The Implementation of a Real-time Monitoring and Prediction System of PM2.5 and Influenza-Like Illness Using Deep Learning

Chao-Tung Yang[1], Lung-Ying Lin[1], Yu-Tse Tsan[2,3], Po-Yu Liu[4], Wei-Cheng Chan[5]

[1] Department of Computer Science, Tunghai University, Taiwan
[2] Department of Emergency Medicine, Taichung Veterans General Hospital, Taiwan
[3] School of Medicine, Chung Shan Medical University, Taiwan
[4] Department of Internal Medicine, Taichung Veterans General Hospital, Taiwan
[5] Division of Occupational Medicine, Department of Emergency Medicine, Taichung Veterans General Hospital, Taiwan
ctyang@thu.edu.tw, didi3310781@gmail.com, janyuhjer@gmail.com, pyliu@vghtc.gov.tw, fufu110017@gmail.com

Abstract

Nowadays, deep learning is very popular and has excellent application performance in various fields. Specifically, in predicting disease outbreaks. It is possible to quickly and accurately predict disease outbreaks, lead to a significant research project in the medical field. However, there is a lack of technological development particularly in real-time monitoring of influenza-like illness (ILI). Moreover, there are still not sufficient ILI prediction models and results, while the incidence of ILI itself is high. In this paper, we propose an analysis and prediction of ILI outbreaks. In this case, deep learning analysis is applied to predict the ILI outbreaks to provide relevant information for relevant areas. The LSTM model is implemented to obtain information on AQI and ILI incident by using government open source data, research, and analysis of ILI and air quality indicators AQI. The visualization model using Highchart's with Django framework to present regional air quality and predicts whether influenza in the area will be outbreaks or not.

Keywords: Deep learning, Influenza-like illness, AQI, LSTM

1 Introduction

Influenza-like illness (ILI) is a medical diagnosis of possible influenza illnesses. It refers to an acute respiratory infection, caused most commonly by influenza viruses but also by other pathogens, such as parainfluenza viruses, adenoviruses, etc. All these kinds of pathogens cause having a common clinical manifestation. According to WHO surveillance case definitions, ILI is defined as an acute respiratory infection with measured fever ≥38℃ and cough, with onset within the last ten days; and severe acute respiratory infection (SARI) is defined as ILI that requires hospitalization [11-13].

The most common clinical manifestations of influenza are fever and cough, which often begins with an abrupt onset after an incubation period of 1 to 2 days. Besides, headache, malaise, or myalgia are the most troublesome symptoms, and the severity is related to the height of the fever. The systemic symptoms usually persist for three days, the typical duration of fever. Most of the impact of influenza is on compromising activities of daily living because of the symptoms, even in young, healthy individuals. It has been estimated that a typical case of influenza, on average, is associated with 5 to 6 days of restricted activity, 3 to 4 days of bed disability, and about three days lost from work or school. Direct medical costs of illness account for only about 20% of the total expenses of a case of influenza, with a significant proportion (30% to 50%) of the economic impact due to loss of productivity. Influenza-associated with decreased job performance in working adults and reduced levels of independent functioning in elderly [14-16].

The contributions of environmental factors to influenza epidemics have been observed in a different setting. In temperate climates in either hemisphere, epidemics occur almost exclusively in the winter (November to April in the Northern Hemisphere and May to September in the Southern Hemisphere). Seasonal periodicity is also observed in tropical climates, with increased activity during periods of low absolute humidity, although influenza can occur throughout the year and seasonal fluctuations are not as marked. The reasons for these seasonal changes are not entirely clear but might be the result of more favorable environmental conditions for virus survival. Seasonality may also be associated with behavioral changes that

*Corresponding Author: Yu-Tse Tsan; E-mail: janyuhjer@gmail.com
DOI: 10.3966/160792642019122007020

may increase transmissions, such as indoor crowding or school attendance. However, many potential environmental factors remain unidentified [20-22].

The relation of air pollutants and influenza activity have long suspected but difficult to confirm previously. The effects of air pollutants and influenza viruses on disease pathogenesis could potentially interact synergistically because both of them affect the respiratory system, from the nasal cavity and nasopharynx to the central airways and alveoli, and also through systemic effects on the cardiovascular system. The transmission of influenza viruses is believed to rely on short-distance dispersion of droplets. However, there is speculation that ambient air pollutants, especially particulate matter ≤ ten μm in aerodynamic diameter, might facilitate the spread of influenza viruses by providing condensation nuclei for the virus droplets. Air pollution is believed to be critical to long-range transmission of influenza viruses [29-30]. Recently, epidemiologic studies showed the impacts of different air pollutants on the influenza activity in different cities worldwide. Fine particles, O3, NO2, are most frequently reported drivers of influenza in these studies. However, all the studies suffer from retrospective, report biases, and small geographic scales [23-25].

In this paper, we propose an analysis and prediction of ILI outbreaks. In this case, deep learning analysis is applied to predict the ILI outbreaks to provide relevant information for relevant areas. The contribution of this paper are listed as follows:

(1) The LSTM model is implemented to obtain information on AQI and ILI incident by using government open source data, research, and analysis of ILI and air quality indicators AQI.

(2) The visualization model using Highchart's with Django framework to present regional air quality and predicts whether influenza in the area will be outbreaks or not.

2 Background Review and Related Works

In this section, we review the visualization, prediction, and evaluation used in our system.

2.1 Visualization

We used Django [7] that is an open source web application framework written in Python. Compared to other languages' Model-View-Controller (MVC) architecture, Django uses MVT's software design pattern to separate the program from the interface design to improve development efficiency. The framework is named after the Gypsy jazz guitarist Django Reinhardt in Belgium, in the hope that the website development will be as elegant as jazz. Django can (and has been) used to build almost any type of website—from content management systems and wikis to social networks and news sites. It works with any client framework and can provide content in almost any format including HTML, RSS, JSON, XML, etc. The website you are reading is based on Django.

Internally, although it provides choices for almost all of the features that may be needed, such as several popular repositories, stencil engines, etc., it can be extended to use other components if needed.

Django has the following features:

Free open source code

- Focus on rapid development, high efficiency
- Compliance with DRY (Don't Repeat Yourself) code, dedicated to easy to understand and elegant code
- Built-in distribution system that allows components in the application to be pre-defined the signals communication with each other.
- A system for extending the capabilities of a functional template engine.

2.2 LSTM Networks

We used deep learning algorithm Long Short-Term Memory (LSTM), a time recurrent neural network (RNN) [1-3]. LSTM solves the problem of gradient vanishing. The difference between LSTM and RNN is that it has more gates. Its function is for information. For the purpose of filtering, the filtering is based on x(t) and h(t-1), and the output of gate is 0 to 1. In LSTM [5], the network constructs three gate control information flows, which are input gates i_t: how much information can flow into memory cell c(t), forgetting gate f(t): last time memory cell c (the information in t) can be accumulated into the current memory cell c(t), and the output gate o(t): how much of the current cell's memory cell (t) can flow into the current hidden state h(t).

At the bottom of Figure 1, there are four function units. The leftmost function is the input of the block. The third one is determined by the gate to determine whether the input value can be passed to the leftmost block. The second one on the left is the input valve. If the output value here is similar to Zero, the valve will be closed and will not enter the next layer. The third one on the left is the forgetting valve. If the output value here is close to zero, the value remembered in the block will be deleted. The fourth, the rightmost input, is the output valve. It can determine whether the input in the block memory can be output.

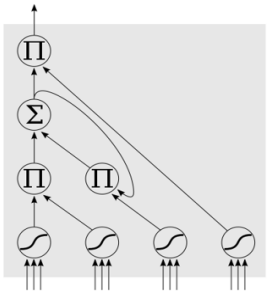

Figure 1. LSTM model

LSTM [6] only solves the gradient disappearance problem and does not solve the gradient explosion problem. Therefore, we must do some operations on the gradient during the training to prevent the gradient explosion problem. That is, the monitor gradient is greater or smaller than a threshold every update. The value is forced to give a fixed value. In the implementation of tensorflow, we can use tf.clip_by_value or tf.clip_by_norm to solve the gradient explosion problem.

2.3 Evaluation

To evaluate about how well our LSTM model can forecast, we used Mean Absolute Percent Error (MAPE). The MAPE values of the training set, the test set and the verification set are respectively tested as the criteria to rate the model. The average of the MAPE is formulated as follows. Table 1 describes the MAPE percentage accuracy status.

$$\sum \frac{\left| \frac{observed_t - predict_t}{observed_t} x100 \right|}{n}$$

Table 1. MAPE assessment

MAPE value	Accuracy of Forecast
Less than 10%	Highly Accurate Forecast
11% to 20%	Good Forecast
21% to 50%	Reasonable Forecast
More than 51%	Inaccurate Forecast

2.4 Related Works

Pulmonary complications are more frequent in elderly than in any other age groups. Morbidity and mortality are very high in individuals with certain high-risk underlying medical conditions, including adults and children with cardiovascular or pulmonary diseases such as asthma, or those requiring regular medical care because of chronic metabolic disease, renal dysfunction, hemoglobinopathies, or immunodeficiency and in those with neurologic conditions that compromise handling respiratory secretions.

In addition to the mortality resulting in part from the pulmonary complications, another unique feature of influenza is the epidemic nature of the disease. The influenza virus has been causing recurrent epidemics of the febrile respiratory disease every 1 to 3 years for at least the past 400 years and usually occurs in outbreaks of varying severity almost every winter in temperate climates and year-round in tropical climates [17-19].

A real-time monitoring and prediction system is crucial to the successful control of an influenza epidemic [10]. Existing methods mostly rely on expensive surveys of health care facilities, typically with lag times of days to weeks for influenza reporting. In light of the highly contagious nature and rapid transmission speed of influenza, the delay makes any meaningful control measures impossible [31-32]. The preliminary data from air pollutants and influenza epidemiologic studies shed light on the potential for the application of early warning system. In the project, we aim to develop a machine-learning model to unravel and predict the hidden dynamics of influenza using real-time air quality index (AQI) data [26-28].

3 System Design and Implementation

In this section, we introduce the system design architecture and implementation of the Real-time Monitoring and Prediction System of PM2.5 and Influenza-Like Illness.

3.1 System Architecture

In this work, we use python to automatically capture information from the Department of Health and Welfare's Disease Control Agency website and store it in the MySQL database. After store the data then we conduct the data analysis using deep learning through TensorFlow and Keras. Return to MySQL. The website part uses django as the web server, the webpage beautifies the bootstrap [9] template, and then uses Django's built-in syntax to get the data from the database and display the graphed data through Highcharts [8]. Figure 2 shows the system architecture.

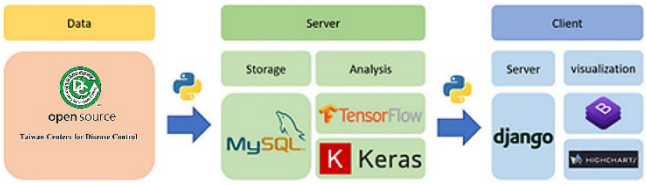

Figure 2. System architecture

3.2 Dataset

At present, the Department of Health and Welfare's Disease Control Bureau uses for outpatient flu visits machine learning algorithms, such as: ARIMA, RF, SVR, XGB, and Ensemble. In our experiments, we will use the deep learning LSTM algorithm to the data of flu-attendances. In this case, we will provide another model option to the experts, physicians, and the general public. Table 2 describes The Department of Health and Welfare's Disease Control Agency model selection comparison.

3.3 LSTM System Flow

LSTM uses "cell station" to enhance current decisions as describe in Figure 4, using three control valves (Gate). LSTM determine the storage and use of:

(1) Memory Add a memory branch. When the time is updated, the current memory is by the C_t represented symbol, and the Forget Gate and Input

Gate are used to

Table 2. The department of health and welfare's disease control agency model selection

Machine Learning Algorithms	Week 37 09/09 to 09/15	Week 38 09/16 to 09/22	Week 39 09/23 to 09/29	Week 40 09/30 to 10/06
Ensemble	44,121	44,506	43,081	44,26
ARIMA	44,02	44,142	42,717	44,357
RF	45,753	47,295	47,401	48,529
SVR	45,109	45,751	45,4	47,471
XGB	44,249	43,687	41,386	43,352

determine whether to update the memory.

(2) Forget Gate (by f_t indicated), if the current value is a new value or a word with the opposite value, then the previous value will be filtered by this valve, otherwise it may be retained in the cell.

(3) Inlet valve (Input Gate, as i_t determines the current input (Input) and the newly generated Memory Cell Candidate whether to join the long-term, this valve is Sigmoid function, the value is converted to 0 or 1.

(4) Output Gate: Determines whether the current value is added to the output. This valve is also a Sigmoid function that converts the value to 0 or 1.

Finally, for long-term memory to be added to the output, usually using the tanh function, the value will fall between [-1, 1], and -1 means to remove long-term memory. Combining three valves, the final formula is as follows:

$$i_t = \sigma(w_{xi} + w_{hi} h_{t-i} + b_i),$$
$$f_t = \sigma(w_{xf\ xt} + w_{hf} h_{t-i} + b_f),$$
$$o_t = \sigma(w_{xo\ xt} + w_{ho} h_{t-i} + b_o),$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \sigma_t(W_c x_t + U_c h_{(t-1)} + b_c),$$
$$h_t = o_t \odot \sigma_h(c_t).$$

We used LSTM parameter system selection as follows:

(1) Hyperparameter Look_back: 14
(2) Batch_size: 32
(3) Loss Function: mean squared error
(4) Optimizer: adam

4 Experimental Results

In this section, we demonstrate our experimental results in three categories: Visualization, Prediction and Evaluation experimental results.

4.1 Visualization

Our project used historical AQI data plus seasonal, wind direction and weather factors to predict the incidence of influenza at the same time, and integrated the prediction results to the general public, hospital physicians and the Department of Health to prevent the spread of disease.

More detailed of AQI and ILI reports make it easier for physicians to see the correlation between ILI and air quality. Also, we use artificial intelligence to predict the class. In the flu outbreak season, the results of the professional analysis will be combined with the above statements to provide the doctors with a view to prevent the ILI epidemic.

We used Highchart to visualize the graph of air quality and ILI data as shown in Figure 3 in weeks' interval.

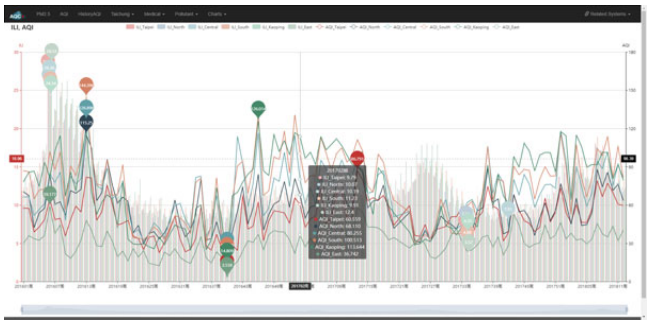

Figure 3. Graph visualization of ILI

Combined with Google Maps, AQI forecasts from all regions of Taiwan are instantly displayed on the webpage, making it easy for people who need to travel to easily check the air quality of the foreign counties and decide whether to travel to the area as shown in Figure 4.

Figure 4. AQI status on the map

The two dash lines describe the average of ILI and AQI value. The red dash line shows that the average of ILI case is 11.45. The blue dash line shows the average of AQI value is 61.37.

According to the region, the AQI is calculated based on National AQI index value, as shown in Figure 5. The map depicts Taiwan regions with AQI level value in each areas distinguished by color. Green is good, with the AQI index between 0-50. Yellow is Moderate, with the AQI value is 5-100. Orange is Unhealthy for Sensitive Groups, with the AQI value is 101-150. Light

Red is Unhealthy, with the AQI value is 151-200. Purple is Very Unhealthy, with the AQI value is 201-300. Dark Red is Hazardous, with the AQI value up to 301.

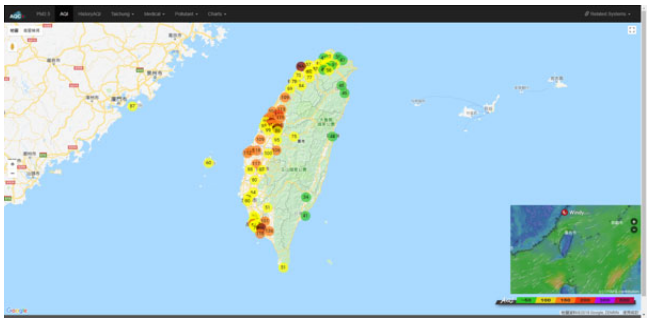

Figure 5. Geographical division details (Central District)

In addition, historical information such as temperature and humidity is added for rough evaluation, as shown in Figure 6.

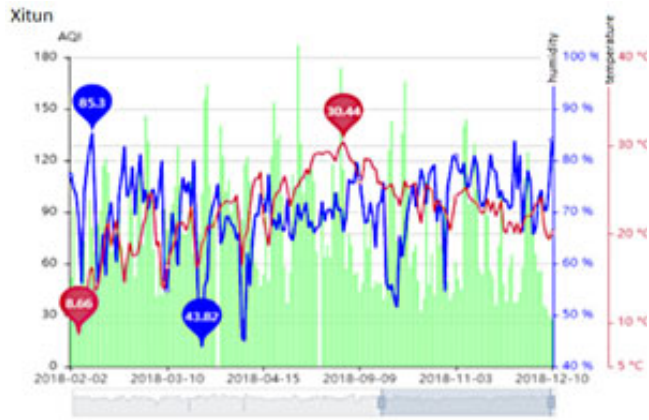

Figure 6. AQI, temperature and humidity historical data (Xitun station)

Our system visualization also uses the influenza outbreak calculation method based on the World Health Organization (WHO). The calculation examines whether the flu-like disease control status is still in a safe state, as shown in Figure 7.

Figure 7. Influenza outbreak line chart in 2017

In this system, we recorded the historical values of ILI and Air Quality Index (AQI) regions, and displayed them in weekly time intervals. In the Figure 8, the bar chart is used to show the change of ILI weekly values. The change of the AQI value is represented by a line graph, which represents the continuous value.

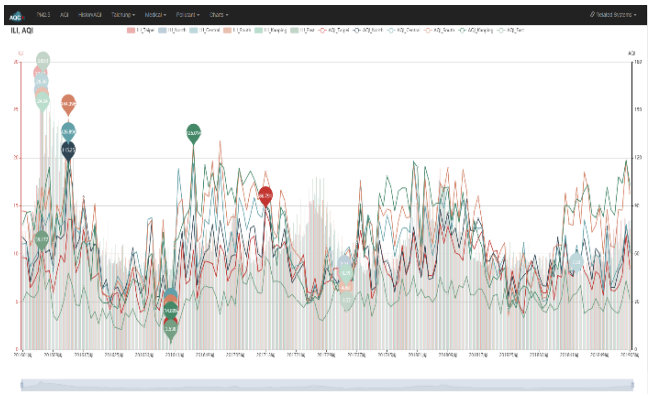

Figure 8. The line chart of ILI and AQI in weekly

In order to more intuitively observe the relationship between PM2.5, influenza-like, pneumonia and enterovirus, we will use the weekly data to graphically present, in Figure 9, we can clearly see the weekly 4 different data to facilitate correlation and change between observations.

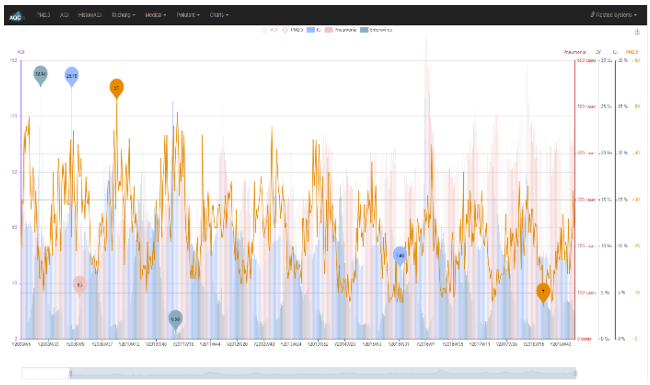

Figure 9. PM2.5 influenza, pneumonia, and enterovirus trend map

4.2 Prediction

We conducted 20 predictions as shown in Table 3 for outpatient-emergency visits in the 36th week of 2018, and recorded the predicted value of each visit and the number of actual visits to the 36 weeks of year 2018.

Table 4 presents the comparison of prediction in 2018, in a three-week outpatient influenza-predicted. In this case, we compare LSTM model with the selected model by the Department of Health and Welfare.

Table 3. LSTM prediction in 36 weeks of year 2018

36 weeks of year 2018 Actual outpatient influenza visits: 44049 cases		
Number of experiment	LSTM prediction	The gap value
1	43988	61
2	43563	486
3	43974	75
4	43793	256
5	42931	1118
6	43519	530
7	43229	820
8	45420	1371
9	43222	827
10	43626	423
11	44254	-205
12	44295	-246
13	43217	832
14	43314	735
15	42936	1113
16	43099	950
17	44186	-137
18	43870	179
19	44532	-483
20	43804	245

Table 4. Different model predictive values Outpatient influenza-like

Model	38 weeks	39 weeks	40 weeks
Ensemble	44506	43081	44260
ARIMA	44142	42717	44357
RF	47295	47401	48529
SVR	45751	45400	47471
XGB	43687	41386	43352
LSTM	45909	47559	49348

- Week 38 (09 / 16 - 09 / 22)
- Week 39 (09 / 23 - 09 / 29)
- Week 40 (09 / 30 - 10 / 06)

We used historical AQI data plus seasonal, wind direction and weather factors to compare with the incidence of disease at the same time (especially the epidemic). Then, calculating the correlation between air quality and disease occurrence, and make the correlation between them. Make predictions about the disease and integrate the predictions to the general public, hospital physicians, and the Department of Health as described in Figure 10 to prevent the spread of the disease. The graph shows the prediction values of PM 2.5, NO2, PM10, O3, and CO within six hours ahead. The black line is NO2, the light blue line is PM 2.5, the green line is PM10, the dark blue line is CO, and the orange line is O3 time series data.

We builds a Linux-based Tensorflow deep learning development environment and uses the Keras development kit for experiments. Using the deep learning technique Recurrent Neural Network (RNN) [3-5] Long Short- Term Memory Model (LSTM) as a learning framework to find the interaction between ILI

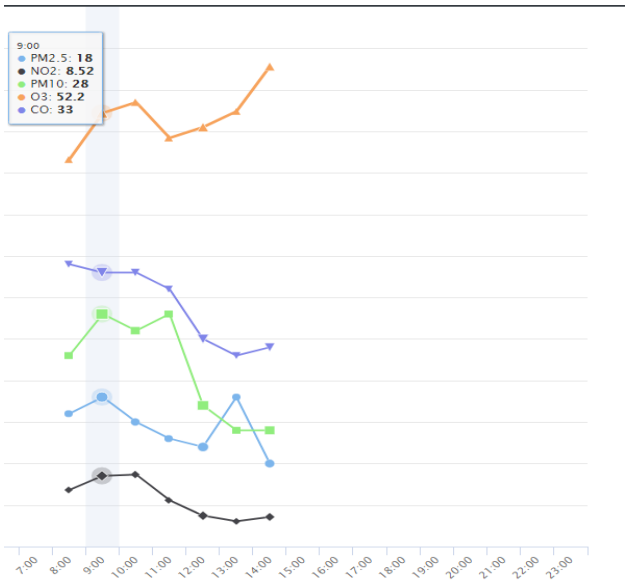

Figure 10. The prediction of AQI

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm_1 (LSTM)                (None, 4, 256)            265216
_________________________________________________________________
lstm_2 (LSTM)                (None, 4, 256)            525312
_________________________________________________________________
time_distributed_1 (TimeDist (None, 4, 1)              257
_________________________________________________________________
flatten_1 (Flatten)          (None, 4)                 0
_________________________________________________________________
dense_2 (Dense)              (None, 128)               640
_________________________________________________________________
dense_3 (Dense)              (None, 4)                 516
=================================================================
Total params: 791,941
Trainable params: 791,941
Non-trainable params: 0
_________________________________________________________________
```

Figure 11. LSTM neural layer architecture diagram

and PM2.5 to predict the future ILI value. Figure 11 shows the detail of LSTM neural layer architecture diagram.

In this experiment, the predicted number of target weeks is set to four weeks, and the time interval of four weeks is predicted in the experiment as shown in Figure 12.

```
X_test
array([[[-0.096855  , -0.26634126],
        [-0.09384899,  0.12839559],
        [-0.11288706,  0.07576401],
        [-0.12992113, -0.34528862]],

       [[-0.09384899,  0.12839559],
        [-0.11288706,  0.07576401],
        [-0.12992113, -0.34528862],
        [-0.15497123, -0.24002547]],

       [[-0.11288706,  0.07576401],
        [-0.12992113, -0.34528862],
        [-0.15497123, -0.24002547],
        [-0.097857  , -0.02949915]],

       [[-0.12992113, -0.34528862],
        [-0.15497123, -0.24002547],
        [-0.097857  , -0.02949915],
        [-0.08483095, -0.18739389]],
Y_test
array([[ 8.52,  9.09,  9.22,  8.33],
       [ 9.09,  9.22,  8.33,  9.23],
       [ 9.22,  8.33,  9.23,  9.05],
       [ 8.33,  9.23,  9.05, 10.27],
       [ 9.23,  9.05, 10.27,  9.97],
```

Figure 12. The schematic diagram of the entirely training and verification data

By using the above data to train the RNN-LSTM model, the project divides the raw data into three categories, namely the training set, the test set, and the verification set. The purpose is to use the training set and the verification set during training, and to verify the identity of the prediction module for new data after the training is completed, the verification set that does not enter the training process is used as the identity of the new data. This standard will be more credible than using a test set as a criterion for judgment. Figure 13 shows the training set data and prediction results.

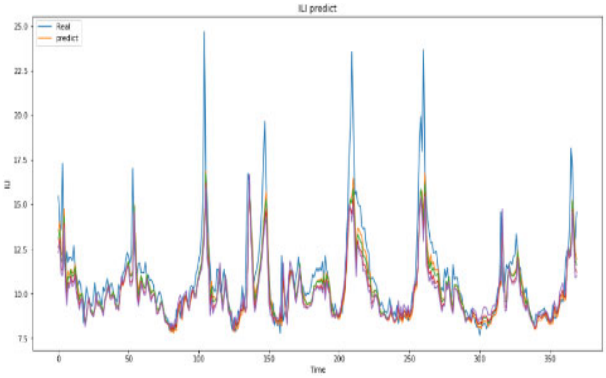

Figure 13. Influenza patient visits graph from 2012 week 1-2018 week 35 in all ages (the training and test result)

We conduct the test and validation of prediction as shown in Figure 14 and Figure 15 in order to evaluate our MAPE value.

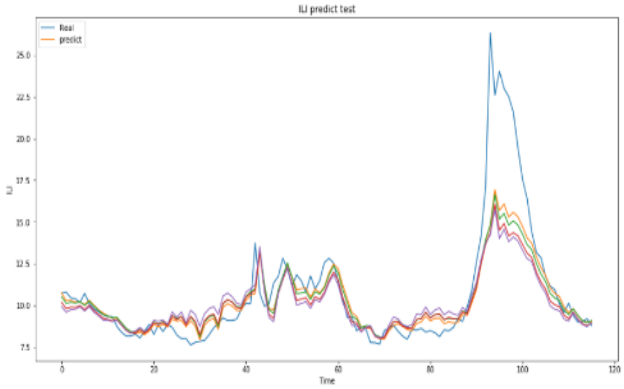

Figure 14. Test set original data and prediction results

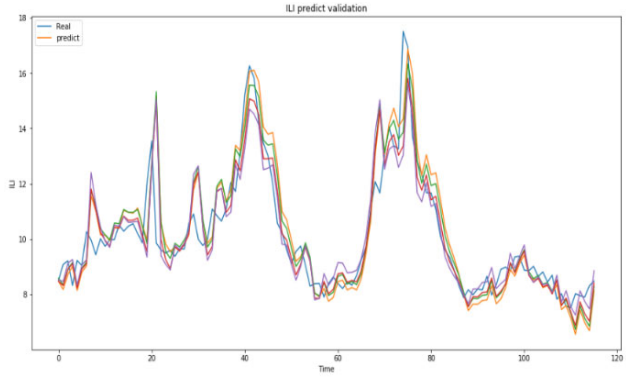

Figure 15. Validation set original data and prediction results

4.3 MAPE Evaluation

In this section we evaluated the MAPE value predictions. We found the predicted MAPE is less than 10%, and the fourth week MAPE value is only slightly above the threshold of 10% which means High Accuracy based on Table 1. Figure 16 describes the training set, test set, and verification set of the MAPE results.

```
#Train MAPE
arrayPreTrain = model.predict(X_train)
mape(Y_train,arrayPreTrain,4)
```

```
[6.9093326871673755, 8.1334512904917, 9.779850121136914, 10.334333663385303]
```

```
#Test MAPE
arrayPreValidation = model.predict(X_validation)
mape(Y_validation,arrayPreValidation,4)
```

```
[7.484256033701232, 9.247477561611289, 9.723925070873488, 11.244605859708637]
```

```
#Validation MAPE
mape(Y_test,arrayPre,4)
```

```
[7.259693369235709, 8.852144309049296, 9.094136027308124, 10.102900041440657]
```

Figure 16. Training set, test set, and verification set MAPE results

5 Conclusion and Future Works

Our study provides the framework of a real-time monitoring and prediction of ILI and AQI system using deep learning. In our experiment, we compared the AQI data and ILI data from open data Taiwan government. The system demonstrates timely predictions and better visualization combined with Google Map. The data from each station and the surrounding areas can be visualized based on colors. We can provide the color difference status of the AQI and ILI outbreaks. There is a differentiated login page between the doctors and public information related AQI and ILI data. The optimization of the model, the simplification of the algorithm, and the selection of data input methods will also be adjusted to fit the system. In the term of prediction, we implemented the LSTM model to predict the influenza-like illness in 36 weeks of 2018. To evaluate our prediction, we use MAPE calculation. We found the predicted MAPE is less than 10%, and the fourth week MAPE value is only slightly above the threshold of 10% which means High Accuracy.

In the future, the system will focus on improving the accuracy of the prediction model, improving and optimizing the algorithm model. Also, this system can be expanded by developing a version of the mobile device. The mobile version has a function that can transmit photos of areas with poor air quality in real time. The system will also increase the discussion area to allow the public to focus on air quality issues. The areas that have not yet installed the station will be used to further increase the coverage of the system. Therefore, people in these areas without the station can

also clearly understand the air quality and ILI outbreaks in the current area.

Acknowledgments

This work was funded by the Ministry of Science and Technology (MOST), Taiwan, under grant number 106-3114-M-029-001-A and 108-2119-M-029-001-A. In addition, this work was also funded in part by the Taichung Veterans General Hospital (TCVGH), Taiwan, under grant number TCVGH-T1077803.

References

[1] A. Graves, Generating *Sequences with Recurrent Neural Networks*, arXiv:1308.0850, August, 2013.

[2] I. Sutskever, O. Vinyals, Q. V. Le, Sequence to Sequence Learning with Neural Networks, *Advances in Neural Information Processing Systems*, Montreal, Canada, 2014, pp. 3104-3112.

[3] R. Pascanu, T. Mikolov, Y. Bengio, On the Difficulty of Training Recurrent Neural Networks, *International Conference on Machine Learning*, Atlanta, GA, USA, 2013, pp. 1310-1318.

[4] K. Hwang, W. Sung, *Online Sequence Training of Recurrent Neural Networks with Connectionist Temporal Classification*, arXiv:1511.06841, November, 2015.

[5] R. Pascanu, C. Gulcehre, K. Cho, Y. Bengio, *How to Construct Deep Recurrent Neural Networks*, arXiv:1312.6026, December, 2013.

[6] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-C. Woo, Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting, *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, Montreal, Canada, 2015, pp. 1-9.

[7] Django, https://docs.djangoproject.com/en/2.1/.

[8] Highcharts, https://www. Highcharts.com/docs.

[9] Bootstrap, https://startbootstrap.com/template-categories/all/.

[10] H. Achrekar, A. Gandhe, R. Lazarus, S.-H. Yu, B. Liu, Predicting Flu Trends Using Twitter Data, *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Shanghai, China, 2011, pp. 702-707.

[11] M. L. Bell, D. L. Davis, T. Fletcher, A Retrospective Assessment of Mortality from the London Smog Episode of 1952: The Role of Influenza and Pollution, *Environmental Health Perspectives*, Vol. 112, No. 1, pp. 6-8, January, 2004.

[12] P. J. Birrell, G. Ketsetzis, N. J. Gay, B. S. Cooper, A. M. Presanis, R. J. Harris, A. Charlett, X.-S. Zhang, P. J. White, R. G. Pebody, D. D. Angelis, Bayesian Modeling to Unmask and Predict Influenza A/H1N1pdm Dynamics in London, *Proceedings of the National Academy of Sciences*, Vol. 108, No. 45, pp. 18238-18243, November, 2011.

[13] G. Boivin, I. Hardy, G. Tellier, J. Maziade, Predicting Influenza Infections during Epidemics with Use of a Clinical Case Definition, *Clinical Infectious Diseases*, Vol. 31, No. 5, pp. 1166-1169, November, 2000.

[14] D. A. Broniatowski, M. J. Paul, M. Dredze, National and Local Influenza Surveillance through Twitter: An Analysis of the 2012-2013 Influenza Epidemic, *PloS One*, Vol. 8, No. 12, e83672, December, 2013.

[15] S. T. Brown, J. H. Tai, R. R. Bailey, P. C. Cooley, W. D. Wheaton, M. A. Potter, R. E. Voorhees, M. LeJeune, J. J. Grefenstette, D. S. Burke, S. M. McGlone, B. Y. Lee, Would School Closure for the 2009 H1N1 Influenza Epidemic Have been Worth the Cost?: A Computational Simulation of Pennsylvania, *BMC Public Health*, Vol. 11, No. 1, Article number 353, May, 2011.

[16] D. L. Chao, M. E. Halloran, V. J. Obenchain, I. M. Longini, FluTE, A Publicly Available Stochastic Influenza Epidemic Simulation Model, *PLoS Computational Biology*, Vol. 6, No. 1, e1000656, January, 2010.

[17] W. C. Cockburn, P. J. Delon, W. Ferreira, Origin and Progress of the 1968-69 Hong Kong Influenza Epidemic, *Bulletin of the World Health Organization*, Vol. 41, No. 3-4-5, pp. 345-348, 1969.

[18] A. Culotta, Towards Detecting Influenza Epidemics by Analyzing Twitter Messages, *Proceedings of the First Workshop on Social Media Analytics*, Washington, DC, USA, 2010, pp. 115-122.

[19] L. E. Davis, G. G. Caldwell, R. E. Lynch, R. E. Bailey, T. D. Chin, Hong Kong Influenza: The Epidemiologic Features of a High School Family Study Analyzed and Compared with a Similar Study during the 1957 Asian Influenza Epidemic, *American Journal of Epidemiology*, Vol. 92, No. 4, pp. 240-247, October, 1970.

[20] A. Flahault, S. Letrait, P. Blin, S. Hazout, J. Menares, A.-J. Valleron, Modelling the 1985 Influenza Epidemic in France, *Statistics in Medicine*, Vol. 7, No. 11, pp. 1147-1155, November, 1988.

[21] E. Goldstein, S. Cobey, S. Takahashi, J. C. Miller, M. Lipsitch, Predicting the Epidemic Sizes of Influenza A/H1N1, A/H3N2, and B: A Statistical Method, *PLoS Medicine*, Vol. 8, No. 7, e1001051, July, 2011.

[22] J. P. Horcajada, T. Pumarola, J. A. Martinez, G. Tapias, J. M. Bayas, M. De la Prada, F. García, C. Codina, J. M. Gatell, M. T. Jiménez de Anta, A Nosocomial Outbreak of Influenza during a Period without Influenza Epidemic Activity, *European Respiratory Journal*, Vol. 21, No. 2, pp. 303-307, February, 2003.

[23] N. P. Johnson, J. Mueller, Updating the Accounts: Global Mortality of the 1918-1920 "Spanish" Influenza Pandemic, *Bulletin of the History of Medicine*, Vol. 76, No. 1, pp. 105-115, Spring, 2002.

[24] J. T. F. Lau, S. Griffiths, K.-C. Choi, C. Lin, Prevalence of Preventive Behaviors and Associated Factors during Early Phase of the H1N1 Influenza Epidemic, *American Journal of Infection Control*, Vol. 38, No. 5, pp. 374-380, June, 2010.

[25] K. Lin, D. Y.-T. Fong, B. Zhu, J. Karlberg, Environmental Factors on the SARS Epidemic: Air Temperature, Passage of Time and Multiplicative Effect of Hospital Infection, *Epidemiology & Infection*, Vol. 134, No. 2, pp. 223-230, April, 2006.

[26] A. S. Mugglin, N. Cressie, I. Gemmell, Hierarchical Statistical Modelling of Influenza Epidemic Dynamics in Space and Time, *Statistics in Medicine*, Vol. 21, No. 18, 2703-2721, September, 2002.

[27] F. Pervaiz, M. Pervaiz, N. A. Rehman, U. Saif, FluBreaks: Early Epidemic Detection from Google Flu Trends, *Journal of Medical Internet Research*, Vol. 14, No. 5, e125, October, 2012.

[28] P. Quenel, W. Dab, Influenza A and B Epidemic Criteria Based on Time-series Analysis of Health Services Surveillance Data, *European Journal of Epidemiology*, Vol. 14, No. 3, pp. 275-285, April, 1998.

[29] L. Simonsen, M. J. Clarke, G. D. Williamson, D. F. Stroup, N. H. Arden, L. B. Schonberger, The Impact of Influenza Epidemics on Mortality: Introducing a Severity Index, *American Journal of Public Health*, Vol. 87, No. 12, pp. 1944-1950, December, 1977.

[30] C. Viboud, P.-Y. Boëlle, F. Carrat, A.-J. Valleron, A. Flahault, Prediction of the Spread of Influenza Epidemics by the Method of Analogues, *American Journal of Epidemiology*, Vol. 158, No. 10, pp. 996-1006, November, 2003.

[31] C. Viboud, T. Tam, D. Fleming, A. Handel, M. A. Miller, L. Simonsen, Transmissibility and Mortality Impact of Epidemic and Pandemic Influenza, with Emphasis on the Unusually Deadly 1951 Epidemic, *Vaccine*, Vol. 24, No. 44-46, pp. 6701-6707, November, 2006.

[32] C. M. Wong, L. Yang, T. Q. Thach, P. Y. K. Chau, K. P. Chan, G. N. Thomas, T. H. Lam, T. W. Wong, A. J. Hedley, J. S. M. Peiris, Modification by Influenza on Health Effects of Air Pollution in Hong Kong, *Environmental Health Perspectives*, Vol. 117, No. 2, pp. 248-253, February, 2009.

Biographies

Chao-Tung Yang is Distinguished Professor of Computer Science at Tunghai University in Taiwan. He received his Ph.D. in computer science from National Chiao Tung University in July 1996. In August 2001, he joined the Faculty of the Department of Computer Science at Tunghai University. He is serving in a number of journal editorial boards, including Future Generation Computer Systems, International Journal of Communication Systems, KSII Transactions on Internet and Information Systems. He has published more than 280 papers in journals, book chapters and conference proceedings. His present research interests are in cloud computing and service, big data, and parallel computing. He is a member of the IEEE Computer Society and ACM.

Lung-Ying Lin is under graduate student at Computer Science, Tunghai University. In August 2018, he joined Taiwan AI academy as teaching assistant. His present research interests are in machine learning, date analyzing, deep learning, computer vision and reinforcement learning.

Yu-Tse Tsan is Assistant Professor of Medicine and Medical Physician in Taichung Veterans General Hospital in Taiwan. He received his Ph.D. in public heath from Institute of Occupational Medicine and Industrial Hygiene, National Taiwan University in June 2013. In August 2000, he joined the Department of Emergency Medicine at Taichung Veterans General Hospital. He ever won the Excellent Paper Awards of Professor Chen Gongbei Memorial Award. His present research interests are in pharmacoepidemiology, environmental and occupational health and big data programming.

Po-Yu Liu is a clinical consultant at the Taichung Veterans General Hospital. He achieved and maintains Board Certification in Internal Medicine, Emergency Medicine, and Infectious Diseases. During his Ph.D. at National Chung Hsing University, he studied environment-microbe-host interactions. He focuses his research on increasing knowledge that will lead to better diagnosis and management of emerging infections.

Wei-Cheng Chan received the B.S. degree in Public health from Kaohsiung Medical University, Kaohsiung in 2009 and M.S. degree in Graduate Institute of Microbiology and Public Health from National Chung Hsing University, Taichung in 2011, Taiwan. Now he is an Assistant in Department of Emergency Medicine Taichung Veterans General Hospital, Taichung, Taiwan. His research interests epidemiology, biostatistics, biomedicine and virology.