with a group of random particles. The sensitivity was 100% and specificity was 80%. Thus the conclusion is PSO performs faster and gets better results.

The ANN technique given in [7] was already utilized in medicine in order to improve the performance of the clinical decision-making tools. Here, ANN has the accuracy prediction as 96.77%.The prediction algorithm which has been employed in this study consists of two stages: the feature reduction through partial least square(PLS) regression and the classification stage by multilayer perceptron (MLP) gets accuracy 96.77, sensitivity 96.15% and specificity 100%.probabilistic neural networks(PNN) gets the sensitivity 100%, specificity 80%, accuracy 96.77%. MLP and PNN topologies are obtained best prediction accuracy 96.77%. Children with recurrent wheezing with asthma at school age have utilized the Asthma Predictive Index(API). In partial least square regression, breathing tests are used i.e.,FEV_1%, $FEF_{25/75}$%. PLS regression is a technique used with data which contain correlated, predictor variables. Three types of configuration were used in MLP and PNN classifier which are the input layer, the hidden layer and output layer.

In [8], the authors using support vector machine classifier. The proposed system can be used in asthma outcome prediction with 95.54% success. Here, two types of process used. The first stage is Principal Component Analysis(PCA). The second stage is Least Square Support Vector Machine(LSSVM). These two techniques are used to find out the accuracy of asthma prediction. Support Vector Machine(SVM) is a classification and regression prediction tool that uses machine learning theory to maximize predictive accuracy while automatically avoiding over-fit to data. Advantage of this process is minimize the complexity of the system and achieve the low computational cost.

[11] contains combination of machine learning algorithms. There are random forest algorithm, adaboost algorithm, artificial neural networks algorithm. The total classification accuracy of the Random Forest and AdaBoosted Random Forest classifiers were achieved 90% of accuracy, sensitivity 90% and specificity 90%. There was not any performance difference between AdaBoosted Random Forest and Random Forest classifiers. However, when compared to multi-layer perceptron neural networks (MLPNN), both of them perform better. The accuracy of MLPNN gets 80% of accuracy, sensitivity 80% and specificity 80%. Prediction of asthma using only the chest sound signals obtained from actual patients using ordinary microphones. Breathing process was used to consider whether the patient get asthma or not. Percentage of forced vital capacity(FVC) and forced expiratory volume in one second(FEV1) values, which are used in pulmonary function test(PFT). In order to perform wavelet analysis of pulmonary sounds, Matlab program was used. In this paper, when compare to MLPNN, random forest and adaboost performs more better.

In [12], the authors proposed the approach based on k-nearest algorithm. Data classification is carried out by using a variety of methods including k-nearest neighbor, support vector machine and random forest, which are popular methods for multi-class diagnosis in the area of pattern recognition. The k-nearest neighbor(KNN) algorithm which is also called the memory-based classification method, is one of easiest and most direct techniques of data mining. The support vector machine(SVM) algorithm as a very efficient method for the purpose of pattern recognition. The random forest algorithm is a general classification method consisting of a large number of decision trees and class outputs. Thus the SVM method was produced best results, i.e., specificity, sensitivity, and accuracy were 0.9934, 0.9737, 0.9870, respectively, when the radial basis function was used.

III. PERFORMANCE ANALYSIS

Three types of tests are using to find out lung capacity of pulmonary function testing. There are FVC, FEV1 and FEF. FVC refers to forced vital capacity and FEV1 refers to forced expiratory volume in one second. FEF refers to forced expiratory flow related to various portion of the FVC curl; modifiers refer to amount of FVC previously exhaled.FEF25%-75% - This measurement describe the amount of air excluded from the lungs during the middle half of the forced vital capacity test. Many physician like to look at this value because it is an display of obstructive disease. In this literature survey, seven types of

algorithms are used to predict whether the person has asthma or not. The algorithms are namely, random forest, adaboost, PNN, LSSVM MLPNN, MLP, and SVM. According to the Sensitivity, Specificity the accuracy level will be calculated.

3.1.Sensitivity

Number of correctly classified recordings belonging to subjects with asthma disease/Total number of recordings belonging to subjects with asthma disease.

3.2.Specificity

Number of correctly classified recordings belonging to healthy subjects/Total number of recordings belonging to healthy subjects.
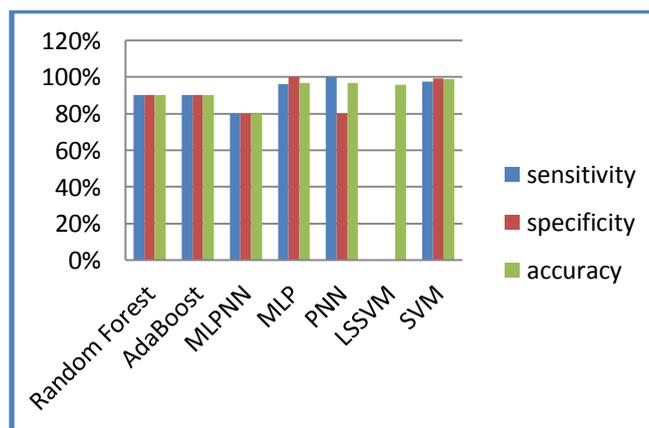
3.3.Accuracy

Number of correctly classified recordings /Total number of recordings.

Table 1: Clinical values about the Asthma prediction tests

FEV1/FVC	FEF	Dignosis
>80%	-	minimal obstructive defect
65-80%	-	mild obstructive defect
50-65%	-	moderate obstructive defect
<50%	-	severe obstructive defect
-	<65%	Asthma

Commonly using the breathing test is named as FEV1/FVC. In medicine field spirometer is used to find out these values.

Chart 1: General analysis of sensitivity, specificity, accuracy of Asthma prediction using the above 7 algorithms.

In this chart, asthma affected people are easily analyzed the accuracy level are focused based an comparison of various classification techniques according to the chart 1 and the highest accuracy has found on the technique SVM.

IV. CONCLUSION

In this study, various classification techniques are used to find out the sensitivity, specificity, accuracy of asthma prediction data. High accuracy achieved through SVM technique that has 98% compare than other algorithms. MLP achieved 100% specificity compare than other algorithms. PNN achieved 100% sensitivity compare than other algorithms. Three types of breathing tests are used in this paper which is FEV1, FVC and FEF. In this paper, are commonly applying the FEV1/FVC test to find out the lung capacity. Lung capacity means find out the various breathing sounds of a particular person. In this survey, among all the existing techniques SVM technique gives the best accuracy. The next accuracy level has been produced by the technique MLP. In Future, the techniques SVM and MLP has to be merged to yield more accuracy level.

REFERENCES

[1] Ali, J, Summer, WR, &Levitzky, MG. (2009). "Pulmonary pathophysiology" (3rd ed.). New York: McGraw-Hill Medical.

[2] BDCN Prasad, P E S N Krishna Prasad and Y Sagar -"A Comparative Study of Machine Learning Algorithms as Expert Systems in Medical Diagnosis (Asthma)", CCSIT 2011, Part I, CCIS 131, pp. 570– 576, 2010.

[3] BDCN Prasad, P E S N Krishna Prasad, Y Sagar: "A Study on Associative neural memories", IJACSA 2010, vol 1 no.

[4] BDCN Prasad, P. E. S. N Krishna Prasad and Y Sagar, "An approach to develop expert systems in medical diagnosis using machine learning algorithms (Asthma) and a performance study", International Journal on Soft Computing, Vol. 2, No. 1, 26-33, 2011.

[5] A. Bush, "Diagnosis of asthma in children under five," *Primary Care Respiratory Journal*, vol. 16, no. 1, pp. 7–15, 2007.

[6] Çatmakaş, Z, Köse, İH, Toker, O, &Öz, HR. (2009). Towards an ARM based low cost and mobile biomedical device test bed for improved multi-channel pulmonary diagnosis. In 4th European Conference of the International Federation for Medical and Biological Engineering (pp. 1108–1112). Berlin Heidelberg: Springer.

[7] E. Chatzimichail, E. Parakakis and A. Rigas, "Predicting asthma outcome using partial least square regression and artificial neural networks" Advances in Artificial Intelligence, Article ID 435321, 1-7, 2013.

[8] E. Chatzimichail, E. Parakakis, M. Sitzimi and A. Rigas, "An intelligent system approach for Asthma prediction in symptomatic preschool children", Computational and Mathematical Methods in Medicine, Article ID 240182, 1-5, 2013.

[9] M Gori, P Frasconi and A Tesi, "Success and failures of Backpropogation: A Theoretical Investigation".

[10]J.R.G.Lansangan and E.B.Barriors, "Principal components analysis of nonstationary time series data, "Statistics and Computing, vol. 19, no. 2, pp173-187, 2009. View at Publisher. View at Google Scholar. View at Scopus.

[11] Nahit Emanet, Halil R Oz, Nazan Bayram and Dursun Delen, "A comparative analysis of machine learning methods for classification type decision problems in healthcare", Decision Analytics, vol. 1, No.: 6, 1-20, 2014.

[12]TahaSamadSoltaniHeris,MostafaLangarizadeh, Zahra Mahmoodvand, Maryam Zolnoori, "Intelligent diagnosis of Asthma using machine learning algorithms", International Research Journal of Applied and Basic Sciences, Vol. 5, No. 1, 140 – 145, 2013.

[13] Xie J, Wang C. 2011 "Using Support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases. "Expert Syst. Appl. 38(5): 5809-5815.

[14] www.decisiontrees.net/node/