Detection Of Acute Respiratory Distress Syndrome Using Support Vector Machine Model In Machine Learning

Agathiya S[1], Diana K[2], Kamali K[3], Dr.Chithra R[4]
[1, 2, 3] Dept of Information Technology
[4]Professor, Dept of Information Technology
[1, 2, 3, 4] K.S.Rangasamy College of Technology, Tiruchengode, Tamil Nadu

Abstract- *When training a machine learning algorithm for a supervised-learning task in some clinical applications, uncertainty in the correct labels of some patients may adversely affect the performance of the algorithm. For example, even clinical experts may have less confidence when assigning a medical diagnosis to some patients because of ambiguity in the patient's case or imperfect reliability of the diagnostic criteria. As a result, some cases used in algorithm training may be mis-labeled, adversely affecting the algorithm's performance.We present Support Vector Machine model along with classification algorithm like random forest,naive bayes and decision tree algorithms for to increase the accuracy level of syndrome detection. We apply supervised learning algorithms. We can able to improve the performance of SVM algorithm to detect the patient with ARDS on hold-on or hold-out and we finalize the accuracy level.*

Keywords- Machine learning algorithms,support vector machine, label uncertainty, acute respiratory distress syndrome, accuracy level.

I. INTRODUCTION

Acute respiratory distress syndrome (ARDS) occurs when fluid builds up in the tiny, elastic air sacs (alveoli) in your lungs. The fluid keeps your lungs from filling with enough air, which means less oxygen reaches your bloodstream. This deprives your organs of the oxygen they need to function.ARDS typically occurs in people who are already critically ill or who have significant injuries. Severe shortness of breath — the main symptom of ARDS — usually develops within a few hours to a few days after the precipitating injury or infection.Many people who develop ARDS don't survive. The risk of death increases with age and severity of illness. Of the people who do survive ARDS, some recover completely while others experience lasting damage to their lungs.Globally, ARDS affects more than 3 million people a year.The condition was first described in 1967.Although the terminology of "adult respiratory distress syndrome" has at times been used to differentiate ARDS from "infant respiratory distress syndrome" in newborns, the international consensus is that "acute respiratory distress syndrome" is the best term because ARDS can affect people of all ages.Acute Respiratory Distress Syndrome (ARDS) can be detected by using algorithms.We applied semisupervised learning algorithms to process the dataset.A supervised learning algorithm learns from labeled training data, helps you to predict outcomes for unforeseen data. .Unsupervised learning is a machine learning technique, where you do not need to supervise the model. It mainly deals with the unlabelled data.Unsupervised learning algorithms allow you to perform more complex processing tasks compared to supervised learning. Although, unsupervised learning can be more unpredictable compared with other natural learning deep learning and reinforcement learning methods.Report of the incidence of ARDS in its different grades of severity vary, to some extent due to the lack of precision in the earlier AECC definition. Incidence estimates range 10–58 cases per 100 000 people, depending on geographical location and on the reporting system used. Using data from a prospective multicentre European cohort study that included 6522 patients treated in ICU, the proportion with ALI and ARDS averaged 7.1% of all patients admitted to critical care. This rose to 12.5% when only patients treated for more than 24 hours in ICU were included. Another study reported that patients with ALI represented 4.5% of all those receiving ventilation at the time of admission to intensive care.In a recent database analysis from a single ICU treating both surgical and medical patients, a decrease in the prevalence of ARDS was reported.Overall, the mortality of patients suffering from ARDS remains unacceptably high despite our extensive knowledge of the pathophysiology of the lung injury and the various multicentre studies of treatment reported to date.

II. EXISTING SYSTEM

Our sampling method outperforms using all available data (no sampling) from the EHR by producing a much balanced dataset for training and minimizing dependencies in each patient's time series data, making it closer to the state of being. We also compared our sampling algorithm to randomly sampling on negative examples to yield a 2:1 negative to positive ratio from each patient. This random sampling method also provides a balanced dataset for training, and as a result, we observed an increase in accuracy and AUROC from all algorithms when compared to training without sampling.

III. PROPOSED SYSTEM

We have proposed this approach in the ARDS dataset which is a binary data implementing SVM algorithm along with classification algorithms.In this proposed approach,we perform the dataset in both test and training model. We have approached this model through three modules -

- Data collection & Data preprocessing
- Performing algorithms
- Output/accuracy level detection.

IV. SYSTEM MODULES

4.1 DATA COLLECTION & PREPROCCESSING

Dataset we have collected from the available source in kaggle website. This dataset consists of a report values of the patients which is binary dataset.We have collected datasets on the basis of certain features like age, patient_id, diabetic,bp,etc.

Preproccessing is an main process which involves the process of data mining. This process is applicable to the data mining and machine learning projects. It includes cleaning,Normalization,transformation and selection and so on. This process is said to be one of the final training set. The data pre-processing process will also include the checking of the quality of a data and this is considered as the very first and foremost form of analysis. Data pre-processing is considered to be the most important phase of machine learning algorithms.

Data cleaning

This Data cleaning method is used to fill the missing values, smooth noisy data, identify or remove outliers and resolve inconsistencies.We used this method for to obtain noise free data for processing the algorithms.

Normalization

This normalization method we used to avoid the null values from our dataset. It removes all the unwanted and null values which fails the condition. If the datasets is too large it takes more time for processing because it contains unused and null values so we present this method for to avoid confusion and to obtain accurate datasets.

Feature selection

A feature selection method is proposed to select a subset of variables in principal component analysis (PCA) that preserves as much information present in the complete data as possible.The results showed that the proposed method successfully identified structure-bearing variables in both data sets and that it leads to a better subset of variables than other studied feature selection methods.In this model,we used five classification algorithms such as linear regression classifier, logistic regression classifier,k-nearest neighbors classifier,decision tree,naive bayes classifier.

Linear regression classifier

Linear Regression is a supervised machine learning algorithm widely used for data analysis. We used this regression technique for to analyze our datasets and to provide continuous constant output.

Logistic regression classifier

Logistic Regression is a Machine Learning algorithm which is used for the classification problems, it is a predictive analysis algorithm and based on the concept of probability.The hypothesis of logistic regression tends it to limit the cost function between 0 and 1.

4.2 MACHINE LEARNING ALGORITHM

Acute respiratory distress syndrome (ARDS) is a serious respiratory condition In which Supervised machine learning predictions may help to predict patients accuracy level

Support vector machine(SVM)

“Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. Support Vectors are simply the co-ordinates of individual observation.We proposed this formulation for account label uncertainty in the classification model in the following manner.

Naive Bayes classifier

It this classification technique based on Bayes' Theorem with an assumption of independence among a predictors. we proposed a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

Random forest classifier

We proposed in this Random forest is a supervised learning algorithm which is used for both classification and regression.we Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them values and finally selects the solution.

Decision Tree classifier

We proposed Decision Tree classifier as a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcomes of the value. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node.

4.3 PREDICTION

We can predict the uncertainty and also we can predict the accuracy level of Acute Respiratory Distress Syndrome (ARDS) datasets using the algorithms and methods mentioned in the above modules.

V. CONCLUSION

In this method,we proposed this project in three modules which provides the result of several classification algorithms.At the end of the process,we achieve the accuracy rate for the Acute Respiratory Distress Syndrome (ARDS) using classification algorithms.We obtain the result in table 1.We determined the Training dataset and prediction Time of Acute Respiratory Distress Syndrome.we achieve the best accuracy level 0.75 in the naive bayes the best F1-Score 0.79 in the Support vector machine and the best precision 0.80 in the Support vector machine.Also we get high performance level of Training and Testing dataset, All output results we get Satisfaction.

VI. ACKNOWLEDGEMENT

We are very proudly rendering our thanks to our Principal **Dr. R. GOPALAKRISHNAN, M.E., Ph.D.,** for the facilities and the encouragement given by him to the progress and completion of our project.

We proudly render our immense gratitude to the Head of the Department **Dr.A.SABARI M.Tech.,M.B.A., Ph.D.,** for his effective leadership, encouragement and guidance in the project.

We are highly indebted to provide our heart full thanks to our guide Professor **Dr.R.CHITHRA,B.E.,M.E.,Ph.D.,** for her valuable ideas, encouragement and supportive guidance throughout the project.

We wish to extend our sincere thanks to all faculty members of our Information Technology Department for their valuable suggestions, kind co-operation and constant encouragement for successful completion of this project.

We wish to acknowledge the help received from various Departments and various individuals during the preparation and editing stages of the manuscript.

REFERENCES

[1] Accounting for Label Uncertainty in Machine Learning for Detection of Acute Respiratory Distress Syndrome,IEEE Journal Biomedical Health Information,Jan 2019.

[2] G.D. Rubenfeld et al, Incidence and outcomes of acute lung injury, N Engl J Med, vol. 353(16), pp. 1685-1693. Oct 2005.

[3] R.M. Sweeney, D.F. McAuley. Acute respiratory distress syndrome, Lancet, vol. 388(10058), pp. 2416-2430. Nov 2016.

[4] G. Bellani et al, Epidemiology, Patterns of Care, and Mortality for Patients With Acute Respiratory Distress Syndrome in Intensive Care Units in 50 Countries, JAMA, vol. 315(8), pp. 788-800. Feb 2016.

[5] B.J. Clark, M. Moss, the Acute Respiratory Distress Syndrome: Dialing in the Evidence? JAMA, vol. 315(8), pp. 759-761. Feb 2016.

[6] M.W. Sjoding, R.C. Hyzy, Recognition and Appropriate Treatment of the Acute Respiratory Distress Syndrome Remains Unacceptably Low, Crit Care Med, vol. 44(8), pp. 1611-1612. Aug 2016.

[7] M.W. Sjoding, Translating evidence into practice in acute respiratory distress syndrome: teamwork, clinical decision support, and behavioral economic interventions, Curr Opin Crit Care. Jul 2017.

[8] H.C. Koenig et al, Performance of an automated electronic acute lung injury screening system in intensive care unit patients, Crit Care Med,vol. 39(1), pp. 98-104. Jan 2001.

[9] G.D. Rubenfeld et al, Interobserver variability in applying a radiographic definition for ARDS, Chest, vol. 116(5), pp. 1347-53. Nov 1999.

[10] M.W. Sjoding et al, Acute Respiratory Distress Syndrome Measurement Error: Potential Effect on Clinical Study Results, Ann Am Thorac Soc,vol. 13(7), pp. 1123-8. Jul 2016.
[11] C.V. Shah et at, An alternative method of acute lung injury classification for use in observational studies, Chest; vol. 138(5), pp. 1054-1061. Nov 2010.
[12] D.F. Nettleton, A. Orriols-Puig, A. Fornells, A study of the effect of different types of noise on the precision of supervised learning techniques, Artificial Intelligence Reviews, Vol. 33(40), 2010.
[13] B. Frenay and M. Verleysen, "Classification in the Presence of Label Noise: A Survey," IEEE Transactions on Neural Networks and Learning Systems, vol. 25(5), pp. 845-869. May 2014.
[14] N. Natarajan et al, Learning with noisy labels," in Neural Information Processing Systems, pp. 1196-1204. December 2013.
[15] Y. Duan and O. Wu, Learning With Auxiliary Less-Noisy Labels, IEEE Transactions on Neural Networks and Learning Systems, vol. 28(7), pp.1716-1721. May 2017.
[16] S. Vembu and S. Zilles, "Interactive Learning from Multiple Noisy Labels," in Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer International Publishing. 2016,pp. 493-508.
[17] X. Yang, Q. Song, Y. Want, A weighted support vector machine for data classification, International Journal of Pattern Recognition and Artificial Intelligence, vol 21, Nov 5 (2007).
[18] E. Osuna, R. Freund, F. Girosi. An improved training algorithm for support vector machines, in Neural Networks for Signal Processing[1997] VII. Proceedings of the 1997 IEEE Workshop (pp. 276-285).
[19] J. Shawe-Taylor et al, "Structural risk minimization over data-dependent hierarchies," IEEE Transactions on Information Theory, vol. 44(5), pp.1926-1940, Sep 1998.
[20] R. Bellazzi and A. Riva, Learning conditional probabilities with longitudinal data, in Working Notes of the IJCAI Workshop Building Probabilistic Networks: Where Do the Numbers Come From, 1995 (pp.7-15).