Computers in Biology and Medicine 134 (2021) 104463

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Automated detection of acute respiratory distress syndrome from chest X-Rays using Directionality Measure and deep learning features

Narathip Reamaroon [a,*], Michael W. Sjoding [b], Jonathan Gryak [a], Brian D. Athey [a], Kayvan Najarian [a], Harm Derksen [c]

[a] Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, United States
[b] Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States
[c] Department of Mathematics, Northeastern University, Boston, MA, United States

ARTICLE INFO

Keywords:
Acute respiratory distress syndrome
Chest X-ray images
Image processing
Machine learning
Deep learning

ABSTRACT

Acute respiratory distress syndrome (ARDS) is a life-threatening lung injury with global prevalence and high mortality. Chest x-rays (CXR) are critical in the early diagnosis and treatment of ARDS. However, imaging findings may not result in proper identification of ARDS due to a number of reasons, including nonspecific appearance of radiological features, ambiguity in a patient's case due to the pathological stage of the disease, and poor inter-rater reliability from interpretations of CXRs by multiple clinical experts. This study demonstrates the potential capability of methodologies in artificial intelligence, machine learning, and image processing to overcome these challenges and quantitatively assess CXRs for presence of ARDS. We propose and describe Directionality Measure, a novel feature engineering technique used to capture the "cloud-like" appearance of diffuse alveolar damage as a mathematical concept. This study also examines the effectiveness of using an off-the-shelf, pretrained deep learning model as a feature extractor in addition to standard features extracted from the histogram and gray-level co-occurrence matrix (GLCM). Data was collected from hospitalized patients at Michigan Medicine's intensive care unit and the cohort's inclusion criteria was specifically designed to be representative of patients at risk of developing ARDS. Multiple machine learning models were used to evaluate these features with 5-fold cross-validation and the final performance was reported on a hold-out, temporally distinct test set. With AdaBoost, Directionality Measure achieved an accuracy of 78% and AUC of 74% – outperforming classification results using features from the histogram (75% accuracy and 73% AUC), GLCM (76% accuracy and 73% AUC), and ResNet-50 (77% accuracy and 73% AUC). Further experimental results demonstrated that using all feature sets in combination achieved the best overall performance, yielding an accuracy of 83% and AUC of 79% with AdaBoost. These results demonstrate the potential capability of using the proposed methodologies to complement current clinical analysis for detection of ARDS from CXRs.

1. Introduction

Acute respiratory distress syndrome (ARDS) is a critical illness with a mortality rate of 40%, affecting 200,000 patients in the United States and 3 million globally each year [1,2]. ARDS is characterized by diffuse alveolar damage, often resulting in widespread edema and fluid buildup in the lungs. The clinical corollary is catastrophic and generally associated with poor outcomes, with such risks increasing with age and severity of illness [3]. Furthermore, early studies have shown that ARDS is associated with the novel coronavirus disease 2019 (COVID-19) – specifically in severe cases where it is hypothesized that the virus travels beyond the upper airway, moving through the lungs and causing a widespread inflammation in the alveoli [4].

Imaging is integral to the care of these critically ill patients. CXR (chest x-ray) features of ARDS usually develop 12–24 h after initial lung insult. Although appearances in manifestation can vary depending upon the stage of the disease, CXRs of patients typically exhibit characteristics of diffuse bilateral opacities with dense consolidation. Because of this, chest x-rays are a critical resource that can support an early diagnosis and provide evidence-based management strategies to patients with

* Corresponding author. Department of Computational Medicine and Bioinformatics, University of Michigan, 100 Washtenaw Avenue, Palmer Commons #2017, Ann Arbor, MI, 48109, United States.
E-mail address: nreamaro@umich.edu (N. Reamaroon).

https://doi.org/10.1016/j.compbiomed.2021.104463
Received 30 November 2020; Received in revised form 15 April 2021; Accepted 28 April 2021
Available online 11 May 2021

ARDS to improve their outcomes [5]. Identification of pulmonary opacification is a requirement for diagnosis of ARDS; however, radiological features by themselves are nonspecific and may not be correlated with clinical findings [6,7]. As a result, inconsistencies in interpretability of chest imaging and poor inter-rater reliability suggest that patients with ARDS are not recognized when they develop this illness. Consequently, they do not receive the therapies proven to reduce mortality [8,9].

We therefore hypothesize that it may be possible to utilize advances in machine learning for the detection of ARDS from chest x-ray scans. The aim of this study was to develop an algorithm capable of extracting features that identify lung injury and examine the strength of deep learning approaches for pathology detection in CXRs.

To analyze CXRs for the presence of ARDS, it is critical to investigate the lung fields for pulmonary opacification – which manifests as a "cloud-like" appearance on radiographs. We propose a method, Directionality Measure, which aims to capture this intuitive notion of diffuse alveolar injury damage and "cloudiness" as a mathematical concept. The basis of this approach is to strongly blur along areas of the image that exhibits directionality and also in regions where there are few details. Artifacts and peripheral structures in the CXR (e.g., ribs, vasculature, medical equipment) typically have features of directionality, while the detail-rich regions of the lungs and areas with opacification do not. This approach enables quantification of lung injury by capturing a diverse set of properties and measurements that may be suitable indicators for ARDS.

In addition to Directionality Measure, this study also examines other features that have been used for similar applications in the detection of related lung diseases. For example, first-order statistics calculated from the histogram are also included in this work. These features have shown promising results in previously published works as a textural descriptor to differentiate a healthy lung field compared to a CXR present with lung injury [10]. Furthermore, we also extracted features from the gray-level co-occurrence matrix (GLCM). These features characterize the texture of an image by considering the spatial relationship and dependencies in the matrix. Although features from the GLCM have been used to train machine learning models for detection of various lung diseases (e.g., pneumonia and atelectasis), we did not find any research or studies applying GLCM features for detection of ARDS [11–13].

We also investigate the use of transfer learning with pretrained neural networks for extracting additional features that can be used to train machine learning algorithms to detect ARDS. Recent studies have indicated that information extracted from certain layers of convolutional neural networks can be very powerful features for use in classification tasks [14]. For example, neurons in the first layer learn features similar to Gabor filters while those from the last layer are more specific to the given learning task [15]. Initializing a network with transferred features from different layers can yield boosts to generalization even after fine-tuning to the target dataset [16]. Previously published works demonstrate this notion of transfer learning and document the success of using these features extracted from intermediate and higher layers of CNNs for recognition tasks that the network was not trained on [16–19]. Furthermore, this approach of using deep learning models trained on large scale, non-medical data to extract features for general medical image recognition tasks via transfer learning has been demonstrated with favorable results by multiple research groups [20,21]. Although several research groups have used pretrained networks to extract features from CXRs, we are not aware of any studies that evaluate the feasibility of using transfer learning in this capacity for detection of ARDS [11,22].

In this study, we evaluate these approaches - Directionality Measure, first-order statistics from the histogram, GLCM, and pretrained deep neural networks - on CXRs obtained from Michigan Medicine and use the extracted features to train machine learning models for the detection of ARDS. Support vector machine (SVM), random forest, AdaBoost, and RUSBoost models were trained and evaluated with 5-fold cross-validation on these features from 2018 CXRs (data from 70% of patients); the final overall performance was then reported on a hold-out test set comprised of the remaining 1060 CXRs (data from 30% of patients). To better understand the strength and contribution of each technique, we present the results of classification when using each feature set separately and also combining all features when training the machine learning algorithms.

2. Data

This study was approved by the Institutional Review Board with a waiver of informed consent. The patient cohort consists of adult patients hospitalized in intensive care units at Michigan Medicine between 2016 and 2017. We retrospectively identified patients with moderate hypoxia (requiring more than 3 L of supplemental oxygen by nasal cannula for at least 2 h) and acute hypoxic respiratory failure (PaO2/FiO2 ratio of $<$300 mm Hg while receiving invasive mechanical ventilation).

This data is comprised of critically ill patients with respiratory failure. The cohort selection criteria were intentionally designed as such, rather than an investigation of ARDS vs. healthy patients, to create a realistic representation of the patient population and clinical settings where these algorithms would be used. Because these data were acquired from hospitalized settings, the CXRs tend to be more complex than a standard chest imaging obtained from controlled studies or outpatient settings. Characteristics of these complexities in the CXRs include varying quality (e.g., dynamic range and sharpness), presence of introduced medical devices, diverse body habitus, and manifestation of disease. Examples of CXRs in this dataset are provided in Fig. 1.

In total, 500 patients were included in this study and 3078 anterior-posterior chest x-rays were obtained. Of this population, 208 patients met the criteria for ARDS after being reviewed independently by multiple clinical experts. Data from 70% of patients (approximately 2018 chest x-rays) were used for model training and validation while the remaining 30% of patients (1060 chest x-rays) were used as the hold-out test set. There are 191 females (mean age of 58 years, 32% ARDS positive) and 309 males in this study cohort (mean age of 57 years, 29% ARDS positive). The cohort demographics are summarized in Tables 1 and 2.

The chest x-rays in this study were reviewed independently by multiple clinical experts to generate the labels used in training the machine learning algorithms. ARDS is a life-threatening condition without a "gold standard" for diagnosis and the inter-rate reliability for correct diagnosis of the illness is only moderate [9]. Because of this, multiple experts were asked to determine whether each chest x-ray is consistent with ARDS and also to provide a confidence level in their diagnosis - high, moderate, slight, or equivocal. This information was converted to scale between 1 and 8 as illustrated in Fig. 2. If the clinical experts' averaged review was below or equal to 4.5, a label of −1 (no ARDS) would be assigned to the chest x-ray. If the averaged review was above 4.5, a label of 1 (ARDS) would be assigned.

3. Methodology

3.1. Pre-processing and lung segmentation

Chest x-rays are acquired in DICOM format and converted to an 8-bit grayscale image. Lung segmentation was performed with total variation-based active contour to identify the region of interest from the CXR [23]. Multiple masks were created for each CXR to represent the two lung fields (i.e., left lung and right lung) and four lung quadrants (e.g., upper-left lung and lower-left lung) to ensure that the extracted features meet one of the clinical criteria for diagnosis of ARDS (bilateral opacities on CXR).

Four distinct feature sets were used to train machine learning models for the detection of ARDS from chest x-ray scans: Directionality Measure, first-order statistics from the histogram, information from a gray-level occurrence matrix, and deep learning features extracted with a

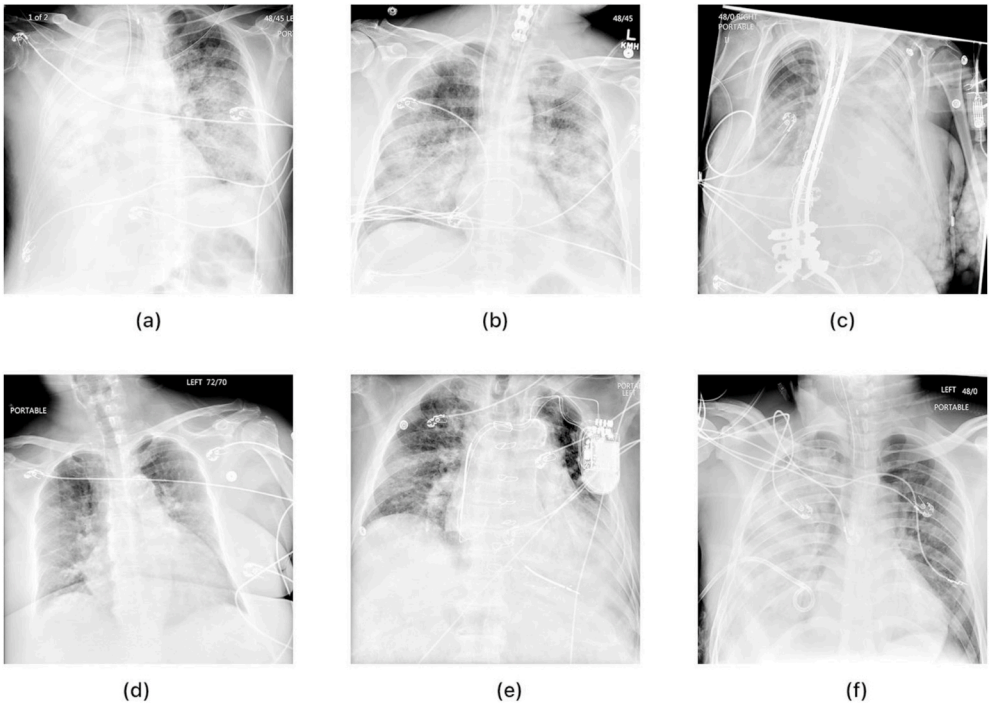

Fig. 1. Examples of chest x-rays from the Michigan Medicine dataset that were annotated as consistent with ARDS or inconsistent with ARDS based on the reviews of multiple clinical experts. Chest x-rays (a), (b), (c) demonstrate the findings of ARDS, which are bilateral airspace disease not explained by effusions, lobar/lung collapse, or nodules. These findings may (b) or may not (a, c) be uniform across both lung fields. Chest x-rays (d), (e), (f) do not demonstrate clear findings of ARDS, either because the lung fields lack clear airspace disease (d) or the disease that is present is unilateral (e, f).

Table 1
Data available from the Michigan Medicine ARDS CXR dataset.

	Patients	Chest X-Rays
Positive	151	909
Negative	349	2169
Total	500	3078

Table 2
Cohort demographic for the Michigan Medicine ARDS CXR dataset.

	n	Age	ARDS	Non-ARDS
Male	309	57.16 ± 16.72	89	220
Female	191	58.46 ± 15.71	62	129
Total	500	57.65 ± 16.32	151	349

pre-trained neural network. Based on relevant works in analytical morphomics, we perform a similar approach of normalization with structural physiology – using a ratio of chest width to sternum width - for selected features [24].

3.2. Feature extraction

3.2.1. Directionality Measure

Lung infiltrates present with a "cloud-like" appearance on CXRs. We propose Directionality Measure, a novel method to capture the intuitive notion of cloudiness as a mathematical concept. For this task, we first exclude normal findings within the CXR – such as ribs, vasculature, and medical equipment (e.g., tubes, cables, prosthetic devices). These artifacts typically have features of directionality, whereas the "clouds" corresponding to lung opacities are non-directional. Therefore, a cloudy lung region without artifacts can be described as follows: a) the average gray value will be above a certain threshold, b) the gray level varies within the region, and c) the gray level is non-directional.

Suppose that $T : [0,a] \times [0,b] \rightarrow [0,1]$ is a grayscale CXR. Consider a window $W = [u-\varepsilon, u+\varepsilon] \times [v-\varepsilon,\ v+\varepsilon]$ of size $2\varepsilon \times 2\varepsilon$ about a point $[u,v]$ in the lung region. Let T_x and T_y be the partial derivatives of T with respect to coordinates x and y. Define

$$A_{xx} = \int_W T_x^2 dx\, dy\ ,\ A_{xy} = \int_W T_x T_y dx\, dy\ ,\ A_{yy} = \int_W T_y^2 dx\, dy$$

then the function $G : [0,a] \times [0,b] \rightarrow [0,1]$ defined as $G(u,v) = \sqrt{A_{xx} + A_{yy}}$ measures the variation in a region. We normalize G such that it has values in [0,1]. The matrix A, defined as

$$A = \begin{bmatrix} A_{xx} & A_{xy} \\ A_{xy} & A_{yy} \end{bmatrix}$$

is non-negative definite. Let $\lambda_1 < \lambda_2$ be the two eigenvalues of A, and define $H(u,v) = \frac{\lambda_1}{\lambda_2}$. Note that H has values between 0 and 1.

Our preliminary analysis and observation have shown that the product $G{\cdot}H$ may already be a suitable indicator for recognition of ARDS. The functions T, G, H, and $G{\cdot}H$ applied to CXRs for a patient diagnosed with ARDS and for a non-ARDS patient are shown in Fig. 3a and Fig. 3b respectively. In both figures, the original CXR is shown in the upper left, G in the lower left, H in the upper right, and $G{\cdot}H$ in the lower right.

First-order statistics and additional measurements, further described in §3.2.2, are extracted from the product $G{\cdot}H$ as features to be used in training machine learning algorithms for detection of ARDS. In total, 72 features were extracted from Directionality Measure.

3.2.2. Histogram features

First-order statistics (mean, max, variance, kurtosis, and skewness) are calculated from the CXR histogram to capture textural properties of the lung fields. Previously published literature has demonstrated that such features extracted from CXRs exhibit significant differences between healthy and injured lung fields [10]. Specifically, higher variance and lower mean values in intensity have been observed in areas with pulmonary opacities when compared to normal, healthy lungs [25]. This evidence suggests that these first-order statistics correspond to the magnitude and the coarseness (or fineness) of the infiltrate [26].

Additional measurements of the histogram are also used as features to capture the density of pulmonary infiltrates by examining local

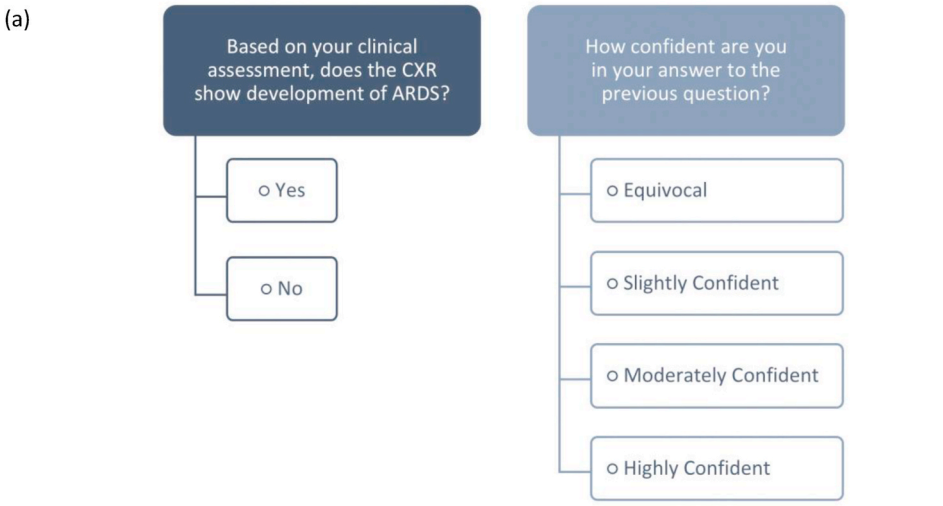

Fig. 2. Label generation of chest x-ray scans. (a) Multiple expert clinicians were asked to independently review patients' CXR and determine if any individuals had ARDS. Clinicians were also rated their confidence of the diagnosis as equivocal, slightly confident, moderately confident, or highly confident. (b) The diagnosis and confidence were converted to a scale between 1 and 8. The final label was generated from aggregating these reviews to ensure correctness and consistency of the diagnosis. A label of −1 (non-ARDS) was assigned if the averaged review was below or equal to 4.5, and a label of 1 (ARDS) was assigned if the averaged review was above 4.5.

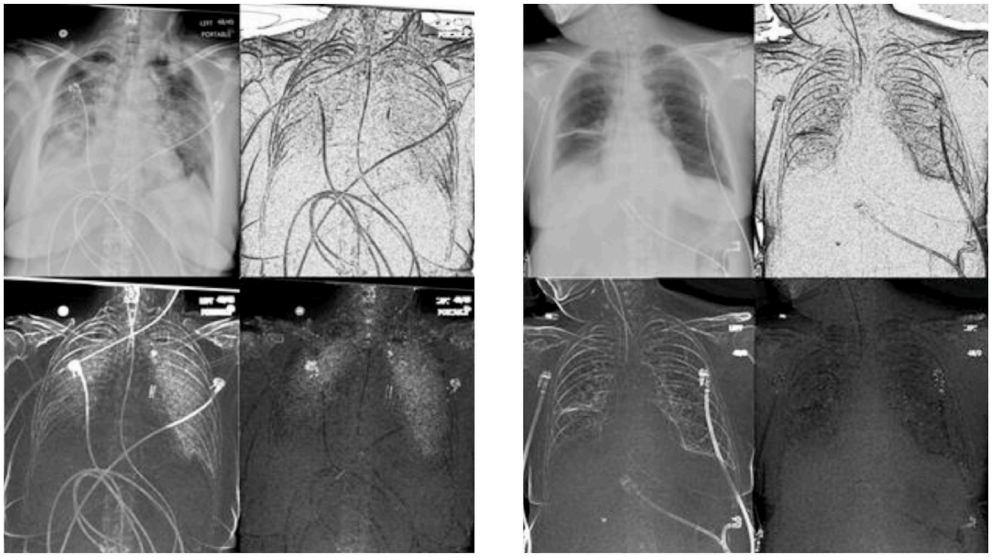

Fig. 3. Directionality Measure applied to CXRs (a) from a patient diagnosed with ARDS and (b) from a non-ARDS diagnosis. The/original CXR is shown in the upper left, *G* in the lower left, *H* in the upper right, and *G·H* in the lower right for both figures.

grayscale distribution. These features include standard deviation of the 5 largest local maxima (peaks), width of the largest peak at half-prominence, gray-level value at the first and second largest peaks, median of the maxima distribution and the frequency of that value, and area under the histogram. Features were separately extracted from the lung fields and lung quadrants. In total, 72 features were extracted from the histogram.

3.2.3. Gray-level Co-Occurrence matrix

The gray-level co-occurrence matrix (GLCM) is a statistical method used to characterize the texture of an image with respect to spatial relationships at the pixel level [27]. The GLCM is defined as a two-dimensional matrix of joint probabilities between pairs of pixels of co-occurring values at specified offsets, which is used to compute second-order statistics [28,29]. Multiple offsets and angles can be defined to increase the sensitivity of capturing pixel relationships of varying direction and distance [12]. Statistics extracted from the GLCM have demonstrated promising results when used as features to train machine learning models for detection of various lung diseases [11,12,30]. The second-order statistics extracted from the GLCMs in this study are contrast, correlation, energy, and homogeneity.

Contrast measures local variation present in an image and returns a

measure of the intensity difference between a pixel and its neighbor over the entire image. For example, a high value of this feature may indicate the presence of edges, noise, etc. This property has been demonstrated to be higher in abnormal presentations within chest radiographs as compared to normal findings [31]. Contrast is defined as

$$\sum_{i,\,j} p_{i,j}\left|i-j\right|^2$$

where i and j represents the x and y coordinates of the GLCM and $p_{i,j}$ is the element i,j of the GLCM.

Correlation measures the linear dependence (joint probability) of pixel pairs and can be interpreted as quantifying the consistency of image textures. A high correlation value indicates the predictability of pixel relationships. We expect that capturing these characteristics of a CXR may be useful as features for this study. Correlation is defined as

$$\sum_{i,j} \frac{p_{i,j}(i-\mu_i)(i-\mu_j)}{\sigma_i \sigma_j}.$$

Energy, also referred to as the angular second moment, measures the uniformity of grayscale distribution of the image. Images with a smaller number of gray levels (e.g., when it is considered very uniform in representation) have larger values of energy. Therefore, we expect this measurement to be lower for abnormal findings and useful for distinguishing between a CXR with an ARDS and one from a non-ARDS diagnosis. Energy is defined as:

$$\sum_{i,j} \left(p_{i,j}\right)^2$$

Homogeneity measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal and can be interpreted as a representation of the scale of local changes in image texture. High values of homogeneity indicate the absence of intra-regional changes and locally homogenous distribution of image textures. Homogeneity is defined as

$$\sum_{i,j} \frac{p_{i,j}}{1+(i-j)^2}.$$

Features were separately extracted from the lung fields and lung quadrants for each of the described GLCM properties. GLCMs are generated with multiple angles (0°, 45°, 90° and 135°) at a pixel distance of 1. A single, "invariant" spatial direction is generated by taking an average of the four directions so that the texture features will not be influenced by the angle of rotation. A total of 24 features were extracted from the GLCMs.

3.2.4. Features from pretrained neural networks

A pretrained ResNet-50 deep learning model is used as to extract features from chest x-ray scans for detection of ARDS. ResNet-50 is a deep convolutional neural network consisting of 50 layers with skip connections to facilitate training deep networks, specifically for optimizing trainable parameters during backpropagation to mitigate the problem of vanishing gradients [32]. Residual networks such as ResNet50 are composed of multiple building blocks with shortcut connections that skip convolutional layers via identity mapping. Each block is composed of 3 convolutional layers that perform downsampling with a stride of 2, followed by batch normalization and rectified linear unit (ReLU) activation. The architecture of ResNet-50 ends with a global average pooling (GAP) layer and a 1000 fully connected layer with softmax activation. The network was trained on over a million images from the ImageNet database [33] to learn rich feature representations and is capable of classifying into 1000 object categories.

Based on previously published work on transfer learning, we propagate CXRs (resized to 244 × 244) through the pretrained network. These layers can be reinterpreted as learned feature extraction layers [34] and activations from the GAP layer prior to the fully connected layer are extracted as feature vectors that can be used to train machine learning models to solve the classification task of this study. In total, 2048 features were extracted from the GAP layer of ResNet-50.

3.3. Machine learning & model validation

The extracted features were used to train multiple machine learning models for the detection of ARDS. A soft-margin support vector machine (SVM) with a linear kernel was used in this study and utilizes Bayesian optimization for tuning the C parameter (to adjust the penalty of misclassification).

We also implement a random forest and additional gradient boosting methods such as adaptive boosting (AdaBoost), random under-sampling boosting (RUSBoost), robust boost, and total boost in this work. These techniques also use Bayesian optimization for hyperparameter tuning to determine the optimal n_{trees}, learning rate, and number of learning cycles. Gradient boosting has often been compared to random forest given the number of similarities between the two techniques. While random forest is known to be more robust to noise and easier to train, boosting techniques maintain a reputation of being more resistant to overfitting, with benchmark results having shown that booting produces better learners than random forests.

Data from 70% of patients (approximately 2018 CXRs) were used for model training and performance evaluated using 5-fold cross-validation ong this training set. The final reported results are from the remaining 30% of patients (1060 CXRs) that were used as the hold-out test set.

4. Results

Performance metrics (accuracy, Area under the Curve (AUC), and F1 score) on the Michigan Medicine dataset for all four feature sets are reported in Table 3. The best performance achieved using an individual feature set was attained with an AdaBoost classifier trained on features derived from Directionality Measure (0.78 accuracy and 0.74 AUC), followed by a RUSBoost classifier also trained on the same feature set (0.77 accuracy and 0.74 AUC). These results also indicate that training on multiple feature sets can yield further improvements. The best overall performance was achieved with AdaBoost trained on all available features from Directionality Measure, first-order statistics in the histogram, GLCM, and deep learning (0.83 accuracy and 0.79 AUC). The second-best performing model when trained with all features is RUSBoost (0.81 accuracy and 0.77 AUC.) Results for additional permutations of combined features (e.g., Directionality Measure and histogram features) are provided in Table 4 and in the supplementary materials (Table S1, S2, S3).

A total of 2216 features were used to generate these results: 72 features from Directionality Measure, 72 features from the histogram, 24 features from the GLCM, and 2048 features from deep learning. To reduce the dimensionality of utilized features, we implemented feature selection with PCA, minimum redundancy maximum relevance (mRMR), and chi-squared test. The first three principal components (98% of variance explained) were used for the PCA approach. The top 100 most important predictors were selected from feature selection with mRMR and chi-squared test. We did not test beyond using the top 100 ranked features since anything beyond that yielded an insubstantial predictor importance score. The results of this analysis are provided in Table 5.

5. Discussion

In this study, we propose and describe Directionality Measure, a novel feature engineering technique used to capture the "cloud-like" appearance of diffuse alveolar damage as a mathematical concept. This work also examines the effectiveness of using a pretrained deep neural network via transfer learning as a feature extractor in addition to standard features extracted from the histogram and GLCM.

Many published algorithms for the detection of various lung

Table 3
Performance metrics (accuracy, AUC, and F1 score) for detection of ARDS using features from Directionality Measure, the histogram, GLCM, deep learning, and a combination of these feature sets.

	Directionality Measure			Histogram			GLCM			Deep Learning (ResNet-50)			All Features Combined		
	Accuracy	AUC	F1	Accuracy	AUC	F1	Accuracy	AUC	F1	Accuracy	AUC	F1	Accuracy	AUC	F1
SVM	0.73	0.66	0.58	0.70	0.62	0.57	0.72	0.65	0.59	0.73	0.67	0.64	0.74	0.73	0.64
Random Forest	0.75	0.66	0.56	0.73	0.64	0.54	0.74	0.65	0.54	0.75	0.65	0.59	0.76	0.71	0.63
AdaBoost	**0.78**	**0.74**	**0.61**	**0.75**	**0.73**	**0.58**	**0.76**	**0.73**	**0.59**	**0.77**	**0.73**	**0.62**	**0.83**	**0.79**	**0.65**
RUSBoost	**0.77**	**0.74**	**0.65**	**0.74**	**0.71**	**0.64**	**0.77**	**0.72**	**0.66**	**0.77**	**0.72**	**0.63**	**0.81**	**0.77**	**0.67**
Robust Boost	0.73	0.69	0.58	0.70	0.66	0.55	0.71	0.68	0.57	0.75	0.70	0.56	0.76	0.73	0.63
Total Boost	0.70	0.68	0.55	0.69	0.64	0.54	0.70	0.67	0.55	0.73	0.70	0.53	0.73	0.68	0.59

Table 4
Additional performance metrics for classification results when using Directionality Measure with additional features.

	Directionality Measure + Histogram			Directionality Measure + GLCM			*Directionality Measure + Deep Learnings (ResNet-50)*			All Features Combined		
	Accuracy	AUC	F1	Accuracy	AUC	F1	Accuracy	AUC	F1	Accuracy	AUC	F1
SVM	0.73	0.68	0.59	0.74	0.69	0.62	0.75	0.70	0.64	0.74	0.73	0.64
Random Forest	0.75	0.66	0.59	0.76	0.68	0.59	0.76	0.67	0.62	0.76	0.71	0.63
AdaBoost	**0.79**	**0.75**	**0.62**	**0.81**	**0.77**	**0.63**	**0.80**	**0.77**	**0.65**	**0.83**	**0.79**	**0.65**
RUSBoost	**0.77**	**0.73**	**0.62**	**0.79**	**0.76**	**0.66**	**0.80**	**0.76**	**0.65**	**0.81**	**0.77**	**0.67**
Robust Boost	0.71	0.70	0.60	0.70	0.70	0.60	0.72	0.69	0.60	0.76	0.73	0.63
Total Boost	0.70	0.69	0.58	0.71	0.68	0.58	0.73	0.68	0.59	0.73	0.68	0.59

Table 5
Comparison of feature selection methods used to reduce dimensionality of available features.

	All Features			PCA (98% Variance Explained)			mRMR (100)			Chi-Squared Test (100)		
	Accuracy	*AUC*	*F1*	*Accuracy*	*AUC*	*F1*	*Accuracy*	*AUC*	*F1*	*Accuracy*	*AUC*	*F1*
SVM	0.74	0.73	0.64	0.73	0.67	0.62	0.70	0.66	0.61	0.79	0.74	0.59
Random Forest	0.76	0.71	0.63	0.74	0.65	0.53	0.75	0.67	0.62	0.79	0.73	0.61
AdaBoost	**0.83**	**0.79**	**0.65**	**0.79**	**0.71**	**0.63**	**0.80**	**0.73**	**0.60**	**0.81**	**0.74**	**0.64**
RUSBoost	**0.81**	**0.77**	**0.67**	**0.76**	**0.68**	**0.58**	**0.76**	**0.72**	**0.62**	**0.80**	**0.76**	**0.65**
Robust Boost	0.76	0.73	0.63	0.72	0.67	0.58	0.76	0.68	0.58	0.75	0.70	0.60
Total Boost	0.73	0.68	0.59	0.72	0.66	0.55	0.76	0.67	0.56	0.75	0.70	0.60

pathologies from chest radiology exists [10–13,25,26]. However, we did not find any studies that particularly focused on acute respiratory distress syndrome. Some of these conditions share similarities in pathology and clinical presentation, but it was unknown whether the features used to detect a particular condition (e.g., sepsis) would also be effective for the detection of ARDS. Our results show that some of these features do in fact work, e.g., first-order statistics from the histogram and GLCM. Furthermore, we demonstrate that the proposed Directionality Measure technique is capable of detecting ARDS and outperforms other techniques that have been used for similar applications. We report that the best overall performance is obtained when the machine learning models are trained with all four features sets combined rather than only having access to each individual feature separately.

We also conducted extensive tests with several feature selection methods, including principal component analysis (PCA), minimum reduction maximum relevance (mRMR), and chi-squared test. The outcome of those experiments demonstrated that better classification results were obtained when using all available features compared to only using the most important features selected by these techniques. These results were surprising at first, since we expected redundancy in the feature space, especially from the 2048 features from deep learning. However, after a more comprehensive analysis the data, we concluded that this is reasonable because many of the tree-based and boosting methods already include feature selection in their implementation. Therefore, models that don't already include this process (e.g., SVM) will benefit the most from feature selection – which is exactly what we observe in Table 5 with SVM when using all available compared to only using the top ranked features from the chi-squared test.

Although ResNet-50 was used to extract the deep learning features, a number of other pretrained deep neural networks were also considered – including ResNet-18, ResNet-101, Inception-v3, U-Net, and VGG19. These networks were primarily chosen based on their publication record and capability in using arbitrary layers for feature extraction. Preliminary results showed that features extracted with ResNet18, ResNet101, and Inception-v3 did not perform as well as ResNet50. As the architectures for U-Net and VGG19 do not contain a GAP layer features would be extracted from the max pooling or convolutional layers. The activations from VGG19's max pooling layer have a dimensionality of $7 \times 7 \times 512$, while the activations from U-Net's ReLU layer are of size $256 \times 256 \times 32$, resulting in almost 2 million features when flattened. We did try multiple tensor decomposition methods to work with this high-dimensional data, including higher-order singular value decomposition (HOSVD), but did not achieve satisfactory performance.

Intuitively, one could argue that the learned weights from the deeper layers should be more specific to the images of the training dataset and the task it was initially trained for. However, a number of publications have reported promising results with features derived from the GAP layer. Furthermore, our internal testing showed comparable results between extracting features from the GAP layer and a shallower layer. Ultimately, we decided to use ResNet-50's GAP layer – which yielded 2048 features.

We recognize that there are several limitations to this study. With more data, it would be worthwhile to investigate training an end-to-end deep learning model directly from CXRs, comparing the effectiveness of this approach to features extracted with a pretrained deep neural network via transfer learning. Another limitation to note is that our dataset is only labeled for binary classification of ARDS or non-ARDS, even though the patients in the non-ARDS cohort still exhibit a degree of respiratory failure. In future work, we would like to examine the

feasibility of multi-label classification to further improve accuracy in diagnosis. We also plan to include additional data, such as signals from physiological waveforms and data from electronic health records, for the detection of ARDS.

6. Conclusion

This study demonstrates the capability of multiple feature extraction methods that can be used to train machine learning algorithms for the detection of ARDS from chest x-ray scans. Our work introduces Directionality Measure, a novel feature engineering technique used to capture the "cloud-like" appearance of diffuse alveolar damage as a mathematical concept. We also examine the effectiveness of transfer learning with pre-trained networks for feature extraction in this context and standard features extracted from the histogram and GLCM. Our results demonstrate that Directionality Measure is capable of detecting ARDS from CXRs and outperforms other techniques that have been used for similar applications. We also report that the best overall performance is obtained when the machine learning models are trained with multiple features sets combined rather than training on each individual feature separately.

We believe this paper makes a significant contribution towards the evaluation of acute respiratory distress syndrome with machine learning and artificial intelligence to improve patient outcomes. In the future, we plan to conduct a larger evaluation and include more respiratory conditions (e.g., pneumonia and chronic obstructive pulmonary disease) by obtaining additional data from Michigan Medicine and other hospital systems.

Declaration of competing interest

An invention disclosure has been filed for the Directionality Measure technology described in this paper.

Acknowledgements

This material is based upon work supported in part by the National Science Foundation under Grant 1722801 and in part by the National Institutes of Health under Grant NHLBI K01HL136687.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2021.104463.

References

[1] G.D. Rubenfeld, E. Caldwell, E. Peabody, J. Weaver, D.P. Martin, M. Neff, E. J. Stern, L.D. Hudson, Incidence and outcomes of acute lung injury, N. Engl. J. Med. 353 (16) (2005 Oct 20) 1685–1693.
[2] G. Bellani, J.G. Laffey, T. Pham, E. Fan, L. Brochard, A. Esteban, L. Gattinoni, F. Van Haren, A. Larsson, D.F. McAuley, M. Ranieri, Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries, Jama 315 (8) (2016 Feb 23) 788–800.
[3] S.R. Desai, Acute respiratory distress syndrome: imaging of the injured lung, Clin. Radiol. 57 (1) (2002 Jan 1) 8–17.
[4] X. Li, X. Ma, Acute respiratory failure in COVID-19: is it "typical" ARDS? Crit. Care 24 (2020 Dec) 1–5.
[5] L.B. Ware, M.A. Matthay, The acute respiratory distress syndrome, N. Engl. J. Med. 342 (18) (2000 May 4) 1334–1349.
[6] ARDS Definition Task Force, V.M. Ranieri, G.D. Rubenfeld, B.T. Thompson, N. D. Ferguson, E. Caldwell, Acute respiratory distress syndrome, Jama 307 (23) (2012 Jun 20) 2526–2533.
[7] S. Sheard, P. Rao, A. Devaraj, Imaging of acute respiratory distress syndrome, Respir. Care 57 (4) (2012 Apr 1) 607–612.
[8] J.M. Peng, C.Y. Qian, X.Y. Yu, M.Y. Zhao, S.S. Li, X.C. Ma, Y. Kang, F.C. Zhou, Z. Y. He, T.H. Qin, Y.J. Yin, Does training improve diagnostic accuracy and inter-rater agreement in applying the Berlin radiographic definition of acute respiratory distress syndrome? A multicenter prospective study, Crit. Care 21 (1) (2017 Dec) 1–8.
[9] M.W. Sjoding, T.P. Hofer, I. Co, A. Courey, C.R. Cooke, T.J. Iwashyna, Interobserver reliability of the Berlin ARDS definition and strategies to improve the reliability of ARDS diagnosis, Chest 153 (2) (2018 Feb 1) 361–367.
[10] J.H. Tan, U.R. Acharya, C. Tan, K.T. Abraham, C.M. Lim, Computer-assisted diagnosis of tuberculosis: a first order statistical approach to chest radiograph, J. Med. Syst. 36 (5) (2012 Oct 1) 2751–2759.
[11] T.A. Akmal, J.C. Than, H. Abdullah, N.M. Noor, Chest X-ray image classification on common thorax diseases using GLCM and AlexNet deep features, International Journal of Integrated Engineering 11 (4) (2019 Sep 5).
[12] N.S. Lingayat, M.R. Tarambale, A computer based feature extraction of lung Nodule in chest x-ray image, International Journal of Bioscience, Biochemistry and Bioinformatics 3 (6) (2013 Nov 1) 624.
[13] S.A. Patil, Texture analysis of TB X-ray images using image processing techniques, J. Biomed. Bioeng. 3 (1) (2012 Mar) 53–56.
[14] A.S. Razavian, H. Azizpour, J. Sullivan, S. Carlsson, CNN features off-the-shelf: an astounding baseline for recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014, pp. 806–813.
[15] H. Lee, R. Grosse, R. Ranganath, A.Y. Ng, Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, in: Proceedings of the 26th Annual International Conference on Machine Learning, 2009 Jun 14, pp. 609–616.
[16] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks? InAdvances in neural information processing systems (2014) 3320–3328.
[17] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2009 Oct 16) 1345–1359.
[18] J. Yue-Hei Ng, F. Yang, L.S. Davis, Exploiting local features from deep networks for image retrieval, InProceedings of the IEEE conference on computer vision and pattern recognition workshops (2015) 53–61.
[19] A.S. Razavian, J. Sullivan, S. Carlsson, A. Maki, Visual instance retrieval with deep convolutional networks, ITE Transactions on Media Technology and Applications 4 (3) (2016) 251–258.
[20] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, M. Nielsen, Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Berlin, Heidelberg, 2013 Sep 22, pp. 246–253.
[21] D.C. Cireşan, A. Giusti, L.M. Gambardella, J. Schmidhuber, Mitosis detection in breast cancer histology images with deep neural networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Berlin, Heidelberg, 2013 Sep 22, pp. 411–418.
[22] Y. Bar, I. Diamant, L. Wolf, H. Greenspan, Deep learning with non-medical training used for chest pathology identification, in: Medical Imaging 2015: Computer-Aided Diagnosis, vol. 9414, International Society for Optics and Photonics, 2015 Mar 20, p. 94140V.
[23] N. Reamaroon, M.W. Sjoding, H. Derksen, E. Sabeti, J. Gryak, R.P. Barbaro, B. D. Athey, K. Najarian, Robust segmentation of lung in chest x-ray: applications in analysis of acute respiratory distress syndrome, BMC Med. Imag. 20 (1) (2020 Dec) 1–3.
[24] K. Chughtai, Y. Song, P. Zhang, B. Derstine, E. Gatza, J. Friedman, L. Hully, C. Inglis, S. Goldstein, J. Magenau, A. Pawarode, Analytic morphomics: a novel CT imaging approach to quantify adipose tissue and muscle composition in allogeneic hematopoietic cell transplantation, Bone Marrow Transplant. 51 (3) (2016 Mar) 446–450.
[25] S. Katsuragawa, K. Doi, Computer-aided diagnosis in chest radiography, Comput. Med. Imag. Graph. 31 (4–5) (2007 Jun 1) 212–223.
[26] L. Monnier-Cholley, H. MacMahon, S. Katsuragawa, J. Morishita, T. Ishida, K. Doi, Computer-aided diagnosis for detection of interstitial opacities on chest radiographs, AJR. American journal of roentgenology. 171 (6) (1998 Dec) 1651–1656.
[27] R.M. Haralick, K. Shanmugam, I.H. Dinstein, Textural features for image classification, IEEE Transactions on systems, man, and cybernetics (6) (1973 Nov) 610–621.
[28] A. Zotin, Y. Hamad, K. Simonov, M. Kurako, Lung boundary detection for chest X-ray images classification based on GLCM and probabilistic neural networks, Procedia Computer Science 159 (2019 Jan 1) 1439–1448.
[29] K.S. Vidya, E.Y. Ng, U.R. Acharya, S.M. Chou, R. San Tan, D.N. Ghista, Computer-aided diagnosis of myocardial infarction using ultrasound images with DWT, GLCM and HOS methods: a comparative study, Comput. Biol. Med. 62 (2015 Jul 1) 86–93.
[30] S.A. Patil, V.R. Udupi, Geometrical and texture features estimation of lung cancer and TB images using chest X-ray database, Int. J. Biomed. Eng. Technol. 6 (1) (2011 Jan 1) 58–75.
[31] Q. Zhao, C.Z. Shi, L.P. Luo, Role of the texture features of images in the diagnosis of solitary pulmonary nodules in different sizes, Chin. J. Canc. Res. 26 (4) (2014 Aug) 451.
[32] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, InProceedings of the IEEE conference on computer vision and pattern recognition (2016) 770–778.
[33] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, L. Fei-Fei, Imagenet: a large-scale hierarchical image database, in: In2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009 Jun 20, pp. 248–255.
[34] E. Rezende, G. Ruppert, T. Carvalho, F. Ramos, P. De Geus, Malicious software classification using transfer learning of resnet-50 deep neural network, in: In2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2017 Dec 18, pp. 1011–1014.