

AWARD NUMBER: W81XWH-20-1-0496

TITLE: Computer Vision Technologies for Rapid Detection of the Acute Respiratory Distress Syndrome

PRINCIPAL INVESTIGATOR: Michael W Sjoding

CONTRACTING ORGANIZATION: University of Michigan, Ann Arbor, MI

REPORT DATE: October 2022

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release; Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE October 2022	2. REPORT TYPE Annual	3. DATES COVERED 01Aug2021-31Jul2022
4. TITLE AND SUBTITLE Computer Vision Technologies for Rapid Detection of the Acute Respiratory Distress Syndrome		5a. CONTRACT NUMBER W81XWH-20-1-0496
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Michael W Sjoding E-Mail: msjoding@umich.edu		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Regents of the University of Michigan 3003 S. State Street Wolverine Tower Ann Arbor, MI 48109-1274		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Development Command Fort Detrick, Maryland 21702-5012		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT This report summarizes the progress made over the second year on the grant: W81XWH2010496/Computer Vision Technologies for Rapid Detection of the Acute Respiratory Distress Syndrome. The Acute Respiratory Distress Syndrome (ARDS) is a critical illness syndrome with a 35% mortality rate. We proposed to develop computer vision technologies powered by deep convolutional neural networks to automatically identify chest x-ray findings consistent with ARDS with expert-level accuracy. During the second year of the grant, we performed several computational analyses to improve our published ARDS model including 1) increasing the model pre-training time, 2) investigating the effect of training a model on larger chest x-ray images, 3) incorporating a lung-segmentation algorithm into our chest x-ray processing pipeline. Of the three strategies, increasing training time resulted in the largest improvement in model validation performance over our base model. We also performed several investigations to understand and compare model performance to human expert performance in identifying ARDS on chest x-rays. We are working towards submitting this for publication. This research is propelling our group towards a new focus of enabling physician and artificial intelligence collaborations in the diagnosis of ARDS, and will likely be the future focus of projects and grants in our group. Overall, we remain on track towards completion of all aspects of the grant within the no-cost extension period.		

15. SUBJECT TERMS Acute Respiratory Distress Syndrome; Deep Convolutional Neural Network; Machine Learning; chest x-rays					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	Unclassified	20	USAMRDC
Unclassified	Unclassified	Unclassified			19b. TELEPHONE NUMBER <i>(include area code)</i>

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

	<u>Page</u>
1. Introduction	3
2. Keywords	3
3. Accomplishments	3
4. Impact	10
5. Changes/Problems	11
6. Products	13
7. Participants & Other Collaborating Organizations	15
8. Special Reporting Requirements	17
9. Appendices	17

1. INTRODUCTION:

The Acute Respiratory Distress Syndrome (ARDS) is a critical illness syndrome affecting patients with trauma, sepsis, pneumonia, and aspiration, and it has a 35% mortality rate. However, up to 65% of patients with ARDS are diagnosed late or missed and as a result, do not receive evidence-based therapies. In this grant, we proposed to develop novel computer vision technologies powered by deep convolutional neural networks to identify chest x-ray findings consistent with ARDS with expert-level accuracy. We proposed to train a deep convolutional neural network using an existing cohort of 2,237 patients and 9,632 chest x-rays, and then test the algorithm on chest x-rays from two validation cohorts, reviewed by critical care physicians and radiologists, and compare the model's performance to individual experts. An innovative aspect of the grant is also to evaluate the benefit of training the model to first localize the lung fields before discriminating ARDS. In addition, we proposed to develop a novel network output display, allowing a user to better interpret and evaluate the network's result. We envision that the technology developed in this grant would be deployed on portable digital imaging systems, supporting and augmenting front-line providers in the prompt diagnosis of ARDS. This system could provide critical diagnostics, enabling rapid identification and triage of patients with ARDS to ensure prompt treatment, ultimately leading to improved patient outcomes.

2. KEYWORDS:

Acute Respiratory Distress Syndrome, Deep Convolutional Neural Network, Machine Learning, Chest x-rays, Detection

3. ACCOMPLISHMENTS:

What were the major goals of the project?

Aim 1: Train a deep convolutional neural network using an existing cohort of 2,237 patients and 9,632 chest x-rays to detect images consistent with ARDS

Major Task 1: Train deep learning models to identify chest x-rays with findings consistent with ARDS using a previously adjudicated cohort of critically ill patients hospitalized at the University of Michigan (U-M). Original timeline: Project month 1-12. **Status: Completed on 11/30/2020**

Task 2: Train a deep learning lung segmentation algorithm to identify lung fields on chest x-rays and combine it with the deep learning ARDS detection model (to be completed in month 6-14 of the proposal. Timeline: Project month 6-12

Status: Initial algorithm completed on 8/6/2021, completed additional work on developing U-net segmented model on 1/15/2022

Aim 2: Test the ARDS deep neural network on chest x-rays from two validation cohorts, reviewed by critical care physicians and radiologists, and compare the model's performance to individual experts

Major Task 1: Internally validate the ARDS deep learning neural network on a temporally distinct cohort of critically ill U-M patients. Timeline: Project month 4-16.

Status: In process, continued efforts to investigate physician and model performances and understand the differences between the two (2 of 3 subtasks completed)

Major Task 2: Externally validate the ARDS deep learning network on critically ill patients hospitalized at the University of Pennsylvania (U-Penn). Timeline: Project month 4-16.

Status: Completed on 11/1/2021 for initial algorithm, Completed for updated version of algorithms on 1/15/2022

What was accomplished under these goals?

For this reporting period describe: 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative); and/or 4) other achievements. Include a discussion of stated goals not met. Description shall include pertinent data and graphs in sufficient detail to explain any significant results achieved. A succinct description of the methodology used shall be provided. As the project progresses to completion, the emphasis in reporting in this section should shift from reporting activities to reporting accomplishments.

Aim 1: Train a deep convolutional neural network using an existing cohort of patients rays to detect images consistent with ARDS

Major Task 1: Train deep learning models to identify chest x-rays with findings consistent with ARDS using a previously adjudicated cohort of critically ill patients hospitalized at the University of Michigan (U-M).

Specific objectives:

1. HRPO/ACURO Approval – 9/28/2020
2. Michigan IRB Approval – 5/8/2020
3. Train an initial deep learning models to detect chest x-rays with ARDS

Major activities and significant results: Building on the success of our initial ARDS model, we studied other strategies for improving the model performance. This included increasing model pre-training from 3 epochs to 15 epochs on the MIMIC-CXR and CheXpert datasets prior to training on our ARDS dataset, and also testing whether a model that analyzed a 512 x 512 pixel image compared to a 320 x 320 pixel image. Both of these efforts required extensive computational efforts and had mixed results. We found that additional pretraining did seem to improve the overall

model’s performance, particularly in the external validation data (AUROC = 0.910 vs 0.880). However, using a larger pixel size image did not meaningfully improve model performance any further. These results are shown in Table 1. Based on these results, we will use the model with additional pre-training in all subsequent results.

Table 1. Model performance after additional pre-training and with larger image size

	Initial model with additional pre-training	New model analyzing larger image size
<i>Pre-processing</i> Resized to	320 x 320	512 x 512
<i>Pre-training</i> Datasets	MIMIC & CheXpert	MIMIC & CheXpert
Initialization Weights	ImageNet	ImageNet
Pretraining epochs	15	15
Best pretraining epoch	15	12
<i>Testing</i> UM Test cohort AUROC (n = 1560)	0.919 (0.907 - 0.93)	0.921 (0.909 - 0.932)
Penn cohort AUROC	0.910 (0.886 - 0.933)	

Major Task 2: Train a deep learning lung segmentation algorithm to identify lung fields on chest x-rays and combine it with the deep learning ARDS detection model.

Specific objectives:

1. Identify a subset of chest x-rays from the training data and annotate the lung fields
2. Train a deep learning algorithm to segment the lung fields in chest x-rays of critically ill adults
3. Combine the lung segmentation and ARDS detection algorithms and evaluate the combined algorithm performance

Major activities and significant results: Using our successfully trained a deep learning model based on a U-net architecture that segments lung fields, we developed a combined lung segmentation and ARDS detection model. This model uses the U-net algorithm to “crop” the image and zoom in on the lung fields. To accomplish this, we also had to apply the same combined algorithm to the CheXpert and MIMIC-CXR images, and pre-train the model on the cropped version of these images. Then, we took the same approach to the cropped ARDS adjudicated images and trained the model to detect ARDS. Overall, we found that this approach did not meaningfully improve performance compared to the new version of our ARDS detection algorithm which was pre-trained for 15 epochs (**Table 2**).

Table 2. Comparing baseline and U-net lung-cropped model on detecting ARDS

	Baseline model with additional pre-training	U-net cropped model
UM test AUROC (n = 1560)	0.919 (0.896-0.940)	0.920 (0.898-0.941)
UM test AUROC 6 physicians (n = 413)	0.936 (0.890-0.970)	0.913 (0.864-0.956)
UPenn test AUC (n = 958)	0.910 (0.886-0.933)	0.905 (0.882-0.927)

Because U-net cropping did not convincingly improve model performance, we developed an alternative approach where we use the U-net model to segment the lungs and then analyze each lung independently and combine the two individual scores to determine whether the patient has ARDS. The approach is illustrated below in **Figure 1**, which we have called the “dual-input DenseNet.” In this approach, the U-net is used to identify each lung field, and then each lung is sent through the DenseNet model separately, and then the scores are combined to determine whether ARDS is present in the image.

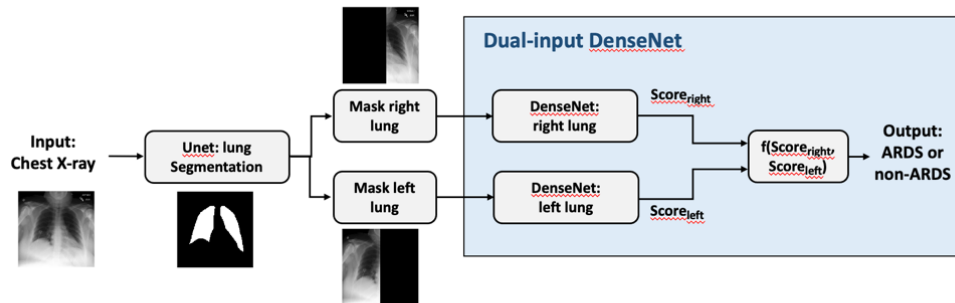


Figure 1. Illustration of the dual-input ARDS detection model

Table 3. Comparison of standard and dual-input densenet model performance on UM test set.

Model	Test set AUROC (n=1624)
Standard ARDS model	0.907
Dual input model: $Score_{right} \times Score_{left}$	0.906
Dual input model: $Average(Score_{right}, Score_{left})$	0.908

Aim 2: Test the ARDS deep neural network on chest x-rays from two validation cohorts, reviewed by critical care physicians and radiologists, and compare the model’s performance to individual physician experts

Major Task 1: Internally validate the ARDS deep learning neural network on a temporally distinct cohort of critically ill U-M patients hospitalized.

Specific objectives:

1. Identify a UM validation cohort
2. Perform 6-physician adjudication of patients in this cohort
3. Evaluate performance of the algorithm compared to individual expert physicians

Major activities and significant results: We have now performed extensive validation of our deep neural network model on a test cohort of chest x-rays that has been each reviewed by at least 6 experts including a chest radiologist. This work included validation of the initial deep neural network model and additional testing of several updated versions of the model. Some of these results are described in the above Tables. Our analysis has shown that the model performs at least the same if not better than physicians overall. In a subset of 414 chest x-rays from 115 adult patients, we found the model had overall higher accuracy in detecting ARDS compared to physicians (0.847 [95%CI 0.826-0.868] vs. 0.807 [95%CI 0.789-0.827]). Physicians had significantly lower average sensitivity (0.732 [95% CI 0.695-0.768]) across the subsets of chest X-rays that each reviewed compared to the AI model's sensitivity (0.791 [95% CI 0.733-0.846]). The AI model also maintained a significantly higher specificity (0.865 [95% CI 0.844-0.888]) than physicians (0.838 [95% CI 0.822-0.854]).

We then performed several further investigations to better understand how the model and physicians perform differently and understand why the model performance better on average than physicians. We wanted to understand whether the model is more consistent at classifying chest x-rays deemed to be relatively easy to interpret correctly compared to physician experts or whether the model is better at classifying chest x-rays deemed more difficult. To perform this analysis, we made use of annotations made by physicians where they characterized the chest x-ray as more difficult to interpret or relatively easy to interpret by quantifying the level of certainty that physicians had when classifying images. In the below **Figure 2**, we found that the model was generally more consistent at classifying the 74.6% of images deemed not as difficult to classify than physicians but was slightly worse at classifying the 25.4% of images that were deemed difficult to classify.

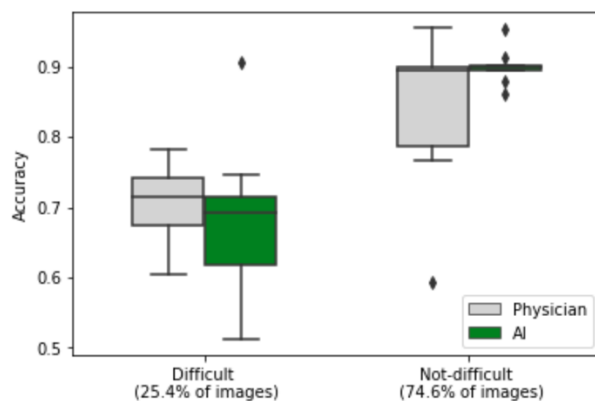


Figure 2. Comparing the performance of expert physicians and deep learning AI model stratified by the difficulty of image interpretation.

We also wanted to understand whether the model and expert physicians had difficulty with the same images or different images. To identify images that the model had more difficulty classifying, we looked at model output probabilities and grouped images as difficult if the model probability was between 0.36 and 0.64. As shown in the below figure, it was rare for physicians and model to both have difficulty classifying images, which occurred in only 1.7% of the images.

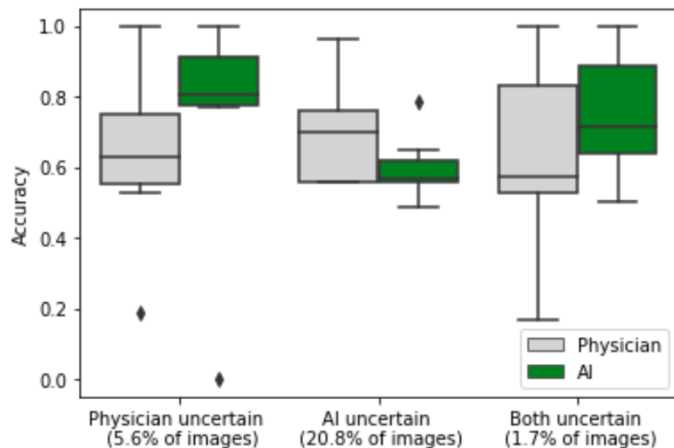


Figure 3. Comparing the performance of expert physicians and deep learning AI model stratified by the uncertainty level in their image interpretation.

From this analysis we concluded that expert physicians and the deep learning model have complementary expertise and having them work together to classify images for ARDS may achieve the best overall performance. This is important insights as we move forward with planning how to most effectively deploy the ARDS model in future work.

Major Task 2: Externally validate the ARDS deep learning network on critically ill patients hospitalized at the University of Pennsylvania (U-Penn).

Specific objectives:

1. Acquire U-Penn chest x-rays
2. Perform additional adjudication of the U-Penn chest x-rays
3. Evaluate the generalization performance of the system.

Major activities and significant results: We have received the chest x-rays from U-Penn and completed additional adjudication of the chest x-rays for ARDS. We have evaluated the generalization performance of the deep learning system and found it to have good external performance. These results are shown in the above **Tables**.

What opportunities for training and professional development has the project provided?

Nothing to report

How were the results disseminated to communities of interest?

We had several opportunities to disseminate our work to communities of interest over the last reporting period. The work on developing the deep learning Model to detect ARDS was presented as a podium talk during a Scientific Symposium at the American Thoracic Society international conference in May 2022. During the talk, the work comparing human and model expertise was presented. Similar work was also presented at Machine Learning in Medicine (MLxMed) a joint seminar series sponsored by the University of Pittsburgh, Carnegie Mellon, and the University of Toronto.

What do you plan to do during the next reporting period to accomplish the goals?

During the final 6 months of this project, we will focus our attention on completing all aspects of specific aim 2. This includes completing the analysis comparing human expert and algorithm performance and submitting this analysis for publication. We will also all complete analysis and

testing of the newest version of the ARDS detection algorithm which incorporates dual lung segmentation and evaluate performance on the UM and UPenn test cohorts.

4. IMPACT:

What was the impact on the development of the principal discipline(s) of the project?

Our work on this project has already had an impact on the field of ARDS research. Our primary research paper published from this project to date has already been cited by several other groups working on related algorithmic development, including the diagnosis of pneumonia or influenza. Additionally, our pediatric collaborators have partnered with our team to modify and apply the adult ARDS detection model to pediatric patients to detect pediatric ARDS. Overall, our work has demonstrated the ability of machine-learning systems to achieve human-level performance at complex diagnostic tasks in clinical medicine and it is a model being replicated by others in the field.

What was the impact on other disciplines?

Other disciplines, especially in computer science and engineering, will likely benefit from the research performed and knowledge gained from analysis performed in this grant to date. We have demonstrated the benefits and limitations of “transfer learning,” a technique in computer science where an algorithm is trained on one type of problem, and then that knowledge is used to help the algorithm more quickly learn another type of problem. Our most recent analysis also uncovers the strengths and limitations of machine-learning algorithms compared to human experts and how they ultimately may complement each other in health care practice. These insights have broad applicability to other disciplines seeking to develop algorithms to support and augment human experts.

What was the impact on technology transfer?

In concert with Innovation partnerships at the University of Michigan, we are in ongoing talks with the company AirStrip, who is considering licensing the ARDS detection algorithm technology and including it among their product line.

What was the impact on society beyond science and technology?

Nothing to report

5. CHANGES/PROBLEMS:

Nothing to report

Changes in approach and reasons for change

Nothing to report

Actual or anticipated problems or delays and actions or plans to resolve them

See below for the description of changes that had a significant impact on expenditures

Changes that had a significant impact on expenditures

During the COVID-19 pandemic, we had difficulty hiring and retaining staff, including our research scientist and data scientist. For this reason, we requested a no-cost extension on 8/30/2021 to extend the grant for an additional year which was approved. This additional time has allowed us to bring new people onto the project (Data Scientist Negar Frazaneh), allowing us to get our project back on track and complete all tasks.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Significant changes in use or care of human subjects

Nothing to Report

Significant changes in use or care of vertebrate animals

Nothing to Report

Significant changes in use of biohazards and/or select agents

Nothing to Report

6. PRODUCTS:

- **Publications, conferen
Journal publications.**

Nothing to report

Books or other non-periodical, one-time publications.

Nothing to report

Other publications, conference papers and presentations.

Nothing to report

- **Website(s) or other Internet site(s)**

A description of the main product that has been developed under the support of this DOD grant can be found here: <https://weil.institute.med.umich.edu/product-portfolio/detect-ards>
This website was created to disseminate the technology to companies who may want to pursue a licensing agreement.

- **Technologies or techniques**

The primary technology, an analytic algorithm that can analyze a chest x-ray and determine whether the patient has ARDS has been created and can be licensed.

- **Inventions, patent applications, and/or licenses**

Nothing to report

- **Other Products**

Nothing to report

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name: Michael Sjoding, MD

Project Role: PI

Researcher Identifier (e.g. ORCID ID): 0000-0002-0535-9659

Nearest person month worked: 1

Contribution to Project: Dr. Sjoding is the PI of the project and oversaw all work related to the project, including clinical annotations of chest x-rays by participating physician reviewers, model development and testing.

Name: Elizabeth Lee, MD

Project Role: Co-I

Researcher Identifier (e.g. ORCID ID): 0000-0001-5404-1447

Nearest person month worked: 1

Contribution to Project: Dr. Lee is a chest radiologist, who oversees and supports the annotation of chest x-ray images for ARDS as well as provides general guidance for the overall scientific direction on the project.

Name: Sardar Ansari, PhD

Project Role: Co-I

Nearest person month worked: 6

Contribution to Project: Dr. Ansari is the chief data scientist on the project, overseeing and carrying out the training of the deep learning model for detection of ARDS and supervising the daily work of the data analyst.

Name: Kevin Ward, MD

Project Role: Co-I

Researcher Identifier (e.g. ORCID ID):

Nearest person month worked: 1

Contribution to Project: Dr. Ward provides clinical expertise to guide the development of the deep learning models, leadership for the proposal in its overall direction, overseeing the clinical data collection and assisting Dr. Sjoding with model evaluation, analysis, and reporting

Name: Negar Frazaneh, PhD

Project Role: Data Scientist

Researcher Identifier (e.g. ORCID ID): 0000-0003-1200-5274

Nearest person month worked: 3

Contribution to Project: Dr. Frazaneh is a data scientist who participated in model training and evaluation

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Nothing to report

What other organizations were involved as partners?

Nothing to report

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS:

QUAD CHARTS:

9. APPENDICES:

Quad Chart attached