Prediction of Outpatient Visits for Upper Respiratory Tract Infections by Machine Learning of PM2.5 and PM10 Levels in Taiwan

Pei-Hsuan Yang[1], Mi-Tren Hsieh[1], Gen-Min Lin[1,2], Mei-Juan Chen[1], *Senior Member, IEEE,* Chia-Hung Yeh[3,4], *Senior Member, IEEE,* Zhi-Xiang Huang[3], Chieh-Ming Yang[1]

1. Department of Electrical Engineering, National Dong Hwa University, Hualien, Taiwan

2. Department of Medicine, Hualien Armed Forces General Hospital, Hualien, Taiwan

3. Department of Electrical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan

4. Department of Electrical Engineering, National Taiwan Normal University, Taipei, Taiwan

***Abstract*— Particulate Matter (PM) 2.5 and PM10 are referred as a mixture of liquid droplets and solid particles in the air with diameters ≤2.5 μm and ≤10 μm, respectively. Both PM2.5 and PM10 can deposit on respiratory tract and trigger inflammatory reactions, which makes the respiratory tract predisposed to infections. The study used machine learning on daily PM2.5 and PM10 levels of consecutive 30 days from the open website datasets of Environment Protection Administration between Dec. 2008 and Dec. 2016 to predict the subsequent one-week outpatient visits for upper respiratory tract infections (URI) from the Centers for Disease Control (CDC) in Taiwan between Jan. 2009 and Dec. 2016. The weekly URI cases were classified by tertile as high, moderate, and low volumes. In general, both URI burden and PM levels peak in winter and spring seasons. The testing used the mid-month dataset of each season (Jan., Apr., Jul., and Oct.), and the training used the other months datasets. In the nationwide data analysis, PM2.5 and PM10 levels input to the multilayer perceptron (MLP) can precisely predict the degree of URI number for the elderly (89.05% and 88.32%, respectively) and the overall population (81.75% and 83.21%, respectively). In conclusion, machine learning of PM2.5 and PM10 levels could accurately predict the burden of outpatient visits for URI in Taiwan.**

Keywords—PM2.5; PM10; upper respiratory infections; machine learning, air pollution

I. INTRODUCTION

Particulate Matter (PM) 2.5 and PM10, also known as particle pollutions are referred to a mixture of liquid droplets and solid particles in the air with diameters ≤2.5 μm and ≤10 μm, respectively. Both PM2.5 and PM10 can deposit on respiratory tract and may trigger inflammatory reactions such as an increase of plasma interleukin-6 and fibrinogen [1]. Inflammation process related to air pollution might decrease innate immunity and predispose robust individuals to illness like acute upper respiratory tract infections (URI). Several observational studies have uncovered that PM2.5 and PM10 levels may be associated with the occurrence of URI [2-3] and increase the risk of death related to hospitalized pneumonia [4]. Machine learning of potential hazardous exposures at baseline has been successfully applied to predicting the occurrence of several clinical diseases like myocardial infarction or the risk of mortality in the previous studies [5]. Accordingly, we utilized the method of machine learning of PM2.5 and PM10 levels to predict the development of acute URI in outpatients.

II. METHODS

A. Data Collection

The datasets of outpatient visits for URI were obtained from the website of Centers for Disease Control (CDC) of Taiwan during the period from December 2008 to December 2016. Clinical physicians made the diagnosis of URI, and the cases of URI were retrieved from the Taiwan Nationwide Health Insurance records based on the ICD-9 codes for URI. The datasets of PM2.5 and PM10 levels were obtained from the Environment Protection Administration, and the levels were measured and collected from all ambient air quality monitoring sites spread around Taiwan.

B. Multilayer Perceptron Model (MLP)

Figure 1 shows the multilayer perceptron (MLP) architecture with forward and back propagation learning algorithm for the predictive accuracy of PM2.5 and PM10 levels to one-week outpatient visits of URI, which is classified by tertile as high, moderate, and low volumes. The testing used mid-month dataset of each season, which account for 33% of all datasets (Jan. (winter), Apr. (spring), Jul. (summer), and Oct. (fall)), and the training used the other months datasets, which account for 67% (Dec. and Feb. (winter), Mar. and May (spring), Jun. and Aug. (summer), and Sep. and Nov. (fall)). Training and testing procedures were repeated multiple times to determine optimal outcomes.

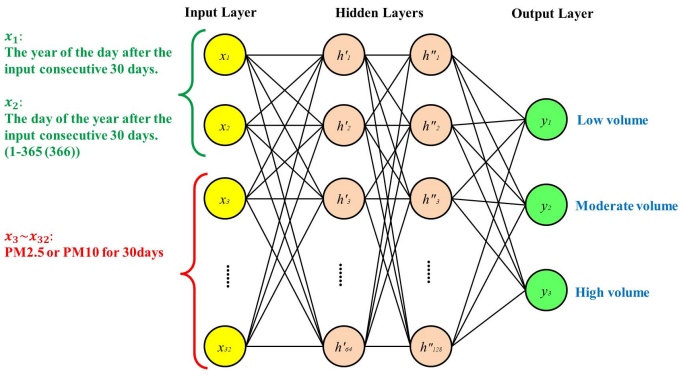

Figure 1. The multilayer perceptron model (MLP) for proposed algorithm

C. Statistical Analysis

Average daily PM2.5 and PM10 levels for consecutive 30 days were the inputs of MLP model to predict the following one-week URI burden. The reason to use one-week volumes of URI as outputs is that the time lag effect [2] of patients to seek medical consultation. Also, the accuracy of machine learning

for the overall and elderly (defined as age ≧65 years) patients were estimated.

III. ANALYSIS AND RESULTS

Figure 2 shows the average daily levels of PM2.5 and PM10 from Dec. 2008 to Dec. 2016 and the average numbers of outpatient visits for URI for the overall and the elderly populations of each month from Jan. 2009 to Dec. 2016. As showed, the PM2.5 and PM 10 levels distribute as a diurnal curve and peak in winter and spring seasons. The PM2.5 levels were between 15 and 46 μg/m^3, and the PM10 levels range between 30 and 100+ μg/m^3. The occurence of total and elderly outpatient visits for acute URI are most prevalent in winter and spring seasons, which is correlated with the PM2.5 and PM10 levels. With regard to the burden of outpatient visits for URI, the monthly overall URI cases range from 35,000 to 70,000, and the monthly elderly URI cases range from 4,800 to 9,000.

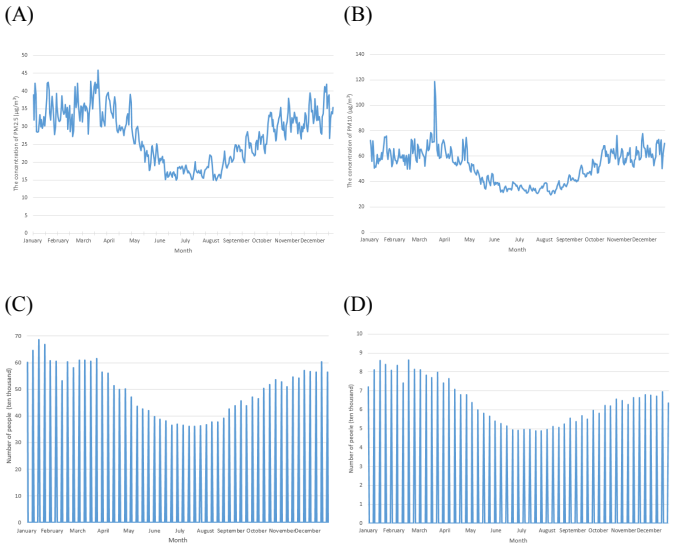

Figure 2.The average daily levels of (A) PM2.5 and (B) PM10 between Dec. 2008 and Dec. 2016 and the average numbers of outpatient visits for URI for (C) the overall and (D) the elderly populations in each month from Jan. 2009 to Dec. 2016.

Table 1 shows the results of accuracy of machine learning of PM2.5 and PM10 levels in predicting the URI burden for the overall and elderly outpatients. PM2.5 and PM10 levels input to the MLP can precisely predict the degree of URI number for the elderly (89.05% and 88.32%, respectively) and the overall population (81.75% and 83.21%, respectively).

Table 1. Accuracy of machine learning of PM2.5 and PM10 levels to predict the URI burden in the overall and elderly outpatients.

Level	Overall Population	Elderly Population
PM2.5	81.75%	89.05%
PM10	83.21%	88.32%

IV. DISCUSSION AND STUDY LIMITATIONS

In previous studies, the hazardous effect of high level of PM2.5 and PM10 exposures on the URI occurrence had been observed [1, 2]. Their methods often applied case-crossover design using the case PM data on the event day to compare with other control PM data on prospective and retrospective days to see the odds of URI risk. To our knowledge, the effect PM has on respiratory tract may be synergic and using one-day PM level to estimate the URI risk have potential bias; our study used a novel procedure of MLP of successive 30-days PM levels to predict URI.

We found the levels of PM2.5 and PM10 peak in winter and spring, which is probably related to several meteorological parameters such as gravity, outdoor temperature, humidity, wind speed, and rain [6]. The URI occurrence may be associated with factors with which the pathogens can grow rapidly and predispose robust individuals to illness. Higher PM levels coincided with the pathogens active seasons could contribute to the high prevalence of URI. In addition, we also showed that the prediction had the best result for the elderly outpatients. This could be explained in part by the fact that the elderly who had many comorbidities were more likely to have acute illness if exposing to multiple air pollutions.

Although the MLP is a well-known machine learning method, we still have a few limitations. First, we used only PM data in this study, and we may need more air pollution data such as SO_2 and other meteorological parameters to make adjustments. Second, the detailed baseline data of the URI patients were lacking, which may result in bias.

V.CONCLUSION

Machine learning of monthly PM2.5 and PM10 levels can accurately predict the weekly burden of overall and elderly outpatient visits for URI in Taiwan.

VI. REFERENCES

[1]M. A. Bind, A. Baccarelli, A. Zanobetti, L. Tarantini , H. Suh, P. Vokonas, and J. Schwartz, "Air pollution and markers of coagulation, inflammation, and endothelial function: associations and epigene-environment interactions in an elderly cohort," Epidemiology, vol. 23, no. 2, pp.332-340, Mar. 2012.

[2]R. Li, N. Jiang, Q. Liu, J. Huang, X. Guo, F. Liu, and Z. Gao, "Impact of air pollutants on outpatient visits for acute respiratory outcomes," Int. J. Environ. Res. Public Health, vol. 14, no. 1, Jan. 2017.

[3]A. H. Sinclair, E. S. Edgerton, R. Wyzga, and D. Tolsma, "A two-time-period comparison of the effects of ambient air pollution on outpatient visits for acute respiratory illnesses, " J. Air Waste Manag. Assoc., vol. 60, no. 2, pp. 163-175, Feb. 2010.

[4]A. Faustini, M. Stafoggia, P. Colais, G. Berti, L. Bisanti, E. Cadum, A. Cernigliaro, S. Mallone, C. Scarnato, F. Forastiere, and et al, "Air pollution and multiple acute respiratory outcomes, " Eur Respir. J. , vol. 42, pp. 304-313, 2013.

[5]B. Ambale-Venkatesh, X. Yang, C. O. Wu, K. Liu, W. G. Hundley, R. McClelland, A. S. Gomes, A. R. Folsom, S. Shea, E. Guallar, D. A. Bluemke, and J. A. C. Lima, "Cardiovascular event prediction by machine learning: The multi-ethnic study of atherosclerosis, " Circ. Res., vol. 121, no. 9, pp. 1092-1101 , Oct. 2017.

[6]J. Wang and S. Ogawa, "Effects of meteorological conditions on PM2.5 concentrations in Nagasaki, Japan," Int. J. Environ. Res. Public Health, vol. 12, no. 8, pp. 9089-9101, Aug. 2015.