



OPEN

Assessing wind field characteristics along the airport runway glide slope: an explainable boosting machine-assisted wind tunnel study

Afaq Khattak¹, Pak-wai Chan², Feng Chen^{1✉} & Haorong Peng³

Aircraft landings are especially perilous when the wind is gusty near airport runways. For this reason, an aircraft may deviate from its glide slope, miss its approach, or even crash in the worst cases. In the study, we used the state-of-the-art glass-box model, the Explainable Boosting Machine (EBM), to estimate the variation in headwind speed and turbulence intensity along the airport runway glide slope and to interpret the various contributing factors. To begin, the wind field characteristics were examined by developing a scaled-down model of Hong Kong International Airport (HKIA) runway as well as and the surrounding buildings and complex terrain in the TJ-3 atmospheric boundary layer wind tunnel. The placement of probes along the glide slope of the model runway aided in the measurement of wind field characteristics at different locations in the presence and absence of surrounding buildings. Next, the experimental data was used to train the EBM model in conjunction with Bayesian optimization approach. The counterpart black box models (extreme gradient boosting, random forest, extra tree and adaptive boosting) as well as other glass box models (linear regression and decision tree) were compared with the outcomes of the EBM model. Based on the holdout testing data, the EBM model revealed superior performance for both variation in headwind speed and turbulence intensity in terms of mean absolute error, mean squared error, root mean squared error and R-square values. To further evaluate the impact of different factors on the wind field characteristics along the airport runway glide slope, the EBM model allows for a full interpretation of the contribution of individual and pairwise interactions of factors to the prediction results from both a global and a local perspective.

Wind shear and turbulence: the invisible enemies. The proximity of wind to an airport's runway has a significant impact on aircraft operations. It should come as no surprise that wind is one of the most important factors influencing the flight path of an aircraft. When discussing the "wind effect," it is important to understand the three broad classifications of wind types¹. Wind can influence takeoff and landing from three different directions. When the wind is blowing in the direction of an aircraft, it is referred to as a "headwind". When the wind is blowing from behind the aircraft, it is referred to as a "tailwind". When the wind is blowing from the side of the aircraft, it is referred to as a "crosswind". It is preferable for the aircraft to land in the direction of the wind, which facilitates the landing by reducing its speed. Likewise, it is preferable for the aircraft to take off in the wind's direction. During takeoff, a headwind increases airflow; consequently, the necessary lift is attained sooner and at lower speeds (the wind speed is added to the aircraft speed). Therefore, less runway is required for safe takeoff and landing. Likewise, the tailwind is preferred because it speeds up the aircraft and aids its propulsion systems, as shown in Fig. 1.

¹Key Laboratory of Infrastructure Durability and Operation Safety in Airfield of CAAC, College of Transportation Engineering, Tongji University, 4800 Cao'an Road, Jiading, Shanghai 201804, China. ²Hong Kong Observatory, 134A Nathan Road, Kowloon, Hong Kong, China. ³The Key Laboratory of Infrastructure Durability and Operation Safety in Airfield of CAAC, Tongji University, 4800 Cao'an Road, Jiading, Shanghai 201804, China. ✉email: fengchen@tongji.edu.cn

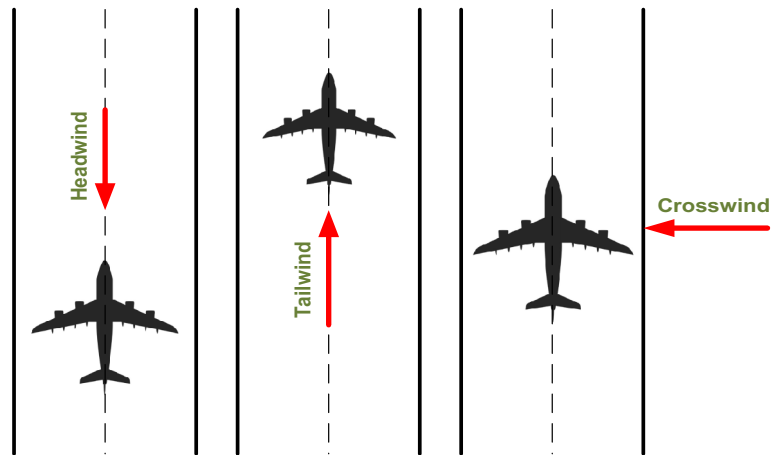


Figure 1. Type of wind effects on the departing and approaching aircraft.

Constant wind speed is preferable from a takeoff and landing safety perspective. Consequently, wind speed is not a problem in and of itself. The issue arises, however, when the wind speed and/or direction fluctuate abruptly^{2,3}. Wind shear is the term for this phenomenon. According to the International Civil Aviation Organization (ICAO), wind shear is defined as a sudden change in wind speed or direction within three nautical miles (~5500 m) of an airport runway threshold at an altitude of 500 m or less⁴. Changes in the aircraft's lift can be significantly impacted by wind shear. The declining and raising headwind shear on an aircraft during an approach is depicted in Fig. 2. When the approaching aircraft is subjected to a declining headwind, the lift is lowered, and the aircraft may possibly land short of the runway. In the case of a rising headwind, the aircraft's airspeed increases in relation to the surrounding air flow, generating more lift and resulting in a flatter angle of descent or even a climb, which ultimately may cause a missed approach⁵.

In addition to wind shear, turbulence may also affect an aircraft's landing. Rapid, irregular air motion is what causes turbulence. It causes jolts or hiccups. In extreme situations, the aircraft may lose control momentarily^{6–8}. When an aircraft encounters turbulence, loose objects within the cabin may become dislodged, and passengers who are not wearing seat belts may sustain injuries⁹. A pilot must take corrective measures to ensure safety in the presence of significant wind shear and turbulence.

Causes of wind shear and turbulence. Two basic factors causing wind shear and turbulence are meteorological elements^{10,11} and the geographical environment inside or around the airport^{12–14} as depicted in Fig. 3. Tropical cyclones, thunderstorms, cold and warm fronts, and jet streams (narrow bands of strong winds) are common weather phenomena that can cause wind shear and turbulence. Sea breezes, strong monsoon winds, and strong winds that climb over hills and man-made structures are additionally recognized to cause low-level wind shear near runways. Similarly, the intensity of turbulence is expected to decline when the wind blows over flat terrain, such as open water, and to increase when the wind blows over sizeable obstacles, such as mountains and buildings^{15,16}.

Assessment of wind shear and turbulence by numerical simulation. Over the years, various studies related to aviation meteorology have been performed to gain an understanding of low-level wind shear and turbulence. Wind shear has been assessed via computational fluid dynamics (CFD) simulations using both Reynolds Averaged Navier–Stokes (RANS) simulations and Large Eddy Simulations (LES). At Hong Kong Inter-

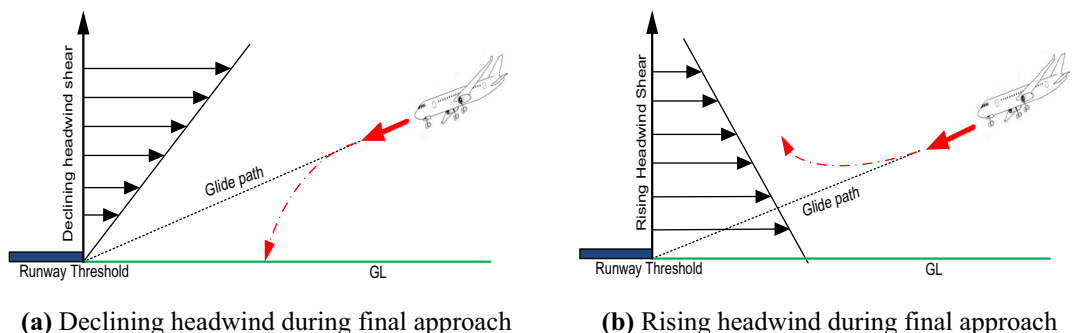


Figure 2. Detrimental Effects of wind shear on the approaching aircraft.

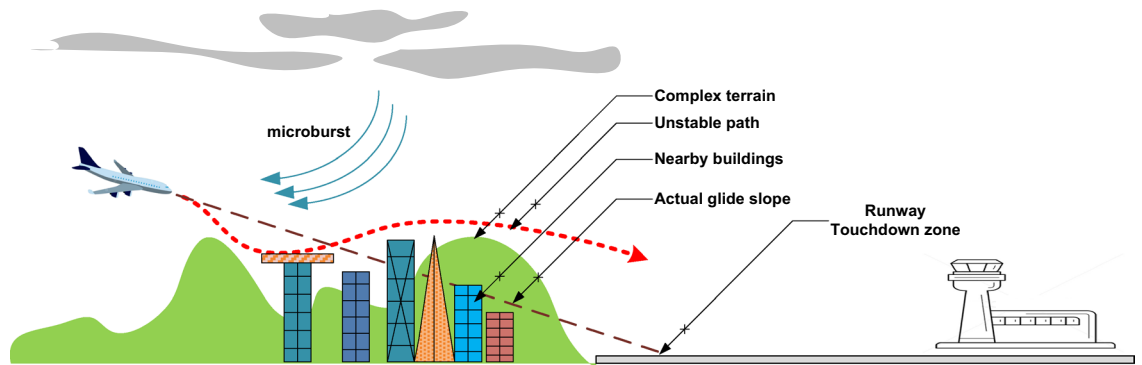


Figure 3. Wind Shear and Turbulence caused by nearby buildings and complex terrain.

national Airport (HKIA), Lei et al.¹⁷ used RANS equations and a LES model based on CFD to simulate terrain-induced wind shear. Similar to this, Chen et al.¹⁸ developed a high-resolution LES model using information from the Weather Research and Forecasting (WRF) model for the detection and analysis of terrain-induced low-level wind shear at HKIA. Boilley and Mahfouf¹⁹ performed numerical simulations with the nonhydrostatic Mesoscale (NH) model to predict the low-level wind shear over Nice airport in France. Rasheed and Sørli²⁰ employed CFD analysis for the terrain-induced turbulence at Kristiansand airport, Kjevik. Zhang et al.²¹ also developed CFD-based terrain-induced wind shear models for Beijing Capital International Airport (BCIA).

Additionally, researchers have used CFD to assess turbulence intensity. Both RANS simulation and LES were used to study the transient nature of terrain-induced flow disturbances along the glide paths of descending aircraft²². Similar to this, Shimoyama et al.²³ used LES to gain insight into the turbulence near Japan's Shonai Airport. According to the aforementioned studies, it has been found that the CFD model has precisely modeled wind shear and turbulence in the close vicinity of the airport. However, these studies were naturally constrained in their duration and/or scope by the exclusive use of simulation models. They employed the RANS equation, which can simulate and forecast the average airport wind field characteristics but cannot gauge the actual wind shear and turbulence intensity.

Applications of wind tunnel studies. The wind tunnel studies offer a viable alternative to numerical simulation models for the evaluation of wind shear and turbulence characteristics surrounding airports. Consequently, the results of wind tunnel experiments can serve as a crucial basis for determining the reliability of numerical simulations. Numerous scholars from various fields have conducted wind tunnel experiments to evaluate the impact of wind on bridges, wind turbines, low- and high-rise building structures, etc. For instance, Diana et al.²⁴ investigated the wind-bridge interaction by employing wind tunnel experiments and considering the aerodynamic phenomena that affect the various bridge components, primarily the deck and towers. He and Zou²⁵ implemented a wind tunnel experiment to examine the aerodynamic problems that arise with train-bridge systems in the presence of wind. Li et al.²⁶ studied the wind characteristics at the bridge site in the deep-cutting gorge using wind tunnel experiments. In the simulated atmospheric boundary layer, the effects of varying incoming wind directions on the wind characteristics at the bridge site were investigated.

Similarly, Su et al.²⁷ demonstrated the influence of array configuration on the power performance of vertical-axis wind turbines through wind tunnel experiments. Tip speed ratio, rotational direction, and position were considered, and the performance of various array configurations was compared. Ice buildup on the blade surface of wind turbines installed in cold, humid climates significantly reduces their performance. Gao et al.²⁸ conducted icing wind tunnel tests to investigate the characteristics of icing under various tip speed ratios and rime ice conditions. Likewise, using wind tunnel experiments, Cai et al.²⁹ evaluated the impact of collective pitch control and cyclic pitch control on the power and loads of floating offshore wind turbines under the condition of uniform wind velocity.

Li et al.³⁰ analyzed the effect of winds on the wooden pagoda in China by employing wind tunnel tests and computing the wind loads on the pagoda. One of the parts of low-rise buildings that is most susceptible to damage in windstorms is the overhang on the roof. Therefore, Wang et al.³¹ conducted field monitoring during typhoons and wind tunnel experiments to analyze the wind pressures on the upper and lower surfaces of the roof overhang in order to investigate the wind-induced pressure distributions and forces on the roof overhang of a typical low-rise building.

Despite the fact that a number of researchers have successfully employed wind tunnel experiments in their respective fields, the most serious drawback of wind tunnel investigations is the high cost of testing as well as the lack of testing infrastructure and time. To achieve the desired outcomes, numerous experiments are needed, carried out in many different kinds of settings. This reduces productivity and wastes both time and money. Empirical modeling techniques must replace experimental work in order to overcome the aforementioned limitations.

Combining machine learning with wind tunnel studies. In recent years, engineering has made great strides with machine learning strategies^{32–35}. This is a result of the growing demand for sophisticated computational methods to process enormous data sets. Several researchers have utilized it to integrate machine learning strategies with the wind tunnel outputs, such as Weng and Paal³⁶ developed a novel machine learning-based

wind pressure prediction model (ML-WPP) for non-isolated low-rise buildings by utilizing the non-isolated low-rise wind tunnel dataset. A machine learning-based predictive model of crosswind vibrations of rectangular cylinders was developed by Lin et al.³⁷. Building pressure patterns were identified using an unsupervised machine learning method by Kim et al.³⁸. Using the data collected from wind tunnel tests, deep learning methods to predict wind pressures on tall buildings were proposed³⁹. Similarly, Wada et al.⁴⁰ used deep reinforcement learning in conjunction with discrete actions to control the pitch of an aerial vehicle based on experiments in a wind tunnel.

Numerous applications have evidently led to the use of machine learning techniques to predict the wind-induced responses of structures. In contrast, its applicability to addressing the impact of wind field characteristics on airport runway glide paths is severely constrained.

Research process. This research combines wind tunnel experiments with machine learning models to estimate wind shear and turbulence intensity along the airport runway glide path in order to overcome the shortcomings of numerical simulations. Nevertheless, one of the most significant drawbacks of machine learning models is their black box nature. They required an explicit post-hoc explanation tool for the interpretation of the model and factors. In this study, we propose an explainable boosting machine (EBM) strategy⁴¹ combined with wind tunnel experiments to develop non-parametric models for estimating variation in headwind speed and turbulence intensity along the airport runway glide path. A scaled model of HKIA's north runway and Lantau Island (south of HKIA) was constructed in the TJ-3 atmospheric boundary layer (ABL) wind tunnel at Tongji University, Shanghai. Since HKIA is particularly susceptible to wind shear and turbulence due to its location and the topography of the area, an analysis of the airport's wind field is warranted^{11,18}. The north runway at HKIA is oriented as 25RA and 07LA; consequently, the glide paths of 25RA and 07LA were subject to different inflow directions. Along the glide paths of both 25RA and 07LA, measurements were made at fixed distances from the runway threshold. The EBM model was trained and evaluated based on the results of wind tunnel tests to estimate variations in headwind speed and turbulence intensity. In addition, the performance of the EBM model in estimating variations in headwind speed, crosswind speed, and turbulence intensity was compared to the performance of other glass box models, including decision tree (DT)⁴² and linear regression⁴³, as well as its counterpart black box models, Random Forest (RF)⁴⁴, Extreme Gradient Boosting (XGBoost)⁴⁵, Adaptive Boosting (AdaBoost)⁴⁶, and Extra Tree (ET)⁴⁷. The Bayesian optimization approach⁴⁸ was employed during the training-validation stage to fine-tune the hyperparameters of machine learning models. Afterwards, EBM was also used to investigate a number of factors and their global and local pairwise interactions. Figure 4 depicts the entire process of wind tunnel experiments combined with machine learning modeling.

Data preparation

Wind field effect at HKIA. The HKIA is located on the island of Lantau, which is subtropical terrain, off the southeast coast of the Chinese mainland as shown in Fig. 5⁴⁹. Numerous observational and modeling studies have shown that the HKIA's complex orography and high land-sea contrast are favorable for the emergence of low-level turbulence and wind shear^{50–52}. Pilot flight reports based on the HKIA indicate that nearly 1 in 500 flights have been subjected to wind shear and turbulence since the airport's opening. 97% of the pilot reports revealed 15 to 25 knots of low-level wind shear. Terrain-induced wind shear was to blame for about 70% of the wind shear that pilots reported⁵³. In addition to the terrain, nearby buildings are also significant sources of wind shear and turbulence⁵⁴, as illustrated in Fig. 6⁵⁵.

The wind tunnel experiments. This study employed wind tunnel experiments to evaluate the turbulence intensity and variations in headwind speeds along the glide path of the north runway at HKIA under various inflow wind conditions. At the State Key Laboratory for Disaster Reduction in Civil Engineering at Tongji University in Shanghai, experiments were carried out in the TJ-3 ABL wind tunnel. This wind tunnel is a low-speed, closed-circuit, vertical-return type. The wind tunnel's test chamber has dimensions of 14 m long, 15 m wide, and 2 m high. For the wind tunnel testing, Lantau Island, the north runway of the HKIA, and the buildings to the south of the runway, with a test range of 27.2 km and an average height of 425.2 m, were taken into consideration. The Lantau island and building were included in the 1:4000 scale model with a diameter of 6.8 m that was used in the wind tunnel for testing purposes, as shown in Fig. 7a. The height within the test range was 0.106 m.

The surrounding terrain model was constructed layer by layer using dense foam with a one-inch texture (equivalent to a difference in actual terrain height of 40 m). The wind tunnel blockage ratio was calculated to be 2.402%, which is less than 5% (preferably for wind tunnel research) and satisfied the requirements of the wind tunnel experiments. The wind direction was varied from 90 to 240 degrees in 15 degree increments, taking into account the typical airflow patterns of east to southeasterly winds and the southwest monsoon of Hong Kong⁵⁶, as shown in Fig. 7b. Therefore, eleven different wind conditions were in operation. The wind directions were given as follows: 0 degrees for the north wind, 90 degrees for the east wind, 180 degrees for the south wind, and 270 degrees for the west wind. Given that aircraft usually perform headwind landings, the intervention of a southwest wind (with wind directions of 190, 210, 225, and 240 degrees) was primarily taken into account for the glide slope of Runway 25RA, and the intervention of an east-southeast wind (with wind directions of 190, 210, 225, and 240 degrees) was taken into consideration for the glide slope of Runway 07LA.

The pink dotted line in Fig. 7b shows that aircraft typically follow a glide slope of 3 degrees in the final three nautical miles before landing on the runway⁵⁷. In wind tunnel experiments, two sets of a total of eight measurement points (a1, a2, a3, a4) and (b1, b2, b3, b4) were placed along the glide paths of runways 07LA and 25RA, respectively. The cobra probes were mounted on bespoke stands, and the installation's overall height was adjusted so that it was in line with the height of the measurement site. The probes were oriented toward the inflow

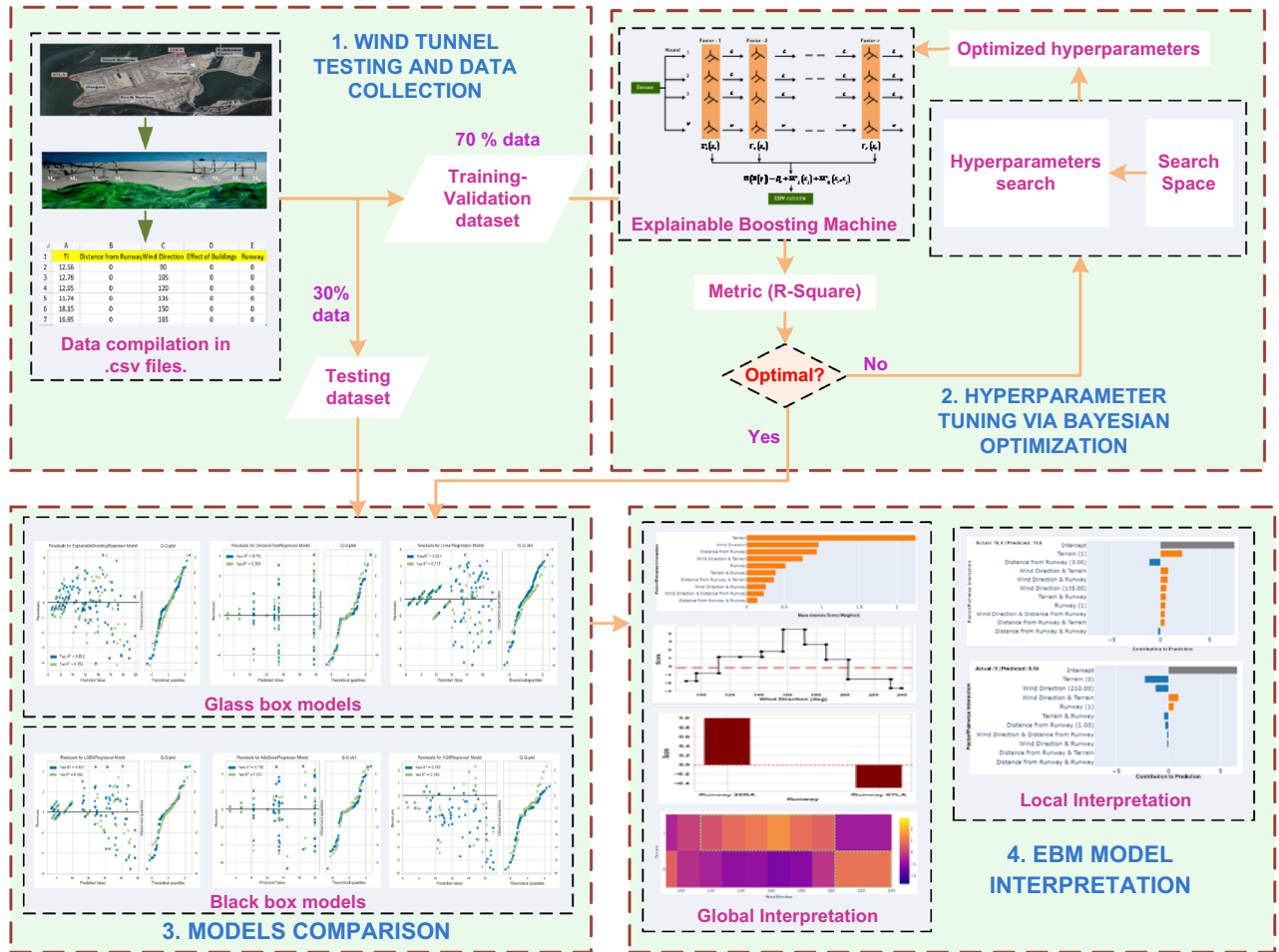


Figure 4. Schematic diagram of wind tunnel experiments combined with machine learning modeling.

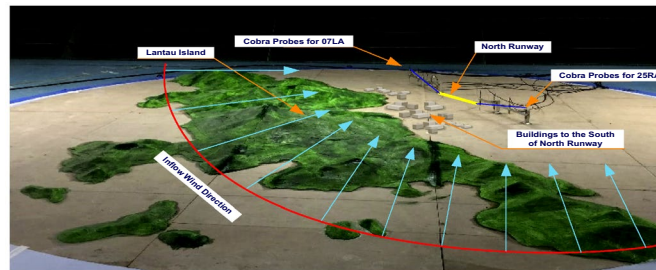


Figure 5. HKIA and Lantau Island.

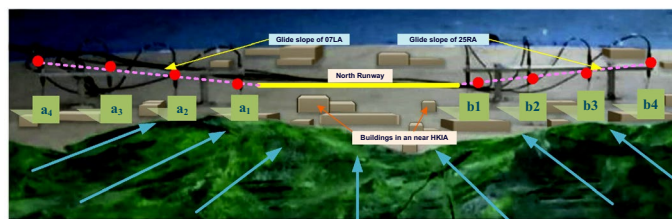
direction after each test of operational conditions. Each working condition was sampled for a total of 65,536 s at a single-point sampling rate of 1000 Hz.



Figures 6. Buildings near and at HKIA.



(a)



(b)

Figure 7. Scaled model of HKIA and Lantau Island.

Development of explainable boosting machine model

An explainable boosting machine (EBM) model was developed to estimate the turbulence intensity and variation in headwind along the airport runway glide slope using data from wind tunnel experiments. The EBM model is an interpretable model, in contrast to other machine learning models. When implementing machine learning models, interpretability can be crucial. By interpreting models, one can gain confidence in them and facilitate their adoption. It may also be useful for debugging the model, and in some cases, explanations for the model’s predictions are required. Figure 8 depicts the EBM model, with factors serving as inputs and outputs. Prior to developing an EBM model, it is necessary to perform label encoding. The factors, which include the effect of buildings at or near HKIA, the orientation of the approach runway, and the distance from the runway, are coded as shown in Table 1.

Fundamental concept of EBM model. The EBM algorithm is based on GAMs, which are universally acknowledged to be the gold standard for comprehensibility. Given that $\Delta = (x_j, y_j)$ is a training dataset with a length of R , that (x_1, x_2, \dots, x_R) are the input vectors with v number of factors, and that y_r is the target factor, then the GAMs takes the form as shown in Eq. (1).

$$\Theta(E[y]) = \beta_0 + \sum \Gamma_j(x_j) \tag{1}$$

where, x_j indicates the j th factor in the factors set, Θ is the link function that conforms the GAM to regression (e.g., $\Theta = \text{identity}$) or classification (e.g., $\Theta = \text{logistic}$), and Γ_r is the attribute function. EBM, which employs bagging and gradient boosting, provides several significant advantages over conventional GAMs. An adequate description would be bagged, augmented, and lowered shallow trees. As depicted in the Fig. 9, the core of the algorithm employs boosted trees.

In EBM, training is carried out repeatedly, with each repetition involving the creation of a distinct boosting procedure for each factor. The factor’s order is irrelevant because of the low learning rates used in this process. In order to determine the best way to determine the contribution of each factor to the model’s prediction, the high number of iterations aims to reduce the effects of co-linearity. Additionally, pairwise interaction terms

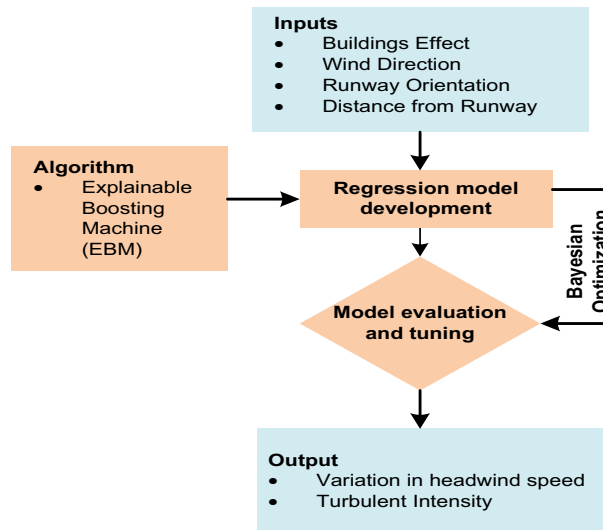


Figure 8. Flow chart of EBM modeling process with with input and output factors.

Factors	Data type	Coding
Turbulence Intensity	Continuous	-
Wind shear (change in the headwind speed)	Continuous	-
Buildings effect	Discrete	1: If the effects of buildings at and near HKIA are taken into account 0: If the effects of buildings at and near HKIA are ignored
Wind direction	Continuous	-
Runway orientation	discrete	1: When using the approaching Runway 25RA's glide slope 0: When using the approaching Runway 07LA's glide slope
Distance from runway	Discrete	0: When the distance between the aircraft and the end of the approaching runway is 0.25 nautical miles (0.25MF) 1: When the distance between the aircraft and the end of the approaching runway is 0.75 nautical miles (0.75MF) 2: When the distance between the aircraft and the end of the approaching runway is 1.25 nautical miles (1.25MF) 3: When the distance between the aircraft and the end of the approaching runway is 1.75 nautical miles (1.75MF)

Table 1. Label encoding of the factors.

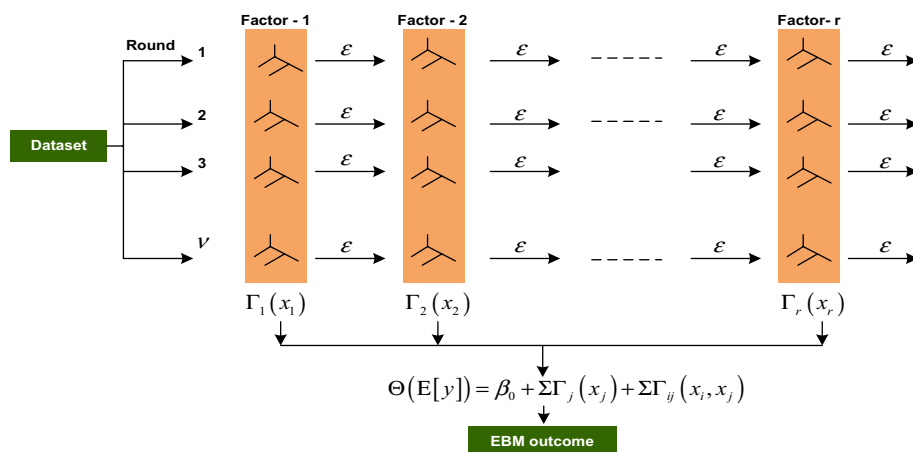


Figure 9. Working principle of EBM algorithm.

can be automatically detected and included by EBM, improving the model's accuracy while maintaining its explainability. The fact that EBM is an additive model allows for the capture and visualization of each factor's contribution, which improves explainability.

Small trees are built sequentially for each iteration of the model, each of which can only use one factor. In a boosting manner, the residual (ε) is updated and a new tree is built using a different factor. Every iteration involves doing this for every factor. When training is finished, we can view all the trees that a particular factor used to build them. Due to the concepts of additivity and modularity, it is feasible to order and visualize the contributions to determine which factor had the greatest influence on the individual prediction. Another advantage of the EBM is the ability to improve accuracy by incorporating pairwise interactions into standard GAMs that becomes GA2Ms with the form and given as Eq. (2).

$$\Theta(E[y]) = \beta_0 + \sum \Gamma_r(x_r) + \sum \Gamma_{rs}(x_r, x_s) \quad (2)$$

Here, the two-dimensional interaction $\Gamma_{rs}(x_r, x_s)$ can be depicted as a heat map on the two-dimensional $x_r - x_s$ plane while still retaining an exceptionally high degree of intelligibility. Initially, GA2M constructs the best GAM possible, and then it searches through the residuals to find all of the possible combinations of interactions, and then it ranks them based on how crucial they are. The last step is to integrate into the model the κ most significant pairwise interactions, where κ is a value determined through the process of cross-validation.

Hyperparameter tuning of EBM via Bayesian optimization. In this study, the process of fine-tuning the hyperparameters of EBM model as well as other counterpart black box and other glass box model was carried out with the aid of a Bayesian optimization strategy. It is one of the algorithms for global optimization, and it finds widespread application in the field of engineering^{58,59}. The objective function that needs to be optimized in order to find the optimal values for the hyperparameters of the model can be expressed as Eq. (3).

$$x^\ominus = \arg \max_{x \in \Lambda_s} \varphi(x) \quad (3)$$

where, x : Hyperparameters of the model, Λ_s : Hyperparameters search space, $\varphi(x)$: Objective function that reflects the association between models' performance and hyperparameters.

In this study, we employed R^2 as the evaluation metric. The purpose of this optimization is to identify the appropriate combination of hyperparameters (x^\ominus) in such a way that the performance of the model $\varphi(x)$ is increased to its full potential. The concepts of the Bayesian theorem are utilized in Bayesian optimization, which includes the following [Eq. (4)].

$$\rho(\varphi|T) = \frac{\rho(T|\varphi)\rho(\varphi)}{\rho(T)} \quad (4)$$

where, φ : Black box function, $\rho(\varphi)$: Prior probability of φ , $\rho(T|\varphi)$: Probability obtained from the observed point, $\rho(T)$: Normalized constant, $\rho(\varphi|T)$: Given the current observation point T, the posterior probability of φ .

If the Gaussian process is used, then the black box function φ will behave in a way that is consistent with the Gaussian distribution. At this very moment, the acquisition function $\Upsilon(x)$, which is based on the expected improvement, looks like this:

$$\Upsilon(x) = \begin{cases} (v(x) - \varphi^+) \Lambda(Z) + \alpha(x)\beta(Z) & \alpha(x) > 0 \\ \max(0, v(x) - \varphi^+) & \alpha(x) = 0 \end{cases} \quad (5)$$

$$Z = \frac{v(x) - \varphi^+}{\alpha(x)} \quad (6)$$

where, Λ : Cumulative distribution function of Gaussian distribution, β : PDF of Gaussian distribution

The process of Bayesian optimization of hyperparameters, the iterative search is as follow:

Iterate the loop ω times ($\omega = 1, 2, \dots$):

1. In accordance to the acquisition function $\Upsilon(x)$ computed by Eq. (4), the hyperparameters combination $x_{\omega+1} = \arg \max \Upsilon(x; T)$ of the next cluster of instances is obtained.
2. Corresponds to an instance $x_{\omega+1}$, compute the performance $\varphi_{\omega+1}$ of the machine learning model.
3. Update $x_{\omega+1}$ to the observed point $T_{1:\omega+1} = \{T_{1:\omega}, (x_{\omega+1}, \varphi_{\omega+1})\}$;
4. The posterior distribution of φ is then updated in accordance to $\rho(\varphi|T) = \frac{\rho(T|\varphi)\rho(\varphi)}{\rho(T)}$

When the loop terminates, the final obtained combination of hyperparameters is the output of Bayesian optimization.

Performance measures. To assess the performance of different models, four distinct metrics could be utilized, i.e., mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and Coefficient of determination (R^2). The MAE is the average absolute value of the individual prediction errors across all instances [Eq. (7)]. MSE is calculated using the average squared difference between observed and predicted values, as shown in Eq. (8). RMSE is the square root of the difference between the observed and predicted

values, according to Eq. (9). R^2 which ranges from 0 to 1, indicates the ability of a model to accurately predict values. R^2 is given by Eq. (10).

$$MAE = \sum_{x=1}^{\zeta} \frac{|\sigma_x - \bar{\sigma}_x|}{\zeta} \quad (7)$$

$$MSE = \frac{1}{\zeta} \sum_{x=1}^{\zeta} (\sigma_x - \bar{\sigma}_x)^2 \quad (8)$$

$$RMSE = \sqrt{\sum_{x=1}^{\zeta} \frac{(\sigma_x - \bar{\sigma}_x)^2}{\zeta}} \quad (9)$$

$$R^2 = 1 - \frac{\sum_{x=1}^{\zeta} (\sigma_x - \bar{\sigma}_x)^2}{\sum_{x=1}^{\zeta} (\sigma_x - \sigma_{avg})^2} \quad (10)$$

where, the total number of data points are denoted by ζ , the actual experimental outcomes are represented by σ_x , the predicted outcomes are represented by $\bar{\sigma}_x$ and σ_{avg} represents the average of predicted outcomes.

Results and discussion

This section presents the development of the EBM model, other glass box models, such as DT and linear regression, and their counterpart black box models, such as RF, XGBoost, AdaBoost, and ET, to estimate variation in headwind speed and turbulence intensity based on data from wind tunnel experiments. In order to achieve better performance, it is crucial to note that the hyperparameters must be optimized prior to the development of these models. To achieve this, the results of the wind tunnel experiments were split into a training-validation set (comprising 70% of the total) and a testing set (comprising 30% of the total)⁶⁰. The development of machine learning models and the tuning of their hyperparameters through Bayesian optimization were both done using the training-validation data. Bayesian optimization aimed to find the set of hyperparameters for various machine learning models that maximizes the R^2 value across a given sample space. Figures 10 and 11 show the progress of the Bayesian optimization process over 50 iterations using both glass box and black box models for variations in headwind speed and turbulence intensity, respectively. They were tuned using the hyperparameters associated with the models at the best R^2 to provide the best prediction model. Table 2 shows the optimal values for the hyperparameters that were obtained for the estimation of variation in headwind speeds, while Table 3 shows the optimal hyperparameters for models that estimate turbulence intensity.

Prediction of variation in headwind speed and turbulence intensity. Once the optimal hyperparameters were determined, holdout or testing data was used to evaluate and compare the performance of various models. Importantly, the dataset was partitioned after being shuffled with 40% and 50% test data, and this demonstrated that the metrics values for model testing remained constant within the 95% confidence interval. In addition to tuning hyperparameters, this was performed to prevent over-fitting. When the size of the test data was altered, no anomalies in the performance metrics were observed. Table 4 presents the performance metrics of various models for estimating the variation in headwind speeds using both the training and test datasets. In terms of MAE (0.346), MSE (0.189), RMSE (0.435), and R^2 (0.928) for the training dataset and MAE (0.571),

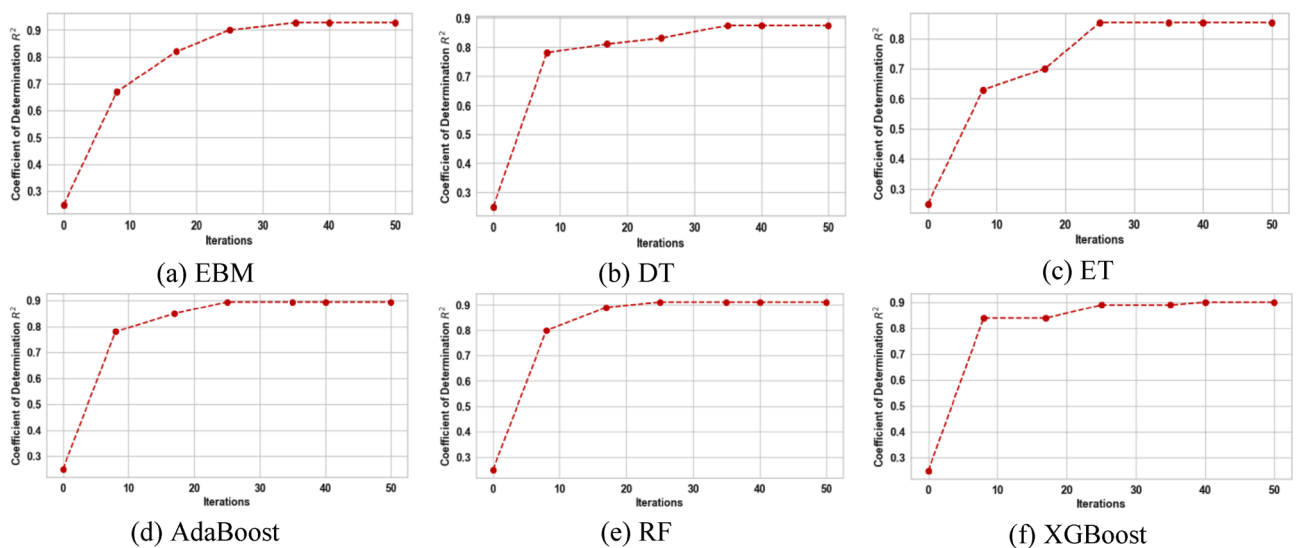


Figure 10. Progress of Bayesian optimization process for variation in headwind speed estimation.

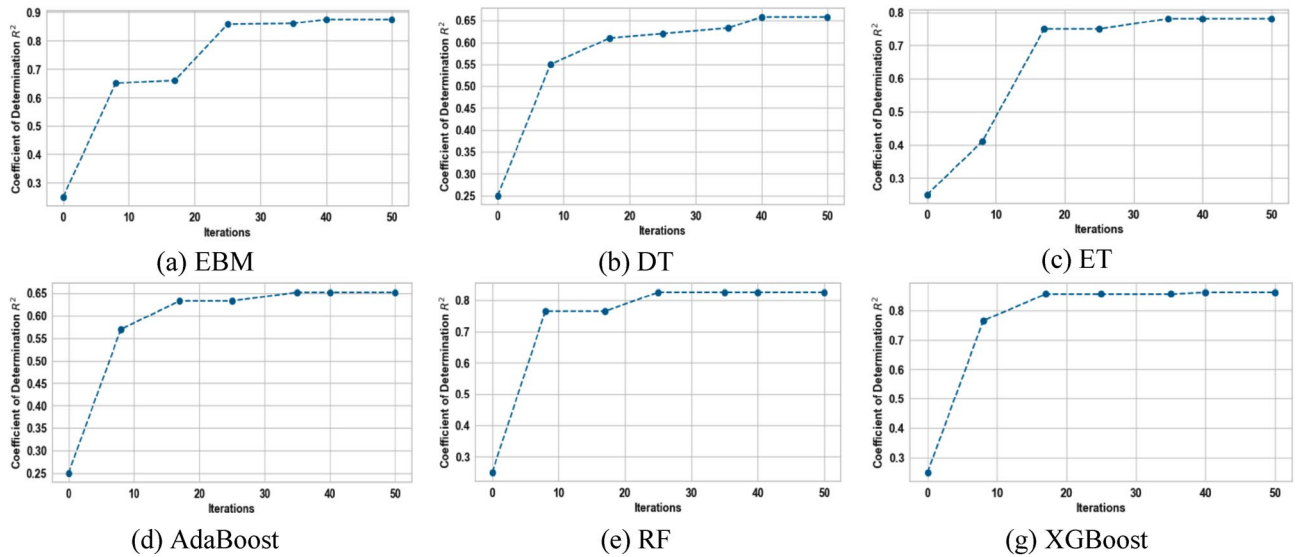


Figure 11. Progress of Bayesian optimization process for the estimation of turbulence intensity.

Model	Algorithm	Hyperparameters	Range	Optimal values
Glass box	EBM	learning_rate	(0.01, 0.5)	0.35
		min_sample_leaf	(3, 10)	5
	DT	max_depth	(2, 10)	3
Black box	ET	max_depth	(2, 10)	4
		n_estimators	(10, 100)	50
	RF	max_depth	(2, 10)	3
		n_estimators	(10, 100)	62
	XGBoost	n_estimators	(10, 100)	38
		learning_rate	(0.01, 0.5)	0.18
	AdaBoost	n_estimators	(10, 100)	25
learning_rate		(0.01, 0.5)	0.19	

Table 2. Optimal hyperparameters for the estimation of variation in the headwind speeds.

Model	Algorithm	Hyperparameters	Range	Optimal values
Glass box	EBM	learning_rate	(0.01, 0.5)	0.42
		min_sample_leaf	(3, 10)	4
	DT	max_depth	(2, 10)	3
Black box	ET	max_depth	(2, 10)	3
		n_estimators	(10, 100)	25
	RF	max_depth	(2, 10)	4
		n_estimators	(10, 100)	20
	XGBoost	n_estimators	(10, 100)	17
		learning_rate	(0.01, 0.5)	0.19
	AdaBoost	n_estimators	(10, 100)	15
learning_rate		(0.01, 0.5)	0.14	

Table 3. Optimal hyperparameters for the estimation of turbulence intensity.

Transparency	Models	Training dataset				Testing dataset			
		MAE	MSE	RMSE	R ²	MAE	MSE	RMSE	R ²
Glass box	EBM	0.346	0.189	0.435	0.928	0.571	0.466	0.682	0.839
	LR	2.861	10.965	3.316	0.268	2.475	9.716	3.116	0.370
	DT	0.424	0.332	0.579	0.874	0.925	1.194	1.091	0.589
Black box	ET	0.495	0.390	0.625	0.852	0.858	0.984	0.994	0.659
	AdaBoost	0.388	0.284	0.0535	0.894	0.931	1.210	1.105	0.584
	RF	0.336	0.152	0.396	0.911	0.828	1.074	1.037	0.630
	XGBoost	0.353	0.250	0.504	0.905	0.830	0.926	0.964	0.680

Table 4. Performance measures of different models for estimating variation in headwind speeds.

MSE (0.466), RMSE (0.682), and R² (0.839) for the testing dataset, the EBM model performed better than other glass box models. Similarly, the XGBoost model outperformed other black box models using testing data, with an MAE of 0.830, an MSE of 0.926, an RMSE of 0.964, and an R² of 0.680. Overall, the EBM model performed better than its black-box counterpart. The relative percent difference between MAE of EBM and its counterpart XGBoost on the testing dataset was 36.974%, MSE was 66.092%, RMSE was 32.265%, and R² was 21.053%.

Table 5 enumerates the outcomes of various models' attempts to estimate the turbulence intensity for the training and test datasets. The EBM model performed better compared to both the black box and the glass box models, with an MAE of 1.705, MSE of 3.994, RMSE of 1.993, and R² of 0.741, using the testing dataset. The LR model performed the worst overall, with MAE, MSE, RMSE, and R² values of 2.516, 9.664, 3.109, and 0.209, respectively.

Similarly, in the case of both variations in headwind speed and turbulence intensity, we also reported and compared the prediction confidence and their ability to extrapolate on unseen data for both black box and glass box models using a residual scatter plot and a residual QQ (quantile–quantile) plot, as shown in Figs. 12 and 13. By analyzing the residual scatter plot's residual patterns, the model's applicability is shown. The discrepancy between the predicted and observed values is represented by residues. In a residual scatter plot, the target factor values are plotted on the horizontal axis, and the residuals are plotted on the vertical axis. When using the EBM model to analyze residuals for variations in headwind speed as well as turbulence intensity, in contrast to other models, the points are closer to the horizontal line at 0.00.

EBM interpretation. Based on its performance metrics utilizing a testing dataset, EBM was inevitably the best model. From the EBM model, we can glean explanations. Consequently, we next utilized the EBM model to illustrate global and local interpretations of various individual and pairwise factor interactions.

Global importance of factors and pairwise interaction. The EBM algorithm is a generalized additivity model that enhances and expands the tree-based model. Due to the persistence of additivity, the contributions of factors can be ranked and plotted to demonstrate impact prediction in both global and local contexts. The exhaustive global explanation of the EBM makes it possible to visualize the effect of each possible combination of factors on the variation in headwind speed and turbulence intensity.

Figure 14a and b provides a summary of the global significance of each individual factor and their pairwise interactions for variation in headwind speed and turbulence intensity, respectively. Wind direction contributed the most to the variation in headwind speed, followed by Runway. In pairwise interaction, the most significant contributor to the variation in headwind speed was the combination of Distance from Runway and Runway. Similarly, wind direction contributed the most to turbulence intensity, followed by runway length and distance from the runway. The combination of wind direction and runway contributed the most to the occurrence of intense turbulence in pairwise interaction. These results are also inline with the previous study^{54,61}

Additional output provided by the proposed EBM model for comprehending the global findings for the variation in headwind speed and turbulence intensity is shown as examples in Figs. 15 and 16, respectively. These

Transparency	Models	Training dataset				Testing dataset			
		MAE	MSE	RMSE	R ²	MAE	MSE	RMSE	R ²
Glass box	EBM	1.051	1.887	1.373	0.874	1.705	3.991	1.993	0.741
	LR	2.901	11.056	3.328	0.317	2.516	9.668	3.109	0.206
	DT	1.807	5.533	2.352	0.658	2.237	7.184	2.685	0.410
Black box	ET	1.442	3.253	1.806	0.781	1.755	4.665	2.158	0.695
	AdaBoost	1.925	5.626	2.372	0.652	1.726	4.694	2.166	0.614
	RF	1.178	2.615	1.6171	0.825	1.609	4.579	2.132	0.706
	XGBoost	1.115	2.246	1.497	0.861	1.490	3.586	1.894	0.705

Table 5. Performance measures of different models for the estimation of turbulence intensity.

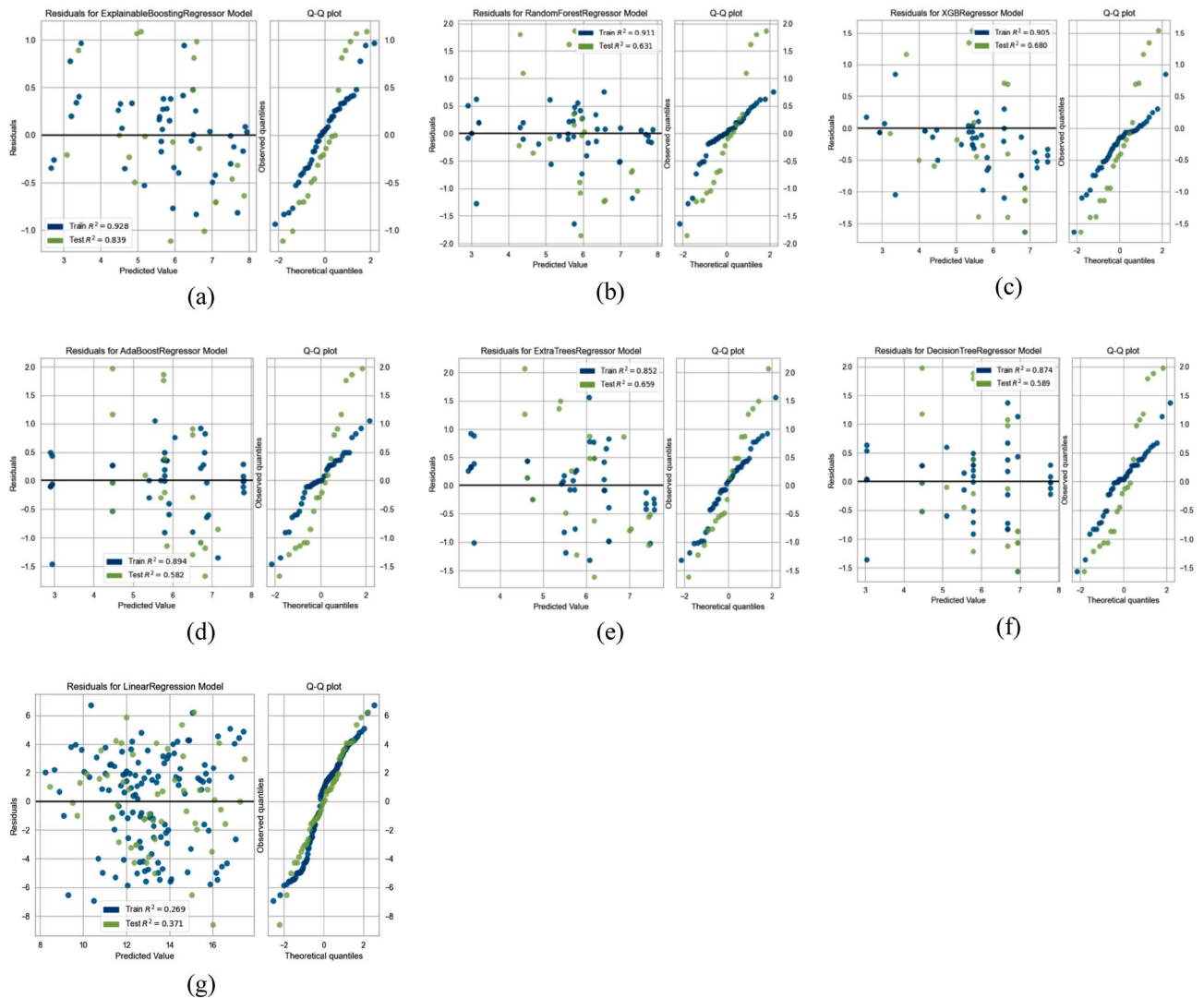


Figure 12. Residual analysis in case of variation in headwind speeds; Glass box models (a–c), Black box models (d–g).

include EBM shape functions and a heat map. For each predictor factor, the gradient boosting procedure yields a large number of decision trees, which are then used to generate the function characterizing the dependent factor's response to the specific predictor factor. One can then graphically represent these functions as a one-dimensional function, with the values of the predictor factors along the horizontal axis and the scores (i.e., the effect of the predictor factor on the prediction) along the vertical axis. When the score is greater than 0, it means that the value of the predictor factor is strongly linked to the outcome. In the case of the variation in headwind speed, Fig. 14a–c shows EBM shape functions and a heat map for the important individual and pairwise interaction factors.

Figure 15a demonstrates that the variation in headwind speed is greatest when the wind direction is between 90° and 135° , as well as between 200° and 240° . The headwind speed along the glide path of Runway 25RA significantly decreases due to the presence of buildings at the airport that are in its path. Runway 25RA persists in being affected by a wind coming from the southwest at an angle ranging from 195° to 240° . Compared to Runway 07LA, Runway 25RA is more likely to experience a greater variation in headwind speed (Fig. 15b). This may be due to the presence of buildings close to Runway 25RA that cause variations in wind speed.

The pairwise interactions between runway and distance from runway are depicted in Fig. 15c. The heat map revealed that the variation in headwind speed is greatest within 0.5 miles of Runway 25RA. Consequently, shorter distances along Runway 25RA have been more susceptible to variations in headwind speed, and pilots may encounter severe wind shear in this area. Pilots must be extremely vigilant during the final approach to Runway 25RA.

Figure 16a–c displays EBM shape functions and a heat map for the crucial individual and pairwise interaction factors in the case of the turbulence intensity. The higher turbulence intensity is more likely to occur for wind directions between 110° and 200° , as shown in Fig. 16a. Complex terrain in this direction blocks the wind, causing it to fluctuate in speed, which ultimately increases the intensity of the turbulence. In contrast to Runway 07LA, Runway 25RA is more likely to encounter higher turbulence intensity than Runway 07LA (Fig. 16b). This may be caused by the presence of building structures close to Runway 25RA that cause wind speed variations. This

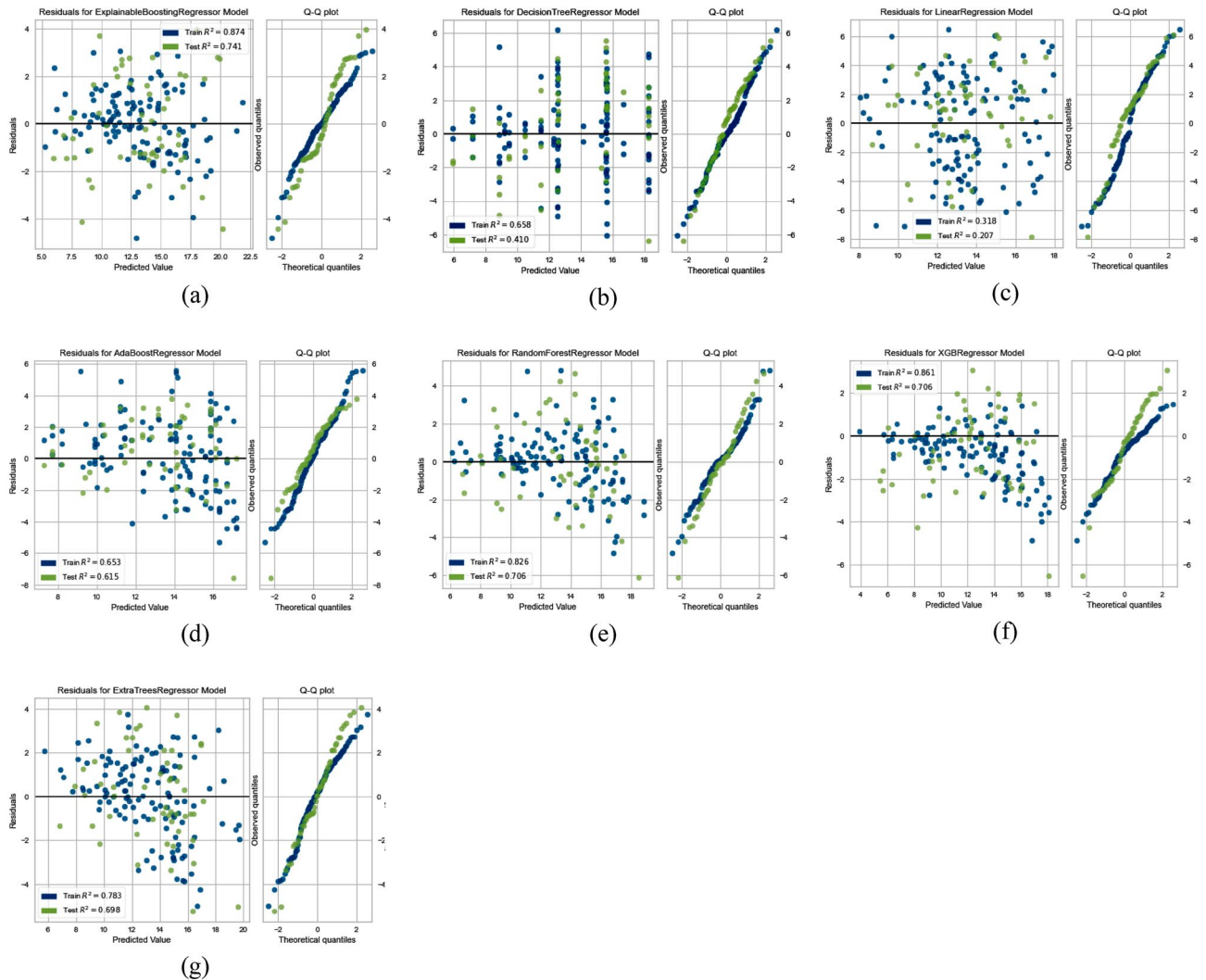


Figure 13. Residual analysis in case of turbulence intensity; Glass box models (a–c), Black box models (d–g).

suggests that nearby structures are restricting the wind close to the runway⁶². The heat map in Fig. 16c shows how the runway and wind direction interact pairwise. It demonstrates that Runway 25RA is most likely to experience high turbulence intensity when the wind direction is between 170 and 220 degrees and that Runway 07LA is most likely to experience high turbulence intensity when the wind direction is between 130 and 170 degrees.

Local factor interpretation. The local explanation, which includes the explanations behind particular predictions in a single case, can be investigated for both variations in headwind speeds and turbulence intensity. The intercept is shown to be constant in gray, while additive terms that have a positive effect on output are shown in orange, and additive terms that have a negative effect on output are shown in blue, demonstrating local factor contribution assessments for each instance. The estimates for each instance are derived from the shape functions and pairwise interactions (heat map) based on the input values. We take into account two arbitrary cases where the predictions are very close to actual values, each for variations in headwind speed and turbulence intensity. Nevertheless, we can perform local interpretation in other instances as well.

The local explanation for a variation in headwind speed prediction is shown in Fig. 17a. It demonstrates that the variation in mean headwind speed was high for a randomly chosen instance at Runway 25RA and a shorter distance from the runway threshold. Similarly, Fig. 17b shows the degree of turbulence for a randomly selected instance, showing that a wind direction of 240 degrees has a negative effect on the turbulence intensity. However, a closer proximity to the runway, the presence of buildings, and Runway 25RA all increase the likelihood of experiencing high turbulence.

Conclusion and future work

In order to estimate the variation in headwind speed and turbulence intensity along the airport runway glide slope based on wind tunnel experiments, this study employs an explainable boosting machine (EBM), a state-of-the-art glass box model. The proposed approach is both explicable and intuitive, and it provides accuracy on

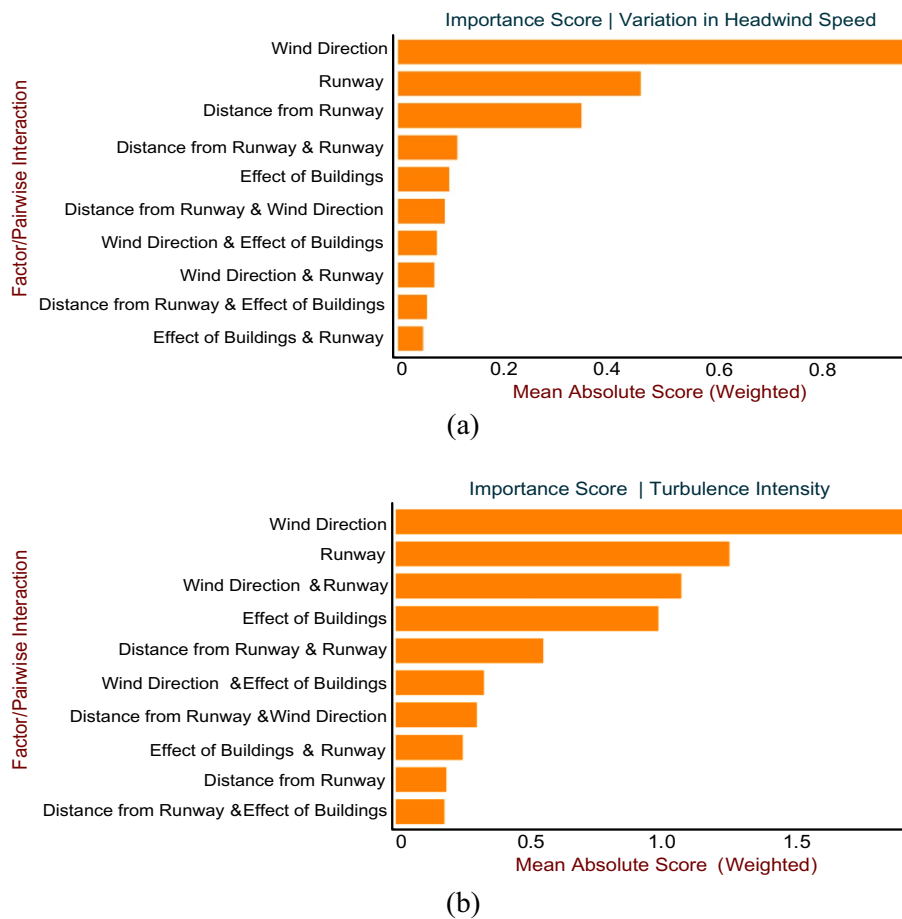


Figure 14. Global Importance of factors and their pairwise interaction; (a) In case of variation in headwind speed; (b) In case of turbulence intensity.

par with black box competitors. Conventional machine learning regression models lack explainability, but the EBM model has been effective in overcoming this limitation. Therefore, the following conclusions can be drawn:

- Based on the testing data set for the variation in headwind speed, the Bayesian optimized EBM model had the best overall performance of all the black box and glass box models, with an MAE of 0.571, an MSE of 0.466, an RMSE of 0.682, and an R^2 of 0.839. Similarly, in the case of holdout dataset turbulence intensity, the Bayesian optimized EBM model outperformed other models with an MAE (1.705), MSE (3.991), RMSE (1.993), and R^2 (0.741).
- The LR model demonstrated the worst performance with MAE (2.475), MSE (9.716), RMSE (3.116) and R^2 (0.370) in case of variation in headwind speed estimation as well as in estimating turbulence intensity with MAE (2.516), MSE (9.668), RMSE (3.109) and R^2 (0.206).
- In the case of individual factors for the variation in headwind speed, EBM-based global interpretation revealed that “wind direction” and “Runway” significantly contributed to the occurrence of high variation in headwind speed. In the case of pairwise interaction, it was found that the combination of “distance from runway” and “runway” contributed more to the high variation in headwind speed.
- The EBM-based interpretation also demonstrated that, in the case of individual factors, “Wind Direction” and “Runway” significantly influenced the occurrence of high turbulence intensity. It was found that the combination of “Wind Direction” and “Runway” contributed more to the high turbulence intensity in the case of pairwise interaction.

This proposed EBM framework could be applied to a comprehensive analysis of wind field characteristics at other airports as well. It is indeed a great resource for individuals associated with civil aviation. However, there are a few recommendations for future research.

- Although this study utilized a number of different input parameters to predict wind field characteristics along the airport runway glide slope, many other parameters, such as atmospheric pressure and temperature, could be considered in future studies.

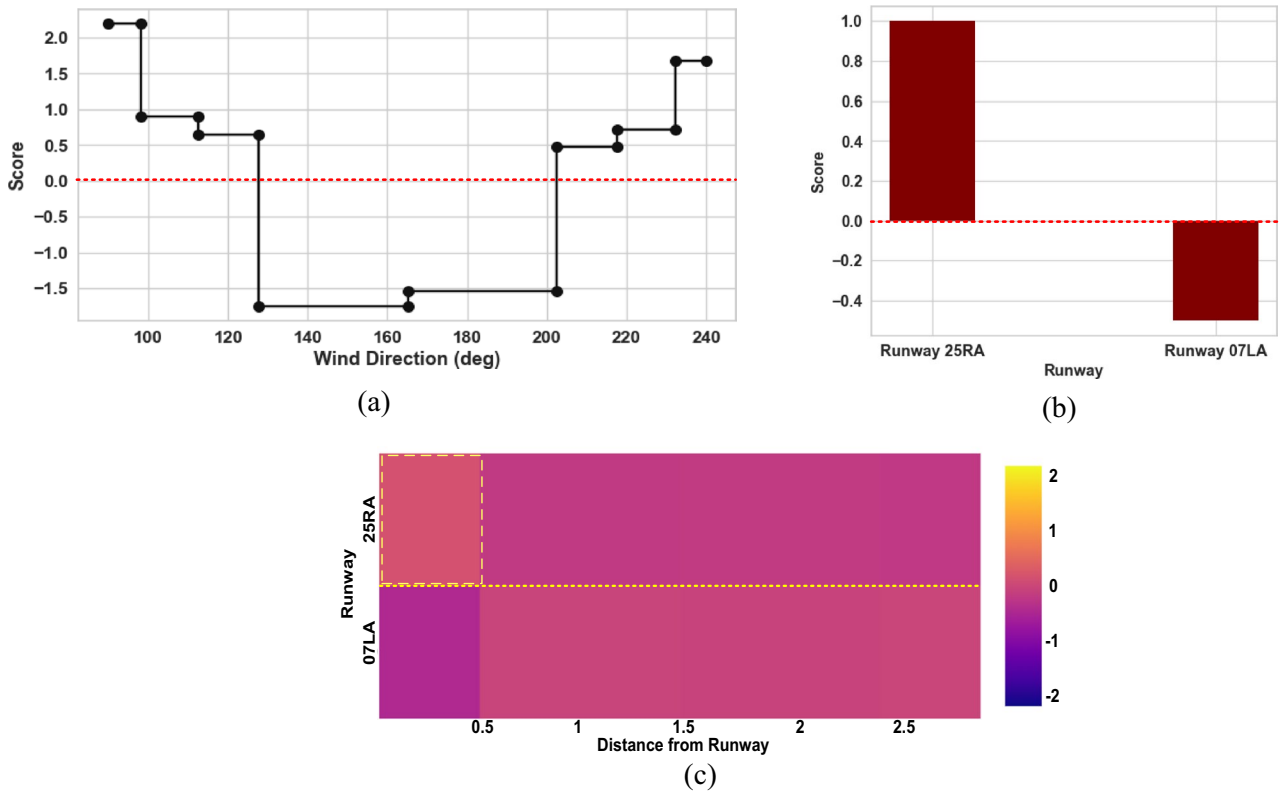


Figure 15. Shape function and heat map obtained via EBM model for mean wind speed; (a) shape function for Distance from Runway factor; (b) shape function for wind direction; (c) heat map for pairwise interaction of Runway and Distance from Runway factors.

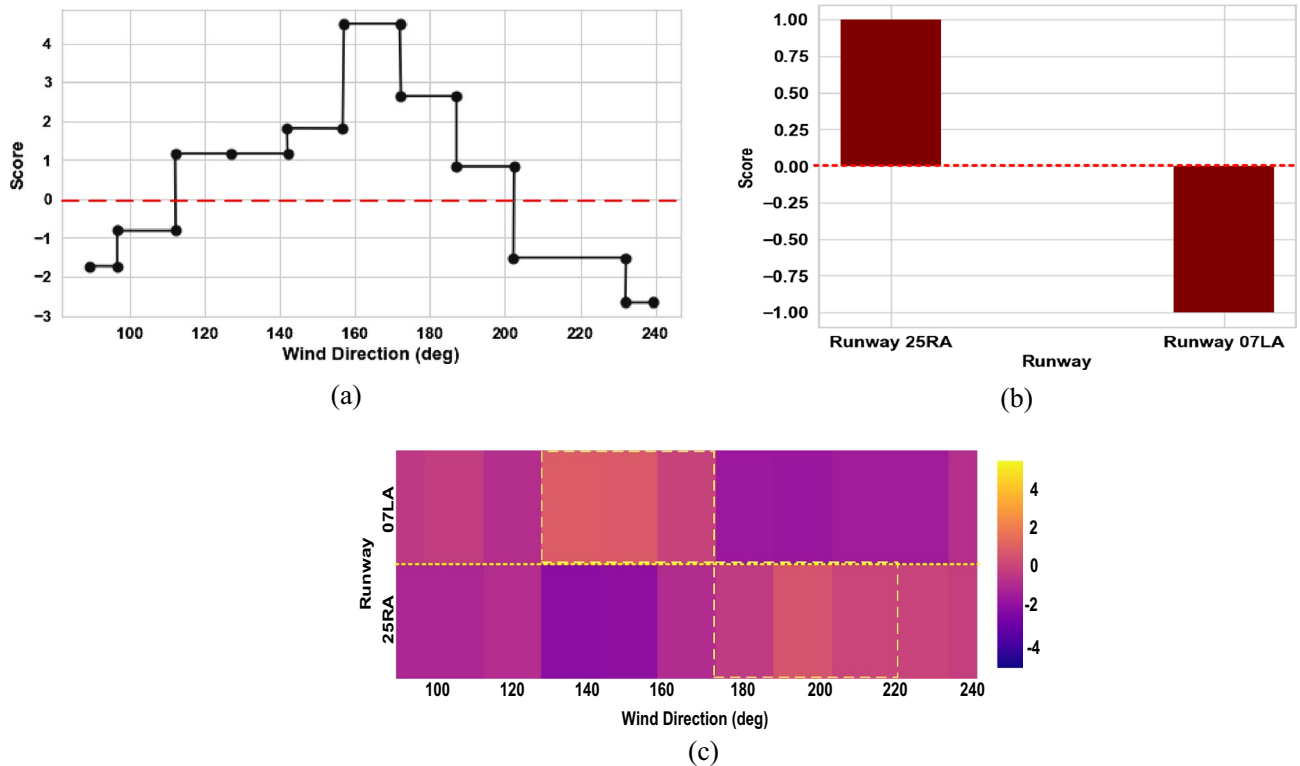


Figure 16. Shape function and heat map obtained via EBM model for turbulence intensity; (a) shape function for wind direction factor; (b) shape function for Runway; (c) heat map for pairwise interaction of Runway and wind direction factors.

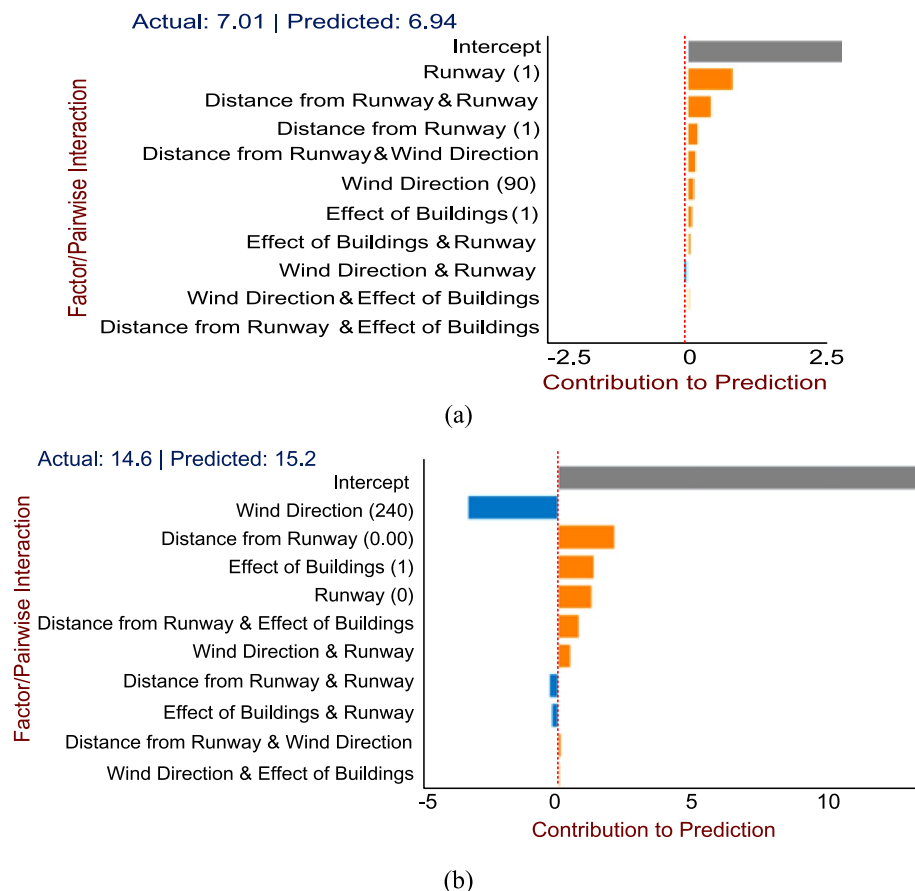


Figure 17. Local Interpretation for randomly selected instances; (a) In case of variation in headwind speed; (b) In case of turbulence intensity.

- In this study, the variation in headwind speed and turbulence intensity along the airport runway glide slopes were the parameters of concern. Similarly, the turbulence integral length scale as well as the variation in crosswind speeds at different points on the glide path can also be considered for future work.

Data availability

The datasets used and analyzed in the current study is available from the corresponding author on reasonable request.

Received: 20 March 2023; Accepted: 5 June 2023

Published online: 06 July 2023

References

1. Sadraey, M. H. *Aircraft Performance: An Engineering Approach* (CRC Press, 2017).
2. Gao, Z. X. & Li, Z. Q. Predictive controller for aircraft landing under wind shear. In *Applied Mechanics and Materials* (vol. 587, 2030–2035). (Trans Tech Publications Ltd., 2014).
3. Chan, P. W. & Zhang, P. *Aviation Applications of Doppler Radars in the Alerting of Wind shear and Turbulence* (IntechOpen, 2012).
4. International Civil Aviation Organization (ICAO). *ICAO Safety Report 2017 Edition* (2017).
5. Bretschneider, L., Hankers, R., Schönhals, S., Heimann, J. M. & Lampert, A. Wind shear of low-level jets and their influence on manned and unmanned fixed-wing aircraft during landing approach. *Atmosphere* **13**(1), 35 (2021).
6. Smalikho, I. N. & Rahm, S. Lidar investigations of the effects of wind and atmospheric turbulence on an aircraft wake vortex. *Atmos. Ocean. Opt.* **23**(2), 137–146 (2010).
7. Kim, J. H. & Chun, H. Y. Statistics and possible sources of aviation turbulence over South Korea. *J. Appl. Meteorol. Climatol.* **50**(2), 311–324 (2011).
8. Morelli, E. & Cunningham, K. Aircraft dynamic modeling in turbulence. In *AIAA Atmospheric Flight Mechanics Conference*, 4650 (2012).
9. Zhao, Y., Liu, Z., Li, X., Zhao, M. & Liu, Y. A modified turbulence model for simulating airflow aircraft cabin environment with mixed convection. In *Building Simulation*, vol. 13, 665–675. (Tsinghua University Press, 2020)
10. Chan, P. W., Shun, C. M. & Wu, K. C. Operational LIDAR-based system for automatic wind shear alerting at the Hong Kong International Airport. In *12th Conference on Aviation, Range, and Aerospace Meteorology*, vol. 6 (2006).
11. Shun, C. M. & Chan, P. W. Applications of an infrared Doppler LiDAR in detection of wind shear. *J. Atmos. Ocean. Tech.* **25**(5), 637–655 (2008).

12. Wilson, W. F., Goodrich, K. R. & Carson, S. *Aircraft Measurement of Terrain-Induced Wind Shear on the Approaches to Juneau Airport* (National Center for Atmospheric Research, 2005).
13. Szeto, K. C. & Chan, P.W. High resolution numerical modelling of wind shear episodes at the Hong Kong International Airport. In *12th Conference on Aviation, Range, and Aerospace Meteorology* (2006).
14. Lane, T. P., Doyle, J. D., Sharman, R. D., Shapiro, M. A. & Watson, C. D. Statistics and dynamics of aircraft encounters of turbulence over Greenland. *Mon. Weather Rev.* **137**(8), 2687–2702 (2009).
15. Chan, P. W. Super-high-resolution numerical simulation of atmospheric turbulence in an area of complex terrain. In *12th Conference on Mountain Meteorology* (2006).
16. Kwong, K. M. & Chan, P. W. LIDAR-based turbulence intensity calculation along glide paths. In *14th Coherent Laser Radar Conference*, 8–13 (2007).
17. Lei, L., Chan, P. W., Li-Jie, Z. & Hui, M. Numerical simulation of terrain-induced vortex/wave shedding at the Hong Kong International Airport. *Meteorol. Z.* **22**(3), 317–327 (2013).
18. Chen, F., Peng, H., Chan, P. W., Huang, Y. & Hon, K. K. Identification and analysis of terrain-induced low-level wind shear at Hong Kong International Airport based on WRF–LES combining method. *Meteorol. Atmos. Phys.* **134**(4), 60 (2022).
19. Boilley, A. & Mahfouf, J. F. Wind shear over the Nice Côte d’Azur airport: Case studies. *Nat. Hazard.* **13**(9), 2223–2238 (2013).
20. Rasheed, A. & Sorli, K. CFD analysis of terrain induced turbulence at Kristiansand airport. *Kjevik. Aviation* **17**(3), 104–112 (2013).
21. Zhang, H. *et al.* Low-level wind shear identification along the glide path at BCIA by the pulsed coherent doppler LiDAR. *Atmosphere* **12**(1), 50 (2020).
22. Tse, K. S. *Numerical Simulations of Terrain-Induced Turbulent Flow Around Hong Kong International Airport* (Doctoral dissertation) (2018).
23. Shimoyama, K., Nakanomyo, H. & Obayashi, S. Airport terrain-induced turbulence simulations integrated with weather prediction data. *Trans. Jpn. Soc. Aeron. Space Sci.* **56**(5), 286–292 (2013).
24. Diana, G., Fiammenghi, G., Belloli, M. A. R. C. O. & Rocchi, D. A. N. I. E. L. E. Wind tunnel tests and numerical approach for long span bridges: The Messina bridge. *J. Wind Eng. Ind. Aerodyn.* **122**, 38–49 (2013).
25. He, X. & Zou, S. Advances in wind tunnel experimental investigations of train–bridge systems. *Tunn. Undergr. Space Technol.* **118**, 104157 (2021).
26. Li, Y., Hu, P., Xu, X. & Qiu, J. Wind characteristics at bridge site in a deep-cutting gorge by wind tunnel test. *J. Wind Eng. Ind. Aerodyn.* **160**, 30–46 (2017).
27. Su, H., Meng, H., Qu, T. & Lei, L. Wind tunnel experiment on the influence of array configuration on the power performance of vertical axis wind turbines. *Energy Convers. Manage.* **241**, 114299 (2021).
28. Guo, W., Shen, H., Li, Y., Feng, F. & Tagawa, K. Wind tunnel tests of the rime icing characteristics of a straight-bladed vertical axis wind turbine. *Renewable Energy* **179**, 116–132 (2021).
29. Cai, C. *et al.* Wind tunnel and numerical study of a floating offshore wind turbine based on the cyclic pitch control. *Renewable Energy* **172**, 453–464 (2021).
30. Li, Y., Sun, P. P., Li, A. & Deng, Y. Wind effect analysis of a high-rise ancient wooden tower with a particular architectural profile via wind tunnel test. *Int. J. Arch. Herit.* **17**(3), 518–537 (2023).
31. Wang, X., Li, Q. & Li, J. Field monitoring and wind tunnel study of wind effects on roof overhang of a low-rise building. *Struct. Control. Health Monit.* **27**(3), e2484 (2020).
32. Barua, L., Zou, B. & Zhou, Y. Machine learning for international freight transportation management: A comprehensive review. *Res. Transp. Bus. Manag.* **34**, 100453 (2020).
33. Nama, M. *et al.* Machine learning-based traffic scheduling techniques for intelligent transportation system: Opportunities and challenges. *Int. J. Commun. Syst.* **34**, e4814 (2021).
34. Khattak, A., Almujiabah, H., Elamary, A. & Matarra, C. M. Interpretable dynamic ensemble selection approach for the prediction of road traffic injury severity: A case study of Pakistan’s national highway N-5. *Sustainability* **14**, 12340 (2022).
35. Dong, S., Khattak, A., Ullah, I., Zhou, J. & Hussain, A. Predicting and analyzing road traffic injury severity using boosting-based ensemble learning models with SHAPley Additive exPlanations. *Int. J. Environ. Res. Public Health* **19**, 2925 (2022).
36. Weng, Y. & Paal, S. G. Machine learning-based wind pressure prediction of low-rise non-isolated buildings. *Eng. Struct.* **258**, 114148 (2022).
37. Lin, P. *et al.* Machine learning-based prediction of crosswind vibrations of rectangular cylinders. *J. Wind Eng. Ind. Aerodyn.* **211**, 104549 (2021).
38. Kim, B., Yuvaraj, N., Tse, K. T., Lee, D.-E. & Hu, G. Pressure pattern recognition in buildings using an unsupervised machine-learning algorithm. *J. Wind Eng. Ind. Aerodyn.* **214**, 104629 (2021).
39. Hu, G. *et al.* Deep learning-based investigation of wind pressures on tall building under interference effects. *J. Wind Eng. Ind. Aerodyn.* **201**, 104138 (2020).
40. Wada, D., Araujo-Estrada, S. A. & Windsor, S. Unmanned aerial vehicle pitch control using deep reinforcement learning with discrete actions in wind tunnel test. *Aerospace* **8**, 18 (2021).
41. Nori, H., Jenkins, S., Koch, P. & Caruana, R. *Interpretml: A Unified Framework for Machine Learning Interpretability*. (2019). <http://arxiv.org/abs/1909.09223>.
42. Rathore, S. S. & Kumar, S. A decision tree regression based approach for the number of software faults prediction. *ACM SIGSOFT Softw. Eng. Notes* **41**(1), 1–6 (2016).
43. Su, X., Yan, X. & Tsai, C. L. Linear regression. *Wiley Interdiscipl. Rev. Comput. Stat.* **4**(3), 275–294 (2012).
44. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
45. Chen, T. & Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794 (2016).
46. Schapire, R. E. *Explaining Adaboost Empirical Inference* 37–52 (Springer, 2013).
47. Ahmad, M. W., Reynolds, J. & Rezgui, Y. Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *J. Clean. Prod.* **203**, 810–821 (2018).
48. Wu, J. *et al.* Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. Electron. Sci. Technol.* **17**(1), 26–40 (2019).
49. Hong Kong Lantau Island. *Gosur Maps*. https://www.gosur.com/map/hong-kong_islands-district_lantau%20land/?ll=22.248861393152396,113.92478936610541&z=11.512420070406215&t=satellite. Accessed 2 May 2023.
50. Chan, P. An event of tail strike of an aircraft due to terrain-induced wind shear at the Hong Kong International Airport. *Meteorol. Appl.* **19**, 325–333 (2012).
51. Chan, P. A significant wind shear event leading to aircraft diversion at the Hong Kong international airport. *Meteorol. Appl.* **19**, 10–16 (2012).
52. Hon, K., Chan, P., Chiu, Y. & Tang, W. Application of short-range LIDAR in early alerting for low-level wind shear and turbulence at Hong Kong International Airport. *Adv. Meteorol.* **2014**, 1–10 (2014).
53. Chan, P. & Hon, K. Observation and numerical simulation of terrain-induced wind shear at the Hong Kong International Airport in a planetary boundary layer without temperature inversions. *Adv. Meteorol.* **2016**, 1–9 (2016).
54. Chen, F., Peng, H., Chan, P.-W. & Zeng, X. Low-level wind effects on the glide paths of the North Runway of HKIA: A wind tunnel study. *Build. Environ.* **164**, 106337 (2019).

55. Hong Kong International Airport. *Gosur Maps*. https://www.gosur.com/map/hong-kong_islands-district_lantau%20land/?ll=22.30716330676063,113.91088761573155&z=13.198816038941512&t=satellite. Accessed 2 May 2023.
56. Chan, P. W. Generation of eddy dissipation rate map at the Hong Kong International Airport based on Doppler LIDAR data. In *12th Conference on Aviation, Range, and Aerospace Meteorology* (2006).
57. Puranik, T.G., Harrison, E., Min, S., Jimenez, H. & Mavris, D.N. General Aviation approach and landing analysis using flight data records. In *16th AIAA Aviation Technology, Integration, and Operations Conference*, 3913 (2016).
58. Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian optimization of machine learning algorithms. *Adv. Neural. Inf. Process. Syst.* **25**, 1–12 (2012).
59. Turner, R. *et al.* Bayesian optimization is superior to random search for machine learning hyperparameter tuning: Analysis of the black-box optimization challenge 2020. In *NeurIPS 2020 Competition and Demonstration Track*, 3–26. (PMLR, 2021).
60. Mina, M., Rezaei, M., Sameni, A., Ostovari, Y. & Ritsema, C. Predicting wind erosion rate using portable wind tunnel combined with machine learning algorithms in calcareous soils, Southern Iran. *J. Environ. Manage.* **304**, 114171 (2022).
61. Chen, F., Peng, H., Chan, P.-W. & Zeng, X. Wind tunnel testing of the effect of terrain on the wind characteristics of airport glide paths. *J. Wind Eng. Ind. Aerodyn.* **203**, 104253 (2020).
62. Chan, P. Case study of a special event of low-level wind shear and turbulence at the Hong Kong International Airport. *Atmos. Sci. Lett.* **1**, e1143 (2022).

Acknowledgements

This research was supported by Research Fund for International Young Scientists (RFIS) of National Natural Science Foundation of China (NSFC) (Grant No. 52250410351), National Foreign Expert Project (Grant No. QN2022133001L), National Natural Science Foundation of China (U1733113) and Shanghai Municipal Science and Technology Major Project (2021SHZDZX0100).

Author contributions

A.K. and P.C. done the data curation. A.K. performed formal analysis. F.C. and A.K. had funding acquisition. F.C., A.K. and H.P. devised the methodology. P.C. done the Validation. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to F.C.

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

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



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

© The Author(s) 2023