Prognostic Factors for Covid-19 on Admission Profile and Air Pollutants

Kikue SATOa,1, Taiki FURUKAWAa, Satoshi YAMASHITAb, Daisuke KOBAYASHIb, Shintaro OYAMAc and Yoshimune SHIRATORIa

aNagoya University Hospital Medical IT Center, Nagoya University, Nagoya, Japan
bGraduate School Medicine/School of Public Health Department of Health Economics and Quality Management, Kyoto University, Kyoto, Japan
cInnovative Research Center for Preventive Medical Engineering, Nagoya University, Nagoya, Japan

Abstract. It has been reported that the severity and lethality of Covid-19 are associated with coexisting underlying diseases (hypertension, diabetes, etc.) and cardiovascular diseases (coronary artery disease, atrial fibrillation, heart failure, etc.) that increase with age, but environmental exposure such as air pollutants may also be a risk factor for mortality. In this study, we investigated patient characteristics at admission and prognostic factors of air pollutants in Covid-19 patients using a machine learning (random forest) prediction model. Age, Photochemical oxidant concentration one month prior to admission, and level of care required were shown to be highly important for the characteristics, while the cumulative concentrations of air pollutants SPM, NO2, and PM2.5 one year prior to admission were the most important characteristics for patients aged 65 years and older, suggesting the influence of long-term exposure.

Keywords. AirPollutionExposure, Covid-19, MachineLearning, Prognostic Factors

1. Introduction and Methods

The health effects of air pollutants on deaths from cardiovascular and respiratory diseases have become clear in many countries [1], and it has been reported that air pollution and other environmental exposures may be included as risk factors for mortality in Covid-19 as well [2]. If prognosis can be predicted based on environmental exposures at the individual level, based on patient characteristics at the time of admission and residential information, it will contribute to the evaluation of causal relationships between health effects and mortality.

DPC (Diagnosis Procedure Combination) is a patient classification method developed in Japan for inpatients in the acute phase of illness, and the nationwide uniform electronic DPC data includes clinical information on patients, information on medical procedures used for patient classification. Using the DPC data, this study included 4,071 cases of Covid-19 (ICD-10: U07.1 or U07.2) who were hospitalized at medical institutions in Aichi Prefecture from April 2020 to April 2021.

Air pollutants covered were nitrogen dioxide (NO2), photochemical oxidants (Ox), suspended particulate matter (SPM), and fine particulate matter (PM2.5). Air pollution

1 Corresponding Author: Kikue Sato, E-mail: satokiku@med.nagoya-u.ac.jp.
concentrations are calculated using monthly averages from each air pollution monitoring station located in designated areas by local governments, and are compared with monitoring stations located in the vicinity of each patient's residence. Cumulative exposures for the previous month and year are obtained from the date of hospitalization for Covid-19. Air pollution constant monitoring data are available at the National Institute for Environmental Studies (Environmental Observatory website: https://tenbou.nies.go.jp/download/).

The following items were extracted as patient profiles at the time of hospitalization and used as variables.

- Inpatient Facility/Infectious Disease Epidemic: Size of hospital beds, Type of facility (university, public, private), and period of infection spread.
- Patient attributes/ profile: gender, age, height, weight, BMI, smoking index, level of care required, and patient residence (city, ward, town, or village).
- Disease information: diabetes, hypertension, obesity, kidney disease, cardiovascular disease CVD, chronic obstructive pulmonary disease COPD comorbidity
- Air pollutants: Nitrogen dioxide NO2, Photochemical oxidant Ox, suspended particulate matter SPM, and fine particulate matter PM2.5 Concentrations for each substance in the month before the day of admission and cumulative concentrations for one year.

The prediction model was constructed using a machine learning random forest with a binary outcome of [death at discharge] or not. Python 3.9.7 was used as the program.

2. Results and Conclusions

Of the eligible cases, the valid data set consisted of 2,610 cases (126 deaths). The objective variable was death at discharge, which was split with the explanatory variables and down-sampled according to the number of cases for the objective variable. Next, a random forest model with 30 decision trees was created by splitting the data into 80% training data and 20% test data. The model had Accuracy of 0.80 (AUC:0.83) and the top features were age, Photochemical oxidant concentration one month prior to admission, and level of care needed. In addition, although the Accuracy was lower for those aged 65 years and older (Accuracy:0.67, AUC:0.62), and the top feature values were cumulative SPM concentration one year prior to admission, cumulative NO2 concentration one year prior to admission, and cumulative PM2.5 concentration one year prior to admission. In a prognostic model for patients with Covid-19, we were able to show the influence of environmental factors as a feature from the patient profile at admission.

References