Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240720

Environmental Geomarker to Assess Impact on Hospitalization

Kikue SATO [a,1], Taiki FURUKAWA[a], Daisuke KOBAYASHI[b], Shintaro OYAMA[c] and Yoshimune SHIRATORI[a]
[a] *Nagoya University Hospital Medical IT Center, Nagoya University, Nagoya, Japan*
[b] *Toyama University Hospital, Department of Community Medical Support, Toyama, Japan*
[c] *Innovative Research Center for Preventive Medical Engineering, Nagoya University, Japan*
ORCiD ID: Kikue Sato https://orcid.org/0009-0000-7405-9687

Abstract. By linking medical real-world data with geographic information, it is possible to evaluate the impact on hospitalization based on these characteristics, such as patient residence information and disease and medical information. In this study, environmental exposure to air pollutants was reported as a risk factor, and predictive models were used to examine factors affecting health. The importance of the characteristics appeared according to the disease, and overall, the patient profile at the time of admission, such as ADL, was shown to be high, but for respiratory diseases, the cumulative concentration of air pollutants NO_2, SPM, and NOx for one year before the onset of admission was the top risk factor for long-term hospitalization, suggesting the influence of exposure due to environmental factors.

Keywords. Geomarker, Air Pollutants, Machine Learning, Predictive Model

1. Introduction

Geomarkers have been defined as “any objective, contextual, or geographic measure that influences or predicts the incidence of outcome or disease.” [1] By linking medical real-world data and geographic information with geographically distributed environmental factors, such as patient residence information and disease and practice information, large-scale epidemiological studies can be conducted. [2] If prognostic predictions can be made based on environmental exposure status, this will contribute to the evaluation of causal relationships between health effects on hospitalization.

2. Methods

This study included cases of "cerebrovascular disease", "cardiovascular disease", " high-blood pressure", "respiratory disease" and "Covid-19," which were given as the most medical resource-intensive diseases in patients admitted to medical institutions in Aichi

[1] Corresponding Author: Kikue Sato, Nagoya University Hospital Medical IT Center, 65 Tsurumai-cho, Showa-ku, Nagoya 466-8560, Japan; E-mail: satokiku@nagoya-u.jp.

Prefecture from April 2020 to March 2022, based on large-scale Administrative Claims Data (DPC data).

Target air pollutants were nitric oxide (NO), nitrogen dioxide (NO2), nitrogen oxides (NOx), photochemical oxidants (OX), suspended particulate matter (SPM), and fine particulate matter (PM2.5). Air pollution concentrations were obtained from air pollution continuous monitoring data from measuring stations located in designated areas by local governments. There are 114 measuring stations in Aichi Prefecture. Nearest neighbor locations were searched by linear distance between the residence and the measuring station, and the cumulative exposure to each air pollutant for one year backward from the starting date of hospitalization was tied to the subject.

A prognostic model based on patient profiles at admission was built using machine learning (random forest). Python 3.9.7 was used for the program.

3. Results, Discussion and Conclusions

There were 126,243 eligible hospitalized patients (8,119 deaths) that could be linked to environmental data, of which 56,965 (2,673 deaths) were valid data sets after excluding unknown codes for each variable and other factors. The predictive model included "death at discharge" and "hospitalization longer than 30 days" as outcomes, split with the explanatory variables, and down-sampled according to the number of cases for the objective variable. Next, a random forest model with decision trees was created by splitting the data into 80% training data and 20% test data.

In terms of risk of death, cardiovascular disease showed good model accuracy, with a correct response rate of 0.82 (AUC: 0.88), and the top features were ambulance transport, ADL at admission, and age, in that order, with the profile at admission being the top feature. For Covid-19, the model had Accuracy of 0.77 (AUC: 0.82), for cerebrovascular disease, the model had Accuracy of 0.72 (AUC: 0.78), and the top features were ADL at admission, ambulance transport, and cumulative SPM concentration year 1 year before admission for both patients. For long-term hospitalization, respiratory disease had a model Accuracy of 0.90 (AUC: 0.71), and the top characteristics were cumulative NO2 concentration for 1 year before hospitalization, cumulative SPM concentration for 1 year before hospitalization, cumulative NOx concentration for 1 year before hospitalization and air pollutants, in order from highest to lowest. Exposure to air pollutants was associated with increased risk of mortality and prolonged hospitalization.

We were able to show the influence of environmental factors as a feature of the patient profile using a disease prognostic model. The use of medical real-world data with geo-markers demonstrates the potential for gaining insight into regional disease patterns.

References

[1] Brokamp C, et al. Decentralized and reproducible geocoding and characterization of community and environmental exposures for multisite studies. J Am Med Inform Assoc. 2018 Mar 1;25(3):309-314.

[2] Erika RM, et al. A Multi-Modal Geomarker Pipeline for Assessing the Impact of Social, Economic, and Environmental Factors on Pediatric Hospitalization. J Am Med Inform Assoc. 2024 May.