Acute Respiratory Distress Syndrome (ARDS) Detection in the Pediatric Intensive Care Unit (PICU) setting Demonstrates High Performance with Transfer Deep Learning

Vahid Khalkhali, PhD, Research Scientist, Children's Hospital of Philadelphia; Michael Welsh, DO; Patricia P. Rafful, MD, PhD; Dana S. Alkhulaifat, MD; Adarsh Ghosh, MD; Saurav Bose, MD; Nadir Yehya, MD; Susan T. Sotardi, MD

Introduction

Acute respiratory distress syndrome (ARDS) is a significant cause of morbidity and mortality in the pediatric intensive care unit (PICU). ARDS diagnosis involves chest X-ray (CXR) criteria combined with clinical and laboratory parameters. Machine learning models have demonstrated utility in the detection of ARDS on chest radiographs.

Hypothesis

Machine learning models for radiology CXR can perform better than humans. We evaluated the performance of deep learning (DL) models to diagnose ARDS based on CXR exams. We performed statistical evaluation performance between the models and two experienced radiologists, and also statistical agreement evaluation between the two radiologists as a gold-standard correlation for comparison.

Methods

In this retrospective, IRB-approved study, we identified 368 children admitted to the PICU with a diagnosis of ARDS, at a large pediatric academic center from 2014 to 2019. A single random radiograph from all patients admitted to the PICU without a diagnosis of ARDS during 2018 was used as the control cohort, (n=1127). The train-validation-test ratio was 60/20/20. Using transfer learning, we utilized pretrained DL structures to diagnose ARDS (PyTorch, version 1.2). The area under the receiver operating characteristic (AUROC) was the main performance metric. Two pediatric radiologists independently assigned labels of ARDS/No ARDS and interrater reliability was calculated. Correlations were calculated using the Pearson correlation coefficient and Cohen’s Kappa. All statistical analyses used Type-I error of 5% and power of 80%.

Results

The interrater reliability between the radiologists was 94.5% (Cohen’s Kappa of 85.8%) for the training cohort. The ARDS diagnostic performance of two radiologists yielded an AUROC (balanced accuracy) of 72.5%, while the DenseNet161 model achieved 86.0% (AUROC of 92.5%) and an ensemble of models reach 83.7% (AUROC of 93.5%). Radiologist diagnoses were only 81% correlated (Cohen’s Kappa 51.6%) with the DenseNet161 and 83% (Cohen’s Kappa 60.5%) with the ensemble. While the Pearson correlation between the two radiologists was high (>90%) on the test set, the difference between the detection of different models was statistically significant ($p < 0.01$). Attention maps show that models are able to capture the regions of interest (Figure 1).

Conclusion

Using transfer DL, we trained models to reliably detect ARDS in the PICU and compared their performance with the diagnostic rates of two experienced radiologists. DL can automatically detect ARDS on chest radiographs, with a performance that parallels those of radiologists.

Statement of Impact

Deep learning detection of ARDS could improve the triage of patients in the intensive care unit before the availability of dedicated pediatric radiologist reads.

Figure 1. Attention maps for CNN: left and right examples are non-ARDS and ARDS, respectively.

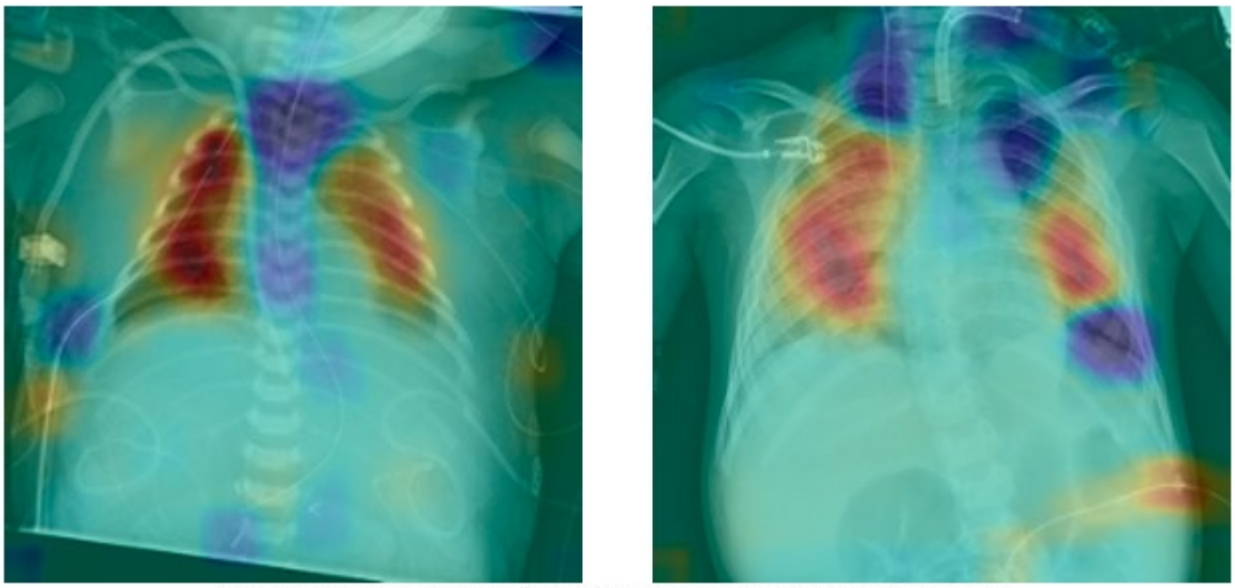

Figure 1. Attention maps for CNN: left and right examples are non-ARDS and ARDS, respectively.

Keywords

Acute Respiratory Distress Syndrome (ARDS)**;** Transfer Deep Learning**;** Chest X-Ray (CXR)