FluDeep: Combing Deep Learning-based Chest X-Ray Severity (FluDeep XR) and Clinical Variables to Predict outcome of Hospitalized Patients with Influenza

Manuscript of table: <https://docs.google.com/document/d/1GLVaBogd-OyZ-0g9xpWpww_QoQ12w0ASgXEjui1txhs/edit?usp=sharing>

**Abstract**

**Background**

Influenza is a major health issue worldwide with an estimated global burden of over five million related hospitalizations of which 30-40% for pneumonia. The creation of a prediction mortality tool to early risk-stratify patients would allow physicians to identify possible critical patients and optimize the allocation of resources. For this purpose, machine learning models could be applied, however previous studies, including only clinical data, failed in designing accurate prediction systems.

**Objective**

We aim to develop the first accurate machine learning model that predicts 30 days mortality of influenza patients, by integrating clinical data with chest X-rays images, using an objective severity score we designed.

**Methods**

We identified all adult hospitalized patients with influenza confirmed by a nasopharyngeal swab and with a chest X-ray in the first 24 hours of presentation to the National Taiwan University hospital in Taipei, Taiwan. We scored, stratified by severity and randomly divided all chest X-rays into training and testing sets (7:3); then, we regressively trained several deep learning models to score the severity of the pathological changes. Next, we combined the severity scores with clinical data, stratified them by 30-day survival, randomly divided into training and testing sets (7:3), and performed feature selection to extract the most important variables. We designed two types of fusion models with different weights of chest-X-ray to predict the risk of disease progression.

**Findings**

765 patients who presented to the National Taiwan University hospital between Jan 9, 2016 and Dec 31, 2018 were identified and 1018 CXR were taken in total. The CXRs are assigned to model training (n=712) or testing sets (n=306). Pre-trained Xception performed the best at scoring the severity of the CXR with a mean square error (MSE) improved from 2.79 to 1.14. Subsequently, the patients were stratified and assigned to the model training (n=535) or testing sets (n=230). The selected features for the prognosis prediction system involved the severity score of CXR images and patients’ demographic features, vital signs and laboratory markers. LightGBM machine learning models had the highest AUROC and F1 scores, and model fusion incorporating EMR clinical data and CXR image severity improved model performance. In addition, late fusion model performed better than early fusion models.When CXR were added to clinical data for severity prediction, the area under the receiver operating characteristic curve (ROC-AUC) increased from 0.859 to 0.904 for the early fusion model and to 0.988 for the late fusion model.

**Interpretation**

To our knowledge, we developed the first deep learning-based model to quantify the severity of influenza CXR images (FluDeep-XR) and the first AI model that combined radiological image severities and clinical data (FluDeep) for prediction of 30-day mortality for influenza pneumonia. This is also the first study to demonstrate that model fusion improved prognosis prediction. Of note, all of the extracted features in our study did not require physician judgment, indicating the potential of establishing an automatic, objective prognostic prediction system for flu patients via the artificial intelligence system we built.

**Keywords:** FluDeep-XR, FluDeep, Flu Severity Prediction Fusion model, Influenza, Chest x-ray, Clinical data, Artificial intelligence, Pre-trained model

**Introduction**

Influenza is a major health issue worldwide, with an estimated global burden of over five million related hospitalizations and 200,000 deaths every year. Among hospitalized influenza patients 30 to 40% present with acute pneumonia, 9% develop acute respiratory distress syndrome (ARDS) and more than 25% require admission to the critical care unit. Even though in the emergency department physicians try to risk-stratify influenza patients, due to the heterogeneous presentation, uncertainty prevails and can affect the correct management of these patients. Thus, it is pivotal to design an accurate outcome prediction tool that would allow doctors to early recognize high-risk mortality influenza patients who need prompt intervention to prevent disease worsening and better allocate resources.

In the last decade, several mortality prediction models for influenza have been developed. However, unlike clinicians who assess the severity of a patient by considering both radiological and clinical data, these models only included the latter, thus exhibiting suboptimal predictive performance with areas under the receiver operator characteristics curve (AUROC) between 0.766 and 0.908. To increase accuracy, scientists tried to integrate the findings of chest X-rays in the analyses, since they have been demonstrated to predict the severity of viral pneumonia and its progression to respiratory failure and death. However, due to the heterogeneity and mostly qualitative nature of the radiological reports, it has been difficult to integrate them in the prognosis prediction systems. To address this problem, artificial intelligence (AI) models could be applied. In fact, these models enable a numerical quantification of the severity of chest X-ray images and, therefore, the possibility to calculate a prognosis in terms of severity or mortality. The feasibility and accuracy of these models have already been demonstrated by several research groups, who applied AI to evaluate chest X-rays to detect ARDS or predict the severity of COVID-19 infection.

Our study aims to design the first machine learning based mortality prediction model for pneumonia due to influenza that combines clinical and radiological data. To do so, we decided to conceive a severity scoring model of influenza chest X-rays which could be applied to enhance the performance of AI prognosis prediction systems.

**Methods**

**Study participants and data collection**

We retrospectively identified all patients with influenza who were admitted to the emergency department of the National Taiwan University Hospital from 16 January 2016 to 9 September 2018. We included all patients with a diagnosis of influenza, defined by the positivity of a rapid influenza diagnostic test, and at least one chest X-ray within the first 24 hours in the emergency department (Figure 1). Patients were excluded from the study if they met the following criteria: (1) less than 20 years of age; (2) on palliative care; (3) had missing data; (4) were transferred to other hospitals or (5) received treatment non-adherent to standard medical guidelines. The Research Ethics Committee Office of National Taiwan University Hospital approved this study. We collected the patients’ chest X-ray images and clinical data from the electronic medical record, including demographic features, comorbidities, vital signs and results of laboratory exams. We considered for vital signs and laboratory tests the values obtained at the initial presentation to the hospital. We de-identified all data to preserve patient confidentiality.

e

**Radiological Severity Score**

We assigned to all chest X-ray images a radiological severity score based on the following scoring system: 1 for normal lung; 2 for hyperinflation and/or bronchial wall thickening; 3 for focal alveolar consolidation limited in a single lobe; 4 for multifocal consolidation; 5 for diffuse alveolar consolidation and/or pleural effusion involving the upper lobe on at least one lung (Table 1). The severity scoring of the chest X-rays was performed by two medical students and later checked by an experienced physician and a senior radiologist. The discrepancies in the score were resolved by discussion between the emergency physician and the senior radiologist; if an agreement was not met, the score was adjudicated by the senior radiologist. The Radiological Severity Score was defined as ground truths for subsequent training of deep learning models.

**Chest X-ray Severity Scoring model development**

For the training of the severity scoring model, we used a total of 1,018 chest X-rays images, divided in the following numbers for categories from 1 to 5, respectively: 243, 138, 182, 187, and 268. The candidate convolutional neural network (CNN) models used in this study (ResNet50, EfficientNet, VGG16, DenseNet121, InceptionV2, InceptionV3 and Xception) underwent pre-training on ImageNet before model training to improve model performance. To allow so, we reshaped all the chest X-ray images to 299x299 pixel size and normalized to the range 0-1. All images were subsequently stratified by the predetermined severity scores and randomly assigned (7:3) to either model training or testing sets to train the CNN models through transfer learning. To fit the purpose of our study, we added three fully connected layers with respectively 256, 32, and 1 nodes at the end of the models. The activation function was linear, and the optimizer was Adam with a learning rate of 0.0002 decaying by 0.00001 every epoch. We called the model with the best performance FluDeep-XR.

**Feature selection**

The importance of all the collected clinical variables represented their contribution to stratifying patients and provided explainability and it is ranked by the same machine learning model that was used for building the prognosis predicting system in the next step.

**Clinical-data-only model development**

All patients were stratified by 30-day mortality and randomly assigned to model training and testing sets (7:3) and only selected clinical data is used while building this model. The candidate machine learning models used were Random Forest, XGBoost Classifier, and LGBM Classifier. The model was called clinical-data-only model.

**Fusion Model development**

We combined clinical data with the radiological severity scoring model to create two fusion models: early fusion and late fusion, as illustrated in Figure 2. Random Forest, XGBoost and LightGMB were the candidate machine learning models. For the early fusion system, the chest X-ray severity scores were input as a clinical feature along with clinical data and the selected features were then directly input into a machine learning model. For the late fusion model, the result of the clinical-data-only model and the radiological severity scoring model (FluDeep-XR) were generated independently, and the two numeric outputs were later combined as the input for a support vector classifier. We added platt scaling at the end of all systems to calibrate the outcome of the AI-based system to predicted probabilities. The fusion model with the most accurate prediction performances for 30-day influenza mortality was called FluDeep. We then designed a graphical user interface for other researchers and clinicians for further studies and applications (Appendix figure 1-2).

**Explainability and Usability**

For explainability purposes, we ranked all the variables contributing to our model and demonstrated them in the order of importance. Furthermore, to enhance usability of the model, we designed a graphic website with a user-friendly interface calculator that allows physicians to upload patients’ data and chest x-ray images to obtain a probability of death within 30 days.

**Statistical analysis**

To evaluate the difference of patients’ characteristics between Survivors and Non-Survivors, we used the Student’s t test for continuous variables and the χ² test for categorical variables. We calculated the mean-square error (MSE) between the labeled value from physicians and the scoring model for the FluDeep-XR model to compare the precision between distinct models and to identify overfitting of the models. We calculated the area under the receiver operating characteristic curve (AUROC), accuracy, sensitivity, specificity, and precision to assess the performance of our clinical-data-only system, early fusion system, late fusion system. We applied DeLong’s tests to compare the AUROC of different AI systems and we used F1 score, precision-recall curve, and calibration plot to evaluate and determine the best architecture of the systems we designed. Furthermore, we compared our new prediction model against existing risk stratification scores for viral pneumonia, such as the MuLBSTA score for multiple viral pneumonia and the 4C mortality score for COVID-19 pneumonia. We also stratified our cohort by our FluDeep-XR scoring model to evaluate the capability of predicting their prognosis only on their first chest X-ray. We determined the cutoffs for sensitivities and specificities of each scoring system by the Youden index and applied DeLong’s tests to compare the AUROC of the different scoring systems and our AI systems.

**Results**

**Characteristics of Participants**

We identified a total of 765 patients admitted to the Emergency Department of the National Taiwan University Hospital between the 16th of January 1, 2016 and the 9th September 2018. Of every patient we collected all chest X-ray images for a total of 1018. We stratified patients in two groups based on mortality at 30 days: Survivors (S) and Non-survivors (N-S). We evaluated the differences in characteristics (demographics, vital signs, laboratory markers, comorbidities, and chest x-ray image severity) between the two groups, as shown in Table 2 Non-survivors were older (mean age S: 65.9 ± 18.2 versus N-S 78.3 ± 12.82 years, p-value<0.0001), had higher CRP levels (S: 8.39 versus N-S:11.98, p-value 0.0003), higher prevalence of hypertension. (S: 43% versus N-S: 62%, p-value 0.0265) and higher average severity scores of first CXR image taken within 24 hours of presentation to the ER (S: 2.98 ± 1.17 ; N-S: 3.92 ± 1.05, p value<0.0001).

**Chest X-ray Severity Scoring model (FluDeep-XR)**

We used six CNN models with and without pre-training with ImageNet to build the radiologic severity scoring model. As shown in Table 3, pre-training improved the performances of both InceptionV3 and Xception, by reducing the MSE. The ImageNet Pre-trained Xception model exhibited the highest performance with the lowest MSE (1.1390) and was called FluDeep-XR.

**Clinical-data-only and Fusion Models (FluDeep)**

We built severity prediction models based solely on clinical data (clinical-data-only model) and fusion models, integrating FluDeep-XR and clinical data, using Random Forest, XGBoost Classifier and LightGBM Classifier (LGBM). Table 4 and Figure 4 present the performance of these models. Despite an AUROC ranging from 0.72 to 0.81, Random Forest-based systems lacked the potential of clinical application due to low precisions and F1 scores. For XGBoost- and LGBM-based, compared to clinical-data-only, fusion systems (early and late) showed improved AUROC and F1 scores, but only the late fusion system demonstrated a significant difference compared to clinical-data-only systems. LGBM-based systems exhibited higher AUROC and F1 scores than the XGBoost-based counterparts. The LGBM-based clinical-data-only, early fusion and late fusion models showed an AUROC of 0.8589, 0.9039 and 0.9876, and F1 scores of 0.4000, 0.6667 and 0.9167, respectively. The late fusion model demonstrated a better-performing precision-recall curve than the early fusion model and was called FluDeep (Figure 5). As displayed in Figure 6 with the calibration plots, the predictive performance of both the early fusion and late fusion models were acceptably stable throughout different predicted probabilities of death within 30 days.

**Explainability of the model**

To better support our model in the real-world setting, we investigated its explainability by identifying and ranking the features contributing the most to FluDeep’s prognosis predictions (Figure 7). The five most important predictors were lactic acid level, FluDeep-XR score, age, C-reactive protein and troponin T levels.

**Discussion**

**Principal Results**

**We designed the first AI mortality prediction system for influenza (FluDeep) that combines clinical and radiological data under the belief that machine learning models should imitate how a physician reasons, maintaining the holistic approach to medicine, while achieving results previously considered impossible. To do so, we conceived the first deep learning model to directly quantify the severity of chest x-ray images (FluDeep-XR). Furthermore, to enhance usability, we designed a graphic website with a user-friendly interface calculator that allows physicians to upload patients’ data and chest x-ray images to obtain a probability of death within 30 days. Compared to previous models which included only either clinical or radiological data, our integrated approach allowed us to create a fusion model with better performance. In fact, after model fusion, the AUROC of the LightGBM-based clinical-data-only model improved from 0.8589 to 0.9039 (early fusion) and 0.9876 (late fusion) (p-value both <0.005), with the late fusion model as a stronger predictor than early fusion (p-value=0.001). Another element that supports our approach is the weight demonstrated by the chest x-rays severity score on prognosis prediction, which ranked as the second most important feature.**

**The idea of combining radiological images to clinical data was derived by recent studies that demonstrated that chest x-rays can be applied to assess the severity of pneumonia in ARDS or COVID-19, using CNN-based models 23,24,39-41,25-27. To clarify whether these images can serve as an independent prognostic factor, we tried to predict the prognosis by only using chest X-ray and determined a cutoff based on the Youden index in our FluDeep-XR model. This resulted in a lower AUROC compared to FluDeep (AUROC 0.9876 vs. 0.6989, p-value <0.001), thus, demonstrating that chest x-rays alone are not an ideal prognostic factor (Figure 8A, Appendix Table 1). Their lack of accuracy could be related to their ability to show delayed pathological changes and thus might not truly reflect the severity of the early stages of pneumonia. However, since they are usually performed in the emergency department, we believed that integrating them in a CNN model would have allowed us to improve accuracy. Thus, to reduce the interobserver variability of chest x-rays, we built a model that automatically generates the severity score of chest x-rays and pre-trained it with other datasets to limit the number of data required and the disparity of distribution between different outcome labels. In fact, previous studies demonstrated that, compared with training from scratch, a pre-trained CNN-based model improved the accuracy of detecting diseases 40,47,48 and the ability of feature extraction49 from medical radiological images. In this study, pre-training models by ImageNet decreased the mean square error of InceptionV3 and Xception by 0.82 and 1.65, which translated to better performance of both models.**

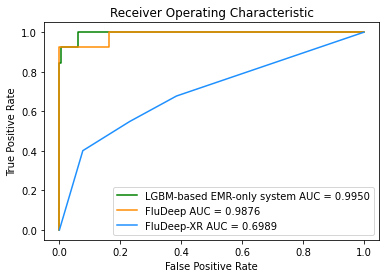
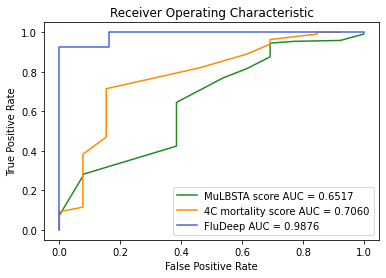
With the accurate prognosis prediction system we developed, physicians in emergency departments can identify patients with a potentially severe disease course earlier, conduct more safe discharges, and may be able to reduce the severity of complications by providing precise treatment. In this way, primary healthcare providers can allocate medical resources more efficiently and effectively, especially during public health crises.

**Comparison with prior work**

Our research presents several differences compared to previous studies. In fact, while we collected a big cohort of patients to firstly develop an objective radiology severity system and then integrate it to clinical data to increase the performance of our model, previous studies failed in stratifying the severity of flu patients, due to smaller cohorts of patients and the use of logistic regression models, which cannot account for the nonlinear relationship between clinical features and outcome. Moreover, not considering chest x-rays, their prediction accuracy was reduced, as demonstrated by Hu et al and Guo et al studies (0.82 and 0.842) that aimed to predict in-hospital influenza mortality and Cheong et al. study (AUROC of 0.757) that aimed to predict ICU patients’ influenza mortality. Furthermore, compared to previous prognosis prediction scores for viral pneumonia (MuLBSTA Score, 4c-mortality score) FluDeep demonstrated better predictive accuracy (AUROC 0.9876 versus 0.6517 of MuLBSTA and 0.7060 of 4c-mortality score, p-value 0.0001 and 0.0074, respectively, Figure 8B).

Prior studies have identified several features related to a worse prognosis, such as early use of a neuraminidase inhibitor, smoking, or coma scale score,13,15,16 which involve physician judgment, and could potentially suffer from inter-rater variability or bias. However, we considered clinical data and chest x-rays severity, which were objective and bypassed physician judgment. This eliminates the subjective evaluation of individual physicians and speeds up the prediction process by running the data on the model in computers. In our study, among the selected features, age, BMI, blood sugar level and decreasing of lymphocytes were proven to be related to severity or prognosis of influenza pneumonia.42-4445,46.

A B

* Figure 8. Comparison of Receiver Operating Characteristic (ROC) curve of Deepflu and other prediction systems.   
  8A Comparison of the ROC curve of LGBM-based clinical-data-only systems, FluDeep-XR, FluDeep. The FluDeep had higher AUC than LGBM-based clinical-data-only systems and FluDeep-XR (AUROC 0.9876 versus 0.8589 and 0.6989, p-value 0.0037 and < 0.0001).   
  8B Comparison of the ROC curve of MuLBSTA score, 4c-mortality score, and FluDeep on our patients. The FluDeep had higher AUC than MuLBSTA score and 4c-mortality score (AUROC 0.9876 versus 0.6517 and 0.7060, p-value 0.0001 and 0.0074).

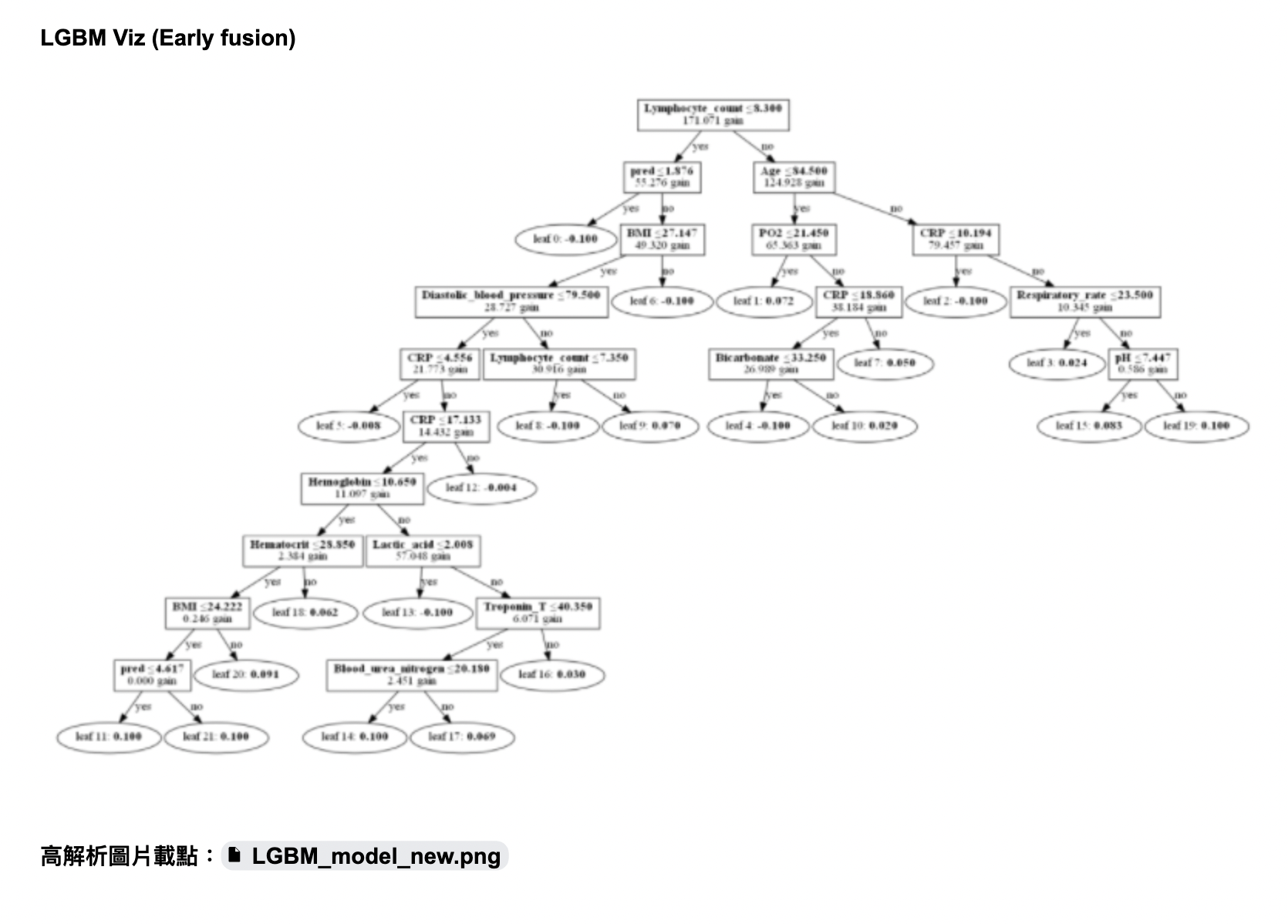
**Limitations**

There are several limitations to our study. Firstly, our cohort only contains patients from a single tertiary academic medical center Which could limit the generalizability of our models and results.[49](https://www.zotero.org/google-docs/?X04HCY) Attempts have been made on our side to obtain external databases for validation for prognosis prediction of viral pneumonia, and access to a Spanish COVID database was obtained. However, we could not identify any influenza database that provides both clinical data and CXR images. Secondly, 30-day survival was used as the cutoff, so long-term impact on survival was not evaluated. Thirdly, the ground truth for CXR severity scoring was performed by experienced radiologists, so inter-rater variability and subjective evaluation were inevitable. Future studies can aim for a larger cohort and datasets from other medical centers to mitigate the limitations. Our system welcomes further external validation or revision.

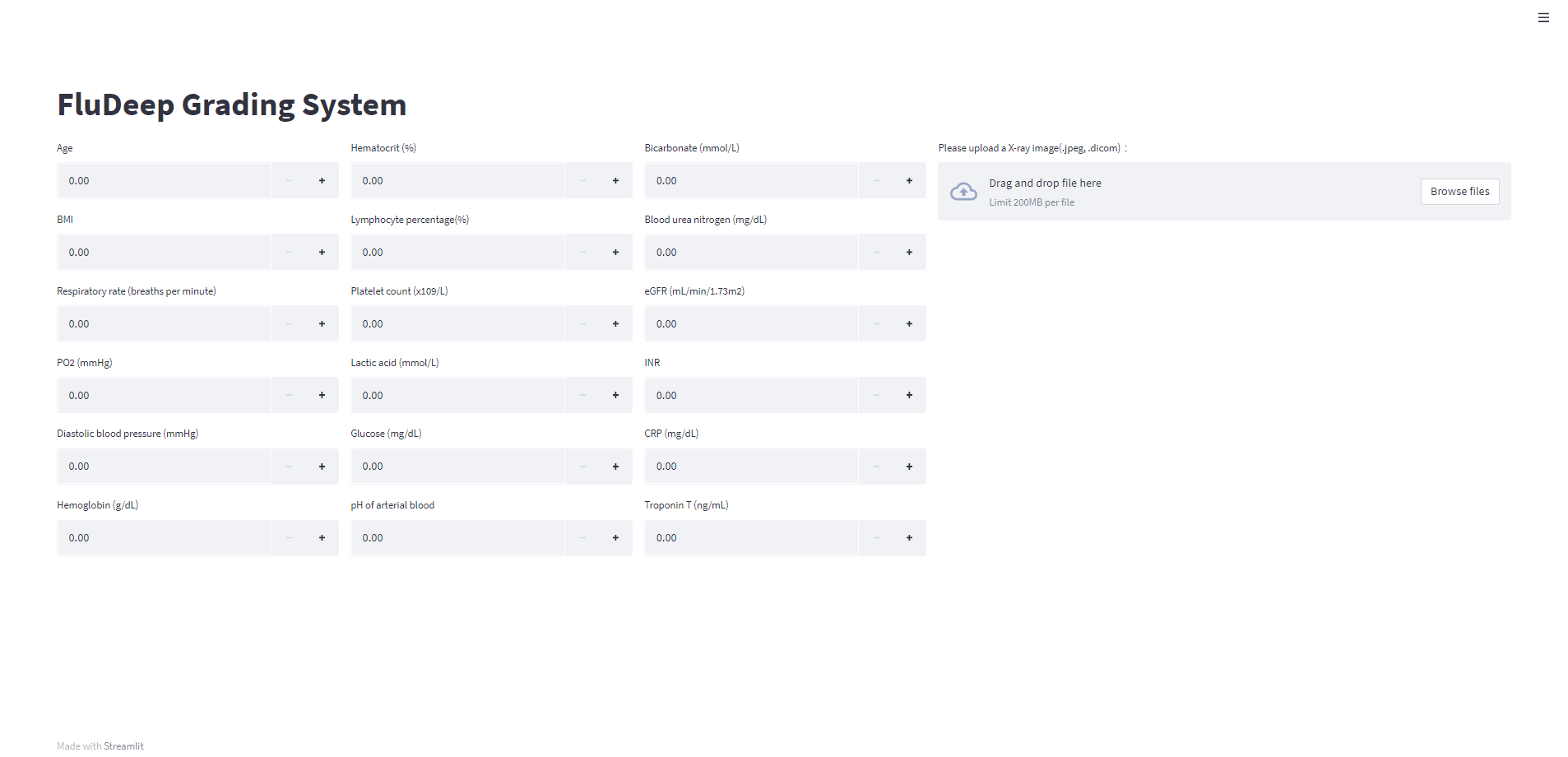
**Conclusion**

**Our study presents the first deep learning-based model to quantify the severity of flu chest x-ray images (FluDeep-XR) and the first AI model that combines radiological images and clinical data (FluDeep) for an accurate prediction of 30-day mortality of influenza pneumonia. This prognosis prediction system could allow physicians to early recognize high-risk mortality influenza patients that need prompt intervention to prevent disease worsening and better allocate resources, especially during influenza outbreaks. Moreover, our study supports the idea of integrating data of different** types **together to obtain more accuracy, imitating physician’s reasoning. This approach could provide valuable insight for future further development and application of fusion models as prognostic prediction tools.**

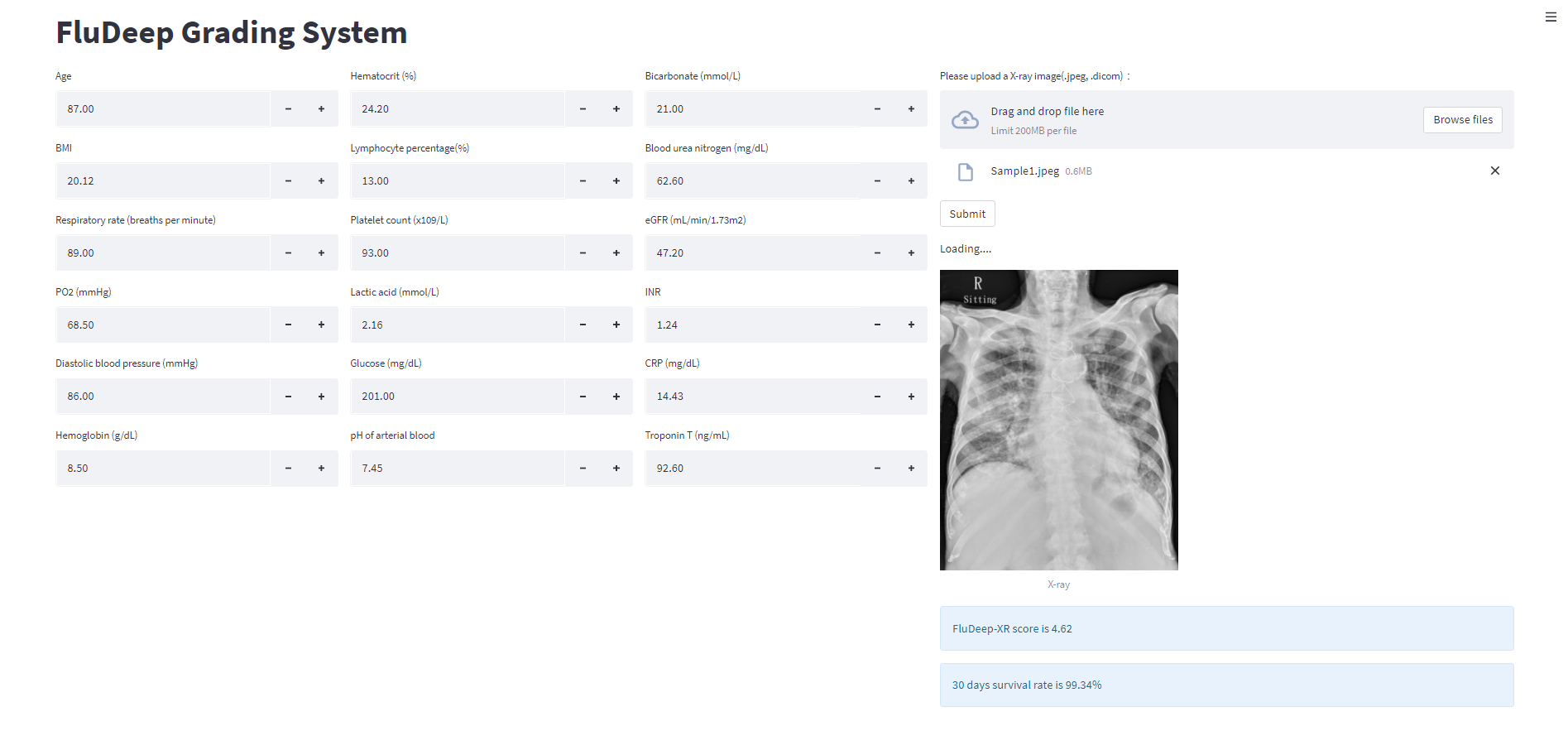
**Appendix**

****

* Appendix figure 1. The Decision tree of the FluDeep prediction system.



* Appendix figure 1. The screenshot for the graphic interface of FluDeep prediction system.



* Appendix figure 2. The screenshot for the result of the graphic interface of the FluDeep prediction system.

| DeepFlu-XR | | | | |
| --- | --- | --- | --- | --- |
| cutoff | F1 | sensitivity | specificity | Youden index |
| >1 | 0.1070 | 1.0000 | 0.0000 | 0.0000 |
| **>2** | **0.1548** | **0.9231** | **0.4009** | **0.3240** |
| >3 | 0.1653 | 0.7692 | 0.5484 | 0.3176 |
| >4 | 0.1758 | 0.6154 | 0.6774 | 0.2928 |

* Appendix Table 1. The performance of FluDeep-XR with different cutoffs on our patients.

| MuLBSTA score | | | | |
| --- | --- | --- | --- | --- |
| cutoff | F1 | sensitivity | specificity | Youden index |
| >0 | 0.1145 | 1.0000 | 0.0737 | 0.0737 |
| >1 | 0.1145 | 1.0000 | 0.0737 | 0.0737 |
| >2 | 0.1311 | 0.9231 | 0.2719 | 0.1950 |
| >3 | 0.1326 | 0.9231 | 0.2811 | 0.2042 |
| >4 | 0.1096 | 0.6154 | 0.4240 | 0.0394 |
| >5 | 0.1143 | 0.6154 | 0.4516 | 0.0670 |
| >6 | 0.1553 | 0.6154 | 0.6221 | 0.2375 |
| **>7** | **0.1633** | **0.6154** | **0.6452** | **0.2606** |
| >8 | 0.1739 | 0.4615 | 0.7696 | 0.2311 |
| >9 | 0.1724 | 0.3846 | 0.8157 | 0.2003 |
| >10 | 0.1818 | 0.3077 | 0.8756 | 0.1833 |
| >11 | 0.2222 | 0.3077 | 0.9124 | 0.2201 |
| >12 | 0.2759 | 0.3077 | 0.9447 | 0.2524 |
| >13 | 0.2308 | 0.2308 | 0.9539 | 0.1847 |
| >14 | 0.0870 | 0.0769 | 0.9585 | 0.0354 |
| >15 | 0.0000 | 0.0000 | 0.9908 | -0.0092 |
| >16 | 0.0000 | 0.0000 | 0.9908 | -0.0092 |
| >17 | 0.0000 | 0.0000 | 0.9954 | -0.0046 |
| >18 | 0.0000 | 0.0000 | 0.9954 | -0.0046 |

* Appendix Table 2. The performance of MuLBSTA scores with different cutoffs on our patients.

| 4C mortality score | | | | |
| --- | --- | --- | --- | --- |
| cutoff | F1 | sensitivity | specificity | Youden index |
| >0 | 0.1070 | 1.0000 | 0.0000 | 0.0000 |
| >1 | 0.1070 | 1.0000 | 0.0000 | 0.0000 |
| >2 | 0.1074 | 1.0000 | 0.0046 | 0.0046 |
| >3 | 0.1083 | 1.0000 | 0.0138 | 0.0138 |
| >4 | 0.1106 | 1.0000 | 0.0369 | 0.0369 |
| >5 | 0.1166 | 1.0000 | 0.0922 | 0.0922 |
| >6 | 0.1106 | 0.9231 | 0.1152 | 0.0383 |
| >7 | 0.1188 | 0.9231 | 0.1843 | 0.1074 |
| >8 | 0.1290 | 0.9231 | 0.2581 | 0.1812 |
| >9 | 0.1395 | 0.9231 | 0.3226 | 0.2457 |
| >10 | 0.1509 | 0.9231 | 0.3825 | 0.3056 |
| >11 | 0.1583 | 0.8462 | 0.4700 | 0.3162 |
| >12 | 0.1930 | 0.8462 | 0.5853 | 0.4315 |
| **>13** | **0.2558** | **0.8462** | **0.7143** | **0.5605** |
| >14 | 0.2373 | 0.5385 | 0.8203 | 0.3588 |
| >15 | 0.2381 | 0.3846 | 0.8894 | 0.2740 |
| >16 | 0.2667 | 0.3077 | 0.9401 | 0.2478 |
| >17 | 0.3200 | 0.3077 | 0.9631 | 0.2708 |
| >18 | 0.2353 | 0.1538 | 0.9908 | 0.1446 |
| >19 | 0.2667 | 0.1538 | 1.0000 | 0.1538 |
| >20 | 0.1429 | 0.0769 | 1.0000 | 0.0769 |

* Appendix Table 3. The performance of 4c-mortality scores with different cutoffs on our patients.

**Acknowledgements**

may

**Conflicts of Interest**

None declared.

**Abbreviations**

AI: artificial intelligence

ARDS: acute respiratory distress syndrome

AUROC: area under the receiver operator characteristics curve

BMI: body mass index

CDC: Centers for Disease Control and Infection

CNN: convolutional neural network

CRP: C-reactive protein

CXR: chest X-ray

DL: deep learning

ED: emergency department

eGFR: estimated glomerular filtration rate

ICU: intensive care unit

INR: international normalized ratio

LGBM: LightGBM model

ML: machine learning

PO2: arterial partial pressure of oxygen

**Reference**

1. Iuliano AD, Roguski KM, Chang HH, et al. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. Lancet. Mar 31 2018;391(10127):1285-1300. doi:10.1016/S0140-6736(17)33293-2

2. Thompson WW, Shay DK, Weintraub E, et al. Mortality associated with influenza and respiratory syncytial virus in the United States. JAMA. Jan 8 2003;289(2):179-86. doi:10.1001/jama.289.2.179

3. Chaves SS, Aragon D, Bennett N, et al. Patients hospitalized with laboratory-confirmed influenza during the 2010-2011 influenza season: exploring disease severity by virus type and subtype. J Infect Dis. Oct 15 2013;208(8):1305-14. doi:10.1093/infdis/jit316

4. Lafond KE, Nair H, Rasooly MH, et al. Global Role and Burden of Influenza in Pediatric Respiratory Hospitalizations, 1982-2012: A Systematic Analysis. PLoS Med. Mar 2016;13(3):e1001977. doi:10.1371/journal.pmed.1001977

5. Martinez A, Soldevila N, Romero-Tamarit A, et al. Risk factors associated with severe outcomes in adult hospitalized patients according to influenza type and subtype. PLoS One. 2019;14(1):e0210353. doi:10.1371/journal.pone.0210353

6. Mata-Marin LA, Mata-Marin JA, Vasquez-Mota VC, et al. Risk factors associated with mortality in patients infected with influenza A/H1N1 in Mexico. BMC Res Notes. Sep 11 2015;8:432. doi:10.1186/s13104-015-1349-8

7. Kalil AC, Thomas PG. Influenza virus-related critical illness: pathophysiology and epidemiology. Crit Care. Jul 19 2019;23(1):258. doi:10.1186/s13054-019-2539-x

8. Jain S, Kamimoto L, Bramley AM, et al. Hospitalized patients with 2009 H1N1 influenza in the United States, April-June 2009. N Engl J Med. Nov 12 2009;361(20):1935-44. doi:10.1056/NEJMoa0906695

9. Aviram G, Bar-Shai A, Sosna J, et al. H1N1 influenza: initial chest radiographic findings in helping predict patient outcome. Radiology. Apr 2010;255(1):252-9. doi:10.1148/radiol.10092240

10. Balbi M, Caroli A, Corsi A, et al. Chest X-ray for predicting mortality and the need for ventilatory support in COVID-19 patients presenting to the emergency department. Eur Radiol. Apr 2021;31(4):1999-2012. doi:10.1007/s00330-020-07270-1

11. Toussie D, Voutsinas N, Finkelstein M, et al. Clinical and Chest Radiography Features Determine Patient Outcomes in Young and Middle-aged Adults with COVID-19. Radiology. Oct 2020;297(1):E197-E206. doi:10.1148/radiol.2020201754

12. Kim HW, Capaccione KM, Li G, et al. The role of initial chest X-ray in triaging patients with suspected COVID-19 during the pandemic. Emerg Radiol. Dec 2020;27(6):617-621. doi:10.1007/s10140-020-01808-y

13. Chen L, Han X, Li YL, Zhang C, Xing X. FluA-p score: a novel prediction rule for mortality in influenza A-related pneumonia patients. Respir Res. May 8 2020;21(1):109. doi:10.1186/s12931-020-01379-z

14. Cheong CW, Chen CL, Li CH, et al. Two-stage prediction model for in-hospital mortality of patients with influenza infection. BMC Infect Dis. May 19 2021;21(1):451. doi:10.1186/s12879-021-06169-6

15. Chung JY, Hsu CC, Chen JH, et al. Geriatric influenza death (GID) score: a new tool for predicting mortality in older people with influenza in the emergency department. Sci Rep. Jun 18 2018;8(1):9312. doi:10.1038/s41598-018-27694-6

16. Guo L, Wei D, Zhang X, et al. Clinical Features Predicting Mortality Risk in Patients With Viral Pneumonia: The MuLBSTA Score. Front Microbiol. 2019;10:2752. doi:10.3389/fmicb.2019.02752

17. Hu CA, Chen CM, Fang YC, et al. Using a machine learning approach to predict mortality in critically ill influenza patients: a cross-sectional retrospective multicentre study in Taiwan. BMJ Open. Feb 25 2020;10(2):e033898. doi:10.1136/bmjopen-2019-033898

18. Grupo de Trabajo Gripe AGdlSEdMICyUC. Spanish Influenza Score (SIS): Usefulness of machine learning in the development of an early mortality prediction score in severe influenza. Med Intensiva (Engl Ed). Mar 2021;45(2):69-79. Spanish Influenza Score (SIS): utilidad del Machine Learning en el desarrollo de una escala temprana de prediccion de mortalidad en la gripe grave. doi:10.1016/j.medin.2020.05.017

19. Albaum MN, Hill LC, Murphy M, et al. Interobserver reliability of the chest radiograph in community-acquired pneumonia. PORT Investigators. Chest. Aug 1996;110(2):343-50. doi:10.1378/chest.110.2.343

20. Loeb MB, Carusone SB, Marrie TJ, et al. Interobserver reliability of radiologists' interpretations of mobile chest radiographs for nursing home-acquired pneumonia. J Am Med Dir Assoc. Sep 2006;7(7):416-9. doi:10.1016/j.jamda.2006.02.004

21. Carrell DS, Schoen RE, Leffler DA, et al. Challenges in adapting existing clinical natural language processing systems to multiple, diverse health care settings. J Am Med Inform Assoc. Sep 1 2017;24(5):986-991. doi:10.1093/jamia/ocx039

22. McCoy AB, Wright A, Eysenbach G, et al. State of the art in clinical informatics: evidence and examples. Yearb Med Inform. 2013;8:13-9.

23. Reamaroon N, Sjoding MW, Gryak J, Athey BD, Najarian K, Derksen H. Automated detection of acute respiratory distress syndrome from chest X-Rays using Directionality Measure and deep learning features. Comput Biol Med. Jul 2021;134:104463. doi:10.1016/j.compbiomed.2021.104463

24. Sjoding MW, Taylor D, Motyka J, et al. Deep learning to detect acute respiratory distress syndrome on chest radiographs: a retrospective study with external validation. Lancet Digit Health. Jun 2021;3(6):e340-e348. doi:10.1016/S2589-7500(21)00056-X

25. Li MD, Arun NT, Gidwani M, et al. Automated assessment of COVID-19 pulmonary disease severity on chest radiographs using convolutional Siamese neural networks. medRxiv. May 26 2020;doi:10.1101/2020.05.20.20108159

26. Wong A, Lin ZQ, Wang L, et al. Towards computer-aided severity assessment via deep neural networks for geographic and opacity extent scoring of SARS-CoV-2 chest X-rays. Sci Rep. Apr 29 2021;11(1):9315. doi:10.1038/s41598-021-88538-4

27. Zhu J, Shen B, Abbasi A, Hoshmand-Kochi M, Li H, Duong TQ. Deep transfer learning artificial intelligence accurately stages COVID-19 lung disease severity on portable chest radiographs. PLoS One. 2020;15(7):e0236621. doi:10.1371/journal.pone.0236621

28. Liang W, Liang H, Ou L, et al. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. JAMA Intern Med. Aug 1 2020;180(8):1081-1089. doi:10.1001/jamainternmed.2020.2033

29. Jiao Z, Choi JW, Halsey K. Prognostication of patients with COVID-19 using artificial intelligence based on chest x-rays and clinical data: a retrospective study (vol 3, pg 286, 2021). Lancet Digital Health. May 2021;3(5):E283-E283. doi:10.1016/S2589-7500(21)00060-1

30. Shamout FE, Shen Y, Wu N, et al. An artificial intelligence system for predicting the deterioration of COVID-19 patients in the emergency department. NPJ Digit Med. May 12 2021;4(1):80. doi:10.1038/s41746-021-00453-0

31. He K. Deep Residual Learning for Image Recognition. arXivorg. 2015;

32. Mingxing T. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 2020;doi:10.48550/arXiv.1905.11946

33. Simonyan K. Very Deep Convolutional Networks for Large-Scale Image Recognition. doi:doi.org/10.48550/arXiv.1409.1556

34. Huang G. Densely Connected Convolutional Networks. IEEE. 2017:2261-2269. doi:10.1109/CVPR.2017.243

35. Szegedy C. Rethinking the Inception Architecture for Computer Vision. IEEE. 6/2016 2016:2818-2826. doi:10.1109/CVPR.2016.308v

36. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. IEEE. 2017:1800-1807. doi:10.1109/CVPR.2017.195

37. Deng J. ImageNet: A large-scale hierarchical image database. IEEE. 2009:248-255. doi:10.1109/CVPR.2009.5206848

38. Lahat D. Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. Proc IEEE. 2015;doi:10.1109/JPROC.2015.2460697

39. Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med. Jun 2020;43(2):635-640. doi:10.1007/s13246-020-00865-4

40. Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. Pattern Anal Appl. 2021;24(3):1207-1220. doi:10.1007/s10044-021-00984-y

41. Rahimzadeh M, Attar A. A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. Inform Med Unlocked. 2020;19:100360. doi:10.1016/j.imu.2020.100360

42. Yu H, Feng Z, Uyeki TM, et al. Risk factors for severe illness with 2009 pandemic influenza A (H1N1) virus infection in China. Clin Infect Dis. Feb 15 2011;52(4):457-65. doi:10.1093/cid/ciq144

43. Louie JK, Acosta M, Samuel MC, et al. A novel risk factor for a novel virus: obesity and 2009 pandemic influenza A (H1N1). Clin Infect Dis. Feb 1 2011;52(3):301-12. doi:10.1093/cid/ciq152

44. Jain S, Benoit SR, Skarbinski J, Bramley AM, Finelli L, Pandemic Influenza AVHIT. Influenza-associated pneumonia among hospitalized patients with 2009 pandemic influenza A (H1N1) virus--United States, 2009. Clin Infect Dis. May 2012;54(9):1221-9. doi:10.1093/cid/cis197

45. Shi SJ, Li H, Liu M, et al. Mortality prediction to hospitalized patients with influenza pneumonia: PO2 /FiO2 combined lymphocyte count is the answer. Clin Respir J. May 2017;11(3):352-360. doi:10.1111/crj.12346

46. Bellelli V, d'Ettorre G, Celani L, Borrazzo C, Ceccarelli G, Venditti M. Clinical significance of lymphocytopenia in patients hospitalized with pneumonia caused by influenza virus. Crit Care. Oct 29 2019;23(1):330. doi:10.1186/s13054-019-2608-1

47. Gupta A, Anjum, Gupta S, Katarya R. InstaCovNet-19: A deep learning classification model for the detection of COVID-19 patients using Chest X-ray. Appl Soft Comput. Feb 2021;99:106859. doi:10.1016/j.asoc.2020.106859

48. Khan AI, Shah JL, Bhat MM. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Comput Methods Programs Biomed. Nov 2020;196:105581. doi:10.1016/j.cmpb.2020.105581

49. Rajpurkar P. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. 2017;doi:10.48550/ARXIV.1711.05225

**A multimodal deep learning approach for outcome prediction in hospitalized influenza patients**

**Introduction**

Influenza is a major health issue worldwide, resulting in high morbidity and mortality especially among adults aged 65 years and older [ ]. Reports released by Centers for Disease Control and Infection (CDC) estimated over 5 million influenza-associated hospitalizations every year globally [ ] and 200 to 600 thousands of deaths every year globally [ ]. A number of risk factors have been identified to be associated with its complications, intensive care unit (ICU) admission and death [ ]. Predicting the outcome of influenza to enable risk stratification is important in regard to the proper allocation of limited medical resources. It could also help physicians make more personalized treatment plans for patients with influenza, and intervene early in the course of disease with more intensive monitoring for patients with high risk for mortality.

About 30-40% of patients hospitalized with influenza presented with acute pneumonia [ ]. Findings on chest x-rays have important predictive value for the severity of viral pneumonia and its progression to respiratory failure and death [ ]. However, current mortality prediction models for influenza only included clinical data as presented in the medical records, and did not utilize chest x-ray images. The major obstacle to including chest x-ray findings in the prediction models is that in practice, radiologists usually do not quantify the severity of pathological changes in the lungs with numbers [ ]. They only report qualitative terms describing the morphological changes and image characteristics, such as increased infiltration, consolidation, vascular engorgement, pulmonary edema, pleural effusion or cardiomegaly.

Based on clinical data from electronic medical records and patient databases alone, several prediction models and scoring systems for mortality from influenza have been developed, including logistic regression models [ ] and machine learning models [ ]. Unfortunately, without incorporation of chest x-ray images, they only exhibited suboptimal predictive performance for influenza mortality, with Area Under the Curve (AUC) falling between XX to OO… for [ ] and for [ ], respectively.

To address this problem, machine learning has been used to identify acute respiratory distress syndrome (ARDS), a late and serious complication of viral pneumonia, on chest x-rays [ ]. A model with convolutional neural network developed by a team at the University of Michigan was able to detect ARDS with an area under the receiver operator characteristics curve (AUROC) of 0·92 (95% CI 0·89–0·94) [ ]. Machine learning also enables quantification of the severity of chest x-ray images and prognostication of mortality from viral pneumonia [ ].

To date, machine learning has not been applied to determine the severity of influenza pneumonia on chest x-ray images, and there exists no mortality prediction model that incorporates both clinical data and chest x-ray images. Recently, two groups of researchers presented machine learning models based on the combination of chest x-ray images and clinical data to predict severity and clinical deterioration of COVID-19 infection [ ]. They both showed that adding information from chest x-ray images to clinical data improved the models’ overall performance, compared to models using clinical variables alone [ ].

In this study, we aimed to build and compare two machine learning models for prediction of mortality from influenza: one based on clinical data, and the other based on a combination of chest x-ray images and clinical data. The three image models used for chest x-ray images are: ARDS score, CVOID-19 severity score and tailored influenza pneumonia severity score.

Introduction

WHAT IS OUR OUTCOME: mortality

2016-2017 mortality 8-10%

2019 4-5%

1. Describe the medical and societal burden of influenza
2. Describe the importance of outcome prediction/risk stratification in the management of flu
3. Describe the suboptimal performance of prediction score based on the clinical data, including labs
4. Briefly summarize the performance of influenza mortality risk prediction score, including recent development in machine learning score
5. Chest X ray is an important diagnostic tool to predict the severity of influenza pneumonia. The current reporting system is based on description of morphological change/contexture change. The severity of the chest X ray image cannot be quantified by a radiologist.
6. The AI model can help quantify the risk of chest X ray image e.g. predict ARDS <https://www.thelancet.com/journals/landig/article/PIIS2589-7500(21)00056-X/fulltext>, or predict COVID-19 pneumonia severity https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7451075/
7. Specify the aim of the study, 1. Build a machine learning score on clinical data, 2. Test whether combining a clinical model or image model would improve the predictive accuracy. Three image models will be used, 1. ARDS score, 2. COVID-19 severity score. 3. Tailored influenza pneumonia severity score (reference)

Method

1. Describe the cohort inclusion and exclusion criteria
2. Describe the data component
3. Describe the machine learning for clinical model
4. Describe the machine learning for image severity model.
5. Describe the method for model fusion
6. Describe the statistical method a. descriptive statistics, b. Metrics of model accuracy c. metric for model comparison including NRI/ IDI, d, best cutoff with corresponding sen/spe, PPV/NPV

Results

Discussion

References