Proceedings on Automation in Medical Engineering
Proc AUTOMED, 2026, Vol. 3, No 1, Article ID 2471, 2 Pages

Infinite Science Publishing

Comparative Analysis of Feature Selection Methods in the use-case of ARDS Classification in Clinical Time-Series Data

S. Fonck[1,3], S. Wedding[1], S. Fritsch[2,3], and A. Stollenwerk[1,3,*]

[1] *Embedded Software - RWTH Aachen University, 52074 Aachen,Germany*
[2] *Jülich Supercomputing Centre, Forschungszentrum Jülich, 52428 Jülich, Germany*
[3] *Center for Advanced Simulation and Analytics - Digital Health, FZ Jülich, 52428 Jülich, Germany*
[*] *Corresponding author, email:stollenwerk@embedded.rwth-aachen.de*

Abstract: This study investigates four feature selection methods (χ2, ANOVA F-test, Lasso, and a Tree-based method) to enhance Acute Respiratory Distress Syndrome (ARDS) classification in time-series intensive care unit data using an existing Random Forest algorithm. While feature selection did not significantly improve ARDS classification performance, using a reduced number of features achieved comparable results to using the entire dataset. This indicates the potential of dimensionality reduction for ARDS classification.

I. Introduction

Acute Respiratory Distress Syndrome (ARDS) is a severe inflammatory lung condition with a high mortality rate of approximately 40 % [1]. With a prevalence of 10 % in mechanically ventilated patients, ARDS continues to be a major challenge in intensive care units (ICU). Since 2012, ARDS is diagnosed using the so-called Berlin definition (BD) [2]. According to the BD, a Horovitz quotient lower than 300 mmHg measured under a positive end-expiratory pressure (PEEP) higher than 5 cmH2O must be present. The horovitz quotient can be calculated by dividing partial pressure of arterial oxygen (PaO2) by fraction of inspired oxygen (FiO2) and represents transpulmonal oxygen transfer capacity of the lungs. One reason for the high mortality rate regarding ARDS could be that many cases are missed by physicians or diagnosed too late for adequate treatment [3]. To prevent the underrecognition of ARDS and reduce mortality rate, Artificial Intelligence (AI) techniques are proposed to classify ARDS and assist treating physicians effectively [4]. Various AI-based classifiers have been published for ARDS classification in time-series data with promising results [5, 6]. When training AI models, the number of features can lead to high computational costs and reduce the performance of trained AI methods, as data can add noise and make calculations more difficult ("*curse of dimensionality*"). Furthermore, the higher the dimensionality of data, the harder it gets to transfer classification results to real world scenarios. To recognize the unimportant and unrelated features and thus ensure efficient training of AI methods, Feature Selection (FS) methods are used [7]. This is particularly relevant in the medical context, as the recording and use of clinical data is associated with many regulations and limitations. In this study, we examine the influence of different FS methods with different numbers of features on the performance of AI models for the classification of ARDS in time-series data. Therefore, we adapted the pipeline for the implementation of a Random Forest (RF) algorithm by Fonck et al. [6].

II. Methods

The study utilized a dataset of 13,067 ICU admissions from the University Hospital Aachen. In consultation with the doctors, 40 parameters related to ARDS were initially selected (comprising vital signs, mechanical ventilator parameters and laboratory results). In further process steps, the data is imputed using forward filling to close gaps. Finally, the ARDS onset is determined based on the first time point at which the Horovitz quotient falls below 300 mmHg, using data points 50,000 seconds before and after the onset. Four FS methods were investigated:

- **χ2**: Evaluates the relationship between each feature and the target classification ARDS. A high χ2 statistic indicates the feature to be informative for the prediction. Users can specify the number of features to be retained for the classification task. [8]
- **ANOVA F-Test**: Compares the variance of a feature between different groups using F-statistic to assess relevance. Users can specify the desired number of features to keep for classification. [9]
- **Lasso**: Adds a penalty term, proportional to the absolute magnitude of the coefficients, to the loss function during training. This promotes reduction in the model by zeroing some coefficients, effectively selecting a subset of features. The number of considered features is determined by the method, by driving the weights of unimportant features to zero. [10]

• **Tree-based**: Uses decision trees to calculate feature importance scores by computing each node‘s gini impurity reduction for every feature. The model is first trained on the complete data set, and afterwards on the features selected by the algorithm. The number of features kept can be indirectly controlled by changing the minimum feature importance threshold. [11]

To evaluate the influence of the FS methods, an RF algorithm was trained with the respective feature sets (all 40, 20, 10 and 5), that were determined by an FS method, and evaluated on a test dataset. Performance was evaluated using accuracy, sensitivity, specificity, and F1-score.

III. Results

The study found that none of the FS methods significantly improved the classification performance of the RF algorithm compared to using all 40 features (see Tab. 1).

Table 1: Comparison of performance of the RF algorithm for different FS methods compared to no FS. Resulting number of used features: F-Test 20, $\chi 2$ 20, Lasso 9, Tree-based 20.

	No FS	F-Test	$\chi 2$	Lasso	Tree-based
Acc	0.900	**0.912**	0.911	0.883	0.816
Sens.	0.777	0.757	0.757	0.767	**0.821**
Spec.	0.916	**0.925**	0.924	0.893	0.816
F1-Score	**0.730**	0.717	0.717	0.629	0.621

Furthermore, no FS method significantly outperformed the others. However, Lasso and Tree performed slightly worse, with a 10 % reduction in F1-score. However, performance remained comparable while reducing the number of features (see Tab. 2 for results using $\chi 2$).

Table 2: Comparison of performance of the RF algorithm for different numbers of features for $\chi 2$.

	40	20	10	5
Acc	0.900	**0.911**	0.906	0.875
Sens.	**0.777**	0.757	0.757	**0.777**
Spec.	0.916	**0.924**	0.918	0.884
F1-Score	**0.730**	0.717	0.707	0.617

IV. Discussion & Conclusions

ARDS is a severe lung condition that is often underdiagnosed. To support treating physicians in the diagnostic process, AI methods for classification are increasingly being published. These methods are usually trained with many features. Our results suggest that FS can be a way to tackle this curse of dimensionality of ICU time-series data for ARDS classification using an RF with no to minor loss in performance. This can help to save resources for data collection. Although no single FS method significantly improved performance, the study demonstrates the potential of using a subset of parameters for an initial assessment of ARDS. As different methods yield different results, it is important to choose an appropriate FS technique for a classification task. In our case, $\chi 2$ and F-Test outperformed Lasso and the Tree based method.

The data provided were not specifically labelled for ARDS leading to the assumption that the dataset includes many missed cases of ARDS as well as some cases where a diagnosis was made without the condition being present. Furthermore, the single used dataset is non-publicly available, which restricts the generalizability of the results. Further research should explore other datasets and AI models to validate these findings and the clinical implications of the selected features as well as their transferability. Moreover, the influence of wrongly labeled or overly broad classified (e.g. outdated ICD-9 codes, not reflecting ARDS) data remains unclear [6]. These labels are decisive for the evaluation of AI methods. Without qualitative ground truth, statements about the corresponding performance of AI methods are not reliable. For small datasets FS itself is likely to introduce a higher computational cost in contrast to the relief of the subsequent algorithms. Lastly, as we only looked at statistical measures to assess the performance, the clinical implications and relevance remain unclear. Furthermore, more or different initial parameters can also be considered to analyze their influence on the classification of ARDS. Finally deep learning-based FS methods should be researched.

AUTHOR'S STATEMENT

Research funding: The authors state no funding involved. Conflict of interest: Authors state no conflict of interest. Ethical approval: Ethical approval was obtained from the local Ethics Committee (EK 102/19).

REFERENCES

[1] G. D. Ashbaugh, D. B. Bigelow, L. T. Petty, and E. B. Levine, “Acute Respiratory Distress in Adults,” The Lancet, vol. 290, no. 7511, pp. 319–323, 1967.

[2] V. M. Ranieri, G. D. Rubenfeld, B. T. Thompson, N. D. Ferguson, E. Caldwell, E. Fan, et al., “Acute Respiratory Distress Syndrome: the Berlin Definition,” JAMA, vol. 307, no. 23, pp. 2526–2533, 2012.

[3] G. Bellani, T. Pham, and J. G. Laffey, “Missed or delayed diagnosis of ARDS: a common and serious problem,” Intensive Care Med, vol. 46, pp. 1180–1183, 2020.

[4] M. Rashid, M. Ramakrishnan, V. P. Chandran, S. Nandish, S. Nair, V. Shanbhag, and G. Thunga, “Artificial intelligence in acute respiratory distress syndrome: a systematic review,” Artificial intelligence in medicine, vol. 131, p. 102361, 2022.

[5] P. Yang, T. Wu, M. Yu, F. Chen, C. Wang, J. Yuan, J. Xu, and G. Zhang, “A new method for identifying the acute respiratory distress syndrome disease based on noninvasive physiological parameters,” PloS one, vol. 15, no. 2, p. e0226962, 2020.

[6] S. Fonck, S. Fritsch, H. Pieper, A. Baron, S. Kowalewski, and A. Stollenwerk, “Retrospective classification of ARDS in ICU time-series data using random forest with a focus on data pre-processing,” IFAC-PapersOnLine, vol. 58, no. 24, pp. 129–134, 2024.

[7] J. Cai, J. Luo, S. Wang, and S. Yang, “Feature selection in machine learning: A new perspective,” Neurocomputing, vol. 300, pp. 70–79, 2018.

[8] S. Ray, K. Alshouiliy, A. Roy, A. AlGhamdi, and D. P. Agrawal, “Chi-squared based feature selection for stroke prediction using azureml,”, pp. 1–6, IETC 2020, IEEE, 2020.

[9] A. Y. Suvorov et.al., “Statistical hypothesis testing: general approach in medical research,” jour, vol. 13, no. 1, pp. 4–13, 2021.

[10] V. Fonti and E. Belitser, “Feature selection using lasso,” VU Amsterdam research paper in business analytics, vol. 30, pp. 1–25, 2017.

[11] A. Suresh and C. Bharathi, “Sentiment classification using decision tree based feature selection,” Intl J of Control Theory and Applications, vol. 9, no. 36, pp. 419–425, 2016.