Fuzzy Systems and Data Mining VIII
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA220380

Nowcasting Influenza-Like Illness (ILI) via a Stacking-Based Ensemble Approach

Xuan WANG[a], Jiajia LUO[a], Zicheng CAO[a], Xiangjun DU[a] and Yang ZHAO[a,1]
[a] *School of Public Health (Shenzhen), Sun Yat-sen University, Shenzhen, China*

Abstract. Monitoring influenza activity can facilitate developing prevention strategies and optimizing public health resource allocation in an effective manner. Traditional influenza surveillance methods usually have a time lag of 1 to 2 weeks. This study concerns the problem of nowcasting influenza-like illness (ILI) by comprehensively incorporating historical ILI records, Internet search data, and tourist flow information. In this study, a set of predictive models are adapted for ILI prediction, including autoregressive integrated moving average (ARIMA), autoregressive with Google search data (ARGO), extreme gradient boosting (XGBoost), and linear regression (LR). To further improve prediction accuracy, a stacking-based ensemble approach is developed to integrate the prediction results from the different models. These methods are validated using the ILI-related data in Taiwan Province of China at both global and city levels. The results show that the stacking-based ensemble approach achieved the best performance in the task of nowcasting, with the least prediction errors at the global level (MAPE=5.6%; RMSE=0.16%; MAE=0.08%). The developed approach is easily tractable and computationally efficient and can be viewed as a feasible alternative to nowcast ILI in areas where influenza activity has no constant seasonal trend.

Keywords. Influenza-like illness, Internet search data, machine learning, ensemble approach

1. Introduction

In the past decades, many countries have been struck down by waves of influenza outbreaks. Worldwide, influenza causes 3 million to 5 million severe cases and approximately 290,000 to 650,000 deaths globally each year, resulting in significant loss of life and property [1]. The influenza situation in the subtropical region is more complex than in other regions, where influenza activity has no constant seasonal trend [2]. Early detection of influenza outbreaks is critical for developing effective strategies for prevention, intervention, and countermeasures, including but not limited to quarantine, vaccination, antiviral campaigns, and optimization of medical resources, which can help reduce the impact of influenza.

In current practice, the influenza-like illness (ILI) surveillance records released by the Centers for Disease Control and Prevention (CDC) usually have a time lag of 1-2 weeks which could be a major challenge for policymakers to accurately estimate epidemics in an efficient real-time manner. Therefore, the effective prediction of current ILI (i.e., nowcasting) has become an essential part of preventing the spread of influenza.

[1] Corresponding Author, Yang ZHAO, School of Public Health (Shenzhen), Sun Yat-sen University, Shenzhen, China; E-mail: zhaoy393@mail.sysu.edu.cn.

Previous studies have employed various forecasting methods for ILI prediction. He et al. [3] explored the application of autoregressive integrated moving average (ARIMA) to predict the positive rate of influenza virus during nine influenza seasons. To incorporate multi-source ILI-related information, Nagar et al. [4] used linear regression (LR) to predict influenza cases by analyzing daily Twitter data from temporal and spatiotemporal. The penalized regression models, such as autoregressive with Google search data (ARGO) [5] and ridge regression [6], have been proposed to settle the overfitting in LR. In addition, using deep learning models such as convolutional neural network (CNN) and long short term memory (LSTM) to forecast influenza has also been studied [7–9].

This paper aims to investigate the predictive capability of machine learning and statistical models in the task of ILI nowcasting and develop an effective prediction approach by incorporating multi-source ILI-related data. In this study, a set of predictive models are adapted for ILI prediction, including ARIMA, ARGO, and extreme gradient boosting (XGBoost) [10], LR. To further improve prediction accuracy, a stacking-based ensemble approach is developed to integrate the prediction results of the individual models. The existing literature has shown that the developed approach could outperform individual predictive models [11-12]. In addition, 10-fold cross-validation is utilized to avoid overfitting in the developed approach. Since cross-validation may lead to look-ahead bias when ARIMA is taken into ensemble learning, ARGO, LR, and XGBoost are used as base learners and support vector machine (SVM) as the meta learner in our developed approach.

The contributions of this paper are mainly two folds: 1) We incorporate three ILI-related data, including historical ILI records, Internet search data, and tourist flow information, for predicting the ILI rate in Taiwan province of China at both global and city levels. 2) We conduct a comprehensive comparative study of the heterogeneous machine learning and statistical models, as well as the stacking-based ensemble approach for ILI nowcasting. This study provides an essential reference for data incorporation and predictive method selection for ILI surveillance.

2. Materials and Methods

2.1. Data Description

The western Pacific region, where Taiwan Province is located, has been reported as one of the three regions with the highest burden of annual influenza-associated deaths [13]. Influenza remains a serious public health threat in Taiwan due to its special geographical location and climatic environment. During 2017–18, Taiwan experienced two severe influenza epidemics caused by A/H3N2 and B viruses, respectively [14]. The data we used to develop the predictive models include weekly ILI records, Google search data, and tourist flow information of Taiwan Province. The details of these data are as follows.

- ILI records. The weekly report of ILI records is from the Taiwan CDC[2] , and the surveillance period covers March 30, 2008, to January 16, 2022. The ILI rate (number of visits of ILI in general outpatient clinics/total number of visits) is calculated to eliminate the impact of demographic changes on influenza data.

[2] Data source: http://www.cdc.gov.tw

- Google search data. In order to ensure the objectivity of the search terms, the keywords utilized are combined from previous studies [15-16], and 17 influenza-related keywords data are selected: "cold", "cough", "fever", "flu", "h1n1", "h3n2", "influenza", "感冒", "喉嚨痛", "咳", "流感", "流感病毒", "流感疫苗", "流感治療", "流行性感冒", "禽流感","傷風". All keywords data are downloaded from the Google public website Google Trends[3].
- The tourist flow information. The tourist flow information, as the indicator of crowd density, is collected from the Taiwan Statistics Information Network[4], including the number of visitors to 98 well-known attractions in Taiwan Province. The correlation between the number of tourists and the ILI rate can be as high as 0.7, which is consistent with the actual situation.

2.2. Study Design

In our case, the response variable is the log-transformed weekly CDC-reported ILI rate, and the covariates are the historical ILI data with lags of 1 to 4 weeks, log-transformed Google search data, and tourist flow data. Each Google search data is added a small value $\delta = 0.5$ before the transformation to avoid the situation of log 0.

We adopt four representative predictive methods for ILI nowcasting, including ARIMA, ARGO, LR, and XGBoost. ARIMA is a classic time series fitting method. LR is a statistical model with outstanding performance in various forecasting activities because of its simplicity and efficiency. ARGO is a regression that introduces time series features for influenza forecasting [5]. XGBoost is a boosting method by integrating multiple regression trees, and its fitting effect can be greatly improved compared to a single regression tree [10]. An ensemble model is developed to construct a Taiwan influenza predictive model to combine the capabilities of each method. A sliding window of 104 weeks is used to train machine learning and statistical models to capture complete information on annual and seasonal trends. We then use the trained model to conduct out-of-sample predictions for the following week and compare the predicted values to the real values reported by the CDC.

2.3. Stacking-based Ensemble Approach

The stacking-based ensemble approach is a technique used to solve problems of classification and regression, which generally integrates a set of heterogeneous models. Figure 1 shows the general framework of the developed approach in this study. In the stacking-based ensemble approach modeling, ARGO, LR, and XGBoost are trained as base learners since ARIMA may lead to look-ahead bias in k-fold cross-validation [11]. SVM is employed as a meta learner given its outstanding performance in solving small samples and nonlinear high-dimensional problems [17].

As shown in Figure 1, the raw data is split into training data and test data. The training data is a dataset with a sliding window of 104 weeks, and the test data is the following observation. Three base learners are trained using 10-fold cross-validation on the training data, and the trained models are also predicted and averaged on the test data. The results predicted by the base learners are used as the input of the meta learner, and

[3] Data source: https://trends.google.com/trends

[4] Data source: www.stat.gov.tw/ct

conduct the final prediction uses the trained meta learner. The developed approach can effectively improve accuracy and reduce biases.

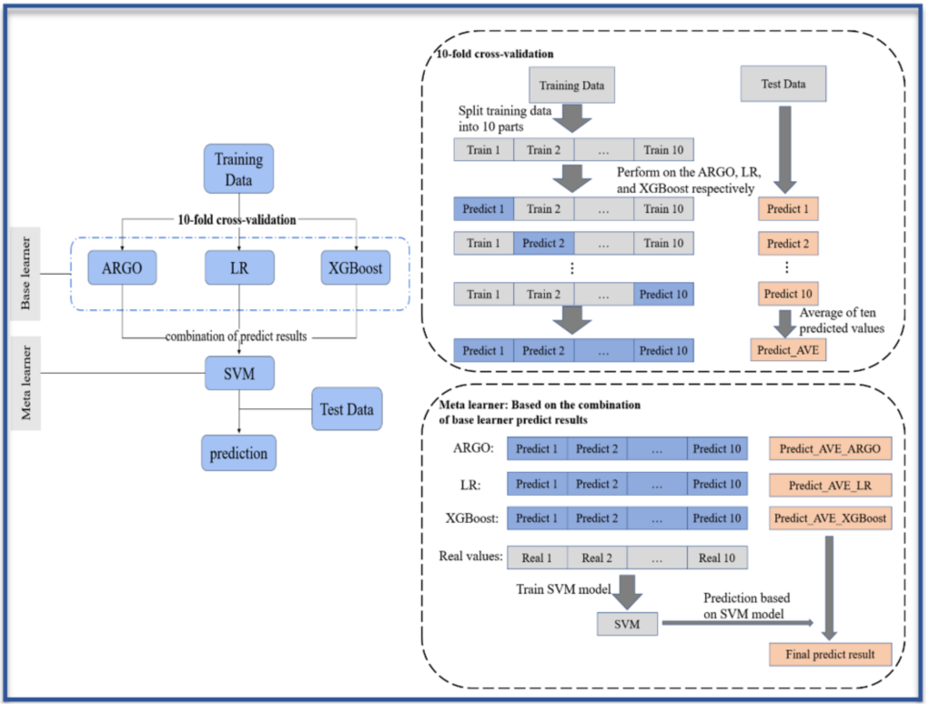

Figure 1. The general framework of the stacking-based ensemble approach.

2.4. Evaluation Metrics

To compare the performance of each model in nowcasting ILI, we use four evaluation metrics to measure the prediction accuracy, including root mean square error (RMSE), mean absolute error (MAE), average absolute percentage error (MAPE), and Pearson correlation coefficient (Corr). The smaller the RMSE, MAE, and MAPE, or the larger the Corr, indicates that the model performs better. For real values $(y_1, y_2, \dots, y_n)$ and the predicted values $(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)$ these metrics are defined as:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{\hat{y}_i - y_i}{y_i}\right|$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i|$$

$$Corr = \frac{cov(\hat{Y}, Y)}{\sigma_{\hat{Y}}\sigma_Y}$$

3. Results

We conduct ILI nowcasting during the period from December 24, 2017, to December 15, 2019, before the outbreak of COVID-19, to evaluate the performance of the five methods at both global and city levels. Since Google search data are not available for all the cities, the six representative cities with Google search data are selected, including Taipei, Tainan, Taichung, Kaohsiung, New Taipei, and Taoyuan.

Table 1 shows the prediction performance of the four individual methods and the stacking-based ensemble approach. It can be seen that the developed approach outperforms the individual methods on the four-evaluation metrics in the task of ILI nowcasting for the whole of Taiwan and five cities. The developed approach performs particularly well on the evaluation metric RMSE, which is sensitive to outliers. Compared to the best-performing individual method, the developed approach reduces RMSE by 7.7% to 26.5%. It is worth noting that the ARGO method outperforms ARIMA, LR, and XGBoost in terms of MAPE, MAE, and RMSE, due to its inherent regularization strategy.

Table 1. The prediction performance of the four individual methods and the stacking-based ensemble approach.

	The whole of Taiwan	**Taipei**	**Tainan**	**Taichung**	**Kaohsiung**	**New Taipei**	**Taoyuan**
MAPE							
ARIMA	0.0630	0.0799	0.0691	0.0690	0.0719	0.0706	0.0748
ARGO	0.0580	0.0726	0.0650	0.0650	0.0692	0.0690	**0.0688**
LR	0.0675	0.0863	0.0696	0.0751	0.0765	0.0746	0.0809
XGBoost	0.0849	0.0987	0.1045	0.0833	0.1084	0.1015	0.0952
Stacking	**0.0560**	**0.0660**	**0.0578**	**0.0633**	**0.0654**	**0.0635**	0.0727
RMSE							
ARIMA	0.0020	0.0020	0.0019	0.0027	0.0017	0.0020	0.0024
ARGO	0.0018	0.0018	0.0018	0.0024	0.0015	0.0019	0.0022
LR	0.0031	0.0029	0.0034	0.0043	0.0030	0.0036	0.0038
XGBoost	0.0021	0.0018	0.0027	0.0024	0.0020	0.0020	0.0023
Stacking	**0.0016**	**0.0013**	**0.0015**	**0.0021**	**0.0014**	**0.0015**	**0.0018**
MAE							
ARIMA	0.0009	0.0009	0.0011	0.0012	0.0009	0.0009	0.0011
ARGO	0.0008	0.0008	0.0010	0.0012	0.0008	0.0009	**0.0011**
LR	0.0009	0.0009	0.0011	0.0013	0.0009	0.0010	0.0012
XGBoost	0.0012	0.0010	0.0016	0.0014	0.0013	0.0013	0.0014
Stacking	**0.0008**	**0.0007**	**0.0009**	**0.0011**	**0.0008**	**0.0008**	0.0011
Corr							
ARIMA	0.8414	0.8007	0.8373	0.8242	0.8007	0.8972	0.8510
ARGO	0.8601	0.8221	0.8448	0.8509	0.8342	0.9034	0.8711
LR	0.8332	0.8235	0.8045	0.8353	0.7630	0.9048	0.8359
XGBoost	0.8394	0.8308	0.7446	0.8406	0.7327	0.9047	0.8661
Stacking	**0.8968**	**0.9062**	**0.8858**	**0.8810**	**0.8590**	**0.9389**	**0.9135**

Figure 2 shows the visualization of the prediction effect at the global level. The difference between the predicted and real values of the ARIMA, ARGO, and developed approach is slight, while the LR and XGBoost show some deviation. It can be seen that the ARIMA, ARGO, and LR methods have problems in capturing irregular turning points and thus may lead to the delayed prediction of outbreaks. XGBoost could not respond immediately when the ILI rate arose in March 2019 by underestimating the trend of influenza outbreaks. The prediction of the developed approach is the closest to the real value but also underestimates the intensity of influenza in March 2019.

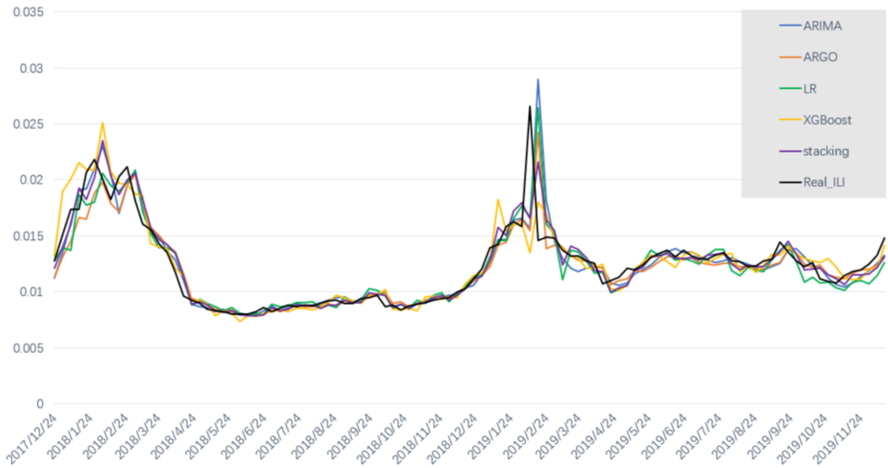

Figure 2. Nowcasting with four individual models and a stacking-based ensemble approach for the whole of Taiwan.

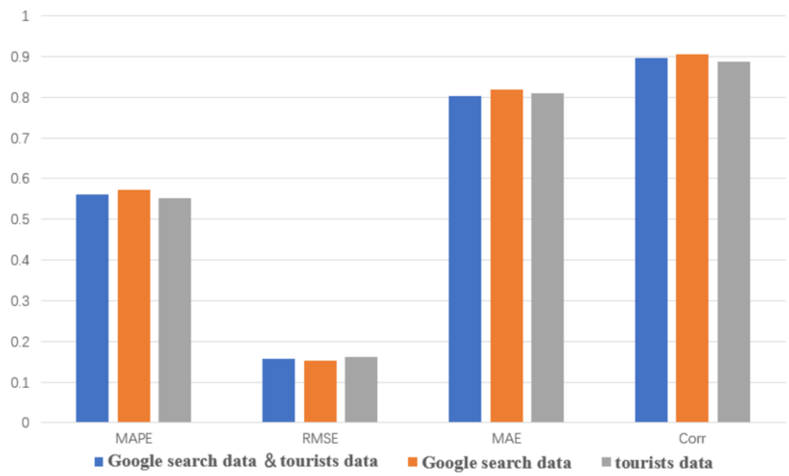

Figure 3. Results of ILI prediction using the stacking-based ensemble approach of different exogenous data in the Whole Taiwan Region. MAPE, RMSE, and MAE are the results after multiplying by 10, 100, and 1000, respectively.

Additionally, we conduct an ablation study to investigate the effects of different covariates on model performance. Figure 3 shows the effect of fixing historical ILI records as a covariate, in combination with other exogenous data, on the prediction results. The difference between the three schemes is negligible, and no scheme is always better than the other two for all evaluation metrics. This result indicates that the contributions of these external variables to the prediction are equivalent to the developed approach.

4. Discussion

In this study, we evaluate the utility of incorporating historical ILI records, Internet search data, and tourist flow information to make real-time ILI predictions. We consider four representative parametric predictive models, including ARIMA, LR, ARGO, and XGBoost; a 104-week sliding window is used for adaptive prediction. These four models are all feasible individual models that can provide a good influenza prediction result under limited conditions. However, each method also has shortcomings, and there is no way to prove which individual predictive model performs best. In this case, we try to fuse the models through an ensemble method. This study performs a stacking-based ensemble approach for LR, ARGO, and XGBoost.

The results show that our developed approach leverages the strengths of the individual models and thus provides robust real-time predictions. Additionally, we care about predictions at the city level, which provide a critical reference for the health resource deployment of the public health department in a small region. However, the Google search data for some cities are missing due to the small population, and other available exogenous data has to be considered as alternative data to conduct research on ILI activity.

We also apply the nowcasting approach to the period from December 8, 2019, to November 28, 2021, after the outbreak of COVID-19, and the results show that the developed approach is the best in ILI nowcasting compared to individual methods due to its robustness. Nevertheless, the ARIMA method performs better on a few evaluation metrics. There may be two reasons for this: 1) After the outbreak of COVID-19, people's attention shifted to COVID-19, which led to changes in search habits, and the role of Google search data decreased accordingly. 2) The ARIMA approach performs better in obtaining a long-term downward trend as the government's anti-epidemic measures significantly reduced the ILI intensity.

5. Conclusion

This study shows that the stacking-based ensemble approach enables more robust and accurate ILI prediction than the individual methods. The developed approach can be viewed as a feasible alternative to nowcast ILI in the areas where ILI activity has no constant seasonal trend. In the future, we will evaluate different ILI predictive methods in application to other geographical regions and incorporate other available ILI-related data such as Twitter posts [18-19]. Besides, at the methodological level, we will attempt to investigate the effect of temporal and spatial dependencies on ILI nowcasting and employ deep learning-based models for evaluating the performance of our developed approach [20].

References

[1] Influenza (Seasonal) [Internet]. [cited 2022 May 31]. Available from: https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)

[2] Yang JR, Hsu SZ, Kuo CY, Huang HY, Huang TY, Wang HC et al. An epidemic surge of influenza A(H3N2) virus at the end of the 2016–2017 season in Taiwan with an increased viral genetic heterogeneity. J Clin Virol. 2018 February 1;99–100:15–21.

[3] He Z, Tao H. Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China: A nine-year retrospective study. Intern J Infect Dis. 2018 September 1;74:61–70.
[4] Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Chunara R, et al. A case study of the New York City 2012-2013 influenza season with daily geocoded twitter data from temporal and spatiotemporal perspectives. Journal of Medical Internet Research. 2014 Oct 20;16(10):e3416.
[5] Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO. Proc Natl Acad Sci USA. 2015 Nov 24;112(47):14473–8.
[6] Priedhorsky R, Daughton AR, Barnard M, O'Connell F, Osthus D. Estimating influenza incidence using search query deceptiveness and generalized ridge regression. Vespignani A, editor. PLoS Comput Biol. 2019 Oct 1;15(10):e1007165.
[7] Watmaha J, Kamonsantiroj S, Pipanmaekaporn L. An integrated climate and spatio-temporal determinant for influenza forecasting based on convolution neural network. In: The 2021 9th International Conference on Computer and Communications Management. New York, NY, USA: Association for Computing Machinery; 2021. p. 104–10. (ICCCM '21).
[8] Kara A. Multi-step influenza outbreak forecasting using deep LSTM network and genetic algorithm. Expert Syst Appl. 2021 October 15;180:115153.
[9] Tsan YT, Chen DY, Liu PY, Kristiani E, Nguyen KLP, Yang CT. The prediction of influenza-like illness and respiratory disease using LSTM and ARIMA. Intern J Environ Res Public Health. 2022 Jan;19(3):1858.
[10] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. p. 785–94.
[11] Cheng HY, Wu YC, Lin MH, Liu YL, Tsai YY, Wu JH, et al. Applying machine learning models with an ensemble approach for accurate real-time influenza forecasting in Taiwan: development and validation study. J Med Internet Res. 2020 August 5;22(8):e15394.
[12] Amorim AN de, Deardon R, Saini V. A stacked ensemble method for forecasting influenza-like illness visit volumes at emergency departments. PLoS One. 2021 Mar 22;16(3):e0241725.
[13] Kang M, Zanin M, Wong SS. Subtype H3N2 influenza a viruses: an unmet challenge in the western pacific. Vaccines. 2022 Jan;10(1):112.
[14] Gong YN, Kuo RL, Chen GW, Shih SR. Centennial review of influenza in Taiwan. Biomed J. 2018 Aug;41(4):234–41.
[15] Xu Q, Gel YR, Ramirez Ramirez LL, Nezafati K, Zhang Q, Tsui KL. Forecasting influenza in Hong Kong with Google search queries and statistical model fusion. Cowling BJ, editor. PLoS One. 2017 May 2;12(5):e0176690.
[16] Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS. Monitoring influenza epidemics in China with search query from Baidu. PLoS One. 2013 May 30;8(5):e64323.
[17] Tian Y, Shi Y, Liu X. Recent advances on support vector machines research. Technol Econ Devel Econ. 2012 April 10;18(1):5–33.
[18] Baker QB, Shatnawi F, Rawashdeh S. Forecasting epidemic diseases with Arabic Twitter data and WHO reports using machine learning techniques. Bulletin of Electrical Engineering and Informatics. 2022 April 1;11(2):739–49.
[19] Poirier C, Hswen Y, Bouzillé G, Cuggia M, Lavenu A, Brownstein JS, et al. Influenza forecasting for French regions combining EHR, web and climatic data sources with a machine learning ensemble approach. PLoS One. 2021 May 19;16(5):e0250890.
[20] He Y, Zhao Y, Chen Y, Yuan H, Tsui K. Nowcasting influenza-like illness (ILI) via a deep learning approach using google search data: An empirical study on Taiwan ILI. Int J Intell Syst. 2022 Mar;37(3):2648–74.