International Journal of Engineering Sciences & Research Technology

(A Peer Reviewed Online Journal)
Impact Factor: 5.164

IJESRT

Chief Editor
Dr. J.B. Helonde

Executive Editor
Mr. Somil Mayur Shah

RESEARCHERID
THOMSON REUTERS

[Shubhangi *al.*, 13(4): April, 2024]
IC™ Value: 3.00

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

MACHINE LEARNING-BASED DETECTION AND COMPARATIVE ANALYSIS OF THE SWIFT RESPIRATORY DISTRESS SYNDROME (SRDS) BASED ON VOCAL

Dr.Shubhangi D C[1], Dr. Baswaraj Gadgay[2], Sumangala Pujari[3]

[1]Department of Computer Science, Vishveshwarayya Technological University CPGS Kalaburagi, Karnataka, India
Email:drshubhangipatil1972@gmail.com

[2]Department of Electronics and Communication, Vishveshwarayya Technological University CPGS Kalaburagi, Karnataka, India
Email:baswaraj_gadgay@vtu.ac.in

[3]Department of Computer Science, Vishveshwarayya Technological University CPGS Kalaburagi, Karnataka, India
Email: Sumangalapujari3@gmail.com

DOI: https://doi.org/10.57030/ijesrt.13.4.1.2024

ABSTRACT

Patients with severe conditions such as sepsis, pneumonia, are at increased risk for developing Swift Respiratory Distress Syndrome (SRDS), fulminant inflammatory lung damage. Unfortunately, many people who acquire SRDS are not diagnosed with the condition and so may not get therapy that might improve their prognosis. Due to the clinical nature of SRDS, diagnostic confusion (label uncertainty) may arise while treating a patient. In addition, a chest x-ray is necessary for the diagnosis; however, this is test that isnt always readily accessible in a clinical context. For this reason, we develop machine learning-based model for assessing the risk of SRDS, using patient's respiration sounds as both data for training and testing, as well as random forest, gradient boosting, & LR, & comparing results with SVM

KEYWORDS: Swift respiratory distress syndrome(SRDS), Random forest , gradient boosting, LR, SVM

1. INTRODUCTION

Tachypnea, severe hypoxemia, poor respiratory compliance, and lung tissue destruction seen on chest radiographs are the hallmarks of ARDS, a clinical condition [20]. Clinical characterization is used to make diagnoses of ARDS & its milder counterpart, acute lung injury (ALI), even if diffuse alveolar disruption is underlying pathological process [21]. In 2012, a new set of criteria, called as "Berlin definition" [22], was established for diagnosing ALI/ARDS in hospitals. More than 75,000 people in the United States lose their lives each year to ARDS [23], and it is linked to an overall mortality rate of 35% to 50% [24]. There is a lack of specific and sensitive methods for early diagnosis of ALI/ARDS, and the disease often progresses rapidly once it has been identified. Patients with ARDS have a far better chance of survival if they are diagnosed and treated quickly, as has been shown in both fundamental & clinical investigations [25]. Predicting ARDS occurrences is just as crucial, but no methods for doing so are being reported on yet. To further aid clinical diagnosis of ARDS, there's an urgent need for creation & testing of a prediction model for ARDS occurrences.

A biomarker is " a measurable and quantifiable attribute used to assess the efficacy of a drug or other treatment intervention, or to monitor the progress of a disease" [26], as defined by the National Institutes of Health in 2001. Pathophysiological pathways may be reflected by biomarkers, which might make them useful for diagnosing ARDS. Therefore, the diagnosis of ARDS may be improved by combining established clinical

IJESRT is licensed under a Creative Commons Attribution 4.0 International License.

classifications with validated biomarkers. Biomarkers may aid in risk stratification, outcome prediction, or act as surrogate endpoints to track therapies beyond just diagnosing ARDS. [27].

2. RELATED WORK

Used EHR data and chest x-rays to diagnose ARDS utilizing the Learning Information (LULUPAPI) paradigm [16], which is based on support vector machine. The test's AUC was improved by 3.55% when compared to SVM.

By the year 2021, SARS-CoV-2 will have infected approximately 165 million individuals throughout the globe, resulting in 3.4 million deaths from acute respiratory distress syndrome (ARDS). There are few aiSR, despite evidence that AI may aid in interpretation of biomedical images like X-rays and CT scans in diagnosis of ARDS [17]. This research aims to shed light regarding RoB of non- randomized AI trial for treating ARDS with use of innovative AP (ai) Bias. In order for our hypothesis to be accepted in low RoB studies, the average score across all studies must be over 80%. Forty-two of finest AI papers were assessed using the PRISMA framework for insight into the RoB. The top 19 research were selected using a raw-cutoff of 1.9 and the AP(ai)Bias paradigm. The data was derived by finding where the cumulative plot of "mean score vs. study" & distribution of scores intersected.

Amongst many intricate human body systems is respiratory system. In this paper[18], the authors suggest an electro-acoustic bio-impedance analysis- based mathematical model for respiratory system.

3. METHODOLOGY

Modeling using a random forest Breiman advocated the use of non-parametric and supervised ensembles machine learning tools as a way to handle classification and regression issues, and random forest has since become a popular machine learning model.6 Using Fisher's discriminant as a linear classifier at each node, the random forest is based on techniques to build forest of binary decision trees. The procedure ina collective decision tree uses a binary arithmetic technique to divide observations in two groups, or branches, & keeps doing this till the "tree" is fully formed ["node purity" is obtained]. Seventy percent of the data was used for model training, while remaining thirty percent was used for validation following random forest modelling. A 5-fold cross-validation procedure was used to get model's hyper-parameters in training set. More specifically, 80% of training set and 20% of testing set were randomly split up and used for testing and training purposes, respectively. To use 5-fold cross- validation, you must repeat this procedure five times, for a total of five possible results. Model's final tuned hyper-parameters were determined by averaging these five separate scores. At last, SRDS may be identified using random forest model.
Random Forest Algorithm:

Input: Patient Test data

1. First, for a given set of test data, randomly generate decision trees (D Trees) and record their results.
2. Add up all complaints from each group.
3. Proclaim winning group to be majority group.

Output: Final predicted SRDS

For tabular datasets, gradient boosting is a common machine learning approach. Great accessibility allows it to handle missing values, anomalies, as well as high cardinality categories upon your features with no any extra consideration, & its power allows it to detect any nonlinear connection between your model goal and features. Even though you may build basic gradient boosting trees with help of popular libraries like XGBoost or LightGBM without understanding the algorithm, you'll need to have some familiarity with it if you want to tune hyper-parameters, customize loss functions, etc., to improve your model's quality.

Gradient Boosting Algorithm

1. Initialize model with a constant value:

IJESRT is licensed under a Creative Commons Attribution 4.0 International License.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

$$F0(x) = \arg\min \sum_{i=1}^{n} L(yi, y)$$

2. For m=1 to M:

2-1. compute residuals

$$r_{in} = \left[\frac{\partial L(yi, F(xi))}{\partial F(xi)}\right]_{F(x)=Fm-1(x)}$$

For I=1….n

2-2. Train regression tree x with 'r' & create terminal node reasions Rjin for j=1,…,Jm

2-3 compute

$$y_{jin} = \arg\min \sum_{x \in Rin} L(yi\, F_{m-1}(x_i) + y) \text{ for } j=1,\ldots,J_m$$

2-4 Update the model:

$$F_m(x) = F_{m-1}(x) + v\sum_{i=1}^{jn} y_{in} 1(x \in R_{jin})$$

Let's demystify this line by line.

Step 1

Initialize model with a constant value:

$$F0(x) = \arg\min \sum_{i=1}^{n} L(yi, y)$$

Making a forecast of the constant number F0 is first order of business. In context of regression, L denotes a squared loss function.

$$L=(y_i-y)^2$$

When we use argmin, we want to find value of that minimizes ΣL (y_i,γ). Let's use our real loss function to calculate the value γ. We are calculating the derivative of ΣL with respect to in order to get the

value of which minimizes ΣL.

IJESRT is licensed under a Creative Commons Attribution 4.0 International License.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

$$\frac{\partial}{\partial\gamma}\sum_{i=1}^{n} r \quad \frac{\partial}{\partial\gamma}\sum_{i=1}^{n}(y_i-\gamma)^2$$

$$=-2\sum_{i=1}^{n}(y_i-\gamma)$$

$$=-2\sum_{i=1}^{n} y_i-2n\gamma$$

Deliberation of γ making $\partial\Sigma L/\partial\gamma$ equivalent to 0.

$$-2\sum_{i=1}^{n} y_i+2n\gamma=0$$

$$n\gamma=\sum_{i=1}^{n} y_i$$

$$\gamma=\frac{1}{n}\sum_{i=1}^{n} y_i=\bar{y}$$

It was discovered as mean of y is value of which minimizes ΣL. That's why we made our first forecast, F_0, based upon y-mean in previous section

$$F_0(x)=\lambda=y$$

Step2

2 for m=1 to M:
All of the operations in step2, from 2-1 to 2-4, are repeated M times. Tiny M specifies index of each tree, while M is the total number of trees being created.

Step 2–1
Compute residuals

$$r_{in}\left[\frac{\partial L(yi,F(xi))}{\partial F(xi)}\right]_{F(x)=Fm-1(x)}$$

For i=1....n

To determine residuals ri, we first take derivative of loss function when compared to prior forecast F and then multiply that result by 1. You can see that ri is calculated for each and every sample i by looking at the subscript index. There may be some confusion about the meaning of the term "ri residuals" in this context. This number is

IJESRT is licensed under a Creative Commons Attribution 4.0 International License.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

negative gradient that tells us how much and in what directions loss function may be reduced. In a moment, you'll see why we're referring to it as a residue. By the way, gradient descent method, which is often used to improve neural networks, is somewhat similar to this method where a gradient is utilized to reduce loss of your model. Here, we will calculate the residuals. F in equation stands for conclusion reached in previous iteration. The initial value is F0. The ri residuals are the focus of our equation solution.

$$r_{im} = \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$
$$= -\frac{\partial (y_i - F_{m-1})^2}{\partial F_{m-1}}$$
$$= 2(y_i - F_{m-1})$$

As a constant, we may eliminate 2 from the equation. Now we're in a r_i. Perhaps now you can understand the name "residuals." Intriguingly, this reveals which negative gradient which reveals direction & amount of loss minimization is just residuals.

Step 2–2

Create a decision node by training a regression tree using x features and r as R_{jin} for j=1,…,J_m

Terminal node (leaf) of tree is signified by letterj, tree index by letter m, & total number of leaves bycapital letter J.

Step 2–3: Compute

$$y_{jin} = \arg\min \sum_{x \in Rin} L(yi F_{m-1}(x_i) + y) \text{ for } j=1,\ldots,J_m$$

Finding γ_j such that loss function is minimized at every terminal node j is our goal. By aggregating lossacross all sample xis associated with terminal node Rj, as shown by $\Sigma x_i \in R_j$. Insert loss function in formula.

$$\gamma_{jm} = \underset{\gamma}{argmin} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma)$$
$$= \underset{\gamma}{argmin} \sum_{x_i \in R_{jm}} (y_i - F_{m-1}(x_i) - \gamma)^2$$

Now finding γ_j making derivative of Σ (*) equivalent to 0.

IJESRT is licensed under a Creative Commons Attribution 4.0 International License.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

$$\frac{\partial}{\partial \gamma} \sum_{x_i \in R_{jm}} (y_i - F_{m-1}(x_i) - \gamma)^2 = 0$$

$$-2 \sum_{x_i \in R_{jm}} (y_i - F_{m-1}(x_i) - \gamma) = 0$$

$$n_j \gamma = \sum_{x_i \in R_{jm}} (y_i - F_{m-1}(x_i))$$

$$\gamma = \frac{1}{n_j} \sum_{x_i \in R_{jm}} r_{im}$$

Please keep in mind that nj refers to total amount of samples in node j's terminal. This implies that average of residuals ri in last node Rj is optimum j which optimizes loss function. That is to say, j represents typical prediction values for regression trees, which are the mean of target values (residuals) at every leaf node.

Step 2–4

$$F_m(x) = F_{m-1}(x) + v \sum_{i=1}^{jn} y_{in} 1(x \in R_{jin})$$

The last process involves revising the forecast of the integrated model F. If variable x is located in the last branch Rj, then value γ_j is selected. Given a single x, there is only one terminal node that it can fit into, hence associated γ_j is appended to original prediction F to get the new prediction F. This is because all terminal nodes are mutually exclusive. To adjust how much the supplementary tree prediction contributes to overall prediction F, v is the learning rate, which may take values between 0 and 1. The likelihood of model overfitting training data is reduced along with impact of extra tree prediction when learning rate is decreased.

3. SYSTEM ARCHITECTURE

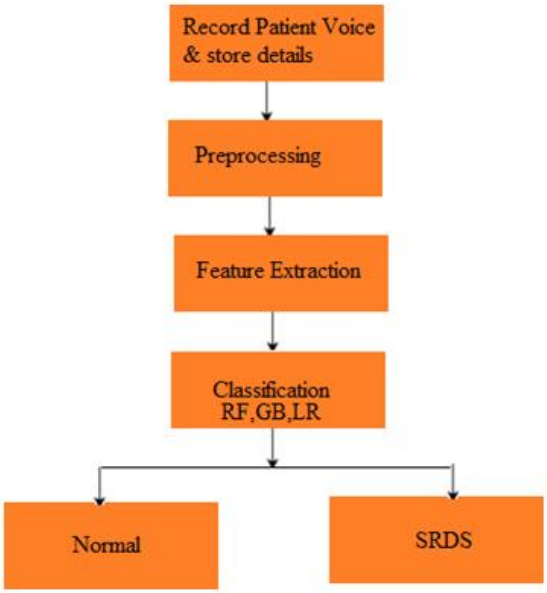

Figure 1: System Architecture

IJESRT is licensed under a Creative Commons Attribution 4.0 International License.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

In this system recording the patients breathing voice and applying pre-processing technique for noise removal purpose and extracts the feature using librosa-MFCC. Then applies the classification RF, GB and LR. Based on this predicts the patient is Normal or SRDS effected.

4. RESULTS AND DISCUSSIONS

Current research utilize linear SVM to model ARDS. We observed that nonlinear kernel (RBF) for SVM produced less stable results in exploratory data analysis. Due to lower performance of SVM, we are using RF, GR and LR machine learning models. In certain clinical settings, a chest x-ray may not be readily accessible. , is necessary for the diagnosis but takes time so immediate result and more accuracy purpose using patients breathing voice. And based on breathing voice the system predicts SRDS or not.

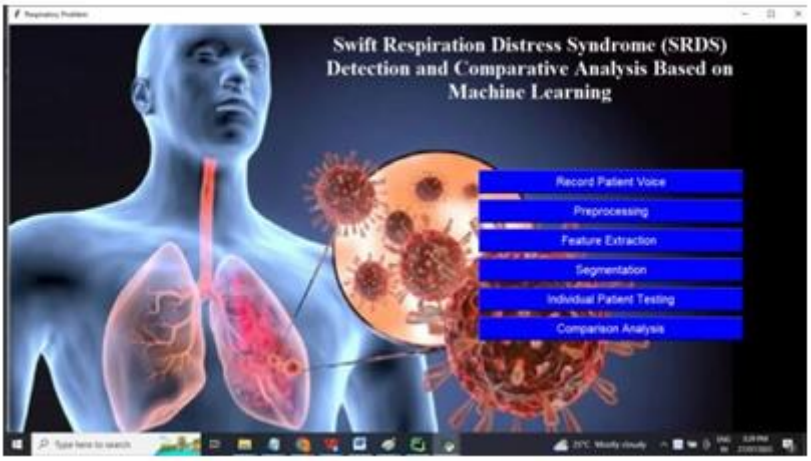

Figure 2: Menu

The menu consists of different options to do the analysis such as Patient details with patient voice, preprocessin, feature extraction, classification

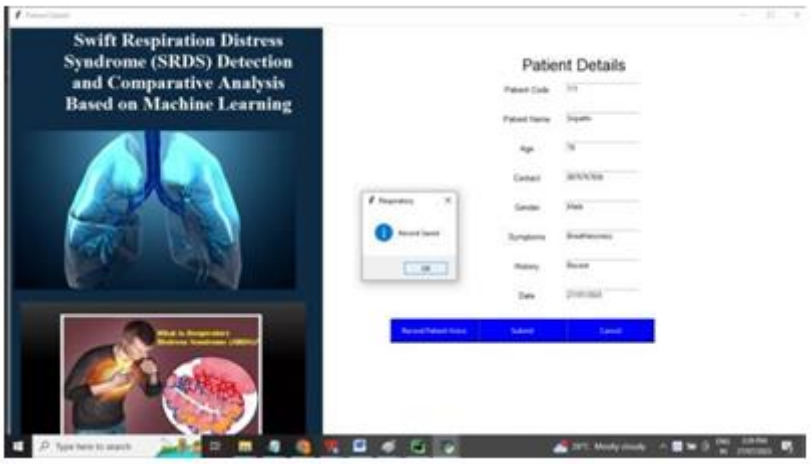

Figure 3: Patient details

Storing the patient details and recording the patient voice

IJESRT is licensed under a Creative Commons Attribution 4.0 International License.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

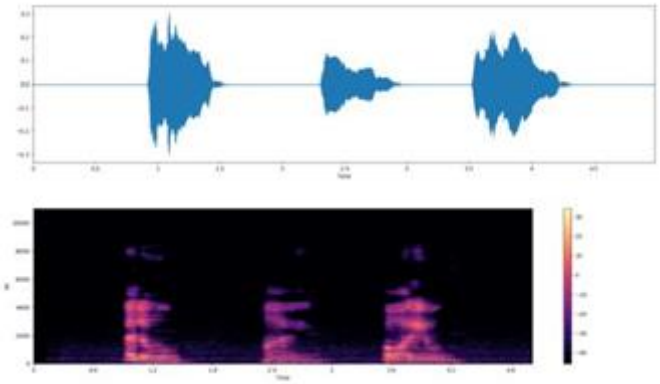

Figure 4: Patient voice

This module records the patient voice and simultaneously displays the voice signal graph as shown

5. CONCLUSION

The results of our analysis indicate as machine learning system is superior at predicting SRDS in cardiac surgery patients. When used effectively, resulting random forest has potential to direct therapeutic decision-making & help in improvement of patients' long-term prognoses.

REFERENCES

1. G. D. Rubenfeld et al., "Incidence and outcomes of acute lung injury," New Engl. J. Med., vol. 353, no. 16, pp. 1685– 1693, Oct. 2005.
2. R. M. Sweeney and D. F. McAuley, "Acute respiratory distress syndrome," Lancet, vol. 388, no. 10058, pp. 2416– 2430, Nov. 2016.
3. G. Bellani et al., "Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries," J. Amer. Med. Assoc., vol. 315, no. 8, pp. 788–800, Feb. 2016.
4. B. J. Clark and M. Moss, "The acute respiratory distress syndrome: Dialing in the evidence?" J. Amer. Med. Assoc., vol. 315, no. 8, pp. 759–761, Feb. 2016.
5. M. W. Sjoding and R. C. Hyzy, "Recognition and appropriate treatment of the acute respiratory distress syndrome remains unacceptably low," Critical Care Med., vol. 44, no. 8, pp. 1611–1612, Aug. 2016.
6. M. W. Sjoding, "Translating evidence into practice in acute respiratory distress syndrome: Teamwork, clinical decision support, and behavioral economic interventions," Current Opin. Critical Care, vol. 23, no. 5, pp. 406–411, Oct. 2017.
7. V. Herasevich et al., "Validation of an electronic surveillance system for acute lung injury," Intensive Care Med., vol. 35, no. 6, pp. 1018–1023, Jun. 2009.
8. H. C. Koenig et al., "Performance of an automated electronic acute lung injury screening system in intensive care unit patients," Critical Care Med., vol. 39, no. 1, pp. 98–104, Jan. 2001.
9. G. D. Rubenfeld et al., "Interobserver variability in applying a radiographic definition for ARDS," Chest, vol. 116, no. 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. Amer. Thoracic Soc., vol. 13, no. 7, pp. 1123–8, Jul. 2016.
11. D. F. Nettleton, A. Orriols-Puig, and A. Fornells, "A study of the effect of different types of noise on the precision of supervised learning techniques," Artif. Intell. Rev., vol. 33, no. 4, pp 275–306, Jan. 2010. [13] B. Frenay and M. Verleysen, "Classification in the presence of label noise: A survey," IEEE Trans. Neural Netw. Learn. Syst., vol. 25, no. 5, pp. 845– 869, May 2014.
12. N. Natarajan et al., "Learning with noisy labels," in Proc. Neural Inform. Process. Syst., Dec. 2013, pp. 1196–1204.
13. [15] Y. Duan and O. Wu, "Learning with auxiliary less-noisy labels," IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 7, pp. 1716–1721, May 2017

IJESRT is licensed under a Creative Commons Attribution 4.0 International License.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

14. Elyas Sabeti; Joshua Drews; Narathip Reamaroon; Jonathan Gryak; Michael Sjoding; Kayvan Najari 2019, "Detection of Acute Respiratory Distress Syndrome by Incorporation of Label Uncertainty and Partially Available Privileged Information."
15. Ashbaugh DG, Bigelow DB, Petty TL, Levine BE. Acute respiratory distress in adults. Lancet. 1967;2:319–23.
16. Tomashefski JF Jr. Pulmonary pathology of acute respiratory distress syndrome. Clin Chest Med. 2000;21:435– 66.
17. Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, Fan E, Camporota L, Slutsky AS. Acute respiratory distress syndrome: the Berlin definition. JAMA. 2012;307:2526–33.
18. Rubenfeld GD, Caldwell E, Peabody E, Weaver J, Martin DP, Neff M, Stern EJ, Hudson LD. Incidence and outcomes of acute lung injury. N Engl J Med. 2005;353:1685–93.
19. Villar J, Blanco J, Kacmarek RM. Current incidence and outcome of the acute respiratory distress syndrome. Curr Opin Crit Care. 2016;22:1–6.
20. Diaz JV, Brower R, Calfee CS, Matthay MA. Therapeutic strategies for severe acute lung injury. Crit Care Med. 2010;38:1644–50.
21. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69:89–95.

IJESRT is licensed under a Creative Commons Attribution 4.0 International License.