Proceedings on Automation in Medical Engineering
Proc AUTOMED, Vol 2, No 1 (2023): Article Ref 742
www.journals.infinite-science.de/automed/article/view/742

Infinite Science Publishing

Implementation of ResNet-50 for the Detection of ARDS in Chest X-Rays using transfer-learning

S. Fonck[1,2*], S. Fritsch[2,3,4], G. Nottenkämper[1], and A. Stollenwerk[1,2]

[1] *Informatik 11 - Embedded Software, RWTH Aachen University, Aachen, Germany*
[2] *SMITH Consortium of the German Medical Informatics Initiative, Aachen, Germany*
[3] *Department of Intensive Care Medicine, University Hospital RWTH Aachen, Aachen, Germany*
[4] *Juelich Supercomputing Centre, Forschungszentrum Juelich, Juelich, Germany*

[*] *Corresponding author, email: fonck@embedded.rwth-aachen.de*

Abstract: Acute Respiratory Distress Syndrome is a severe condition with high morbidity and mortality. The current standard for the diagnosis of ARDS was proposed by the Berlin-Definition in 2012. However, studies have shown, that ARDS is often recognized too late or not at all. Smart methods, like machine learning algorithms, may help clinicians to identify ARDS earlier and therefore initiate the appropriate therapy. To address the imaging assessment of the Berlin-Definition, a deep learning model for the detection of ARDS in x-rays is proposed. The model achieved an AUC score of 92.6%, a sensitivity of 87% and a specificity of 97%.

I. Introduction

Acute Respiratory Distress Syndrome (ARDS) is a severe condition causing pulmonary damage and respiratory failure [1]. The large observational *LUNG SAFE* study has shown, that ARDS is often recognized too late or not at all [2], even though clear criteria for the diagnosis, the Berlin-Definition was published in 2012 [3]. Artificial intelligence methods are suitable for the detection of serious diseases, as they allow the monitoring of large amounts of data and the notification of physicians about potential diagnoses, p.e. ARDS. For the diagnosis of ARDS, the Berlin Definition requires the use of imaging techniques, which is usually a time-consuming process including the consultation of a radiologist. In recent years, deep learning (DL) has become widely used for analysis (e.g.: classification, prediction or segmentation) of medical images, such as chest x-rays (CXR) or CT scans [4]. D. Moses provides a well-researched overview of DL in the medical field, i.e. for the detection of pneumonia, pneumothorax and various other diseases [5]. The detection of ARDS has been addressed in different studies [6-8]. Zaglam et al. extracted statistical and spectral features from CXRs and classified these using a support vector machine (SVM) [6]. Reamaroon et al. trained different machine learning algorithms, like SVM and Random-Forest to classify ARDS patients from directionality measure and DL features extracted from chest CXRs [7]. Sjoding et al. proposed a DL model using the DenseNet-121 architecture, which was trained with transfer-learning [8]. We propose a DL model based on the ResNet-50 architecture [9] for the detection of ARDS in CXRs, which may help accelerating the diagnostic process. The model is trained using publicly available datasets ensuring reproducibility.

II. Concept and Data

Conceptually, we used the ResNet-50 as the model's architecture, which is a convolutional neural network (CNN), specifically implemented for image recognition [9]. The model is firstly trained for the detection of pneumonia (directly associated with ARDS and with frequently similar imaging), as the database with ARDS labels is relatively small. Afterwards, we used transfer-learning by unfreezing the last block of the model to enable the detection of ARDS. The final results are then visualized using saliency maps providing interpretability and post-hoc analysis (see Figure 1). The performance is measured using the area under the receiver operating characteristic curve (AUC), Accuracy, Sensitivity, Specificity and Precision.

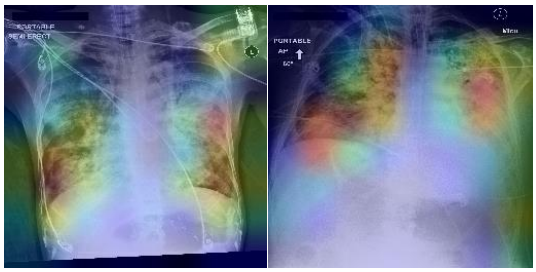

Figure 1: Saliency map for CXRs of two ARDS-patients, where the model correctly identified ARDS.

For the training and evaluation of the DL model, we used two publicly available datasets: CheXpert (C) and MIMIC-CXR (M) [10, 11]. Both datasets contain CXRs images (C: 224.316 / M: 377.110) for various patients (C: 65.240 / M: 65.379). The databases contain labels for 14 different diseases, created using a natural language processing algorithm, that screened the physician's notes [10, 11]. MIMIC-CXR can be combined with the MIMIC-IV

database, that contains vital signs, laboratory results and various information about the patient's treatment [12]. Neither of these datasets contain labels for ARDS. The training for pneumonia detection was done using a subset of CheXpert including 4.684 CXRs labeled with pneumonia and 17.000 images with no findings. As the databases contain no labels for ARDS, we searched the MIMIC-CXR dataset for radiographs of patients in whom ARDS was mentioned in the physician's report for the transfer-learning. These 550 images were annotated by a trained radiologist for ARDS and no-ARDS in addition to the confidence interval of how sure the physician was with the decision to include label uncertainty [13]. The resulting dataset contains 163 radiographs for ARDS and 387 for No-ARDS (Unweighted). Additionally, we weighted the dataset based on the confidence level for ARDS by duplicating and augmented the according high-confidence ARDS-images resulting in 267 CXRs for ARDS and 387 for no findings (Weighted). The datasets are split into training, test and evaluation data (60%/20%/20%).

III. Results & Discussion

The ResNet-50 trained with datasets for pneumonia achieved moderate performance (see Table 2).

AUC	Acc.	Sens.	Spec.	Prec.
0.77	0.78	0.75	0.81	0.83

Table 2: Accuracy, Sensitivity, Specificity and Precision are given for a threshold of 0.5.

The results for the detection of ARDS in CXRs after transfer-learning using evaluation data can be found in Table 3 for the unweighted/weighted dataset respectively.

Dataset	AUC	Acc.	Sen.	Spec.	Prec.
Unweighted (U)	0.85	0.84	0.89	0.82	0.88
Weighted (W)	0.926	0.923	0.87	0.96	0.97

Table 3: Accuracy, Sensitivity, Specificity and Precision are given for a threshold of 0.5.

Compared to related work, ResNet-50 trained with the weighted dataset performs equally well, even surpassing other works in Specificity (see Table 4).

	Zaglam et al. [6]	Reamaroon et al. [7]	Sjoding et al. [8]	ResNet-50 (W)
AUC	-	0.79	**0.93**	0.926
Sen.	**0.906**	-	0.88	0.87
Spec.	0.865	-	0.83	**0.97**

Table 4: AUC, Sensitivity, Specificity (if available) are given for the trained ResNet and related works.

Our research also has limitations: The data set for training for ARDS detection was only annotated by one radiologist. By involving additional domain experts and an appropriate consensus process, more robust and generalizable results could be obtained. The training data set for ARDS detection is small compared to the data used for pneumonia detection. In further work, we aim to annotate more images to enable a more comprehensive classification. In addition, there are images for which the model's assessment visualized in the saliency map is not comprehensible, although the classification is correct. This circumstance should be addressed by external validation involving further experts.

IV. Conclusions

ARDS is a critical condition that requires imaging assessment for diagnosis. This process is usually very time-consuming and requires additional consultation. With the help of neural networks, this process can be accelerated and the specialist staff can be supported. We have trained a CNN based on the well-established ResNet-50 architecture to detect ARDS in CXRs using transfer-learning. Our results can keep up with those from related literature and even surpass them in specificity. In future work we would like to incorporate bigger datasets annotated by more physicians and experts to increase the robustness and transferability of the results. Additionally, more architectures of CNNs (like the UNet, DenseNet-121) or Transformer Models can be included providing an extended classification and assessment of different methods. Merging the methods could further support the diagnostic process. Additional processing of the image data, such as gray level adjustment, could also increase the transferability of the trained model to external data sets from different X-ray tubes. A comparative analysis of different loss functions and optimizers included in the classification process as well as an hyperparameter tuning, could further improve the automatic detection of ARDS in CXRs.

Author's statement

Research funding: The research is funded by the German Federal Ministry of Education and Research with the grant codes: 01ZZ1803B, 01ZZ1803K and 01ZZ1803M. Conflict of interest: Authors state no conflict of interest. Ethical approval: Not applicable. Acknowledgements: The authors thank Maximilian Schulze-Hagen, M.D., M.Sc., for annotating the images used for ARDS detection.

References

[1] D. G. Ashbaug et al. *Acute Respiratory Distress in Adults*, The Lancet, 1967

[2] G. Bellani et al., *Epidemiology, patterns of care, and mortality for patients with Acute Respiratory Distress Syndrome in Intensive Care Units in 50 countries*, Journal of the American Medical Association, 2016

[3] G. D. Rubenfeld et al., *Acute Respiratory Distress Syndrome – The Berlin Definition*, American Medical Association; 307, 2012

[4] E. Calli et al., *Deep learning for chest X-ray analysis: A survey*, Medical Image Analysis; 72, 2021

[5] D. A. Moses, *Deep learning applied to automatic disease detection usingchest X-rays*, Journal of Medical Imaging and Radiation Oncology; 65, 2021

[6] N. Zaglam et al., *Computer-aided diagnosis system for the Acute Respiratory Distress Syndrome from chest radiographs*, Computers in Biology and Medicine; 52, 2014

[7] N. Reamaroon et al., *Automated Detection of Acute Respiratory Distress Syndrome from Chest X-Rays Using Directionality Measure and Deep Learning Features*, Computers in Biology and Medicine; 134, 2021

[8] M. W. Sjoding et al., *Deep learning to detect acute respiratory distress syndrome on chest radiographs: a retrospective study with external validation*, Lancet Digital Health; 3, 2021

[9] K. He et al., *Deep Residual Learning for Image Recognition*, IEEE Conference on Computer Vision and Pattern Recognition, 2016

[10] J. Irvin et al., *CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison*, Conference on Artificial Intelligence, 2019

[11] A. E. W. Johnson et al., *MIMIC-CXR-JPG, A large publicly available Database of labeled Chest Radiographs*, 2019

[12] A. Johnson et al., *MIMIC-IV (version 2.1)*, PhysioNet, 2022

[13] N. Reamaroon et al., *Accounting for Label Uncertainty in Machine Learning for Detection of Acute Respiratory Distress Syndrome*, Biomedical and Health Informatics; 23, 2019