ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

Contextual Learning Approach to Improve Diagnostic Accuracy for Hybrid Disease

R.DELPHINE CLARE[1], PG Scholar, Department of Computer Science and Engineering,
R.V.S Educational Trust's Group Of Institution, Dindigul, Tamil Nadu.

Dr.J.AVANIJA[2] ,ME,Ph.D[2],Head of the Department, Department of Computer Science and Engineering,
R.V.S Educational Trust's Group Of Institution, Dindigul, Tamil Nadu.

Abstract- **Clinicians need to routinely make management decisions about patients who are at risk for a disease such as multiple diseases. This paper designs a clinical decision support system detecting multiple diseases. Prediction of disease is proved using proposed data mining algorithm for improve prediction accuracy. The project proposes a multiple diseases predict system with high accuracy using Neural Network, Naive bayes, Decision tree, Association functional rule, Regression tree. To predict multiple diseases through multiple data mining algorithms for medical application. The performance of the BPN classifier was evaluated in terms of training performance and classification accuracies.**

Keywords-**Contextual Learning technique, Data mining Algorithms, Computer aided Disease diagnosis system, multiple diseases.**

I. INTRODUCTION

1.1 OVERVIEW OF DISEASE DIAGNOSIS

Disease Diagnosis is one of the applications where data mining tools are proving successful results. Now a day's more number of disease leading cause of death all over the world. Using multiple data mining techniques in the diagnosis of all disease has been comprehensively investigated showing acceptable levels of accuracy. However, using data mining techniques to identify a suitable treatment for all disease patients has received less attention. This paper identifies gaps in the research on all disease diagnosis and treatment. It proposes a model to systematically close those gaps to discover. Applying multiple data mining techniques to all disease treatment data can provide as reliable performance as that achieved in diagnosis all diseases. The health care industry collects huge amount of health care data which, unfortunately, are not mined to discover hidden information for effective decision making. Discover of hidden patterns and relationship often goes unexploited. Advance data mining techniques can helps to resolve these problems.

1.2 INTELLIGENT HYBRID DISEASE PREDICTION SYSTEM:

In this paper a prototype Intelligent Hybrid Disease Prediction System (IHDPS) using data mining techniques, namely, Neural Network, Naive Bayes, Decision tree, Association functional rule, Regression tree has been developed. Results show that each technique has its unique strength in realizing the objectives of the defined mining goals. Intelligent Hybrid Disease Prediction System can answer complex what if queries which traditional decision support systems cannot. Using medical profiles such as age, sex, Hoarseness, Random, Asbestos, Tobacco smoking, Marijuana, Recurring Inflammation, Family history, chest pain type it can predict the likelihood of patients getting all diseases. It enables significant knowledge, e.g. patterns, relationship between medical factors related to all diseases, to be established. Intelligent Hybrid Disease Prediction System is Web-based, user friendly, scalable, reliable and expandable.
There are three predictive focus of disease prognosis

- Prediction of disease susceptibility
- Prediction of disease recurrence and
- Prediction of disease survivability.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

II .LITERATURE REVIEW

The authors of the paper focused on to improve health awareness. Current methods of detecting disease such as computed tomography scans are time-consuming. Michelle Thurston et al use the electronics nose detection system for detection of disease [10]. The author of the paper focus on proposed Mining and Integrating Reliable Decision Rules for Imbalanced Gene Expression Data Sets [4]. The authors of the paper focus on automated 3-D Segmentation of Lungs with Lung Cancer in CT Data Using a Novel Robust Active Shape Model Approach [8]. The author of the paper focused on SEER related processing using Ensample data mining technique to find out the some diseases [1].

III. EXISTING SYSTEM DISADVANTAGES

The single data mining technique is used to diagnose all diseases. There is no previous research that identifies which data mining technique can provide more reliable accuracy in identifying suitable treatment for all disease patients. Practical use of healthcare database systems and knowledge discovery is difficult in all disease diagnosis. Practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Hospitals do not provide the same quality of service even though they provide the same type of service. There is no previous research that identifies which data mining technique can provide more reliable accuracy in identifying suitable treatment for all disease patients. It takes more time consumption for practical use of healthcare database systems. Not applicable for detect the diseases in short time. Structure of disease function is hard to find some diseases.

IV.PROPOSED WORK

In proposed method applying multiple data mining techniques to identifying suitable treatments for all disease patients. Multiple data mining techniques to apply all disease diagnosis benchmark dataset to establish baseline accuracy for each single data mining technique in the diagnosis of all disease patients. Apply the same multiple data mining techniques used in all disease diagnosis to all disease treatment dataset to investigate if single data mining techniques can achieve equivalent (or better) results in identifying suitable treatments as that achieved in the diagnosis. Hybrid data mining techniques to all disease diagnosis benchmark dataset to establish baseline accuracy for each hybrid data mining technique in the diagnosis of all disease patients.

V. DATA MINING PROCESS

Data mining methods are extracting patterns from data. The patterns can be discovered depend upon the data mining tasks applied. Generally, there are two types of data mining tasks: *descriptive data mining tasks* that describe the general properties of the existing data, and *predictive data mining tasks* that attempt to do predictions based on available data. Data mining can be done on data which are in quantitative, textual, or multimedia forms.

Data mining applications can use different kind of parameters to examine the data. They include association (patterns where one event is connected to another event), sequence or path analysis (patterns where one event leads to another event), classification (identification of new patterns with predefined targets) and clustering (grouping of identical or similar objects).

(1) *Problem definition:* The first step is to identify goals. Based on the defined goal, the correct series of tools can be applied to the data to build the corresponding behavioural model.

(2) *Data exploration:* If the quality of data is not suitable for an accurate model then recommendations on future data collection and storage strategies can be made at this. For analysis, all data needs to be consolidated so that it can be treated consistently.

(3) *Data preparation:* The purpose of this step is to clean and transform the data so that missing and invalid values are treated and all known valid values are made consistent for more robust analysis.

(4) *Modelling:* Based on the data and the desired outcomes, a data mining algorithm or combination of algorithms is selected for analysis. These algorithms include classical techniques such as statistics, neighbourhoods and clustering but also next generation techniques such as decision trees, networks and rule based algorithms. The specific algorithm is selected based on the particular

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

objective to be achieved and the quality of the data to be analysed.

(5) *Evaluation and Deployment:* Based on the results of the data mining algorithms, an analysis is conducted to determine key conclusions from the analysis and create a series of recommendations for consideration.

VI. DIAGNOSTIC RECOMMENTATION ALGORITHM

The main idea of our diagnostic recommendation approach is to adaptively cluster patients based on related contexts and then learn the best action for each patient cluster.

NEURAL NETWORK:

It contains *class* of functions F, learning means using a set of *observations* to find $f^* \in F$ which solves the task in some *optimal* sense. This entails defining a cost function $C : F \to \mathbb{R}$ such that, for the optimal solution f^*, $C(f^*) \leq C(f)\ \forall f \in F$.These same type of methodology used to discover the disease accuracy. Neural networks are biologically inspired highly interconnected cells. The perception is the simplest architecture which has one neuron and a learning method. More sophisticated architecture is multi-layer neural networks (MLP) which one or more neurons connected at different layers. Neural networks can be trained to learn a classification task and to predict diseases

NAÏVE BIASE CLASSIFICATION:

Bayes' theorem, the conditional probability can be followed as

$$p(C_k|\mathbf{x}) = \frac{p(C_k)\ p(\mathbf{x}|C_k)}{p(\mathbf{x})}$$

From a Bayesian viewpoint, a classification problem can be written as the problem of finding the class with maximum probability given a set of observed attribute values. Such probability is seen as the posterior probability of the class given the data, and is usually computed using the Bayes theorem. Estimating this probability distribution from a training dataset is a difficult problem.

DECISION TREE:

Decision Trees (DTs) are a non-parametric supervised learning method used for classification. The main aim is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. The structure of decision tree is in the form of a tree. Decision trees classify instances by starting at the root of the DISEASE and moving through it until a other problems. Decision trees are commonly used in operations research, mainly in decision analysis. Some of the advantages are they can be easily understand and interpret, robust, perform well with large datasets, able to handle both numerical and categorical data. Decision-tree learners can create over-complex trees that do not generalize well from the training data is one the limitation.

ASSOCIATION FUNCTIONAL RULE:

Association is a data mining function that discovers the probability of the co-occurrence of items in a collection. The relationships between co-occurring items are expressed as association rules. Association rules are often used to analyse sales transactions. Here patients' symptoms are associated to different direction. And it gives the accurate result. Association rules represent a promising technique to improve heart disease prediction. Unfortunately, when association rules are applied on a medical data set, they produce an extremely large number of rules. Most of such rules are medically irrelevant and the time required to find them can be impractical, four constraints were proposed to reduce the number of rules: item filtering, attribute grouping, maximum item set size, and antecedent/consequent rule filtering. When association rules are applied on a medical data set, they produce an extremely large number of rules. Most of such rules are medically irrelevant and the time required to find them can be impractical. A more important issue is that, in general, association rules are mined on the entire data set without validation on an independent sample

REGRESSION TREE:

Regression-type problems are generally those where we attempt to predict the values of a continuous variable from one or more continuous and/or categorical predictor variables. The regression tree mostly used to finding the variables of the disease. Logistic regression is a generalized form of linear regression. It is used primarily for predicting the binary or multi-class dependent variables. As the response variable is

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

discrete, it cannot be modeled directly by linear regression. Therefore, rather than predicting point estimate of the event itself, it builds the model to predict the odds of the occurrence. In a two-class problem, odds greater than 50% would mean that the case is assigned to the class designated as ''1'' and ''0'' otherwise. While logistic regression is a very powerful modeling tool, it assumes that the response variable (the log odds, not the event itself) is linear with respect to the predictor variables.

VII. CONCLUSION AND FUTURE WORK

In this paper examines the clinical decision support system proposed using multiple Data mining algorithms. The existing records are collected and maintained as a big medical dataset stored in data base. The process of multiple diseases diagnosis and Prediction is represented as a decision making with multiple data mining algorithms identified the disease and accuracy of the results. For using this algorithms easily predict the all types of disease with effectively and time consumption is less. In Future Work, applying data mining techniques in identifying suitable treatments for diseases with different patients. Apply various multiple data mining techniques to use various diseases.

REFERENCES

[1] Ankit Agrawal, sanchit Misra, Ramanathan Narayanan, Lalitha Polepeddi, Alok Choudhary (2011) "A Lung Cancer Outcome Calculator Using Ensemble Data Mining on SEER Data"

[2] Anitha, Balachandran (2013) "An efficient optimization based lung cancer pre-diagnosis system with aid of feed forward back propagation neural network (ffbnn)"

[3] Fatma Taher (2011) "Lung cancer detection by using fuzzy clustering methods"

[4] Hualong Yu Jun Ni2, Yuanyuan Dan, Sen Xu (2012) "Mining and Integrating Reliable Decision Rules for Imbalanced Cancer Gene Expression Data Sets"

[5] Krishnaiah, V.Krishnaiah, Dr.G.Narsimha, Dr.N.Subhash Chandra (2013) "Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques"

[6] Lihua Li et al (2009) "Data mining techniques for cancer detection using serum proteomic profiling"

[7] Rajneet Kaur (2013) "Using Some Data Mining Techniques to Predict the Survival Year of Lung Cancer Patient"

[8] Shanhuri son, Christian Bauer, and Reinhard Beichel member IEEE (2012) "Automated 3-D Segmentation of Lungs With Lung Cancer in CT Data Using a Novel Robust Active Shape Model Approach"

[9] Topon Kumar Paul (2009) "Prediction of Cancer Class with Majority Voting Genetic Programming Classifier Using Gene Expression Data"

[10] Vanessa Hetman, Hiang Ping Chan, Michelle Thurston, Paul Jackson, Craig Lewis, Deborah Yates, Graham Bell, Paul S.Thomas (2010) "Breath Analysis of Lung Cancer Patients Using an Electronic Nose Detection System"