Annals of Emergency Medicine

Developing and Validating Clinical Feature-Based Machine Learning Algorithms to Predict Influenza Infection in Influenza-like Illness Patients

--Manuscript Draft--

Manuscript Number:	
Full Title:	Developing and Validating Clinical Feature-Based Machine Learning Algorithms to Predict Influenza Infection in Influenza-like Illness Patients
Article Type:	Original Research
Section/Category:	Infectious Disease
Corresponding Author:	Kuan-Fu Chen, M.D., Ph.D. Chang Gung University Taoyuan, TAIWAN
First Author:	Shang-Kai Hung
Order of Authors:	Shang-Kai Hung
	Chin-Chieh Wu
	Avichandra Singh
	Jin-Hua Li
	Christian Lee
	Eric H. Chou
	Andrew Pekosz
	Richard Rothman
	Kuan-Fu Chen, M.D., Ph.D.
Additional Information:	
Question	**Response**
Annals adheres to the International Committee of Medical Journal Editor's specific requirements regarding authorship (see www.icmje.org). On behalf of all authors, I verify that all listed authors meet the www.icmje.org authorship requirements.	Yes
As corresponding author, I verify that everyone who contributed to this manuscript is either listed as an author or acknowledged as a contributor, and that the title page details any professional writing assistance or others paid to provide manuscript support.	Yes
Did anyone other than the listed authors have control over the data, over how the data were analyzed or interpreted, or over the wording or conclusions?	No

Has a prior version of this manuscript or study been submitted to Annals under a different submission sequence and/or manuscript number?	No
Have any data used in this study been included as part of another manuscript (whether published, submitted, or in preparation)?	No
If the manuscript involves human or animal subjects, does the methods section include verification of approval or exemption by an institutional review or ethics board? (Exceptions only for analyses of third party anonymized databases with pre-existing IRB approval or exemption.)	Yes
If the manuscript includes one of the following study designs, is it compliant with guidelines for such research? Study designs include: 1) Randomized controlled trials - CONSORT*; 2) Chart reviews - Kaji: AnnEmergMed 2014;64:292-298; 3) Diagnostic accuracy studies - STARD*; 4)Observational studies - STROBE*; 5) Systematic reviews - PRISMA*. *For details see: www.equator-network.org	Yes
If the manuscript is a clinical trial, was it pre-registered? If yes, please ensure that the registration number is listed on the title page. If not, please recognize that Annals may be unable to publish your study on this basis alone. (See the section “Obligation to register clinical trials” at www.icmje.org.)	Not applicable
Annals adheres to ICMJE guidelines regarding conflicts of interest; before completing this section please review them at www.icmje.org. For any aspect of the submitted manuscript, did any authors or their institutions receive:	None of the above for any author
Do any authors have any of the following financial relationships in the past 36 months with entities in the bio-medical arena that could be perceived to	None of the above for any author

influence, or that give the appearance of potentially influencing, what you wrote in the submitted work? Please err on the side of full disclosure.	
On behalf of all authors, I transfer all copyright ownership to the American College of Emergency Physicians (ACEP), in the event that the work is published in any medium. ACEP recognizes that works prepared by officers or employees of the US government, as part of their official government duties, are in the public domain. Authors reserve all proprietary rights (other than copyright) including patent rights relating to the subject matter of the article and the right for personal use in lectures, lecture notes, and exhibits. For all other uses, authors must seek written permission from ACEP and the publisher, Elsevier Inc. Annals provides an Open Access option, allowing authors to retain copyright and to make their article immediately freely accessible. Should you select this option after your article has been accepted, you will be billed for an open access publication fee; for more information, please refer to https://www.elsevier.com/journals/annals-of-emergency-medicine/0196-0644/open-access-options All submitted material that has been published elsewhere is accompanied by the written consent of the author and publisher. Photographs and any materials that contain personal information are accompanied by their signed release authorizing publication. By means of checking the box below, on behalf of all authors, I signify acceptance of the terms and conditions listed above. I have also been made aware of the journals' policies with respect to funding agency requirements such as the NIH Public Access policy, and the rapid	Agree

publication “Articles in Press” service.		
Suggested Reviewers:		
Opposed Reviewers:		
Funding Information:	Ministry of Science and Technology, Taiwan (109-2314-B-182-036)	Dr. Kuan-Fu Chen
	Chang Gung Memorial Hospital (CMRPG2K0211)	Dr. Kuan-Fu Chen
	Chang Gung Memorial Hospital (CMRPG2K0241)	Dr. Kuan-Fu Chen
	Chang Gung Memorial Hospital (CMRPG2L0281)	Dr. Kuan-Fu Chen
	National Institutes of Health (HHSN272201400007C)	Dr. Kuan-Fu Chen
	National Institutes of Health (75N93021C00045)	Dr. Kuan-Fu Chen
	Ministry of Science and Technology, Taiwan (109-2327-B-182-002)	Dr. Kuan-Fu Chen

Kuan-Fu Chen, MD, PhD
Associate Professor,
Department of Emergency Medicine
Chang-Gung Memorial Hospital

長庚紀念醫院
Chang Gung Memorial Hospital

5 Fu-Shin Street, Gueishan District, Taoyuan City, Taiwan, 333
Telephone: 886-3-3281200 #2505; Email: kfchen@cgmh.org.tw

January 24, 2022

To Whom It May Concern:

We submit our manuscript, entitled "**Developing and Validating Clinical Feature-Based Machine Learning Algorithms to Predict Influenza Infection in Influenza-like Illness Patients**." This original article, as submitted or its essence in another version, is not under consideration for publication elsewhere and will not be published elsewhere while under consideration by the *Annals of Emergency Medicine*. All authors have made substantive contributions to all the following: (1) the conception and design of the research or collected, analyzed and interpreted the data, (2) drafting the article or revising it critically for important intellectual content, (3) final approval of the version to be submitted. The authors do not have any conflict of interest to disclose. We follow the **T**ransparent **R**eporting of a multivariable prediction model for **I**ndividual **P**rognosis **o**r **D**iagnosis Statement (TRIPOD) recommendation to report this study.

In this prospective international, multi-center cohort study, we comprehensively recruited 2,189 patients with influenza-like illness (ILI) in the emergency departments. A predesigned questionnaire was used to thoroughly obtain the participants' medical history, travel and exposure history, symptoms, and signs. We systematically evaluated seven machine learning (ML) algorithms in predicting influenza infection in ILI patients. XGBoost had the best performance, achieving a strong 0.82 area under the receiver operating characteristic curve. Importantly, this model outperformed traditional scoring systems that currently guide clinical decision making. Furthermore, we validated body temperature, cough, and rhinorrhea as significant predictors of influenza infection, and identified controversial or new features like sore throat, week of the season, pulse rate, and oxygen saturation as important variables. We believe this article would contribute to future risk stratification and ML modeling in the field of emergency medicine.

We appreciate your consideration of this manuscript.

Sincerely,

Kuan-Fu Chen, MD, PhD
Associate Professor,
Department of Emergency Medicine
Chang-Gung Memorial Hospital

5 Fu-Shin Street, Gueishan District, Taoyuan City, Taiwan, 333
Telephone: 886-3-3281200 #2505; Email: kfchen@cgmh.org.tw

Kuan-Fu Chen, MD, PhD, Associate Professor,
Department of emergency medicine, Chang Gung Memorial Hospital and Chang Gung University

Developing and Validating Clinical Features-Based Machine Learning Algorithms to Predict Influenza Infection in Influenza-like Illness Patients

Authors:

Shang-Kai Hung, MD[1], mm200411800@cgmh.org.tw

Chin-Chieh Wu, PhD[2], wujinja@gmail.com

Avichandra Singh, PhD[2], avichandra0420@gmail.com

Jin-Hua Li[2], anna71729@gmail.com

Christian Lee, MSc[3], christian.lee@bswhealth.org

Eric H. Chou, MD[3], eric.chou@bswhealth.org

Andrew Pekosz[4], apekosz1@jhu.edu

Richard Rothman[5], rrothma1@jhmi.edu

Kuan-Fu Chen, MD, PhD[1,2,6] kfchen@cgmh.org.tw

Author Affiliations:

[1] Department of Emergency Medicine, Chang Gung Memorial Hospital, Linkou, Taiwan

[2] Clinical Informatics and Medical Statistics Research Center, Chang Gung University, Taoyuan, Taiwan

[3] Department of Emergency Medicine, Baylor Scott and White All Saints Medical Center, Fort Worth, TX, USA

[4] W. Harry Feinstone Department of Molecular Microbiology and Immunology, The Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205 USA,

[5] Department of Emergency Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205 USA,

[6] Department of Emergency Medicine, Chang Gung Memorial Hospital, Keelung, Taiwan

Corresponding Author:

Kuan-Fu Chen, MD, PhD

***Address for reprints:**

Department of Emergency Medicine, Chang Gung Memorial Hospital, Chang Gung University.

No. 5, Fu-Shin Street, Gueishan Village,

Taoyuan, Taiwan 333

TEL: 886-3-3281200 #2505)

Email: kfchen@cgmh.org.tw

Meetings: Parts of this research have been presented in the American College of Emergency Physicians Conference (ACEP) at the San Diego Conference Center, October 1st – 4th, 2018, San Diego.

Funding sources/Disclosures: This study was supported by the Ministry of Science and Technology and Chang Gung Memorial Hospital in Taiwan (MOST 109-2314-B-182-036, MOST 109-2327-B-182-002, CMRPG2K0211, CMRPG2K0241 and CMRPG2L0281) and National Institutes of Health in the US (HHSN272201400007C and 75N93021C00045). The funder of this study had no role in study design, data collection, data analysis, data interpretation, writing of the report, or in the decision to submit for publication. The authors had full access to all the data, and the corresponding author was responsible for the decision to submit for publication.

Conflict of interest: There are no conflicts of interest to declare.

Word Count: 3487

Author Contribution: Conceptualization: SKH, CCW, KFC; Data curation: CCW and AS; Formal analysis: AS, CCW, and KFC; Figure preparation: CCW; Investigation: KFC; Writing-original draft: SKH and KFC; Writing- review & editing: SKH and KFC

Running Title: Machine Learning Algorithms to Predict Influenza

Data Availability Statement: The datasets generated and analyzed during the current study are available at https://github.com/wujinja-cgu/FLu_ML_prediction.

Acknowledgments: This research was supported by Chang Gung Memorial Hospital. We are thankful to our colleagues who provided their expertise, which greatly assisted the research, although they may not agree with all the interpretations provided in this paper.

Developing and Validating Clinical Feature-Based Machine Learning Algorithms to Predict Influenza Infection in Influenza-like Illness Patients

Abstract

Objective:

Seasonal influenza poses a significant risk, and patients can benefit from early diagnosis and treatment. However, underdiagnosis and undertreatment remain widespread. We developed and compared clinical feature-based machine learning (ML) algorithms that can accurately predict influenza infection in emergency departments (EDs) among patients with influenza-like illness (ILI).

Methods:

We conducted a prospective cohort study in five EDs in the US and Taiwan from 2015 to 2020. Adult patients visiting the EDs with symptoms of ILI were recruited and tested by real-time RT-PCR for influenza. We evaluated seven ML algorithms and compared their results with previously developed clinical prediction models.

Results:

Out of the 2,189 enrolled patients, 1,104 tested positive for influenza. The eXtreme Gradient Boosting achieved superior performance with an area under the receiver operating characteristic curve of 0.82 (95% confidence interval [CI]=0.79–0.85), with a sensitivity of 0.92 (95% CI=0.88–0.95), specificity of 0.89 (95% CI=0.86–0.92), and accuracy of 0.72 (95% CI=0.69–0.76) in the testing set over cut-offs of 0.4, 0.6 and 0.5, respectively. These results were superior to those of previously proposed clinical prediction models. The model interpretation revealed that body temperature, cough,

rhinorrhea, and exposure history were positively associated with and the days of illness and influenza vaccine were negatively associated with influenza infection. We also found the week of the influenza season, pulse rate, and oxygen saturation to be associated with influenza infection.

Conclusions:

The clinical feature-based ML model outperformed conventional models for predicting influenza infection.

Keywords: Influenza infection, Machine learning, Influenza-like illness, Prediction model

Introduction

Background

Seasonal influenza poses a significant risk. In 2017, over 300,000 deaths resulted from influenza-associated respiratory diseases, and nearly 10 million patients were hospitalized because of influenza-associated lower respiratory tract infections worldwide.[1, 2] Influenza was also the principal infectious disease in Europe between 2009–2013.[3] Furthermore, influenza poses an even more significant threat to low-income countries in sub-Saharan Africa, western Pacific, and Southeast Asia.[2]

The CDC recommends early antiviral medication, especially for patients with complications or those requiring hospitalization for influenza infection.[4] Early (≤ 48 hours after illness onset) antiviral treatment should also be considered for otherwise healthy symptomatic outpatients to reduce the duration of symptoms. Timely diagnosis and early administration of antiviral agents decrease the illness duration, viral transmission, symptom severity, hospitalization risk, antibiotic usage, and mortality rate.[5, 6] However, according to a 2015 study, only 29% of the laboratory-confirmed influenza infection patients who were hospitalized or presented to emergency departments (EDs) were clinically diagnosed with fever and respiratory symptoms.[7-10] Another study published in 2019 found that over 60% of influenza-like illness (ILI) patients who were presented to primary care settings with posthoc laboratory-confirmed influenza infection did not initially receive an antiviral agent.[6] This discrepancy is even more widespread in high-risk groups that would benefit significantly from prompt treatment.[5, 6, 11] Therefore, influenza's undertreatment is common in EDs and primary care settings.

Importance

Diagnosing influenza is clinically challenging. The general symptoms often overlap with other respiratory virus infections. The presentations vary among patients of different age groups and in various clinical settings.[12, 13] Therefore, many clinicians rely on the rapid influenza diagnostic test (RIDT) to screen for potential influenza infection in EDs and other primary care settings.[14, 15] However, because its low sensitivity, RIDT is not recommended in the updated CDC guidelines anymore.[16] Although other laboratory methods with higher accuracy, namely RT-PCR assays, rapid molecular assays other nucleic acid amplification tests, and viral cultures, are available, clinical gestalt is still required to direct the diagnostic pathway for a sustainable health care environment.[17] However, previous studies demonstrated that sole reliance on clinical diagnosis had poor sensitivity.[18-20] Clinical prediction models that incorporate several different symptoms may be helpful but are still insufficient with a wide range of sensitivities (36–80%) and specificities (78–98%).[18-20]

Therefore, researchers are turning to machine learning (ML) algorithms, which are objective and have replicable approaches, for integrating multiple features to improve diagnostic ability in conditions such as sepsis, urinary tract infection, acute myocardial infarction, and congestive heart failure.[21-25] Previous studies have shown that ML algorithms outperform expert-built classifiers in predicting the influenza infection.[26] However, these studies were either conducted retrospectively or only used restrictive feature sets with no reliable validations.[19, 20, 26-28]

Goals of this investigation

The primary aim of this study was to develop and compare clinical feature-based ML algorithms that can predict influenza infection among patients with ILI in EDs. The

secondary aim was to compare the performance of clinical feature-based ML algorithms with previously developed clinical prediction models.

Methods

Study design and setting

We conducted a prospective cohort study in two hospitals in the US and three hospitals in Taiwan at the Centers of Excellence for Influenza Research and Surveillance. The Institutional Review Board of Johns Hopkins University (IRB00135664, IRB00041233, IRB00141101, IRB00052743, and IRB00091667) and the Chang Gung Medical Foundation (201406930B0) approved the study. The hospitals where the patients were enrolled ranged from community hospitals to medical centers in both the suburbs and the metropolitan areas. Appendix 1 provides a detailed description of these hospitals. To report our study, we followed the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis.[29]

Participant enrollment and data collection

Adult patients were presented to the study EDs hospitals between 2015 and 2020 with ILI symptoms modified from the WHO were eligible for this study. ILI was defined as a documented or reported fever and any of the three respiratory symptoms, including cough, headache, or sore throat, within the past seven days before the ED visit.[30] Patients who could not provide written informed consent, were currently incarcerated, or were previously enrolled in the study during the same influenza season were excluded. After informed consent, patient demographic information, comorbidities, history of exposure to a confirmed influenza infection patient in the past five days, travel, vaccination, and medication history, initial vital signs at triage, date of ED visit, and clinical symptoms were collected by research coordinators. The coordinators also reviewed the medical history in the electronic medical record and obtained a nasopharyngeal (NP) swab for the influenza PCR test. The PCR was

performed using a next-generation, fully automated, and integrated system, Cepheid® Xpert Flu Assay multiplex real-time PCR (Cepheid), with an overall sensitivity and specificity of 98–100%.[31] Additionally, external features, including the dominant viral subtype of the season, the week of the influenza season, and visits during the influenza season, were also collected for model development.

Data preprocessing and partitioning

We first randomly partitioned our dataset into a 75% training dataset and a 25% testing dataset, stratified by influenza infection status. The training dataset was used for feature selection and model development. We performed ten-fold cross-validations on the training dataset to fine-tune the hyperparameters. The testing dataset was kept aside for performance validation. To facilitate model training, we normalized the continuous features.

Feature selection

To extract important candidate features and remove noisy or redundant features that may result in an inefficient, impractical, or overfitting model, we adopted a wrapper method with the Boruta algorithm to rank the predictive influenza infection features. We repeated the Boruta algorithm 300 times for practical reasons and chose the top 20 ranked features.

Machine learning-based models

For model development, we evaluated the seven ML algorithms, including tree- and ensemble-based models: eXtreme Gradient Boosting (XGBoost), conditional random forest (CForest), random forest (RF), and RANdom forest GEneRator

(RANGER); a distance-based model: support vector machine (SVM); neural network-based models: artificial neural networks (ANN) and deep learning models with two, three, and four layers. Appendix 2 provides a detailed description and the specific hyperparameters for each algorithm.

Performance evaluation and interpretation

After model development, the candidate models and the previously developed clinical prediction models developed by Zimmerman *et al.*, Anderson *et al.*, and Dugas *et al.* were compared using the testing dataset.[19, 20, 32] Performance was evaluated on the area under the receiver-operating characteristic curves (AUROC). We further adopted modifiable cut-offs for the candidate ML algorithm to evaluate corresponding sensitivities and specificities. To interpret the ML models, we used the Shapley additive extension (SHAP) values to illustrate the direction and strength of the selected features in the final model. Additionally, we plotted a calibration curve and estimated the calibration by the Pearson Correlation for goodness-of-fit diagnosis, which enabled us to qualitatively compare the predicted probability of influenza infection to the empirical probability.

Statistical analyses

Demographics and clinical characteristics of recruited patients are presented as mean (standard deviation) or median (interquartile range) for continuous features and counts and percentages for discrete features. ML modeling was performed using R (version 4.0.1, R Foundation for Statistical Computing, Vienna, Austria) with the Boruta[33] and Caret packages.[34] Deep learning modeling was performed using Python (Python Software Foundation, Wilmington, DE) with sequential, dense, and dropout

packages. Prediction models were compared using R with the DeLong method of the pROC package.[35] All statistical tests were two-sided, and statistical significance was defined as a *p*-value of <0.05.

Subgroup analyses

We evaluated the final model in the subgroups of different countries (Taiwan and the US) and different subtype-dominant seasons (H1N1 and H3N2). To compare model performance in different subgroups, we applied the following two methods: first, we applied the model developed from the overall training dataset to the testing set of each subgroup; then retrained the model in different subgroups of the training dataset and then evaluated model performance in different subgroups of the testing set.

Results

Patient characteristics

Among 2,189 patients with ILI recruited from 2015 to 2020, 1,104 patients tested positive for influenza infection (50%). These patients were over 40 years (median; 50 and IQR: 29–54) and were women (55%, Table 1 and Appendix 3). Patients with influenza infection were older and more likely to have a history of exposure to a confirmed influenza infection patient in the past five days and receive an antiviral agent within 30 days. They were also less likely to receive an influenza vaccine that year and to have taken antibiotics within the past 30 days (Table 1), but were more likely to visit the ED earlier and visit during the influenza season. Patients with influenza infection had higher body temperatures, pulse rates, respiratory rates, and lower oxygen saturation at triage. They had significantly more respiratory symptoms, including a cough, with or without sputum, a sore throat, rhinorrhea, shortness of breath, and wheezing, but less nausea and stomach pain.

Clinical feature-based machine learning algorithms

We selected the top 20 most important clinical features from 115 features using the wrapper with 300 iterations (Figure 1 and **Appendix 4**), including patient demographics (body height and weight), past medical history (chronic lung disease and influenza vaccination), influenza infection-related history (travel history, exposed to human with confirmed influenza infection within the past five days, antiviral agent administered within 30 days before the hospital visit, the week of the influenza season, and visit during influenza season), symptoms (day of illness, cough, cough with sputum, rhinorrhea, sore throat, and sinus pain), and signs (body temperature, pulse rate, oxygen saturation, systolic blood pressure, and respiratory rate).

We then used these clinical features to develop and compare seven ML algorithms (Table 2). The XGBoost model performed the best in the testing set (accuracy: 0.72 (95% confidence interval [CI]=0.68–0.76); AUROC, 0.82 (95% CI=0.79–0.85); final hyperparameter setting is reported in Appendix 2). Among different cut-offs, XGBoost had a sensitivity of 0.92 (95% CI=0.88–0.96), specificity of 0.89 (95% CI=0.86–0.92) and accuracy of 0.72 (95% CI=0.69–0.76) in the testing set over cut-offs of 0.4, 0.6 and 0.5, respectively (Appendix 5). The calibration plot of the XGBoost algorithm revealed a good correlation (Pearson Correlation coefficients = 0.97) between the predicted and observed probabilities of influenza infection in the testing dataset (Appendix 6).

The SHAP summary plot in Figure 1 shows that body temperature, cough, rhinorrhea, and exposure history were highly positively associated with influenza infection in the final XGBoost model. In contrast, the day of illness and a history of influenza vaccination were negatively associated with influenza infection. In addition, the week of the influenza season, pulse rate, and oxygen saturation were strongly associated with influenza infection.

Comparison with clinical prediction models

We then compared the performance of the XGBoost model with that of previously proposed models for predicting influenza infection among patients with ILI. Our XGBoost model performed significantly better than the other models proposed by Zimmerman *et al.*, Anderson *et al.*, and Dugas *et al.* in the testing dataset (AUROC, 0.60, 0.61, and 0.65, respectively; all p values < 0.001; Table 3 and Figure 2). We also found that the model proposed by Zimmerman *et al.* has the best sensitivity of 0.96

(95% CI=0.93–0.98) but the worst specificity (0.23 [95% CI=0.19–0.28]) and AUROC (0.60 [95% CI=0.56–0.64]).

Subgroup analysis

Applying the developed model from the overall training dataset to each subgroup, we found that the XGBoost model performed better in the Taiwan subgroup than in the US (AUROC, 0.78 (95% CI=0.72–0.84) vs. 0.72 (95% CI=0.59–0.85), p=0.004) but similarly between H1N1 and H3N2 dominant seasons (0.84 [95% CI=0.79–0.90] vs. 0.81 (95% CI=0.75–0.87), p=0.062). Additionally, we retrained the model separately for different subgroups. The performance of the XGBoost models improved slightly in the US subgroup (0.83 [95% CI=0.79–0.88]) but not in the Taiwan subgroup and different subtype dominant seasons (Appendix 7).

Implementation

We also developed an applet that could be easily linked with electrical medical records to enhance clinical utility, available at the following hyperlink. (https://cgmher.shinyapps.io/shinyapp_for_flu_prediction/, **Appendix 8**).

Limitations

Here are some limitations of our research. First, the model performance based on the AUROC differed between the two subgroups of recruiting countries. The differences in the retrained models for these subgroups remained. We hypothesized that the disparity was because of different dominant season-specific influenza viral subtypes or different respiratory pathogen compositions between the enrolled sites.[36-38] However, the subgroup analysis of varying subtype dominant seasons was almost similar. Models must be retrained or updated in different subgroups with larger sample sizes. Second, previous research has shown that up to 40% of influenza infection patients, particularly the elderly, may be afebrile.[39, 40] The inclusion criteria derived from the modified WHO ILI definition would invariably result in a different study population. We, therefore, caution readers about selecting patient groups before applying our findings. Third, we used up to 20 features for prediction, more than the ILI definitions or traditional clinical prediction models. However, most of these features are simple and easy to integrate with existing electronic medical record systems. With the assistance of modern information technology, we can still implement our model. Finally, we conducted the surveillance before the COVID-19 outbreak, which may impact applicability and should be used cautiously. The performance of the clinical features warrants further investigation in the new era of COVID-19.

Discussion

This is the first prospective binational multicenter study to use clinical feature-based ML algorithms to predict influenza infection in patients with ILI. The XGBoost model outperformed the other seven ML algorithms and three previously developed clinical prediction models with an AUROC of 0.82 (95% CI=0.77–0.87) and achieved a sensitivity of 0.92 (95% CI=0.88–0.95) and specificity of 0.89 (95% CI=0.86–0.92) with different cut-offs. Body temperature, cough, rhinorrhea, and exposure history were positively associated with influenza infection, whereas days of illness and history of influenza vaccine were negatively associated with influenza infection. Furthermore, the infection was found to be strongly associated with the week of the influenza season, pulse rate, and oxygen saturation.

The major strength of our study is that we conducted a prospective multicenter study with a comprehensive collection of data from patients with ILI in front-line health care settings. The first crucial challenge in developing machine learning models is assembling a representative and diverse dataset. Previous efforts that relied on retrospective extraction from medical records may have missed essential features that were not routinely recorded, lowering the model's quality.[41] In our study, well-trained research coordinators used a predefined and comprehensive questionnaire for every enrolled patient to minimize potential recall bias and maximize model reliability. Another advantage is that it spanned four influenza seasons in multiple scaled EDs in two countries. This reduced the potential spectrum bias related to different age distributions, clinical presentations, and time distributions between viral subtypes.[42, 43] We also compared state-of-the-art ML algorithms with available clinical prediction

models designed to predict influenza infection, eliminating the need for a historical control that may represent an invalid comparison.

Two recent studies attempted to predict influenza infections using the classification and regression tree methods; however, their results were suboptimal. The first study conducted in 2016 by Zimmerman *et al.* included three features (fever, cough, and fatigue) in their prediction model and found a sensitivity of 84%, specificity of 49%, and AUROC of 0.69.[20] Anderson *et al.*(2018) incorporated four features, including cough, rhinorrhea, chills, and body aches, to achieve a sensitivity of 52.1%, specificity of 82.9%, and AUROC of 0.689.[19] However, in our testing data set, our XGBoost significantly outperformed these two models (AUROC, 0.82, 0.60, and 0.61, respectively; both $p < 0.001$). In contrast, we found that fatigue, chills, and body aches were similar between patients with ILI, and with and without influenza infection. We hypothesized that the subjective definitions of these symptoms might contribute to the limited results of these studies. However, the usefulness of external features and the advantage of our XGBoost model in dealing with complex cases could be attributed to improved performance.

In agreement with previous studies, our findings showed that cough, rhinorrhea, and sore throat were positively associated with influenza infection. This supports the current understanding of cough as an important feature in influenza patients.[30] The influenza virus is thought to infect epithelial cells in the airway, causing inflammation and cytokine release from the host immune system. Furthermore, sore throat was previously included in the WHO ILI definition but was removed in 2011 owing to conflicting evidence.[44-46] A sore throat in young children is challenging to diagnose.[11] However,

we demonstrated that sore throat has a significant predictive value, implying that further revision of the definition based on age groups is warranted and could facilitate a more prompt and precise prediction of influenza infection.[26, 27]

We also evaluated another clinical prediction model developed by Dugas *et al.*, composed of only three clinical symptoms, including fever, cough, and headache.[32] Similarly to the original study, our findings indicated that sensitivity was high (84%) but at the expense of low specificity (34%). This clinical prediction model was developed using data from a single influenza season and focused solely on high-risk patients, which could explain its low specificity. Furthermore, the presence or absence of a headache shows questionable value in distinguishing adult patients with influenza and other pathogens.[32] Our data revealed that headaches did not differ significantly between ILI patients with and without influenza infection. Another explanation is that simple prediction models frequently disregard more complex interactions between clinical features. The ability of ML algorithms to handle high-order interactions and nonlinear relationships is a significant advantage that can enhance prediction models.

Similar studies used natural language processing (NLP) and free text reports to detect influenza, identical to our effort to use clinical features to predict influenza infection. Pineda and Tsui demonstrated, in two retrospective studies, respectively, that NLP and ML algorithms could be used to extract features from electronic health records to detect influenza infection.[26, 27] However, the model developed by Pineda *et al.* was trained in a single health system without external validation. Furthermore, the two studies used non-ILI patients as control groups, which may have overestimated model performance.[28]

In contrast to previous clinical prediction models that relied solely on patient signs and symptoms, we discovered that external features such as the week of the influenza season, the season of the year, and the day of illness have a significant role to play in our XGBoost model. These factors are not intuitive, and their nonlinear relationship with other features makes embedding them into conventional prediction models difficult. However, modern ML algorithms and text-mining techniques make this process possible and therefore warrant further investigation.

Given the current availability of rapid influenza diagnostic tools such as point-of-care testing and RT-PCR, the need for a model to effectively predict influenza infection arises. However, clinical gestalt is still required for a sustainable health care environment to increase pre-test probability and direct testing toward specific pathogens.[17] Our model only included clinical and external features that could be obtained easily in front-line health care settings and used in resource-limited regions and pandemic preparedness in the future.

In conclusion, we created and validated a clinical feature-based ML algorithm to predict influenza infections among ILI patients in the ED. We demonstrated that the XGBoost model outperformed the other seven selected ML algorithms and surpassed previously designed models. We created an applet that can be easily linked with electronic medical records to improve clinical utility based on our findings. However, in the new era of COVID-19, our model needs further validation and modification, which can serve as future research directions.

Table 1. Characteristics of enrolled patients with influenza-like illness

*Been exposed to human with confirmed influenza infection within the past 5 days

Table 2. Performance of machine learning algorithms based on top 20 features

Table 3. Comparison of performance between the XGBoost model and other clinical prediction models

***: P -value< 0.001 represents the difference between XGBoost, Zimmerman, Anderson, and Dugas compared to the corresponding testing dataset, respectively. ©Accuracy, sensitivity and specificity of the XGBoost model were measured under the cutoffs of 0.5, 0.4 and 0.6, respectively.

Figure 1. Shapley Additive exPlanations (SHAP) summary plot to explain the feature importance obtained by the XGBoost algorithm

Human exposure: State of having been exposed to people with confirmed influenza infection within the past 5 days; Influenza season: Period during which there is a prevalent outbreak of influenza; Influenza vaccine: Received influenza vaccination of the year; Travel history: Travel history within the past 30 days; Influenza antivirals: State of having taken an influenza antiviral agent within the past 30 days

Figure 2. Receiver-operating characteristic curves of the XGBoost model and the other clinical prediction models

Appendix 1. Characteristics of the Johns Hopkins Centers of Excellence for Influenza Research and Surveillance (JHCEIRS) network hospitals in the US and Taiwan

Appendix 2. Detailed descriptions of ML algorithms

Appendix 3. Flow chart for enrollment

Appendix 4. Overall variables used for model development.

Appendix 5. Performance of the XGBoost model with different cut-off

Appendix 6. Calibration plot for the XGBoost algorithm

Pearson Correlation coefficients = 0.97

Appendix 7. Performance of the XGBoost model in the subgroup analyses

Appendix 8. Examples of the applet.

*Both H1N1 and H3N2 were circulating in the 2018-1-2019 season.

Table 1. Characteristics of enrolled patients with influenza-like illness

Mean/N (SD/%)	Overall (n=2,189)	Influenza Negative (n=1,085)	Influenza Positive (n=1,104)
Demographics and related medical history			
Age, years (median, (IQR))	40 (29-54)	40 (29-53)	40.5 (30-55)
Female	1203 (55.0)	610 (56.2)	593 (53.7)
BMI, kg/m^2	27.30 (7.29)	27.33 (7.33)	27.27 (7.26)
Received influenza vaccination of the year	686 (31.3)	385 (35.5)	301 (27.3)
Exposed to influenza*	292 (13.3)	83 (7.6)	209 (18.9)
Travel history within the past 30 days	568 (25.9)	264 (24.3)	304 (27.5)
Taken antibiotics within the past 30 days	525 (24.0)	296 (27.3)	229 (20.7)
Taken influenza antiviral agent within the past 30 days	217 (9.9)	86 (7.9)	131 (11.9)
Day of illness of ILI while visiting the EDs	3 (2-5)	4 (2-5)	3 (2-4)
Visiting during influenza season	1372 (62.7)	633 (58.3)	739 (66.9)
Recruited during H1N1 dominant season	1261 (57.6)	537 (49.5)	724 (65.6)
Recruited in the US	1103 (50.4)	575 (53.0)	528 (47.8)
Triage vital sign			
Body temperature, °C	37.69 (1.10)	37.49 (1.11)	37.92 (1.05)
Pulse rate, bpm	103.94 (19.06)	102.78 (19.31)	105.25 (18.70)
Respiratory rate, bpm	18.68 (3.38)	18.52 (2.64)	18.87 (4.05)
Oxygen saturation, %	96.64 (2.88)	96.84 (2.80)	96.39 (2.95)
Past medical history			
Chronic lung disease	572 (26.1)	295 (27.2)	277 (25.1)
HIV	140 (6.4)	71 (6.5)	69 (6.2)
Autoimmune disease	91 (4.2)	51 (4.7)	40 (3.6)
Clinical symptoms			
Cough	1908 (87.2)	849 (78.2)	1059 (95.9)
Cough with sputum	1410 (64.4)	613 (56.5)	797 (72.2)
Headache	1768 (80.8)	890 (82.0)	878 (79.5)
Sore throat	1462 (66.8)	689 (63.5)	773 (70.0)
Body aches	1816 (83.0)	886 (81.7)	930 (84.2)
Rhinorrhea	1541 (70.4)	696 (64.1)	845 (76.5)
Shortness of breath	1469 (67.1)	698 (64.3)	771 (69.8)
Sinus pain	520 (23.8)	267 (24.6)	253 (22.9)
Wheezing	968 (44.2)	452 (41.7)	516 (46.7)
Fatigue	1943 (88.8)	951 (87.6)	992 (89.9)
Nausea	1110 (50.7)	574 (52.9)	536 (48.6)
Diarrhea	638 (29.1)	332 (30.6)	306 (27.7)
Stomach pain	763 (34.9)	410 (37.8)	353 (32.0)

*Been exposed to human with confirmed influenza infection within the past 5 days

Table 2. Performance of machine learning algorithms based on top 20 features

Algorithm	Dataset	Accuracy (95% C.I.)	AUROC (95% C.I.)
XGBoost	Training	0.74 (0.71-0.76)	0.82 (0.80-0.84)
	Testing	0.72 (0.68-0.76)	0.82 (0.79-0.85)
Ranger	Training	0.97 (0.96-0.98)	0.99 (0.99-1.00)
	Testing	0.71 (0.67-0.74)	0.79 (0.75-0.82)
Random Forest	Training	1.00 (0.99-1.00)	1.00 (1.00-1.00)
	Testing	0.70 (0.66-0.73)	0.78 (0.75-0.82)
Cforest	Training	0.83 (0.81-0.85)	0.92 (0.90-0.93)
	Testing	0.69 (0.66-0.73)	0.77 (0.73-0.80)
SVM	Training	0.75 (0.72-0.77)	0.84 (0.82-0.86)
	Testing	0.70 (0.66-0.73)	0.77 (0.73-0.80)
Artificial Neural Network	Training	0.74 (0.72-0.76)	0.82 (0.80-0.84)
	Testing	0.70 (0.67-0.74)	0.76 (0.72-0.79)
Deep Learning (2-layers)	Training	0.70 (0.67-0.72)	0.79 (0.77-0.81)
	Testing	0.67 (0.63-0.70)	0.75 (0.71-0.79)
Deep Learning (3-layers)	Training	0.69 (0.67-0.71)	0.76 (0.73-0.78)
	Testing	0.68 (0.64-0.72)	0.73 (0.69-0.77)
Deep Learning (4-layers)	Training	0.67 (0.64-0.69)	0.73 (0.71-0.76)
	Testing	0.65 (0.61-0.68)	0.71 (0.67-0.75)

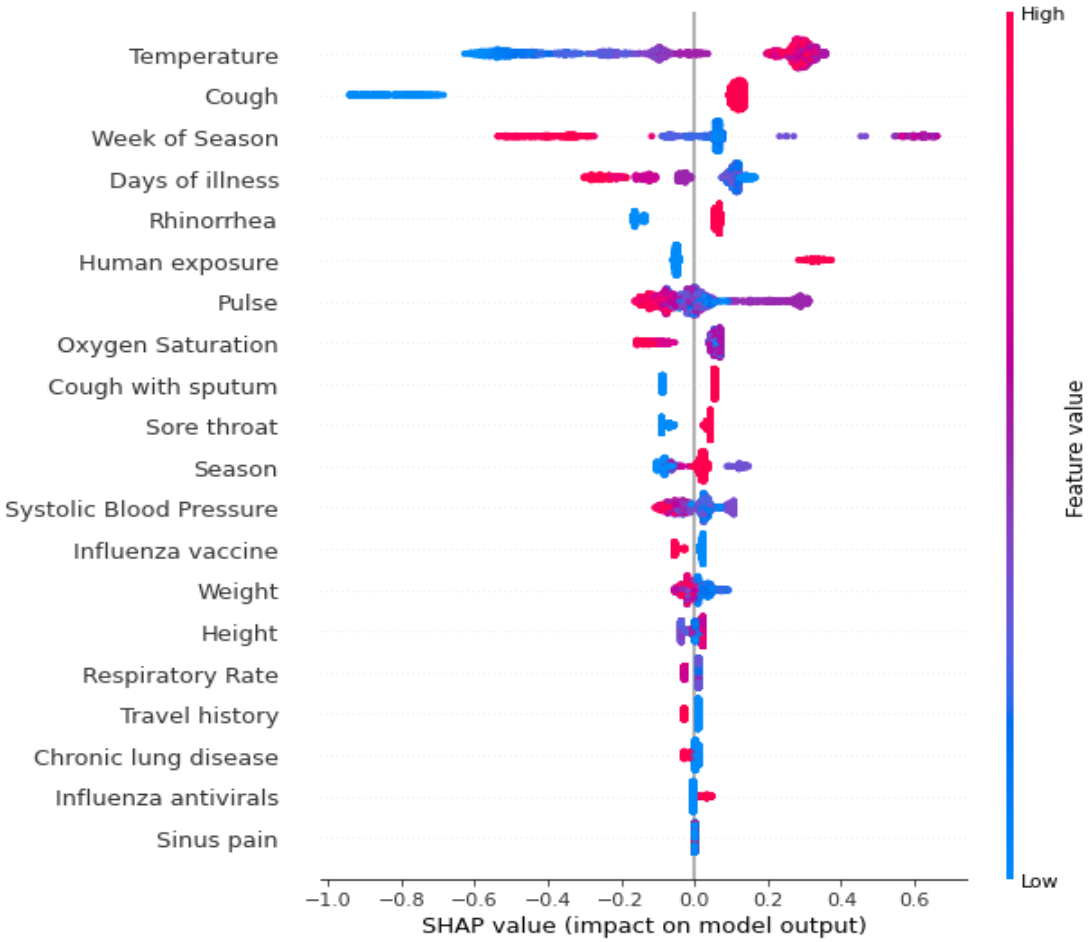
Table 3. Comparison of performance between the XGBoost model and other clinical prediction models

Models	Dataset	Accuracy (95% C.I.)	Sensitivity (95% C.I.)	Specificity (95% C.I.)	AUROC (95% C.I.)
XGboost	Training	0.74 (0.71-0.76)[©]	0.91 (0.89-0.93)[©]	0.88 (0.85-0.90)[©]	0.82 (0.80-0.84)
	Testing	0.72 (0.69-0.76)[©]	0.92 (0.88-0.95)[©]	0.89 (0.86-0.92)[©]	0.82 (0.79-0.85)
Zimmerman	Training	0.59 (0.56-0.61)	0.96 (0.94-0.97)	0.21 (0.18-0.24)	0.59 (0.56-0.62)
	Testing	0.60 (0.56-0.64)	0.96 (0.93-0.98)	0.23 (0.19-0.28)	0.60 (0.56-0.64) ***
Anderson	Training	0.64 (0.61-0.66)	0.55 (0.51-0.58)	0.72 (0.69-0.75)	0.67 (0.64-0.70)
	Testing	0.57 (0.53-0.61)	0.52 (0.46-0.58)	0.63 (0.57-0.68)	0.61(0.56-0.66) ***
Dugas	Training	0.59 (0.56-0.61)	0.84 (0.82-0.87)	0.33 (0.30-0.36)	0.61 (0.55-0.67)
	Testing	0.59 (0.55-0.64)	0.84 (0.80-0.88)	0.34 (0.28-0.40)	0.65 (0.62-0.68)***

***: P -value< 0.001 represents the difference between XGBoost, Zimmerman, Anderson, and Dugas compared to the corresponding testing dataset, respectively.

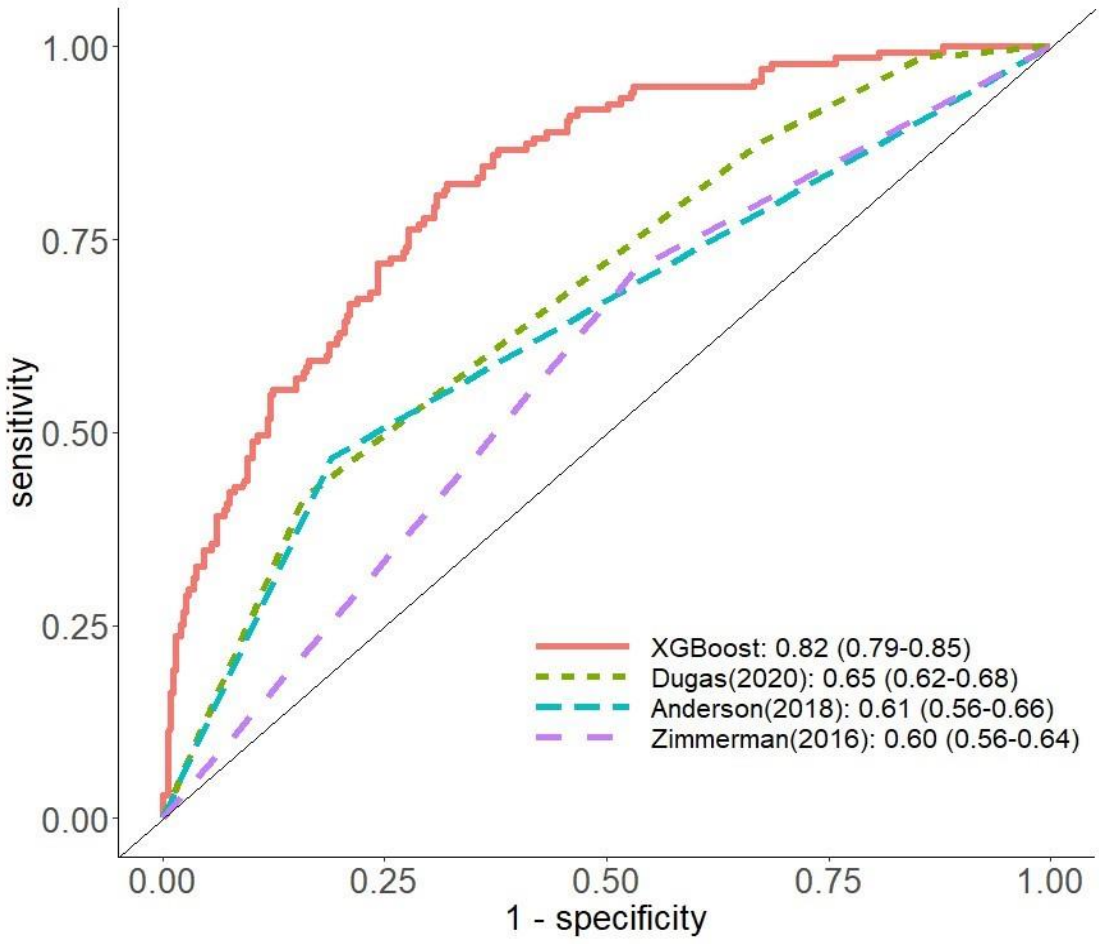
[©]Accuracy, sensitivity and specificity of the XGBoost model were measured under the cutoffs of 0.5, 0.4 and 0.6, respectively.

Figure 1. Shapley Additive exPlanations (SHAP) summary plot to explain the feature importance obtained by the XGBoost algorithm



Human exposure: State of having been exposed to people with confirmed influenza infection within the past 5 days; Influenza season: Period during which there is a prevalent outbreak of influenza; Influenza vaccine: Received influenza vaccination of the year; Travel history: Travel history within the past 30 days; Influenza antivirals: State of having taken an influenza antiviral agent within the past 30 days

Figure 2. Receiver-operating characteristic curves of the XGBoost model and the other clinical prediction models



Appendix 1. Characteristics of the Johns Hopkins Centers of Excellence for Influenza Research and Surveillance (JHCEIRS) network hospitals in the US and Taiwan

Information about JHCEIRS (Johns Hopkins Centers of Excellence for Influenza Research and Surveillance) network hospitals in the US and Taiwan
Two hospitals in US including Johns Hopkins Hospital (1177 beds with 100,000 annual ED visits) and Johns Hopkins Bayview Medical Center (448 beds with 60,000 annual ED visits); Three hospitals in Taiwan including Linkou Chang Gung Memorial Hospital (3406 beds with 160,000 annual ED visits), Keelung Chang Gung Memorial Hospital (1089 beds with 68,000 annual ED visits) and Taipei Chang Gung Memorial Hospital (252 beds with 17,000 annual ED visits). Johns Hopkins Hospital and Linkou Chang Gung Memorial Hospital are tertiary referral medical center located in urban area; Johns Hopkins Bayview Medical Center and Keelung Chang Gung Memorial Hospital are regional hospitals located in suburb area of a metropolitan city; Taipei Chang Gung Memorial Hospital is a community hospital located in capital city of Taiwan.

Appendix 2. Detailed descriptions of ML algorithms

Algorithms	Description	Final hyperparameter setting	Software/Package
XGBoost	eXtreme Gradient Boosting (XGBoost) is an algorithm for structured or tabular data. It applies boosting methods to further improve performance on those mis-classified observations.	learning_rate =0.03814 n_estimators=2354 max_depth=63 min_child_weight=1 gamma=11 subsample=0.177 colsample_bytree=0.81 objective= 'binary:logistic' scale_pos_weight=1	Python/xgboost
Ranger	RANdom forest GEneRator could build models quickly and find out optimal parameter values using parameter tuning. Ranger is used when dealing with high dimensional data and expect a memory efficient fast implementation of Random Forest.	ROC was used to select the optimal model using the largest value. The final values used for the model were mtry = 2, splitrule = gini and min.node.size = 11.	R/Caret
Random Forest	Random forest forms bagged decision tree models which split on a subset of features on each split. By creating a multitude of decision trees to be trained on training dataset, the output	ROC was used to select the optimal model using the largest value. The final value used for the model was mtry = 4.	R/Caret

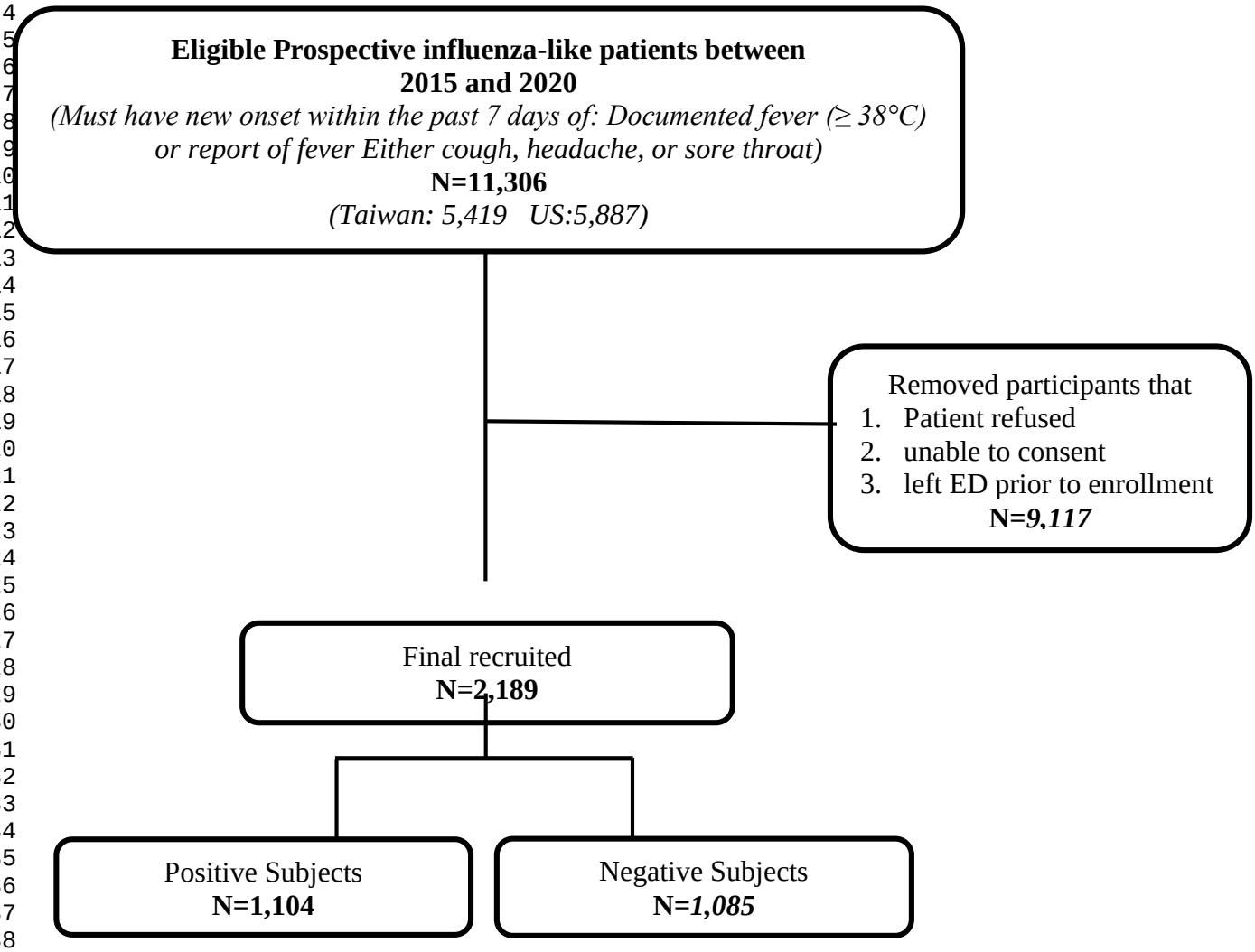
	predictive value is the mode of the classes (classification) or mean prediction (regression) of the individual tree. By averaging away the variances of a number of trees, it helps reducing the high variance derived from a single tree.		
Cforest	Conditional random forest (Cforest) is computationally more expensive and better than the Random forest package in terms of accuracy. CForest uses out-of-bag data to provide higher accuracy. It then uses weighted average of the trees to get the final ensemble. CForest provides more reliable predictions by producing unbiased trees.	ROC was used to select the optimal model using the largest value. The final value used for the model was mtry = 6.	R/Caret
SVM	A support-vector machine constructs hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outlier detection. When the	ROC was used to select the optimal model using the largest value. The final values used for the model were sigma = 0.04938311 and C = 0.3276326.	R/Caret

	datasets to be discriminated are not linearly separable in a finite-dimensional space, SVM help mapping into a much higher-dimensional space, making the separation and classification easier in that space.		
Artificial Neural Network	Artificial Neural Networks are complex and flexible nonlinear systems. Hidden nodes which link features with the outcomes, allowing nonlinear interactions among the features.	ROC was used to select the optimal model using the largest value. The final values used for the model were size = 19 and decay = 1.342004.	R/Caret

Deep Learning	Deep learning method make use of artificial neural networks which construct multiple layers to progressively extract higher level features from the raw input. For supervised learning tasks, it translates the data into compact intermediate representations akin to principal components and derive layered structures that remove redundant	2-layers:
		`model = Sequential()`
		`model.add(Dense(40, input_dim=20, kernel_initializer='he_uniform', activation='softplus'))`
		`model.add(Dropout(0.2))`
		`model.add(Dense(80, kernel_initializer='normal', activation='softplus'))`
		`model.add(Dropout(0.1))`
		`model.add(Dense(1, activation='sigmoid'))`
		3-layers:
		`model = Sequential()`
		`model.add(Dense(20, input_dim=20, kernel_initializer='glorot_normal', activation='relu', kernel_constraint=maxnorm(9)))`
		`model.add(Dropout(0))`
		`model.add(Dense(30, kernel_initializer='normal', activation='linear'))`
		`model.add(Dropout(0.1))`
		`model.add(Dense(50, kernel_initializer='he_normal', activation='relu'))`
		`model.add(Dropout(0))`
		`model.add(Dense(1, activation='sigmoid'))`
		4-layers:
			Python/Sequential; Dense; Dropout; SGD

	engineering features thus can serve as validation of other machine learning algorithm.	`model = Sequential()`
		`model.add(Dense(90, input_dim=20, kernel_initializer='glorot_normal', activation='tanh'))`
		`model.add(Dropout(0))`
		`model.add(Dense(20, kernel_initializer='he_normal', activation='tanh'))`
		`model.add(Dropout(0.1))`
		`model.add(Dense(20, kernel_initializer='normal', activation='softsign'))`
		`model.add(Dropout(0.2))`
		`model.add(Dense(80, kernel_initializer='he_uniform', activation='relu'))`
		`model.add(Dropout(0))`
		`model.add(Dense(1, activation='sigmoid'))`	

Appendix 3. Flow chart of patient enrollment



Appendix 4. Overall variables used for model development.

Number	Variables
1	age
2	Height
3	Weight
4	BMI
5	Body temperature
6	Pulse rate
7	Respiratory rate
8	Systolic blood pressure
9	Oxygen saturation
10	Day of illness
11	Visits hospital during influenza season
12	Week of the influenza season of visit
13	Female
14	Pregnancy
15	Breastfeeding
16	The subject had a menses in the past 12 months
17	The method used for birth control
18	The subject is currently taking hormone replacement therapy
19	Primary living situation
20	Employed
21	The highest level of education
22	Received influenza vaccination of the year
23	Exposed to influenza
24	Exposed to poultry
25	Exposed to wild birds
26	Exposed to swine
27	Large Farm
28	Backyard Flock
29	Farm
30	Food Preparation
31	Slaughterhouse
32	Exposed to other
33	Exposed to human with confirmed influenza
34	Travel history within the past 30 days
35	Cough
36	Cough with sputum
37	Increase in sputum production
38	Headache
39	Sore throat
40	Body aches

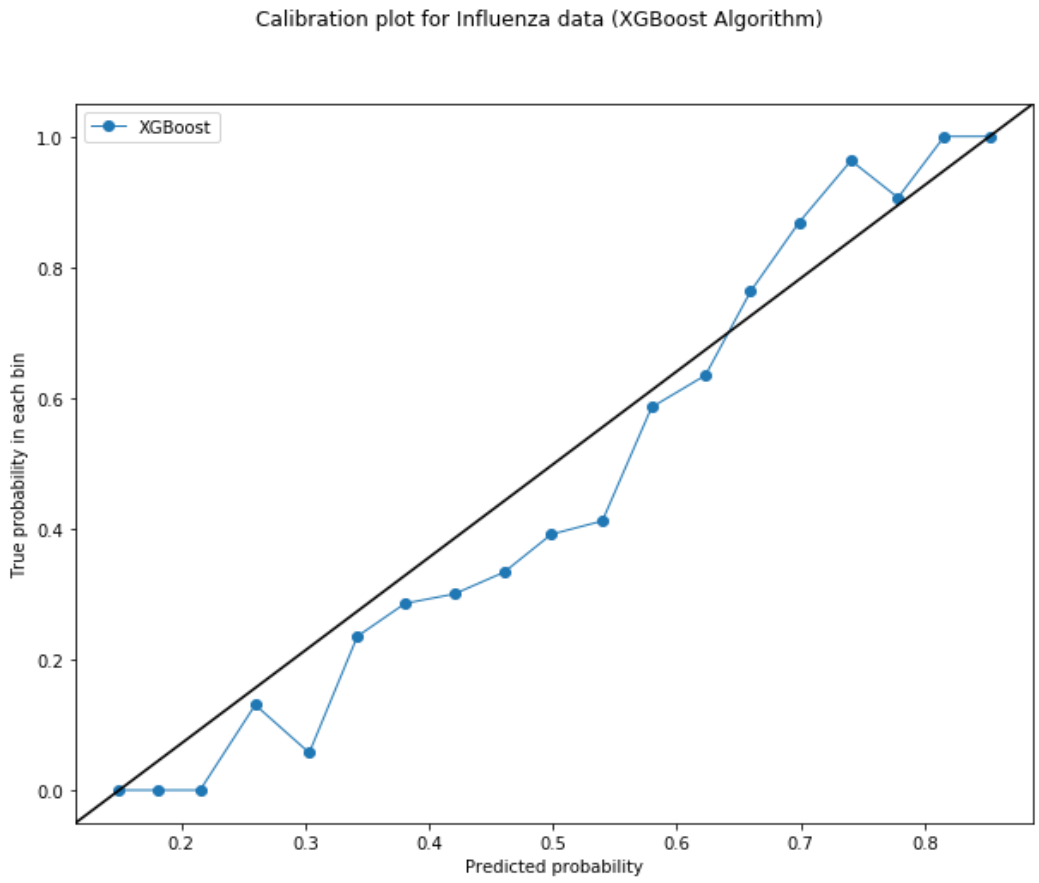
41	Rhinorrhea
42	Shortness of breath
43	Sinus pain
44	Wheezing
45	Fatigue
46	Able to get out of bed
47	Chest pain
48	Chest hurt during breath
49	Shaking chills
50	Loss of appetite
51	Nausea
52	Diarrhea
53	Stomach pain
54	Conjunctivitis
55	Other clinical symptoms
56	Admitted to the hospital within the past 30 days
57	Any antibiotics was taken within the past 30 days
58	Any influenza antivirals were taken within the past 30 days
59	Currently taking steroids
60	Any medications are taking that suppress their immune system
61	Chronic lung disease
62	Asthma
63	COPD
64	Cystic Fibrosis
65	Other chronic lung disease
66	Cardiovascular disease
67	Coronary Artery disease
68	Congestive Heart Failure
69	Cardiomyopathy
70	Valvular disease
71	Congenital Heart disease
72	Other Cardiovascular disease
73	Renal disease
74	End Stage Renal disease
75	Other Renal disease
76	Dialysis
77	Hepatic disease
78	Cirrhosis
79	Hepatitis B
80	Hepatitis C
81	End Stage Liver disease
82	Other Hepatic disease

83	Endocrine/Metabolic disorders
84	Diabetes
85	Thyroid Disorder
86	Other Endocrine/Metabolic disorder
87	Hematologic disease
88	Sickle Cell disease
89	Lymphoma
90	Leukemia
91	Other Hematologic disease
92	Neurological disorders
93	Stroke
94	Seizure/Epilepsy
95	Intellectual Disability
96	Multiple Sclerosis
97	Spinal Cord disease or Injury
98	Peripheral Nerve disease
99	Cerebral Palsy
100	Other Neurological disorders
101	Polycystic ovarian syndrome
102	Endometriosis
103	Primary ovarian insufficiency
104	Other Polycystic ovarian syndrome
105	HIV/AIDS
106	A recent CD4 count within the last 12 months
107	Autoimmune disorder
108	Cancer
109	Chemotherapy
110	Chemotherapy medications available
111	Subject currently treated with radiation
112	Subject had an organ transplant
113	Subject suffered from any other medical conditions not mentioned above
114	H1N1
115	H3N2

Appendix 5. Performance of the XGBoost model with different cut-off

	Training set			Testing set		
Cut-off	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
0	0.52	1.00	0	0.47	1.00	0
0.05	0.52	1.00	0	0.47	1.00	0
0.10	0.52	1.00	0	0.47	1.00	0
0.15	0.52	1.00	0.01	0.48	1.00	0.01
0.20	0.54	1.00	0.05	0.50	1.00	0.05
0.25	0.58	0.99	0.14	0.53	1.00	0.11
0.30	0.62	0.98	0.23	0.57	0.99	0.19
0.35	0.66	0.95	0.34	0.62	0.97	0.31
0.40	0.70	0.91	0.47	0.65	0.92	0.42
0.45	0.72	0.86	0.57	0.70	0.86	0.55
0.50	0.74	0.78	0.69	0.72	0.77	0.68
0.55	0.74	0.69	0.78	0.75	0.70	0.79
0.60	0.72	0.57	0.88	0.73	0.54	0.89
0.65	0.67	0.42	0.94	0.69	0.41	0.95
0.70	0.61	0.27	0.98	0.66	0.29	0.98
0.75	0.54	0.12	1.00	0.61	0.17	0.99
0.80	0.50	0.03	1.00	0.56	0.06	1.00
0.85	0.48	0.01	1.00	0.53	0.01	1.00
0.90	0.48	0	1.00	0.53	0	1.00
0.95	0.48	0	1.00	0.53	0	1.00
1	0.48	0	1.00	0.53	0	1.00

Appendix 6. Calibration plot for the XGBoost algorithm.



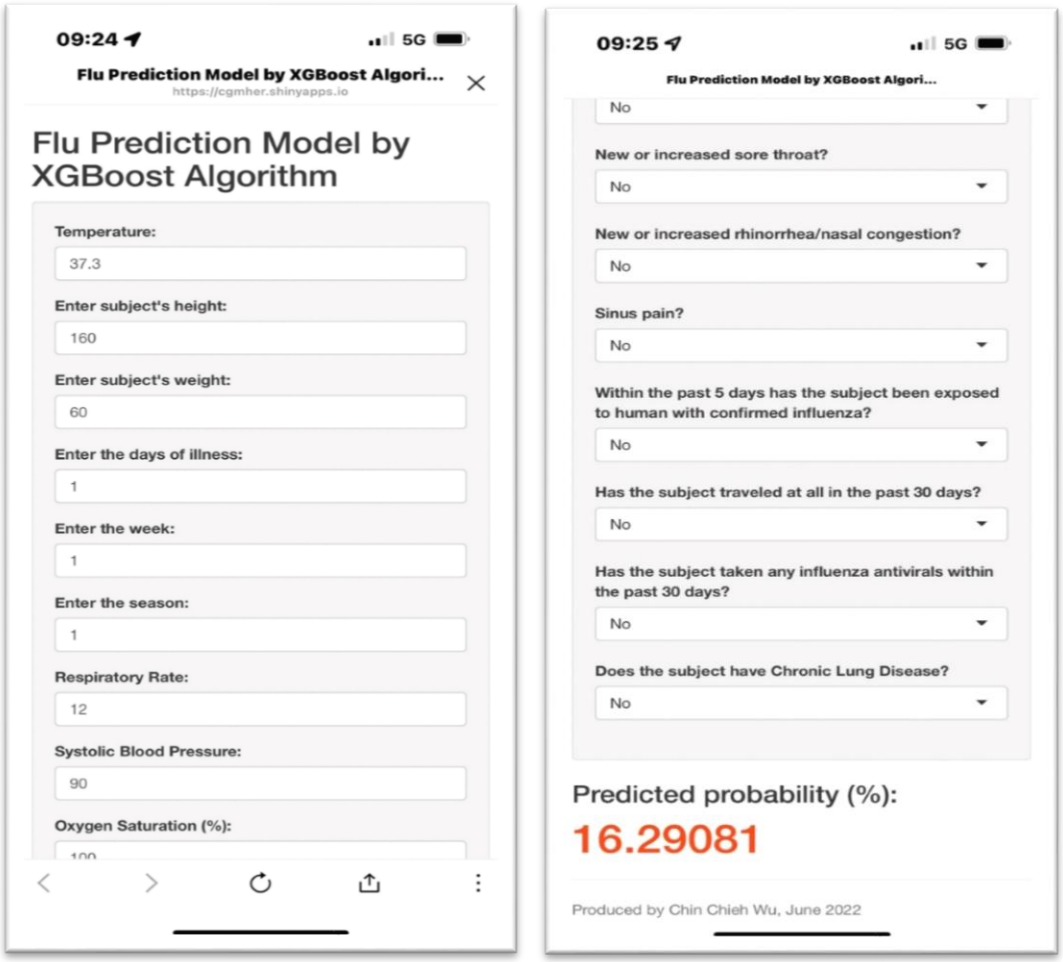
Pearson Correlation coefficients = 0.97

Appendix 7. Performance of the XGBoost model in the subgroup analyses

			(Apply overall model for subgroups)			(Re-training in subgroups)		
Category	**dataset**	**Sample size**	**Training (AUROC) (95% C.I.)**	**Testing (AUROC) (95% C.I.)**	**p-value**	**Training (AUROC) (95% C.I.)**	**Testing (AUROC) (95% C.I.)**	**p-value**
Country	**Taiwan**	1,086	---	0.78 (0.72–0.84)	p=0.004	0.81 (0.78–0.84)	0.75 (0.69–0.80)	p=0.024
	US	1,103	---	0.72 (0.59–0.85)		0.86 (0.83–0.88)	0.83 (0.79–0.88)	
Subtype	**H1N1**	1,261*	---	0.84 (0.79–0.90)	p=0.062	0.88 (0.85–0.90)	0.79 (0.74–0.84)	p=1.000
	H3N2	1,397*	---	0.81 (0.75–0.87)		0.84 (0.82–0.86)	0.79 (0.75–0.83)	

*Both H1N1 and H3N2 were circulating in the 2018-1-2019 season.

Appendix 8. Examples of the applet.



Reference

1. Iuliano AD, Roguski KM, Chang HH, et al. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *The Lancet.* 2018;391:1285-1300.
2. Troeger CE, Blacker BF, Khalil IA, et al. Mortality, morbidity, and hospitalisations due to influenza lower respiratory tract infections, 2017: an analysis for the Global Burden of Disease Study 2017. *The Lancet Respiratory Medicine.* 2019;7:69-89.
3. Cassini A, Colzani E, Pini A, et al. Impact of infectious diseases on population health using incidence-based disability-adjusted life years (DALYs): results from the Burden of Communicable Diseases in Europe study, European Union and European Economic Area countries, 2009 to 2013. *Euro Surveill.* 2018;23.
4. Centers for Disease Control and Prevention NCfIaRDN. Influenza Antiviral Medications: Summary for Clinicians2021.
5. Hartman L, Zhu Y, Edwards KM, Griffin MR, Talbot HK. Underdiagnosis of Influenza Virus Infection in Hospitalized Older Adults. *J Am Geriatr Soc.* 2018;66:467-472.
6. Fowlkes AL, Steffens A, Reed C, Temte JL, Campbell AP, Influenza Incidence Surveillance Project Working G. Influenza Antiviral Prescribing Practices and the Influence of Rapid Testing Among Primary Care Providers in the US, 2009-2016. *Open Forum Infect Dis.* 2019;6:ofz192.
7. Muthuri SG, Myles PR, Venkatesan S, Leonardi-Bee J, Nguyen-Van-Tam JS. Impact of neuraminidase inhibitor treatment on outcomes of public health importance during the 2009-2010 influenza A(H1N1) pandemic: a systematic review and meta-analysis in hospitalized patients. *J Infect Dis.* 2013;207:553-563.
8. Muthuri SG, Venkatesan S, Myles PR, et al. Effectiveness of neuraminidase inhibitors in reducing mortality in patients admitted to hospital with influenza A H1N1pdm09 virus infection: a meta-analysis of individual participant data. *The Lancet Respiratory Medicine.* 2014;2:395-404.
9. Uyeki TM, Bernstein HH, Bradley JS, et al. Clinical Practice Guidelines by the Infectious Diseases Society of America: 2018 Update on Diagnosis, Treatment, Chemoprophylaxis, and Institutional Outbreak Management of Seasonal Influenzaa. *Clin Infect Dis.* 2019;68:e1-e47.
10. Katzen J, Kohn R, Houk JL, Ison MG. Early Oseltamivir After Hospital Admission Is Associated With Shortened Hospitalization: A 5-Year Analysis of Oseltamivir Timing and Clinical Outcomes. *Clin Infect Dis.* 2019;69:52-58.
11. Miller MR, Peters TR, Suerken CK, Snively BM, Poehling KA. Predictors of Influenza Diagnosis Among Patients With Laboratory-Confirmed Influenza. *J Infect Dis.* 2015;212:1604-1612.
12. Campe H, Heinzinger S, Hartberger C, Sing A. Clinical symptoms cannot predict influenza infection during the 2013 influenza season in Bavaria, Germany. *Epidemiol Infect.* 2016;144:1045-1051.
13. Lam PP, Coleman BL, Green K, et al. Predictors of influenza among older adults in the emergency department. *BMC Infect Dis.* 2016;16:615.
14. Centers for Disease Control and Prevention NCfIaRDN. Centers for Disease Control and Prevention. Guide for considering influenza testing when influenza viruses are circulating in the community.2020.

15. Chartrand C, Leeflang MM, Minion J, Brewer T, Pai M. Accuracy of rapid influenza diagnostic tests: a meta-analysis. *Ann Intern Med.* 2012;156:500-511.

16. Dugas AF, Valsamakis A, Atreya MR, et al. Clinical diagnosis of influenza in the ED. *Am J Emerg Med.* 2015;33:770-775.

17. Lee BY, McGlone SM, Bailey RR, et al. To Test or to Treat? An Analysis of Influenza Testing and Antiviral Treatment Strategies Using Economic Computer Modeling. *PLOS ONE.* 2010;5:e11284.

18. Ebell MH, Afonso A. A systematic review of clinical decision rules for the diagnosis of influenza. *Ann Fam Med.* 2011;9:69-77.

19. Anderson KB, Simasathien S, Watanaveeradej V, et al. Clinical and laboratory predictors of influenza infection among individuals with influenza-like illness presenting to an urban Thai hospital over a five-year period. *PLoS One.* 2018;13:e0193050.

20. Zimmerman RK, Balasubramani GK, Nowalk MP, et al. Classification and Regression Tree (CART) analysis to predict influenza in primary care patients. *BMC Infect Dis.* 2016;16:503.

21. Delahanty RJ, Alvarez J, Flynn LM, Sherwin RL, Jones SS. Development and Evaluation of a Machine Learning Model for the Early Identification of Patients at Risk for Sepsis. *Ann Emerg Med.* 2019;73:334-344.

22. Taylor RA, Moore CL, Cheung KH, Brandt C. Predicting urinary tract infections in the emergency department with machine learning. *PLoS One.* 2018;13:e0194085.

23. Than MP, Pickering JW, Sandoval Y, et al. Machine Learning to Predict the Likelihood of Acute Myocardial Infarction. *Circulation.* 2019.

24. Wu CC, Hsu WD, Islam MM, et al. An artificial intelligence approach to early predict non-ST-elevation myocardial infarction patients with chest pain. *Comput Methods Programs Biomed.* 2019;173:109-117.

25. Awan SE, Sohel F, Sanfilippo FM, Bennamoun M, Dwivedi G. Machine learning in heart failure: ready for prime time. *Curr Opin Cardiol.* 2018;33:190-195.

26. Lopez Pineda A, Ye Y, Visweswaran S, Cooper GF, Wagner MM, Tsui FR. Comparison of machine learning classifiers for influenza detection from emergency department free-text reports. *J Biomed Inform.* 2015;58:60-69.

27. Tsui F, Wagner M, Cooper G, et al. Probabilistic case detection for disease surveillance using data in electronic medical records. *Online J Public Health Inform.* 2011;3.

28. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155:529-536.

29. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ.* 2015;350:g7647.

30. Fitzner J, Qasmieh S, Mounts AW, et al. Revision of clinical case definitions: influenza-like illness and severe acute respiratory infection. *Bull World Health Organ.* 2018;96:122-128.

31. Novak-Weekley SM, Marlowe EM, Poulter M, et al. Evaluation of the Cepheid Xpert Flu Assay for rapid identification and differentiation of influenza A, influenza A 2009 H1N1, and influenza B viruses. *J Clin Microbiol.* 2012;50:1704-1710.

32. Dugas AF, Hsieh YH, LoVecchio F, et al. Derivation and Validation of a Clinical Decision Guideline for Influenza Testing in 4 US Emergency Departments. *Clin Infect Dis.* 2020;70:49-58.

33. Kursa MB, Rudnicki WR. Feature Selection with the Boruta Package. *J Stat Softw.* 2010;36:1 - 13.

34. Kuhn M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software.* 2008;28:1 - 26.

35. Figueroa RL, Zeng-Treitler Q, Kandula S, Ngo LH. Predicting sample size required for classification performance. *BMC Med Inform Decis Mak.* 2012;12:8.

36. Mandelia Y, Procop GW, Richter SS, Worley S, Liu W, Esper F. Dynamics and predisposition of respiratory viral co-infections in children and adults. *Clin Microbiol Infect.* 2021;27:631.e631-631.e636.

37. Shih HI, Wang HC, Su IJ, et al. Viral Respiratory Tract Infections in Adult Patients Attending Outpatient and Emergency Departments, Taiwan, 2012-2013: A PCR/Electrospray Ionization Mass Spectrometry Study. *Medicine (Baltimore).* 2015;94:e1545.

38. Chiu SC, Lin YC, Wang HC, et al. Surveillance of upper respiratory infections using a new multiplex PCR assay compared to conventional methods during the influenza season in Taiwan. *Int J Infect Dis.* 2017;61:97-102.

39. Falsey AR, Baran A, Walsh EE. Should clinical case definitions of influenza in hospitalized older adults include fever? *Influenza Other Respir Viruses.* 2015;9 Suppl 1:23-29.

40. Smith BJ, Price DJ, Johnson D, et al. Influenza With and Without Fever: Clinical Predictors and Impact on Outcomes in Patients Requiring Hospitalization. *Open Forum Infect Dis.* 2020;7:ofaa268.

41. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med.* 2019;380:1347-1358.

42. <2012 BMC Infectious Diseases Differing clinical characteristics between influenza strains among young healthy adults in the tropics.pdf>.

43. Ang LW, Cui L, Mak TM, et al. Differential age-specific distribution of influenza virus types and subtypes in tropical Singapore, 2011 to 2017. *J Med Virol.* 2019;91:1415-1422.

44. Organization WH. World Health Organization & WHO Global Influenza Programme. (2011). WHO global technical consultation: global standards and tools for influenza surveillance. Geneva, Switzerland2011.

45. Call SA, Vollenweider MA, Hornung CA, Simel DL, McKinney WP. Does this patient have influenza? *Jama.* 2005;293:987-997.

46. Kasper MR, Wierzba TF, Sovann L, Blair PJ, Putnam SD. Evaluation of an influenza-like illness case definition in the diagnosis of influenza among patients with acute febrile illness in Cambodia. *BMC Infect Dis.* 2010;10:320.

Table 1. Characteristics of enrolled patients with influenza-like illness

*Been exposed to human with confirmed influenza infection within the past 5 days

Table 2. Performance of machine learning algorithms based on top 20 features

Table 3. Comparison of performance between the XGBoost model and other clinical prediction models

***: P -value< 0.001 represents the difference between XGBoost, Zimmerman, Anderson, and Dugas compared to the corresponding testing dataset, respectively. ◎Accuracy, sensitivity and specificity of the XGBoost model were measured under the cutoffs of 0.5, 0.4 and 0.6, respectively.

Figure 1. Shapley Additive exPlanations (SHAP) summary plot to explain the feature importance obtained by the XGBoost algorithm

Human exposure: State of having been exposed to people with confirmed influenza infection within the past 5 days; Influenza season: Period during which there is a prevalent outbreak of influenza; Influenza vaccine: Received influenza vaccination of the year; Travel history: Travel history within the past 30 days; Influenza antivirals: State of having taken an influenza antiviral agent within the past 30 days

Figure 2. Receiver-operating characteristic curves of the XGBoost model and the other clinical prediction models

Appendix 1. Characteristics of the Johns Hopkins Centers of Excellence for Influenza Research and Surveillance (JHCEIRS) network hospitals in the US and Taiwan

Appendix 2. Detailed descriptions of ML algorithms

Appendix 3. Flow chart for enrollment

Appendix 4. Overall variables used for model development.

Appendix 5. Performance of the XGBoost model with different cut-off

Appendix 6. Calibration plot for the XGBoost algorithm

Pearson Correlation coefficients = 0.97

Appendix 7. Performance of the XGBoost model in the subgroup analyses

Appendix 8. Examples of the applet.

*Both H1N1 and H3N2 were circulating in the 2018-1-2019 season.

Appendix 1. Characteristics of the Johns Hopkins Centers of Excellence for Influenza Research and Surveillance (JHCEIRS) network hospitals in the US and Taiwan

Information about JHCEIRS (Johns Hopkins Centers of Excellence for Influenza Research and Surveillance) network hospitals in the US and Taiwan
Two hospitals in US including Johns Hopkins Hospital (1177 beds with 100,000 annual ED visits) and Johns Hopkins Bayview Medical Center (448 beds with 60,000 annual ED visits); Three hospitals in Taiwan including Linkou Chang Gung Memorial Hospital (3406 beds with 160,000 annual ED visits), Keelung Chang Gung Memorial Hospital (1089 beds with 68,000 annual ED visits) and Taipei Chang Gung Memorial Hospital (252 beds with 17,000 annual ED visits). Johns Hopkins Hospital and Linkou Chang Gung Memorial Hospital are tertiary referral medical center located in urban area; Johns Hopkins Bayview Medical Center and Keelung Chang Gung Memorial Hospital are regional hospitals located in suburb area of a metropolitan city; Taipei Chang Gung Memorial Hospital is a community hospital located in capital city of Taiwan.

Appendix 2. Detailed descriptions of ML algorithms

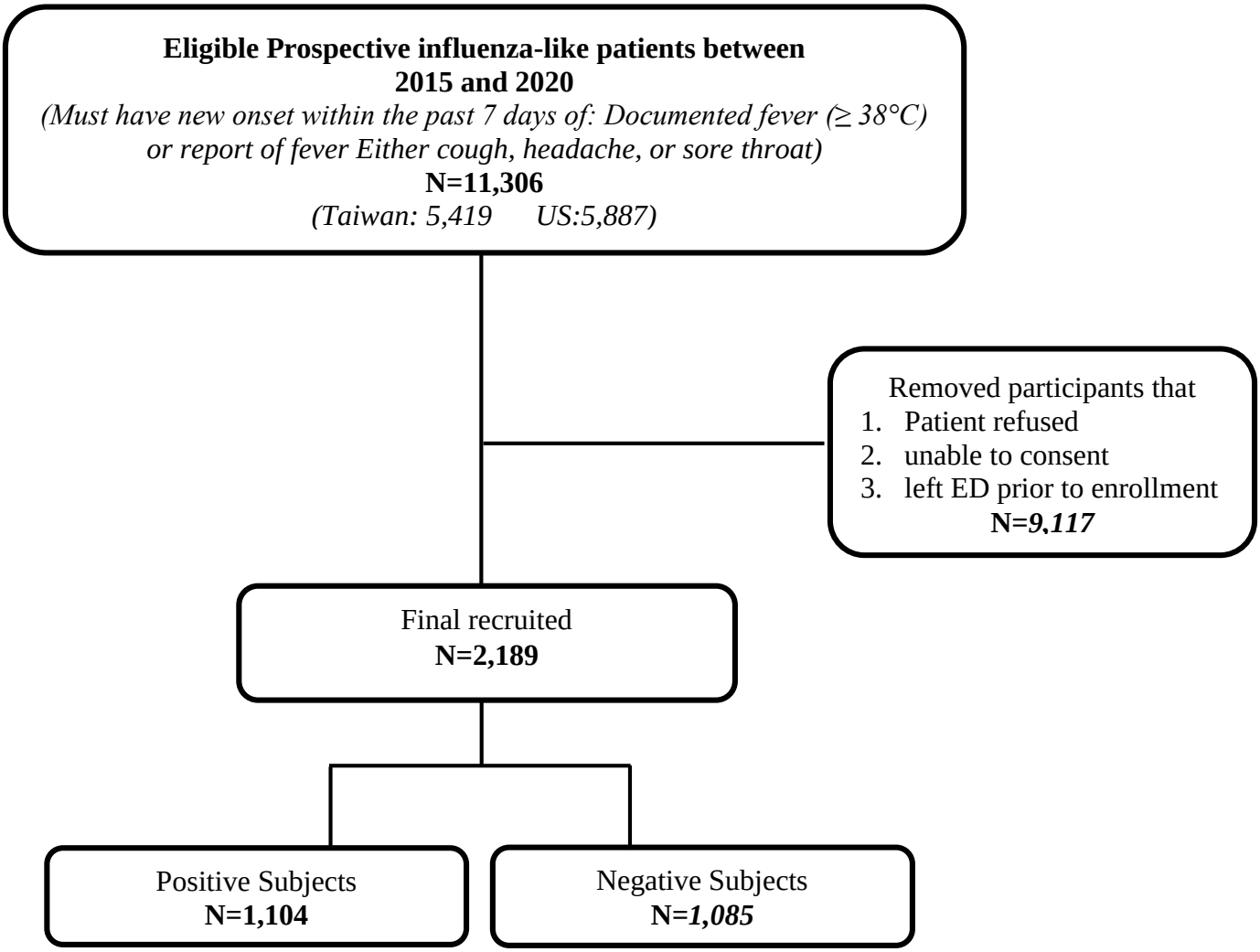
Algorithms	Description	Final hyperparameter setting	Software/Package
XGBoost	eXtreme Gradient Boosting (XGBoost) is an algorithm for structured or tabular data. It applies boosting methods to further improve performance on those mis-classified observations.	learning_rate =0.03814 n_estimators=2354 max_depth=63 min_child_weight=1 gamma=11 subsample=0.177 colsample_bytree=0.81 objective= 'binary:logistic' scale_pos_weight=1	Python/xgboost
Ranger	RANdom forest GEneRator could build models quickly and find out optimal parameter values using parameter tuning. Ranger is used when dealing with high dimensional data and expect a memory efficient fast implementation of Random Forest.	ROC was used to select the optimal model using the largest value. The final values used for the model were mtry = 2, splitrule = gini and min.node.size = 11.	R/Caret
Random Forest	Random forest forms bagged decision tree models which split on a subset of features on each split. By creating a multitude of decision trees to be trained on training dataset, the output predictive value is the mode of the classes (classification) or mean prediction (regression) of the individual tree. By averaging away the variances of a number of trees, it helps reducing the high variance derived from a single tree.	ROC was used to select the optimal model using the largest value. The final value used for the model was mtry = 4.	R/Caret
Cforest	Conditional random forest (Cforest) is computationally more expensive and better than the Random forest package in terms of accuracy. CForest uses out-of-bag data to provide higher accuracy. It then uses weighted average of the trees to get the final ensemble. CForest provides more reliable predictions by producing unbiased trees.	ROC was used to select the optimal model using the largest value. The final value used for the model was mtry = 6.	R/Caret

SVM	A support-vector machine constructs hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outlier detection. When the datasets to be discriminated are not linearly separable in a finite-dimensional space, SVM help mapping into a much higher-dimensional space, making the separation and classification easier in that space.	ROC was used to select the optimal model using the largest value. The final values used for the model were sigma = 0.04938311 and C = 0.3276326.	R/Caret
Artificial Neural Network	Artificial Neural Networks are complex and flexible nonlinear systems. Hidden nodes which link features with the outcomes, allowing nonlinear interactions among the features.	ROC was used to select the optimal model using the largest value. The final values used for the model were size = 19 and decay = 1.342004.	R/Caret

Deep Learning	Deep learning method make use of artificial neural networks which construct multiple layers to progressively extract higher level features from the raw input. For supervised learning tasks, it translates the data into compact intermediate representations akin to principal components and derive layered structures that remove redundant engineering features thus can serve as validation of other machine learning algorithm.	2-layers:
		`model = Sequential()`
		`model.add(Dense(40, input_dim=20, kernel_initializer='he_uniform', activation='softplus'))`
		`model.add(Dropout(0.2))`
		`model.add(Dense(80, kernel_initializer='normal', activation='softplus'))`
		`model.add(Dropout(0.1))`
		`model.add(Dense(1, activation='sigmoid'))`
		3-layers:
		`model = Sequential()`
		`model.add(Dense(20, input_dim=20, kernel_initializer='glorot_normal', activation='relu', kernel_constraint=maxnorm(9)))`
		`model.add(Dropout(0))`
		`model.add(Dense(30, kernel_initializer='normal', activation='linear'))`
		`model.add(Dropout(0.1))`
		`model.add(Dense(50, kernel_initializer='he_normal', activation='relu'))`
		`model.add(Dropout(0))`
		`model.add(Dense(1, activation='sigmoid'))`
		4-layers:
		`model = Sequential()`
		`model.add(Dense(90, input_dim=20, kernel_initializer='glorot_normal', activation='tanh'))`
		`model.add(Dropout(0))`
		`model.add(Dense(20, kernel_initializer='he_normal', activation='tanh'))`
		`model.add(Dropout(0.1))`
		`model.add(Dense(20, kernel_initializer='normal', activation='softsign'))`
		`model.add(Dropout(0.2))`
		`model.add(Dense(80, kernel_initializer='he_uniform', activation='relu'))`
		`model.add(Dropout(0))`
		`model.add(Dense(1, activation='sigmoid'))`
			Python/Sequential; Dense; Dropout; SGD

5

Appendix 3. Flow chart of patient enrollment



Appendix 4. Overall variables used for model development.

Number	Variables
1	age
2	Height
3	Weight
4	BMI
5	Body temperature
6	Pulse rate
7	Respiratory rate
8	Systolic blood pressure
9	Oxygen saturation
10	Day of illness
11	Visits hospital during influenza season
12	Week of the influenza season of visit
13	Female
14	Pregnancy
15	Breastfeeding
16	The subject had a menses in the past 12 months
17	The method used for birth control
18	The subject is currently taking hormone replacement therapy
19	Primary living situation
20	Employed
21	The highest level of education
22	Received influenza vaccination of the year
23	Exposed to influenza
24	Exposed to poultry
25	Exposed to wild birds
26	Exposed to swine
27	Large Farm
28	Backyard Flock
29	Farm
30	Food Preparation
31	Slaughterhouse
32	Exposed to other
33	Exposed to human with confirmed influenza
34	Travel history within the past 30 days
35	Cough
36	Cough with sputum
37	Increase in sputum production
38	Headache
39	Sore throat

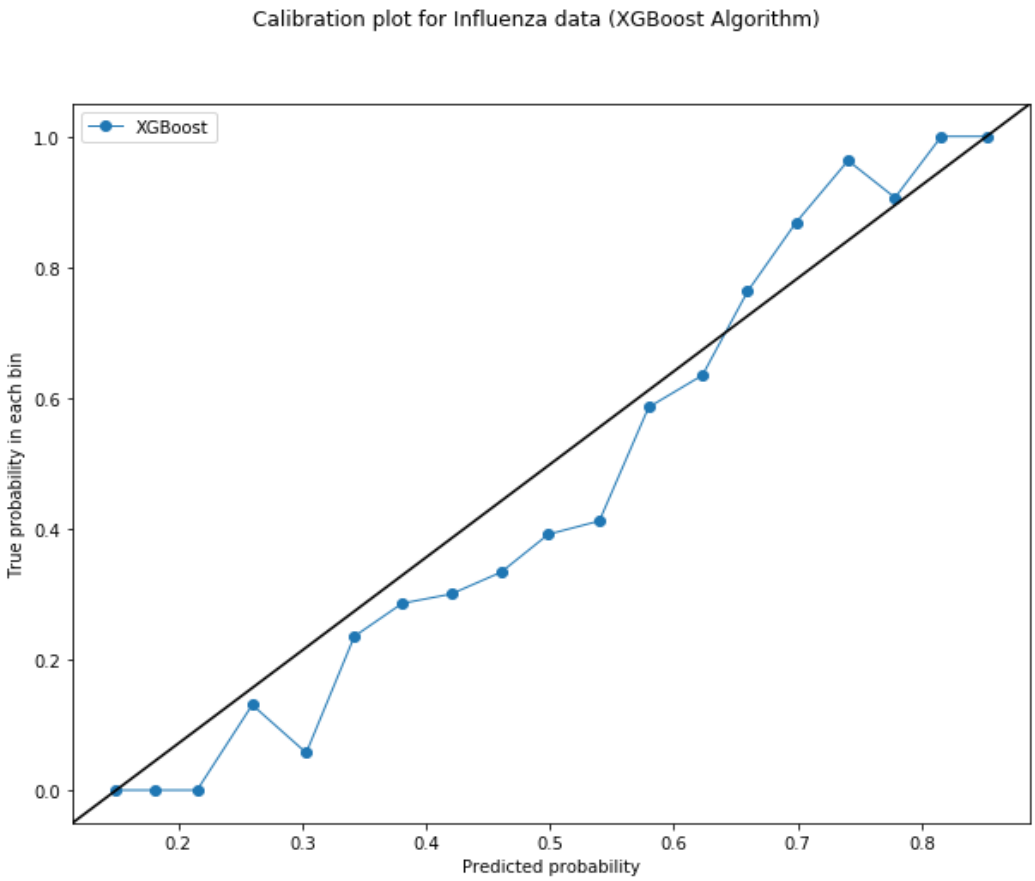
40	Body aches
41	Rhinorrhea
42	Shortness of breath
43	Sinus pain
44	Wheezing
45	Fatigue
46	Able to get out of bed
47	Chest pain
48	Chest hurt during breath
49	Shaking chills
50	Loss of appetite
51	Nausea
52	Diarrhea
53	Stomach pain
54	Conjunctivitis
55	Other clinical symptoms
56	Admitted to the hospital within the past 30 days
57	Any antibiotics was taken within the past 30 days
58	Any influenza antivirals were taken within the past 30 days
59	Currently taking steroids
60	Any medications are taking that suppress their immune system
61	Chronic lung disease
62	Asthma
63	COPD
64	Cystic Fibrosis
65	Other chronic lung disease
66	Cardiovascular disease
67	Coronary Artery disease
68	Congestive Heart Failure
69	Cardiomyopathy
70	Valvular disease
71	Congenital Heart disease
72	Other Cardiovascular disease
73	Renal disease
74	End Stage Renal disease
75	Other Renal disease
76	Dialysis
77	Hepatic disease
78	Cirrhosis
79	Hepatitis B
80	Hepatitis C
81	End Stage Liver disease

82	Other Hepatic disease
83	Endocrine/Metabolic disorders
84	Diabetes
85	Thyroid Disorder
86	Other Endocrine/Metabolic disorder
87	Hematologic disease
88	Sickle Cell disease
89	Lymphoma
90	Leukemia
91	Other Hematologic disease
92	Neurological disorders
93	Stroke
94	Seizure/Epilepsy
95	Intellectual Disability
96	Multiple Sclerosis
97	Spinal Cord disease or Injury
98	Peripheral Nerve disease
99	Cerebral Palsy
100	Other Neurological disorders
101	Polycystic ovarian syndrome
102	Endometriosis
103	Primary ovarian insufficiency
104	Other Polycystic ovarian syndrome
105	HIV/AIDS
106	A recent CD4 count within the last 12 months
107	Autoimmune disorder
108	Cancer
109	Chemotherapy
110	Chemotherapy medications available
111	Subject currently treated with radiation
112	Subject had an organ transplant
113	Subject suffered from any other medical conditions not mentioned above
114	H1N1
115	H3N2

Appendix 5. Performance of the XGBoost model with different cut-off

	Training set			Testing set		
Cut-off	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
0	0.52	1.00	0	0.47	1.00	0
0.05	0.52	1.00	0	0.47	1.00	0
0.10	0.52	1.00	0	0.47	1.00	0
0.15	0.52	1.00	0.01	0.48	1.00	0.01
0.20	0.54	1.00	0.05	0.50	1.00	0.05
0.25	0.58	0.99	0.14	0.53	1.00	0.11
0.30	0.62	0.98	0.23	0.57	0.99	0.19
0.35	0.66	0.95	0.34	0.62	0.97	0.31
0.40	0.70	0.91	0.47	0.65	0.92	0.42
0.45	0.72	0.86	0.57	0.70	0.86	0.55
0.50	0.74	0.78	0.69	0.72	0.77	0.68
0.55	0.74	0.69	0.78	0.75	0.70	0.79
0.60	0.72	0.57	0.88	0.73	0.54	0.89
0.65	0.67	0.42	0.94	0.69	0.41	0.95
0.70	0.61	0.27	0.98	0.66	0.29	0.98
0.75	0.54	0.12	1.00	0.61	0.17	0.99
0.80	0.50	0.03	1.00	0.56	0.06	1.00
0.85	0.48	0.01	1.00	0.53	0.01	1.00
0.90	0.48	0	1.00	0.53	0	1.00
0.95	0.48	0	1.00	0.53	0	1.00
1	0.48	0	1.00	0.53	0	1.00

Appendix 6. Calibration plot for the XGBoost algorithm.



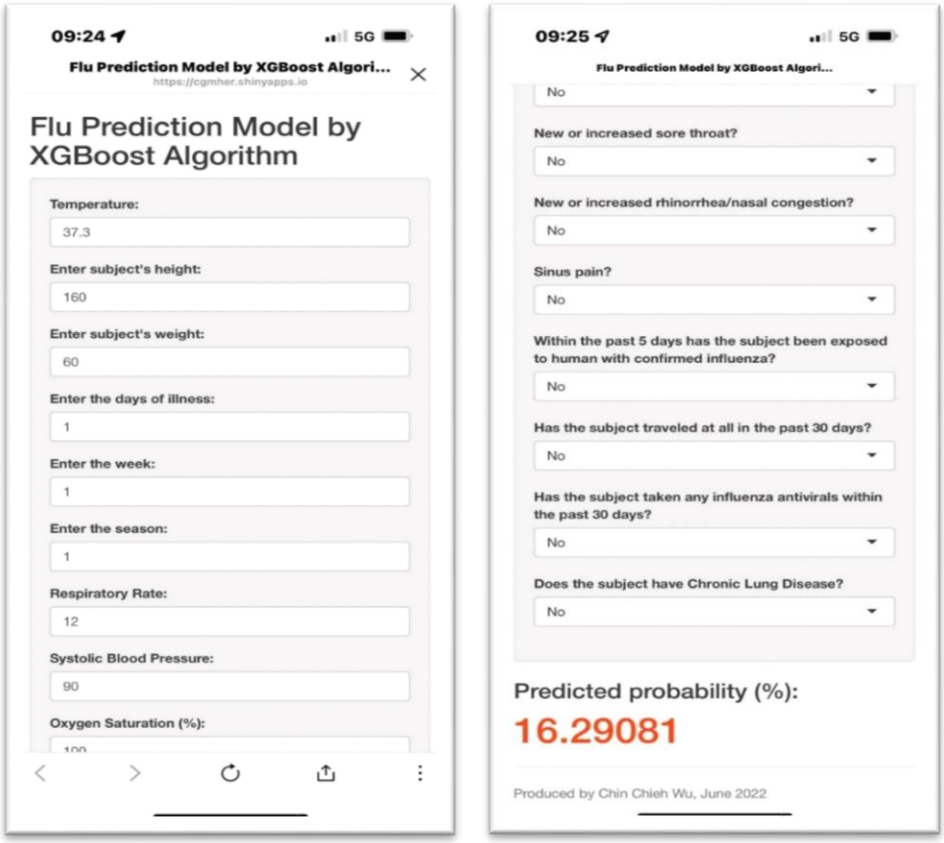
Pearson Correlation coefficients = 0.97

Appendix 7. Performance of the XGBoost model in the subgroup analyses

			(Apply overall model for subgroups)			(Re-training in subgroups)		
Category	**dataset**	**Sample size**	**Training (AUROC) (95% C.I.)**	**Testing (AUROC) (95% C.I.)**	**p-value**	**Training (AUROC) (95% C.I.)**	**Testing (AUROC) (95% C.I.)**	**p-value**
Country	**Taiwan**	1,086	---	0.78 (0.72–0.84)	p=0.004	0.81 (0.78–0.84)	0.75 (0.69–0.80)	p=0.024
	US	1,103	---	0.72 (0.59–0.85)		0.86 (0.83–0.88)	0.83 (0.79–0.88)	
Subtype	**H1N1**	1,261*	---	0.84 (0.79–0.90)	p=0.062	0.88 (0.85–0.90)	0.79 (0.74–0.84)	p=1.000
	H3N2	1,397*	---	0.81 (0.75–0.87)		0.84 (0.82–0.86)	0.79 (0.75–0.83)	

*Both H1N1 and H3N2 were circulating in the 2018-1-2019 season.

Appendix 8. Examples of the applet.



TRIPOD

TRIPOD Checklist: Prediction Model Development

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	5-6
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	7-8
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	8-9
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	10
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	10
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	Appendix1
	5b	Describe eligibility criteria for participants.	10
	5c	Give details of treatments received, if relevant.	NA
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	10
	6b	Report any actions to blind assessment of the outcome to be predicted.	NA
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	10-11
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	NA
Sample size	8	Explain how the study size was arrived at.	NA
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	10-11
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	10-11
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	11-12
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	12
Risk groups	11	Provide details on how risk groups were created, if done.	NA
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Appendix3
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	14
Model development	14a	Specify the number of participants and outcome events in each analysis.	14
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	NA
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	14
	15b	Explain how to the use the prediction model.	15-16
Model performance	16	Report performance measures (with CIs) for the prediction model.	15-16
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	17
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	18-21
Implications	20	Discuss the potential clinical use of the model and implications for future research.	21-22 Appendix8
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	Appendix1-8
Funding	22	Give the source of funding and the role of the funders for the present study.	3

We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.