Science of the Total Environment 770 (2021) 144746

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Application of data science methods to identify school and home risk factors for asthma and allergy-related symptoms among children in New York

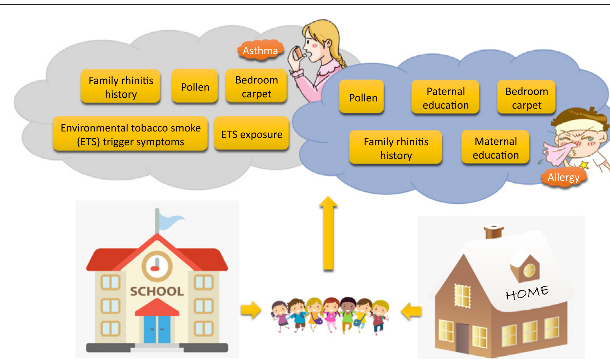
Xinlei Deng [a], George Thurston [b], Wangjian Zhang [a], Ian Ryan [a], Connie Jiang [c], Haider Khwaja [a,d], Xiaobo Romeiko [a], Tia Marks [a], Bo Ye [e], Yanji Qu [f], Shao Lin [a,e,*]

[a] Department of Environmental Health Sciences, University at Albany, State University of New York, Rensselaer, NY 12144, USA
[b] Department of Environmental Medicine, School of Medicine, New York University, New York, NY 12144, USA
[c] Department of Statistics, College of Arts and Sciences, University of Virginia, Charlottesville, VA 22904, USA
[d] Wadsworth Center, New York State Department of Health, Albany, NY 12201, USA
[e] Department of Epidemiology and Biostatistics, University at Albany, State University of New York, Rensselaer, NY 12144, USA
[f] Guangdong Cardiovascular Institute, WHO Collaborating Center for Research and Training in Cardiovascular Diseases, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510080, China

HIGHLIGHTS

- Family rhinitis, pollen and carpet were the top risk factors for asthma and allergy.
- Age <9, family rhinitis and carpet jointly increased the risk of asthma.
- Pollen, solvents, and carpet jointly increased the risk of allergy.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 11 September 2020
Received in revised form 9 December 2020
Accepted 21 December 2020
Available online 23 January 2021

Editor: SCOTT SHERIDAN

Keywords:
Asthma
Allergy
Home and school environment
Machine-learning
Risk factors

ABSTRACT

Objectives: Few studies have comprehensively assessed multiple environmental exposures affecting children's health. This study applied machine-learning methods to evaluate how indoor environmental conditions at home and school contribute to asthma and allergy-related symptoms.
Methods: We randomly selected 10 public schools representing different socioeconomic statuses in New York State (2017–2019) and distributed questionnaires to students to collect health status and home-and school-environmental exposures. Indoor air quality was measured at school, and ambient particle exposures ($PM_{2.5}$ and components) were measured using real-time personal monitors for 48 h. We used random forest model to identify the most important risk factors for asthma and allergy-related symptoms, and decision tree for visualizing the inter-relationships among the multiple risk factors with the health outcomes.
Results: The top contributing factors identified for asthma were family rhinitis history (relative importance: 10.40%), plant pollen trigger (5.48%); bedroom carpet (3.58%); environmental tobacco smoke (ETS) trigger symptom (2.98%); and ETS exposure (2.56%). For allergy-related symptoms, plant pollen trigger (10.88%), higher paternal education (7.33%), bedroom carpet (5.28%), family rhinitis history (4.78%), and higher maternal education (4.25%) were the strongest contributing factors. Conversely, primary heating with hot water radiator was negatively (−6.86%) associated with asthma symptoms. Younger children (<9 years old) with family history

* Corresponding author at: Department of Environmental Health Sciences, School of Public Health, University at Albany, State University of New York, USA.
E-mail address: slin@albany.edu (S. Lin).

https://doi.org/10.1016/j.scitotenv.2020.144746

of rhinitis and carpeting in the bedroom were the prominent combined risk factors for asthma. Children jointly exposed to pollen, solvents, and carpeting in their home tended to have greater risks of allergy-related symptoms, even without family history of rhinitis.

Conclusion: Family rhinitis history, bedroom carpet, and pollen triggers were the most important risk factors for both asthma and allergy-related symptoms. Our new findings included that hot-water radiator was related to reduced asthma symptoms, and the combination of young age, rhinitis history, and bedroom carpeting was related to increased asthma symptoms. Further studies are needed to confirm our findings.

1. Introduction

Asthma and allergy-related symptoms are important health concerns prevalent among primary school-aged children (Nyenhuis et al., 2019; Kagan et al., 2003). Based on the U.S. National Health Interview Survey (NHIS) in 2017, current asthma prevalence among children is 8.4%, which is higher than the 7.7% among adults (Asthma Surveillance Data, 2019). The prevalence was even higher among children in New York State (NYS), with a prevalence of 8.7% in 2014 and increased to 10.0% with the highest-burden in New York City in 2015 (New York State Asthma Dashboard, 2020). Approximately 400,000 people died from asthma worldwide in 2015 and the financial burden on patients with asthma in U.S. increased to US$ 3636 per person-year (Soriano et al., 2017; Ehteshami-Afshar et al., 2016). A similar burden was observed for allergies, which usually shared environmental causes and symptoms with asthma (Allergies and asthma, 2019). From the 2018 NHIS, the incidences of respiratory allergies and skin allergies among children were 9.3% and 12.2%, respectively. In the U.S., allergies are the 6th leading cause of chronic disease. The annual cost of allergies exceeded $18 billion, and more than 50 million Americans have allergies issues every year (Allergy Facts, 2020).

Numerous studies have been conducted to investigate the potential predictors for allergy-related and asthma symptoms. It has been suggested that kitchen ventilation, mold, dust reservoirs, indoor emission of volatile organic compounds (VOCs), and indoor dust particles are associated with elevated risks of allergies and asthma (Toskala and Kennedy, 2015; Burbank et al., 2017). However, several research gaps should be acknowledged in existing studies: 1) Although school is an important place of exposure where children spend a lot of time, fewer prior studies have evaluated the risk factors of allergy-related and asthma symptoms in school settings. In particular, the independent and joint health effects from home and school environmental exposures have been less examined, especially while controlling for all other risk factors and/or confounders; 2) Most previous findings are based on traditional statistical methods such as clustering approaches and logistic or linear regression, which suffer from collinearity problems. When multiple highly correlated factors and outliers are included in the same model, it leads to non-convergence and the statistical analysis becomes impossible. On the other hand, Random Forest, a machine-learning method, can identify top contributors, handle large numbers of variables simultaneously in one model, as well as deal with collinearity problems and outliers; and 3) Few past studies have comprehensively evaluated potential environmental risk factors, including those self-reported or objectively measured in both indoor and outdoor settings (Svendsen et al., 2018; Gupta et al., 2018). Finally, little is known about what combination of these risk factors poses the highest risk for children's health.

To fill the knowledge gaps in this study, we aimed to: 1) assess the association of numerous home and school environment exposures with asthma and allergy symptoms; 2) identify the most important contributors or predictors while controlling all other variables using the sophisticated machine-learning methods instead of traditional models; and 3) evaluate the most important combinations of the risk factors.

2. Methods

2.1. Study participant enrollment and data collection

We conducted the Student Health and Physical Environment (SHAPE) study between August 2017 and May 2019 except for summer vacation using a cross-sectional design in 10 public schools selected from the NYS Capital District Region. The study period was selected because the students were attending school during these periods. The public schools from the NYS Capital District Region were randomly selected using stratified sampling method based on sociodemographic strata at the school level to represent different race/ethnicity proportions, free lunch rates, and urbanicity proportions in NYS public schools. Once school district approval was ascertained from the Superintendent, each volunteer school Principal was asked to randomly recruit two teachers and eight or more students from each participating class to participate in the SHAPE project.

Participation in this study consisted of: 1) completing a questionnaire regarding demographics, health status including asthma and allergy-related symptoms, environmental exposures at home, and in the school in the past school year (completed by parents for primary school students); 2) conducting quantitative measurements through personal monitoring for 48 h to assess indoor air quality (CO, CO_2, $PM_{2.5}$, and VOCs) in fall and spring semesters respectively; and 3) completing an activity and respiratory symptom diary during the week.

We pilot tested the comprehension and literacy level of the questionnaire among a small sample of parents and teachers before conducting the survey. To promote participation and explain the questionnaire, the Principal Investigator (PI) and project staff held a special information session to explain the study procedures and questionnaire. Two or more staff went to the parent conference and the special session classroom teachers were arranged to interpret and respond to any questions the parents had. Several parents also communicated with the PI to raise their concerns or questions.

In terms of quantitative air pollution monitoring, we took real-time air pollution measurements for 48-h in 10 public schools and used personal air monitors to measure exposure among 85 students, due to limited resources. $PM_{2.5}$ was measured using real-time particulate matter AirBeam air samplers (Singer and Delp, 2018). CO_2 was measured using a Lascar EL-USB-CO easylog monitor and total VOC concentrations (TVOC) were measured using a DataRam. With these personal monitors, we collected hourly concentrations of $PM_{2.5}$, CO, CO_2, and TVOCs. As this study was funded by the US Environmental Protection Agency (EPA) STAR grant, we had already developed a 60-page Standard Operating Procedure and a 30-page quality assurance/quality control (QA/QC) plan according to EPA guidelines before the study began. The portable personal air monitors were assessed, calibrated, and validated using the laboratory control sample following the detailed QA/QC before each sampling trip. Only the devices that passed the calibration and validation steps were used.

2.2. Definitions and data source for allergy and asthma symptoms

In this study, information regarding asthma and allergy-related symptoms was obtained from household questionnaires and respiratory symptom diaries. We defined asthma symptoms as reporting any of the 15 questions from the Study of Asthma and Allergies in Childhood (ISAAC) questionnaire (Q1, Q1.1–1.6, Q2-Q9 of the Supplement 1). Since there is no gold standard definition of asthma, especially among children, we used the ISAAC questionnaire in this study, which has been validated and used widely in various epidemiologic studies around the world (Asher et al., 2006; Asher et al., 1995) to collect asthma and allergy-related symptom data. Apart from the asthma definition, we used typical allergy symptoms to define 'allergy-related symptoms' in this study including congestion, sneezing, eye irritation, skin irritation, and sinus problems. If any of those criteria were met, the student was considered to have allergy-related symptoms. The ISAAC questionnaire has been found to maximize the value of epidemiological research regarding asthma and allergy-related diseases by establishing a standardized methodology (Asher et al., 2006; Asher et al., 1995). Similar to asthma, the students who reported having allergy-related symptoms were double-checked via their allergy-related symptom diaries.

2.3. Sources and definition of exposures

Information regarding 82 potential risk factors or exposures for asthma or allergy-related symptoms was obtained from the questionnaire. As listed in the questionnaire in Supplement 1, there were 53 questions/sub-questions (Q10-Q36) regarding home exposures, including general home conditions, source/proxy for home-related allergens, source/proxy for home-related VOCs, and source/proxy for home-related ventilation. The majority of the home questions provided different alternative answers and 10 questions were yes/no questions. Most of these questions were converted to dummy variables to target specific exposures, leaving very few questions with multiple alternatives. We only merged a few variables with several alternatives. The recall period was 12 months prior to the survey, but some of the more general questions were not time-specific. In addition, the questions related to the school environment were actually answered by the children attending the schools. The parents were required to ask their children to ascertain the answers to these questions: for example, "How does your child feel about the indoor air quality in his or her school building?" and "How does your child feel about the outdoor air quality in his or her school building?", etc. Other school environmental factors, including the level of $PM_{2.5}$, CO, CO_2, and TVOCs, were obtained from the personal monitoring data.

These 84 variables included: *socioeconomic status (SES) variables*, *family history of diseases*, and *home and school environmental exposures*, and they were categorized according to the specific definitions as described in Table 1 and Supplement 2. Furthermore, continuous air pollution measurement variables were dichotomized using the 75th percentile of the distribution as a threshold for exposure. Only the measurements between 9 am and 3 pm, suggested by the activity pattern form filled out by the students, were used in this study to represent the school exposure.

2.4. Statistical analysis

We first calculated the percentage distribution for each variable. Univariate logistic regression as a single exposure model was used to calculate the prevalence ratio (PR) for each variable based on the original dataset. Therefore, the PR and percentage distribution were just for the initial description.

Within the dataset, the percentage of total missing values was 2.93%. We used the multivariate feature imputation method to fill in missing values. To assess the robustness of the multivariate feature imputation, we carried out a sensitivity analysis based on records with complete information only. We added 50% missing values, filled the missing values with 5 methods, including zero imputation, one imputation, median imputation, mode imputation, and multivariate feature imputation, and estimated the AUC score using the random forest model for each method and the records with complete information only (Supplement 6).

Table 1
Asthma and allergy symptoms in relation to SES characteristics of school-aged children in NYS, 2017–2018.

Variables	Class	Asthma symptoms		Allergy symptoms	
		N (%)[a]	PR[b] (95%CI[c])	N (%)	PR (95%CI)
Family SES					
Race	Other	9(40.91%)	0.79(0.29,2.11)	7(31.82%)	0.50(0.17,1.36)
	White	28(46.67%)	Ref (1.00)	29(48.33%)	Ref (1.00)
Hispanics	Yes	3(30.00%)	0.69(0.19,2.58)	1(10.00%)	0.46(0.12,1.80)
	No	24(50.00%)	Ref (1.00)	23(47.92%)	Ref (1.00)
Sex	Female	28(52.83%)	2.20(0.87,5.61)	24(45.28%)	1.43(0.57,3.58)
	Male	10(33.33%)	Ref (1.00)	11(36.67%)	Ref (1.00)
Insurance	Public	19(53.35%)	1.61(0.67,3.97)	14(37.84%)	0.64(0.26,1.55)
	Private	17(39.53%)	Ref (1.00)	21(48.84%)	Ref (1.00)
Maternal job status	Employed	29(45.31%)	1.27(0.43,3.75)	29(45.31%)	1.54(0.50,4.76)
	Unemployed	7(41.18%)	Ref (1.00)	6(35.29%)	Ref (1.00)
Parental job status	Employed	28(42.42%)	0.79(0.24,2.55)	31(46.97%)	3.42(0.67,17.47)
	Unemployed	5(45.45%)	Ref (1.00)	2(18.18%)	Ref (1.00)
Parental education	>High school	21(52.5%)	2.04(0.82,5.10)	24(60.00%)	4.79(1.66,13.80)
	≤High school	13(34.21%)	Ref (1.00)	8(21.05%)	Ref (1.00)
Maternal education	>High school	22(50.00%)	1.60(0.66,3.86)	25(56.82%)	3.54(1.37,9.18)
	≤High school	14(38.89%)	Ref (1.00)	9(25.00%)	Ref (1.00)
Age	–	10.27 ± 1.14	0.94(0.64,1.37)	10.27 ± 1.14	0.73(0.48,1.09)
Grade	≥5th grade	25(41.67%)	0.62(0.24,1.60)	23(38.33%)	0.54(0.21,1.40)
	<5th Grade	13(54.17%)	Ref (1.00)	13(54.17%)	Ref (1.00)
Health status					
Family rhinitis history	Yes	17(68.00%)	4.42(1.59,12.31)	15(60.00%)	2.87(1.08,7.58)
	No	15(30.00%)	Ref (1.00)	15(30.00%)	Ref (1.00)
Family asthma history	Yes	15(53.57%)	1.91(0.74,4.93)	13(46.43%)	1.50(0.58,3.87)
	No	17(36.17%)	Ref (1.00)	17(36.17%)	Ref (1.00)

[a] % is represent the percentage of children with asthma/allergy related symptoms in a certain group.
[b] PR is for prevalence ratio.
[c] CI is for confidence interval.

Table 2
Exposure characteristics of children aged 8–14 in home and school, 2017–2018, NYS.

Variables	Class	Asthma symptoms		Allergy symptoms	
		N (%)[a]	PR[b] (95%CI[c])	N (%)	PR (95%CI)
General home conditions					
Clean bedroom	>1 time/wk[d]	18(40.91%)	0.17(0.01,1.29)	16(36.36%)	0.38(0.05,2.53)
	1–3 times/mon	16(45.71%)	0.21(0.01,1.60)	17(48.57%)	0.63(0.08,2.53)
	<1 time/wk	4(80.00%)	Ref (1.00)	3(60.00%)	Ref (1.00)
Carpet in house	Yes	26(42.62%)	0.87(0.32,2.32)	30(49.18%)	3.13(1.03,9.53)
	No	10(45.45%)	Ref (1.00)	5(22.73%)	Ref (1.00)
Bedroom carpet	Yes	26(53.06%)	2.26(0.93,5.51)	27(55.10%)	3.68(1.44,9.44)
	No	12(33.33%)	Ref (1.00)	9(25.00%)	Ref (1.00)
Bedroom parquet flooring	Yes	11(37.93%)	0.66(0.26,1.64)	8(27.59%)	0.38(0.15,1.00)
	No	27(48.21%)	Ref (1.00)	28(50.00%)	Ref (1.00)
Tile flooring	Yes	4(36.36%)	0.67(0.18,2.49)	3(27.27%)	0.47(0.11,1.90)
	No	34(45.95%)	Ref (1.00)	33(44.59%)	Ref (1.00)
Home type	Multiple families	15(40.54%)	0.74(0.31,1.76)	15(40.54%)	0.88(0.36,2.09)
	Single family	23(47.92%)	Ref (1.00)	21(43.75%)	Ref (1.00)
Home location	Near busy roads	19(39.58%)	0.70(0.28,1.71)	18(37.50%)	0.72(0.29,1.78)
	On a quiet street	16(48.48%)	Ref (1.00)	15(45.45%)	Ref (1.00)
House year	1900–1979	13(43.33%)	1.53(0.26,12.28)	12(40.00%)	0.67(0.11,4.13)
	>1980	16(57.14%)	2.67(0.44,21.63)	16(57.14%)	1.33(0.21,8.36)
	Don't know	7(35.00%)	1.08(0.16,9.20)	5(25.00%)	0.33(0.05,2.30)
	<1900	2(33.33%)	Ref (1.00)	3(50.00%)	Ref (1.00)
Time of residence	>10 Years	10(37.04%)	0.63(0.25,1.61)	14(51.85%)	1.76(0.70,4.43)
	<10 years	28(48.28%)	Ref (1.00)	22(37.93%)	Ref (1.00)
No. of people in house	>4 people	31(44.93%)	0.92(0.30,2.83)	30(43.48%)	1.14(0.37,3.56)
	<4 people	7(46.67%)	Ref (1.00)	6(40.00%)	Ref (1.00)
No. of rooms	>5 rooms	29(46.77%)	1.65(0.58,4.67)	26(41.94%)	0.94(0.35,2.54)
	<5 rooms	7(33.33%)	Ref (1.00)	9(42.86%)	Ref (1.00)
Bed change	>10 days	9(42.86%)	0.88(0.33,2.39)	10(47.62%)	1.49(0.55,4.01)
	<10 days	28(45.9%)	Ref (1.00)	24(39.34%)	Ref (1.00)
Garage	Yes	16(53.33%)	1.71(0.70,4.21)	13(43.33%)	1.06(0.43,2.62)
	No	22(40.00%)	Ref (1.00)	23(41.82%)	Ref (1.00)
Mattress	New	34(43.59%)	0.58(0.12,2.77)	31(39.74%)	0.26(0.05,1.45)
	Used	4(57.14%)	Ref (1.00)	5(71.43%)	Ref (1.00)
ETS exposure	Yes	6(85.71%)	6.28(0.81,48.40)	5(71.43%)	2.35(0.49,11.19)
	No	31(41.33%)	Ref (1.00)	31(41.33%)	Ref (1.00)
ETS trigger symptoms	Yes	7(100.00%)	–	5(71.43%)	3.79(0.69,20.78)
	No	31(39.74%)		31(39.74%)	Ref (1.00)
Water leakage	Yes	9(60.00%)	1.87(0.61,5.76)	9(60.00%)	2.11(0.68,6.50)
	No	29(42.03%)	Ref (1.00)	27(39.13%)	Ref (1.00)
Visible mold	Yes	6(54.55%)	1.48(0.41,5.30)	5(45.45%)	1.10(0.31,3.92)
	No	32(43.84%)	Ref (1.00)	31(42.47%)	Ref (1.00)
Smell of mold	Yes	1(50.00%)	0.93(0.06,14.92)	1(50.00%)	1.03(0.06,16.47)
	No	37(45.68%)	Ref (1.00)	35(43.21%)	Ref (1.00)
Source/proxy for home-related allergens					
Pets	Yes	24(45.28%)	1.14(0.47,2.80)	24(45.28%)	1.31(0.53,3.24)
	No	13(41.94%)	Ref (1.00)	12(38.71%)	Ref (1.00)
Spiders	Yes	20(47.62%)	1.26(0.54,2.97)	22(52.38%)	2.28(0.95,5.49)
	No	18(41.86%)	Ref (1.00)	14(32.56%)	Ref (1.00)
Mice	Yes	9(50.00%)	1.31(0.46,3.72)	11(61.11%)	2.64(0.91,7.69)
	No	29(43.28%)	Ref (1.00)	25(37.31%)	Ref (1.00)
Cockroaches	Yes	1(33.33%)	0.61(0.05,6.97)	2(66.67%)	2.82(0.25,32.41)
	No	37(45.12%)	Ref (1.00)	34(41.46%)	Ref (1.00)
Dog trigger	Yes	1(33.33%)	0.61(0.05,6.97)	1(33.33%)	0.67(0.06,7.70)
	No	37(45.12%)	Ref (1.00)	35(42.68%)	Ref (1.00)
Cat trigger	Yes	3(50.00%)	1.257(0.24,6.62)	5(83.33%)	7.74(0.86,69.46)
	No	35(44.3%)	Ref (1.00)	31(39.24%)	Ref (1.00)
Pollen trigger	Yes	20(60.61%)	2.91(1.18,7.16)	23(69.70%)	6.90(2.61,18.24)
	No	18(34.62%)	Ref (1.00)	13(25.00%)	Ref (1.00)
Source/proxy for home-related VOCs					
Ammonia clean product	Yes	0(0.00%)	–	1(25.00%)	0.44(0.04,4.39)
	No	38(46.91%)		35(43.21%)	Ref (1.00)
Bleach clean product	Yes	19(43.18%)	0.88(0.37,2.07)	17(38.64%)	0.73(0.31,1.73)
	No	19(46.34%)	Ref (1.00)	19(46.34%)	Ref (1.00)
Spot remover	Yes	12(48.00%)	1.21(0.47,3.08)	12(48.00%)	1.39(0.54,3.54)
	No	26(43.33%)	Ref (1.00)	24(40.00%)	Ref (1.00)
Spray clean product	Yes	30(41.67%)	0.45(0.13,1.50)	31(43.06%)	1.21(0.36,4.06)
	No	8(61.54%)	Ref (1.00)	5(38.46%)	Ref (1.00)
Polish clean product	Yes	6(24.00%)	0.28(0.10,0.79)	10(40.00%)	0.87(0.34,2.25)
	No	32(53.33%)	Ref (1.00)	26(43.33%)	Ref (1.00)
Bathroom deodorizers	Yes	20(37.04%)	0.36(0.13,0.99)	25(46.30%)	1.35(0.50,3.62)
	No	16(64.00%)	Ref (1.00)	10(40.00%)	Ref (1.00)
Air fresheners	Yes	31(46.97%)	1.30(0.44,3.84)	30(45.45%)	1.57(0.52,4.74)
	No	7(41.18%)	Ref (1.00)	6(35.29%)	Ref (1.00)

Table 2 (*continued*)

Variables	Class	Asthma symptoms		Allergy symptoms	
		N (%)[a]	PR[b] (95%CI[c])	N (%)	PR (95%CI)
Incense candles	Yes	31(48.44%)	1.91(0.62,5.87)	29(45.31%)	1.45(0.49,4.30)
	No	5(29.41%)	Ref (1.00)	6(35.29%)	Ref (1.00)
Oil based paints	Yes	2(28.57%)	0.43(0.06,2.92)	1(14.29%)	0.19(0.03,1.33)
	No	31(46.97%)	Ref (1.00)	33(50%)	Ref (1.00)
Solvents	Yes	10(52.63%)	1.31(0.46,3.69)	11(57.89%)	1.53(0.57,4.10)
	No	25(44.64%)	Ref (1.00)	24(42.86%)	Ref (1.00)
Pesticide use	Yes	7(46.67%)	1.01(0.33,3.08)	10(66.67%)	2.99(0.93,9.60)
	No	31(44.93%)	Ref (1.00)	26(37.68%)	Ref (1.00)
Interior decoration	Yes	17(48.57%)	1.28(0.54,3.06)	14(40.00%)	0.87(0.36,2.10)
	No	21(42.86%)	Ref (1.00)	21(42.86%)	Ref (1.00)
Water resistant paint	Yes	3(37.50%)	0.72(0.16,3.23)	1(12.50%)	0.17(0.02,1.46)
	No	35(45.45%)	Ref (1.00)	35(45.45%)	Ref (1.00)
Water soluble paint	Yes	6(40.00%)	0.79(0.25,2.46)	5(33.33%)	0.63(0.20,2.03)
	No	32(45.71%)	Ref (1.00)	31(44.29%)	Ref (1.00)
New furniture	Yes	10(45.45%)	1.04(0.39,2.76)	9(40.91%)	0.92(0.35,2.47)
	No	28(44.44%)	Ref (1.00)	27(42.86%)	Ref (1.00)
Interior decoration other	Yes	5(55.56%)	1.63(0.41,6.55)	4(44.44%)	1.10(0.27,4.42)
	No	33(43.42%)	Ref (1.00)	32(42.11%)	Ref (1.00)
Source/proxy for home-related ventilation					
Major heating (Water radiator)	Yes	3(17.65%)	0.20(0.05,0.77)	4(23.53%)	0.35(0.10,1.17)
	No	35(51.47%)	Ref (1.00)	32(47.06%)	Ref (1.00)
Major heating (Forced Air)	Yes	20(48.78%)	1.38(0.58,3.24)	22(53.66%)	2.48(1.03,6.00)
	No	18(40.91%)	Ref (1.00)	14(31.82%)	Ref (1.00)
Major heating (Vented)	Yes	1(33.33%)	0.61(0.05,6.97)	2(66.67%)	2.82(0.25,32.41)
	No	37(45.12%)	Ref (1.00)	34(41.46%)	Ref (1.00)
Major heating (Baseboard)	Yes	5(45.45%)	1.04(0.29,3.70)	2(18.18%)	0.26(0.05,1.29)
	No	33(44.59%)	Ref (1.00)	34(45.95%)	Ref (1.00)
Major heating (Floor wall)	Yes	3(37.50%)	0.72(0.16,3.23)	2(25.00%)	0.42(0.08,2.22)
	No	35(45.45%)	Ref (1.00)	34(44.16%)	Ref (1.00)
Major heating (Wood Coal)	Yes	4(100.00%)	–	2(50.00%)	1.38(0.19,10.31)
	No	34(41.98%)		34(41.98%)	Ref (1.00)
Secondary Heating (Floor wall)	Yes	1(50.00%)	1.24(0.08,20.56)	1(50.00%)	1.37(0.08,22.69)
	No	37(44.58%)	Ref (1.00)	35(42.17%)	Ref (1.00)
Secondary Heating (Vented)	Yes	4(40.00%)	0.80(0.21,3.08)	5(50.00%)	1.42(0.38,5.32)
	No	34(45.33%)	Ref (1.00)	31(41.33%)	Ref (1.00)
Secondary Heating (Unvented)	Yes	2(50.00%)	1.25(0.17,9.31)	2(50.00%)	1.38(0.19,10.31)
	No	36(44.44%)	Ref (1.00)	34(41.98%)	Ref (1.00)
Secondary Heating (Wood Coal)	Yes	0(0.00%)	–	1(33.33%)	0.67(0.06,7.70)
	No	38(46.34%)		35(42.68%)	Ref (1.00)
Secondary Heating (Baseboard)	Yes	4(57.14%)	1.73(0.36,8.23)	5(71.43%)	3.79(0.69,20.78)
	No	34(43.59%)	Ref (1.00)	31(39.74%)	Ref (1.00)
Air Conditioner	Yes	34(42.50%)	0.19(0.02,1.73)	34(42.5%)	1.11(0.18,7.00)
	No	4(80.00%)	Ref (1.00)	2(40.00%)	Ref (1.00)
AC type	Central AC	15(46.88%)	1.25(0.52,3.03)	18(56.25%)	2.08(0.83,5.20)
	Window units	20(41.67%)	Ref (1.00)	17(35.42%)	Ref (1.00)
Air Cleaner	Yes	6(60.00%)	2.02(0.53,7.74)	5(50.00%)	1.42(0.38,5.32)
	No	32(42.67%)	Ref (1.00)	31(41.33%)	Ref (1.00)
Humidifier	Yes	15(50.00%)	1.39(0.57,3.40)	14(46.67%)	1.31(0.54,3.22)
	No	23(41.82%)	Ref (1.00)	22(40.00%)	Ref (1.00)
Dehumidifier	Yes	11(52.38%)	1.51(0.56,4.06)	11(52.38%)	1.72(0.64,4.63)
	No	27(42.19%)	Ref (1.00)	25(39.06%)	Ref (1.00)
Stove Type	Gas	18(37.50%)	0.49(0.20,1.17)	18(37.5%)	0.61(0.25,1.46)
	Electric	20(55.56%)	Ref (1.00)	18(50.00%)	Ref (1.00)
Stove Fan	Yes	30(46.15%)	1.29(0.46,3.56)	26(40.00%)	0.67(0.24,1.83)
	No	8(40.00%)	Ref (1.00)	10(50.00%)	Ref (1.00)
Heating (Electricity)	Yes	9(36.00%)	0.60(0.23,1.57)	11(44.00%)	1.10(0.43,2.82)
	No	29(48.33%)	Ref (1.00)	25(41.67%)	Ref (1.00)
Heating (Natural Gas)	Yes	23(40.35%)	0.59(0.24,1.46)	22(38.6%)	0.63(0.25,1.57)
	No	15(53.57%)	Ref (1.00)	14(50.00%)	Ref (1.00)
Heating (Oil Fuel)	Yes	5(71.43%)	3.41(0.62,18.67)	3(42.86%)	1.02(0.21,4.88)
	No	33(42.31%)	Ref (1.00)	33(42.31%)	Ref (1.00)
Heating (Wood Fuel)	Yes	2(50.00%)	1.25(0.17,9.31)	2(50.00%)	1.38(0.19,10.31)
	No	36(44.44%)	Ref (1.00)	34(41.98%)	Ref (1.00)
Heating (Other Fuel)	Yes	3(100.00%)	–	2(66.67%)	2.82(0.25,32.41)
	No	35(42.68%)		34(41.46%)	Ref (1.00)
School characteristics					
$PM_{2.5}$	Yes	11(37.93%)	0.66(0.26,1.67)	9(31.03%)	0.46(0.18,1.19)
	No	26(48.15%)	Ref (1.00)	27(50.00%)	Ref (1.00)
CO	Yes	6(46.15%)	0.93(0.30,2.82)	7(53.85%)	1.82(0.58,5.66)
	No	30(44.78%)	Ref (1.00)	26(38.81%)	Ref (1.00)
TVOC	Yes	15(45.45%)	1.04(0.41,2.64)	16(48.48%)	1.70(0.69,4.23)
	No	19(45.24%)	Ref (1.00)	15(35.71%)	Ref (1.00)

(continued on next page)

Table 2 (*continued*)

Variables	Class	Asthma symptoms		Allergy symptoms	
		N (%)[a]	PR[b] (95%CI[c])	N (%)	PR (95%CI)
CO2	Yes	15(45.45%)	1.04(0.41,2.64)	16(48.48%)	1.70(0.69,4.23)
	No	19(45.24%)	Ref (1.00)	15(35.71%)	Ref (1.00)
Class noise	Noisy	7(30.43%)	0.47(0.16,1.30)	10(43.48%)	1.14(0.41,3.07)
	Fair	25(48.08%)	Ref (1.00)	21(40.38%)	Ref (1.00)
Air Quality Outside School	Good	24(41.38%)	0.71(0.24,2.06)	24(41.38%)	1.11(0.38,3.40)
	Fair	9(50.00%)	Ref (1.00)	7(38.89%)	Ref (1.00)
Air Quality School	Good	21(43.75%)	1.20(0.47,3.16)	21(43.75%)	1.40(0.54,3.74)
	Fair	11(39.29%)	Ref (1.00)	10(35.71%)	Ref (1.00)

[a] % is represent the percentage of children with asthma/allergy related symptoms in a certain group.
[b] PR is for prevalence ratio.
[c] CI is for confidence interval.
[d] wk. is for week.

After addressing the missing values, we adopted random forest model, one of the most established machine-learning methods, to screen important variables through a grid search with 10-fold cross-validation. Since these random forest models will use 70% of all variables and data to build every single tree, we can obtain more robust results through these models. The grid search with 10-fold cross-validation was used to identify the optimal combination of parameters including tree depth, minimum sample for splits, and the number of trees for the random forest model. By adding a random term, variables with greater relative importance than the random term were identified and selected. The direction of relative importance for each variable was determined from the increase or decrease of their partial dependence.

Although the random forest model is more robust, it is not easy to interpret the relationships among selected variables because each tree built in a random forest only considers 70% of variables and data. Instead, we used decision tree models to visualize and interpret the relationships, which is a unique advantage of the decision tree model. In this case, we built decision tree models by only adding variables with greater relative importance than the random term from the random forest models. To ensure the interpretability of the decision tree model, we randomly split the dataset into a training dataset and testing dataset according to a ratio of 8 to 2 for each outcome. We again used a grid search together with 10-fold cross-validation to tune the model's parameters, such as tree depth and a minimum sample of splits, as well as to overcome the over-fitting problem. Finally, the performance of decision tree models was assessed with a test dataset as external validation. The whole analysis was conducted by using "Scikit-learn" in Python and "APML" in R (Deng et al., 2020; Pedregosa et al., 2011).

3. Results

3.1. Descriptive results and crude association

The information from those who completed all components, including questionnaires (completed by parents), symptom diaries, activity diaries, and personal air pollution monitoring, are used in this paper. Among the participants aged 8–14, 38 children reported having asthma symptoms and 36 reported having allergy-related symptoms. As shown in Table 1, family history of rhinitis was significantly associated with asthma and allergy-related symptoms (Asthma: PR = 4.42, 95%CI = 1.59–12.31; Allergy: PR = 2.87, 95%CI = 1.08–7.58). Children with more educated fathers or mothers were more likely to report allergy-related symptoms (Father: PR = 4.79; 95% CI = 1.66–13.80; Mother: PR = 3.54; 95%CI = 1.37–9.18). No other factors were significantly associated with allergy-related or asthma symptoms based on PRs.

Table 2 shows the PRs of environmental exposures with respect to five aspects: general home conditions, source/proxy of home-related allergens, source/proxy for home-related VOCs, source/proxy for home-related ventilation, and school characteristics. For general home conditions, pollen trigger was found as a risk factor for both asthma and

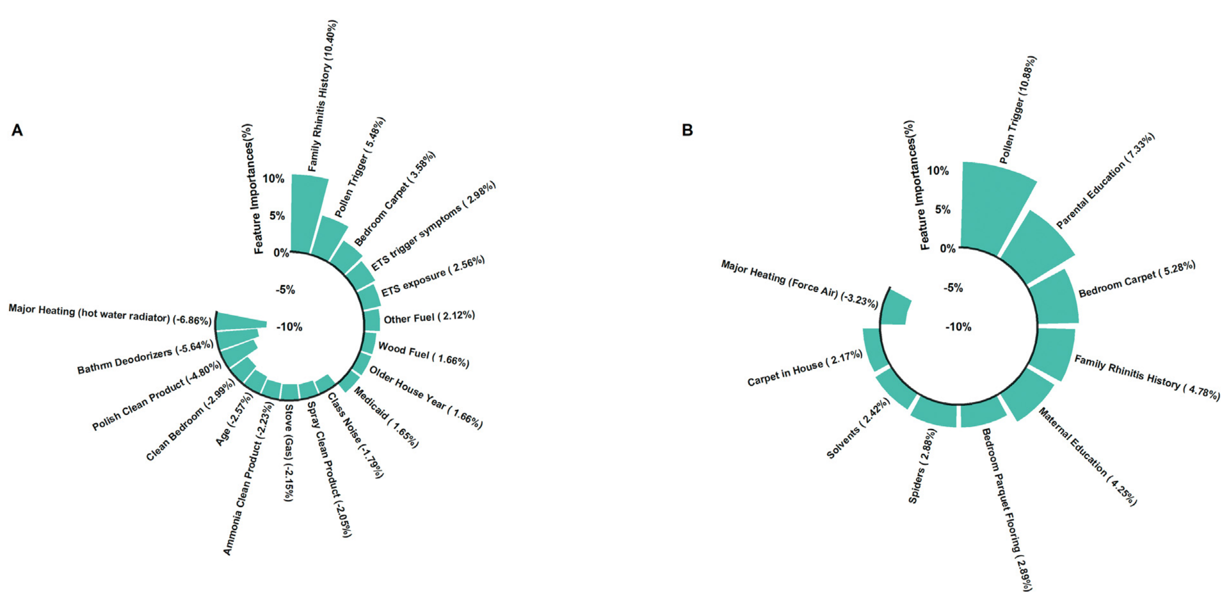


Fig. 1. Selected variables' relative importance based on random forest models. A is for asthma symptoms. B is for allergy symptoms. Values for each variable were calculated based on relative importance. Only variables with relative importance greater than random term were selected.

allergy-related symptoms (Asthma: PR = 2.91, 95%CI = 1.18–7.16; Allergy: PR = 6.90, 95%CI = 2.61–18.24). More specifically, carpeting in the house (PR = 3.13, 95%CI = 1.03–9.53) and bedroom (PR = 3.68, 95%CI = 1.44–9.44) were similarly associated with allergy-related symptoms. In terms of source/proxy of home-related VOCs and ventilation, using polish cleaning products (PR = 0.28, 95%CI = 0.10–0.79), bathroom deodorizers (PR = 0.36; 95%CI = 0.13–0.99), and using a hot water radiator for primary heating (PR = 0.20, 95%CI = 0.05–0.77) were negatively associated with asthma symptoms, while primary heating with forced air was positively associated with allergy-related symptoms(PR = 2.48, 95%CI = 1.03–6.00). Interestingly, no school characteristics were significantly associated with asthma or allergy-related symptoms.

3.2. Dominant contributing factors

As presented in Fig. 1, we applied the random forest model to select factors with greater contributions than random terms ($Random_{asthma}$ = 1.53%, $Random_{allergy}$ = 2.11%) based on the whole dataset. In Fig. 1A, among positive contributions, family history of rhinitis (10.40%) contributed most to asthma symptoms, followed by pollen trigger (5.48%), bedroom carpet (3.58%), ETS trigger symptoms (2.98%), ETS

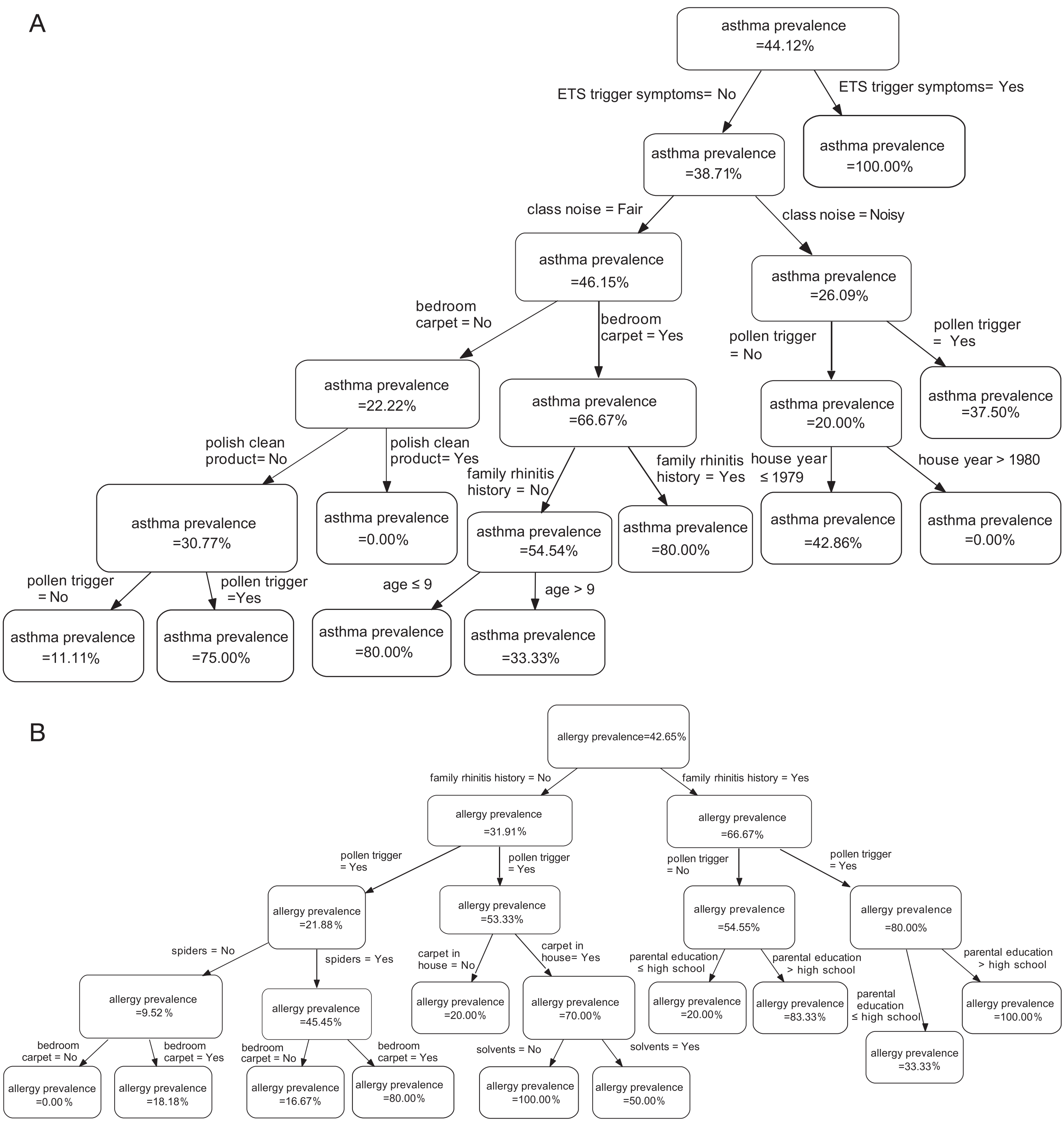


Fig. 2. Relationships among selected variables from decision tree models. A is for asthma symptoms. B is for allergy symptoms.

exposure (2.56%), heating using other fuel types (2.12%), heating using wood (1.66%), older housing (1.66%), and Medicaid insurance (1.65%). Conversely, primary heating with hot water radiator was negatively (−6.86%) associated with asthma symptoms, followed by bathroom deodorizers (−5.64%), using polish cleaning products (−4.8%), cleaning bedrooms frequently (−2.99%), older age (−2.57%), using ammonia-based cleaning products (−2.23%), using a gas stove (−2.15%), spray cleaning products (−2.05%), and class noise (−1.79%).

Similarly, factors associated with allergy-related symptoms are presented in Fig. 1B. Pollen trigger had the greatest contribution (10.88%), followed by higher paternal education (7.33%), bedroom carpet (5.28%), family history of rhinitis (4.78%), higher maternal education (4.25%), bedroom parquet flooring (2.89%), spiders (2.88%), solvents (2.42%), and carpet in house (2.17%). Surprisingly, only primary heating using forced air (−3.23%) negatively contributed to allergy-related symptoms.

3.3. Interaction or relationships among selected variables

Fig. 2 illustrates the interpretation of relationships or interactions among selected factors. In Fig. 2A, the ETS trigger symptoms acted as a leading factor, participating in all paths to asthma symptoms. The prevalence of asthma symptoms decreased or increased following each dsicrete path, which showed a good ability to differentiate between the selected factors. For example, the asthma prevalence of the original group was 44.12%. In the high-risk subgroup with ETS trigger symptoms, the prevalence increased to 100%. Another potential high-risk group were those without ETS trigger symptoms who reported exposure to bedroom carpet and a family history of rhinitis, with their prevalence of asthma increasing from 44.12% to 80.00%. The same effect occurred for allergy-related symptoms, as shown in Fig. 2B. Family history of rhinitis and pollen trigger played important roles in all paths to allergy-related symptoms. Again, the capacity to differentiate performed very well. A potential high-risk group included those with a family history of rhinitis, pollen trigger, and higher paternal education. The prevalence of allergy-related symptoms in this group rose from 42.65% to 100.00%.

3.4. Model performance

Fig. 3 shows the ROC of the decision tree model for asthma and allergy-related symptoms. For asthma symptoms, the AUC of this decision tree model was 0.797 for internal validation, 0.753 for 10-fold cross-validation, and 0.812 for external validation. For allergy-related symptoms, the AUC was 0.819 for internal validation, 0.790 for 10-fold cross-validation, and 0.829 for external validation. When the three types of validation were close in value to each other, and all above 0.7, we assessed that these two models overcame over-fitting problems and performed well.

4. Discussion

4.1. Top contributing factors for asthma symptoms

4.1.1. Home environment

In this study, we found that bedroom carpet, older building, using wood fuel, and other fuel (such as wood pellets and kerosene) are home-related risk factors for asthma symptoms. Potential interpretations include increased airborne allergens such as mite and dust for potential associations between asthma and carpeted floors in the bedroom, and elevated exposure to fuel combustion smoke for associations between asthma and cooking indoors with wood and coal (Barry et al., 2010; Vicendese et al., 2015). Interestingly, an India National Sample Survey reported that rural areas highly dependent on solid fuels (80.5%) had a higher risk of bronchial asthma compared to urban areas (Faizan and Thakur, 2019). According to Solis-Soto's report, around 19% of school-aged children with asthma in Bolivia live in homes primarily fueled by wood or coal. Gas stoves are the primary source of indoor nitrogen dioxide, a common indoor pollutant. Studies show that people who cook with gas are more likely to have wheezing, breathlessness, asthma attacks, and hay fever compared to those who cook with other methods (Casas et al., 2012). Moreover, our results from the decision tree suggest that the risk of asthma increases in older buildings, which is consistent with some studies in other countries (Wang et al., 2019). This observation may be explained by increased numbers of dust mites and molds in these buildings combined with inadequate heating, ventilation, and increased concentrations of biological and chemical contaminants in older houses as compared to newer ones. Levels of bacterial toxins called endotoxins in house dust have been shown to be directly related to asthma symptoms (Esty and Phipatanakul, 2018). In addition, the use of lead-based paint in older buildings may be a potential reason for this observation in the US.

On the other hand, primary heating using hot water radiator, cleaning bedrooms, private insurance, clean products using spray, ammonia and polish, stoves using gas for cooking, class noise, and use of bathroom deodorizers were found to be “protective” factors for asthma symptoms. Compared with other types of heating equipment, we observed that primary heating using a hot water radiator may be negatively associated (i.e. protective) with asthma symptoms. Primary heating using hot water radiators may reduce exposure to hot dry air which may trigger asthma symptoms (Tabka et al., 1988). Another reason could be that using hot water radiators does not produce multiple

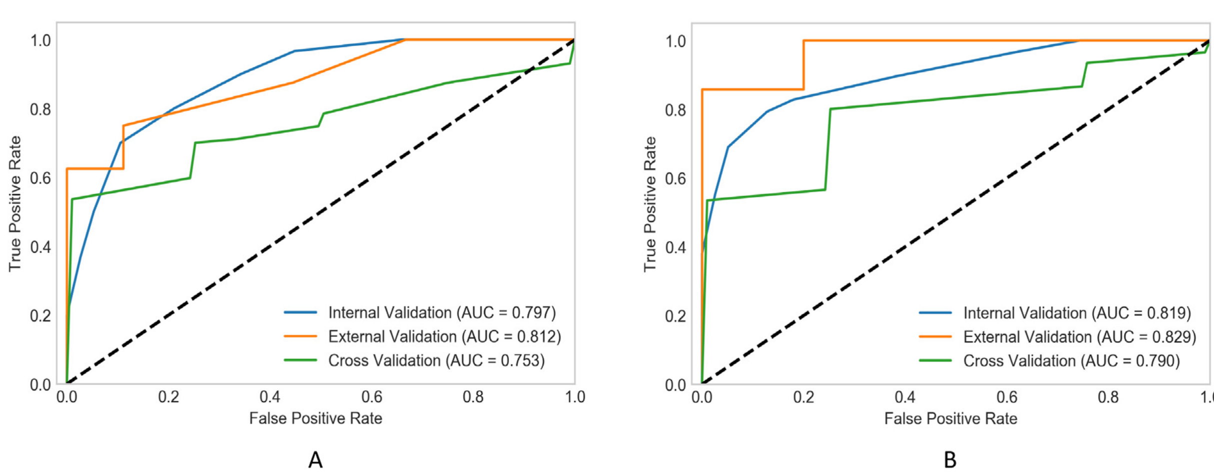


Fig. 3. ROC curves of decision tree for asthma symptom and allergy symptom. A is for asthma symptoms. B is for allergy symptoms.

air pollutants comparing with heating by coal stoves. Cleaning bedrooms was suggested to be a potential protective factor for asthma because it could reduce the amount of dust mites allergens (Morgan et al., 2004; Tsurikisawa et al., 2013). Interestingly, private insurance seemed to be negatively associated with asthma in children. As an alternative to private insurance, Medicaid is typically utilized by disabled and/or low-income individuals. This is important because poverty may be related to asthma or allergy-related symptoms. A previous study reported that public insurance was associated with higher asthma readmission odds (OR = 1.81, 95% CI = 1.46–2.24) (Nkoy et al., 2018). Another study reported that, compared to patients with private insurance, those with public insurance were more likely to visit the emergency department (ED) (35%, 49%, and 45%, respectively; $p < 0.001$) (Hasegawa et al., 2016). Interestingly, our results did show a "protective" effect associated with cleaning products and bathroom deodorizers, which seems different from some existing findings. We double-checked the usage of polish cleaning products among the children with doctor-diagnosed asthma. Among these children, 84% of families reported not using polish clean products. Similarly, 85% parents who had children with asthma symptoms from the ISSAC questionnaire also reported not using polish clean products. However, we found that previous findings discussing the hazardous effect of cleaning products were usually obtained among workers who are continuously exposed to these products in their routine work with high frequency and duration of exposures (Folletti et al., 2017). In contrast, the students with asthma and their parents in our study might avoid such exposures. A potential explanation, then, may be the protective behavior of these families, i.e., the parents avoiding using these cleaning products if their child(ren) had asthma symptoms. As for class noise, we previously found that it was a risk factor for asthma in children when using a sample size of 280 (Palumbo et al., 2018). Again, this conflict may be due to reporting bias from parents, since they could only assess school noise levels by asking their children. This misclassification problem could be minimized if we increased the sample size and adopted an alternative study design in the future.

4.1.2. Demographic characteristics and school environment

Family history of rhinitis, pollen trigger, ETS trigger symptoms, and ETS exposure were found to be risk factors for asthma. According to a genome-wide association study, asthma and allergy rhinitis are linked in different chromosomal regions. Allergy rhinitis is also an important risk factor for pediatric asthma in population-based studies (Chinnakkannan et al., 2017; Mastrorilli et al., 2016). A cross-sectional, multi-center survey consisting of 3217 children indicated that family history of atopy (asthma, allergy rhinitis, and eczema) is an important and significant risk factor for allergy rhinitis (Chinratanapisit et al., 2018). Furthermore, the results regarding pollen trigger, ETS trigger symptoms, and ETS exposure are also compatible with other studies (Solis-Soto et al., 2013; Linneberg et al., 2002). Surprisingly, even though we know the relationship has been well-established, a study reported that around 60% of participants in five European countries still experience asthma-related symptoms triggered by smoking (Gautier and Charpin, 2017). This means public awareness needs to be enhanced.

Notably, age was considered a protective factor for asthma. Approximately a third of children had asthma-related symptoms in their first three years of life, and by the age of six, most children stopped wheezing, especially in children with mild asthma. The severity of childhood symptoms, especially in children with mild asthma, tended to diminish or even disappear as they entered early puberty (van Aalderen, 2012).

However, we did not find any significant associations between asthma and air pollutants measured from the personal monitor during school hours.

4.2. Top contributing factors for allergy symptoms

4.2.1. Home environment

Our results regarding allergy-related symptoms suggest that home-related factors, such as bedroom carpet, bedroom parquet flooring, carpet in house, spiders, and using solvents, are positively associated with allergy-related symptoms. This finding could be related to the indoor environmental quality, since previous studies reported that indoor environmental quality could be affected by carpet, parquet flooring, spiders, and solvents, which in turn arouses public health concerns (Svendsen et al., 2018; Becher et al., 2018; Schlink et al., 2010). More interestingly, we found that carpeting in bedrooms had a greater association with allergy-related symptoms compared to the presence of carpeting elsewhere in the house. Although we don't know the exact reason for this, a possible explanation could be that children usually spend more time in their bedroom throughout the day and bedrooms are necessarily a smaller enclosed space, which likely contributes to the greater concentration of contaminants.

Furthermore, primary heating using forced air was negatively associated with allergy-related symptoms. Some studies also reported that primary heating using forced air didn't exacerbate allergy or asthma symptoms (Dales et al., 1992; Castro-Rodriguez et al., 2016). The exact explanation remains unknown. It could be that primary heating using forced air results in drier air, which leads to less mold. However, whether primary heating using forced air has a negative or protective effect on allergy-related symptoms is still unclear and needs further investigation.

4.2.2. Demographic characteristics & school environment

Similar to asthma symptoms, we also found associations between potential risk factors like pollen trigger, family history of rhinitis, and allergy-related symptoms in this study. It is also important to highlight that higher education of parents was "positively" associated with allergy-related symptoms in children. Interestingly, previous findings suggested that children of parents with a low level of education are at a higher risk of asthma, in part, because parents affected the frequency of their children's fruit, vegetable, and meat consumption (Lewis et al., 2017; Andrusaityte et al., 2017). Hence, the immunity and nutrition status of their children could be weakened. However, one such study reported that low parental education was associated with a decreased risk of inhalant allergy and itchy rash, and an increased prevalence of wheezing and nighttime dry cough in school children (Gehring et al., 2006). In this study, considering the influence of reporting bias, an alternative explanation is that parents with higher education may be more aware of allergies, may pay more attention to their children, and are therefore more prone to report allergy-related symptoms. Therefore, the causality of the association between allergy-related symptoms and education level of parents remains unclear and needs to be further studied. Nevertheless, we did not find significant relationships between allergy-related symptoms and the air pollutants measured through personal monitoring in-school.

4.3. Interpretations of variables' inter-relationship

By using a decision tree to examine relationships among different factors, the external validations of these two models indicated that they performed very well, even with our small sample size. This further confirms that our models can be applied to other populations to some extent. Our study suggests that ETS trigger symptoms and pollen exposure are the leading independent risk factors for asthma symptoms. More importantly, ETS trigger symptoms was involved in all pathways to asthma symptoms, and every student with ETS trigger symptoms reported having asthma symptoms, which suggested that ETS trigger symptoms was the most important risk factor for asthma symptoms. For allergy-related symptoms, family history of rhinitis and pollen exposure largely contributed to a higher prevalence of allergy-related symptoms among children. Interestingly, the findings from the decision tree analysis could describe the inter-relationship among the risk factors and demonstrate the joint effects of multiple factors, which could help us identify potential high-risk groups. For instance, we found that the prevalence of asthma symptoms significantly increased among students who were younger (<9 years old), had a family history of rhinitis, and had carpeting in their bedroom. For allergy-related symptoms, family allergy rhinitis and pollen trigger were the top two leading risk factors. These findings are consistent with previous studies (Solis-Soto et al., 2013; Linneberg et al., 2002; Gautier and Charpin, 2017). We also observed that potential high-risk children who were jointly exposed to pollen, solvents, and carpeting in their home greatly increased their risks of allergy-related symptoms, even without family history of allergy. In other words, the findings from the decision tree model provide a potential pathway describing how and which two or more risk factors jointly affect asthma or allergy-related symptoms. This, in turn, can help us identify high-risk or sensitive groups and guide environmental health agencies and schools to plan interventions to reduce the prevalence of asthma and allergy-related symptoms.

4.4. Overall model performance and advantages

This may be the first study to thoroughly evaluate the potential risk factors from home and school that influence asthma and allergies using machine-learning methods. The random forest model has many unique advantages. Compared to traditional models, it can rank the variables' importance and reduce the dimension of datasets based on the Gini coefficient (Hastie et al., 2009). In addition, using a random forest can effectively reduce bias and variance based on its random property, since it selects subgroups of the training dataset and variables (Hastie et al., 2009). With external validation, the results of such models are usually more robust and generalizable. Moreover, random forest models can address multiple highly correlated variables, which is a formidable problem for traditional models, and in real world settings. This is especially important in this area of research since environmental factors are usually correlated with one another. For example, in our study, carpeting in the house and in the bedroom were highly correlated with each other. However, the relative importance of carpeting in the bedroom was higher than in the house for both allergy-related and asthma symptoms.

A strength of the decision tree model is that it can easily visualize the relationships among important variables, and thereby promote a better understanding of these relationships. Usually, machine-learning models are criticized as a type of "black-box" because they do not provide an easily interpretable method to quantify the reflected relationship between exposures and health outcomes (Nori et al., 2019). Unlike other machine-learning models, tree-model-based algorithms, like random forest models, do provide visible ways for us to interpret results (Hastie et al., 2009; Karimi and Hamilton, 2011). However, random forest models will miss some variables because due to randomly selecting them, generating many different trees and resulting in interpretation difficulties. In this case, the decision tree has its own advantages. By building a single tree, it can supplement the weakness of the random forest model and help us better understand the relationships among selected variables.

4.5. Strengths and limitations

In terms of strengths, our study included 84 variables allowing us to provide a comprehensive understanding of their relationships. Also, by using machine-learning models, we were able to explore potential and highly correlated environmental variables. The questionnaires and air pollutant measurements compensated for each other by providing exposure information from different perspectives. The questionnaires thoroughly covered all potential indoor environmental factors at home and in school by using the standard ISAAC questionnaire. For air pollutant measurements, we adopted several devices, including a 48-hour personal air quality monitor, to provide real-time, comprehensive exposure measurement data at home, in school, on the bus, and during other activities.

Despite the strengths of our study, some limitations need to be acknowledged: 1) Recall/reporting bias is always a concern when using questionnaires in epidemiology studies. To address this concern, we used the following approaches to minimize potential recall biases: a) the ISAAC questionnaire is an internationally standardized and validated asthma/allergy questionnaire, which we used to define asthma and severe asthma in this study. We also used multiple, different asthma questions to confirm if the cases defined were correct; b) similar, yet different wording was used for some asthma and home exposure questions that allowed us to check for response consistency; c) our questionnaires provided multiple alternative answers by defining time in the past year, frequency for some variables (home cleaning), and providing scales from one-to-five to judge air quality and noise, which may reduce misreporting and/or recall bias by asking for detailed information instead of simple affirmative ("yes" or "no") answers; and d) the air pollution data measured using the personal monitors helped validate the self-reported perception of air quality in school. 2) This study may not be able to detect many new risk factors compared with existing studies, potentially due to the sample size limit. For the same reason, adding or subtracting a few subjects may influence the results. However, this problem has been greatly minimized because the random forest models utilized in this study used 70% of all variables and data to build every single tree. When building more trees, false positive and outliers issues are minimized. With hundreds, or even thousands of trees, we can generate more robust results compared with traditional methods. In addition, the 10-fold cross-validation used in this study is a good strategy to deal with this problem. 3) Potential misclassification of asthma and allergy may be a concern as we used ISAAC questionnaires to define asthma and allergy-related symptoms without clinical diagnosis for each case. However, we focused on "asthma or allergy-related symptoms" rather than clinical definitions in the current study. To verify the accuracy of our case definition, we checked the number and severity of symptoms of each case. We found that 74% of the children who were defined to have asthma symptoms in this study reported having more than two asthma-related symptoms. The other 26% of children reported having typical or severe asthma symptoms, such as daily activities being affected or limited, and/or sleep disruption. In addition, we used the respiratory symptom diaries to track the consistency of the symptoms reported in the past week via the questionnaires. In summary, all children who reported asthma symptoms either had more than two symptoms or severe asthma symptoms in the past year using the ISAAC questionnaire. Similarly, all students defined as having allergy-related symptoms had multiple symptoms and most had them recently. This validation suggests that both asthma and allergy-related symptoms defined in this study are relatively accurate and reliable. 4) Some confounders were not available in this study. Although sensitization to pollen and house dust mites may have played roles with regard to students' asthma and allergy-related symptoms, data regarding these factors was not available. In addition, data regarding the frequency of different exposures was, unfortunately, not available. However, this information will be collected in our future school projects.

5. Conclusion

The findings of this study suggest that pollen trigger, bedroom carpet, and family history of rhinitis are associated with increased prevalence of both asthma and allergy-related symptoms. Children exposed to ETS trigger symptoms, wood fuel, older home buildings, and Medicaid insurance had increased risks of asthma prevalence, while those having carpet in house, having parquet flooring in the bedroom, being exposed to spiders, or using solvents tended to be more prone to allergy-related symptoms. The risk of asthma was exaggerated by the joint effects of young age, carpet in bedroom, and family history of rhinitis. Meanwhile, multiple exposures including family allergy rhinitis, pollen trigger, carpet in both the house and bedroom, being exposed to spiders, and solvents could jointly increase allergy risk. Some factors including bathroom deodorizers and polish cleaning products were adversely associated with asthma symptoms possibly due to protective behaviors.

CRediT authorship contribution statement

Xinlei Deng: Conceptualization, Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **George Thurston:** Methodology, Investigation, Writing – review & editing, Project administration, Funding acquisition. **Wangjian Zhang:** Software, Writing – original draft, Writing – review & editing. **Ian Ryan:** Data curation, Writing – original draft, Writing – review & editing. **Connie Jiang:** Data curation, Writing – original draft. **Haider Khwaja:** Methodology, Investigation, Writing – review & editing. **Xiaobo Romeiko:** Validation, Investigation, Writing – review & editing. **Tia Marks:** Writing – original draft. **Bo Ye:** Writing – original draft. **Yanji Qu:** Writing – original draft. **Shao Lin:** Conceptualization, Methodology, Validation, Investigation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no competing interests regarding the publication of this manuscript.

Acknowledgments

This work was supported by EPA (Environmental Protection Agency) STAR grant (83563602). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders. The funders played no role in the study design, data collection, analysis and interpretation, in the writing of the report, or in the decision to submit this article for publication. We would also like to thank the participants, whose generous time and effort made this study possible.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.144746.

References

Allergies and asthma, 2019. Double trouble - Mayo Clinic. Accessed December 24. https://www.mayoclinic.org/diseases-conditions/asthma/in-depth/allergies-and-asthma/art-20047458.

Allergy Facts, 2020. ACAAI Public Website. Accessed September 2. https://acaai.org/news/facts-statistics/allergies.

Andrusaityte, S., Grazuleviciene, R., Petraviciene, I., 2017. Effect of diet and maternal education on allergies among preschool children: a case-control study. Environ. Res. 159, 374–380. https://doi.org/10.1016/j.envres.2017.08.020.

Asher, M.I., Keil, U., Anderson, H.R., et al., 1995. International study of asthma and allergies in childhood (ISAAC): rationale and methods. Eur. Respir. J. 8 (3), 483–491. https://doi.org/10.1183/09031936.95.08030483.

Asher, M.I., Montefort, S., Björkstén, B., et al., 2006. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC phases one and three repeat multicountry cross-sectional surveys. Lancet. 368 (9537), 733–743. https://doi.org/10.1016/S0140-6736(06)69283-0.

Asthma Surveillance Data, 2019. CDC. Accessed December 24. https://www.cdc.gov/asthma/asthmadata.htm.

Barry, A.C., Mannino, D.M., Hopenhayn, C., Bush, H., 2010. Exposure to indoor biomass fuel pollutants and asthma prevalence in southeastern Kentucky: results from the burden of lung disease (BOLD) study. J Asthma. 47 (7), 735–741. https://doi.org/10.3109/02770903.2010.485661.

Becher, R., Øvrevik, J., Schwarze, P.E., Nilsen, S., Hongslo, J.K., Bakke, J.V., 2018. Do carpets impair indoor air quality and cause adverse health outcomes: a review. Int J Environ Res Public Health. 15 (2). https://doi.org/10.3390/ijerph15020184.

Burbank, A.J., Sood, A.K., Kesic, M.J., Peden, D.B., Hernandez, M.L., 2017. Environmental determinants of allergy and asthma in early life. J. Allergy Clin. Immunol. 140 (1), 1–12. https://doi.org/10.1016/j.jaci.2017.05.010.

Casas, L., Tischer, C., Tiesler, C., et al., 2012. Association of gas cooking with children's respiratory health: results from GINIplus and LISAplus birth cohort studies. Indoor Air 22 (6), 476–482. https://doi.org/10.1111/j.1600-0668.2012.00784.x.

Castro-Rodriguez, J.A., Forno, E., Rodriguez-Martinez, C.E., Celedón, J.C., 2016. Risk and protective factors for childhood asthma: what is the evidence? J Allergy Clin Immunol Pract 4 (6), 1111–1122. https://doi.org/10.1016/j.jaip.2016.05.003.

Chinnakkannan, S.K., Singh, M., Das, R.R., Mathew, J.L., Saxena, A.K., 2017. Association of allergic rhinitis and sinusitis with childhood asthma. Indian Pediatr. 54 (1), 21–24. https://doi.org/10.1007/s13312-017-0990-9.

Chinratanapisit, S., Suratannon, N., Pacharn, P., Sritipsukho, P., Vichyanond, P., 2018. Prevalence and risk factors of allergic rhinitis in children in Bangkok area. Asian Pacific J allergy Immunol. https://doi.org/10.12932/AP-120618-0337 Published online.

Dales, R.E., Fyfe, M., Schweitzer, I., 1992. Does forced air heating exacerbate asthma? J. Allergy Clin. Immunol. 90 (5), 803–807. https://doi.org/10.1016/0091-6749(92)90105-B.

Deng, Xinlei, Zhang, W., Lin, S., 2020. APML: An Approach for Machine-Learning Modelling. Published online. https://doi.org/10.13140/RG.2.2.27638.42563.

Ehteshami-Afshar, S., FitzGerald, J.M., Doyle-Waters, M.M., Sadatsafavi, M., 2016. The global economic burden of asthma and chronic obstructive pulmonary disease. Int J Tuberc Lung Dis. 20 (1), 11–23. https://doi.org/10.5588/ijtld.15.0472.

Esty, B., Phipatanakul, W., 2018. School exposure and asthma. Ann. Allergy Asthma Immunol. 120 (5), 482–487. https://doi.org/10.1016/j.anai.2018.01.028.

Faizan, M.A., Thakur, R., 2019. Association between solid cooking fuels and respiratory disease across socio-demographic groups in India. J Heal Pollut. 9 (23). https://doi.org/10.5696/2156-9614-9.23.190911.

Folletti, I., Siracusa, A., Paolocci, G., 2017. Update on asthma and cleaning agents. Curr. Opin. Allergy Clin. Immunol. 17 (2), 90–95. https://doi.org/10.1097/ACI.0000000000000349.

Gautier, C., Charpin, D., 2017. Environmental triggers and avoidance in the management of asthma. J Asthma Allergy. 10, 47–56. https://doi.org/10.2147/JAA.S121276.

Gehring, U., Pattenden, S., Slachtova, H., et al., 2006. Parental education and children's respiratory and allergic symptoms in the pollution and the young (PATY) study. Eur. Respir. J. 27 (1), 95–107. https://doi.org/10.1183/09031936.06.00017205.

Gupta, R.S., Warren, C.M., Smith, B.M., et al., 2018. The public health impact of parent-reported childhood food allergies in the United States. Pediatrics 142 (6). https://doi.org/10.1542/peds.2018-1235.

Hasegawa, K., Stoll, S.J., Ahn, J., Kysia, R.F., Sullivan, A.F., Camargo, C.A., 2016. Association of insurance status with severity and management in ED patients with asthma exacerbation. West. J. Emerg. Med. 17 (1), 22–27. https://doi.org/10.5811/westjem.2015.11.28715.

Hastie, Trevor, Tibshirani, Robert, Friedman, J., 2009. The Elements of Statistical Learning The Elements of Statistical LearningData Mining, Inference, and Prediction, Second Edition. https://doi.org/10.1007/978-0-387-84858-7.

Kagan, R.S., Joseph, L., Dufresne, C., et al., 2003. Prevalence of peanut allergy in primary-school children in Montreal, Canada. J Allergy Clin Immunol. 112 (6), 1223–1228. https://doi.org/10.1016/j.jaci.2003.09.026.

Karimi, K., Hamilton, H.J., 2011. Generation and interpretation of temporal decision rules. Computer (Long Beach Calif). 3, 1–17.

Lewis, K.M., Ruiz, M., Goldblatt, P., et al., 2017. Mother's education and offspring asthma risk in 10 European cohort studies. Eur. J. Epidemiol. 32 (9), 797–805. https://doi.org/10.1007/s10654-017-0309-0.

Linneberg, A., Henrik Nielsen, N., Frølund, L., Madsen, F., Dirksen, A., Jørgensen, T., 2002. The link between allergic rhinitis and allergic asthma: a prospective population-based study. The Copenhagen allergy study. Allergy Eur J Allergy Clin Immunol. 57 (11), 1048–1052. https://doi.org/10.1034/j.1398-9995.2002.23664.x.

Mastrorilli, C., Posa, D., Cipriani, F., Caffarelli, C., 2016. Asthma and allergic rhinitis in childhood: what's new. Pediatr. Allergy Immunol. 27 (8), 795–803. https://doi.org/10.1111/pai.12681.

Morgan, W.J., Crain, E.F., Gruchalla, R.S., et al., 2004. Results of a home-based environmental intervention among urban children with asthma. N. Engl. J. Med. 351 (11), 1068–1080. https://doi.org/10.1056/NEJMoa032097.

New York State Asthma Dashboard, 2020. Accessed July 14. https://webbi1.health.ny.gov/SASStoredProcess/guest?_program=/EBI/PHIG/apps/asthma_dashboard/ad_dashboard&p=sh.

Nkoy, F.L., Stone, B.L., Knighton, A.J., et al., 2018. Neighborhood deprivation and childhood asthma outcomes, accounting for insurance coverage. Hosp Pediatr. 8 (2), 59–67. https://doi.org/10.1542/hpeds.2017-0032.

Nori, H., Jenkins, S., Koch, P., Caruana, R., 2019. Interpret ML: A Unified Framework for Machine Learning Interpretability. Published online. http://arxiv.org/abs/1909.09223. (Accessed 23 December 2019).

Nyenhuis, S.M., Akkoyun, E., Liu, L., Schatz, M., Casale, T.B., 2019. Real world assessment of asthma control and severity in children, adolescents, and adults with asthma: Relationships to care settings and co-morbidities. J Allergy Clin Immunol Pract. https://doi.org/10.1016/j.jaip.2019.10.032 Published online.

Palumbo, J.R., Lin, S., Lin, Z., et al., 2018. Assessing associations between indoor environment and health symptoms in Romanian school children: an analysis of data from the SINPHONIE project. Environ. Sci. Pollut. Res. 25 (9), 9186–9193. https://doi.org/10.1007/s11356-018-1568-3.

Pedregosa, F., Varoquaux, G., Gramfort, A., et al., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Schlink, U., Thiem, A., Kohajda, T., Richter, M., Strebel, K., 2010. Quantile regression of indoor air concentrations of volatile organic compounds (VOC). Sci. Total Environ. 408 (18), 3840–3851. https://doi.org/10.1016/j.scitotenv.2009.12.002.

Singer, B.C., Delp, W.W., 2018. Response of consumer and research grade indoor air quality monitors to residential sources of fine particles. Indoor Air 28 (4), 624–639. https://doi.org/10.1111/ina.12463.

Solis-Soto, M.T., Patiño, A., Nowak, D., Radon, K., 2013. Association between environmental factors and current asthma, rhinoconjunctivitis and eczema symptoms in school-aged children from Oropeza Province – Bolivia: a cross-sectional study. Environ. Health 12 (1), 95. https://doi.org/10.1186/1476-069X-12-95.

Soriano, J.B., Abajobir, A.A., Abate, K.H., et al., 2017. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Respir. Med. 5 (9), 691–706. https://doi.org/10.1016/S2213-2600(17)30293-X.

Svendsen, E.R., Gonzales, M., Commodore, A., 2018. The role of the indoor environment: Residential determinants of allergy, asthma and pulmonary function in children from a US-Mexico border community. Sci Total Environ. 616–617, 1513–1523. https://doi.org/10.1016/j.scitotenv.2017.10.162.

Tabka, Z., Jebria, A.B., Vergeret, J., Guenard, H., 1988. Effect of dry warm air on respiratory water loss in children with exercise-induced asthma. Chest. 94 (1), 81–86. https://doi.org/10.1378/chest.94.1.81.

Toskala, E., Kennedy, D.W., 2015. Asthma risk factors. Int Forum Allergy Rhinol. 5 (Suppl. 1), S11–S16. https://doi.org/10.1002/alr.21557.

Tsurikisawa, N., Saito, A., Oshikata, C., Nakazawa, T., Yasueda, H., Akiyama, K., 2013. Encasing bedding in covers made of microfine fibers reduces exposure to house mite allergens and improves disease management in adult atopic asthmatics. Allergy, Asthma Clin Immunol. 9 (1). https://doi.org/10.1186/1710-1492-9-44.

van Aalderen, W.M., 2012. Childhood asthma: diagnosis and treatment. Scientifica (Cairo). 2012, 1–18. https://doi.org/10.6064/2012/674204.

Vicendese, D., Dharmage, S.C., Tang, M.L.K., et al., 2015. Bedroom air quality and vacuuming frequency are associated with repeat child asthma hospital admissions. J Asthma. 52 (7), 727–731. https://doi.org/10.3109/02770903.2014.1001904.

Wang, J., Zhao, Z., Zhang, Y., et al., 2019. Asthma, allergic rhinitis and eczema among parents of preschool children in relation to climate, and dampness and mold in dwellings in China. Environ. Int. 130, 104910. https://doi.org/10.1016/j.envint.2019.104910.