



RESEARCH ARTICLE

WILEY

Applicability of internet search index for asthma admission forecast using machine learning

Li Luo¹ | Chengcheng Liao¹ | Fengyi Zhang¹ | Wei Zhang² |
Chunyang Li² | Zhixin Qiu³ | Debin Huang⁴

¹Business School, Sichuan University, China

²West China Biomedical Big Data Center, West China Hospital, Sichuan University, China

³Respiratory and Critical Care Medicine, West China Hospital, Sichuan University, China

⁴Chengdu Medical Insurance Administration, China

Correspondence

Fengyi Zhang, Business School, Sichuan University, China.

Email: zhangfengyi90@163.com

Funding information

National Natural Science Foundation of China, Grant/Award Numbers: 71172197, 71131006 and 71532007

Summary

Objective: This study aimed to determine whether a search index could provide insight into trends in asthma admission in China. An Internet search index is a powerful tool to monitor and predict epidemic outbreaks. However, whether using an internet search index can significantly improve asthma admissions forecasts remains unknown. The long-term goal is to develop a surveillance system to help early detection and interventions for asthma and to avoid asthma health care resource shortages in advance.

Methods: In this study, we used a search index combined with air pollution data, weather data, and historical admissions data to forecast asthma admissions using machine learning.

Results: Results demonstrated that the best area under the curve in the test set that can be achieved is 0.832, using all predictors mentioned earlier.

Conclusion: A search index is a powerful predictor in asthma admissions forecast, and a recent search index can reflect current asthma admissions with a lag-effect to a certain extent. The addition of a real-time, easily accessible search index improves forecasting capabilities and demonstrates the predictive potential of search index.

KEYWORDS

asthma hospitalization, internet search, machine learning

1 | INTRODUCTION

Asthma is a common chronic respiratory disease that imposes substantial burdens on individuals and society.^{1,2} China has the largest population suffering from asthma; further, it also shows the highest incidence of asthma prevalence globally.³ In addition, the incidence of asthma is time-varying with seasonal climate and air conditions, which leads to considerable variations in asthma admissions with regard to time and severity. In such a scenario, medical institutions generally cannot meet the demand of asthma-related medical resources in a timely and precise manner, and a shortage in asthma-related health care resources may occur, which results in timely medical aid being unavailable. Hence, the prediction of asthma-related hospital admissions can enable health care resource managers to make more informed decisions about resource allocation and scheduling, improve the quality of health care services, and consequently protect the right of asthma patients to access timely health care and reduce the societal burden of asthma.⁴⁻⁷

Several previous studies on asthma admissions forecast have been conducted. Moustiris et al⁸ applied and developed artificial neural network models to predict the total weekly number of childhood asthma admissions in the greater Athens area in Greece. This study found that given more corresponding input data, an artificial neural network could provide a more reliable forecast for childhood asthma admissions. In order to predict daily hospital asthma admissions in London, Soyiri et al^{9,10} demonstrated a multi-stage quantile regression approach and also developed 2 related negative binomial models compared with a naïve seasonal model to predict excess demand for health care services, using retrospective data from the Hospital Episode Statistics, weather, and air quality. A negative binomial regression model was also used to measure the association between daily asthma hospital admissions and ambient air pollution concentrations.¹¹ In all existing asthma admissions forecast researches, historical asthma admissions, weather, and air pollutants are the common predictors for asthma admissions forecast. Some researchers have proved that the number of asthma admissions today can be predicted based on historical admissions.^{9,10,12-14} In addition, other studies have also proved that asthma admissions are related to weather and air pollutants.¹⁵⁻²⁵ Most experts agree that among various meteorological factors, respirable particulate matter (PM₁₀), fine particulate matter (PM_{2.5}), nitrogen dioxide (NO₂), and sulfur dioxide (SO₂) contribute the most to asthma epidemics.²⁶⁻³⁰

Furthermore, with advances in technology, the Internet has been integrated into people's daily lives, especially in China. The China Statistical Report on Internet Development had predicted that China would be the world's largest Internet community by December 31, 2016. Among Internet users in China, 95.1% of 731 million Internet users would access the Internet via mobile phone.³¹ Internet data gathered from such users can not only monitor behavior, but also can support behavioral forecast. Artola et al³² used the index on Internet search for measuring the relative frequency of keywords associated with traveling to forecast tourism inflows into Spain. Amuri et al³³ utilized the Google Jobs Index as a leading indicator to make the forecast of unemployment dynamics in the United States more accurate. Jun et al³⁴ used search traffic information from Google Insights to analyze actual consumers' attitudes regarding a product. As the capability to monitor and predict events emerges, Google, Yahoo, and other search engines have been widely used in many fields, such as within the social science, economic, scientific, and technological fields, to forecast individuals' behavior.³⁵⁻⁴⁹

Google Trends and other search engines have also demonstrated their capabilities in health care-related prediction. Disease surveillance via Internet in the past decade has provided new data sources for prediction. The earliest and most well-known application of forecasts is the use of Google Flu Trends to predict influenza trends between 2003 and 2007. Since then, several studies have focused on forecasting seasonal influenza epidemics by analyzing Internet search queries.⁵⁰⁻⁵⁷ Ginsberg et al⁵⁸ tracked influenza-like illnesses among the population of the United States and proposed a method for analyzing a large number of Google search queries to accurately estimate the influenza activity in each area of the United States. Li et al⁵⁹ used data from the Dengue Baidu Search Index to help develop a predictive model of dengue fever in China. Wang et al⁶⁰ combined the local Internet search data from Google Trends with original data collected from hospitals in Taiwan, to forecast the incidence of dementia and dementia-related outpatient visits. Fatima et al³⁷ showed that 4 out of 5 users use the Internet to find personalized health care information related to certain diseases and related treatments. In fact, Internet search data have already been utilized to predict morbidity of

seasonal diseases and the behavior of patients, which could be used as a means to improve the forecast accuracy of hospital admissions. For example, Ekström et al⁶¹ used Internet search data to forecast emergency department (ED) visits for the following day. In their research, the correlation between the number of website visits between 6 PM and midnight and ED visits the following day was proved to be significant. The large population of web search user groups makes search queries a new way to monitor and predict behavior related to health care. Thus, it is rational to consider Internet search data as a new supplement to traditional surveillance data,⁶² which could enable public health officials and health care professionals to be better prepared in responding to related admissions epidemics.

When Internet search data are employed to make predictions of health care epidemics and provide early warnings, these insights can prove to be extremely valuable. However, Internet search data is just one of the factors that can forecast asthma admissions; there are several other related factors. Hence, a more feasible way to forecast asthma admissions is to combine search index (SI) and traditional factors.

In this study, we combined SI, air pollution data, weather data, and historical admissions data to forecast asthma admissions using machine learning. This practical forecast model could provide advanced warning for medical institutions, and health care resource managers can schedule the corresponding resources according to the forthcoming demand, and thus guarantee the accessibility of timely health care resources. Based on our research, a surveillance system to help early detection and interventions for asthma can be implemented in advance to avoid asthma health care resource shortages.

2 | RESEARCH DESIGN

2.1 | Data

We obtained the asthma hospital admissions data from the Human Resources and Social Security Bureau of a city located in Southwest China, which corresponds to asthma hospital admissions between 1 January 2014 to 31 December 2014 (ie, 365 days of continuous data), with 6813 admission records. The city has approximately 12.8 million residents whose data have been recorded. "Asthma hospital admissions" denotes the admissions that involved a primary diagnosis of asthma (ICD-10: J45.001, J45.005, J45.901, J45.903).

We obtained the daily (24-hour average) air pollution concentration data from the Environment Monitoring Center database,⁶³ which included PM_{2.5}, PM₁₀, SO₂, and NO₂ concentrations between January 1, 2014 to December 31, 2014 (365 days). All data were recorded in kg/m³ but converted to mg/m³ for carbon monoxide and µg/m³ for other pollutants.

Meteorological data were obtained from the website of the city's Meteorological Bureau,⁶⁴ which were based on averaged daily measurements from the weather monitoring sites across that city between 1 January 2014 to 31 December 2014 (365 days). In this study, only the highest and lowest temperatures were considered and recorded in °C.

In China, Baidu (<http://index.baidu.com/>) is the most popular search engine, and it is preferred by approximately 80.54% of all Internet users.⁶⁵ Some researchers have proved that Internet search query data from Baidu can be used to monitor and predict the epidemics in China. We collected the SI on asthma from Baidu, the search engine that holds the biggest market share in China.^{66,67} Baidu Search Index provided Baidu query volume data (absolute numbers) of asthma-related online searches in that city between 1 January 2014 and 31 December 2014 (365 days).

2.2 | Data analysis

Data analysis involved 2 stages, namely, descriptive statistics and a forecast process. First, we performed descriptive statistical analyses on air pollution data, meteorological data, SI, and historical data. Next, maximal information coefficient (MIC) was employed to measure the correlation between admissions status and each predictor, considering lags. Admissions status includes 2 statuses: normal and excess. "Normal" status refers to a scenario in which the number of asthma admissions within a certain day is lower than the capacity limit, while "excess" refers to a scenario in

which the number of asthma admissions are higher than the capacity limit. In our experiment, the capacity limit was defined as 80% of the highest admissions.

In the forecast process, considering the aforementioned 4 air pollutants (PM_{2.5}, PM₁₀, SO₂, and NO₂) widely considered in asthma admissions forecast, we define these 4 predictors as basic predictors, and the scheme with only basic predictors as the baseline scheme. Besides, other predictors such as highest temperature, lowest temperature, historical admissions, and SI are defined as optional predictors. We have tried 16 (2⁴) schemes, each of which combines the baseline scheme and at least 1 optional predictor. Lag effects were also considered in this experiment. When we used the data from the previous day to N days before, we consider that the scheme's lag is N. For each scheme, the lag varies between 1 and 14, and we define the lag with best performance as the optimal lag. In other words, to achieve the best scheme with an optimal lag, we performed 224 (16 × 14) train-test experiments with different schemes and lags, and the train-test ratio was 8:2. We employed XGBoost as the classifier to predict normal status or excess. Area under the curve (AUC) is the most popular statistic for model comparison, and we used it as the performance indicator in this research.

3 | RESULTS

Table 1 summarizes the basic statistics of corresponding data. A total of 7503 admissions for asthma were recorded from January 1, 2014 to December 31, 2014. The daily admissions varied between 2 and 44 with a mean of 20.56. During the study period, the average of daily concentrations was 71.86 µg/m³ for PM_{2.5}, 116.28 µg/m₃ for PM₁₀, 17.45 µg/m³ for SO₂, 52.15 µg/m³ for NO₂, 21.11°C for highest temperature, 13.35°C for lowest temperature, and 184.10 for SI, respectively.

As mentioned earlier, the SI corresponds with asthma admissions numbers. According to Figure 1, SI showed trends similar to admissions amount, and many of its peaks and troughs appeared a few days before that of admissions amount. For further research on this relation, we employed MIC to measure the tightness between SI and asthma admissions numbers, and its results are listed in Table 2. We found that SI achieved the largest MIC among all predictors for almost all lags. Exceptions only appeared when compared with PM₁₀ and lag bigger than 7. This shows that it is reasonable to use SI as a predictor to predict excess admissions numbers.

Percentage refers to the percentage of admissions and Internet search in the whole year.

Table 3 described the performance of each scheme, and the SI plays an important role in analyzing it. Disparity appeared in several aspects, and we analyzed it from the following aspects: optimal lag, performance in train set, and performance in test set.

As mentioned in the data analysis subsection, we assessed the performance of each scheme using AUC. The maximal AUC in the training set was 1, while the minimum was 0.706. Almost all AUCs were bigger than 0.9, which is a common standard for excellent classifiers. When taking SI into consideration, non-search index (NSI) group ranged from 0.807 to 1, while SI group ranged from 0.706 to 1. Data within Table 4 demonstrate that there is no significant difference between the SI group and NSI group from the perspective of AUC in training set.

TABLE 1 Statistics of corresponding data

	Observed	Minimum	Maximum	Mean	SD
PM _{2.5}	365	10	396	71.86	52.34
PM ₁₀	365	20	562	116.28	71.68
SO ₂	365	3	61	17.45	9.65
NO ₂	365	20	109	52.15	15.73
Highest temperature	365	4	34	21.11	7.56
Lowest temperature	365	-2	24	13.35	7.40
Search index	365	14	282	184.10	28.21
Admission amount	365	2	44	20.56	6.52

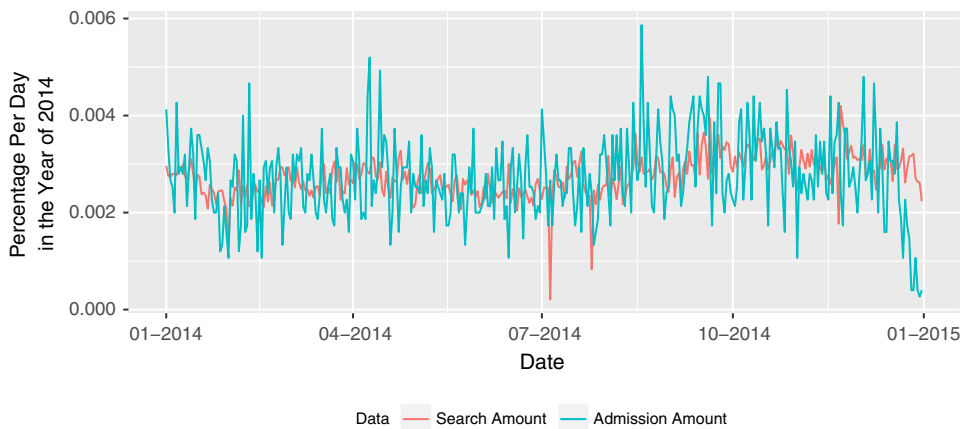


FIGURE 1 Historical data of search index and admissions [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Maximal information coefficients between each predictor and admissions number considering lags

Lag	NO ₂	SO ₂	PM2.5	PM10	Low	High	Search Index
0	-0.063	0.006	-0.121	-0.116	0.049	0.024	0.165
1	-0.063	-0.027	-0.121	-0.12	0.074	0.045	0.146
2	-0.076	-0.034	-0.102	-0.112	0.081	0.052	0.155
3	-0.094	-0.1	-0.117	-0.122	0.106	0.061	0.121
4	-0.054	-0.087	-0.115	-0.111	0.107	0.069	0.115
5	-0.024	-0.079	-0.1	-0.11	0.098	0.06	0.105
6	-0.049	-0.067	-0.089	-0.1	0.107	0.062	0.129
7	-0.051	0.009	-0.09	-0.096	0.112	0.075	0.119
8	-0.064	-0.058	-0.089	-0.107	0.117	0.09	0.1
9	-0.057	-0.093	-0.109	-0.129	0.111	0.094	0.094
10	-0.084	-0.074	-0.086	-0.113	0.12	0.094	0.086
11	-0.064	-0.058	-0.089	-0.107	0.117	0.09	0.1
12	-0.028	-0.061	-0.085	-0.105	0.124	0.083	0.07
13	-0.005	-0.056	-0.064	-0.087	0.123	0.098	0.05
14	-0.036	-0.051	-0.105	-0.121	0.131	0.11	0.103

Considering the performance aspect in the test set, AUCs of all schemes ranged from 0.740 to 0.832. Almost all AUCs were bigger than 0.7, which is a common standard for a good classifier. When taking SI into consideration, the NSI group ranged from 0.775 to 0.829, while SI group ranged from 0.710 to 0.832. Data presented in Table 4 demonstrate that there is no significant difference between the SI group and NSI group in the aspect of AUC in the test set.

The overall optimal lags ranged from 3 to 13. NSI group ranged from 3 to 10, while SI group ranged from 8 to 13. In particular, the schemes with maximal optimal lag all belonged to NSI group, while those with minimal optimal lag belonged to SI group. In fact, almost all optimal lags of NSI group were larger than that of SI group. Data given in Table 4 demonstrate that there is a significant difference between the SI group and NSI group in the aspect of optimal lag.

The scheme that involved all optional predictors is considered as the best, owing to its outstanding performance (AUC: 0.832). When compared with the scheme without any optional predictor (AUC: 0.800), it shows considerable improvement. However, not all schemes considering only 1 optional predictor excel the one without any optional

TABLE 3 Optimal performance for each scheme considering lags

High	Low	Amount	Search Index	AUC In Test Set	Optimal Lag	AUC In Train Set
Y	Y	N	N	0.762	13	0.991
Y	Y	Y	N	0.809	13	0.991
Y	Y	N	Y	0.748	3	0.991
Y	Y	Y	Y	0.832	8	0.991
N	Y	N	N	0.745	10	0.991
N	Y	Y	N	0.743	14	0.762
N	Y	N	Y	0.74	13	1
N	Y	Y	Y	0.815	8	0.991
N	N	N	N	0.8	13	0.996
N	N	N	Y	0.829	9	0.986
N	N	Y	Y	0.761	10	1
N	N	Y	N	0.792	13	1
Y	N	N	N	0.775	9	0.807
Y	N	N	Y	0.71	8	0.706
Y	N	Y	Y	0.789	10	1
Y	N	Y	N	0.789	8	0.948

Y, considered; N, not considered.

TABLE 4 Disparity between NSI group and SI group

	AUC in Train Set		AUC in Test Set		Optimal Lag	
	Mean	P	Mean	P	Mean	P
NSI	0.9581	0.6712	0.777	0.5241	11.625	0.0176
SI	0.9358		0.778		8.625	

P: P value of t-test, and the null hypothesis is the corresponding values of NSI is bigger than SI.

predictor. Among them, only the scheme with SI performed better than the baseline model with an AUC of 0.829, and it is much closer to the best scheme. When the other 3 optional predictors are included in the scheme, the improvement is just 0.003, while the other 3 schemes with 3 specific optional predictors each performed worse than the baseline scheme.

4 | DISCUSSION

Search index has several advantages. First, SI is a free and easily accessible search tool. Second, SI (including Google and Baidu) is now available in more than 70 countries worldwide and thus allows the tracking of searches in different languages and regions. Finally, SI is updated daily, which, in the context of health care concerns, permits more frequent monitoring of patients' symptoms and could facilitate early detection of asthma admissions. In this regard, SI can be used to notify health policy makers and hospital managers to prepare for upcoming outpatient visits, where screening work is expected to achieve early detection and treatment and thus save lives while reducing health care costs.

As we mentioned in the introduction, Google Trends first provided an excellent example in 2003 to 2007 to illustrate the contribution of Internet search data on the prediction of infectious diseases.⁵⁸ However, Google Trends failed to successfully predict the seasonal and pandemic influenza in the USA during the 2012/2013 season.⁶⁸ It has been debated that Internet-based queries might misrepresent the epidemic curve in practice.^{69,70} Some

researchers analyzed the reasons for this failure in the Google Trends model, suggesting that the Internet-based query system can be used as a supplement to, but not a substitute for, traditional data collection and analysis.⁶⁵ Hence, in this study, we considered a SI in a particular city, coupled with air pollution, meteorological, and historical admissions data. We consider all schemes employing air pollution as basic predictors and lowest temperature, highest temperature, historical admissions, and SI as optional predictors.

In addition, XGBoost, a powerful machine learning method, was used as the classifier to predict asthma admissions. Although machine learning has been shown to deliver competitive performance in modeling and forecasting a wide range of complex natural phenomena and disparate social systems, its utility for asthma admissions studies is underexplored. To our knowledge, this is the first study that employs machine learning methodology for forecasting of asthma admissions and explores its predictive performance.

In general, our proposed approach can be a feasible alternative to forecast asthma admissions in any region. The method could produce accurate asthma admissions forecasts and provide policymakers with sufficient time to enhance health risk management strategies, and can help hospitals plan their admissions resources accordingly and lower their costs by optimizing supplies and staffing. It can also improve service quality by decreasing admissions wait times and overcrowding.

5 | CONCLUSIONS

We have developed a practical, city-based forecast model using machine learning and SI. The addition of easily accessible SI improves the forecasting capabilities and demonstrates the predictive potential of SI. This practical forecast model could provide warning regarding potentially high numbers of asthma admissions for medical institutions in advance, thus allowing time to implement an appropriate response to increase in patient volumes.

There are other interesting findings. By splitting all conclusions into NSI group and SI group, 2 conclusions can be drawn: (1) Introducing SI as a predictor cannot significantly improve performance; (2) However, it can significantly reduce the lag effects, which means that when SI considered, only recent information is needed to make accurate prediction for asthma admissions. This may be because of SI's timeliness.

Our research has certain limitations. These are: (1) Our data only cover 1 city; however, data collected from multiple cities can be more convincing. (2) Asthma admissions defined in this research were not age specific; that is to say, we did not consider different groups such as young and old. This is an important consideration, because patients of different ages have different health care demands in terms of capacity requirements as well as service type; and (3) Other social media data can also be used in asthma admissions forecasts. These limitations need to be addressed in future studies.

Our tool is ultimately designed for clinical application. Integrating this potentially powerful forecasting tool into medical institutions requires additional clinical feasibility and effectiveness studies so that such forecasts can be suitably linked with clinical intervention schemes designed to mitigate overcrowding.

In the future, we also plan to evaluate our approach by applying it to other geographical regions, and to incorporate local socio-demographic information and networks of everyday social contacts. To assess a hierarchical social communication structure of contacts, we can use various offline information sources, such as census, school enrollment and public transportation data, and georeferenced online data from sources such as Meetup and Twitter.

ETHICAL APPROVAL

The study does not involve human subjects and adheres to all current laws of China.

ACKNOWLEDGEMENT

We acknowledge the efforts from the West China Biomedical Big Data Center, West China Hospital, Sichuan University, including providing research data, data extraction, cleaning, and standardization, supervising no sensitive

data being reported, and coordination services for the publication of this article. The study was funded by the National Natural Science Foundation of China (no. 71532007, no. 71131006, and no. 71172197).

CONFLICT OF INTERESTS

None declared.

ORCID

Fengyi Zhang  <http://orcid.org/0000-0002-6261-2171>

REFERENCES

1. Akinbami LJ, Moorman JE, Liu X. Asthma prevalence, health care use, and mortality: United States, 2005-2009.[J]. *Nat Health Stat Rep*. 2011;12(32):1.
2. Zahran HS, Bailey C, Garbe P. Vital signs: asthma prevalence, disease characteristics, and self-management education: United States, 2001-2009[J]. *MMWR Morb Mortal Wkly Rep*. 2011;60(17):547.
3. Chen ZH, Wang PL, Shen HH. Asthma research in China: a five-year review[J]. *Respirology*. 2013;18(S3):10-19.
4. Li G, Jing H, Xu G, et al. The short term burden of ambient fine particulate matter on chronic obstructive pulmonary disease in Ningbo, China[J]. *Environ Health*, 2017, 16(1):54.
5. Bryden C, Bird W, Titley HA, Halpin DMG, Levy ML. Stratification of COPD patients by previous admission for targeting of preventative care.[J]. *Respir Med*. 2009;103(4):558-565.
6. Rodó X, Ballester J, Dan C, et al. Association of Kawasaki disease with tropospheric wind patterns.[J]. *Sci Rep*. 2011;1(6057):152.
7. Tsai PF, Lin FM. An application of multi-attribute value theory to patient-bed assignment in hospital admission management: an empirical study.[J]. *J Healthcare Eng*. 2014;5(4):439-456.
8. Moustris KP, Douros K, Nastos PT, et al. Seven-days-ahead forecasting of childhood asthma admissions using artificial neural networks in Athens, Greece[J]. *Int J Environ Health Res*. 2012;22(2):93-104.
9. Soyiri IN, Reidpath DD, Sarran C. Forecasting asthma-related hospital admissions in London using negative binomial models.[J]. *Chron Respir Dis*. 2013;10(2):85-94.
10. Soyiri IN, Reidpath DD, Sarran C. Forecasting peak asthma admissions in London: an application of quantile regression models[J]. *Int J Biometeorol*. 2013;57(4):569-578.
11. Restrepo CE, Simonoff JS, Thurston GD, et al. Asthma hospital admissions and ambient air pollutant concentrations in New York City[J]. *J Environ Prot*. 2012;17(6).
12. Zhang Y, Wang S, Liu Y. Application of ARIMA model on predicting monthly hospital admissions and hospitalization expenses for respiratory diseases[J]. *Chin J Health Stat*. 2015.
13. Gouveia N, Fletcher T. Respiratory diseases in children and outdoor air pollution in S?O Paulo, Brazil: a time series analysis.[J]. *Occup Environ Med*. 2000;57(7):477-483.
14. Hua J, Yin Y, Peng L, du L, Geng F, Zhu L. Acute effects of black carbon and PM_{2.5} on children asthma admissions: a time-series study in a Chinese city[J]. *Sci Total Environ*. 2014;481(1):433-438.
15. Anderson HR, Butland BK, Donkelaar AV, et al. Satellite-based estimates of ambient air pollution and global variations in childhood asthma prevalence[J]. *Environ Health Perspect*. 2012;120(9):1333-1339.
16. Gerke AK, Yang M, Tang F, Foster ED, Cavanaugh JE, Polgreen PM. Association of hospitalizations for asthma with seasonal and pandemic influenza[J]. *Respirology*. 2014;19(1):116-121.
17. McLaren J, Williams ID. The impact of communicating information about air pollution events on public health[J]. *Sci Total Environ*. 2015;538:478-491.
18. Makra L, Puskás J, Matyasovszky I, et al. Weather elements, chemical air pollutants and airborne pollen influencing asthma emergency room visits in Szeged, Hungary: performance of two objective weather classifications[J]. *Int J Biometeorol*. 2015;59(9):1269-1289.
19. Spickett JT, Brown HL, Rumchev K. Climate change and air quality: the potential impact on health. Asia-Pacific journal of public health. *Asia Pac J Public Health*. 2011;23(2_suppl):37S-45S.
20. Sheffield PE, Knowlton K, Carr JL, Kinney PL. Modeling of regional climate change effects on ground-level ozone and childhood asthma.[J]. *Am J Prev Med*. 2011;41(3):251-257.

21. Deschamps KM. Associations between daily asthma hospital admissions and ambient air pollutants in Montreal, 1992 to 1999[J]. *Geography Planning & Environment*, 2003.
22. Aabe T, Tokuda Y, Ohde S, et al. The relationship of short-term air pollution and weather to ED visits for asthma in Japan. [J]. *Am J Emerg Med*. 2009;27(2):153-159.
23. Cakmak S, Dales RE, Coates F. Does air pollution increase the effect of aeroallergens on hospitalization for asthma?[J]. *J Allergy Clin Immunol*. 2012;129(1):228-231.
24. Miller MA, Levisky ME, Enslow W, Rosin A. Randomized evaluation of octreotide vs prochlorperazine for ED treatment of migraine headache[J]. *Am J Emerg Med*. 2009;27(2):160-164.
25. Sun HL, Chou MC, Lue KH. The relationship of air pollution to ed visits for asthma differ between children and adults. *Am J Emerg Med*. 2006;24(6):709-713.
26. Eken C. The relationship of air pollution to ED visits for asthma differs between children and adults. *Am J Emerg Med*. 2007;25(7):852.
27. Tecer LH, Alagha O, Karaca F, Tuncel G, Eldes N. Particulate matter (PM(2.5), PM(10-2.5), and PM(10)) and children's hospital admissions for asthma and respiratory diseases: a bidirectional case-crossover study[J]. *J Toxicol Environ Health A*. 2008;71(8):512-520.
28. Mar TF, Jansen K, Shepherd K, Lumley T, Larson TV, Koenig JQ. Exhaled nitric oxide in children with asthma and short-term PM2.5 exposure in Seattle[J]. *Environ Health Perspect*. 2005;113(12):1791-1794.
29. Yun RS, Jin SU, Hong LY, et al. Effects of environmental pm2.5 pollution on the morbidity of asthma and lung cancer. *J Labour Med*. 2002;23(4):367-378.
30. Weinmayr G, Romeo E, Sario MD, et al. Short-term effects of PM₁₀ and NO₂ on respiratory health among children with asthma or asthma-like symptoms: a systematic review and meta-analysis[J]. *Environ Health Perspect*. 2010;118(4):449-457.
31. the 39th China Statistical Report on Internet Development. available:<http://www.cnnic.net.cn/hlwfzyj/hlwxbzg/hlwjtjb/201701/P020170123364672657408.pdf>
32. Artola C, Pinto F, García PDP. Can internet searches forecast tourism inflows?[J]. *Int J Manpow*. 2015;36(1):103-116.
33. D'Amuri F, Marcucci J. 'Google it!' *Forecasting the US Unemployment Rate with a Google Job Search Index*[J]. Social Science Electronic Publishing; 2009.
34. Jun SP, Park DH, Yeom J. The possibility of using search traffic information to explore consumer product attitudes and forecast consumer preference[J]. *Technol Forecast Soc Change*. 2014;86(340):237-253.
35. Li X, Pan B, Law R, Huang X. Forecasting tourism demand with composite search index[J]. *Tour Manag*. 2017;59:57-66.
36. Penna ND, Huang H. Constructing a consumer confidence index for the US using web search volume[J]. 2009.
37. Fatima I, Halder S, Saleem MA, et al. Smart CDSS: integration of social media and interaction engine (SMIE) in healthcare for chronic disease patients[J]. *Multimed Tool Appl*. 2015;74(14):5109-5129.
38. Li S, Peng G, Wang J. Search queries based sales forecasting on Taobao[C]. *IEEE, 2014:223-228*. International Conference on Computer and Information Science. IEEE, 2014:223-228.
39. Webb GK. Forecasting U.S. home foreclosures with an index of Internet keyword searches. *Value Creation in E-Business Management. Americas Conference on Information Systems, Amcis 2009, Sigebiz Track, San Francisco, Ca, Usa, August 6-9, 2009. Selected Papers* (Vol.36, pp.196-203). DBLP; 2009.
40. Petroligis TI, Pinson P. Early warnings of extreme winds using the ECMWF extreme forecast index[J]. *Meteorol Appl*. 2014;21(2):171-185.
41. Teodoro MF. Industrial production index forecast: methods comparison[J]. 2012;1504(1):1103-1107.
42. Xu W, Zheng T, Li Z. A neural network based forecasting method for the unemployment rate prediction using the search engine query data[C]. *IEEE, international conference on E-business engineering*. IEEE Computer Society, 2011:9-15.
43. Xu W, Li Z, Cheng C, Zheng T. Data mining for unemployment rate prediction using search engine query data[J]. *SOCAL*. 2013;7(1):33-42.
44. Hassani H, Silva ES. Forecasting energy data with a time lag into the future and Google trends[J]. *Int J Energy Stat*. 2017;4(4). 1650020
45. Askitas N. Predicting road conditions with internet search[J]. *PLoS One*. 2016;11(8):e0162080.
46. Rivera R. A dynamic linear model to forecast hotel registrations in Puerto Rico using Google trends data[J]. *Tour Manag*. 2016;57:12-20.
47. Fantazzini D, Toktamysova Z. Forecasting German car sales using google data and multivariate models. *Int J Prod Econ*. 2015;170:97-135.

48. Smith P. Google's MIDAS touch: predicting UK unemployment with Internet search data[J]. *J Forecast*. 2016;35(3):n/a-n/a):263-284.
49. Vicente MR, López-Menéndez AJ, Pérez R. Forecasting unemployment with internet search data: does it help to improve predictions when job destruction is skyrocketing?[J]. *Technol Forecast Soc Change*. 2015;92(92):132-139.
50. Zou X, Zhu W, Yang L, Shu Y. [Google flu trends—the initial application of big data in public health][J]. *Zhonghua Yu Fang Yi Xue Za Zhi*. 2015;49(6):581-584.
51. Moss R, Zarebski A, Dawson P, McCaw JM. Forecasting influenza outbreak dynamics in Melbourne from Internet search query surveillance data[J]. *Influenza Other Respi Viruses*. 2016;10(4):314-323.
52. Araz OM, Bentley D, Muelleman RL. Using Google flu trends data in forecasting influenza-like-illness related ED visits in Omaha, Nebraska.[J]. *Am J Emerg Med*. 2014;32(9):1016-1023.
53. Moss R, Zarebski A, Dawson P, et al. Retrospective forecasting of the 2010-2014 Melbourne influenza seasons using multiple surveillance systems[J]. *Epidemiol Infect*. 2017;145(1):156-169.
54. Xu Q, Gel YR, Ramirez Ramirez LL, Nezafati K, Zhang Q, Tsui KL. Forecasting influenza in Hong Kong with Google search queries and statistical model fusion[J]. *Plos One*. 2017;12(5):e0176690.
55. Bardak B, Tan M. Prediction of influenza outbreaks by integrating Wikipedia article access logs and Google flu trend data [C]. IEEE, International Conference on Bioinformatics and Bioengineering. IEEE, 2015:1-6.
56. Nsoesie E, Marathe M, Brownstein J. Forecasting peaks of seasonal influenza epidemics.[J]. *PLOS Currents*. 2013;5.
57. Preis T, Moat HS. Adaptive nowcasting of influenza outbreaks using Google searches[J]. *R Soc Open Sci*. 2014;1(2):140095.
58. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature*. 2009;457(7232):1012-1014.
59. Li Z, Liu T, Zhu G, et al. Dengue Baidu search index data can improve the prediction of local dengue epidemic: a case study in Guangzhou, China[J]. *PLoS Negl Trop Dis*. 2017;11(3):e0005354.
60. Wang HW, Chen DR, Yu HW, Chen YM. Forecasting the incidence of dementia and dementia-related outpatient visits with Google trends: evidence from Taiwan[J]. *J Med Internet Res*. 2015;17(11):e264.
61. Ekström A, Kurland L, Farrokhnia N, Castrén M, Nordberg M. Forecasting emergency department visits using internet data.[J]. *Ann Emerg Med*. 2015;65(4):436-442.e1.
62. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google flu: traps in big data analysis[J]. *Science*. 2014;343(6176):1203-1205.
63. Chengdu Environment Monitoring Center (EMC). Available:database,http://www.cdemc.org
64. Chengdu Meteorological Bureau. Available: http://www.cdmb.gov.cn
65. Search Engine Market Share in China. 2017 Available: http://gs.statcounter.com/search-engine-market-share/all/china/#monthly-201704-201704-bar
66. Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS. Monitoring influenza epidemics in China with search query from Baidu. *PLoS One*. 2013;8(5):e64323. doi: <https://doi.org/10.1371/journal.pone.0064323> PMID: 23750192.
67. Gu Y, Chen F, Liu T, et al. Early detection of an epidemic erythromelalgia outbreak using Baidu search data[J]. *Sci Rep*. 2015;5(1):12649.
68. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing Google flu trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales[J]. *PLoS Comput Biol*. 2013;9(10):e1003256.
69. Milinovich GJ, Williams GM, Clements ACA, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases[J]. *Lancet Infect Dis*. 2014;14(2):160-168.
70. Althouse BM, Scarpino SV, Meyers LA, et al. Enhancing disease surveillance with novel data streams: challenges and opportunities[J]. *EpJ Data Sci*. 2015;4(1):17.

How to cite this article: Luo L, Liao C, Zhang F, et al. Applicability of internet search index for asthma admission forecast using machine learning. *Int J Health Plann Mgmt*. 2018;1-10. <https://doi.org/10.1002/hpm.2525>