ISBN 978-989-9121-31-7
2nd International Conference on Challenges in Engineering, Medical, Economics and Education: Research & Solutions (CEMEERS-23b)
Istanbul (Turkiye) Nov. 23-24, 2023

Benefits of Deep Learning Chest X-ray Features in the Mortality Prediction of COVID-19 associated ARDS Patients

Galvão T[1], Cysneiros A[2], Jorge PM[1], Bento L[3], Domingues N[1]

[1]Instituto Politécnico de Lisboa/ Instituto Superior de Engenharia de Lisboa
[2]Nova Medical School, Universidade de Lisboa
[3]Unidade de Urgência Médica, Hospital de São José, Centro Hospitalar Universitário Lisboa Central;

Abstract: *Acute respiratory distress syndrome associated with COVID-19 (ARDS-COV19) is a severe pulmonary syndrome leading to acute respiratory failure. ARDS is complex and heterogeneous, with patients frequently needing invasive mechanical ventilation (IMV) in intensive care units (ICUs). The identification of risk groups is crucial for precision medicine, although the lack of diagnostic methods can be limiting. Chest radiography is a qualitative and accessible imaging examination routinely used in ICU settings. The development of a multivariate and quantitative classifier based on radiomics is essential for predicting the mortality of patients under IMV and to evaluate the possible benefits of imaging features in the risk stratifications of this patients. For this purpose, 87 ARDS- COV19 patients from an ICU, with an average age of 64 ±11,6 years, of whom 62.07% were male, were included. The mortality rate was 48.28%. Chest X-rays from the 1st and 3rd days of IMV were collected, pre-processed, and concatenated. Deep learning features were then extracted using a pre-trained convolutional neural network (CheXnet). These features were combined with clinical variables (CV) to build four machine learning models based on logistic regression (LogReg) and multilayer perceptron (MLP) classifiers. Age, the PaO2/FiO2 ratio on the 3rd day of IMV, and a deep learning image feature (DLF_258) were selected by filter techniques to be used in two of the models. The remaining models were created only using the CV variables (Age, PaO2/FiO2). Using 10-fold cross validation the models which included the DLF_258 showed 89% (LogReg) and 82% (MLP) probability of having better accuracy than CV models with a 77% (MLP) and 86% (LogReg) chance of improved F1-Score. In the LogReg based models, AUC had a 74% chance of being improved with the added image features while on the MLP models only a 55% chance was noted. In cross-validation the combined LogReg models provided an AUC of 0.77, 95% CI [0.67, 0.85], an accuracy of 0.75, 95% CI [0.67, 0.85], and an F1 score of 0.73, 95% CI [0.62, 0.82], while the MLP models provided an AUC of 0.78, 95% CI [0.68, 0.86], an accuracy of 0.77, 95% CI [0.67, 0.85], and an F1 score of 0.76, 95% CI [0.67, 0.84]. Despite promising results, the sample size was limited, and external testing is lacking. Therefore, data collection and subsequent validation are essential.*

Keywords: *ARDS, ICU, COVID-19, X-ray, Mortality;*

1. Introduction

Acute respiratory distress syndrome (ARDS) is a heterogeneous complication that could benefit from precision medicine. The definition of ARDS has evolved over the years. Presently, there is medical consensus on the Berlin definition, which relies on the assessment of chest radiographs (CXR) and arterial blood gas analysis. The evaluation of these gases allows for the calculation of the ratio between the partial pressure of oxygen and the inspired oxygen (PaO2/FiO2, or PF ratio), which, in turn, depends on the value of positive end-expiratory pressure (PEEP) for each respiratory cycle. These values are essential in the current diagnosis and severity assessment of the described syndrome. ARDS can thus be defined as an acute pulmonary syndrome involving bilateral infiltration and hypoxemia, not justifiable by left ventricular dysfunction, fluid overload, or chronic lung disease.

Despite its seemingly simple definition, ARDS is a syndrome of considerable complexity, as its expression and severity are co- associated with various pathologies and clinical disorders. ARDS is a common, heterogeneous, and severe syndrome, with most patients requiring invasive mechanical ventilation (IMV) in intensive care units (ICUs). Few therapeutic options have demonstrated benefits in patient prognosis beyond lung-protective ventilation, and this limitation may stem from the underlying heterogeneity of ARDS.

The creation of new methods for identifying disease subphenotypes and classifying/assessing the severity of ARDS is thus essential and largely represents the current state of the art in achieving precision medicine. A subphenotype is a subgroup of patients that present a higher risk of worse disease outcomes or share similar responses to medical procedures. Recent studies have identified two subtypes with different clinical and biological characteristics using latent class analysis: the hypo-inflammatory and hyper-inflammatory subtypes with distinct response to IMV parameters.

Recently, a new typology of ARDS has been considered, namely ARDS associated with COVID-19. The COVID-19 pandemic has caused increases in mortality and morbidity worldwide. Patients with COVID-19 are at a higher risk of requiring IMV and continue to be a current concern. Recent research as shown benefits of using corticosteroids in patients with this type of infection, which has not been proven to work in common ARDS. This demonstrates the possibility of subphenotypes associated with COVID-19 ARDS, proving that studying this syndrome in a homogenous population (COVID-19) may be beneficial. Various studies on protein biomarkers have shown a better understanding of ARDS and may offer a possible pathway to personalized approach and disease severity evaluation, however, acquisition of these biomarkers may present complications in patients undergoing IMV.

Therefore, an alternative method to these invasive diagnostic procedures is necessary. The solution may lie in the untapped quantitative diagnostic potential of medical images, as in the case of chest radiography (CXR). Portable CXR is an economical, readily available diagnostic method used daily in ICUs for ARDS diagnosis and characterization, as well as assessing the severity COVID-19. CXR is strongly associated with ARDS, with its qualitative assessment forming the basis of its definition. Some studies have analyzed chest imaging patterns and the severity of ARDS. The Radiographic Assessment of Lung Edema (RALE) score was developed as a semi-quantitative measure of lung edema in ARDS patients, with higher RALE scores being independently associated with lower PaO_2/FiO_2 ratios and worse survival. A change in RALE during the first days after the onset of ARDS is also independently associated with survival. An analysis of ICU patients under invasive ventilation found that the RALE score has excellent diagnostic accuracy for ARDS, with an area under the receiver operating characteristic curve (AUC) of 0.91. Despite the clear advantages, RALE remains a semi-quantitative method, depending on the subjective assessment of chest radiographs. This fact makes the method potentially time-consuming and subject to inter/intra-operator variabilities. Therefore, there is interest in an automatic method for quantitatively predicting mortality risk using chest radiographs to subsequently identify risk groups of the syndrome.

Machine learning in radiology is a growing research topic and has been successfully used in ARDS diagnosis with replicable quantitative computational methods. Radiomics methodologies allow the extraction of quantifiable features from medical images, either through classical image texture statistical methods or deep learning methods, for various classification, segmentation, or patient prognosis prediction tasks.

Studies focused on the categorization or prediction of ARDS severity and mortality use different angles, from biomarkers to lung physiology and chest imaging. However, few studies combine clinical data and portable CXR. The ones that do, mostly focus on mortality prediction of COVID-19 ICU patients, not considering the ARDS definition in the selected population. These studies show improved model performance metrics when using CXR image features, although they rarely provide the probability of this finding being true considering model generalization to different training groups, validation groups and sample sizes.

The present study aims to evaluate the possible benefits of using

CXR deep learning image features in COVID-19 ARDS mortality prediction using Bayesian statistics and machine learning models.

Conclusions - With the aim of creating an automatic method for predicting mortality in the early days of IMV in ICUs, ventilation parameters, blood gas values, and chest radiographs from the 1st and 3rd days of IMV were used to build a generalizable machine learning models with promising performance metrics. Due to the large volume of images produced daily in ICU care and their economic efficiency, chest radiography is an attractive proposal for investigating mortality prediction models in ARDS patients.

It was possible to extract 1024 numerical features for mortality prediction in COVID-19 patients with ARDS. This was done through image preprocessing (segmentation and contrast/noise optimization) and automatic extraction of deep learning features using a pre-trained CNN for CXR classification (CheXnet). These features were then concatenated with clinical variables to train machine learning models based on logistic regression and multilayer perceptron. After statistical analysis of the variables, the most important ones were selected using filtering techniques and cross-validation, avoiding overfitting and the "curse of dimensionality." It was found that the use of patient age, PF ratio on the 3rd day of IMV, and a deep learning CXR feature provided the best accuracy and AUC in both classifiers. It was also verified that introducing imaging features during model training provided a 89% (LogReg) and 82% (MLP) chance of improving model accuracy and a 77% (MLP) and 86% (LogReg) chance of improving F1-Score. Regarding the LogReg models, there was a 63% chance of improvement considering a 5% difference in accuracy negligeable and a 71% chance of improvement considering a 5% difference in F1-score negligeable. Thus, the model performance boost provided by the addition of imaging features may be superior to 5% considering this metrics.

Through the analysis of SHAP values from trained models, it was found that the deep learning image feature was more important than the PF ratio in predicting patient mortality during training, indicating the potential importance of radiomics in ARDS.

Using 10-fold cross-validation the combined LogReg models provided an AUC of 0.77, 95% CI [0.67, 0.85], an accuracy of 0.75, 95% CI [0.67, 0.85], and an F1 score of 0.73, 95% CI [0.62, 0.82], while the MLP models provided an AUC of 0.78, 95% CI [0.68, 0.86], an accuracy of 0.77, 95% CI [0.67, 0.85], and an F1 score of 0.76, 95% CI [0.67, 0.84]

The low sample dataset and lack of a external testing set was the main limitation of this study, introducing a higher risk of bias and variance. The fact that there was no training or fine-tuning of the CNN due to the limited amount of data also hinders the identification of what the deep learning feature represents through common techniques like Gradient-weighted Class Activation Mapping, producing a black-box algorithm. However, the qualitative analysis of the contribution of this feature in classification showed the possibility that it is associated with modifications of radiological patterns on the 1st and 3rd days of IMV. The input of concatenated images into the CNN is also significantly different from the original images the CheXNet was trained on, with the possibility of producing unoptimized imaging features for the current task.

This limitations should serve as impetus for future work since the results provided by Bayesian statistics seem to suggest that the introduction of radiological information has a high probability of promoting the identification of risk groups in ARDS-COV19. The constructed models were able to achieve promising performance metrics in cross-validation using information available in any ICU, regardless of economic and professional resources. It is expected that the models will soon be prospectively validated and be useful in selecting patients at higher risk of mortality.

Future work could reside in fine-tuning an end-to-end CNN or exploring the use of hand-crafted radiomics features. This type of radiomics information may produce more interpretable results while requiring less training data. Different Image preprocessing methodologies should also be considered in order to provide the best results for the task at hand and not just considering those that have worked previously for other goals.

The combination of quantitative radiomics methods with daily clinical and biochemical assessment in ICUs may thus be the future of personalized medicine in the context of ARDS-COV19, serving as a map to reveal what its heterogeneity may hide.

Acknowledgments- This research was funded by grant DSAIPA/DS/0117/2020, from FCT, Portugal