© 2020 | International Journal of Scientific Research in Science and Technology
IJSRST | Volume 7 | Issue 2 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
DOI : https://doi.org/10.32628/IJSRST207225

Prediction of ARDS Syndrome and CAD Using Multilayer Perceptron

M. Mohanasundari.*[1], Prasanth. S[2]

*[1]Department of CSE, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India

[2]PG Scholar, Department of CSE, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India

ABSTRACT

Acute Respiratory Distress Syndrome (ARDS) and Coronary artery heart Disease (CAD) isa critical condition occurring in ill patients. Our proposed system is to predict ARDS and CAD in hospitalized patients using only physiological signals as heart rate and breathing rate. This based on hypothesis testing is developed to detect whether subjects signals deviate from their initial states. The approach is applied on mechanically ventilated subjects in the MIMIC II database. Hybrid method for CAD diagnosis, including risk factor identification using correlation based feature subset selection with particle swam optimization search method and K-Means clustering algorithms. This proposed system is increasing the efficiency and accuracy of predicting the ARDS and CAD diseases.

Keywords : Multilayer Perceptron, MIMIC Database, Medical Datasets, Preprocessing, Value Imputation

I. INTRODUCTION

Data Mining is an analytic process designed to explore data (usually large amounts of data - typically business or market related) in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. The ultimate goal of data mining is prediction - and predictive data mining is the most common type of data mining and one that has the most direct business applications. The process of data mining consists of three stages: (1) the initial exploration, (2) model building (3) deployment

Acute respiratory distress disease and coronary artery disease Statistics suggest the possibility of diagnosing nearly 2.5 lakhs new cases in India by the year 2015. Prognosis thus takes up a significant role in predicting the course of the Disease even in women who have not succumbed to the Disease but are at a greater risk to. Classification of the nature of the Disease based on the predictor features will enable oncologists to predict the possibility of occurrence of acute respiratory distress Disease for a new case. The dismal state of affairs where more people are conceding to the sway of acute respiratory distress Disease, in spite of remarkable advancement in clinical science and therapy is certainly perturbing. The motivation for research on classification, to accurately predict the nature of acute respiratory distress disease and Coronary artery heart disease. This proposes work mainly focuses on building an efficient classifier for the Wisconsin Prognostic Acute respiratory distress Disease (WPDC) data set and Coronary artery heart disease from the UCI machine learning repository.

II. RELATED WORK

A. DIETARY FAT REDUCTION AND ACUTE RESPIRATORY DISTRESS DISEASE OUTCOME

An algorithm to perform outlier detection on time-series data is developed, the intelligent outlier detection algorithm (IODA). This algorithm treats a time series as an image and segments the image into clusters of interest, such as "nominal data" and "failure mode" clusters. The algorithm uses density clustering techniques to identify sequences of coincident clusters in both the time domain and delay space, where the delay space representation of the time series consists of ordered pairs of consecutive data points taken from the time series. "Optimal" clusters that contain either mostly nominal or mostly failure-mode data are identified in both the time domain and delay space. A best cluster is selected in delay space and used to construct a "feature" in the time domain from a subset of the optimal time-domain clusters. Segments of the time series and each datum in the time series are classified using decision trees. Depending on the classification of the time series, a final quality score or quality index for each data point is calculated by combining a number of individual indicators. The performance of the algorithm is demonstrated via analyses of real and simulated time-series data.

B.THE BURDEN OF CANCER

Detecting outlier efficiently is an active research issue in data mining, which has important applications in the field of fraud detection, network intrusion detection, monitoring criminal activities in electronic commerce, etc. Because of the sparsely of high dimensional data, it is reasonable and meaningful to detect the outliers in suitable projected subspaces. Subspace and outliers in the subspace as anomaly subspace and projected outlier respectively. Many efficient algorithms have been proposed for outlier detection based on different approaches, but there are few literatures on projected outlier detection for high dimensional data sets with mixed continuous and categorical attributes. In this system, a novel projected outlier detection algorithm is proposed to detect projected outliers in high-dimensional mixed attribute data set.

C.CANCER TREATMENT

Interestingness measures play an important role in data mining, regardless of the kind of patterns being mined. These measures are intended for selecting and ranking patterns according to their potential interest to the user. Good measures also allow the time and space costs of the mining process to be reduced. The interestingness measures for rules and summaries, classifies them from several perspectives, compares their properties, identifies their roles in the data mining process, gives strategies for selecting appropriate measures for applications, and identifies opportunities for future research in this area.

D. ACUTE RESPIRATORY DISTRESS

The memory consumption which permits the discovery of all outliers by scanning the dataset 3 times. The upper bound turns out to be extremely low in practice. Since the actual memory capacity of a realistic DBMS is typically larger, we develop a novel algorithm, which integrates our theoretical findings with carefully designed heuristics that leverage the additional memory to improve I/O efficiency. Our technique reports all outliers by scanning the dataset at most twice, and significantly outperforms the existing solutions by a factor up to an order of magnitude.

III. METHODOLOGY

A. EXISTING SYSTEM

It has classified only undesirable effect of changing a Acute respiratory distressCoronary artery heart

disease patient's existing test data groups, potentially undoing the patient's own manual efforts in organizing her history. It involves a high computational cost, have to repeat a large number of attribute test data group similarity computations for every new test data.

Existing medical systems, including hospital management systems and decision making systems, focus on collecting and mining the entire medical data. The entire patient records are loaded and all factors are considered. The medical data cannot be easily analyzed, because for generating a probabilistic rating, not only symptoms but also factors like test results, current epidemics, medical history, external climate conditions, and various other factors are required, which may or may not be present in the report.

The C4.5 model that we previously developed. On the basis of this new measure, a data mining algorithm was developed to mine the causal relationship between drugs and their associated with risk for CAD heart Disease. The exclusive causal-leverage was employed to rank the potential causal associations between each of the selected drugs. Algorithm could effectively make known ADRs rank high among all the symptoms in the database.

DRAWBACKS

- Existing systems have failed to utilize and understand the importance of misdiagnosis.
- Improper classification may provide wrong results
- Poor performance
- Complex data processing to find Acute respiratory distress prediction
- The data retrieval based on user requirement is not done
- This relation-type information, however, is often not readily available in Acute respiratory distress prediction

B. PROPOSED WORK

In the proposed work user will search for the heart Disease diagnosis (heart Disease and treatment related information) by giving symptoms as a query in the search engine. These symptoms are preprocessed to make the further process easier to find the symptoms keyword which helps to identify the heart Disease quickly.

The symptoms which keyword is matched with the stored medical input database to identify the multiple heart Diseases related to that keyword. Multiple heart Diseases is identified, it will make the pattern matching about the multiple heart Diseases and also find the probability of heart Diseases. Then the heart Disease will make a differential diagnosis to find the heart Disease accuracy.

The keyword which is a preprocessed symptom is matched with the heart Diseases stored in the local database to identify the corresponding heart Disease related to those symptoms given by the user. This has to search a record database of more than 20000 heart Diseases and even more symptoms, which is very time consuming, so CFS+PSO classification was applied to classify heart Diseases features into subgroups.

If a group of symptoms match higher preference is given to that subgroup and searching in that new smaller subgroup thus reduces database access. In pattern recognition, CFS with PSO Feature Selection algorithm is a method for classifying objects based on closest training examples in the feature space. CFS+PSO are a type of instance-based learning, or lazy learning where the function is only

approximated locally and all computation is deferred until classification. This feature has been identified as the most suitable for the present system.

Methods that can accurately predict acute respiratory distress Disease are greatly needed and good prediction techniques can help to predict acute respiratory distress Disease more accurately. In this system, it used two feature selection methods, forward selection (FS) and backward selection (BS), to remove irrelevant features for improving the results of Acute respiratory distress Disease prediction. The results show that feature reduction is useful for improving the predictive accuracy and density is irrelevant feature in the dataset where the data had been identified on full field digital mammograms collected at the UCI Repository. In addition, decision tree (DT), support vector machine—Multi Layer Perceptron (SVM-MLP) and their ensembles were applied to solve the Acute respiratory distress Disease diagnostic problem in an attempt to predict results with better performance. The results demonstrate that ensemble classifiers are more accurate than a single classifier.

The proposed framework MLP based on disease prediction is shown to be effective in addressing this prediction. The framework suggests a novel way of network classification: first, capture the latent affiliations of actors by extracting disease prediction based on network connectivity, and next, apply extant data mining techniques to classification based on the extracted prediction. In the initial study, modularity maximization was employed to extract disease prediction. The superiority of this framework over other representative relational learning methods has been verified with acute respiratory distress prediction acute respiratory distress data.

1. SUPPORT VECTOR MACHINE:

Feature reduction applies a mapping of the multidimensional space into a space of lower dimensions. Feature extraction includes features construction, space dimensionality reduction, sparse representations, and feature selection all these techniques are commonly used as preprocessing to machine learning and statistics tasks of prediction, including pattern recognition. Although such problems have been tackled by researchers for many years, there has been recently a renewed interest in feature extraction. The feature space having reduced features truly contributes to classification that cuts preprocessing costs and minimizes the effects of the „peaking phenomenon" in classification. Thereby improving the overall performance of classifier based intrusion detection systems.

2. MULTI LAYER PERCEPTRON:

Classification is the type of Data mining, which deals with the problematic things by recognizing and detecting features of infection, among patients and forecast that which technique shows top performance, on the base of WEKA's outcome. Five techniques have been used in this paper. These techniques uses Explorer interface and it depends on dissimilar techniques NB, REP Tree, RT, J48 and MLP.

IV. ARCHITECTURE

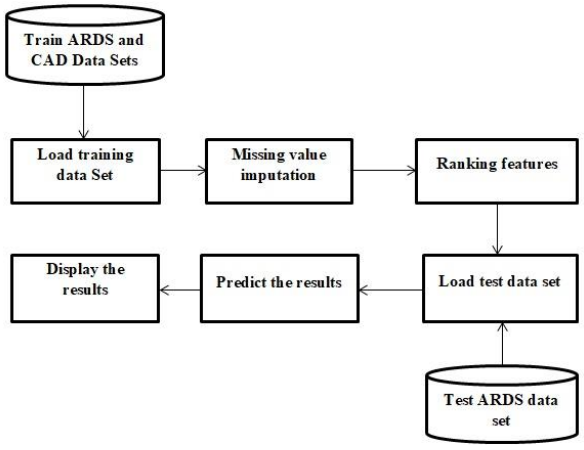


Fig 1. Prediction of ARDS and CAD

V. MODULE DESCRIPTION

A. DATA PRE-PROCESSING

The Wisconsin Prognostic Cleave Land Train Dataset is downloaded from the UCI Machine Learning Repository website and saved as a text file. This file is then imported into Excel spreadsheet and the values are saved with the corresponding attributes as column headers. The missing values are replaced with appropriate values. The ID of the patient cases does not contribute to the classifier performance. Hence it is removed and the outcome attribute defines the target or dependent variable thus reducing the feature set size to 33 attributes. The algorithmic techniques applied for feature relevance analysis and classification are elaborately presented in the following sections.

B. SMO FEATURE SELECTION

The generic problem of supervised feature selection can be outlined as follows. Given a data set {(xi, yi)} ni=1 where xi ∈ Rd and yi∈ {1, 2…c}, we aim to find a feature subset of size m which contains the most informative features. The two well-performing feature selection algorithms on the WPBC dataset are briefly outlined below.

C. LOGISTIC REGRESSION RISK ANALYSIS

When the number of descriptors is very large for a given problem domain, a learning algorithm is faced with the problem of selecting a relevant subset of featuresBackward regression includes regression models in which the choice of predictor variables is carried out by an automatic procedure. The iterations of the algorithm for logistic regression are given in steps as stated as follows.

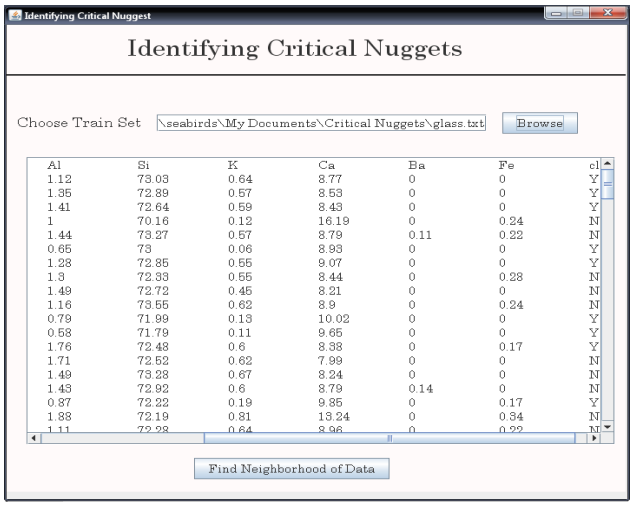
Step 1: The feature set with all 'ALL' predictors.
Step 2: Eliminate predictors one by one.
Step 3: 'ALL' models are learnt containing 'ALL-1' descriptor each.

D. FEATURE REDUCTION BY SMO

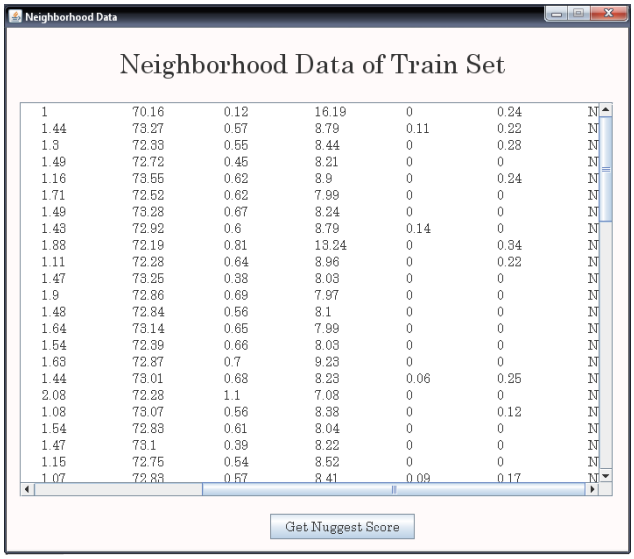
Feature reduction applies a mapping of the multidimensional space into a space of lower dimensions. Feature extraction includes features construction, space dimensionality reduction, sparse representations, and feature selection all these techniques are commonly used as preprocessing to machine learning and statistics tasks of prediction, including pattern recognition. Although such problems have been tackled by researchers for many years, there has been recently a renewed interest in feature extraction. The feature space having reduced features truly Contributes to classification that cuts preprocessing costs and minimizes the effects of the 'peaking phenomenon' in classification. Thereby improving the overall performance of classifier based intrusion detection systems. The commonly used dimensionality reduction methods include supervised approaches such as linear discriminant analysis (LDA), unsupervised ones such as Sequential minimal Optimization (SMO), and additional spectral and manifold learning methods [9]. It converts a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.

VI. SCREENSHOTS FOR OUTPUT

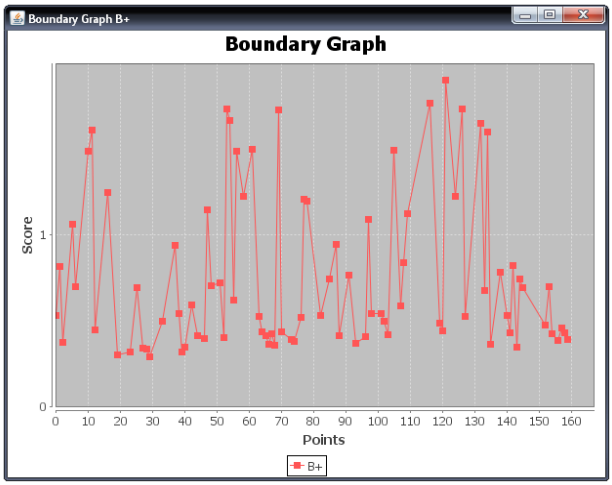
A. FINDING NEIGHBOURHOOD OF DATA

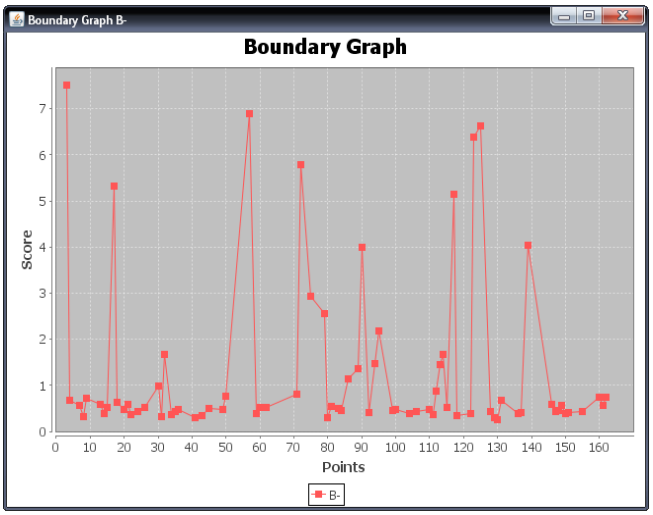


B. NEIGHBOURHOOD DATA OF TRAIN SET



C. BOUNDARY GRAPH





VII. CONCLUSION

Medical diagnosis is an important area of research which helps to identify the occurrence of a heart Disease. The system, making use of various techniques mentioned, will in turn exposes the root heart Disease and lung diseases along with the set of most probable Diseases which have similar symptoms. The project presents a novel hybrid model to identify and confirm CAD and ARDS cases at low cost by using clinical data that can be easily collected at hospitals. Complexity of the system is decreased by reducing the dimensionality of the data set with PSO. It provides reproducible and objective diagnosis, and hence can be a valuable adjunct tool in clinical practices. Results are comparably, promising and therefore the proposed hybrid method will be helpful in heart Disease diagnostics. Experiment results demonstrate the superiority of the proposed hybrid method with regard to prediction accuracy of CAD and ARDS with the features selected by CFS &PSO, we need only a few clinical data to apply this model. The accuracy can be further increased with more data instances.

VIII. REFERENCES

[1]. Blackburn G L,Wang K A (2007) "Dietary fat reduction and Acute respiratory distress Disease outcome: results from the Women's Intervention Nutrition Study (WINS)", IEEE International Journal of Computer Science and Engineering, vol. 32, pp.512.

[2]. Boffetta P, Hashibe M (2006). "The burden of cancer attributable to alcohol drinking", IEEE Journal of Cancer Research and Treatmentvol.90, pp.119.

[3]. Pasche (2010). "Cancer Genetics", (Cancer Treatment and ResearcSh). Berlin: Springer. pp. 19–20.

[4]. Collaborative Group on Hormonal Factors in Acute respiratory distress Disease (2002), "Acute respiratory distress Disease and breastfeeding",IEEE International Journal for Cancer Research and Treatment, vol.15.

[5]. Walker G, (2005),"Predicting Acute respiratory distress Disease survivability: a comparison of three data mining methods", Artificial Intelligence in Medicine, vol. 34, pp. 113-127.

[6]. Ferro, Roberto (2012), "Pesticides and Acute respiratory distress Disease ",IEEEInternational Journal for Cancer Research and Treatment, vol.76.

[7]. Gage M, Wattendorf, D(2012), "Translational advances regarding hereditary Acute respiratory distress Disease syndromes". IEEE International Journal of Computer Science and Engineering, vol. 90.

[8]. Grey N and Sener S. (2006),"Reducing the global cancer burden", IEEE Journal of Computer Science and Engineering. Vol.45, pages 201-202.

[9]. Gupta, S.; Kumar, D., Sharma, A (2011). "Data Mining Classification Techniques Applied for Acute respiratory distress Disease Diagnosis and Prognosis".IEEE Journal of Computer Science and Engineering. Vol.23, pages 1191-1193.

[10]. Hendrick, RE (2010). "Radiation doses and cancer risks from breast imaging studies", IEEE International Journal for Cancer Research and Treatment, vol. 57.

Cite this article as :

Sh M. Mohanasundari, Prasanth. S, "Prediction of ARDS Syndrome and CAD Using Multilayer Perceptron", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 7 Issue 2, pp. , March-April 2020. Available at doi : https://doi.org/10.32628/IJSRST207225

Journal URL : http://ijsrst.com/IJSRST207225