

A Survey on Prediction of Health Diseases Using Classification Techniques

CH. Kavitha ¹, N. Harini ²

Assistant Professor, Department of CSE, Avanthi Institute of Engineering and Technology, Vizianagaram, India¹

Assistant Professor, Department of CSE, Avanthi Institute of Engineering and Technology, Vizianagaram, India²

ABSTRACT: Technology in medical sciences has led to the development of huge number of hospitals and health care centers. Data-mining algorithms are utilized for analyzing and extracting data in various areas and to obtain meaningful data from it, that is raw data into useful information. Classification is a prominent data mining techniques which add same set of labels into a class. This paper summarizes Decision tree, artificial neural networks, association classification, K-Nearest Neighbors and genetic algorithms and also identifies different approaches associated with the implementation of data mining for diagnosing various health problems. These algorithms help in diagnosing and predicting the diseases like heart diseases, liver diseases, and cancer.

KEYWORDS: Classification, ANN, Hybrid DT, KNN, KEEL tool, Decision Tree, Bayesian Classification, SVM , Fuzzy-Genetic Algorithm, WK-ELM

I. INTRODUCTION

Data mining play a key role in knowledge discovery to obtain essential information [1]. It is the process of detecting and extracting hidden information, patterns and specific data connections with certain predictions. The techniques are classified as supervised and unsupervised learning. Supervised learning relies on labeled data in the training data set.

Unsupervised learning refers to organizing the information on non labeled data in the form of clusters. There are certain algorithms in the learning methods as shown in the figure1.

Data mining is the area of research for identifying meaningful information from large datasets of health organizations. Classification techniques are used to examine different diseases on certain factors like food production, purity of water, agriculture factors, stress environment as shown the following figure.

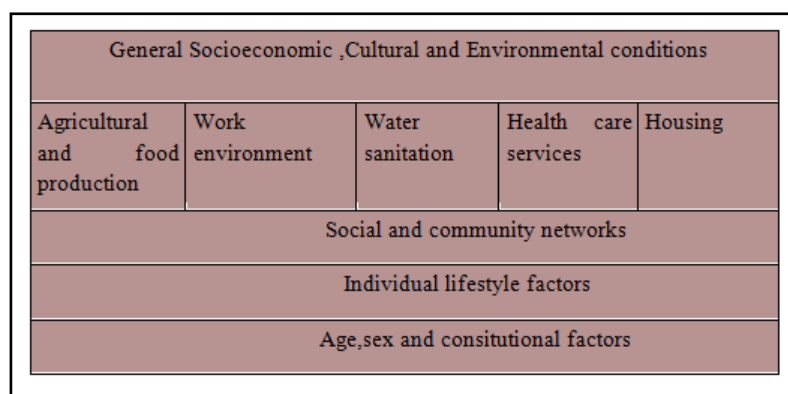


Fig. 1. Classification applications

This paper illustrates the summary of different techniques and tools in classification. The paper is arranged as follows: Section2 illustrates related work and section3 is conclusion.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 8, August 2019

II. RELATED WORK

In [2] C4.5 algorithm used for handling missing attributes and tree post-pruning is performed. C4.5 is a decision tree technique and combined with ANN (Artificial Neural Networks) and named as hybrid DT for the result. This technique is applied on the data that is pruned to remove noise. The result is validated by using tenfold validation test after determining the correctness, sensitivity and specificity of each algorithm. This determines the algorithm that works best and faster in identifying the symptoms that cause heart disease for provided dataset.

In [3] UCI machine learning used breast cancer Wisconsin data sets that are extracted from fine needle aspirate (FNA) of breast dimension. The data is given as input for Bayesian classification and decision tree.

A decision tree classification algorithm J48 builds a tree model using classification and leaf nodes are constructed using data subsets. J48 is implemented in WEKA tool [4] which is implemented in following basic steps a) same class name is given that belong to the same class of labelled tree. b) The gain is calculated for each attribute. c) entropy is calculated as a measure of disorder. Number of occurrences is not defined well in the data sets from the remaining instances. So pruning is applied and data sets are given as input to decision tree algorithm for reducing the classification mistakes and for better accuracy. Bayesian classification is a probabilistic model, statistical and supervised learning method used to solve diagnostic health diseases. Bayes is a well built algorithm to edify noise input instances, the text document in sentiment analysis.

WEKA tool is used for conducting experiments. The breast cancer [5] data set is considered as input for Bayesian classification and decision tree algorithms. Each information can have about 36 attributes with the number instances 320. Methodologies used are

1. Correctly Classified Instances (CCI) compares decision tree algorithm with Bayesian and shows the accuracy.
2. Kappa Statistic (KS) calculates to identify the similar data by two individual tools.
3. Mean absolute error (MAE) measures how near estimates or predictions to the existing outputs.
4. Root mean square error (RMS) used to similarities between two data sets.

In [6] K-Nearest Neighbor (K-NN) classifier identifies the nearest neighbor data points. KNN used in data related to health science, cluster analysis, pattern identification and online marketing etc. The research based on this algorithm on health diseases is to predict the conditions and performs analysis between cardiovascular diseases and hypertension and determinant conditions of various chronic diseases.

SVM separates the data points of the classifiers of original input data using a hyper plane. The difficulty is reduced by separation of data points in original input data by mapping original finite data into new higher dimensional data. The certain kernel functions such as polynomial, Gaussian, sigmoid SVM used for separation between data points using hyper plane with support vectors. Using the SVM classification, diabetic patients low and high risk conditions are predictable.

In [7] an artificial neural network (ANN) [8] very widely used for medical diagnosis and in biomedical systems. It is like computational model of a human brain. There was no need to provide a specific algorithm and the solution is not restricted to linear form. It easily recognizes the disease using scans. The elements or neurons on operation are interconnecting processing neurons that are inspired by biological nervous systems. The connections between elements largely determine the network functions. It is used for diagnosis diseases like nephritis diseases and heart diseases [9].

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 8, August 2019

It identifies the affected and non-affected patients. Feed-forward back propagation neural network is used as a classifier to distinguish between infected or non-infected person in different cases. In supervised learning, the network computation model is based on the input the output patterns are obtained. The neural networks follow iterative process such that it adjusts the connection weights to match its output with the actual output.

In [10] there are certain commonly used optimization methods such as Genetic Algorithms (GAs), Particle swarm optimization (PSO), Ant Colony optimization to find the network weights. GA effectively explores large spaces with back propagation network (BPN) for determining the various parameters such as number of hidden nodes. Fuzzy-genetic [11] is a combined logic that is used for diagnosis of breast cancer. With the high classification performance the output diagnosis is measured with high confidence. Fuzzy-genetic approach is highly effective to diagnose diseases for effective results.

In [12] used a Genetic Algorithm Wavelet Kernel Extreme Learning Machine [13] for ECG signal classification. WK-ELM classification gives better results compared with traditional ELM classifier. Wavelet kernel parameters can be adjusted for better results where as other kernels are fixed and cannot be adjusted. The most optimal values of WKELM parameters are rearranged by an optimization technique, Genetic algorithm. By using this algorithm most suitable solution is identified for a problem. GA technique has replicated natural survival of the fittest, that are comprised in the generation of new chromosomes. This paper tells about chromosomes and population (colony). The fitness function determines which generation is good that generates new solutions used as operators. Selection (reproduction), crossover and mutation are the three basic operators used to improve global optimized network parameters as mentioned in the figure.

In [14] KEEL tool is used that contains four modules: data management (creating data sets), experiments (creating attractive interfaces), educational (teaching part for designing desired experiment) and extending functionalities. KEEL [15] (Knowledge Extraction based on Evolutionary Learning) is an open source (GPLv3) Java software tool that can be used for knowledge discovery. In experiment module k-nearest neighbor (KNN) classification algorithm is used and CHC-TSS evolutionary algorithm is applied to instant selection algorithm. Friedman test is used to align the data. For education module to get required result, decision tree algorithm is applied. A confusion matrix and accuracy of algorithms are obtained from the modules. The table of the matrix tells the efficiency of the classifier used to test the data of heart disease. If the exit value is 0 then the classification has been done correctly. This concludes effectiveness and efficiency of the various algorithms.

III. CONCLUSION

This paper summarizes the survey on various classification tools and techniques. C4.5 algorithm is used for handling missing attributes and combined with ANN that identifies the symptoms of heart disease. It is effectively used biomedical systems. KNN classification used in KEEL tool to identify efficiency of classifiers and data is tested. A decision tree and Bayesian algorithms are used in WEKA tool to diagnose breast cancer on datasets. SVM classification uses hyper plane to identify the risk factors of diabetic patients. Genetic Algorithm WK-ELM is useful in areas of medical sciences at ECG signal classification. Fuzzy-Genetic algorithm explores large spaces with BPN used for diagnosing breast cancer. Therefore, this paper gives information about different tools and algorithms of classification used in medical diagnosis.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 8, August 2019

REFERENCES

1. Anjum Asma and Gihan Nagib, 'Energy Efficient Routing Algorithms for Mobile Ad Hoc Networks--A Survey', International Journal of Emerging Trends & Technology in computer Science, Vol.3, Issue 1, pp. 218-223, 2012.
2. Hong-ryeol Gill, Joon Yoo1 and Jong-won Lee2 , 'An On-demand Energy-efficient Routing Algorithm for Wireless Ad hoc Networks', Proceedings of the 2nd International Conference on Human. Society and Internet HSI'03, pp. 302-311, 2003.
3. S.K. Dhurandher, S. Misra, M.S. Obaidat, V. Basal, P. Singh and V. Punia, 'An Energy-Efficient On Demand Routing algorithm for Mobile Ad-Hoc Networks', 15th International conference on Electronics, Circuits and Systems, pp. 958-9618, 2008.
4. DilipKumar S. M. and Vijaya Kumar B. P. , 'Energy-Aware Multicast Routing in MANETs: A Genetic Algorithm Approach', *International Journal of Computer Science and Information Security (IJCSIS)*, Vol. 2, 2009.
5. AlGabri Malek, Chunlin LI, Z. Yang, Naji Hasan.A.H and X.Zhang , 'Improved the Energy of Ad hoc On- Demand Distance Vector Routing Protocol', International Conference on Future Computer Supported Education, Published by Elsevier, IERI, pp. 355-361, 2012.
6. D.Shama and A.kush, 'GPS Enabled E Energy Efficient Routing for Manet', International Journal of Computer Networks (IJCN), Vol.3, Issue 3, pp. 159-166, 2011.
7. Shilpa jain and Sourabh jain , 'Energy Efficient Maximum Lifetime Ad-Hoc Routing (EEMLAR)', international Journal of Computer Networks and Wireless Communications, Vol.2, Issue 4, pp. 450-455, 2012.
8. Vadivel, R and V. Murali Bhaskaran, 'Energy Efficient with Secured Reliable Routing Protocol (EESRRP) for Mobile Ad-Hoc Networks', *Procedia Technology* 4, pp. 703- 707, 2012.
9. Nobuo Ezaki, Marius Bulacu Lambert , Schomaker , "Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons", Proc. of 17th Int. Conf. on Pattern Recognition (ICPR), IEEE Computer Society, pp. 683-686, vol. II, 2004
10. Mr. Rajesh H. Davda1, Mr. Noor Mohammed, " Text Detection, Removal and Region Filling Using Image Inpainting", International Journal of Futuristic Science Engineering and Technology, vol. 1 Issue 2, ISSN 2320 – 4486, 2013
11. Uday Modha, Preeti Dave, " Image Inpainting-Automatic Detection and Removal of Text From Images", International Journal of Engineering Research and Applications (IJERA), ISSN: 2248-9622 Vol. 2, Issue 2, 2012.
12. Muthukumar S, Dr.Krishnan .N, Pasupathi.P, Deepa. S, "Analysis of Image Inpainting Techniques with Exemplar, Poisson, Successive Elimination and 8 Pixel Neighborhood Methods", International Journal of Computer Applications (0975 – 8887), Volume 9, No.11, 2010