Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

International Journal of Medical Informatics 158 (2022) 104662

Contents lists available at ScienceDirect

International Journal of Medical Informatics

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

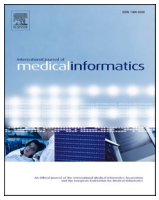


Image and structured data analysis for prognostication of health outcomes in patients presenting to the ED during the COVID-19 pandemic

Liam Butler [a], Ibrahim Karabayir [b,c,d], Mohammad Samie Tootooni [d], Majid Afshar [e], Ari Goldberg [d], Oguz Akbilgic [b,d,*]

[a] Stony Brook University, Stony Brook, NY 11794, USA
[b] Wake Forest School of Medicine, Winston-Salem, NC 27157, USA
[c] Kirklareli University, Turkey
[d] Loyola University Chicago, Maywood, IL 60153, USA
[e] University of Wisconsin-Madison, WI 53792, USA

ARTICLE INFO

Keywords:
COVID-19
ARDS
ICU
Mortality
Machine/deep learning
Chest radiographs

ABSTRACT

Background: Patients admitted to the emergency department (ED) with COVID-19 symptoms are routinely required to have chest radiographs and computed tomography (CT) scans. COVID-19 infection has been directly related to the development of acute respiratory distress syndrome (ARDS) and severe infections could lead to admission to intensive care and increased risk of death. The use of clinical data in machine learning models available at time of admission to ED can be used to assess possible risk of ARDS, the need for intensive care (admission to the Intensive Care Unit; ICU) as well as risk of mortality. In addition, chest radiographs can be inputted into a deep learning model to further assess these risks.

Purpose: This research aimed to develop machine and deep learning models using both structured clinical data and image data from the electronic health record (EHR) to predict adverse outcomes following ED admission.

Materials and Methods: Light Gradient Boosting Machine (LightGBM) was used as the main machine learning algorithm using all clinical data including 42 variables. Compact models were also developed using the 15 most important variables to increase applicability of the models in clinical settings. To predict risk (or early stratified risk) of the aforementioned health outcome events, transfer learning from the CheXNet model was also implemented on the available data. This research utilized clinical data and chest radiographs of 3,571 patients, 18 years and older, admitted to the emergency department between 9th March 2020 and 29th October 2020 at Loyola University Medical Center.

Main Findings: The research results show that we can detect COVID-19 infection (AUC = 0.790 (0.746–0.835)), predict the risk of developing ARDS (AUC = 0.781 (0.690–0.872), risk stratification of the need for ICU admission (AUC = 0.675 (0.620–0.713)) and mortality (AUC = 0.759 (0.678–0.840)) at moderate accuracy from both chest X-ray images and clinical data.

Principal Conclusions: The results can help in clinical decision making, especially when addressing ARDS and mortality, during the assessment of patients admitted to the ED with or without COVID-19 symptoms.

1. Introduction

The novel SARS-CoV-2 (COVID-19) virus has quickly spread globally and was classified as a world pandemic by the World Health Organization (WHO) on 11th of March 2020 [6]. There are currently more than 2 million people infected with this virus. It is reported that this virus has high probability of causing severe acute respiratory distress syndrome (ARDS) and therefore requires early identification and treatment [6]. Since the outbreak of the COVID-19 pandemic, almost 124 million people have been infected and more than 2.73 million people have died from COVID-19 infection [24]. A substantial number of infected individuals arrive at the emergency department (ED) with hypoxic respiratory failure from COVID-19, with greater prevalence among individuals 65 years of age and older [3,5].

Testing for the COVID-19 virus has evolved with multiple assays achieving high sensitivity and greater distribution of testing, including

* Corresponding author at: Wake Forest School of Medicine, Section on Cardiovascular Medicine, 1 Medical Center Blvd, Winston-Salem, NC 27157, USA.
E-mail address: oakbilgi@wakehealth.edu (O. Akbilgic).

https://doi.org/10.1016/j.ijmedinf.2021.104662
Received 9 July 2021; Received in revised form 23 November 2021; Accepted 6 December 2021
Available online 9 December 2021

general symptom assessment and saliva. As part of routine care for patients presenting to the emergency department with COVID-19 symptoms, assessment include chest imaging with radiographs and computed tomography (CT) scans. Image-based deep learning models can rapidly identify outcomes of the infection in different organs such as lungs, which can be a substantial aid to clinicians for time-sensitive diagnosis [4]. COVID-19 infections have been linked to the development of acute respiratory distress syndrome (ARDS) which results in decreased lung space and loss of lung tissue aeration [19]. The use of chest radiographs can be inputted into a deep learning model [12] to predict the probability of infection in each patient as well as facilitate triage to the intensive care unit (ICU) and also signal the risk for mortality.

This research aimed to develop machine and deep learning models using both structured clinical data and image data from the electronic health record (EHR) to distinguish the following events: (1) COVID-19 infections; (2) acute respiratory distress syndrome in patients with and without COVID-19; (3) need for ICU admission and (4) in-hospital death. Further deep learning models to determine probability of (2), (3) and (4) were also developed using image analysis from chest radiographs. An additional aim of this research was to develop compact models, with similar accuracy to the 'full' models, using a smaller number of variables to increase the generalizability of model implementation in clinical settings.

2. Materials & methods

2.1. Cohort

The cohort used in this study included patients admitted to the emergency department at Loyola University Medical Centre, Maywood, Illinois, USA between 9th March 2020 and 29th October 2020. A total of 3,571 patients were included in the study. The inclusion criteria stipulated that all adults over 18 years of age were encountered in the emergency department, were tested for COVID-19 and received chest radiographic imaging. This research was approved by an Institutional Review Board (IRB).

2.2. Outcomes

The main outcome of this research was the risk prediction of i) COVID-19 infection, ii) acute respiratory distress syndome (ARDS), iii) requirement for ICU admission and iv) risk of mortality at time of emergency department (ED) admission. COVID-19 infection was identified and confirmed from polymerase chain reaction (PCR) tests. In cases where a rapid test was previously utilized by patients, a PCR was still performed by the hospital and it is this COVID-19 diagnosis from the PCR testing that was used in this research. ARDS, following the Berlin definition, is an acute diffuse lung injury that can increase pulmonary vascular permeability. ARDS can be identified through chest radiographs due to the development of hypxemia and bilateral radiographic opacities [1,13,19]. The algorithm included vital signs, laboratory data, ventilator data, and keywords from the chest imaging that met the requirement for hypoxemia and bilateral infiltrates without a primary cardiogenic etiology for edema. The algorithm was updated from the original studies to include CT reports, minimum peak end-expiratory pressure of 5 cm H_2O on the ventilator and meeting the criteria within seven days of hospital presentation for primary ARDS. To avoid discrepancies in timing of ICU admission between patients, we standardized categorization for ICU admission to include patients that had been admitted directly to ICU or transferred to ICU at any time following admission to ED. Finally, mortality was classed as any in-hospital death.

2.3. Risk factors

A total of 42 structured data variables from the EHR were included as risk factors in the first machine learning models. The cohort data included demographics such as age, gender, race and ethnicity, and initial measurement of oxygen saturation levels. The data also included comorbidities that were present on admission, including but not limited to heart disease (cardiovascular disease, congestive heart failure etc.), lung disease (chronic obstructive pulmonary disease (COPD), pulmonary circulation disorders etc.), liver and renal diseases as well as abusive drug or alcohol intake. This data was obtained using prespecified ICD10 codes. A statistical summary of the variables used in this reseach is provided in Table S1 in Supplementary Materials. The dataset was comprehensive and only had minimal amounts of missing data with 30 patients having no recorded ethnicity and 2 patients having no recorded first oxygen levels. Due to the low amount of missing data, these two variables were imputed using the most frequent category or the median of the cohort.

2.4. Chest imaging data

This research also used chest radiographs, imported in DICOM format. One frontal radiograph for each patient was exported in *.jpeg* format. The dataset included 3571 images, with 1 chest radiograph per patient. Pre-trained deep learning models on X-ray and CT scans already exist (e.g. CheXNet) [18] and can be used as a baseline to assess risk associated with ARDS, need for ICU admission and possibly risk of mortality following COVID-19 infection (American College of Chest Physicians, 2020). The chest radiographs were rescaled to the same size (between 0 and 1) as that implemented in the original CheXNet model [18]. The images were resized to 224×224 size for training the transfer learned fine-tuned CheXNet-DenseNet121 model. An image data generator was used to load the images and the associated labelling, stored in a *.csv* file, for each of the 4 outcomes. Image data was augmented during training to ensure that the model can classify images appropriately in cases when images either contain noise or are shifted or rotated as well as to reduce overfitting of the data.

2.5. Outcome prediction using clinical risk factors

We randomly split the study cohort into 80% for model building and 20% as a holdout test set. Data imbalance was present across the four outcomes assessed. Since the case-rates for COVID-19 are highly variable between hospitals, we maintained the prevalence of cases. We aimed to provide a training corpus that reflect a real world setting; therefore, we avoided case-control matching or up-sampling of cases. Hyperparamaters for a Light Gradient Boosting Machine (LGBM) were tuned using 5-fold cross validation on the 80% training set to achieve the optimal area under the receiver operating characteristics (AUC). The optimal hyperparameters were identified using the best combination from a parameter grid search. The parameters provided included ranges for the number of estimators, the number of leaves, maximum depth of the resulting tree, regularization terms and subsampling frequency, to reduce overfitting of the models. In addition to the AUC, sensitivity and specificity were also reported. Predictions were adjusted based on specific thresholds to balance specificity and sensitivity of each model. The integration of cross-validation methods have been widely used in machine learning studies and have shown that this strategy can increase model performance when compared to a 80%/20% training-holdout split with no cross-validation [26]. The trained model was tested on the 20% holdout data. All comparisons and evalutions of the models were based on the AUC statistics obtained on the holdout data. Models were compared using the DeLong Test, a non-parametric test for comparison of independent AUCs of the models [8]. For each outcome, a 'shapley additive explanations' (SHAP) variable importance analysis (Molnar, 2020) was performed on each model and the 15 top most important variables were identified as the best predictors for each of the four outcomes. SHAP is a method to explain individual predictions as well as showing the global positive and negative relationships of the predictors with the outcome (Molnar, 2020). With SHAP, global

interpretations of the model are consistent with the local explanations (for each observation), since the importance analysis is based on the combined 'Shapley' values of the global interpretations (Molnar, 2020). In this research we provided SHAP summary plots which combine the global importance of the variables and the effects of these variable on the outcome.

2.6. Outcome prediction from radiograph images using CheXNet

CheXNet is a pretrained convolutional neural network (CNN) model developed to process chest radiograph (X-Ray) images and detect 14 classes namely: Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Pleural Thickening and Hernia [18]. The CheXNet model was developed on 120,000 chest radiograph images available at "https://github.com/jrzech/reproduce-chexnet" and "https://github.com/brucechou1983/CheXNet-Keras", with the full image data set available from the open access National Institute of health (NIH) database at https://nihcc.app.box.com/v/ChestXray-NIHCC [22]. We used the existing CheXNet model on our chest radiograph images and obtained risk predictions for all 14 categories. The resulting 14 predictions were used as individual predictors for the four outcomes to identify possible CheXNet categories associated with the outcomes of interest. The highest resulting AUCs from each of the 14 classes were identified and used. This means that we further developed models that also included a select number of predicted lung disease classes, namely consolidation, infiltration and pneumonia, adopted from CheXNet predictions that are associated with higher risks, rather than using all 14 classes from CheXNet predictions. The model used pre-trained weights from ImageNet and training utilized DenseNet121 architecture, with an initial learning rate of 0.001, reducing with a factor of 10 when validation loss did not decrease after each epoch. The imaging training model used a batch size of 64.

2.7. Transfer learning to create CheXNet-Cov19

Our study then used transfer learning to obtain a CheXNet based deep learning model, namely CheXNet-Cov19, that predicts risk for the four outcomes of interest (COVID-19 infection, ARDS, ICU admission and risk of mortality), with each outcome being a probability between 0 and 1 for the associated risk. To achieve this, we first altered the CheXNet architecture by replacing the 14 nodes outcome layer with four nodes representing COVID-19 infection, ARDS development, ICU admission and mortality. We then initialized all parameters of this new artitecture from CheXNet except the last fully connected layer. We then re-trained and fine-tuned this novel CNN model on 70% of our radiograph images while using 10% of images for validation. The weights for each class for each outcome was based on total counts and class positive counts. The initial learning rate was set to 0.0001, reducing with a factor of 10 when validation loss did not decrease after each epoch. The batch size was set to 64. The final trained model was tested on the same 20% holdout dataset that was used in earlier predictive models.

2.8. Integration of chest radiographs and clinical data

We then integrated radiograph-based predicted risks to each of the final models built on clinical data. We did this by investigating combinations (or ensembles) of various risk predictions and/or risk factors using the machine learning algorithm Light Gradient Boosting Machines (LGBM). In this ensemble approach, we built final prediction models on the same 80% model building dataset (previously discussed) and evaluated the models on the same 20% holdout dataset for streamlined comparisons.

The models and related analyses were performed using the Python programming language and the associated code will be available on a github repository.

3. Results

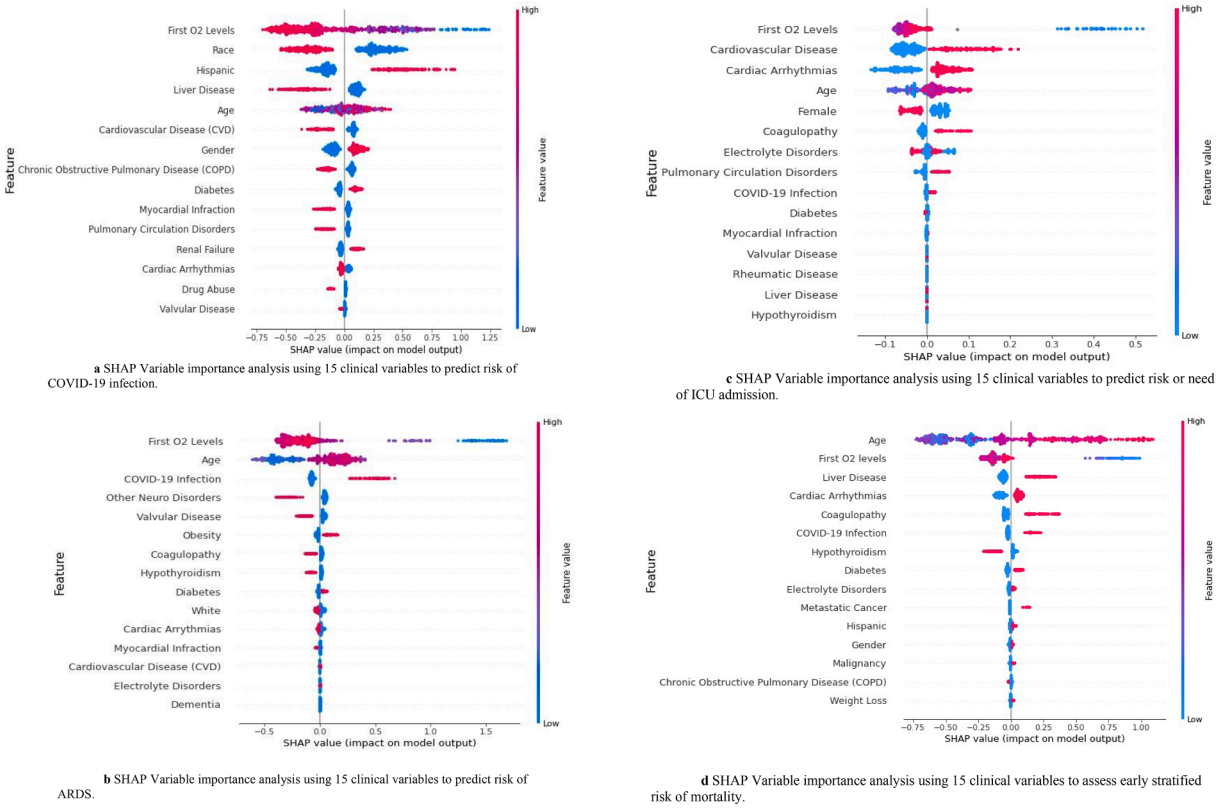
The cohort was composed of 3,571 patients of which 1,907 (53.40%) were male, 1,605 (44.95%) were of white race, and 944 (26.44%) of Hispanic ethnicity. The mean age of the cohort, with standard deviation, was 56.2S3 ± 20.54. Oxygen levels were taken when patients first arrived at ED with mean oxygen saturation of 96.27%±5.35. There were 789 patients (22.09%) diagnosed with COVID-19 infection with laboratory confirmation through assays, 260 patients (7.28%) developed ARDS and 963 patients (26.97%) were admitted to an ICU. From those admitted to an ICU, 435 patients were admitted directly to ICU with the remaining 528 patients being admitted to another ward first. 293 patients (8.20%) died in the hospital, with 212 patients dying in the ICU. In addition, from the 963 patients admitted to ICU, 245 had developed ARDS. The summarized statistics are detailed in Table 1 seperately for COVID-19 positive and negative patients. The details of all 42 risk factors are provided in Supplementary Materials Table 1.

3.1. Outcome prediction using clinical risk factors

Models built on the full clinical data set (all 42 risk factors) resulted in moderate accuracies to predict COVID-19 infection with an AUC of 0.790 (0.746–0.835); ARDS with an AUC of 0.753 (0.675–0.831); ICU admission with an AUC of 0.675 (0.620–0.713); and in-hospital death with an AUC of 0.683 (0.606–0.761). The 15 top most important predictors for all four outcomes from the SHAP variable importance analysis are provided in Table S. 2 as supplementary material. Compact models were re-built for each outcome using only these 15 clinical risk factors. The use of more compact models can allow for generalizability in situations where medical information is not as comprehensive in other clinical settings. The compact models provided prediction accuracies with an AUC of 0.775 (0.730–0.821) for COVID-19 infection, 0.721 (0.641–0.802) for ARDS, 0.658 (0.611–0.702) for ICU admission and AUC of 0.755 (0.669–0.841) for mortality. Compared to models using all 42 variables, the DeLong test showed that the compact models for need of ICU admission and mortality were significantly better ($p < 0.05$) but not significantly different for COVID-19 infection ($p = 0.762$) and ARDS ($p = 0.071$). Fig. 1(a-d) provides the variable importance analysis results for models aimed to predict risk for COVID-19 infection, ARDS, need for ICU admission and risk of mortality based on the 15 most important clinical predictors. First oxygen (O_2) levels are the most important predictors for COVID-19 infection (Fig. 1a), ARDS (Fig. 1b) and ICU admission (Fig. 1c), and second-most important predictor for risk of mortality (Fig. 1d) superceded by age. Interestingly, race and Hispanic ethnicity both feature amongst the most important contributors to the COVID-19 infection models. As expected, COVID-19 infection is a major contributor, as expected, to the development of ARDS but is less important in the risk stratification of ICU admission and mortality. Age is an important predictor for each of the four outcomes, albeit in varying degrees.

Table 1
Summary statistics comparing COVID-19 positive and COVID-19 negative patients who i) developed ARDS, ii) were admitted to ICU and iii) died.

	Total (N = 3571)	COVID positive (N = 789)	COVID negative (N = 2782)	p-value
No. patients developed ARDS, N (%)	260 (7.28)	101(12.80)	159 (5.72)	<0.001
No. patients admitted to ICU, N (%)	963 (26.97)	243 (30.80)	720 (25.88)	0.006
No patients who died, N (%)	293 (8.20)	91(11.53)	202 (7.26)	<0.001



a SHAP Variable importance analysis using 15 clinical variables to predict risk of COVID-19 infection.

b SHAP Variable importance analysis using 15 clinical variables to predict risk of ARDS.

c SHAP Variable importance analysis using 15 clinical variables to predict risk or need of ICU admission.

d SHAP Variable importance analysis using 15 clinical variables to assess early stratified risk of mortality.

Fig. 1. **a** SHAP Variable importance analysis using 15 clinical variables to predict risk of COVID-19 infection. **b** SHAP Variable importance analysis using 15 clinical variables to predict risk of ARDS. **c** SHAP Variable importance analysis using 15 clinical variables to predict risk or need of ICU admission. **d** SHAP Variable importance analysis using 15 clinical variables to assess early stratified risk of mortality.

3.2. Outcome prediction from X-Ray images using CheXNet

The AUCs to predict ARDS were 0.624 for Atelectasis, 0.547 for Cardiomegaly, 0.625 for Effusion, 0.735 for Infiltration, 0.557 for Mass, 0.540 for Nodule, 0.702 for Pneumonia, 0.696 for Pneumothorax, 0.711 for Consolidation, 0.697 for Edema, 0.499 for Emphysema, 0.536 for Fibrosis, 0.586 for Pleural Thickening and 0.471 for Hernia. The categories with highest AUC for ARDS prediction, namely infiltration, pneumonia and consolidation, were used in further models. The use of these three variables was deemed fit since these share similarities in disease with ARDS. While this type of data will not be available at time of admission to ED, the implementation of CheXNet and image-based models can be available at the time when the chest radiograph is available.

As part of the aims of this research, additional models to predict each of the four outcomes were built by adding the predictions from the three named CheXNet categories. The inclusion of predicted infiltration, pneumonia and consolidation to the defined 15 variables resulted in prediction accuracies with an AUC of 0.748 (0.681–0.814) for COVID-19 infection, 0.694 (0.596–0.793) for ARDS, 0.675 (0.611–0.702) for ICU transfer and 0.748 (0.661–0.835) for mortality. Compared to the compact models using only 15 clinical variables, using DeLong Test, the AUCs were significanly higher for all of the outcomes under study with $p < 0.001$.

We also built respective models using demographic data, first oxygen levels, and covid infection (for ARDS, ICU admission and mortality) together with infiltration, pneumonia and consolidation to assess the power of these X-Ray specific predictors on predicting each of the four classes. The use of just demograpics, first oxygen levels and the three CheXNet predicted classes resulted in AUC of 0.730 (0.663–0.797). When COVID-19 infection was added as a predictor for ARDS, ICU admission and mortality, AUC scores for ARDS increased to 0.702 (0.604–0.800) and 0.750 (0.664–0.837) for risk of mortality. The results for each developed model are provided in Table 2. Furthermore, calibration was performed for the best machine learning models using clinical data, namely COVID-19 infection, ARDS and ICU using all 42 variables and using 15 clinical variables for risk of mortality. The calibration curves are provided in Figure S.1 (a-d; supplementary material). Calibration of the models only minimally improved, if at all, predictions from the associated models. We acknowledge that in some cases the presented models can over- or under-estimate the predicted probabilities for each outcome using clinical variables (e.g. FigureS.1b and d). This could be due to the imbalance between positive cases and the rest of the patient cohort within the dataset for each of the four outcomes in the study.

3.3. Transfer learning to create CheXNet-Cov19

CheXNet-Cov19 model provided prediction accuracies with an AUC of 0.712 (0.627–0.797) for COVID-19 infection, 0.741 (0.658–0.824) for risk of ARDS, and risk stratification of 0.665 (0.578–0.752) for ICU admission and 0.759 (0.678–0.840) for mortality (Table 3). Table 3 also highlights and compares the results obtained when building different models on i) all variables + predicted risks, ii) 15 top variables + predicted risks and iii) 15 top variables + predicted risks + predicted infiltration, consolidation and pneumonia.

The inclusion of predicted risk to the model built on all the predictors resulted in AUC of 0.776 (0.685–0.867) for ARDS, AUC of 0.663 (0.609–0.716) for need of ICU admission and AUC of 0.736 (0.69–0.781) for risk of mortality. Similar results were obtained when the number of variables within the model was 15 variables + predicted risks. The inclusion of risk of infiltration, consolidation and pneumonia predicted from the CheXNet CNN resulted in AUCs of 0.728 (0.632–0.824), 0.675 (0.622–0.727) and 0.758 (0.672–0.844) for ARDS, need of ICU

Table 2
Models to predict COVID-19 infection, ARDS, ICU admission and risk of mortality using i) all available clinical variables, ii) 15 top clinical variables, iii) 15 top clinical variables in addition to the three predicted classes from CheXNet and iv) demographics + first oxygen levels + 3 predict CheXNet classes. AUC = area under the receiver operating characteristics, SPE = specificity, SEN = sensitivty.

Model Type	COVID-19			Acute Respiratory Distress Syndrome (ARDS)*			ICU Admission*			Mortality*		
	AUC	SPE	SEN	AUC	SPE	SEN	AUC	SPE	SEN	AUC	SPE	SEN
All 42 clinical variables	0.790 (0.746–0.835)	0.76 (0.67–0.75)	0.7 (0.60–0.80)	0.753 (0.675–0.831)	0.75 (0.66–0.84)	0.60 (0.50–0.70)	0.675 (0.620–0.713)	0.6 (0.50–0.70)	0.65 (0.55–0.75)	0.683 (0.606–0.761)	0.66 (0.56–0.76)	0.61 (051–0.71)
Top 15 clinical variables	0.775 (0.730–0.821)	0.71 (0.61–0.81)	0.71 (0.61–0.81)	0.721 (0.641–0.802)	0.75 (0.66–0.84)	0.47 (0.37–0.57)	0.658 (0.611–0.702)	0.59 (0.49–0.69)	0.55 (0.45–0.65)	0.755 (0.669–0.841)	0.68 (0.58–0.78)	0.72 (0.62–0.82)
15 clinical variables + CheXNet risk predictions of (consolidation + infiltration + pneumonia)	0.748 (0.681–0.814)	0.72 (0.62–0.82)	0.71 (0.61–0.81)	0.694 (0.596–0.793)	0.67 (0.57–0.77)	0.67 (0.57–0.77)	0.675 (0.622–0.727)	0.65 (0.55–0.75)	0.65 (0.55–0.75)	0.748 (0.661–0.835)	0.74 (0.64–0.83)	0.70 (0.60–0.80)
Demographics + first oxygen levels + CheXNet risk predictions of (consolidation + infiltration + pneumonia)	0.730 (0.663–0.797)	0.74 (0.84–0.84)	0.61 (0.51–0.71)	0.702 (0.604–0.800)	0.63 (0.53–0.73)	0.64 (0.54–0.74)	0.657 (0.603–0.710)	0.61 (0.51–0.71)	0.64 (0.54–0.74)	0.750 (0.664–0.837)	0.76 (0.67–0.85)	0.65 (0.55–0.75)

* Models include COVID-19 infection binary status as a predictor.

Table 3
Models to predict ARDS, ICU admission and risk of mortality using i) transfer learning from CheXNet, ii) all clinical variables + CheXNet risk predictions, iii) 15 top clinical variables + the predicted risk and iv) top 15 clinical variables in addition to predicted risks as well as risk of consolidation, infiltration and pneumonia (from initial CheXNet model). AUC = area under the receiver operating characteristics, SPE = specificity, SEN = sensitivty.

Model Type	COVID-19			Acute Respiratory Distress Syndrome (ARDS)*			ICU Admission*			Mortality*		
	AUC	SPE	SEN	AUC	SPE	SEN	AUC	SPE	SEN	AUC	SPE	SEN
CheXNet-Cov19 using chest X-Rays only	0.71 (0.63–0.80)	0.69 (0.56–0.79)	0.56 (0.46–0.66)	0.74 (0.66–0.82)	0.73 (0.63–0.83)	0.63 (0.53–0.73)	0.67 (0.58–0.75)	0.58 (0.48–0.68)	0.63 (0.53–0.73)	0.76 (0.68–0.84)	0.71 (0.61–0.71)	0.62 (0.52–0.72)
CheXNet-Cov19 Predicted risk* + All clinical variables				0.77 (0.69–0.87)	0.62 (0.52–0.72)	0.53 (0.43–0.63)	0.66 (0.61–0.72)	0.54 (0.44–0.64)	0.67 (0.57–0.77)	0.74 (0.69–0.78)	0.61 (0.51–0.71)	0.73 (0.63–0.83)
CheXNet-Cov19 Predicted Risk + 15 clinical variables *				0.78 (0.69–0.87)	0.71 (0.61–0.80)	0.53 (0.43–0.63)	0.65 (0.59–0.70)	0.55 (0.45–0.64)	0.67 (0.57–0.77)	0.75 (0.66–0.84)	0.71 (0.61–0.80)	0.7 (0.60–0.80)
CheXNet-Cov19 predicted Risk + 15 clinical variables + consolidation + infiltration + pneumonia *				0.73 (0.63–0.82)	0.69 (0.59–0.79)	0.69 (0.59–0.79)	0.68 (0.62–0.73)	0.72 (0.62–0.82)	0.58 (0.48–0.68)	0.758 (0.672–0.844)	0.71 (0.61–0.80)	0.67 (0.57–0.77)

* Models include COVID-19 infection binary status as a predictor.

admission and mortality, respectively.

3.4. Subgroup analysis

We implemented a subgroup analysis for 'CheXNet-Cov19 using chest X-Rays only' model and summarized our results in Table 4. Table 4 shows that our proposed predictive model identifies a subset of patients who are at higher risk for each of the four outcomes compared to entire analytic sample of patients admitted to ED.

3.5. Sensitivity analysis

We further implemented a sensitivity analysis on the holdout data for all four outcomes. To do that, separately for each outcome, we identified the patients with top 20% highest predicted risk and calculated the observed percentage of the occurrence of each outcome (Table 5).

4. Discussion

The novel COVID-19 virus has taken the world by surprise and has been recognized as a pandemic due to its ability of quick infection, transmission, as well as mutation into more transmissible variants [7]. Numerous scientific communities have taken to study the virus and its effects on human health with results showing that the virus predominantly attacks the lungs and was the cause of a large influx of patients to intensive care and a consequently very high mortality rate [2]. Even more so, lung infections caused by this viral infection differs from 'typical' pneumonia, and identification of viral pneumonia has been reportedly difficult [11]. While testing for COVID-19 infections has progressed substantially to very rapid testing with the main aim of containing viral infection by encouraging people to self-quarantine, there is still a need for early prediction of possible severe outcomes which can result in ICU admission and/or mortality if left undetected or untreated. People with severe infections and symptoms, especially those of older age, are generally admitted to the emergency department during which a series of chest X-rays are taken in order to be assigned appropriate treatment by a consulting physician, especially if there was determinable development of acute respiratory distress syndrome [9]. However, there has been a recorded bottleneck between time of taking chest X-rays, diagnosis of severity of infection and admission to ICU, unless severe cases have been recognized by ambulatory services [14]. The increase in clinical and radiographic data collected by different hospitals and research groups has provided a pathway for the use of AI to predict risks for these four different but related outcomes [21,25]. Some research has also used blood-based biomarkers to examine the association between COVID-19 infections and the possible need for ICU admission [10]. However, while such research compares different machine learning algorithms to predict specific outcomes at high AUC, there is a notable difficulty in developing a generalizable system that can output predicted risks for each of the four outcomes using a single model. In addition, the use of transfer learning from a widely accessible chest radiographic database and pre-developed CNN model can also provide an important pathway to combine ARDS-specific lung damage to COVID-19 infection, need for ICU admission based on a cascading effect brought about by a combination of conditions as well as the possibility of mortality.

In our study, we developed a series of models that used both clinical variables using machine learning as well as chest X-rays (in a CNN) to try and predict risk of ARDS, the possible need for ICU admission, and risk of mortality. Our models were built on patient data available during admission to the emergency department as well as the utilization of transfer learning, using CheXNet [18], to predict lung-specific diseases most notably infiltration, consolidation and pneumonia. These 3 conditions were added to the most important risk predictors to assess risk of ARDS, need for ICU admission and risk of mortality. In addition, the comprehensive set of models developed in this research aim to classify all four outcomes based on the same dataset and can be implemented at the first point of data availability. Our mortality models do not include admission to ICU as possible predictor since ICU admission can be delayed due to bed unavailability or unassessed levels of symptomatology, increasing the generalizability of the models.

Prediction of risk of ARDS resulted in a moderate AUC of 0.721 (0.641–0.802) using solely 15 short-listed clinical variables without the addition of predicted risk of infiltration, consolidation or pneumonia while risk of mortality was predicted with an AUC of 0.755 (0.669–0.841) without the addition of these risk variables. Variable importance analyses for each model shows that first oxygen levels and age are amongst the most important predictors. This is what was expected since COVID-19 infection reduces oxygen levels within the blood, as well as older patients being more susceptible to infection because of higher proportion of comorbidities [16]. However, variable importance analysis for ARDS prediction models show that COVID-19 infection is indeed within the top three important predictors, conferring the association between COVID-19 viral infection and acute respiratory distress syndome, adding face validity to the models developed. On the other hand, variable importance analysis for models built to assess the need for ICU admission (Fig. 1b), the first oxygen levels retain their importance followed by heart disease. When compiling risk stratification of mortality, Fig. 1c shows that age is, unsurprisingly, the largest contributor to mortality, followed by oxygen levels and liver and cardiac diseases. COVID-19 infections might have potentially had a role in increased risk of mortaltiy due to its higher importance shown in Fig. 1c. The first set of models developed in this research (Table 2), use demographic and clinical variables that are, generally, readily available when patients are admitted to ED. The slightly lower AUC in our study, compared to other research [10,25] offers a balanced tradeoff with the reduced use of invasive or unaccessible medical information, allowing clinicians to make informed decisions without the need to wait for further laboratory examination, such as, blood sample analysis [10,17].

When these risk variables for infiltration, consolidation, or pneumonia were added to the models, the AUC did not change much for prediction of COVID-19 infection, while AUC for prediction of ARDS with inclusion of these external variables decreased to 0.694 (0.596–0.793), presumably due to the similarities and overlap between these three variables and the broad definition of acute respiratory distress syndrome [23]. The results following the addition of predicted risks for ARDS, ICU admission and mortality to the short-listed clinical variables in the initial models showed a notable increase in accuracy for prediction of ARDS (0.781) but retained similarities to models built only on short-listed clinical variables for mortality (0.751). However, we note that stratified risk of mortality solely from chest radiographs is slightly higher (0.759). The models built to predict possibility of mortality also include within them predicted risk of ARDS and need for ICU admission. The reason for this is because our developed models can be utilized at time of deployment of chest radiographs at first time of availability in a clinical setting, potentially helping clinicians determine best course of treatment. The inclusion of extracted information from chest radiographs has been highlighted by researchers [21,25] and while some current research uses such information, the use of raw chest radiographs

Table 4
AUC statistics with 95% CI for X-Ray only model.

Outcome/ Subgroup	Male	Female	Non-white	White
COVID19	0.73 (0.64–0.82)	0.68 (0.58–0.78)	0.74 (0.66–0.82)	0.66 (0.54–0.78)
ARDS	0.75 (0.64–0.86)	0.73 (0.60–0.86)	0.78 (0.68–0.88)	0.66 (0.61–0.82)
ICU Admission	0.74 (0.66–0.82)	0.79 (0.68–0.89)	0.81 (0.74–0.88)	0.67 (0.55–0.79)
Death	0.66 (0.59–0.73)	0.66 (0.59–0.73)	0.69 (0.63–0.75)	0.61 (0.53–0.69)

Table 5
Sensitivity analysis on the holdout data for all four outcomes.

Outcome	Actual Prevalence on Holdout (n = 565)	Observed prevalence in top 20% highest predicted risk patients (n = 113)
COVID19	11.3%	17.7%
ARDS	6.4%	14.2%
ICU Admission	26.2%	44.1%
Death	8.5%	20.4%

increases ease of inclusion in clinical workflow. We acknowledge that admission to intensive care is somewhat subjective by the trained physician, but it would be possible to include the predicted risk in the model for increased precaution by decreasing the time spent at emergency department when the patient has a high noted risk of requiring intensive care [15].

Our sensitivity analysis showed that our proposed models can identify a small subset of patients who are at elevated risk for outcomes. For example, although the overall mortality rate in the cohort was 8.5%, the mortality rate among the patients within the top 20% predicted death risk was 20.4%. For ICU admission, the overall observed prevalence among ED admissions was 26.2%, while the sensitivity analysis showed an expected prevalence of 44.1% among the patients within the top 20% highest ICU admission risk prediction. From the perspective of clinical implementation, it is important to note that intensive care units during a pandemic can be overwhelmed. For future prospective analysis and potential implementation, we suggest using the compacted models that were developed to increase generalizability of model implementation and to reduce clinical burden, especially in clinical settings. The results from the DeLong test showed that compact models with predicted risk transferred from image analysis are preferred over the original models and can better increase awareness and help clinicians in decision making to assess need for ICU admission and risk of mortality. However, the benefit of development of multiple models, as has been done in this research, is to benefit clinicians in predicted risk at multiple time points from first admission of patients to ED. At time of admission of to ED, the first associated predictions or risk stratifications can be achieved from clinical variables based on patient demographics and medical history. Following the 'first wave' of predictions, the use of chest radiograph-based models can provide further risk information to the clinicians as a 'second wave'. The combined predicted/stratified risk results obtained from both AI waves can substantially help clinicians in the determination of best course of action, especially when patients are at an increased risk of requiring ICU admission or death. Whilst predicted at lower AUCs due to its subjectivity, model predictions and classification patient groups in the high risk can be monitored closely and subjected to further planning in medical needs. There is additional scope to utilize the incorporation of image analysis and prediction on mobile technologies such as smartphones. The use of predictive models incorporated into such technologies can potentially help clinicians to incorporate machine learning/deep learning efforts into the workflow.

5. Limitations

Our research has some limitations. Firstly, this study is a single site study and therefore requires external validation before its use in clinical practice. Also, there is a data imbalance in the number of patients with COVID-19 infections and ARDS, which could have reduced the models' training and performance when using clinical variables and image analysis. The data included, though in a small proportion, patients that had developed ARDS but were not diagnosed with COVID-19 infection. In addition, predicted need of ICU admission scored a comparatively lower AUC. This could be because of ICU admission not being directly related to the predictive variables that were collected and used. ICU admission is largely subjective based on clinicians' professional advice in terms of severity of illness, bed availability and hospitalization costs [20].

6. Conclusions

Our research results show that predicted risk of COVID-19 infection, ARDS and mortality can be achieved at moderate accuracy from both chest X-ray images as well as with the addition of clinical variables. This research offers clinical applicability in that it provides a developed tool that can assess chest X-rays at the time that they are taken, which can substantially help clinicians in decision making as well as reduce burden on intensive care units. Since our dataset also included COVID-19 negative patients, it proposes a methodological approach of assessing ARDS and risk of mortality irrespective of COVID-19 status. Furthermore, the use of solely X-ray data in AI models can potentially be improved with increased computational power, high batch size analysis as well as the use of a larger cohort size.

7. Summary points

- COVID-19 has high probability of causing severe acute respiratory syndrome (ARDS) and subsequent increase risk for in-hospital death.
- During the course of the pandemic, COVID-19 has hospitalized a large number of infected patients who had to be admitted to the intensive care unit (ICU).
- Image-based deep learning models can identify outcomes of the infection, such as in the lungs, and early assessment of infection can aid clinicians for time-sensitive diagnostics, including the need for ICU admission.
- The use of clinical data and chest radiographs can be inputted into a machine and deep learning models to predict risk of each clinical outcome and facilitate clinical decision making.
- Light Gradient Boosting Machines (LightGBM) was used as the main machine learning algorithm to develop models to predict 4 main outcomes: COVID-19 infection, development of ARDS, ICU admission and risk of mortality.
- Models using 42 clinical variables and compact clinical models limited to 15 clinical variables were developed.
- Transfer learning from the CheXNet model was implemented on chest radiographs to predict the 4 aforementioned outcomes.
- Results from this research can help in clinical decision making, during the assessment of patients admitted to the emergency department with or without COVID-19 symptoms.

CRediT authorship contribution statement

OA, MA, and MST concptualized the study. LB implemented data curation. LB and IK carried out formal analysis. LB created the original draft, all authords contributed to writing, review, and editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Loyola University Chicago, Loyola University Medical Centre and all associated personnel in their contirbution to data

acquisiton.

Funding Sources

This project was funded (PIs Akbilgic, Tootooni, Afshar, and Goldberg) by CHOIR: Center for Health Outcomes and Informatics Research, Loyola University Chicago.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2021.104662.

References

[1] H.C. Azzam, S.S. Khalsa, R. Urbani, C.V. Shah, J.D. Christie, P.N. Lanken, B. D. Fuchs, Validation study of an automated electronic acute lung injury screening tool, J. Am. Med. Inform. Assoc. 16 (4) (2009) 503–508.

[2] D. Baud, X. Qi, K. Nielsen-Saines, D. Musso, L. Pomar, G. Favre, Real estimates of mortality following COVID-19 infection, Lancet. Infect. Dis 20 (7) (2020) 773.

[3] V. Bhatnagar, R.C. Poonia, P. Nagar, S. Kumar, V. Singh, L. Raja, P. Dass, Descriptive analysis of COVID-19 patients in the context of India, J. Interdisc. Math. (2020) 1–16.

[4] I. Castiglioni, D. Ippolito, M. Interlenghi, C.B. Monti, C. Salvatore, S. Schiaffino, A. Polidori, D. Gandola, C. Messa, F. Sardanelli, Artificial intelligence applied on chest X-ray can aid in the diagnosis of COVID-19 infection: a first experience from Lombardy, Italy, MedRxiv, 2020.

[5] Chavez, S., Long, B., Koyfman, A., & Liang, S. Y. (2020). Coronavirus Disease (COVID-19): A primer for emergency physicians. *The American journal of emergency medicine*.

[6] D. Cucinotta, M. Vanelli, WHO declares COVID-19 a pandemic, Acta Bio Medica: Atenei Parmensis 91 (1) (2020) 157.

[7] A.C. Darby, J.A. Hiscox, Covid-19: variants and vaccination. In: British Medical Journal Publishing Group (2021).

[8] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (3) (1988) 837–845.

[9] M. Dreher, A. Kersten, J. Bickenbach, P. Balfanz, B. Hartmann, C. Cornelissen, A. Daher, R. Stöhr, M. Kleines, S.W. Lemmen, The characteristics of 50 hospitalized COVID-19 patients with and without ARDS, Deutsches Ärzteblatt International 117 (10) (2020) 271.

[10] L. Famiglini, G. Bini, A. Carobene, A. Campagner, F. Cabitza, Prediction of ICU admission for COVID-19 patients: a Machine Learning approach based on Complete Blood Count data. 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), 2021.

[11] P.G. Gibson, L. Qin, S.H. Puah, COVID-19 acute respiratory distress syndrome (ARDS): clinical features and differences from typical pre-COVID-19 ARDS, Med. J. Aust. 213 (2) (2020) 54–56.

[12] O. Gozes M. Frid-Adar H. Greenspan P.D. Browning H. Zhang W. Ji A. Bernheim E. Siegel Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis. (2020) arXiv preprint arXiv:2003.05037.

[13] H.C. Koenig, B.B. Finkel, S.S. Khalsa, P.N. Lanken, M. Prasad, R. Urbani, B. D. Fuchs, Performance of an automated electronic acute lung injury screening system in intensive care unit patients, Crit. Care Med. 39 (1) (2011) 98–104.

[14] T. Li, Z. Han, B. Wei, Y. Zheng, Y. Hong, J. Cong,. Robust screening of covid-19 from chest x-ray via discriminative cost-sensitive learning. (2020). *arXiv preprint arXiv:2004.12592*.

[15] X. Ma, D. Vervoort, Critical care capacity during the COVID-19 pandemic: global availability of intensive care beds, J. Crit. Care 58 (2020) 96.

[16] S. Niu, S. Tian, J. Lou, X. Kang, L. Zhang, H. Lian, J. Zhang, Clinical characteristics of older patients infected with COVID-19: A descriptive study, Arch. Gerontol. Geriatr. 89 (2020), 104058.

[17] J.C. Quiroz, Y.-Z. Feng, Z.-Y. Cheng, D. Rezazadegan, P.-K. Chen, Q.-T. Lin, L. Qian, X.-F. Liu, S. Berkovsky, E. Coiera, Development and validation of a machine learning approach for automated severity assessment of COVID-19 based on clinical and imaging data: Retrospective study, JMIR Med. Inform. 9 (2) (2021), e24572.

[18] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, Matthew, & Andrew. (2017). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *arXiv pre-print server*. https://doi.org/None arxiv:1711.05225.

[19] V.M. Ranieri, G.D. Rubenfeld, B.T. Thompson, N.S. Ferguson, E. Caldwell, E. Fan, L. Camprota, A.S. Slutsky, Acute Respiratory Distress Syndrome: The Berlin Definition, JAMA 307 (23) (2012) 2526–2533, https://doi.org/10.1001/jama.2012.5669.

[20] R. Robert, R. Coudroy, S. Ragot, O. Lesieur, I. Runge, V. Souday, A. Desachy, J.-P. Gouello, M. Hira, M. Hamrouni, Influence of ICU-bed availability on ICU admission decisions, Ann. Intensive Care 5 (1) (2015) 1–7.

[21] I. Shiri, M. Sorouri, P. Geramifar, M. Nazari, M. Abdollahi, Y. Salimi, B. Khosravi, D. Askari, L. Aghaghazvini, G. Hajianfar, Machine learning-based prognostic modeling using clinical data and quantitative radiomic features from chest CT images in COVID-19 patients, Comput. Biol. Med. 132 (2021), 104304.

[22] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R.M. Summers, (2017, 2017). ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases.

[23] A.B. Weinacker, L.T. Vaszar, Acute respiratory distress syndrome: physiology and new management strategies, Annu. Rev. Med. 52 (1) (2001) 221–237.

[24] WHO. (2021). *WHO Coronavirus (COVID-19) Dashboard*. Retrieved 13/04 from https://covid19.who.int/.

[25] W. Xu, N.-N. Sun, H.-N. Gao, Z.-Y. Chen, Y. Yang, B. Ju, L.-L. Tang, Risk factors analysis of COVID-19 patients with ARDS and prediction based on machine learning, Sci. Rep. 11 (1) (2021) 1–12.

[26] S. Yadav, S. Shukla, Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. 2016 IEEE 6th International conference on advanced computing (IACC), 2016.