Research Article

A New Air Quality Prediction Framework for Airports Developed with a Hybrid Supervised Learning Method

Yong Tian ¹, Weifang Huang,^{1,2} Bojia Ye,¹ and Minhao Yang¹

¹College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

²Operation Supervisory Center, Civil Aviation Administration of China, Beijing 100710, China

Correspondence should be addressed to Yong Tian; tianyong@nuaa.edu.cn

Received 14 February 2019; Revised 28 March 2019; Accepted 2 April 2019; Published 15 April 2019

Academic Editor: Alicia Cordero

Copyright © 2019 Yong Tian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to reduce the air pollution impacts by aircraft operations around airports, a fast and accurate prediction of air quality related to aircraft operations is an essential prerequisite. This article proposes a new framework with a combination of the standard assessment procedure and machine learning methods for fast and accurate prediction of air quality in airports. Instead of taking some specific pollutant as concerned metric, we introduce the air quality index (AQI) for the first time to evaluate the air quality in airports. Then, following the standard assessment procedure proposed by International Civil Aviation Organization (ICAO), the airports AQIs in different scenarios are classified with consideration of the airport configuration, actual flight operations, aircraft performance, and related meteorological data. Taking the AQI classification results as sample data, several popular supervised learning methods are investigated for accurately predicting air quality in airports. The numerical tests implicate that the accuracy rate of prediction could reach more than 95% with only 0.022 sec; the proposed framework and the results could be used as the foundation for improving air quality impacts around airports.

1. Introduction

Air quality degradation in airports is of concern due to the intensive growth of air traffic over years and its associated environmental risk to human health [1]. Recently, premature mortalities due to aircraft emissions within landing and takeoff (LTO) regime are estimated at about 4000 early deaths globally every year with monetized damages of \$8.19 billion US dollars [2]. Airports located within nonattainment areas are required to implement measures to bring pollution levels into compliance [3]. Given that the airport-related sources of emissions are recognized to contribute to the degradation of air quality in airports by International Civil Aviation Organization [4], the Air Traffic Management (ATM) research community has been busy investigating different methods to reduce the air quality impacts of aircraft operations at airports; a fast and accurate prediction of air quality in airports with consideration of actual operations is an essential prerequisite.

Notably, significant achievements have been made in air quality evaluation at the airport level over the past two

decades. The statistical methods had been proposed by many researchers to quantify the air quality impacts of aircraft activities at the airport level [5–10]. Carslaw et al. [11] presented approaches that aim to detect and quantify the airport contribution to NO_x concentrations for a network of seven measurement sites close to the airport. Adamkiewicz et al. [12] used extensive monitoring of nitrogen dioxide (NO₂) with land-use regression (LUR) modeling techniques to determine the impact of proximity to the airport and local traffic on these concentrations. Carruthers et al. [13] conducted a quasi-Gaussian dispersion model nested within a trajectory model to calculate air quality. Lobo et al. [14] proposed the methodology for real-time measurements of aircraft engine specific particulate matter (PM) emissions and found that mass-based emissions for older technology engines were 3× higher at takeoff than at idle state. Hsu et al. [15, 16] used high-resolution monitoring and flight activity data to quantify contributions from LTO to ultrafine particulate matter concentrations. Diez et al. [17] proposed a statistical approach for identifying air pollutant mixtures associated with aircraft departures at Los Angeles

International Airport. The statistical methods could make fast and accurate predictions for the air pollutants in airports with observed data, but it is hard to quantify exactly the contribution from actual operations, which would be the key foundation for reducing the air quality impacts caused by aircraft operations in airports.

In order to assist member states in implementing best practices with respect to airport-related air quality, ICAO released the Airport Air Quality Manual and proposed a standard airport air quality assessment procedure related to airport operations, which includes two main areas: emissions inventory calculation and dispersion modeling of pollution concentrations [4]. In addition, some mature modeling systems are also recommended for pollution concentration assessment, such as AEDT, ADMS, ALAQS, and LASPORT. These systems can provide an efficient evaluation of air quality in airports with consideration of airport configuration, actual flight operations, aircraft performance and related meteorological data, etc. However, the assessment procedure is complex and time-consuming, which may take data collection time of several weeks and the execution time may vary from few minutes to tens of minutes for each running. Consider the frequent changes of the weather conditions and stochastic factors to actual operations in airports; the efficiency is facing huge concerns when these systems are used for real-time prediction.

In this research, motivated by recent improvements of accurate and efficiency made in artificial intelligence [18–22], we propose a new framework with a combination of the standard assessment procedure and machine learning methods for fast and accurate predicting air quality in airports. Instead of taking some specific pollutant as concerned metric, we introduce the air quality index (AQI) for the first time to evaluate the air quality in airports. Then, following the standard assessment procedure proposed by International Civil Aviation Organization (ICAO), the airports AQIs in different scenarios are classified with consideration of the airport configuration, actual flight operations, aircraft performance, and related meteorological data. Taking the AQI classification results as sample data, several popular supervised learning methods are investigated for developing the best prediction model with the highest accuracy.

The rest of the paper is organized as follows: Section 2 describes the data sources and types used in the research. Section 3 describes the proposed framework for air quality prediction in airports. Section 4 provides a Nanjing airport case study of the proposed methodology. Finally, Section 5 presents a summary and conclusions from the study.

2. Data Source

2.1. Actual Operations Data. The actual operations data including the scheduled, planned, and actual arrival, departure, and taxi operations at concerned airport are necessary for describing air traffic movements in the airport. The Civil Aviation Administration of China (CAAC) makes the data every year for analyzing of the performance of Chinese air transportation system. The flight plans are submitted by the

operators to all air traffic centers where the aircraft is planned to fly; then the observed data would be recorded along with the plans' executions. The actual operations data include aircraft type, operator, departure airport, arrival airport, taxi-out time, taxi-in time, data and time of the flight, and duration of the flight. Such data is processed to yield traffic volume data for fuel burns and emissions inventory calculation.

2.2. Aircraft Performance Data. Flights operational data acquired from actual operations data are useless alone in the study. We still need additional aircraft performance data for calculating the fuel burns and emissions inventory. Two sources of aircraft performance data used in the research are the Base of Aircraft Data (BADA) and the Aircraft Engine Emissions Databank (EDB) [23, 24].

BADA is a collection of files which specifies operation performance parameters for primary aircraft types published by EUROCONTROL. The information is designed for use in trajectory simulation and prediction algorithms within the domain of Air Traffic Management (ATM). For calculation of the fuel burns for different types of aircraft, some engine related information should be extracted from the database. The EDB is published by ICAO and contains information on exhaust emissions of those engines that have entered production. It provides the engine's fuel flows and emission factors at various operating modes. EDB includes almost all kinds of engines used on civil aviation aircraft; the number of engines, the fuel flow rate of the engine under various operating conditions (taxi, takeoff, climb, and approach), and the emission indexes of nitrogen oxides and carbon monoxide are used in this research. Therefore, the data has high credibility, accuracy, and authority.

2.3. Meteorological Data. The third source of data, meteorological data of airport, is very important for dispersion modeling airport air quality assessment. Airports themselves issue Aviation Routine Weather Report (METAR) and Terminal Area/Aerodrome Forecast (TAF) which summarize local weather conditions. Both METAR and TAF data contain information on temperature, wind speed and direction, visibility, precipitation, cloud height, cloud cover, humidity, and pressure. Generally, METAR data are observational data and issued roughly hourly to ensure that reports keep up with changing weather conditions while TAF data are forecast data and updated every 6 hours. Thus, a collection of hourly observations of METAR data covering the concern airport would comprise an ideal dataset to evaluate airport air quality.

Hourly and daily statistics of four meteorological factors of METAR, temperature, humidity, wind speed, and pressure, were manually exported from web-interface of Weather Underground service. These variables are important parameters in dispersion modeling. Combining these weather variables with the actual operations data and aircraft performance data, we could assess emissions dispersion and air quality by adopting the publicly available and internationally recognized methods and yield label data for supervised learning where the goal is to classify the air quality in airport.

TABLE 1: AQI and health implications.

AQI	AQI levels	Descriptor
0 to 50	Level 1	good
51 to 100	Level 2	fine
101 to 150	Level 3	mildly polluted
151 to 200	Level 4	moderately polluted
201 to 300	Level 5	heavily polluted
301 to 500	Level 6	seriously polluted

3. Methodology

We proposed an integration air quality prediction framework in airports, which includes two main parts: air quality classification in airports and the supervised learning for AQI prediction. The air quality classification in airport is mainly composed of air quality index calculation method, airport emissions inventory evaluation, and airport emissions dispersion modeling. In the supervised learning for prediction part, all the AQI level results for the airport are used as training examples; the supervised learning method is used to analyze the training data and produce an inferred function.

3.1. Air Quality Classification in Airports

3.1.1. Air Quality Index Calculation Method. Air quality refers to the condition of the air within our surrounding, and the AQI is commonly used by government to indicate the level of air pollution to the public [25]. Accordingly, we introduce AQI to evaluate the air quality in airports.

AQI is a dimensionless index that describes the status of air quality quantitatively; different countries have their own air quality indices. In China, the Ministry of Environmental Protection (MEP) has developed an AQI in 2012 which is used as the metric in our research [26]. This AQI is a number on a scale from 1 to 500, where a low value means good air quality and a high value means bad air quality. It can be calculated either per hour or per 24 hours, and the value can be also divided into six levels indicating increasing levels of health concern with different descriptors. For example, an AQI over 300 represents serious pollution while below 50 means the air quality is good, as shown in Table 1. Since people usually pay more attention to the health implications, the AQI level is used as prediction aim in the study.

To calculate the AQI level, an individual air quality index (IAQI) is assigned to concerned pollutant first; then the highest value of IAQI is identified as the final AQI, as

$$AQI = \max \{IAQI_1, IAQI_2, \dots, IAQI_n\} \quad (1)$$

The concerned pollutants in AQI include carbon monoxide (CO), sulfur dioxide (SO₂), nitrogen dioxide (NO_x), and suspended particulates; the IAQI for each concerned pollutant is calculated by

$$IAQI_p = \frac{IAQI_{Hi} - IAQI_{Lo}}{BP_{Hi} - BP_{Lo}} (C_p - BP_{Lo}) + IAQI_{Lo} \quad (2)$$

where $IAQI_p$ is the individual air quality index for pollutant p , C_p is the rounded concentration of pollutant p , BP_{Hi} represents the breakpoint that is greater than or equal to C_p , BP_{Lo} represents the breakpoint that is less than or equal to C_p , $IAQI_{Hi}$ indicates the individual air quality index value corresponding to BP_{Hi} , and $IAQI_{Lo}$ indicates the individual air quality index value corresponding to BP_{Lo} . The key input of (2) is the rounded concentration of concerned pollutant C_p , which is calculated based on the emissions inventory and dispersion. Thus, we need to evaluate the airport emissions inventory first and then compute the pollutants concentration based on emissions dispersion model for the airport.

The aircraft related pollutant emissions in airports include NO_x, CO, HC, CO₂, and SO₂. Since the CO, NO₂, and SO₂ are main concerned air pollutants according to MEP, this paper composites these three pollutant emissions to estimate air quality. The IAQI breakpoints of different air pollution are shown in Table 2.

3.1.2. Airport Emissions Inventory Evaluation. In the airport area, the main source of emissions is from the aircraft movement, which is defined as aircraft landings and takeoffs (LTO) by International Civil Aviation Organization (ICAO). The LTO cycle is based on times in mode data during high activity periods at major airports for four modes of engine operation: taxi/idle, takeoff, climb-out, and approach [4]. LTO operations have a significant contribution (up to 70%) to an airport's pollutant emission inventory [3, 27, 28]. Other sources of emissions associated with aircraft movements considered in the study include the onboard auxiliary power units (APU) and aerospace ground support equipment (GSE). The ground transport travelling to and from airports is not considered in the airport.

Firstly, we follow the approach used by Martini et al. [29] in evaluating aircraft pollution and considering that aircraft only affects local air pollution when operating along the landing takeoff LTO cycle. The LTO cycle is split into four stages: idle (taxiing or standing with engines on), takeoff, climb (up to 3000 ft), and approach (from 3000 ft to landing). The mass emission rates of aircraft are determined at specific percentages of rated thrust with different reference time by certificating authority as Table 3. In the operations, the time of taxi mode may change with different airports while the time of other three modes remain relatively stable. In our study, we use the observed taxi-in and taxi-out time instead of the reference taxi time to improve the accuracy of the evaluation.

We computed the LTO emissions produced by each aircraft/engine combination by matching BADA and EDB databases together. The BADA database contains information regarding the engine type and the number of engines on each aircraft model while the EDB database provides the emission factor for the four LTO phases and for each engine model. To compute the total quantity ($P_{p,LTO}$) of pollutant p , produced by an aircraft during the LTP cycle, we apply the following equation:

$$P_{p,LTO} = N_{EN} \times \sum_{k=1}^4 t_k \times F_k \times E_{p,k} \quad (3)$$

TABLE 2: IAQI breakpoints of different air pollutants.

IAQI	CO 24-hour mean ($\mu\text{g}/\text{m}^3$)	NO ₂ 24-hour mean ($\mu\text{g}/\text{m}^3$)	SO ₂ 24-hour mean ($\mu\text{g}/\text{m}^3$)	CO 1-hour mean ($\mu\text{g}/\text{m}^3$)	NO ₂ 1-hour mean ($\mu\text{g}/\text{m}^3$)	SO ₂ 1-hour mean ($\mu\text{g}/\text{m}^3$)
0	0	0	0	0	0	0
50	2	40	50	5	100	150
100	4	80	150	10	200	500
150	14	180	475	35	700	650
200	24	280	800	60	1200	800
300	36	565	1600	90	2340	1600
400	48	750	2100	120	3090	2100
500	60	940	2620	150	3840	2620

TABLE 3: Reference emissions LTO cycle.

LTO operating mode	Thrust setting	Time in operating mode (min)
Take-off	100%	0.7
Climb	85%	2.2
Approach	30%	4.0
Taxi	7%	26

where N_{EN} represents the number of engines on an aircraft, t_k is the time duration (s) at phase k , F_k is the fuel flow rate (kg/s) of a single engine of the aircraft at phase k , and $E_{p,k}$ is the pollutant p emission index (g/kg) of the aircraft at phase k .

The APU emissions ($P_{p,APU}$) of each aircraft are calculated using the following formula:

$$P_{p,APU} = N_{APU} \times EU_p \times t_{APU} \quad (4)$$

where N_{APU} represents the number of APU on the aircraft; EU_p is the pollutant p emission index (g/kg) of the APU. t_{APU} is the operating time of the APU.

The GSE emissions ($P_{p,GSE}$) for the aircraft are calculated using the following formula:

$$P_{p,GSE} = \sum_{j=1}^m EA_{p,j} \times t_{j,GSE} \quad (5)$$

where j represents the type of the GSE, m is the total number of GSE types, $EA_{p,j}$ is the pollutant p emission index (g/kg) of the GSE j , and $t_{j,GSE}$ is the operating time of the GSE j .

3.1.3. Airport Emissions Dispersion Modeling. Based on the airport emissions inventory, the pollutants concentration of the airport can be computed by pollutants dispersion modeling. There are several alternative models are recommended by ICAO, among which the Aviation Environment Design Tool (AEDT) [30, 31] developed by Federal Aviation Administration (FAA) is used in the study for concentration calculations.

AEDT is a software system that models aircraft performance in space and time which can estimate noise, fuel consumption, and air quality consequence. The emissions dispersion module of AEDT uses the US National Environmental Protection Agency (EPA) Regulatory Model (AERMOD) for pollution dispersion modeling. Based on the diffusion statistics theory, the core of AERMOD air quality model is Gaussian diffusion model, which assumes that pollutant concentration distribution is subject to Gaussian distribution in a certain extent. The model system can be used for various emission sources including point source, volume source, and area source. In this study, the taxiing process of aircraft scenes is regarded as surface pollution source, and the taxiing process of aircraft takeoff is regarded as line pollution source. The diffusion formula of AEDT model under different conditions is as follows:

(1) General diffusion formula:

$$\rho_T(x, y, z) = f \cdot \rho(x, y, z) + (1 - f) \cdot \rho(x, y, z_a) \quad (6)$$

$$\rho(x, y, z) = \frac{Q}{U} \cdot p(y, x) \cdot p(z, x) \quad (7)$$

$$f = 0.5 \cdot (1 + \Phi) \quad (8)$$

$$\Phi = \frac{\int_0^H \rho(x, y, z) dz}{\int_0^\infty \rho(x, y, z) dz} \quad (9)$$

$$z_a = z - z_i \quad (10)$$

where $\rho_T(x, y, z)$ is the total concentration; $\rho(x, y, z_a)$ is the elevation of topographic plume; Φ is the ratio of the plume mass to the total plume mass; Q is the discharge rate of the source; U is the effective wind speed; $p(y, x)$ and $p(z, x)$, respectively, represent the probability density function of horizontal and vertical concentration distribution. f is the weight function; z_a is the effective height; z_i is the topographic height value of this point.

(2) Diffusion formula for convective boundary layer:

$$\rho(x, y, z) = \rho_d(x, y, z) + \rho_r(x, y, z) + \rho_p(x, y, z) \quad (11)$$

$$\rho_d(x, y, z) = \frac{Q}{2\pi U \sigma_y} \cdot \exp \left[-\frac{y^2}{2\sigma_y^2} \right] \cdot \sum_{j=1}^0 \sum_{m=0}^{\infty} \frac{\lambda_j}{2\sigma_j} \cdot \left\{ \exp \left[-\frac{(z - h_j - 2mz_j)^2}{2\sigma_j^2} \right] + \exp \left[-\frac{(z + h_j + 2mz_j)^2}{2\sigma_j^2} \right] \right\} \quad (12)$$

where $\rho(x, y, z)$ is the total concentration of tobacco; $\rho_d(x, y, z)$ is direct discharge concentration of pollution sources; $\rho_r(x, y, z)$ is the emission concentration of virtual source; $\rho_p(x, y, z)$ is the concentration discharge of the coil source; λ_j is the Gaussian distribution weight coefficient; λ_1 is the upflow, λ_2 is the downflow; h_j is the effective source height; σ_j is the vertical diffusion coefficient.

(3) Stable boundary layer diffusion formula:

$$\rho(x, y, z) = \frac{Q}{U} \cdot F_z \cdot F_y \quad (13)$$

$$F_z = \frac{1}{\sqrt{2\pi}\sigma_z} \cdot \sum_{n=-\infty}^{\infty} \left\{ \exp \left[-\frac{(z - h_p + 2nh_z)^2}{2\sigma_z^2} \right] + \exp \left[-\frac{(z + h_p + 2nh_z)^2}{2\sigma_z^2} \right] \right\} \quad (14)$$

$$F_y = \frac{1}{\sqrt{2\pi}\sigma_y} \cdot \exp \left[-\frac{y^2}{2\sigma_y^2} \right] \quad (15)$$

where $\rho(x, y, z)$ is the total concentration of the plume; F_z is the dilution of smoke plume; F_y is the distribution of smoke plumes; h_p is the plume height; h_z is the maximum height of vertical mixing layer; σ_y and σ_z are the diffusion parameter of the plume in horizontal and vertical directions.

To calculate the pollutants concentration, beside terrain information around the airport, the major influence factors are the emissions inventory for different pollution and the main meteorological information including temperature, barometric pressure, wind speed, and humidity which can be obtained from the METAR data.

3.2. Supervised Learning for AQI Prediction. Based on the airport air quality classification method introduced in Section 3.1, we can construct the training set of air quality for the specific airport. Now we are interested in investigating different supervised learning algorithm to predict the air quality with different airport operational features in this section.

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labeled training data consisting of a set of training examples [32, 33]. The main steps to solve a given problem are as follows:

- (i) Determine the type of training examples. Before doing anything else, the user should decide what kind of data is to be used as a training set.
- (ii) Gather a training set. The training set needs to be representative of the real-world use of the function. Thus, a set of input objects is gathered and corresponding outputs are also gathered, either from human experts or from measurements.
- (iii) Determine the input feature representation of the learned function. The accuracy of the learned function depends strongly on how the input object is represented.
- (iv) Determine the structure of the learned function and corresponding learning algorithm. For example, one may choose to use support vector machines or random forest.
- (v) Complete the design. Run the learning algorithm on the gathered training set. Some supervised learning algorithms require the user to determine certain control parameters.
- (vi) Evaluate the accuracy of the learned function. After parameter adjustment and learning, the performance of the resulting function should be measured on a test set that is separate from the training set.

A wide range of supervised learning methods are available to solve the air quality classification problem described above, and each method has its own strengths and weaknesses. However, subject to the dimensionality of the input space, noise of the output values, function complexity, and bias-variance tradeoff, no single one can guarantee the best solutions for the proposed problem. In this study, four widely used methods including the support vector machines (SVM), logistic regression, random forests, and ensemble learning methods are investigated to solve our problem. For each classifier, the parameters are set and tuned for the application while trying to find the tradeoff between accuracy and running time; a 10-fold cross validation procedure is used to avoid overfitting and improve the prediction accuracy.

Logistic regression is a technique borrowed by machine learning from the field of statistic [34]. It measures the relationship between the categorical dependent variable and two or more independent variables by estimating probabilities using a logistic function, which is the cumulative logistic distribution. The linear logistic regression classifier is implemented first. The SVM is a discriminative classifier formally defined by a separating hyperplane [35]. Given the labeled data, the method outputs an optimal hyperplane which categorizes new examples. It can efficiently perform both linear and nonlinear classification by using kernel trick to maximum-margin hyperplanes. Two kernels including the linear kernel (SVM-L) and exponential kernel (SVM-E) are tested in this study. Ensemble methods use multiple learning algorithms to obtain better predictive performance other than using any learning method alone. There exist several ensemble methods; the random forests which are based on a collection of decision trees, built up with some elements of

FIGURE 1: Nanjing Lukou International Airport layout.

random choices, are tested in this study [36, 37]. The Artificial Neural Network (ANN) is not an algorithm, but rather a popular framework for many different machine learning algorithms to work together and process complex data inputs. It learns to perform tasks by considering examples, generally without being programmed with any task-specific rules. In this study, a multiplier perceptron (MLP) which is a class of feedforward neural networks is explored for classification [38].

The different classification methods are implemented with Python as the programming language. The training and testing were executed on a Windows workstation with 2.10Ghz Intel Xeon E5-2620 and 16GB RAM.

4. Numerical Test

4.1. Data Preprocess. The proposed air quality classification framework is applied to a case study of the Nanjing Lukou International Airport (ZSNJ) using the actual operations data in the year 2017 from CAAC. ZSNJ is the second largest airport in eastern China. It now has more than 135 domestic air routes and about 25 international air routes operated by over 25 airlines. As illustrated in Figure 1, the airport consists of two 3600-meter parallel runways RWY06/24 and RWY07/25.

In the year 2017, there are 102,548 arrival flights and 102,520 departure flights in ZSNJ, and the actual operations data we collected including the flight number, aircraft type, departure airport, arrival airport, departure time, arrival time, off-block time, and on-block time. As mentioned in Section 3.1.2, the ground transport travelling to and from airports is not considered in the airport. In order to accurately estimate the aircraft emissions, the actual taxi-out time and the taxi-in time in ZSNJ are used in LTO cycle. The taxi-in time of a flight is calculated by departure time minus off-block time, and the taxi-in time is calculated by on-block time

FIGURE 2: Taxi time in Nanjing Lukou International Airport.

minus arrival time. Figure 2 illustrates the taxi-in and taxi-out time statistic data in ZSNJ.

By statistical analysis, there are 43 types of aircraft operating in the whole year; eight of them which include the A319, A320, A321, B733, B734, B737, B738, and B752 account for 94.5% of the operations in the whole year. Combining the BADA and the EDB database together, the flow rates and engine emissions index of CO and NO_x for these aircraft types are shown in Table 4. According to the Airport Air Quality Manual (Doc 9889) released by ICAO [4], the emissions of SO₂ are only related to the aircraft fuel consumption which is 1g/kg for all engine modes, and the converter efficiency from NO_x to NO₂ is set as 10%. Also, the values representative of APU and GSE emissions for each aircraft operation can also be found in the document and will not be listed here.

TABLE 4: Fuel rate and engine emissions index.

Aircraft type and number of engines	Engine type	Fuel rate and Engine emissions index									
		Fuel rate (kg/sec)			EI CO (g/kg)			EI NO _x (g/kg)			Idle
		T/O	C/O	App	Idle	T/O	C/O	App	Idle	T/O	
A319 (2)	V2522-A5	0.971	0.817	0.311	0.118	0.57	0.67	2.60	13.42	24.5	4.5
A320 (2)	V2500-A1	1.113	0.924	0.334	0.124	0.55	0.55	0.77	7.76	37.13	5.91
A321 (2)	CFM56-5B2/2P	1.39	1.11	0.39	0.13	0.7	1.6	21.4	35.5	29.4	4.2
B733 (2)	CFM56-3-B1	0.946	0.792	0.290	0.114	0.9	0.95	3.8	34.4	17.7	3.9
B734 (2)	CFM56-3C-1	1.154	0.954	0.336	0.124	0.9	0.9	3.1	26.8	20.7	4.3
B737 (2)	JT8D-17A	1.173	0.9344	0.3304	0.1401	1.07	1.16	2.88	12.46	19.1	3.2
B738 (2)	CFM56-7B26	1.221	0.999	0.338	0.113	0.20	0.60	1.60	18.80	28.80	4.70
B752 (2)	RB211-535E4	1.86	1.51	0.57	0.19	1.01	1.23	1.71	15.44	52.7	4.3

FIGURE 3: Daily AQI and flight volume for ZSNJ in 2017.

FIGURE 4: Daily AQI and flight volume for ZSNJ in 2017 ordered by AQI.

Also, we obtained daily and hourly data for four meteorological factors which are the temperature ($^{\circ}\text{F}$), humidity (%), wind speed (mph), and pressure (Hg) in ZSNJ from the Weather Underground website. There are 365 daily average observations and 8760 hourly observations from January 1, 2017, to December 31, 2017, and all data are used as inputs after normalization of ratings by z-scores.

4.2. Prediction Results Analysis. Take the AQI level as metric, and use the airport air quality classification method

introduced in Section 3.1; we calculated the AQI levels for ZSNJ which were then used as the training set for the airport air quality prediction. Figure 3 illustrates the daily AQI levels and the flight volume for ZSNJ in 2017. The bottom axis reflects the day time, the green polyline represents the air quality with vertical coordinates on the left, and the column chart denotes the volume of arrival (blue) and departure (yellow) flights with vertical coordinates on the right. We reorder the AQI for ZSNJ in 2017 in Figure 4, and it can be seen that 61 days stayed at AQI level 1, 196 days stayed at AQI

FIGURE 5: Sample hourly AQI and flight volume for ZSNJ in 2017.

TABLE 5: Classifier accuracy, training, and testing time.

Machine learning method	Classification Accuracy		Training and Running Time	
	daily	hourly	daily	hourly
Logistic Regression	0.6465	0.4745	0.1233s	0.067s
SVM	0.6570	0.5342	10.9212s	0.023s
MLP	0.6558	0.5479	0.1442s	0.021s
Random Forest	0.9070	0.9315	0.1419s	0.025s

level 2, 87 days stayed at AQI level 3, 18 days stayed at AQI level 4, 2 days stayed at AQI level 5, and 1 day stayed at AQI level 6. In total, 94.2% daily AQIs in 2017 stay at level 1, level 2, or level 3, with 5.8% days achieving as high as level 5 and level 6. Generally, the air quality in ZSNJ remains in acceptable range.

Similarly, Figure 5 shows one-week sample data of the hourly AQI and flight volume for ZSNJ from February 1 to February 7 in 2017. The hourly AQIs level show relatively larger fluctuations with more bad air quality data in total. In 2017, there are 8760 hours evaluated in total for ZNSJ airport, with about 94.8% hours staying at AQI level 1, level 2, or level 3, with 5.2% hours achieving as high as level 5 or level 6.

Take the evaluation results as training data; several supervised learning methods are used to develop a classifier for predicting the AQI of ZNSJ. The prediction analysis focuses on investigating the feasibility of predicting the air quality classification with selected features. Also, the results of the analysis compare the supervised learning methods in terms of accuracy, training, and testing times. In the light of the knowledge-based feature selection method [22], both weather-related features and operation features are selected for prediction. The explored features for prediction include the number of departure flights, the number of arrival flights, temperature, humidity, wind speed, and pressure. Logistic

regression, SVM, MLP, and random forest are applied to classify daily and hourly AQI in ZSNJ; the results of analysis compare the 4 classifier performances in terms of accuracy, training, and running time on predicting the airport air quality. To avoid overfitting and improve the prediction accuracy, we implemented 10-fold cross validation procedures for the nonensemble methods, and 10 different training sets for the random forest method. The prediction accuracy, training, and testing time for different methods are illustrated in Table 5.

Figure 6 illustrates the performance result distributions in terms of accuracy for different machine learning methods and compares them for the two air quality datasets. For each machine learning method, a box-plot is chosen to represent the accuracy distribution, whose values can be observed on the left y-axis. Box-plot is a standardized form to illustrate data distribution. The bottom and top of the box, respectively, represent the 25th and 75th percentiles, and the line inside the box denotes the median. The bottom and top lines indicate the minimum and maximum values of the data considered. To show the variations of successful method, the lower accuracy value limit was intentionally set to 0.3. In summary, the ensemble method random forest provides the best accuracy results, respectively, giving 90.7% and 91.37% mean accuracy for daily dataset and hourly dataset. For the

FIGURE 6: Performance result distributions in terms of accuracy.

FIGURE 7: Sensitivity analysis for random forest method for hourly dataset.

nonensemble methods, all the accuracy means are less than 70% for both datasets, and the MLP shows the best accuracy of the three methods. Additionally, the accuracy of hourly dataset results appears to be better than daily dataset results.

Figure 7 illustrates the sensitivity analysis for random forest method used for hourly AQI prediction. Two important parameters which are the proportion of training set and the maximum depth per tree are adjusted to find the near optimal performance for the prediction accuracy rate. In the figure, the proportion of training set varies from 0.1 to 0.9 in 0.1 steps while the maximum depth per tree changes from 1 to 14 in 1 step. Three metrics which are the training accuracy rate, prediction accuracy rate, and the running time are analyzed with the adjustment of different parameters setting. Generally, both the training and prediction accuracy rates increase with the increase of maximum depth per tree. However, higher proportion of training set does not show significant impacts on the three metrics. Take the parameters setting for the maximum prediction accuracy rate in the sensitivity analysis. When the proportion of training set is 0.8 and the maximum depth per tree is 12, the training accuracy rate can reach as high as 98.29% while the prediction accuracy rate can reach 95.64% with only 0.022 seconds as running time.

4.3. Feature Importance Analysis. The analysis of feature importance enables us to discriminate and select the most impactful features on the classification results. Random forest method is selected to extract the features since it provided the best prediction results for air quality classification in the study case. Feature importance provides a score that indicates how valuable each feature was in the construction of the random forest trees within the model, and the higher score implicates more important feature in modeling.

All six features including the number of departure flights, the number of arrival flights, temperature, humidity, wind speed, and pressure are analyzed in the study, and both daily and hourly air quality data are investigated. Figure 8(a) shows that, for the daily air quality classification case, temperature is primarily dominating the other features and it is followed by the humidity and pressure. Since Nanjing is a city that has four distinct seasons and the temperatures vary obviously with season, this should be reason why temperature is more important than other factors. Also, the humidity has obvious correlation with precipitation; this may be why it is the second important feature. Figure 8(b) shows that, for the hourly air quality classification case, pressure is primarily dominating the other features and it is followed by the temperature and humidity. This may implicate that the temperature and

FIGURE 8: Feature importance for air quality classification.

humidity are usually not changed much in one day, and the pressure can be the most important feature in air quality classification. Theoretically, the wind speed is an important feature which may affect the pollutants concentration; the results implicate that it is not the most important one in ZSNJ airport. Of special note is the fact that the departure and arrival flights do not occupy the dominating features in either daily or hourly classification. This may denote that although the aircraft is the source of pollutants in ZSNJ, the meteorological features have greater influence on the air quality classification.

5. Conclusion

Given recent improvements of accuracy and efficiency made in artificial intelligence, this paper proposes a framework for air quality classification in airports, which is aligned with exploring the applicability and effectiveness of several popular supervised learning algorithms. A novel hybrid classification approach is proposed starting with the airport air quality assessment, followed by supervised learning for classification. Also, the framework is applied to a case study of the Nanjing Lukou International Airport for numerical test; prediction results and feature importance are analyzed in the study. The main findings in the paper are summarized below.

The proposed air quality classification method could predict the AQI level for airports, and the study case of ZSNJ in 2017 shows that most daily AQIs stay around level 2 and level 3, with few days achieving level 5 and level 6 while the hourly AQIs levels show relatively larger fluctuations with

more bad air quality data. For the air quality prediction, the ensemble method random forest provides the best accuracy results, respectively, giving 90.7% and 91.37% mean accuracy for daily dataset and hourly dataset, and the prediction accuracy rate of hourly dataset can even reach 95.64% with only 0.022 seconds as running time after the sensitivity analysis. According to the feature importance analysis, temperature is primarily dominating the other features and it is followed by the humidity and pressure in daily classification, while the pressure is primarily dominating the other features and it is followed by the temperature and humidity in the hourly air quality classification. In general, the meteorological features have greater influence on the air quality classification in ZSNJ, which is worthy of attention in alleviating the environmental impacts of aircraft in airports.

This present paper contributes to the field of ATM by bringing and adapting machine learning methods to solve air quality classification issue in airports. Future research may refer to optimization of aircraft operations in airports for environmental consideration.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Yong Tian and Weifang Huang conceived and designed the experiments. Yong Tian performed the experiments. Weifang Huang and Minhao Yang analyzed the data. Bojia Ye contributed analysis tools. Yong Tian and Bojia Ye wrote the paper.

Acknowledgments

This research was cosupported by the Natural Science Foundation of China (61671237), Natural Science Foundation of Jiangsu Province, China (No. BK20160798), and China Postdoctoral Science Foundation (2018M632308).

References

- [1] WHO, "WHO—7 million premature deaths annually linked to air pollution," 2014, <http://www.who.int/mediacentre/news/releases/2014/airpollution/en/>.
- [2] S. H. L. Yim, G. L. Lee, I. H. Lee et al., "Global, regional and local health impacts of civil aviation emissions," *Environmental Research Letters*, vol. 10, no. 3, Article ID 34001, 2015.
- [3] A. Ashok, H. Balakrishnan, and S. R. H. Barrett, "Reducing the air quality and CO₂ climate impacts of taxi and takeoff operations at airports," *Transportation Research Part D: Transport and Environment*, vol. 54, pp. 287–303, 2017.
- [4] ICAO (International Civil Aviation Organization), *Airport Air Quality Manual*, 1st edition, 2011, Doc 9889.
- [5] A. Unal, Y. Hu, M. E. Chang, M. T. Odman, and A. G. Russell, "Airport related emissions and impacts on air quality: application to the atlanta international airport," *Atmospheric Environment*, vol. 39, no. 32, pp. 5787–5798, 2005.
- [6] K. N. Yu, Y. P. Cheung, T. Cheung, and R. C. Henry, "Identifying the impact of large urban airports on local air quality by nonparametric regression," *Atmospheric Environment*, vol. 38, no. 27, pp. 4501–4507, 2004.
- [7] G. Schürmann, K. Schäfer, C. Jahn et al., "The impact of NO_x, CO and VOC emissions on the air quality of Zurich airport," *Atmospheric Environment*, vol. 41, no. 1, pp. 103–118, 2007.
- [8] D. Westerdahl, S. A. Fruin, P. L. Fine, and C. Sioutas, "The Los Angeles International Airport as a source of ultrafine particles and other pollutants to nearby communities," *Atmospheric Environment*, vol. 42, no. 13, pp. 3143–3155, 2008.
- [9] R. E. Dodson, E. A. Houseman, B. Morin, and J. I. Levy, "An analysis of continuous black carbon concentrations in proximity to an airport and major roadways," *Atmospheric Environment*, vol. 43, no. 24, pp. 3764–3773, 2009.
- [10] S. Hu, S. Fruin, K. Kozawa, S. Mara, A. M. Winer, and S. E. Paulson, "Aircraft emission impacts in a neighborhood adjacent to a general aviation airport in southern California," *Environmental Science & Technology*, vol. 43, no. 21, pp. 8039–8045, 2009.
- [11] D. C. Carslaw, S. D. Beevers, K. Ropkins, and M. C. Bell, "Detecting and quantifying aircraft and other on-airport contributions to ambient nitrogen oxides in the vicinity of a large international airport," *Atmospheric Environment*, vol. 40, no. 28, pp. 5424–5434, 2006.
- [12] G. Adamkiewicz, H.-H. Hsu, J. Vallarino, S. J. Melly, J. D. Spengler, and J. I. Levy, "Nitrogen dioxide concentrations in neighborhoods adjacent to a commercial airport: a land use regression modeling study," *Environmental Health: A Global Access Science Source*, vol. 9, no. 1, article no. 73, 2010.
- [13] D. Carruthers, C. McHugh, M. Jackson, and K. Johnson, "Developments in ADMS-airport to take account of near field dispersion and applications to Heathrow airport," *International Journal of Environment and Pollution*, vol. 44, no. 1-4, pp. 332–341, 2011.
- [14] P. Lobo, D. E. Hagen, and P. D. Whitefield, "Measurement and analysis of aircraft engine PM emissions downwind of an active runway at the Oakland international airport," *Atmospheric Environment*, vol. 61, pp. 114–123, 2012.
- [15] H.-H. Hsu, G. Adamkiewicz, E. Andres Houseman et al., "The relationship between aviation activities and ultrafine particulate matter concentrations near a mid-sized airport," *Atmospheric Environment*, vol. 50, pp. 328–337, 2012.
- [16] H.-H. Hsu, G. Adamkiewicz, E. A. Houseman, D. Zarubiak, J. D. Spengler, and J. I. Levy, "Contributions of aircraft arrivals and departures to ultrafine particle counts near Los Angeles international airport," *Science of the Total Environment*, vol. 444, pp. 347–355, 2013.
- [17] D. M. Diez, F. Dominici, D. Zarubiak, and J. I. Levy, "Statistical approaches for identifying air pollutant mixtures associated with aircraft departures at Los Angeles international airport," *Environmental Science & Technology*, vol. 46, no. 15, pp. 8229–8235, 2012.
- [18] C. S. Bosson and T. Nikoleris, "Supervised learning applied to air traffic trajectory classification," in *Proceedings of the AIAA Information Systems-AIAA Infotech Aerospace Kissimmee*, Florida, Fla, USA, January 2018.
- [19] N. Xu, G. Donohue, K. B. Laskey, and C.-H. Chen, "Estimation of delay propagation in the national aviation system using Bayesian networks," in *Proceedings of the 6th USA/Europe Air Traffic Management Research and Development Seminar, ATM 2005*, pp. 353–363, June 2005.
- [20] M. Gariel, A. N. Srivastava, and E. Feron, "Trajectory clustering and an application to airspace monitoring," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1511–1524, 2011.
- [21] P. Balakrishna, R. Ganesan, and L. Sherry, "Accuracy of reinforcement learning algorithms for predicting aircraft taxi-out times: A case-study of Tampa Bay departures," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 6, pp. 950–962, 2010.
- [22] K. D. Kuhn, "A methodology for identifying similar days in air traffic flow management initiative planning," *Transportation Research Part C: Emerging Technologies*, vol. 69, pp. 1–15, 2016.
- [23] Eurocontrol Experimental Center (EEC), "User Manual for the Base of Aircraft Data (BADA), Revision 3.9," EEC Note No. 11/06/15-25, 2011.
- [24] International Civil Aviation Organization (ICAO), "ICAO engine exhaust emissions databank," 2018, <https://www.easa.europa.eu/easa-and-you/environment/icao-aircraft-engine-emissions-databank>.
- [25] D. Mintz, *Technical Assistance Document for the Reporting of Daily Air Quality - the Air Quality Index (AQI)*, Tech. Research Triangle Park, U.S. Environmental Protection Agency, 2016.
- [26] China Ministry of Environmental Protection (MEP), "Technical regulation on ambient air quality index," 2012, http://www.mee.gov.cn/gkml/hbb/bgg/201203/t20120302_224146.htm.
- [27] D. Irvine, L. Budd, S. Ison, and G. Kitching, "The environmental effects of peak hour air traffic congestion: the case of London

- Heathrow airport,” *Research in Transportation Economics*, vol. 55, pp. 67–73, 2016.
- [28] O. Zaporozhets and K. Synylo, “Improvements on aircraft engine emission and emission inventory asesessment inside the airport area,” *Energy*, vol. 140, pp. 1350–1357, 2017.
 - [29] G. Martini, D. Scotti, and N. Volta, “Including local air pollution in airport efficiency assessment: A hyperbolic-stochastic approach,” *Transportation Research Part D: Transport and Environment*, vol. 24, pp. 27–36, 2013.
 - [30] C. Roof, A. Hansen, G. Fleming et al., “Aviation environmental design tool (AEDT) system architecture,” 2007.
 - [31] J. Koopmann, A. Zubrow, A. Hansen, S. Hwang, M. Ahearn, and G. Solman, *Aviation Environmental Design Tool (AEDT) 2b User Guide*, 2016.
 - [32] R. Caruana and A. Niculescu-Mizil, “An empirical comparison of supervised learning algorithms,” in *Proceedings of the ICML 2006: 23rd International Conference on Machine Learning*, pp. 161–168, June 2006.
 - [33] M. I. Jordan and D. E. Rumelhart, “Forward models: Supervised learning with a distal teacher,” *Cognitive Science*, vol. 16, no. 3, pp. 307–354, 1992.
 - [34] P. Peduzzi, J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstein, “A simulation study of the number of events per variable in logistic regression analysis,” *Journal of Clinical Epidemiology*, vol. 49, no. 12, pp. 1373–1379, 1996.
 - [35] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “Training algorithm for optimal margin classifiers,” in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory (COLT ’92)*, pp. 144–152, July 1992.
 - [36] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
 - [37] A. Liaw and M. Wiener, “Classification and regression by random Forest,” *The R Journal*, vol. 2, no. 3, pp. 18–22, 2002.
 - [38] H. Trevor, T. Robert, and F. Jerome, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, NY, USA, 2009.

Submit your manuscripts at
www.hindawi.com