AI prognostication tool for severe community-acquired pneumonia and covid-19 respiratory infections

Han Leong Goh, Tian Yi Lu, Lin Zou, Clarence Peng Yu Ngoh, Zhiyu Wang, Juan Zhen Koh, Christine Gia Lee Ang, Zhiyan Fu, Andy Wee An Ta
Integrated Health Information System (IHiS)
Singapore, Singapore
{hanleong.goh, andy.ta}@ihis.com.sg

Charlene Jin Yee Liew, Jessica Lishan Quah, Venkataraman Narayan
Changi General Hospital (CGH)
Singapore, Singapore
{Charlene.liew.j.y, Jessica.quah.l.s}@singhealth.com.sg

ABSTRACT

As of 23 May 2020, 5.21 million cases of infection and 338,000 deaths worldwide had been reported. Ethical considerations in scarcity would suggest that hospital resources should be prioritized for patients who are most ill. These considerations are becoming more urgent in the current global COVID-19 pandemic due to an unprecedented surge for hospital care. COVID-19 has pushed healthcare systems to the edge and spur rapid development of AI health informatics solutions to fight against the pandemic. Many clinical prognostic scoring tools to aid clinicians in the disposition and severity assessment of COVID-19 have since been developed or proposed. However, most of these tools leverage on either radiology images or electronic medical records (EMR) data. In this paper, we presented a novel tool that augments Chest X-Ray images and electronic medical records data in a single end-to-end machine learning model. We also demonstrate that CAP (Community Acquired Pneumonia) patients have a similar severity characteristic with COVID-19 patients therefore a predictive model trained on CAP cohort is also applicable on COVID-19 patients. The final ensemble model developed has an AUC of 0.71 and accuracy of 0.768 when tested on real world clinical COVID-19 data, this performance is higher when compared to model developed using only Chest X-Ray or EMR data alone. In addition, we also shared the system architecture of a simple to implement standalone application that is well integrated with the radiology information system and picture archiving & communication system (RIS-PACS) clinical workflow.

CCS CONCEPTS

• Computing methodologies → Neural networks; • Applied computing → Health informatics.

KEYWORDS

Clinical Decision Support, pneumonia, COVID-19, Chest X-Ray, Electronic Medical Record, Machine Learning, Neural Network.

ACM Reference format:

Han Leong Goh, Charlene Jin Yee Liew, Jessica Lishan Quah, Tian Yi Lu, Lin Zou, Clarence Peng Yu Ngoh, Zhiyu Wang, Juan Zhen Koh, Christine Gia Lee Ang, Venkataraman Narayan, Zhiyan Fu and Andy Wee An Ta. 2020. AI prognostication tool for severe community-acquired pneumonia and covid-19 respiratory infections.

1 INTRODUCTION

Community-acquired pneumonia (CAP) is the leading cause of morbidity and mortality worldwide. There were estimated 6.8 million episodes of hospital admissions for CAP worldwide in 2015 [1]. This has resulted in estimated 1.1million in-hospital deaths, predominantly amongst older adults.

As CAP is a common disease, multimodal bedside clinical scoring tools to aid clinicians in disposition and severity assessment of CAP are well-validated [2]. These include CURB-65 and Pneumonia Severity Index, which estimate the 30-day mortality of a pneumonia episode. The severity of the disease may then be estimated, to prompt the clinician to consider the options of ambulatory care, inpatient care, or the use of critical care facilities for severely ill patients. However, these scoring systems are less robust in a geriatric population [3] and carry an intrinsic assumption that the availability of healthcare resources and facilities are limitless. Realistically, the peacetime availability of inpatient hospital beds and critical care units are dependent on the socioeconomic status of a country [4]. These considerations are becoming more urgent in the current global COVID-19 pandemic due to an unprecedented surge for inpatient care and critical care support [5].

As COVID-19 is a form of CAP, there is an unmet need for unique population-based automated CAP and COVID-19 assessment tools that can rapidly stratify disease severity. This would enable clinicians to efficiently triage patients focusing on local healthcare resource availability. The efficacy of deep learning-based algorithms based on chest imaging has been

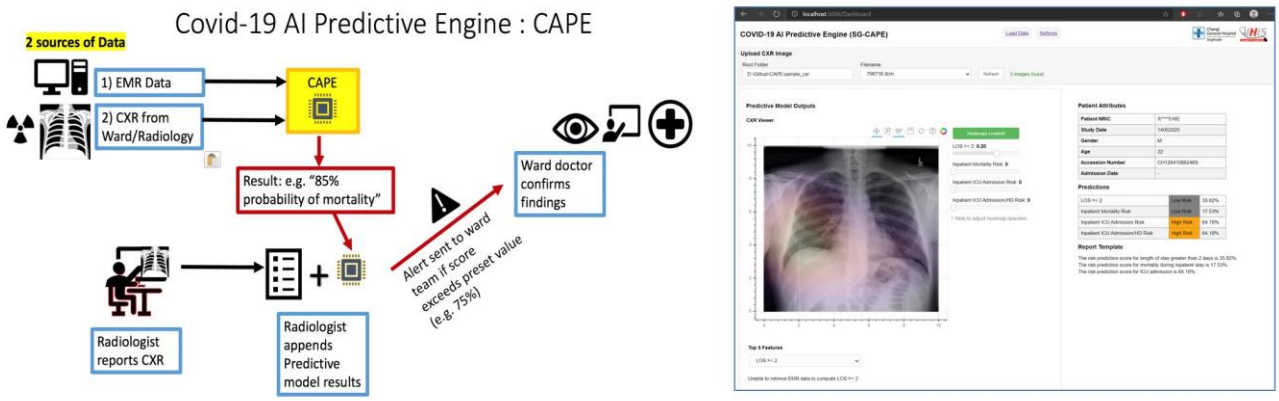

Figure 1: Proposed End-to-end AI application CAPE© that augment CXR and EMR data, risk score from CAPE© will serve as a decision support tool for radiologist, who will append the model result and their assessment in the structured report, which will be disseminated to clinicians downstream via EMR or RIS-PACS system

demonstrated to be useful for pneumonia diagnosis [6] and discrimination of viral from bacterial pneumonia [7]. Machine learning techniques utilizing EMR data could predict whether patients at emergency departments required critical care services [8], with the ROC of this algorithm outperforming traditional widely validated manual scoring systems. Disease-specific models have also been used to aid in disposition of patients with asthma and chronic obstructive pulmonary disease at the emergency department. [9]

In this paper, we present the first machine learning model that augment both Chest X-Ray image and Electronic Medical Health Record (EMR) data in an end-to-end AI application, which can discriminate between mild CAP/COVID-19 and severe CAP/COVID-19 on the first day of inpatient admission for clinical decision support. We developed CAP/COVID-19 AI Predictive Engine (CAPE) using 2,169 Chest X-Ray images and ~200k data points. We had launched the pilot run of CAPE© in Changi General Hospital (CGH), a tertiary acute hospital in Singapore since 9 May 2020.

The main contributions of this paper are as follows:

- We benchmark the performance of various models based on Chest X-Ray data, EMR data and Chest X-Ray + EMR data. We further give an insight on the relative importance of Chest X-Ray images compared with other variables from EMR. This allows clinical researchers to have a deeper understanding of the key predictors for CAP/COVID-19 patient prognosis.

- We proposed an easy to implement system architecture that integrates well with the radiologist workflow using conventional RISP-ACS systems. Flexibility in system architecture and good user interface are design considerations for user adoption and iteration of solution with agility and speed. This can potentially help fellow practitioners in their design of related CAP/COVID-19 AI applications.

- We also established the feasibility and observed the change in model performance when using model trained on CAP cohort patients tested on COVID-19 patients.

2 METHOD

Figure 1 gives an overview on the system architecture and user interface of CAPE© and how it is integrated into the clinician workflow. This section will elaborate on the machine learning techniques and statistical evaluation methods used. The models were implemented in Keras, version 1.3.0 29, lightGBM 2.3.2 and Scikit-learn, version 0.19.1 and Python, version 3.7 (Python Software Foundation).

2.1 Problem Formulation

The objective is to predict pneumonia patients with mild disease and a short inpatient length of stay. The point of prediction is on the day when patient was admitted to the ward, termed as Day 0. We defined it as binary classification problem with the target outcome label defined as patient discharge within 2 days of inpatient admission (excluding the day of admission). As the dataset is imbalanced, this was addressed by using a weighted binary cross entropy loss [9], which ensures that both classes contribute equally to the loss and weight updates will not be biased towards the majority class.

$$loss^i = loss^i_{pos} + loss^i_{neg}$$
$$loss^i_{pos} = -1 \times weight^i_{pos} \times y^i \times log(\hat{y}^i + \epsilon)$$
$$loss^i_{neg} = -1 \times weight^i_{neg} \times (1 - y^i) \times log(1 - \hat{y}^i + \epsilon)$$
where ϵ = a tiny positive number,
$loss_{pos}$ = the loss where the actual label is positive,
$loss_{neg}$ = the loss where the actual label is negative

2.2 Chest X-Ray Image – Convolution Neural Network

Transfer Learning is the use of pre-trained models that are either fine-tuned on the underlying data or used as feature extractors to aid in visual recognition tasks [10]. These models transfer the knowledge gained while learning generic features from large-scale image datasets like ImageNet to different tasks and are especially useful when there is insufficient training data available [11]. For this task, we used a pre-trained image classification network - Xception with ImageNet weights. Xception is an extension of the Inception architecture which replaces the standard Inception modules with depth-wise separable convolutions [12]. We applied transfer learning by freezing the first 12 blocks, except the batch normalization layers, and training only the last 2 blocks. This allows the network to apply the genetic features and learn features specific to Chest X-Rays. Finally, global average pooling was applied before a fully connected layer was added for the binary classification.

2.3 Electronic Medical Record – Gradient Boosting Machine (GBM)

We acquired ~100 independent variables from the EMR that are hypothesized to be related to severe disease in COVID-19 [13]. The tabular dataset was segmented into 4 primary categories: demographic characteristics, past hospital utilization information, past medical conditions and lab test result on the day of admission. In the feature engineering step, the predictors generated for the machine learning algorithms were based on literature review and input from clinicians.

The features derived from the utilization history includes variables like number of admissions, number of specialist outpatient clinic (SOC) visits, number of emergency department (ED) visits and number of admissions to medical ward in the past 2 years. For past medical conditions, we identified comorbidities based on the list in the Charlson Comorbidity Index (CCI) using inpatient diagnosis information in the past 2 years prior to the index admission. We further localized the International Classification of Diseases (ICD) code that is associated with CCI comorbidities by using Ministry of Health (Singapore)'s Chronic Disease Management Programme classification [14].

From the prevailing literature related to COVID19 [15-19], the following list of lab tests are used in the model, namely Albumin, Bilirubin, C reactive protein, Creatinine, Lymphocytes, Neutrophil and Troponin-T.

For the prediction using EMR tabular data we used LightGBM - a fast, distributed, high performance gradient boosting method based on decision tree algorithms. In many use cases and Kaggle competitions, LightGBM has shown to provide better performance than other gradient boosting models such as the popular XGBoost.

2.4 Ensemble the Mixed data

As shown in Figure 2, we experimented with different ensemble approaches to tackle the task: (i) Multiple inputs model where we encode the EMR data in a fully connected layer and concatenate this with the image model, (ii) Image and EMR data are trained independently and then combined into a single weighted risk score through logistic regression, (iii) The probabilities generated from image model is used as input to the gradient boosting tree.

For EMR data used in (i), we carefully included only features that were ranked high on the variable importance of independent EMR model. They are: troponin, average pro-calcitonin, average C-reactive protein, average bilirubin, age, CCI for myocardial infarction, CCI for cancer, CCI for coronary heart disease, and finally, length of stay for the past 12 months of INP, SOC and ED visits.

To date mixed data is still an open research area and as such we benchmark the performances of the different ensemble approaches to select the most effective mixed data model to deploy.

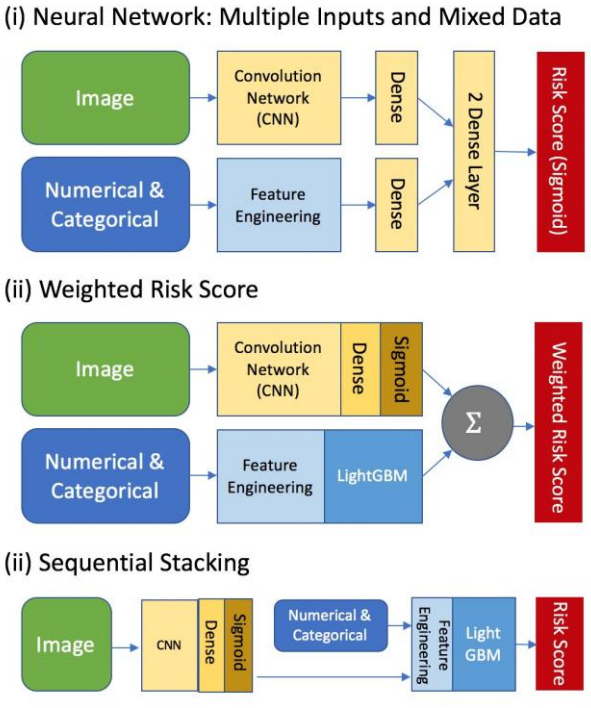

Figure 2: Various ensemble approaches experimented

2.5 Statistical Analysis

To determine model discrimination, the receiver operating characteristic (ROC) were used, with 95% confidence intervals by bootstrapping the observations on the test data. To understand how the different risk thresholds will affect the operation, performance metrics like sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and number of true and false positives were also generated. The true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) from confusion matrix were used to calculate:

- **Sensitivity** $= \frac{TP}{TP+FN}$
- **Specificity** $= \frac{TN}{TN+FP}$
- **Precision** $= \frac{TP}{TP+FP}$
- **Accuracy** $= \frac{(TP+TN)}{TP+TN+FP+FN}$

3 DATA AND PREPROCESSING

3.1 Data

A retrospective study cohort encompassed adult patients admitted to a tertiary acute hospital in Singapore from 1st January 2019 to 30th March 2020 was used to develop the predictive model. The inclusion criteria for the study cohort are patients admitted by Emergency Department with diagnosis as Pneumonia (using ICD-10 coding).

The study cohort consists of 3,048 Chest X-Ray images from 2,902 unique adult patients. The study cohort was de-identified, validated and tested with split-sample validation. For the training and validation set, we randomly selected 80% of inpatient episodes for model training and held out 20% for validation. The 20% validation set was used for tuning and optimizing hyperparameters such as dropout and learning rate.

Three separate sets "training", "validation" and "test" sets were used. The "training & validation" set were clustered based upon time, patients admitted from 1 January 2019 to 31 December 2019. The test set was created from patients admitted from 1 January 2020 to 31 March 2020. The test set spilt by calendar month was used to ensure temporal generalizability of the models.

In situation where there was overlap between the test set and the training & validation dataset, for example, a patient presented twice for a chest radiograph, once in June 2019 and once in February 2020, the record was moved from the test set to the training and validation set. This was to prevent the deep learning model from being tested on previously learnt training data.

For CAP and COVID-19 patients with X-Ray images, the corresponding EMR features were selected including demographic, past hospital utilization information, past medical conditions and lab test results on the day of admission.

3.2 Images preprocessing and augmentation

All images were preprocessed by center-cropping, resizing to dimensions of 244 x 244 pixels, followed by histogram equalization. All pixel values were scaled to between 0 and 1. Various image augmentation techniques were also applied to the training data, such as random transformations, zoom, brightness adjustment and random eraser.

4 RESULTS

4.1 Performance of the different ensembles on the COVID-19 and CAP Test Set

The highest accuracy and AUC were achieved by Ensemble (iii) model - a combination of CNN + LightGBM, utilizing both Chest X-Ray and EMR data (accuracy = 0.745, AUC = 0.767) as shown in Table 1 (Test set CAP). The model maintains a similar

Method of prediction	AUC	TP	FN	TN	FP	Sensitivity	Specificity	PPV	NPV	Accuracy
EMR	0.755	121	165	461	62	0.423	0.881	0.661	0.726	0.719
Chest X-ray	0.741	176	110	408	115	0.615	0.780	0.605	0.788	0.722
Ensemble (i) Neural Network Mixed data inputs	0.717	147	139	416	107	0.696	0.795	0.579	0.750	0.696
Ensemble (ii) Weighted risks score	0.769	146	140	453	70	0.510	0.866	0.676	0.764	0.740
Ensemble (iii) Sequential Stacking	0.767	157	129	446	77	0.549	0.853	0.671	0.776	0.745

Table 1: Model performance of different ensemble approach on CAP test set

Method of prediction	AUC	TP	FN	TN	FP	Sensitivity	Specificity	PPV	NPV	Accuracy
EMR	0.664	7	7	37	18	0.500	0.673	0.280	0.841	0.638
Chest X-ray	0.638	9	5	33	22	0.643	0.60	0.29	0.868	0.609
Ensemble (i) Neural Network Mixed data inputs	0.705	11	3	32	23	0.786	0.582	0.324	0.914	0.623
Ensemble (ii) Weighted risks score	0.692	7	7	44	11	0.5	0.8	0.389	0.863	0.739
Ensemble (iii) Sequential Stacking	0.710	7	7	46	9	0.500	0.836	0.438	0.868	0.768

Table 2: Model performance of different ensemble approach on COVID19 test set

performance (accuracy = 0.768, AUC = 0.710) when tested on COVID-19 test set in Table 2.

4.2 Model interpretation

Figure 3 shows the top 10 variables importance from the Gradient Boosting Machine using Ensemble (iii) model. Figure 4 shows the activation heat map gradient-weighted class activation map (Grad-CAM) [20]. Heatmaps highlight the activated areas during decision making and renders more interpretability of the model, pointing physicians to regions that have abnormal physiological features.

4.2.1 Variables Importance. Variable importance chart indicates the relative weightage of the impact each independent variable has on the model in making the prediction. We found that the most important variable is the predictive risk score from Chest X-Ray images generated using neural networks.

4.2.2 Activation Heat maps. Figure 4 illustrates the Chest X-Ray of a man with a positive RT-PCR test for COVID-19 viral infection. Figure 4(A) Frontal chest radiograph upon admission, Figure 4(B) shows the deep-learning model heatmap superimposed over the image showing pneumonia-related features.

5 SYSTEM IMPLEMENTATION

5.1 Overview of Application Features

Having obtained positive results from the retrospective study, we successfully deployed these predictive models to be integrated with the Community Acquired Pneumonia (CAP) clinical pathway at the tertiary acute hospital. The purpose-built application, named **C**AP/**C**OVID-19 **AI** **P**redictive **E**ngine or **CAPE**©, is a specialized DICOM viewer that embeds the results of the predictive models, providing radiologists with clinical decision support. While we can leverage on many DICOM applications/module available in open source, however, substantial effort is still needed to contextualize and modify in ways to meet our requirements. We will elaborate on the various application modules as follows:

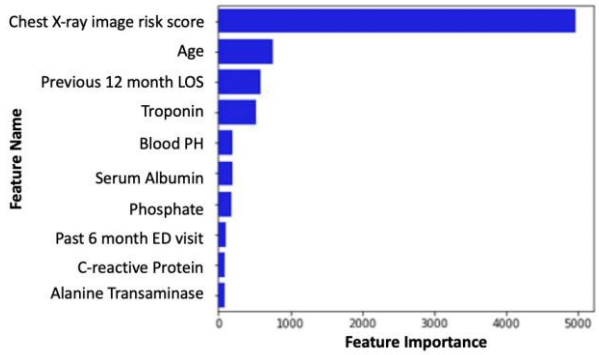

Figure 3: Top 10 variables importance

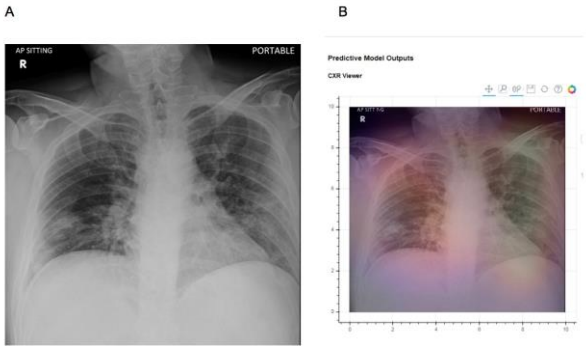

Figure 4: Generating explainable results of deep-learning interpretation by the utilization of heat maps

DICOM Load Module
Allows user to select and load DICOM images from the local filesystem.

Patient Attributes Module
Displays selected patient attributes from DICOM metadata and data from the EMR system. This includes demographic information and admission history.

Predictions Module
Displays the predicted risk scores for each outcome, shown as a percentage. If the risk score exceeds a user-configurable threshold, it will be visually highlighted as "High Risk". The default thresholds are set based on the ROC curves, but can be adjusted up or down by the clinicians to adjust for greater sensitivity or specificity.

Chest X-Ray and Class Discriminating Heat Map Module
Displays the Chest X-Ray images loaded from a standard DICOM file. We provide zoom, pan and image export functionalities. To provide interpretability for the predictive models, radiologists can choose to overlay a "class discriminating" heat map that highlights regions of the image contributing most to the predictions. The opacity of each overlay layer can be controlled independently.

EMR Top Features Module
Displays the top 5 features from EMR data contributing most to the predictions. Similar to heatmaps, these features provide interpretability to the radiologists about the key data points that were evaluated.

Text Reporting Module
Displays a configurable report that summarizes the results of the predictive models as a paragraph of text. This facilitates copy-pasting into other reporting tools, or radiology reports.

5.2 Deployment Approach

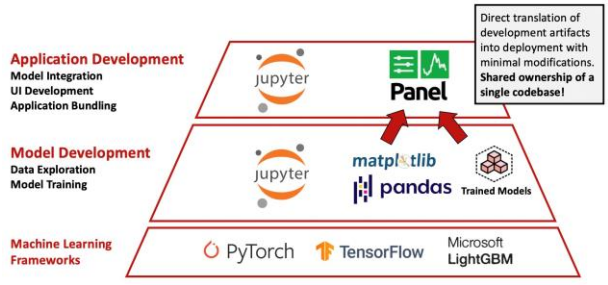

Figure 5: Tech stack to facilitate rapid time-to-deployment

CAPE© is implemented as a standalone desktop application that can be installed on radiology workstations, effectively deploying the predictive models "to the edge". This deployment approach was also preferable from a data security perspective, since patient data does not leave the hospital's network. We chose to implement the application in Python, which was a natural choice for integrating our Tensorflow-based DNN models. To reduce translation effort between development time (using Jupyter Notebooks) and deployment time, we utilized an open source tool, Panel, to layout widgets from IPython notebooks into a web-based application, as well as to add user interactivity. This allowed application development teams and data scientists to have shared ownership of a single code base, accelerating our time-to-delivery while facilitating more agile development practices.

6 DISCUSSION

The results of this study demonstrate that a predictive model that utilizing both Chest X-Ray and EMR data (AUC 0.710) performs better than a model developed using only Chest X-Ray (AUC 0.638) or EMR (AUC 0.664) data alone. Our review of different Ensemble approach shows that running CNN on Chest X-ray to generate a risk score and its subsequent input an independent variable to gradient boosting model give the best model performance results.

Notably, from the variable importance chart, we ascertain that Chest X-Ray data has a higher predictive power to discriminate between mild vs severe pneumonia. This finding can translate to significant improvements in the clinical workflow for CAP patients, whereby the utilization of a single modality: chest radiography with machine learning, can be a quick and relatively easier method for triaging and right siting care in resource-constrained countries.

In addition, we have demonstrated that a predictive model built using CAP patient data may be applied on patients with COVID-19. Both CAP and COVID-19 test cohorts shown similar ROC curves as shown in Figure 6.

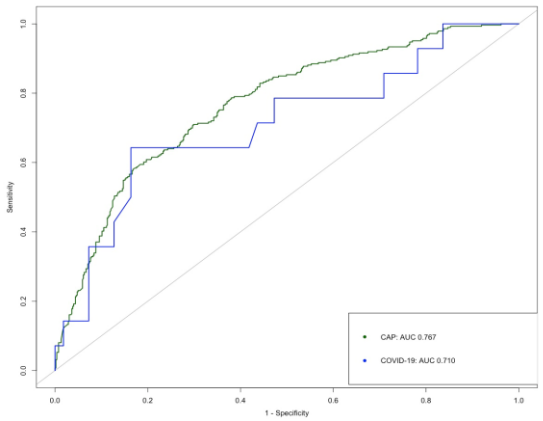

Figure 6: ROC curve of the model on CAP and COVID-19 patient cohort

7 CONCLUSION

We developed a model which can provide clinical decision support to discriminate between mild and severe forms of CAP and COVID-19 respiratory infections. Since 9 May 2020, we had deployed CAPE© in a tertiary acute hospital in Singapore, currently we are conducting prospective validation on the model. The key to our novel approach is in utilizing Chest X-Ray and EMR data in a single algorithm. To ensure its generalizability, we aim to validate the same model at other tertiary healthcare institutions. The authors welcome collaborators interested in deploying this software for use internationally. As such our immediate future work is to work with other Public Health Institutions to expand the pilot run for the model. From the system perspective, we plan to integrate a DICOM Service Class Provider (SCP) module to CAPE©, allowing the application to directly receive Chest X-Ray images from the radiography modalities. This will make it more seamless for radiologists to view those images with the risk scores in real-time.

ACKNOWLEDGMENTS

We thank Prof. Tan Chor Chuan from MOH Office for Healthcare Transformation (MOHT), Prof. Siau Chuin, Ms. Selina Seah, Prof. Teo Eng Kiong, A/Prof. Angeline Poh, Dr. Goh Min Liong, Mr. Tsai Koh Tzan, Ms. Sandhiya Ramanathan from Changi General Hospital (CGH), Mr. Sajda Rajesh Kishanchand, Mr. Ahmed Arshi from Phillips, Mr. Bruce Liang, Mr. Steve Lee, Mr. Chua Chee Yong, Mr. Mervin Chan, Mr. Cheng Lam Peh and Ms. Yanli Lei from Integrated Health Information Systems (IHiS) for the guidance and support.

REFERENCES

[1] Ting Shi, Angeline Denouel, Anna K Tietjen, et al. 2019. Global and Regional Burden of Hospital admissions for Pneumonia in Older Adults: A Systematic Review and Meta-Analysis. *The Journal of Infectious Diseases.* https://doi.org/10.1093/infdis/jiz053

[2] Otavio T. Ranzani, Leandro Utino Taniguchi, and Antoni Torres. 2018. Severity scoring systems for pneumonia: current understanding and next steps. *Current Opinion in Pulmonary Medicine* 24, 3 (2018), 227-236.

[3] Kyeongman Jeon and Hongseok Yoo. 2017. Functional status and mortality prediction in community-acquired pneumonia. *Respirology* 22, 7 (2017), 1400-1406.

[4] A. Rhodes, P. Ferdinande, H. Flaatten, B. Guidet, P.G. Metnitz, and R.P. Moreno. 2012. The variability of critical care bed numbers in Europe. *Intensive Care Medicine* 38, 10 (2012), 1647-1653.

[5] Yaseen M. Arabi, Srinivas Murthy, and Steve Webb. 2020. COVID-19: a novel coronavirus and a novel challenge for critical care. *Intensive Care Med* 46, 5 (2020), 833-836.

[6] Eui Jin Hwang, Sunggyun Park, Kwang-Nam Jin, et al. 2019. Development and Validation of a Deep Learning–Based Automated Detection Algorithm for Major Thoracic Diseases on Chest Radiographs. *JAMA Network Open* 2, 3 (2019).

[7] Daniel S. Kermany, Michael Goldbaum, Wenjia Cai, et al. 2018. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* 172, 5 (2018), 1122-1131.

[8] Da-Young Kang, Kyung-Jae Cho, Oyeon Kwon, et al. 2020. Artificial intelligence algorithm to predict the need for critical care in prehospital emergency medical services. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* 28, 1 (2020).

[9] Pranav Rajpurkar and Andrew Ng. 2020. *Coursera Deep learning AI medicine specialization.* Retrieved from https://www.deeplearning.ai/ai-for-medicine/.

[10] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. 2014. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition, *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Columbus, OH, 2014, pp. 512-519.

[11] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2017. *Deep learning.* MIT Press, Cambridge, MA.

[12] Francois Chollet. 2017. Xception: Deep Learning with Depthwise Separable Convolutions, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 1800-1807.

[13] COVID-19 Open Research Dataset Challenge (CORD-19). Retrieved from https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge

[14] Wee An Ta, Kheng Hock Lee, Sweet Fun Wong, et al. 2018. Development and implementation of nationwide predictive model for admission prevention: System architecture & machine learning, *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, Las Vegas, NV, 2018, pp. 303-306.

[15] Fei Zhou, Ting Yu, Ronghui Du, et al. 2020. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* 395, 10229 (2020), 1054–1062.

[16] Laure Wynants, Ben Van Calster, Gary S Collins, et al. 2020. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* (2020). https://doi.org/10.1136/bmj.m1328

[17] Dong Ji, Dawei Zhang, Jing Xu, et al. 2020. Prediction for Progression Risk in Patients with COVID-19 Pneumonia: the CALL Score. *Clinical Infectious Diseases* (2020). https://doi.org/10.1093/cid/ciaa414

[18] Jiao Gong, Jingyi Ou, Xueping Qiu, et al. 2020. A Tool for Early Prediction of Severe Coronavirus Disease 2019 (COVID-19): A Multicenter Study Using the Risk Nomogram in Wuhan and Guangdong, China. *Clinical Infectious Diseases* (2020). https://doi.org/10.1093/cid/ciaa443

[19] Lin Li, Lixin Qin, Zeguo Xu, et al. 2020. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. *Radiology* 296, 2 (2020).

[20] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2019. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision* 128, 2 (2019), 336–359.