



OPEN

Correlation between air pollution and prevalence of conjunctivitis in South Korea using analysis of public big data

Sanghyu Nam^{1,4}, Mi Young Shin^{2,3,4}, Jung Yeob Han¹, Su Young Moon¹, Jae Yong Kim¹, Hungwon Tchah¹ & Hun Lee¹✉

This study investigated how changes in weather factors affect the prevalence of conjunctivitis using public big data in South Korea. A total of 1,428 public big data entries from January 2013 to December 2019 were collected. Disease data and basic climate/air pollutant concentration records were collected from nationally provided big data. Meteorological factors affecting eye diseases were identified using multiple linear regression and machine learning analysis methods such as extreme gradient boosting (XGBoost), decision tree, and random forest. The prediction model with the best performance was XGBoost (1.180), followed by multiple regression (1.195), random forest (1.206), and decision tree (1.544) when using root mean square error (RMSE) values. With the XGBoost model, province was the most important variable (0.352), followed by month (0.289) and carbon monoxide exposure (0.133). Other air pollutants including sulfur dioxide, PM₁₀, nitrogen dioxides, and ozone showed low associations with conjunctivitis. We identified factors associated with conjunctivitis using traditional multiple regression analysis and machine learning techniques. Regional factors were important for the prevalence of conjunctivitis as well as the atmosphere and air quality factors.

Conjunctivitis is as commonly presenting disease at ophthalmology clinics, caused mainly by viral infections, allergic reactions, or atopy. Environmental factors have also been implicated in incidences of conjunctivitis¹. Consistent contact with the ocular surface in eyes allows toxins to directly access ocular structures and cause conjunctivitis-like symptoms². Additionally, the effects of environmental pollution on human health can vary depending on the composition of and degree and time of exposure to air pollutants^{3,4}.

Previous studies have focused on evaluating the association between air pollution and health problems related to respiratory organs and cardiovascular vessels^{5,6}. However, air quality can affect not only the respiratory and cardiovascular systems, but also the ocular surface of the eye, with which air comes into direct contact. Air pollutants such as ozone, nitrogen dioxide, and sulfur dioxide have been associated with conjunctivitis⁷. Furthermore, one study found relationships between the levels of particulate matter with aerodynamic diameter < 10 μm (PM₁₀) and emergency room visits for keratoconjunctivitis, ischemic heart disease, and stroke in Korea⁸. For medical and health care for conjunctivitis, studies on the prevalence of ocular surface disease like keratoconjunctivitis need to be conducted to report the relationship between the prevalence of conjunctivitis and air and atmosphere quality, as well as population factors such as region, number of people, age, and gender. However, there has been no nationwide study evaluating the relationship between various air pollutants and conjunctivitis.

Therefore, in the present study, we investigated how changes in weather and population factors can affect the prevalence of conjunctivitis using public big data provided by various Korean governmental institutions. Furthermore, we determined whether air pollution increases the risk of conjunctivitis by using machine learning prevalence prediction models.

¹Department of Ophthalmology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea. ²Graduate School of Public Health, Catholic University of Korea, Seoul, Korea. ³Seoul Metropolitan Office of Education, Seoul, Korea. ⁴These authors contributed equally: Sanghyu Nam and Mi Young Shin. ✉email: yhun777@gmail.com

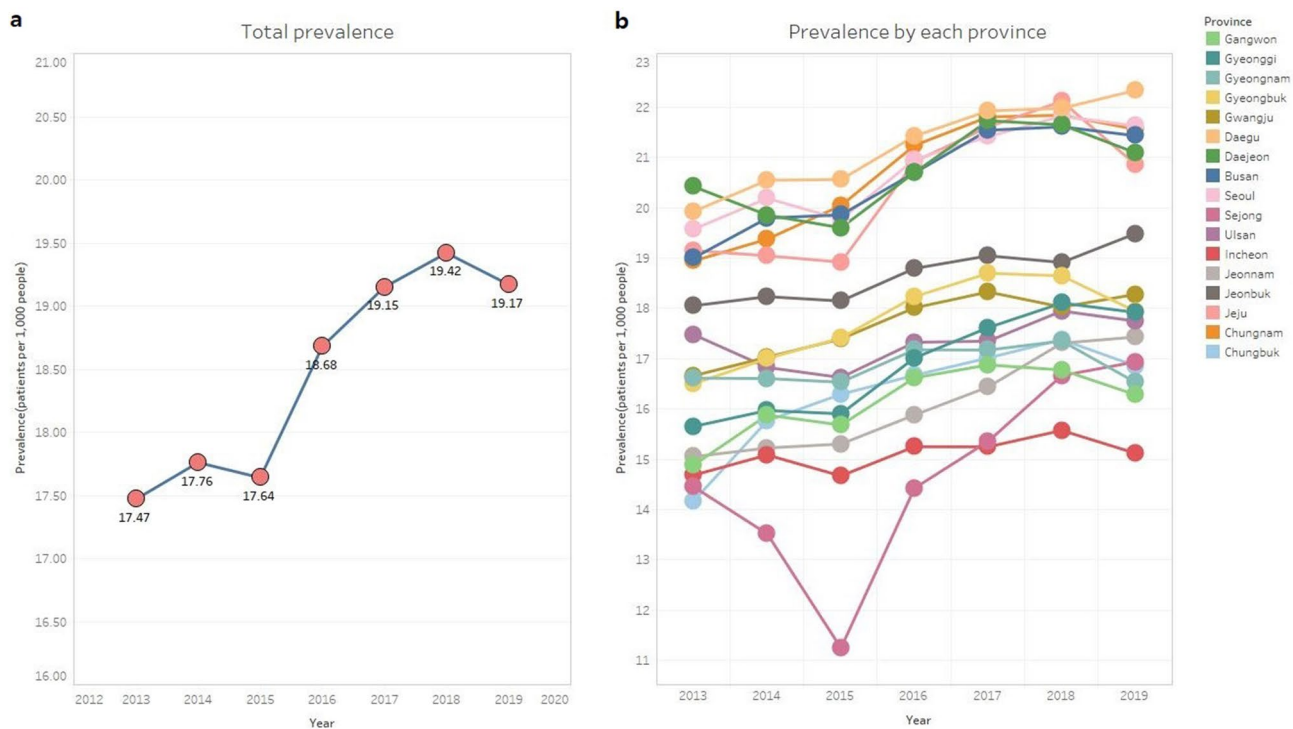


Figure 1. Prevalence of conjunctivitis. **(a)** Prevalence of conjunctivitis by year (number of patients per 1,000 people). The number of patients increased from 2013 to 2019. **(b)** Prevalence by each province.

Results

According to the annual prevalence trends, 19.17 patients per 1,000 people were diagnosed with conjunctivitis in 2019 compared to 17.47 patients per 1,000 people in 2013. The number of patients per year tended to increase from 2013 to 2019, with a mild decrease in 2015 and 2019 (Fig. 1a). Prevalence by each province also showed a steady upward curve. In some regions, the prevalence decreased in 2015, but increased again from 2016 (Fig. 1b).

Figure 2 shows the prevalence of conjunctivitis and weather parameters by month in each region of Korea. The monthly prevalence of conjunctivitis peaked in May and September in all provinces (Fig. 2a). The prevalence tended to increase as winter changed to summer with peaks between seasons. The mean temperature was highest in July and August and lowest in January and February (Fig. 2b). All regions showed similar trends in the mean temperature. The daily temperature difference was highest in spring and fall, but some provinces, including Jeju Island, Busan City, and Incheon City, showed smaller temperature differences because they were coastal areas (Fig. 2c). The average wind speed did not show much change by month; only in the winter season in Jeju Island, relatively high wind speeds were observed compared to those in the other provinces (Fig. 2d).

Figure 3 shows monthly air quality data by region. In all provinces, PM_{10} levels remained high from winter to spring, decreased starting in May with the lowest levels in August, and increased again to high levels from September to spring (Fig. 3a). Other air quality variables including concentrations of nitrogen dioxide, carbon monoxide, and sulfur dioxide, showed low levels in summer and high levels in winter (Fig. 3b–d). Levels of sulfur dioxide were especially unique in Ulsan city and highest in summer (Fig. 3d). This result may be because Ulsan city is heavily industrialized. The concentration of ozone was highest in spring and decreased from summer through winter in all provinces (Fig. 3e).

Pearson's correlation coefficient analysis was performed to evaluate the relationship between the prevalence of conjunctivitis and independent variables (Table 1). The results showed positive correlations with average temperature, humidity, precipitation, and ozone concentrations; negative correlations were described for daily temperature differences, average wind speeds, and concentrations of sulfur dioxide, nitrogen dioxide, carbon monoxide, and PM_{10} .

In the multiple regression analysis, the coefficient of determination was 0.8789. Based on the high predictive power of the multiple regression analysis, we assessed the best performance prevalence prediction model among machine learning techniques including extreme gradient boosting (XGBoost), decision trees, and random forest. The outcome performances of each model were compared using root mean square error (RMSE), and the training and test set ratio was 9:1. As a result, model performance was shown in the order of XGBoost (1.180), multiple regression (1.195), random forest (1.206), and decision tree (1.544) (Table 2).

According to the scatterplots showing the difference between real and predictive values in machine learning predictions, XGBoost's predictions were best suited to real values. The decision trees had the lowest fit among the other models, similar to that in previous studies (Fig. 4)⁹. Based on the results from XGBoost prediction model, which had the best predictive power, the most important variables were province (gain value: 0.352), month (0.289), and carbon monoxide level (0.133; Table 3).

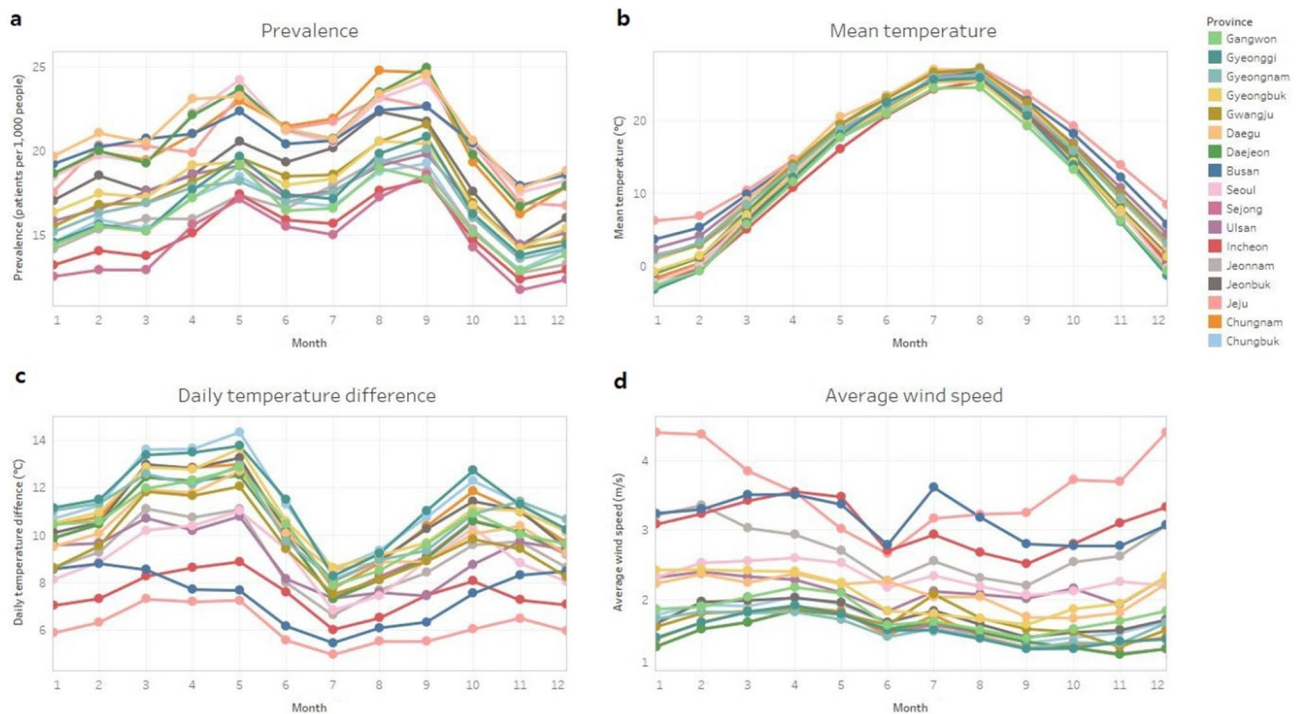


Figure 2. Prevalence of conjunctivitis and weather parameters by month in each region. **(a)** Prevalence of conjunctivitis, **(b)** mean temperature, **(c)** mean daily temperature difference, and **(d)** mean wind speed.

Discussion

In the present study, based on countrywide public big data, we evaluated the effects of weather and air quality variables on the prevalence of conjunctivitis and compared the performance of predictive modeling. There have been previous studies on correlations between air pollution and various diseases, such as keratoconjunctivitis, ischemic heart disease, stroke, and respiratory diseases^{8,10}. Although there are various datasets relating to eye diseases, it is well known that the ocular surface, including the cornea, is always exposed to the air, and subsequently, symptoms of conjunctivitis and air pollutants are always associated¹¹. Therefore, in this study, we selected ocular surface diseases, such as keratoconjunctivitis, conjunctivitis, and blepharoconjunctivitis, to analyze their association with environmental factors.

The prevalence of conjunctivitis showed an increasing trend from 2013 to 2019. Based on a monthly analysis, the prevalence was the highest during spring and fall with two peaks in May and September and was the lowest in winter. This finding aligns with those of a previous study¹², during which the prevalence of allergic conjunctivitis increased from spring to fall in accordance with other increased allergen levels such as those of dust and pollen.

Among the predictive models, the XGBoost model showed the best performance, followed by multiple regression analysis, random forest, and decision tree modeling. The most important variable according to the XGBoost model was province, followed by month and carbon monoxide level. Notably, region was not estimated as an effective factor in a previous study conducted in Korea⁸. This difference may be attributed to different climatic factors, air quality factors, and medical systems in each province. We believe that further research on regional prediction models is necessary.

The second most important factor was the month of the year. As previously mentioned, prevalence differed from month to month with higher rates during the spring and fall. It is notable that the monthly impact was greater than the impact of other climatic or air quality factors. These climatic and air quality factors are comprehensively reflected in each monthly period. Therefore, considering the month as a sole factor, it may be most important compared to other climatic and air quality factors because it can predict the characteristics of the climate and air quality itself.

Among air pollutants, carbon monoxide was most highly associated with the prevalence of conjunctivitis (0.133) when compared to the associations of sulfur dioxide (0.016), PM₁₀ (0.017), nitrogen dioxide (0.013), and ozone (0.019). A few previous studies have shown that carbon monoxide has minor effects on the prevalence of conjunctivitis. One report showed an association between carbon monoxide levels and emergency room visits for asthma¹⁰, and another reported a positive association between carbon monoxide levels and the prevalence of conjunctivitis¹³. In contrast, Chang et al. reported that carbon monoxide had only a non-significant influence on nonspecific conjunctivitis cases in outpatient visits, due to the absence of ocular irritation as a consequence of carbon monoxide exposure¹⁴. According to our study, conjunctivitis and carbon monoxide were negatively correlated, and to our knowledge, it is the only study that has shown negative correlation results. We believe that increases in carbon monoxide levels are closely related to increased use of fuels for heating during cold seasons. The concentrations of carbon monoxide decrease during the summer and increase in the winter. Our results showed that concentrations of carbon monoxide remain low from April to September and then increase from

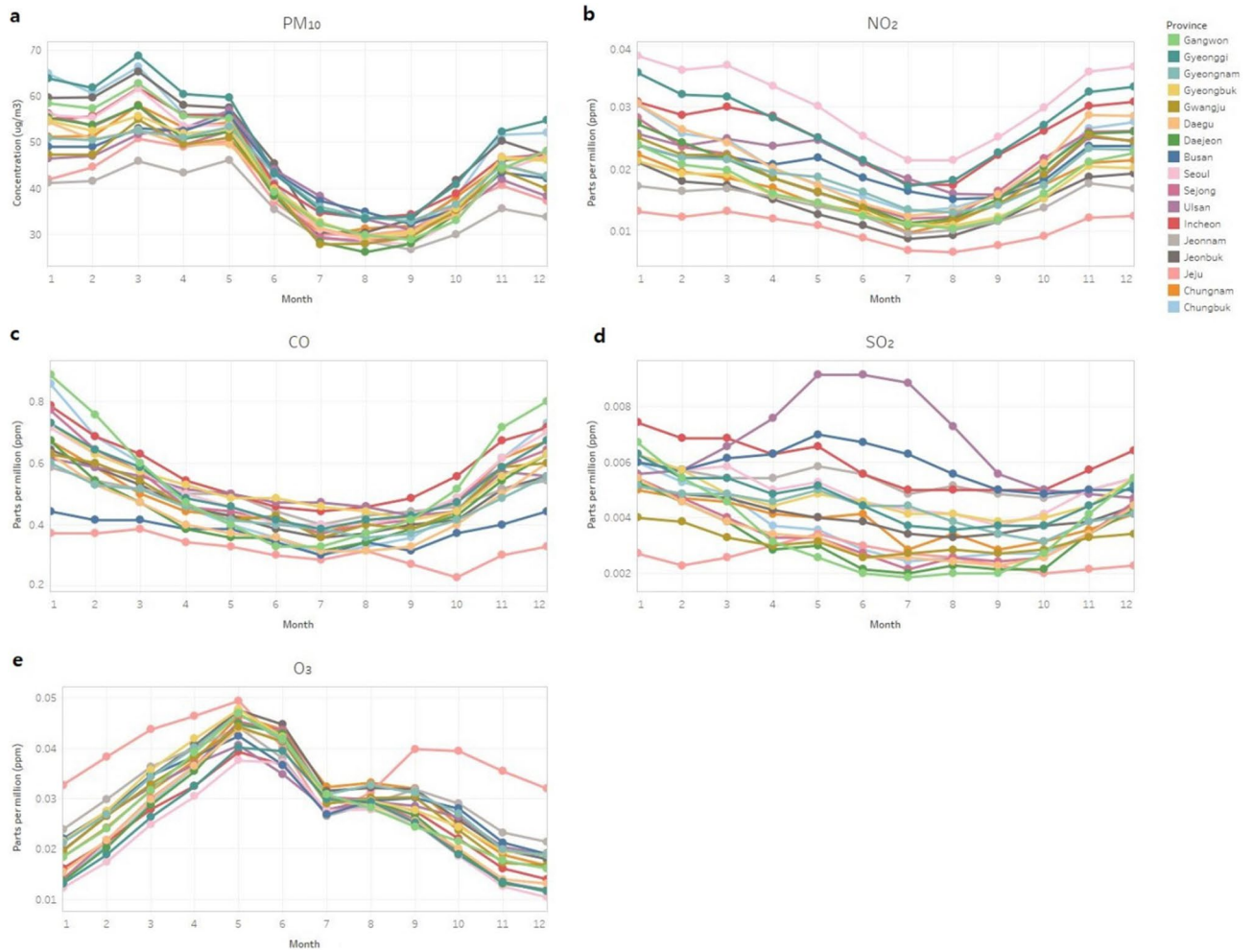


Figure 3. Air quality parameters by month in each region. (a) concentration of sulfur dioxide, (b) concentration of nitrogen dioxide, (c) concentration of carbon monoxide, (d) concentration of PM_{10} , and (e) concentration of ozone.

	1	2	3	4	5	6	7	8	9	10	11
1. Regional Prevalence		0.52	-0.11	0.13	0.29	0.04	-0.28	-0.37	0.43	-0.56	-0.26
2. Temperature	0.52		-0.26	0.61	0.61	-0.07	-0.33	-0.64	0.54	-0.74	-0.58
3. Daily temperature difference	-0.11	-0.26		-0.50	-0.43	-0.45	-0.01	0.22	0.16	0.23	0.48
4. Humidity	0.13	0.61	-0.50		0.59	-0.16	-0.27	-0.55	0.05	-0.39	-0.58
5. Precipitation	0.29	0.61	-0.43	0.59		0.01	-0.25	-0.49	0.21	-0.48	-0.49
6. Wind speed	0.04	-0.07	-0.45	-0.16	0.01		0.25	-0.04	0.21	-0.13	0.07
7. SO_2	-0.28	-0.33	-0.01	-0.27	-0.25	0.25		0.51	-0.18	0.50	0.43
8. NO_2	-0.37	-0.64	0.22	-0.55	-0.49	-0.04	0.51		-0.55	0.67	0.55
9. O_3	0.43	0.54	0.16	0.05	0.21	0.21	-0.18	-0.55		-0.55	0.00
10. CO	-0.56	-0.74	0.23	-0.39	-0.48	-0.13	0.50	0.67	-0.55		0.53
11. PM_{10}	-0.26	-0.58	0.48	-0.58	-0.49	0.07	0.43	0.55	0.00	0.53	

Table 1. Correlation coefficients of variables using correlation analysis between prevalence and temperature or air quality parameters. Average temperature, humidity, precipitation, and ozone showed positive correlation, daily temperature difference, average wind speed, sulfur dioxide, nitrogen dioxide, carbon monoxide and PM_{10} showed negative correlation. SO_2 , sulfur dioxide; NO_2 , nitrogen dioxide; O_3 , Ozone; CO, carbon monoxide; PM_{10} , levels of particulate matter with aerodynamic diameter < 10 μm . Significant values are in bold.

Model	RMSE
Multiple linear regression	1.195
XGBoost	1.180
Random forest	1.206
Decision tree	1.544

Table 2. Comparison of modeling techniques on root mean square error values. RMSE, root mean square error.

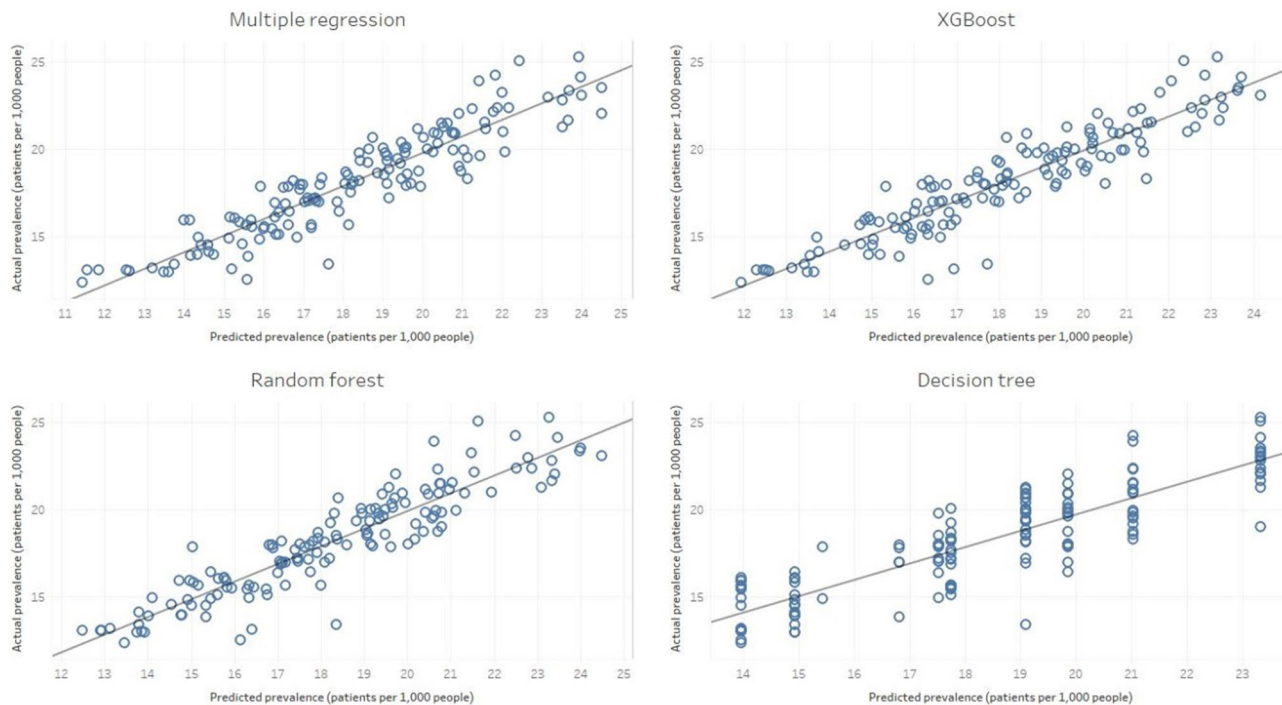


Figure 4. Predicted prevalence to actual prevalence for each model. The XGBoost model shows the most accurate prediction model and the decision tree model shows the least accurate prediction.

Variable	Gain
Province (local)	0.352
Month	0.289
CO	0.133
Temperature	0.060
Humidity	0.030
Wind speed	0.030
Precipitation	0.022
Temperature fluctuation	0.021
O ₃	0.019
PM ₁₀	0.017
SO ₂	0.016
NO ₂	0.013

Table 3. Variables of importance in the XGBoost prediction model. CO, carbon monoxide; O₃, Ozone; PM₁₀, levels of particulate matter with aerodynamic diameter < 10 μm; SO₂, sulfur dioxide; NO₂, nitrogen dioxide.

October to March. The prevalence of conjunctivitis begins to increase in April, peaks in May and September, and decreases from October to March. This change is thought to be the result of similarity in monthly trends rather than a direct association between carbon monoxide and conjunctivitis.

PM₁₀ is a complex component comprised of metal compounds such as nickel, aluminum, silicon, and titanium dioxide, which are correlated with ocular symptoms¹⁵. Lu et al. reported that PM₁₀ is associated with conjunctivitis¹⁶, but another study found no association between the two¹⁴. Automobile exhaust is the main source of atmospheric sulfur dioxide and nitrogen dioxide¹⁷. One Brazilian study found a clear dose–response relationship between the nitrogen dioxide level and goblet cell hyperplasia, suggesting morphological changes in the conjunctival epithelium as an adaptive response to chronic environmental injury¹⁸. Sulfur dioxide was significantly associated with conjunctivitis during outpatient hospital and emergency room visits^{13,19}.

Ozone is an important factor in “summer smog,” generated at ground level by photochemical reactions involving ultraviolet radiation within the atmospheric mixture of nitrogen oxide and hydrocarbons derived from vehicular emissions. Atmospheric concentrations of ozone and nitrogen oxide have been linked to asthma and other airway inflammatory diseases^{20,21}. Ozone can induce an inflammatory response in the ocular surfaces in mouse models and in cultured human conjunctival epithelial cells²². Moreover, exposure to ozone exacerbates the detrimental effects on the integrity of the ocular surface, caused by conjunctival allergic reactions and further increases the inflammatory response²³.

The results of correlations between conjunctivitis and air pollutants are inconsistent. Fu et al.¹³ revealed a significant risk of nitrogen dioxide for the prevalence of conjunctivitis, while Jamaludin et al.²⁴ did not. With regard to PM₁₀, Chang et al.¹⁴ revealed PM₁₀ to be significantly associated with conjunctivitis risk. However, in a different study conducted by Chiang et al.⁷, nitrogen dioxide had no significant effect on the risk of conjunctivitis. Fu et al.¹³ revealed that the correlation between sulfur dioxide and conjunctivitis risk was significant. Previous meta-analyses of five air pollutants (PM₁₀, sulfur dioxide, carbon monoxide, nitrogen dioxide, and ozone) showed a positive correlation between these pollutants and conjunctivitis²⁵. We propose that the contradictory results may be attributable to the study design. Our results are different from those of previous studies, with carbon monoxide being negatively correlated with conjunctivitis. This finding is believed to be due to the slight difference in the analyses methods and origins of data relating to climatic factors and air quality.

In this study, administrative district demographics, weather data, air quality data, and disease data were collected; research was conducted after pre-processing data for effective use and statistical analysis. Machine learning techniques allow users to form guidelines and create new insights using public data. Although ecological analysis has limitations in application to individuals, this study allowed us to obtain individual diagnostic data and variables for subsequent research into weather factors and predictive models for eye disease.

Our study had some limitations. First, the information regarding actual clinical examinations was unavailable in the claims data. Biological factors other than ambient air quality that can cause eye diseases were also undetermined. The International Standard Disease Classification (ICD-10) diagnoses may not be precise enough to reflect the true etiology of conjunctival disease. Additionally, this study used second-hand data to evaluate associations between environmental exposures and diseases; we assumed that the participants were exposed to the same levels of air pollutants as reflected in the measurements of their residential regions. Thus, it is possible that the risk was underestimated²⁶.

In conclusion, we demonstrated associations between weather factors and the prevalence of conjunctivitis via large-scale analyses of nationally provided big data. Traditional multiple regression analysis and machine learning techniques were used to identify the best prediction model. With the best prediction performance by the XGBoost model, region (province), month, and carbon monoxide concentration were found to be the important variables contributing to the prevalence of conjunctivitis. It is meaningful that the association of carbon monoxide among air pollutants was high, and it is also important that regional and monthly factors were related to conjunctivitis along with air quality factors. Consideration of these variables would be helpful for detection and management of conjunctivitis in the clinical field.

Methods

Study object and data source. This study used information from health insurance claims obtained by the Korean Statistical Information Service (KOSIS) and daily meteorological records from the Korea Meteorological Administration (KMA) and the Korea Environment Corporation (Air Korea). The KOSIS provided data from 17 provinces including data on population by province. Basic climate data from the KMA included monthly 24-h weather data regarding average temperature, highest and lowest temperatures, relative humidity, rainfall, and wind speeds. Air Korea provided climate data including concentrations of PM₁₀, nitrogen dioxide, sulfur dioxide, carbon monoxide, and ozone. The ambient PM₁₀ concentration was measured by total 600 air quality monitoring networks, urban air monitoring networks (495), national background concentration networks (11), suburban air monitoring networks (27), road-side air monitoring networks (52), and port air monitoring networks (15). All subjects were assumed to be exposed to the same levels of air pollutants as measured by permanent weather monitoring. The National Ambient Air Quality Standards of South Korea provided by the National Institute of Environmental Research are added in Supplementary Table 1.

Categories of eye diseases were defined using the ICD-10 and collected using the Health Insurance Review and Assessment Service (HIRA)^{27,28}. Disease categories were allergic conjunctivitis, acute conjunctivitis, chronic conjunctivitis, lacrimal gland disorders, blepharconjunctivitis, keratoconjunctivitis and other nonspecific conjunctivitis. Cases of infectious conjunctivitis from pathogens such as adenovirus, herpes virus, meningococcus, gonococcus, acanthamoeba, and trachoma and other bacterial conjunctivitis were excluded. The number of patients diagnosed with the disease was counted and converted to the regional prevalence using local population counts, which was set as a dependent variable.

Variable	Description	Data Source
Province*	17 provinces (Si, Do)	KOSIS
Population	Population by province	KOSIS
Temperature (°C)	Mean temperature by province, round the number to 2 places	KMA
Highest temperature (°C)	Mean highest temperature by province, round the number to 2 places	KMA
Lowest temperature (°C)	Mean lowest temperature by province, round the number to 2 places	KMA
Temperature difference (°C)	Mean daily temperature difference by province, round the number to 2 places	KMA
Humidity (%)	Mean relative humidity by province, round the number to 2 places	KMA
Precipitation (mm)	Monthly total precipitation by province, round the number to 2 places	KMA
Wind speed (m/s)	Mean wind speed by province, round the number to 2 places	KMA
SO ₂ (ppm)	Concentration of SO ₂ by province, round the number to 4 places	Air Korea
NO ₂ (ppm)	Concentration of NO ₂ by province, round the number to 4 places	Air Korea
O ₃ (ppm)	Concentration of O ₃ by province, round the number to 4 places	Air Korea
CO (ppm)	Concentration of CO by province, round the number to 4 places	Air Korea
PM ₁₀ (µg/m ³)	Concentration of PM ₁₀ by province	Air Korea

Table 4. Basic variables from government-provided big data. SO₂, sulfur dioxide; NO₂, nitrogen dioxide; O₃, Ozone; CO, carbon monoxide; PM₁₀, levels of particulate matter with aerodynamic diameter < 10 µm; KOSIS, Korean Statistical Information Service; KMA, Korea Meteorological Administration.

Administrative district demographics, meteorological data, air quality data, and disease data were collected from January 2013 to December 2019. All variables analyzed are presented in Table 4, and data pre-processing was conducted for effective data use and statistical analysis. The Institutional Review Board of Asan Medical Center (University of Ulsan College of Medicine) instead of approved the waiver of reviewing this study (2021-0173). This study was conducted according to the ethical principles outlined in the Declaration of Helsinki. The requirement for obtaining informed consent was waived.

Statistical analysis and machine learning analysis. Many fields utilize machine learning²⁹, and active research is underway in the health sector to utilize machine learning to analyze cancer survival³⁰ and predict emergency room admission³¹. Furthermore, medical big data have been used to develop personalized medicine for dry eye disease³². In our study, conjunctivitis prevalence was set as a dependent variable, and meteorological, air quality, and demographic factors were independent variables. By analyzing prevalence patterns, influencing factors were identified and predictive modeling performed. In this process, exploratory data analysis (EDA) on each variable was conducted to examine each characteristic and identify its impact on prevalence. Finally, the relationship between prevalence and each variable was identified using traditional analysis methods, such as multiple linear regression analysis and machine learning analysis. Machine learning analyses included XGBoost, decision tree, and random forest methods. The total numbers of data sets for analysis were 1428. The machine learning analysis model was maintained at a 90% training set (number of set = 1288) and 10% test set (number of set = 140). The performance of each model was evaluated using RMSE values. The statistical analysis incorporated regression analysis to define correlation factors between independent variables. All statistical analyses were performed using R software (version 3.6.1). Statistical significance was defined as $P < 0.05$.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 30 January 2022; Accepted: 23 May 2022

Published online: 16 June 2022

References

- Miyazaki, D. *et al.* Epidemiological aspects of allergic conjunctivitis. *Allergol. Int.* **69**, 487–495. <https://doi.org/10.1016/j.alit.2020.06.004> (2020).
- Johnson, G. The environment and the eye. *Eye* **18**, 1235–1250. <https://doi.org/10.1038/sj.eye.6701369> (2004).
- Cohen, A. J. *et al.* The global burden of disease due to outdoor air pollution. *J. Toxicol. Environ. Health A* **68**, 1301–1307. <https://doi.org/10.1080/15287390590936166> (2005).
- Sharma, R. K. & Agrawal, M. Biological effects of heavy metals: an overview. *J. Environ. Biol.* **26**, 301–313 (2005).
- Huang, Y.-C.T. & Ghio, A. J. Vascular effects of ambient pollutant particles and metals. *Curr. Vasc. Pharmacol.* **4**, 199–203. <https://doi.org/10.2174/15701610677698351> (2006).
- Pope Iii, C. A. *et al.* Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA* **287**, 1132–1141. <https://doi.org/10.1001/jama.287.9.1132> (2002).
- Chiang, C.-C., Liao, C.-C., Chen, P.-C., Tsai, Y.-Y. & Wang, Y.-C. Population study on chronic and acute conjunctivitis associated with ambient environment in urban and rural areas. *J. Exposure Sci. Environ. Epidemiol.* **22**, 533–538. <https://doi.org/10.1038/jes.2012.37> (2012).
- Lee, J. Relationships between PM10 level and emergency room visits for keratoconjunctivitis, ischemic heart disease and strokes. Unpublished master's thesis, Keimyung University (2017).

9. Pesantez-Narvaez, J., Guillen, M. & Alcañiz, M. Predicting motor insurance claims using telematics data—XGBoost versus logistic regression. *Risks* **7**, 70. <https://doi.org/10.3390/risks7020070> (2019).
10. Akinbami, L. J., Lynch, C. D., Parker, J. D. & Woodruff, T. J. The association between childhood asthma prevalence and monitored air pollutants in metropolitan areas, United States, 2001–2004. *Environ. Res.* **110**, 294–301. <https://doi.org/10.1016/j.envres.2010.01.001> (2010).
11. Norn, M. Pollution keratoconjunctivitis: a review. *Acta Ophthalmol.* **70**, 269–273. <https://doi.org/10.1111/j.1755-3768.1992.tb04136.x> (1992).
12. Leonardi, A., Castegnaro, A., Valerio, A. L. G. & Lazzarini, D. Epidemiology of allergic conjunctivitis: clinical appearance and treatment patterns in a population-based study. *Curr. Opin. Allergy Clin. Immunol.* **15**, 482–488. <https://doi.org/10.1097/ACI.000000000000204> (2015).
13. Fu, Q. *et al.* Air pollution and outpatient visits for conjunctivitis: A case-crossover study in Hangzhou, China. *Environ. Pollut.* **231**, 1344–1350. <https://doi.org/10.1016/j.envpol.2017.08.109> (2017).
14. Chang, C.-J., Yang, H.-H., Chang, C.-A. & Tsai, H.-Y. Relationship between air pollution and outpatient visits for nonspecific conjunctivitis. *Investig. Ophthalmol. Vis. Sci.* **53**, 429–433. <https://doi.org/10.1167/iovs.11-8253> (2012).
15. Patil, N. A., Gade, W. & Deobagkar, D. D. Epigenetic modulation upon exposure of lung fibroblasts to TiO₂ and ZnO nanoparticles: alterations in DNA methylation. *Int. J. Nanomed.* **11**, 4509. <https://doi.org/10.2147/IJN.S110390> (2016).
16. Lu, P. *et al.* Short-term exposure to air pollution and conjunctivitis outpatient visits: A multi-city study in China. *Environ. Pollut.* **254**, 113030. <https://doi.org/10.1016/j.envpol.2019.113030> (2019).
17. D'amato, G., Liccardi, G., D'amato, M. & Holgate, S. Environmental risk factors and allergic bronchial asthma. *Clin. Exp. Allergy* **35**, 1113–1124. <https://doi.org/10.1111/j.1365-2222.2005.02328.x> (2005).
18. Novaes, P. *et al.* Ambient levels of air pollution induce goblet-cell hyperplasia in human conjunctival epithelium. *Environ. Health Perspect.* **115**, 1753–1756. <https://doi.org/10.1289/ehp.10363> (2007).
19. Szyzkowicz, M., Kousha, T. & Castner, J. Air pollution and emergency department visits for conjunctivitis: a case-crossover study. *Int. J. Occup. Med. Environ. Health* **29**, 381–393. <https://doi.org/10.13075/ijom.1896.00442> (2016).
20. Atkinson, R. *et al.* Short-term associations between outdoor air pollution and visits to accident and emergency departments in London for respiratory complaints. *Eur. Respir. J.* **13**, 257–265. <https://doi.org/10.1183/09031936.99.13225799> (1999).
21. Peters, J. M. *et al.* A study of twelve Southern California communities with differing levels and types of air pollution: I. Prevalence of respiratory morbidity. *Am. J. Respir. Crit. Care Med.* **159**, 760–767. <https://doi.org/10.1164/ajrccm.159.3.9804143> (1999).
22. Lee, H. *et al.* Effects of ozone exposure on the ocular surface. *Free Radic. Biol. Med.* **63**, 78–89. <https://doi.org/10.1016/j.freeradbiomed.2013.05.006> (2013).
23. Lee, H., Kim, E. K., Kim, H. Y. & Kim, T. I. Effects of exposure to ozone on the ocular surface in an experimental model of allergic conjunctivitis. *PLoS ONE* **12**, e0169209. <https://doi.org/10.1371/journal.pone.0169209> (2017).
24. Jamaludin, A. R. B. *et al.* Correlational study of air pollution-related diseases (asthma, conjunctivitis, URTI and dengue) in Johor Bahru, Malaysia. *Malays. J. Fundam. Appl. Sci.* **13**, 354–361 (2017).
25. Chen, R. *et al.* Global associations of air pollution and conjunctivitis diseases: a systematic review and meta-analysis. *Int. J. Environ. Res. Public Health* **16**, 3652. <https://doi.org/10.3390/ijerph16193652> (2019).
26. Armstrong, B. G. Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup. Environ. Med.* **55**, 651–656. <https://doi.org/10.1136/oem.55.10.651> (1998).
27. Malerbi, F. K., Martins, L. C., Saldiva, P. H. N. & Braga, A. L. F. Ambient levels of air pollution induce clinical worsening of blepharitis. *Environ. Res.* **112**, 199–203. <https://doi.org/10.1016/j.envres.2011.11.010> (2012).
28. Galor, A., Kumar, N., Feuer, W. & Lee, D. J. Environmental factors affect the risk of dry eye syndrome in a United States veteran population. *Ophthalmology* **121**, 972–973. e971, doi:<https://doi.org/10.1016/j.ophtha.2013.11.036> (2014).
29. Hah, D. W., Kim, Y. M. & Ahn, J. J. A study on KOSPI 200 direction forecasting using XGBoost model. *Korean Data Inform. Sci. Soc.* **30**, 655–669 (2019).
30. Lynch, C. M. *et al.* Prediction of lung cancer patient survival via supervised machine learning classification techniques. *Int. J. Med. Informatics* **108**, 1–8. <https://doi.org/10.1016/j.ijmedinf.2017.09.013> (2017).
31. Hong, W. S., Haimovich, A. D. & Taylor, R. A. Predicting hospital admission at emergency department triage using machine learning. *PLoS ONE* **13**, e0201016. <https://doi.org/10.1371/journal.pone.0201016> (2018).
32. Inomata, T. *et al.* Using medical big data to develop personalized medicine for dry eye disease. *Cornea* **39**(Suppl 1), S39–s46. <https://doi.org/10.1097/ico.0000000000002500> (2020).

Author contributions

Conception and design of study: S.H.N., M.Y.S., H.L.; acquisition of data: M.Y.S., H.L.; analysis of data: S.H.N., M.Y.S., H.L.; drafting the manuscript: S.H.N., M.Y.S., J.Y.H., S.Y.M., J.Y.K, H.T., H.L.; revising the manuscript critically for important intellectual content: S.H.N., M.Y.S., J.Y.H., S.Y.M., J.Y.K, H.T., H.L.; approval of the version of the manuscript to be published: S.H.N., M.Y.S., J.Y.H., S.Y.M., J.Y.K, H.T., H.L.

Funding

This work was supported by the Korea Medical Device Development Fund, granted by the Korean government (the Ministry of Science and ICT; the Ministry of Trade, Industry, and Energy; the Ministry of Health and Welfare, the Ministry of Food and Drug Safety) (Project number: 9991006821, KMDF_PR_20200901_0148), by the Korean Fund for Regenerative Medicine, funded by the Ministry of Science and ICT; the Ministry of Health and Welfare (21C0723L1-11, Republic of Korea), and by a grant from the Asan Institute for Life Sciences, Asan Medical Center, Seoul, Korea (2022IP0019-1,2021IP0061-2).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-13344-5>.

Correspondence and requests for materials should be addressed to H.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022