



#### **Review Article**

# The Use of Machine Learning **Decision Tree Algorithms in Phenotyping Acute Respiratory** Distress Syndrome (ARDS) Based on Clinical, Radiological, and **Biological Heterogeneity-A Review**

## Moumita Chakraborty\*

Assistant Professor, Department of Respiratory Care Technology, Narayana College of Health Sciences, Bengaluru, India

#### **More Information**

#### \*Address for correspondence:

Moumita Chakraborty, Assistant Professor, Department of Respiratory Care Technology, Narayana College of Health Sciences, Bengaluru, India, Email: chakrabortymoumita27@gmail.com

Submitted: June 16, 2025 Approved: July 07, 2025 Published: July 08, 2025

How to cite this article: Chakraborty M. The Use of Machine Learning Decision Tree Algorithms in Phenotyping Acute Respiratory Distress Syndrome (ARDS) Based on Clinical, Radiological, and Biological Heterogeneity-A Review. J Pulmonol Respir Res. 2025; 9(2): 026-030. Available from:

https://dx.doi.org/10.29328/journal.jprr.1001070

Copyright license: © 2025 Chakraborty M. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.





#### Abstract

Background: Acute Respiratory Distress Syndrome (ARDS) is a clinically, radiologically, and biologically heterogeneous condition. This variability contributes to diagnostic challenges and inconsistent responses to therapy. Identifying homogeneous subgroups or phenotypes within ARDS may enhance precision medicine and therapeutic targeting.

Objective: This review evaluates the utility of decision tree-based supervised machine learning (ML) algorithms—specifically CART, Random Forest, and AdaBoost—in phenotyping ARDS using clinical, radiological, and biological data.

Methods: A comprehensive literature search was conducted between December 2023 and March 2024 using PubMed and Google Scholar. Search terms included 'decision tree in ARDS', 'phenotype in ARDS', and 'ML in hypo- and hyperinflammatory ARDS'. Twenty-six relevant articles were included, comprising original studies and reviews.

Results: Decision tree-based models have demonstrated significant potential in classifying ARDS subtypes using routine clinical variables, radiographic features, and biomarker profiles. These algorithms have shown strong predictive performance in differentiating inflammatory phenotypes, forecasting mortality, and enabling early ARDS prediction.

Conclusion: Decision tree algorithms offer a promising approach to ARDS phenotyping by leveraging routinely available data. Their interpretability and predictive accuracy may aid in translating complex biological insights into bedside clinical decision-making, advancing personalized care in critical illness.

## Introduction

Acute Respiratory Distress Syndrome (ARDS) is a clinical syndrome that comprises diffuse lung inflammation and edema, which leads to rapidly progressive acute respiratory failure of type I (hypoxemic) [1,2]. It is characterized by pulmonary edema of non-cardiogenic origin, diffuse alveolar damage, and inflammatory cell infiltration [3]. The ARDS definition has continuously evolved since the first description as a 'syndrome' by Ashbaugh and colleagues in 1967 [4].

In 2012, Berlin's definition simplified the terminology of

ARDS; stratified its severity based on PaO2/FiO2 of ≤100, 101-200, and 201-300 mmHg as mild, moderate, and severe, respectively, requiring positive end-expiratory pressure (PEEP) of >= 5 cm  $H_2O$ . The onset has been defined as 1 week from the clinical insult and presence of bilateral chest infiltrates on chest X-ray, not caused by heart failure [5].

However, this approach of defining ARDS had some limitations as there was no linear relationship found between PaO<sub>2</sub> and FiO<sub>2</sub> [6]. Also, it does not account for the relationship of PaO<sub>2</sub>/FiO<sub>2</sub> and the predictable structural changes in the alveolar-capillary membrane [7]. The definition also had a



major pitfall as it was validated using only cohort trials and fewer prospective randomised controlled trials [8]. Also, certain studies [9,10], such as the LUNG SAFE study [8] and various other studies have highlighted the fact that ARDS diagnosis may go unrecognized [11].

The occurrence of ARDS is around 10% of all Intensive Care Unit (ICU) patients, and the estimated mortality is about 40% [12]. The etiology of ARDS has been categorised as direct and indirect causes, with pneumonia being the major etiology. Different etiologies of ARDS can result in varied pathological and biological changes, leading to complex clinical and biological heterogeneity. Studies have shown that heterogeneity is a central factor contributing to failed randomized controlled trials in ARDS [13-15]. Hence, certain factors highlight the need for stratifying ARDS into homogenous subsets or phenotypes, diverse etiologies, clinical heterogeneity, complexity in radiographic lung morphology, biomarker profiles, and varied response to pharmacotherapy [9,10,16-19].

Machine learning(ML) has been proven to be a powerful and effective tool in various fields, including healthcare, to diagnose diseases, predict the severity of infections, estimate the likelihood of hospital readmissions, etc. [20,21] The potential for ML has been increasingly recognised in the field of critical care medicine, such as in the prediction of ARDS development [22,23].

One such type of machine learning algorithm is the decision tree. Various studies have shown the effectiveness of decision trees in the classification of data sets into homogeneous groups for better decision making [24,25]. In this review article, we aim to elucidate the use of decision trees in phenotyping ARDS based on clinical, radiological, and biological features.

### **Clinical phenotyping in ARDS**

Clinical phenotypes can group ARDS patients by shared etiology, time-course, or radiographic presentation.

Two subphenotypes of ARDS were identified in the HARP-2 randomised cohort study: hyper- and hypoinflammatory types, with distinct clinical and biological features and disparate clinical outcomes involving 539 patients. The hyperinflammatory subphenotype had improved survival with simvastatin compared with placebo [26]. Calfee, et al. [27] classified ARDS patients into classic hyperinflammatory and hypoinflammatory types, for 1022 patients: 473 in the ARMA cohort and 549 in the ALVEOLI cohort. Independent latent class models indicated that a two-class (ie, two subphenotypes) model was the best fit for both cohorts. But the definition of this biological phenotype requires the use of plasma biomarkers as class-defining variables, such as sTNFR-1 and interleukin (ILs), which are not routinely available and cannot be quickly quantified at the bedside.

Therefore, the clinical applicability of this classification system may be limited. This represents a barrier to the clinical implementation of phenotypes [27].

In order to address this limitation, a study by Pratik Sinha, et al. used standard clinical and laboratory parameters to categorise patients from 3 prior clinical trials into inflammatory subtypes. The performance of the optimised gradient boosted model (GBM) classifier model - a type of decision tree model was evaluated on a fourth separate dataset, by comparing the classes it assigned to the 'gold-standard' LCA classes. For the hypoinflammatory class, the GBM model (with a probability threshold of 0.5) gave the correct answer (assuming LCA is correct) in 98% of cases (460/468), but for the Hyperinflammatory class, it was only correct in 63% of cases (175/277). The combined accuracy for both classes was 85% [28].

In a study by Sidney Le, et al. data from 9919 patients were retrospectively analyzed from the Medical Information Mart of Intensive Care III (MIMIC III) database, using the XG Boost gradient boost decision tree model using routinely used clinical data. The classifier algorithm attained an AUROC value of 0.905, indicating the possibility to predict ARDS up to 48 hours prior to its onset [29].

Yang P, et al. in their study to identify ARDS based on noninvasive physiological parameters, concluded that XGBoost (Decision tree) had the best performance of ARDS identification with the sensitivity of 84.03%, the specificity of 87.75% and the AUC of 0.9128 [30].

Yui Bai, et al. conducted a retrospective study using two large databases, eICU and MIMIC IV, to explore the clinical phenotypes of sepsis-related ARDS patients and treatment response. Firstly, the early diagnosis model data was trained and tested in the eICU database. Simultaneously, the patients were clustered using predictable variables (clinical data). Also, the clinical outcome among clusters was identified. The results were further validated in the MIMIC-IV databases to assess the reproducibility of results. Among the 5 machine learning algorithms that had been trained, Adaboost –a type of decision tree was the best performing model with an AUROC of 0.895 [31].

The efficacy of the decision tree was also identified in a retrospective study on post-operative ARDS by Jianmin Ling. In the study, a total of 1065 patients were included. Clinical variables, along with laboratory variables, were used for latent profile analysis (LPA). The LPA identified three subtypes of postoperative ARDS based on clinical features and respiratory compliance. Patients in profile 1 were mainly accepted for neurosurgery, profile 2 and 3 were treated with orthopedic and vascular or thoracic surgery, respectively. The XG-Boost model (decision tree) effectively predicted mortality with an AUC of 0.935, which was comparatively higher than scores such as SOFA (0.622), APACHE 2 (0.629), SLIP (0.579), and SLIP-2 (0.550) [32].



## Radiological phenotyping in ARDS

Radiological phenotyping relies on identifying distinct imaging patterns, primarily via chest radiographs and computed tomography (CT). Imaging serves not only to support diagnosis but also to differentiate focal versus nonfocal patterns of lung injury, which has prognostic and therapeutic relevance.

CT imaging can distinguish between focal (lobar) and non-focal (diffuse) patterns. This distinction was notably utilized in the LIVE trial, where an ML-based decision tree model identified subphenotypes using basic radiological and ventilator parameters [23]. These included tidal volume,  $PaO_2/FiO_2$  ratio, and peak airway pressure. The model differentiated between focal and non-focal ARDS with high sensitivity and specificity.

The recognition of these patterns can inform ventilator settings. For example, patients with non-focal ARDS may respond better to recruitment maneuvers and higher PEEP, while those with focal ARDS may experience harm from overdistention [21,29]. However, the LIVE trial demonstrated increased mortality in the misclassified group, underscoring the importance of accurate phenotyping.

Radiographically, ARDS has largely been described as two phenotypes, nonfocal/diffuse ARDS and focal/lobar ARDS, based on morphologic characteristics on computed tomography (CT) [33].

Supporting a biologic basis to radiographic phenotypes, another study reported a strong association of plasma concentrations of the epithelial biomarker RAGE and nonfocal CT-based lung-imaging patterns in patients with ARDS [34].

#### **Biological phenotyping in ARDS**

Biological phenotyping uses biomarkers such as cytokines, surfactant proteins, and endothelial injury markers to classify ARDS subgroups.

Latent class analysis has historically been used to define hyper- and hypo-inflammatory phenotypes, with consistent replication across trials (e.g., ARMA, FACTT, ALVEOLI) [18]. These phenotypes show distinct cytokine profiles, disease severity, and treatment response.

Decision tree models can replicate LCA-based classifications with fewer inputs. A 2019 study demonstrated that a decision tree using just three variables—IL-8, protein C, and bicarbonate—achieved >95% accuracy in identifying the hyperinflammatory phenotype [18,22]. This simplification enhances bedside utility and reduces the need for large biomarker panels.

Our understanding of the heterogeneity of critical illness syndromes has improved with the use of mathematical and statistical methods, such as cluster analysis and latent class analysis (LCA). However, all these ARDS studies used plasma biomarkers as class-defining variables, such as sTNFR-1 and interleukins (ILs), which are not routinely available and cannot be quantified rapidly at the bedside. Thus, the clinical applicability of this classification system may be limited [35-45].

## Conclusion

ARDS is a highly heterogeneous syndrome that has defied one-size-fits-all therapeutic approaches. Machine learning, particularly decision tree-based algorithms, offers an interpretable and practical means of phenotyping ARDS into clinically meaningful subgroups. These models can integrate multidimensional data—including clinical variables, imaging, and biomarkers—into simple yet accurate classification tools. As ML integration into the ICU progresses, decision tree-based phenotyping could form the backbone of precision medicine strategies in ARDS, improving diagnostic accuracy, risk stratification, and individualized treatment.

### References

- Bos LDJ, Ware LB. Acute respiratory distress syndrome: causes, pathophysiology, and phenotypes. Lancet. 2022;400(10358):1145-1156.
   Available from: https://doi.org/10.1016/s0140-6736(22)01485-4
- Fernando SM, Ferreyro BL, Urner M, Munshi L, Fan E. Diagnosis and management of acute respiratory distress syndrome. CMAJ. 2021;193(21):E761-E768. Available from: https://doi.org/10.1503/cmaj.202661
- Bai Y, Xia J, Huang X, Chen S, Zhan Q. Using machine learning for the early prediction of sepsis-associated ARDS in the ICU and identification of clinical phenotypes with differential responses to treatment. Front Physiol. 2022;13:1050849. Available from: https://doi.org/10.3389/fphys.2022.1050849
- Matthay MA, Thompson BT, Ware LB. The Berlin definition of acute respiratory distress syndrome: Should patients receiving high-flow nasal oxygen be included? Lancet Respir Med. 2021;9(8):933-936.
   Available from: https://doi.org/10.1016/s2213-2600(21)00105-3
- Yuan X, Pan C, Xie J, Qiu H, Liu L. An expanded definition of acute respiratory distress syndrome: Challenging the status quo. J Intensive Med. 2023;3(1):62-64. Available from: https://doi.org/10.1016/j.jointm.2022.06.002
- Sayed M, Riaño D, Villar J. Novel criteria to classify ARDS severity using a machine learning approach. Crit Care. 2021;25:150. Available from: https://ccforum.biomedcentral.com/articles/10.1186/s13054-021-03566-w
- Villar J, Pérez-Méndez L, Kacmarek RM. The Berlin definition met our needs: no. Intensive Care Med. 2016;42:648–650. Available from: https://link.springer.com/article/10.1007/s00134-016-4242-6
- Bellani G, Laffey JG, Pham T, Fan E, Brochard L, Esteban A, et al. Epidemiology, Patterns of Care, and Mortality for Patients With Acute Respiratory Distress Syndrome in Intensive Care Units in 50 Countries. JAMA. 2016;315(8):788–800. Available from: https://doi.org/10.1001/jama.2016.0291
- Bellani G, Pham T, Laffey JG. Missed or delayed diagnosis of ARDS: a common and serious problem. Intensive Care Med. 2020;46(6):1180-1183.
   Available from: https://doi.org/10.1007/s00134-020-06035-0
- Le S, Pellegrini E, Green-Saxena A, Summers C, Hoffman J, Calvert J, et al. Supervised machine learning for the early prediction of acute respiratory distress syndrome (ARDS). J Crit Care. 2020;60:96-102. Available from: https://doi.org/10.1016/j.jcrc.2020.07.019



- ARDS Definition Task Force; 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(23): 2526-33. Available from: https://doi.org/10.1001/jama.2012.5669
- McNicholas BA, Rooney GM, Laffey JG. Lessons to learn from epidemiologic studies in ARDS. Curr Opin Crit Care. 2018;24:41–48. Available from: https://doi.org/10.1097/mcc.00000000000000473
- Matthay MA, McAuley DF, Ware LB. Clinical trials in acute respiratory distress syndrome: challenges and opportunities. Lancet Respir Med. 2017;5(6):524–34. Available from: https://doi.org/10.1016/s2213-2600(17)30188-1
- Reilly JP, Calfee CS, Christie JD. Acute Respiratory Distress Syndrome Phenotypes. Semin Respir Crit Care Med. 2019;40(1):19–30. Available from: https://doi.org/10.1055/s-0039-1684049
- Ruan SY, Huang CT, Chien YC, Kuo LC, et al. Etiology-associated heterogeneity in acute respiratory distress syndrome: a retrospective cohort study. BMC Pulm Med. 2021;21:183. Available from: https:// bmcpulmmed.biomedcentral.com/articles/10.1186/s12890-021-01557-9
- Wilson JG, Calfee CS. ARDS Subphenotypes: Understanding a Heterogeneous Syndrome. Crit Care. 2020;24:102. Available from: https://doi.org/10.1186/s13054-020-2778-x
- Thille AW, Esteban A, Fernández-Segoviano P, Rodriguez JM, Aramburu JA, Peñuelas O, et al. Comparison of the Berlin definition for acute respiratory distress syndrome with autopsy. Am J Respir Crit Care Med. 2013;187(7):761–7. Available from: https://doi.org/10.1164/rccm.201211-1981oc
- Alipanah N, Calfee CS. Phenotyping in acute respiratory distress syndrome: state of the art and clinical implications. Curr Opin Crit Care. 2022;28(1):1-8. Available from: https://doi.org/10.1097/mcc.0000000000000003
- Schwager E, Jansson K, Rahman A, Schiffer S, Chang Y, Boverman G, et al. Utilizing machine learning to improve clinical trial design for acute respiratory distress syndrome. npj Digit Med. 2021;4:133. Available from: https://doi.org/10.1038/s41746-021-00505-5
- Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. 2020:25–60. Available from: https://doi.org/10.1016/B978-0-12-818438-7.00002-2
- Sarke IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science. 2021. Available from: https://doi.org/10.1007/s42979-021-00592-x
- Ding XF, Li JB, Liang HY, Wang ZY, Jiao TT, Liu Z, et al. Predictive model for acute respiratory distress syndrome events in ICU patients in China using machine learning algorithms: a secondary analysis of a cohort study. J Transl Med. 2019;17:326. Available from: https://doi.org/10.1186/s12967-019-2075-0
- Sinha P, Churpek MM, Calfee CS. Machine learning classifier models can identify acute respiratory distress syndrome phenotypes using readily available clinical data. Am J Respir Crit Care Med. 2020;202:996–1004. Available from: https://doi.org/10.1164/rccm.202002-0347oc
- Oo AN, Naing T. Decision Tree Models for Medical Diagnosis.
  International Journal of Trend in Scientific Research and Development.
  2019 Apr;3(3):1697-1699. Available from: https://www.slideshare.net/slideshow/decision-tree-models-for-medical-diagnosis/151269112
- 25. Azar AT, El-Metwally SM. Decision tree classifiers for automated medical diagnosis. Neural Comput & Applic. 2013;23:2387–2403. Available from: https://link.springer.com/article/10.1007/s00521-012-1196-7
- Calfee CS, Delucchi KL, Sinha P, et al. Acute respiratory distress syndrome subphenotypes and differential response to simvastatin: secondary analysis of a randomised controlled trial. Lancet Respir Med. 2018;6(9):691-698. Available from: https://www.thelancet.com/journals/lanres/article/PIIS2213-2600(18)30177-2/abstract
- 27. Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA, et al. Subphenotypes in acute respiratory distress syndrome: Latent class

- analysis of data from two randomised controlled trials. Lancet Respir Med. 2014;2(8):611–620. Available from: https://doi.org/10.1016/s2213-2600(14)70097-9
- Sinha P, Churpek MM, Calfee CS. Machine Learning Classifier Models Can Identify Acute Respiratory Distress Syndrome Phenotypes Using Readily Available Clinical Data. Am J Respir Crit Care Med. 2020;202(7):996-1004. Available from: https://doi.org/10.1164/rccm.202002-0347oc
- Le S, Pellegrini E, Green-Saxena A, Summers C, Hoffman J, Calvert J, Das R. Supervised machine learning for the early prediction of acute respiratory distress syndrome (ARDS). J Crit Care. 2020;60:96-102. Available from: https://doi.org/10.1016/j.jcrc.2020.07.019
- Yang P, Wu T, Yu M, Chen F, Wang C, Yuan J, et al. A new method for identifying the acute respiratory distress syndrome disease based on noninvasive physiological parameters. PLoS One. 2020;15(2):e0226962. Available from: https://doi.org/10.1371/journal.pone.0226962
- Bai Y, Xia J, Huang X, Chen S, Zhan Q. Using machine learning for the early prediction of sepsis-associated ARDS in the ICU and identification of clinical phenotypes with differential responses to treatment. Front Physiol. 2022;13:1050849. Available from: https://doi.org/10.3389/fphys.2022.1050849
- 32. Ling J, Liu H, Yu D, Wang Z, Fang M. Three subtypes of postoperative ARDS that showing different outcomes and responses to mechanical ventilation and fluid management: A machine learning and latent profile analysis. Heart Lung. 2023;62:135-144. Available from: https://doi.org/10.1016/j.hrtlng.2023.07.007
- Puybasset L, Cluzel P, Gusman P, Grenier P, Preteux F, Rouby JJ; CT Scan ARDS Study Group. Regional distribution of gas and tissue in acute respiratory distress syndrome. I. Consequences for lung morphology. Intensive Care Med. 2000;26(07):857–869. Available from: https://doi.org/10.1007/s001340051274
- 34. Mrozek S, Jabaudon M, Jaber S, Paugam-Burtz C, Lefrant JY, Rouby JJ, et al; Azurea network. Elevated plasma levels of sRAGE are associated with nonfocal CT-based lung imaging in patients with ARDS: a prospective multicenter study. Chest. 2016;150(05):998-1007. Available from: https://doi.org/10.1016/j.chest.2016.03.016
- 35. Liu X, Jiang Y, Jia X, Ma X, Han C, Guo N, et al. Identification of distinct clinical phenotypes of acute respiratory distress syndrome with differential responses to treatment. Crit Care. 2021;25:320. Available from: https://ccforum.biomedcentral.com/articles/10.1186/s13054-021-03734-y
- Ruan SY, Huang CT, Chien YC, Huang CK, Chien JY, Kuo LC, et al. Etiologyassociated heterogeneity in acute respiratory distress syndrome: a retrospective cohort study. BMC Pulm Med. 2021;21:183. Available from: https://doi.org/10.1186/s12890-021-01557-9
- Sinha P, Delucchi KL, McAuley DF, O'Kane CM, Matthay MA, Calfee CS. Development and validation of parsimonious algorithms to classify acute respiratory distress syndrome phenotypes: a secondary analysis of randomised controlled trials. Lancet Respir Med. 2020;8(3):247-257. Available from: https://doi.org/10.1016/s2213-2600(19)30369-8
- Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects.
   BMJ. 2018;363:k4245. Available from: https://doi.org/10.1136/bmj.k4245
- 39. Singhal L, Garg Y, Yang P, Tabaie A, Wong Al, Mohammed A, et al. eARDS: A multi-center validation of an interpretable machine learning algorithm of early onset Acute Respiratory Distress Syndrome (ARDS) among critically ill adults with COVID-19. PLoS One. 2021;16(9):e0257056. Available from: https://doi.org/10.1371/journal.pone.0257056
- Le S, Pellegrini E, Green-Saxena A, Summers C, Hoffman J, Calvert J, Das R. Supervised machine learning for the early prediction of acute respiratory distress syndrome (ARDS). J Crit Care. 2020;60:96-102. Available from: https://doi.org/10.1016/j.jcrc.2020.07.019
- 41. Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision Trees: An Overview

The Use of Machine Learning Decision Tree Algorithms in Phenotyping Acute Respiratory Distress Syndrome (ARDS) Based on Clinical, Radiological, and Biological Heterogeneity-A Review



- and Their Use in Medicine. J Med Syst. 2002;26:445–463. Available from: https://doi.org/10.1023/a:1016409317640
- 42. Chang W, Liu Y, Xiao Y, Yuan X, Xu X, Zhang S, Zhou S. A Machine-Learning-Based Prediction Method for Hypertension Outcomes Based on Medical Data. Diagnostics (Basel). 2019;9(4):178. Available from: https://doi.org/10.3390/diagnostics9040178
- 43. Quinlan JR. Induction of decision trees. Mach Learn. 1986;1:81–106. Available from: https://link.springer.com/article/10.1007/BF00116251
- 44. Bos LDJ, Artigas A, Constantin JM, Hagens LA, Heijnen N, Laffey JG, et al. Precision medicine in acute respiratory distress syndrome: workshop report and recommendations for future research. Eur Respir Rev. 2021;30(159):200317. Available from: https://doi.org/10.1183/16000617.0317-2020
- 45. Pennati F, Aliverti A, Pozzi T, Gattarello S, Lombardo F, Coppola S, et al. Machine learning predicts lung recruitment in acute respiratory distress syndrome using single lung CT scan. Ann Intensive Care. 2023;13:60. Available from: https://doi.org/10.1186/s13613-023-01154-5