Taylor & Francis
Taylor & Francis Group

Applied Artificial Intelligence
An International Journal

ISSN: 0883-9514 (Print) 1087-6545 (Online) Journal homepage: www.tandfonline.com/journals/uaai20

Assessment of Wind Shear Severity in Airport Runway Vicinity using Interpretable TabNet approach and Doppler LiDAR Data

Afaq Khattak, Jianping Zhang, Pak-Wai Chan & Feng Chen

To cite this article: Afaq Khattak, Jianping Zhang, Pak-Wai Chan & Feng Chen (2024) Assessment of Wind Shear Severity in Airport Runway Vicinity using Interpretable TabNet approach and Doppler LiDAR Data, Applied Artificial Intelligence, 38:1, 2302227, DOI: 10.1080/08839514.2024.2302227

To link to this article: https://doi.org/10.1080/08839514.2024.2302227

Published online: 10 Jan 2024.

APPLIED ARTIFICIAL INTELLIGENCE
2024, VOL. 38, NO. 1, e2302227 (27 pages)
https://doi.org/10.1080/08839514.2024.2302227

Taylor & Francis
Taylor & Francis Group

Assessment of Wind Shear Severity in Airport Runway Vicinity using Interpretable TabNet approach and Doppler LiDAR Data

Afaq Khattak[a], Jianping Zhang[b], Pak-Wai Chan[c], and Feng Chen[a]

[a]The Key Laboratory of Infrastructure Durability and Operation Safety in Airfield of CAAC, College of Transportation Engineering, Tongji University, Jiading, Shanghai, China; [b]The Second Research Institute of Civil Aviation Administration of China, Civil Unmanned Aircraft Traffic Management Key Laboratory of Sichuan Province, Chengdu, China; [c]Hong Kong Observatory, Hong Kong, China

ABSTRACT

Severe wind shear events near airport runways pose serious safety risks and are a growing concern in civil aviation. Identification of severe wind shear risk factors may enhance aviation safety. Although rare, severe wind shear impacts safety by affecting the airspeed, lift, and maneuverability of aircraft. This study presents TabNet, a novel deep learning technique coupled with Bayesian optimization (BO) to predict wind shear severity in the runway vicinity using Doppler LiDAR data from Hong Kong International Airport. To address imbalanced wind shear data, it was first processed by resampling techniques and then used as input to TabNet. The analysis demonstrated that Bayesian-tuned TabNet (BO-TabNet) with SVM-SMOTE-processed data led to better performance compared to other strategies. The TabNet architecture employs the attention mechanism to enable model-specific interpretability. Analysis showed that the most important contributing factor was the summer season, followed by the wind shear encounter location (1 nautical miles from the runway at the departure end). Additionally, a more comprehensive model-agnostic LIME method was used to elucidate the model from a local perspective. By predicting severe wind shear and assessing contributing factors, aviation stakeholders can proactively manage and mitigate the associated risks, leading to safer and more efficient operations.

ARTICLE HISTORY
Received 25 August 2023
Revised 10 December 2023
Accepted 2 January 2024

Introduction

The impact of local weather conditions significantly affects the operational activities of airlines. Adverse weather conditions can give rise to flight disruptions, including delays, cancellations, and accidents (Borsky and Unterberger 2019; Choi et al. 2016; Gultepe et al. 2019). In the field of aviation, the term

CONTACT Feng Chen fengchen@tongji.edu.cn The Key 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

"wind shear" is commonly used to describe an occurrence that involves an abrupt alteration in the speed or trajectory of wind. The meticulous evaluation of wind conditions holds significant importance in the field of aviation, particularly in the context of the landing phase, owing to the considerable risks it poses. According to the International Civil Aviation Organization (ICAO), wind shear is defined as a consistent alteration in the velocity of a headwind or tailwind, with a minimum magnitude of 15 knots occurring within a vertical distance of 1600 ft. (Airport Council International 2017). The issue at hand relates to a safety-critical occurrence, as it has the potential to cause aircraft to deviate from their intended trajectory, thus presenting potential risks to both incoming and outgoing aircraft as shown in Figure 1.

Numerous instances have been documented wherein wind shear magnitudes surpassing 25 knots have been observed at different airports worldwide. The term used to classify these occurrences of wind shear is commonly known as "severe wind shear." The susceptibility of Hong Kong International Airport (HKIA) to wind shear has been extensively stated in prior research (Chan 2012; Shun and Chan 2008).

This location lies in the northern vicinity of Lantau Island and exhibits a diverse topography, characterized by low-lying regions with an average altitude of 300 m and elevated mountain summits reaching up to 900 m. The occurrence of wind shear has been recorded in a significant proportion of pilot flight reports (PIREPs) at Hong Kong International Airport (HKIA) during the period from its inauguration in 1998 to 2015 (Chan 2017). The reports provided data on wind shear magnitudes ranging from 15 to 25 knots, which accounted for 97.70% of all documented occurrences. There have been multiple recorded instances of wind shear with magnitudes equal to or greater than 25 knots, constituting approximately 2.30% of all reported wind shear events.

To detect wind shear in the runway vicinity, several major airports worldwide have installed a number of different meteorological instruments, including Terminal Doppler Weather Radar (TDWR), ground-based anemometer

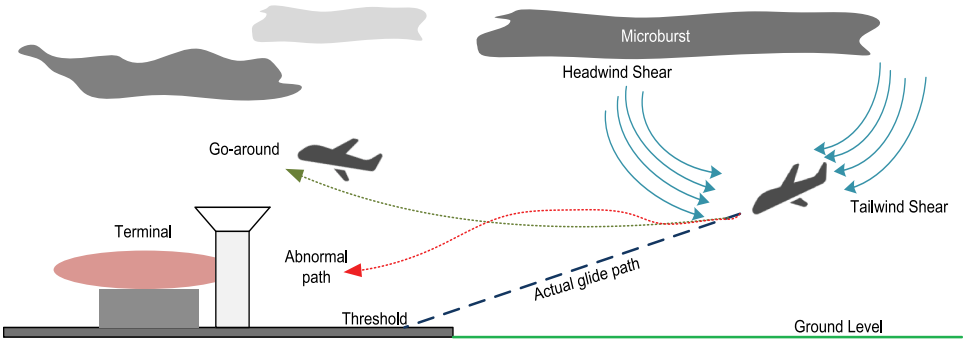

Figure 1. Effect of wind shear on approaching aircraft.

networks, wind profilers, and Doppler Light Detection and Ranging (Doppler LiDAR) systems (Nechaj et al. 2019; Ryan, Saputro, and Sopaheluwakan 2023; Zhang et al. 2019). However, only a few airports globally, such as those in Japan, Germany, France, China, and Singapore, have implemented these technologies. The significant expenses associated with the operation and maintenance of these technologies limit their adoption (Thobois, Cariou, and Gultepe 2019). Moreover, although they have shown efficacy in on-site detection of wind shear, these technologies are unable to predict the occurrence of subsequent wind shear events or identify the specific risk factors that contribute to their occurrence (Chen et al. 2020). Compared to the high cost of acquiring and maintaining wind shear detection instruments, developing and implementing statistical and artificial intelligence (AI)-based prediction models can be more cost-effective. Once trained, these models can provide predictions without the need for expensive hardware installations. By leveraging advanced algorithms, AI model can learn from historical data and identify patterns that may contribute to wind shear occurrences. This can lead to improved accuracy in predicting wind shear events, enabling better preparedness and risk mitigation. Therefore, it is important to acknowledge that suitably trained expert systems of AI, when equipped with a sufficient dataset, possess the capability to aid researchers in aviation safety by accurately and expeditiously estimating the severity of wind shear (Khattak et al. 2022). However, the research domain pertaining to wind shear modeling and prediction in the vicinity of airport runways has been recognized as a particularly challenging area within the field of civil aviation (Chen et al. 2020). Recently, the incorporation of AI, particularly machine learning and deep learning techniques, has witnessed an increased prevalence in the domain of aviation-related decision-making (Alkhamisi and Mehmood 2020; Cankaya et al. 2023). Various machine learning and deep learning techniques have been employed in the field of aviation meteorology. These include Extreme Gradient Boosting (XGBoost) (Khattak et al., 2023b), Artificial Neural Network (ANN) (Coburn, Arnheim, and Pryor 2022), Chaotic Oscillatory-based Neural Networks (CONN) (Liu, Kwong, and Chan 2012), Explainable Boosting Machine (EBM) (Khattak et al., 2023), as well as analytical approaches such as Principal Component Analysis (PCA) combined with K-Means (Mizuno, Ohba, and Ito 2022), Self-Paced Ensemble (SPE) utilizing Extreme Gradient Boosting (XGBoost) as the base estimator (Khattak et al., 2023), Extra Tree (ET) Regression (Dhief et al. 2020), Long Short-Term Memory (LSTM) (Schultz and Reitmann 2019), and Random Forest (Luo et al. 2021), among others. Table 1 illustrates the summary of literature review of state-of-the-art AI algorithm in aviation meteorology and operational safety.

In the past decade, deep learning has demonstrated notable advancements in various domains that encompass structured data, including but not limited to images, text, audio, and time series (Pouyanfar et al. 2018). The utilization

Table 1. Summary of literature review of state-of-the-art AI algorithm in aviation meteorology and operational safety.

Aspects	Description	Performance Matrix	Optimal AI model	Literature
Meteorological Aspects	Time series forecasting of severe wind shear based on HKIA-based Doppler LiDAR data	R^{2*}	Extreme Gradient Boosting (XGBoost)	Khattak et al. (2023b)
	Short-term forecasting of wind gusts at airport	MAE*	Artificial Neural Network (ANN)	Coburn, Arnheim, and Pryor (2022)
	Flight delays prediction based on flight meteorological data	Accuracy, Loss and Gradient values	DCNN* and SE-DenseNet*	Qu et al. (2020)
	Wind shear and turbulence forecasting	RMSE*	Chaotic Oscillatory-based Neural Networks (CONN)	Liu, Kwong, and Chan (2012)
	Assessment of wind field along the glide slope of airport runway.	RMSE and R^2	Explainable Boosting Machine (EBM)	Khattak et al. (2023a)
	Turbulence risk prediction model based on meteorological data	Welch's test	Principal Component Analysis (PCA)	Mizuno, Ohba, and Ito (2022)
Safety and Operational Aspects	Predicting wind shear-induced missed approaches at Hong Kong International Airport based on the Pilot Reports	G-Mean*	Self-Paced Ensemble (SPE) with Extreme Gradient Boosting (XGBoost) as base estimator	Khattak et al., 2023
	Prediction of aircraft landing time at Singapore Changi International Airport	MAE and RMSE	Extra Tree (ET) Regression	Dhief et al. (2020)
	Prediction of aircraft boarding time	MSE	Long Short-Term Memory (LSTM)	Schultz and Reitmann (2019)
	Forecasting the exact aircraft ground time of aircraft at airport	Classification Accuracy	Random Forest (RF)	Luo et al. (2021)

*R^2: Coefficient of Determination, RMSE:Root Mean Squared Error, MAE: Mean Absolute Error; G-Mean: Geometric Mean; DCNN: Dual-channel Convolutional Neural Network, SE-DenseNet: Squeeze and Excitation-Densely Connected Convolutional Network

of structured data is highly advantageous due to the presence of deep learning-based canonical architectures, which enable the efficient conversion of raw data into meaningful and comprehensible representations. The obstacles to optimizing a large number of hyperparameters, addressing class imbalance in the dataset, and interpreting AI algorithms are crucial. Grid search (GS) is a commonly employed hyperparameter tuning technique in AI applications (Bergstra et al. 2011). However, it is notorious for its lengthy computational time. One method that enhances the efficiency of grid search (GS) is the random search (RS) approach (Bergstra and Bengio 2012). The approach employed in this approach entails the random selection of data points from the search space of hyperparameters. The search continues till the specified allocation is depleted. Although more efficient than a GS, this strategy remains time-consuming in comparison to alternative approaches (Wicaksono and Supianto 2018). The Bayesian optimization (BO) algorithm has become

popular for its efficient global optimization of hyperparameters in AI algorithms (Cho et al. 2020; Snoek, Larochelle, and Adams 2012). The presented approach is a Sequential Model-Based Optimization technique that has been specifically developed to effectively ascertain the global optimum while working within the constraints of a restricted number of experimental trials. The process of choosing the subsequent trial point for sampling is designed to achieve a harmonious equilibrium between exploration and exploitation. Moreover, the problem of class imbalance occurs when there is a substantial discrepancy in the sample sizes among various classes. When majority classes outnumber minority classes, AI classification algorithms have difficulty classifying them accurately (Alahmari 2020). The classifiers' predictive accuracy decreases significantly for the minority class. Resampling methods, including SMOTE (Chawla et al. 2002), SVM-SMOTE (Yang et al. 2018), Borderline-SMOTE (Han, Wang, and Mao 2005), and SMOTE-ENN (Batista, Prati, and Monard 2005) can be utilized to tackle these concerns. The diagnostic capabilities of AI algorithms are influenced by various factors, and it is important to consider their impact. The absence of interpretability in artificial intelligence algorithms is a significant limitation. Several post-hoc techniques, including Partial Dependency Analysis (PDA) (Molnar 2020), Local Interpretable Model-Agnostic Explanations (LIME) (Palatnik de Sousa, Maria Bernardes Rebuzzi Vellasco, and Costa da Silva 2019), and the Shapley Additive explanations (SHAP) (Lundberg and Lee 2017), can be employed in conjunction with AI algorithms to enhance interpretability.

This research focuses on the efficient prediction of wind shear severity using a deep learning strategy while addressing the issue of class imbalance as well as interpretability as shown in Figure 2. To achieve this, we utilized wind shear tabular data extracted from the Doppler LiDAR at Hong Kong International Airport. We applied multiple resampling methods to handle the imbalance of the wind shear data. Subsequently, we employed a cutting-edge deep tabular learning architecture called "TabNet," which was fine-tuned using Bayesian techniques (Arik and Pfister 2021).

The strategy took inspiration from the beneficial aspects of deep learning and sought to address its limitations when applied to tabular data. BO is used as a means to enhance the hyperparameter tuning of the TabNet architecture. The BO-TabNet architecture achieves superior performance in classification (McDonnell et al. 2023; Yan et al. 2021). In addition to precisely predicting findings from tabular data, it also exhibits interpretability. This addresses a major drawback of traditional deep learning architectures, which are known for their limited interpretability.

The main objective of the present research is to develop Bayesian-optimized TabNet (BO-TabNet) system by utilizing Doppler LiDAR data

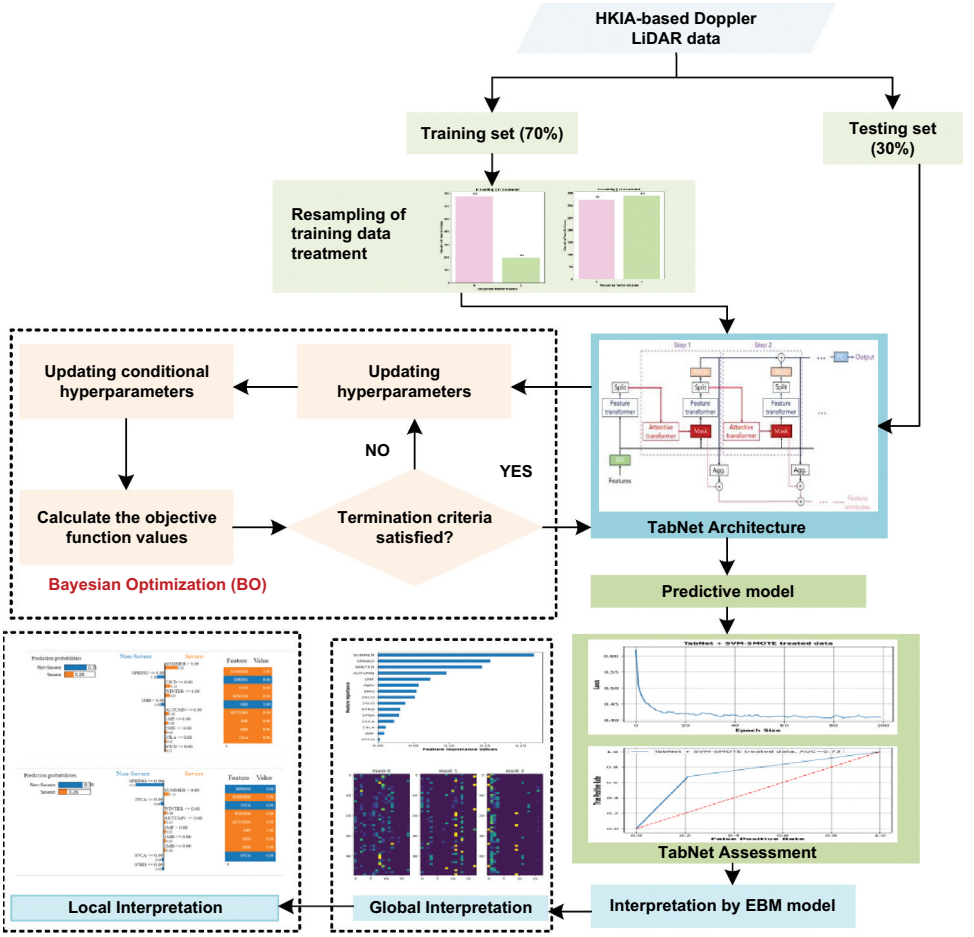

Figure 2. Framework of proposed BO-TabNet strategy for prediction and interpretation of wind shear severity.

in order to improve the accuracy of wind shear severity classification. This paper presents the primary contributions as follows:

- A deep learning classifier, referred to as TabNet, has been introduced and utilized for the purpose of classifying wind shear severity in tabular data. The current approach has been utilized using LiDAR data obtained from HKIA, which underwent several data resampling strategies. The application of BO has significantly improved the learning ability of the TabNet architecture through the optimization of its hyperparameters.
- The efficacy of our proposed BO-TabNet has been assessed in relation to other contemporary classification models.
- TabNet's sequential attention system has been utilized to explain wind shear severity classification results on both a local and a global scale.

Furthermore, the utilization of a model-agnostic explainable LIME approach interprets the outcomes of the model from a local perspective.

The rest of the parts of the article are structured in the following manner: Section 2 provides a comprehensive account of the study site and the technique that was employed for data extraction from the Doppler LiDAR at HKIA. This is followed by a theoretical overview of TabNet, Bayesian Optimization, and the LIME approach. Section 3 provides the study's results and corresponding discussions. Section 4 concludes by providing a summary of the research findings.

Data and Methods

Study Location and Data Retrieval from LiDAR

The Hong Kong International Airport (HKIA), also referred to as HKIA, is situated on Lantau Island, which is surrounded by water on three of its sides. In the southern region of the airport lies a mountainous terrain characterized by elevated plateaus that exceed an altitude of 900 m relative to sea level, as shown in Figure 3. A number of studies based on observations and forecasts have indicated that the Hong Kong International Airport (HKIA) has a unique combination of land-sea contrast, complex topography, and nearby buildings (Figure 4), which establish ideal circumstances for the occurrence of wind shear (Szeto and Chan 2006). This airport is one of the most severely affected by wind shear on a global level (Chan 2006). The present study

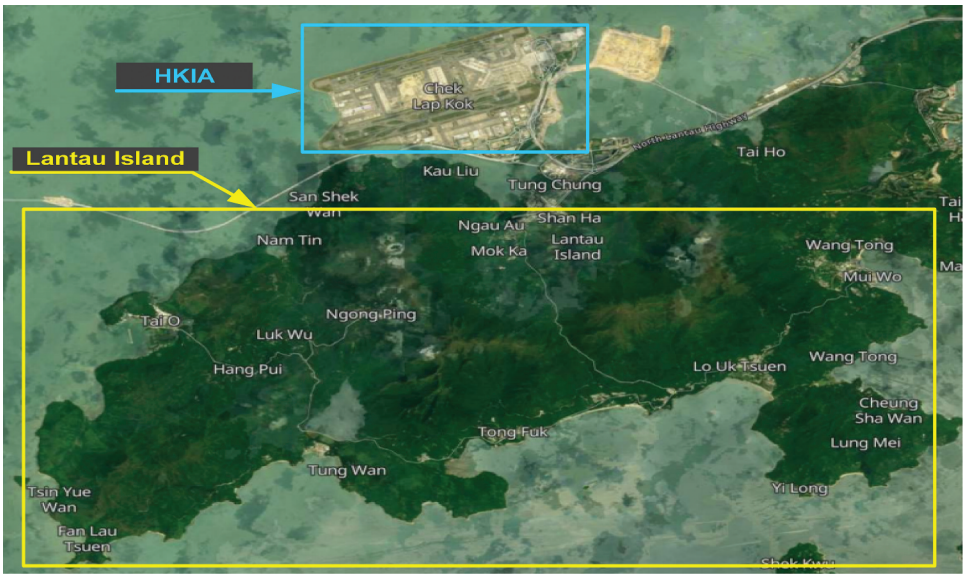

Figure 3. Hong Kong international airport and surrounding Lantau Island.

Figure 4. Buildings near and at Hong Kong international airport.

utilizes wind shear data collected at Hong Kong International Airport (HKIA) through the implementation of two long-range Doppler LiDAR systems. It is capable of determining the magnitude of wind shear events and reporting their location of occurrence. The Doppler LiDAR's radial resolution, referred to as the physical range gate, is 100 m, while its operational wavelength is approximately 1.5 microns in the infrared spectrum. The maximum achievable radial velocity is around 40 m/s. Under ideal weather conditions and without any impediments like low clouds, it is possible to observe objects within a radius that ranges from 10 to 15 km. Doppler LiDAR systems can be customized to perform glide-path scans for both landing and takeoff trajectories, in addition to routine fixed-elevation scans (plan-position indicator). To attain the desired objective, it is imperative to synchronize the vertical (elevation) and horizontal (azimuth) shifts of the laser scanner's head. The northern LiDAR system scans the runway using different configurations, such as 07LA, 25RA, 07LD, and 25RD. The South LiDAR system assesses the four arrangements of the southern runway, specifically 25LA, 07LA, 07RD, and 25LD. Wind shear across each runway can be estimated using radial velocity data collected from glide-path scans. The average duration for scanning each runway is approximately 1 min.

The utilization of a Plan Position Indicator (PPI) scan of the south runway LIDAR at the HKIA is shown in Figure 5a. The scan was performed with an azimuth angle of 3° in relation to the horizon. Figure 5b displays a visual representation of a radial velocity plot that has been obtained through the PPI scan. A significant expanse of winds was observed to be flowing in a direction opposite to the prevailing east-southeast airflow, situated in the western and southern regions relative to the specified coordinates. This area is located at a distance of three nautical

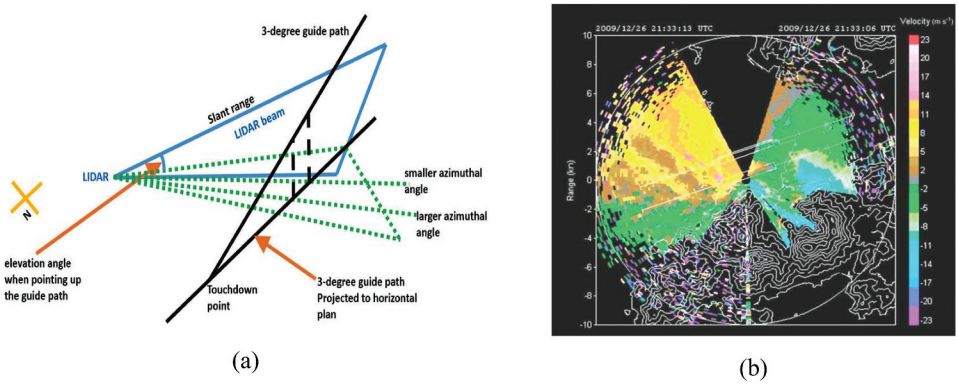

Figure 5. Wind shear detected by Doppler LiDAR; (a) PPI scan @ 3 degree glide path; (b) Radial velocity plot.

miles (equivalent to 5.6 km) in the west-southwest direction from the westernmost point of the southern runway. The depicted region is visually characterized by its green hue.

It is imperative to comprehend that the precise horizontal positioning of the wind shear encounter is also of paramount importance. As depicted in Figure 6, the areas where wind shear events take place are denoted by the aviation terminology of RWY, MD, or MF. The runway is visually represented by a rectangular shape in the dark gray color, and it is labeled as RWY. The dark red rectangles positioned on the right side of the runway indicate the distance to the final approach in miles (1MF corresponds to 1 nautical mile to the final approach). Likewise, the dark red rectangles on the left side of the diagram symbolize the distance from the end of the runway, where takeoff occurs. As an illustration, a two-mile final (2MF) denotes a specific distance of two nautical miles from the threshold of the runway at the arrival end, visually depicted by a pink circle. Table 2 illustrates the sample wind shear data obtained from HKIA-based Doppler LiDAR.

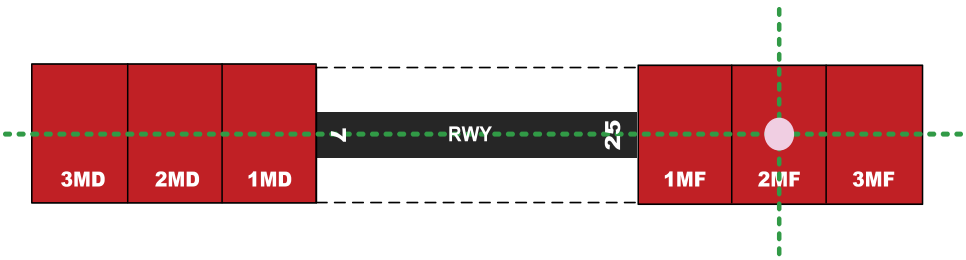

Figure 6. Schematic representation of the horizontal wind shear encounter location.

Table 2. Sample wind shear data obtained from HKIA-based Doppler LiDAR.

Wind Shear Occurrence Date	Wind Shear Occurrence Time	Wind Shear Occurrence Runway	Wind Shear Horizontal encounter Location	Wind shear Magnitude (±) (knots)
2017-01-02	13:31	07RD	RWY	−24
2018-05-18	22:05	25LA	2MF	+21
—	—	—	—	—
—	—	—	—	—
2020-12-31	23:35	07RA	1MF	−32
2021-04-06	17:48	25RD	3MD	+15
—	—	—	—	—
—	—	—	—	—
2022-05-19	13:37	25LA	2MF	−30
2022-08-21	06:14	07RD	RWY	+25

Data Processing

In order to build a wind shear severity model, an extensive amount of data was required. Given this context, we first obtained the wind shear data from the Doppler LiDAR installed at HKIA. Subsequently, we applied a filtering procedure in order to categorize wind shear events as either "Severe-Wind Shear" or "Non-Severe-Wind Shear." Particularly, any wind shear event with a magnitude equal to or exceeding 25 knots was categorized as "Severe-Wind Shear," while all other wind shear events were labeled as "Non-Severe-Wind Shear." The binary classification problem involved assigning the occurrence of the "Severe-Wind Shear" event as 1 and the occurrence of "Non-Severe-Wind Shear" as 0, as illustrated by Eq.1.

$$\text{Wind Shear Severity} = \begin{cases} 1 : \text{Severe} - \text{Wind Shear}, & \text{Wind shear magnitude} \geq 25\,\text{knots} \\ 0 : \text{Non} - \text{Severe} - \text{Wind Shear}, & \text{Wind shear magitude} = 14 - 24.9\,\text{knots} \end{cases} \tag{1}$$

While the occurrence of severe wind shear instances was significantly lower compared to non-severe wind shear, the latter category continues to hold significant importance when it comes to aviation safety analysis. The features of the data set were encoded using the one-hot encoding strategy, as indicated in Table 3. The dataset underwent a processing step wherein each nominal value was transformed into an individual column, and this new column was assigned a binary value of either 0 or 1. The number of nominal values corresponds to the number of columns present in the table. An instance of generating a matrix with eight columns involves utilizing a nominal factor known as "Runway" that encompasses eight unique values (07RA, 07LA, 25LA, 07LD, 25RA, 07RD, 25RD, 25LD).

Theoretical Overview of TabNet, BO, and LIME

This section presents the theoretical frameworks of TabNet and BO. To access frameworks of bench marked classification algorithms, readers can consult the following: Fully Convolutional Network (FCN) (Ben-Cohen et al. 2016) and Extreme Gradient Boosting (XGBoost) (Chen and Guestrin 2016).

Table 3. One-hot encoding of wind shear data from HKIA-based Doppler LiDAR.

Factor	Codes and Description
Runways Orientation	
07LA	If the occurrence of wind shear is noticed at Runway 07LA, the value is 1; otherwise, it is 0.
07LD	If the occurrence of wind shear is noticed at Runway 07LD, the value is 1; otherwise, it is 0.
07RA	If the occurrence of wind shear is noticed at Runway 07RA, the value is 1; otherwise, it is 0.
25RD	If the occurrence of wind shear is noticed at Runway 25RD, the value is 1; otherwise, it is 0.
25LA	If the occurrence of wind shear is noticed at Runway 25LA, the value is 1; otherwise, it is 0.
25LD	If the occurrence of wind shear is noticed at Runway 25LD, the value is 1; otherwise, it is 0.
25RA	If the occurrence of wind shear is noticed at Runway 25RA, the value is 1; otherwise, it is 0.
25RD	If the occurrence of wind shear is noticed at Runway 25RD, the value is 1; otherwise, it is 0.
Encounter Location	
1MD	If the wind shear event occurs within a distance of 1 nautical mile from the departure runway threshold, assign a value of 1. Otherwise, assign a value of 0.
2MD	If the wind shear event occurs within a distance of 2 nautical miles from the departure runway threshold, assign a value of 1. Otherwise, assign a value of 0.
3MD	If the wind shear event occurs within a distance of 3 nautical miles from the departure runway threshold, assign a value of 1. Otherwise, assign a value of 0.
1MF	If the wind shear event occurs within a distance of 1 nautical mile from the approaching runway threshold, assign a value of 1. Otherwise, assign a value of 0.
2MF	If the wind shear event occurs within a distance of 2 nautical miles from the approaching runway threshold, assign a value of 1. Otherwise, assign a value of 0.
3MF	If the wind shear event occurs within a distance of 3 nautical miles from the approaching runway threshold, assign a value of 1. Otherwise, assign a value of 0.
RWY	If the wind shear event occurs at Runway then the assign value is 1.
Season of the Year	
Winter	If the wind shear occurrence happens during the winter season, the value is 1; otherwise, it is 0.
Spring	If the wind shear occurrence happens during the spring season, the value is 1; otherwise, it is 0.
Summer	If the wind shear occurrence happens during the summer season, the value is 1; otherwise, it is 0.
Autumn	If the wind shear occurrence happens during the autumn season, the value is 1; otherwise, it is 0.
Time of the Day	
Day	The occurrence of a wind shear event is denoted as 1 if it happens during daytime, and 0 if it occurs at nighttime.
Night	1: The occurrence of a wind shear event is denoted as 1 if it happens during nighttime, and 0 if it occurs at daytime.

TabNet

TabNet is a deep learning model that functions through sequential multi-step processing, as defined by Arik and Pfister (Arik and Pfister 2021). The utilization of this particular deep architecture enhances the approach to feature selection and boosts the ability to acquire and comprehend high-dimensional features. During each iteration at the n^{th} step, a feature vector with D-dimensions is processed, and the resulting output is directed to a block known as the Feature Transformer block (FTB). FTB consists of multiple layers that can either be propagated across decision steps or specific to a particular decision step. Each block is composed of fully-connected layers,

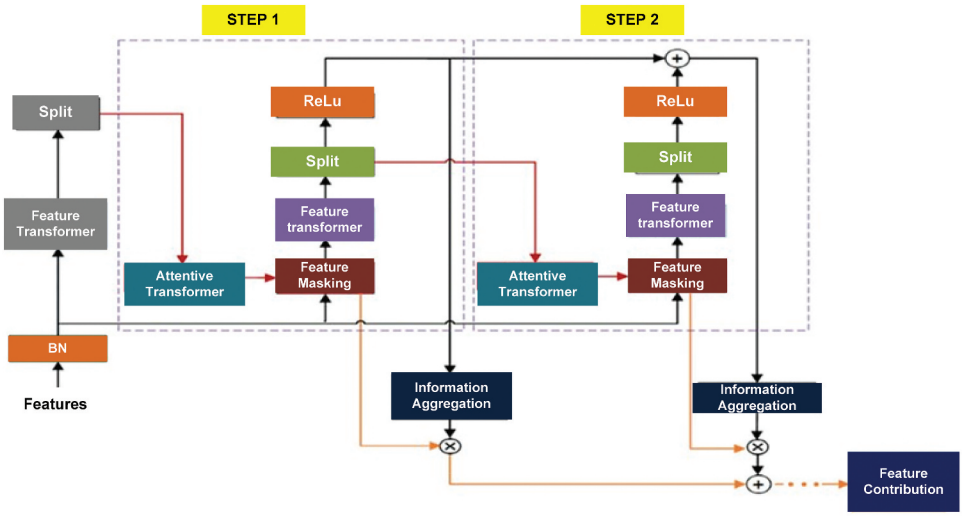

Figure 7. TabNet architecture.

a batch normalization layer, and a Gated Linear Unit (GLU) activation. Furthermore, the Generalized Linear Unit (GLU) is associated with a normalization residual connection, which plays a role in stabilizing the variance throughout the network. The implementation of this multi-layered block enables the process of feature selection and improves the parameter functionality of the network. Figure 7 provides a comprehensive illustration of the TabNet architecture.

The proposed approach comprises multiple elements, specifically an attentive transformer, mask, feature transformer, split node, and ReLu activation, which are present in each iteration. The steps are organized in a sequential manner, progressively increasing until reaching N steps, at which point they are linked to a fully connected layer to produce the output. The Attentive Transformer model integrates various components, such as a fully connected layer, batch normalization, prior scale, and sparse-max dimensionality reduction. The mask function produces significant feature contributions for the purpose of aggregation. When $M_{b,j}[i] = 0$, there is no feature contribution. The Feature Transformer is linked to the Attentive Transformer and Mask, enabling effective feature selection at each iteration. The Attentive Transformer is a computational module comprising several layers, encompassing fully connected layers and batch normalization layers. The presentation of the Attentive Transformer formulation and its corresponding masking procedure is demonstrated in Eq.2.

$$\mathrm{a[i-1]} : \mathrm{M[i]} = \mathrm{sparsemax(P[i-1].h_i([a-1]))} \tag{2}$$

In the given context, $\mathrm{a[i-1]}$represents the preceding stage, $P[\mathrm{i}]$ denotes the prior scale, and $\mathrm{h_i}$ refers to a trainable function. The Attentive Transformer

incorporates two fundamental components, namely the sparsemax activation function and the prior. The dimensionality of feature vectors is reduced through the utilization of sparsemax, which introduces sparsity. Subsequently, these features are projected onto a probability map within Euclidean space. The inclusion of a probability for each projected feature vector enhances the interpretability of the model. The term $P[\mathrm{i}]$ in the context refers to the prior scale term, which represents the prominence of a particular feature in the preceding steps. Its definition shown by Eq. 3:

$$P[\mathrm{i}] = \prod_{j=1}^{i} (\gamma - \mathrm{M}[j]) \tag{3}$$

The concept of γ delineates the correlation between the implementation of a particular attribute at a singular decision point or across multiple decision points. When γ equals 1, the feature is implemented at the specified step and at multiple steps when γ equals 0. The Attentive Transformer algorithm employs a mechanism to identify and prioritize the most significant features, which are then used to construct the transformed feature vector. These selected features are subsequently passed through the learnable Mask, denoted as $\mathrm{M}[j]$. The utilization of the Mask enhances the interpretability of the model and provides additional enhancements to the process of feature selection as compared to the Attentive Transformer. The j^{th} feature of the b^{th} sample is defined by $\mathrm{M}_{b,j}[j]$. When $\mathrm{M}_{b,j}[j]$ is equal to zero, the feature does not make any contribution at that particular step. The process of combining these Masks at each iteration results in a coefficient that assigns a value to the significance of each step in a final determination.

Bayesian Optimization

The process of hyperparameter tuning through trial and error is laborious and frequently yields sub-optimal outcomes. Therefore, it is imperative to employ robust tuning approaches, particularly when the objective of optimization is to locate the highest value at the point of sampling for an unfamiliar function as shown by Eq. 4.

$$\Psi+ = \arg\max_{\Psi \in \Delta} \vartheta(\hbar) \tag{4}$$

In the present context, the symbol Ψ is employed to denote the sampling point, whereas it signifies the search space of said sampling point Ψ. The function ϑ represents an unspecified objective function, and $\Psi+$ denotes the location at which this objective function is to be optimized for maximum value. The approach of performing BO consists of two primary steps (Frazier 2018).

- First, the BO attempts to construct a surrogate function for ϑ by randomly selecting a subset of data points. In this study, the surrogate function has been updated using a Gaussian process (GP) to create the posterior distribution over ϑ. The use of GP is justified by its high flexibility, robustness, accuracy, and analytical traceability.
- Initially, the BO procedure endeavors to create a surrogate function for the target function, denoted as ϑ, by employing a random selection process to choose a subset of data points. The surrogate function in this study has been enhanced through the utilization of GP in order to generate the posterior distribution over ϑ. The utilization of GP is justified due to its notable attributes such as high flexibility, robustness, accuracy, and analytical traceability.
- Subsequently, the posterior distribution obtained from the initial step is employed to derive an acquisition function that serves the dual purpose of exploring unexplored regions within the search space and exploiting regions that have already been identified as yielding optimal outcomes. The processes of exploration and exploitation are ongoing, and the surrogate model continues to be updated with new findings until an establishment until the termination criterion is met. The primary aim is to enhance the performance of the acquisition function, particularly the expected improvement metric, for the purpose of identifying the subsequent sampling point.

Interpretation via LIME

The present study examines interpretation strategies that are specific to particular models as well as those that are applicable across multiple models. In order to analyze both local and global interpretations, we employed a model-specific feature importance and attentional transformer mask within the TabNet framework. The LIME technique (Palatnik de Sousa, Maria Bernardes Rebuzzi Vellasco, and Costa da Silva 2019) was employed to obtain local interpretations of the model. One of the most appealing attributes of LIME are its accessibility and simplicity. The goal of this approach is to work with an interpretable and understandable depiction of the input data for individuals. As a result, the output of LIME consists of a collection of explanations that highlight the significance of each attribute in predicting the outcome of a given data instance. In order to provide further insight into an observation, the LIME approach employs a process of iteratively perturbing the observations, resulting in the generation of replicated feature data. Subsequently, the perturbed data is utilized to generate predictions through the application of a prediction model, such as TabNet. Subsequently, the perturbed data is subjected to a comparison with the original data points in the dataset, wherein the Euclidean distance is computed to ascertain the extent of separation

between the perturbed data point and the initial observation. This signifies the input feature(s) that the model deems significant for generating predictions. The main goal is to generate an explainer that is reliable and comprehensible. The objective function, as denoted by Eq.5, is minimized by LIME in order to accomplish this task.

$$\Gamma(x) = \arg\max_{o \in O} \Theta(u, o, \Pi_x) + \Lambda(\mathrm{o}) \tag{5}$$

where u illustrates the actual model, o denotes the interpretable model,$x$represent the original data points, Π_x represent measure of proximity from every permutations to actual data points, $\Theta(u, o, \Pi_x)$element is a gauge of unfaithfulness of o in approximating u in the locality given by Π_x, and $\Lambda(o)$is the model complexity measure.

Performance Measures

The performance of BO-TabNet in terms of its predictive and classification abilities can be assessed by utilizing five metrics: precision, recall, F1-Score, Matthews Correlation Coefficient (MCC), Geometric mean (G-Mean), and receiver operating characteristic (ROC) curve. The specific information pertaining to each item is provided in the following section.

Precision

Precision refers to the potential of a model to accurately predict a positive outcome. The Eq.6 demonstrates the calculation of the ratio between the number of accurate positive predictions, referred to as "true positives," and the total number of positive predictions, which includes both "false positives" and "true positives."

$$\text{Precision} = \frac{t_{\mathrm{p}}}{t_{\mathrm{p}} + f_{\mathrm{p}}} \tag{6}$$

Recall

The measure of recall assesses the ability of the model to correctly identify instances that belong to the positive class. The computation involves determining the ratio of accurately classified positive instances to the total number of original positive instances, as represented by Eq. 7. The recall value is observed to increase when a greater number of positive instances are accurately classified.

$$\text{Recall} = \frac{t_{\mathrm{p}}}{t_{\mathrm{p}} + f_{\mathrm{n}}} \tag{7}$$

Geometric Mean

The geometric mean, commonly referred to as the G-mean, is estimated by taking the square root of the product of recall values for each class as shown in Eq.8. This metric aims to optimize accuracy across all classes while ensuring consistent performance across the entire dataset.

$$\mathrm{G-Mean} = \sqrt{\left(\frac{t_p}{t_p + f_n}\right)\left(\frac{t_n}{f_p + t_n}\right)} \tag{8}$$

Matthews Correlation Coefficient

This metric pertains to the quality of binary classification. The metric evaluates both true and false positives and negatives, rendering it a well-balanced measure that is applicable even in scenarios with substantial variations in class sizes. The value of MCC ranges from −1 to 1. The value of 1 illustrates the perfect agreement. The computation can be performed using Eq.9.

$$\mathrm{MCC} = \frac{t_p \times t_n - f_p \times f_n}{\sqrt{(t_p + f_p)(t_p + f_n)(t_n + f_p)(t_n + f_n)}} \tag{9}$$

Receiver Operating Characteristic (ROC) Curve

ROC metrics have been devised to employ predicted probabilities for each class, rather than requiring a binary decision threshold value. In an ROC curve, the horizontal axis corresponds to the false-positive rates (FPRs), while the vertical axis corresponds to the true-positive rates (TPRs). The area under the ROC curve (AU-ROC) is a numerical measure ranging from 0 to 1, where a higher value signifies a greater level of accuracy in the model.

Results and Discussion

Data acquired from HKIA-based Doppler LiDAR were initially subjected to a cleaning process and assessed for the presence of any missing values. Subsequently, the data was divided into separate training and testing datasets in order to mitigate the potential problems associated with data leakage. Due to the absence of a prescribed protocol for data partitioning ratio, the datasets were subjected to a random partitioning process, resulting in a division of 70% for model training and 30% for model testing purposes. During the model's Bayesian optimization stage, the training data (70%) that had been cleaned were divided into 10 equal subsets. This partitioning was done randomly and was used for 10-fold cross-validation. The model training and tuning process utilized

Table 4. Optimal hyperparameters for the TabNet model with different data treatment strategies.

Data Treatment	Optimal Hyperparameters for TabNet model
Untreated Data	optimizer = Adam, n_d = 65, n_a = 65, lambda_sparse = 3.95E–5; gamma = 1.1, n_shared = 3, n_steps = 2, learning_rate = 0.12, mask_type = sparsemax, max_epochs = 100,
SVM-SMOTE	learning_rate = 0.06, n_steps = 3, n_d = 54, n_a = 52, lambda_sparse = 4.30E–5, gamma = 1.16, n_shared = 2, mask_type = sparsemax, optimizer = Adam, max_epochs = 100
Borderline-SMOTE	learning_rate = 0.08, n_steps = 2, n_d = 62, n_a = 59, lambda_sparse = 3.96E–5, gamma = 1.20, n_shared = 2, mask_type = sparsemax, optimizer = Adam, max_epochs = 100
Near Miss	learning_rate = 0.11, n_steps = 2, n_d = 67 n_a = 65, lambda_sparse = 4.30E–5, gamma = 1.11, n_shared = 3, mask_type = sparsemax, optimizer = Adam, max_epochs = 100
ADASYN	learning_rate = 0.17, n_steps = 1, n_d = 61, n_a = 60, lambda_sparse = 4.22E–5, gamma = 1.24, n_shared = 2, mask_type = sparsemax, optimizer = Adam, max_epochs = 100
SMOTE-ENN	learning_rate = 0.14, n_steps = 2, n_d = 62, n_a = 59, lambda_sparse = 4.13E–5, gamma = 1.0, n_shared = 2, mask_type = sparsemax, optimizer = Adam, max_epochs = 100

a total of nine subsets, while one additional subset was allocated for validation purposes. The procedure was iterated a total of 10 instances. A portion of 30% from the processed data was allocated for the purpose of evaluating the model.

Hyperparameter Tuning of TabNet via BO

The TabNet model has demonstrated superior performance compared to traditional machine learning models when evaluated with datasets from different fields of study (Chang et al. 2023; Liu 2023; McDonnell et al. 2023). Despite its robust classification capabilities, it is highly parametric. Improperly tuned hyperparameters may lead to sub-optimal outcomes. Hence, it is necessary to determine the most suitable hyperparameters prior to training the model. This study utilized the BO algorithm with a GP surrogate model to achieve this objective. The hyperparameters of TabNet were optimized by tuning them within their respective search space. The hyperparameters encompass several key components: 1) the masking function, which can be either entmax or sparsemax; 2) the width of the decision prediction layer, denoted as n_d; 3) the width of the attention embedding for each mask, referred to as n_a; 4) the number of steps in the architecture, represented by n_steps; 5) the coefficient for feature re-usage in the masks, denoted as gamma; 6) the number of shared GLU at each step, indicated by n_shared; 7) and the sparsity loss coefficient (lambda_sparse). The optimal hyperparameters are chosen based on the evaluation metric, i.e., highest area under the ROC curve (AU-ROC). The hyperparameters that were ultimately selected for different data treatments are shown in Table 4. Similarly, the loss vs. Epoch size plots for TabNet model with different resampling strategies are shown in Figure 8.

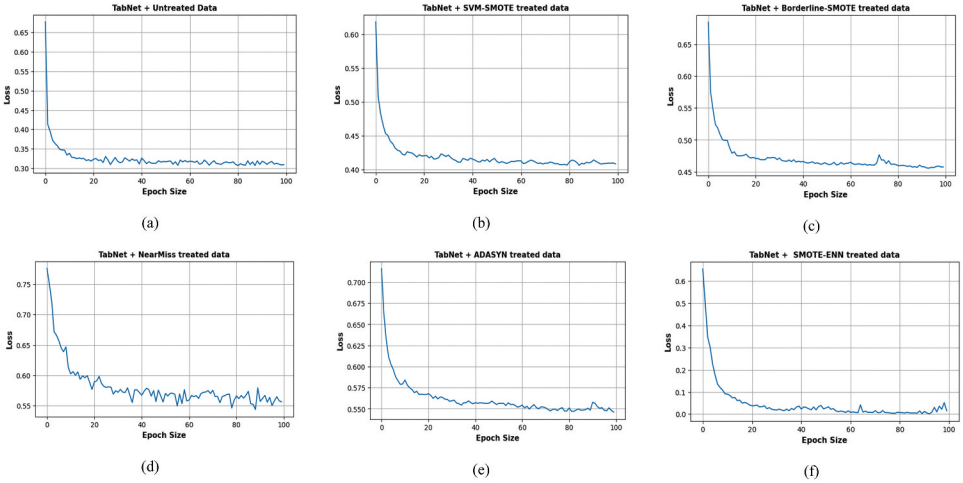

Figure 8. Loss vs. Epoch size for BO-tuned TabNet; (a) TabNet + untreated data; (b) TabNet + SVM-SMOTE treated data; (c); TabNet + borderline-SMOTE treated data; (d) TabNet + Near Miss treated data; (e) TabNet + ADASYN treated data; (d) TabNet + SMOTE-ENN treated data.

Table 5. Comparison of confusion matrix outcomes of BO-TabNet with other models and with different data treatment strategies.

Data Treatment Strategy	Models	True Negative (TN)	False Positive (FP)	False Negative (FN)	True Positive (TP)
Untreated Data	BO-TabNet	1804	13	225	71
	FCN	1807	10	233	63
	XGBoost	1812	5	241	55
SVM-SMOTE	BO-TabNet	1430	387	95	201
	FCN	1510	307	111	185
	XGBoost	1527	290	118	178
Borderline-SMOTE	BO-TabNet	1256	561	80	216
	FCN	1321	496	95	201
	XGBoost	1330	487	100	196
Near Miss	BO-TabNet	1317	500	104	192
	FCN	1395	422	119	177
	XGBoost	1317	500	123	173
ADASYN	BO-TabNet	1378	439	92	204
	FCN	1423	394	107	189
	XGBoost	1431	386	110	186
SMOTE-ENN	BO-TabNet	983	834	76	220
	FCN	1054	763	98	198
	XGBoost	1074	743	105	191

Performance Prediction and Comparison

The classification results of true positive, true negative, false positive, and false negative were obtained from the confusion matrix values of both the treated and untreated models (Table 5). The aforementioned findings were employed in the calculation of the evaluation metrics in order to facilitate the comparison of models. The metrics of precision, recall, F1-Score, and G-Mean are frequently utilized in the domain of imbalanced classification problems. A preference is given for indicators with a higher magnitude. The metrics

Table 6. Performance measures of BO-TabNet with different resampling strategies.

Model with data treatment strategy	Class	Precision	Recall	F1-Score	MCC	G-Mean
BO-TabNet + Untreated Data	Non-Severe	0.89	0.99	0.94	0.12	0.49
	Severe	0.85	0.24	0.37		
	Average	0.88	0.89	0.89		
BO-TabNet +SVM-SMOTE treated data	Non-Severe	0.94	0.79	0.86	0.36	0.73
	Severe	0.34	0.68	0.45		
	Average	0.85	0.77	0.80		
BO-TabNet + Borderline-SMOTE treated data	Non-Severe	0.94	0.69	0.80	0.33	0.71
	Severe	0.28	0.73	0.40		
	Average	0.85	0.70	0.74		
BO-TabNet + Near Miss treated data	Non-Severe	0.93	0.72	0.81	0.23	0.69
	Severe	0.28	0.65	0.39		
	Average	0.84	0.71	0.75		
BO-TabNet + ADASYN treated data	Non-Severe	0.94	0.76	0.84	0.32	0.72
	Severe	0.32	0.69	0.43		
	Average	0.85	0.75	0.78		
BO-TabNet + SMOTE-ENN treated data	Non-Severe	0.93	0.54	0.68	0.18	0.63
	Severe	0.21	0.74	0.33		
	Average	0.83	0.57	0.63		

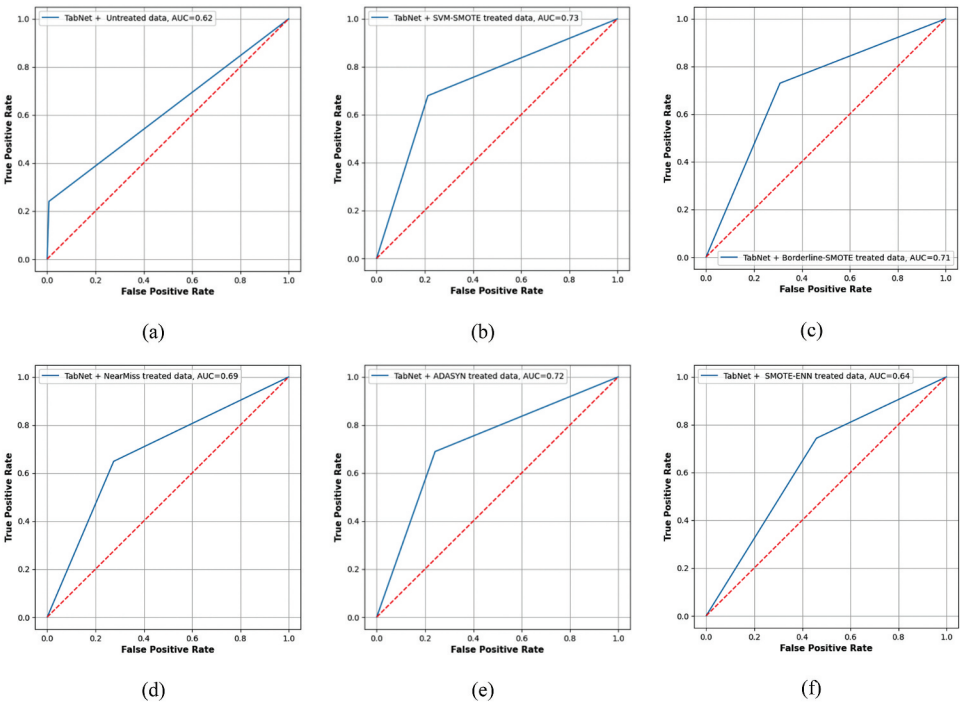

Figure 9. ROC curve for BO-tuned TabNet; (a) TabNet + untreated data; (b) TabNet + SVM-SMOTE treated data; (c); TabNet + borderline-SMOTE treated data; (d) TabNet + Near Miss treated data; (e) TabNet + ADASYN treated data; (d) TabNet + SMOTE-ENN treated data.

are displayed in Table 6 and ROC curves are illustrated in Figure 9. When the BO-tuned TabNet model was applied to untreated Doppler LiDAR data based on HKIA, it achieved Precision, Recall, F1-Score, MCC, G-Mean, and AU_ROC measures of 0.88, 0.89, 0.86, 0.12, 0.49, and 0.62, respectively.

Various resampling strategies were subsequently utilized to process the data, and it was noted that the employment of BO-TabNet with SVM-SMOTE on the treated data yielded superior results compared to other methods, with a higher G-Mean value of 0.73 and an AU_ROC of 0.73. The aforementioned values exhibited the highest magnitude among all assessed competing models. The observed enhancements in these metrics demonstrate the efficacy of resampling techniques in enhancing the classification of wind shear severity.

Interpretation of TabNet

The outcomes achieved by the proposed BO-TabNet model utilizing SVM-SMOTE-processed data were noteworthy. However, in order to enhance its reliability and practicality, interpretation strategies have been implemented. Interpretations were obtained through the utilization of both model-specific and model-agnostic techniques, specifically the LIME method. To adopt a model-specific strategy, we employed the permutation-based feature importance measure of TabNet as well as the mask-based feature importance measure of TabNet. As shown in Figure 10a, the summer season turned out to be highly influential in the likelihood of severe wind shear. It was then followed by the spring season. The reason for severe wind shear in the summer at HKIA can be attributed to a combination of meteorological factors specific to the region. Hong Kong is located in a region prone to tropical cyclones during the summer months. These cyclones bring strong winds and atmospheric instability, which can create significant wind shear conditions near HKIA. The interaction between the cyclone's circulation and

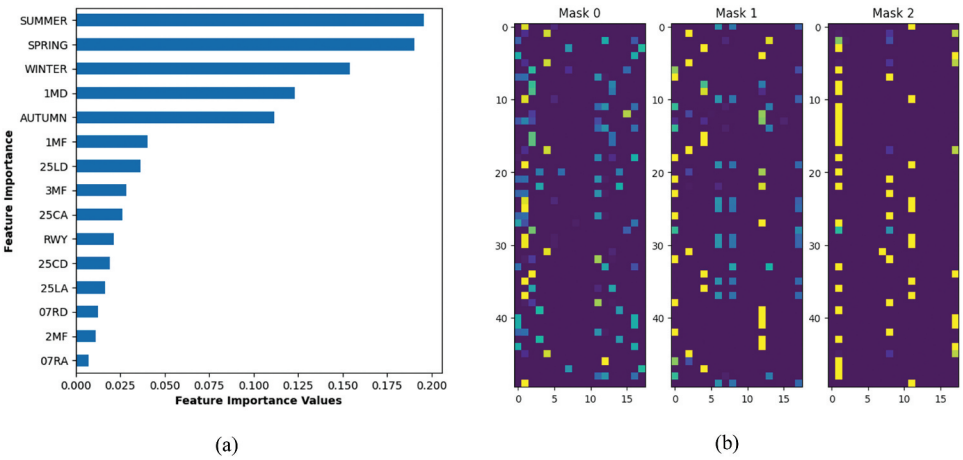

Figure 10. Global interpretation; (a) permutation-based feature importance via TabNet; (b) mask-based feature of TabNet.

the local topography can further enhance wind shear effects (Chan and Li 2020; Leung et al. 2018). In addition, summer in Hong Kong is characterized by high temperatures and high humidity, which can lead to the development of convective weather, including thunderstorms. Thunderstorms often bring strong updrafts and downdrafts, creating turbulent conditions and wind shear near HKIA (Chen et al. 2020).

Similarly, in the context of wind shear encounter location, 1MD was more prone to the occurrence of severe wind shear events. Among all the runways at HKIA for both departures and arrivals, Runway 25LD was most susceptible to the occurrence of severe wind shear events. Several factors, including the airport's location and local meteorological conditions, contribute to the occurrence of severe wind shear near the runway. As mentioned previously, HKIA is located in a region characterized by complex terrain, encompassed by hills and mountains. As air flows over and around these elevated features, it can create variations in wind speed and direction. These changes in wind patterns can result in severe wind shear very close to the runway (Stocker et al. 2022). HKIA is also located near the coast of the South China Sea, and its proximity to it can contribute to the development of strong wind shear as well. The interaction between the land and sea breezes, as well as the influence of nearby waterbodies, can generate turbulent wind conditions near the runway (Hon and Chan 2022).

In the context of mask-based feature importance, the attention masks in TabNet allocate importance scores to features at each decision step. The aforementioned scores may be employed to ascertain the local importance of features for every tested instance. The current study involved the allocation of three decision steps, which led to the creation of three local feature importance masks, denoted as mask 0, mask 1, and mask 2, corresponding to steps 1, 2, and 3, respectively. The heatmap depicted in Figure 10b showcases the local significance of feature mask scores for the first 50 instances.

Local interpretability is necessary for understanding the prediction process in individual instances. Figure 11 displays the local interpretability analysis conducted using LIME. The analysis is capable of providing model interpretations for individual instances (any i^{th} instance). Figure 11 displays bar graphs that represent the contribution of each feature to the prediction and classification of the corresponding i^{th} instance. The LIME strategy is notable for its ability to accurately determine the influence of individual features on the outcome by presenting it as a probability. If for example, $i = 43$, the BO-TabNet showed that there was a 98% chance that the event was a severe-wind shear event. In this instance, spring, autumn, summer, and runway 25CA resulted most in the likelihood of severe wind shear. Similarly, for instance, if $i = 72$, the BO-TabNet showed that there was a 71% chance that the event was a non-severe-wind shear event. In this instance, winter, 2MD, and autumn were the three contributing factors.

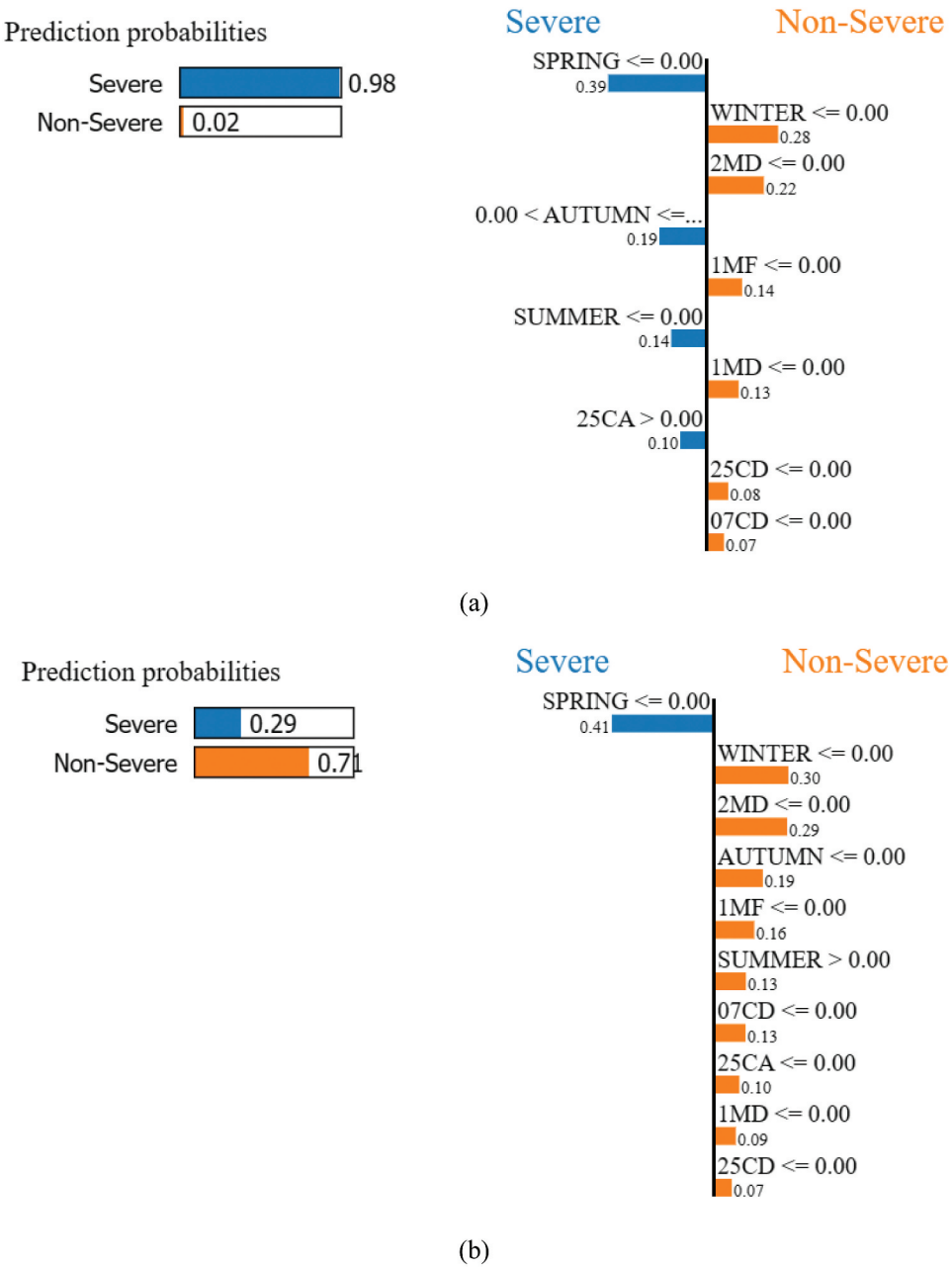

Figure 11. Local interpretation by LIME; (a) Correct classification of a random instance (*i*=43) as severe wind shear; (b) Correct classification of a random instance (*i*=72) as non-severe wind shear.

Conclusion and Recommendations

This study presents a novel TabNet, a cutting-edge deep learning model, for the prediction of wind shear severity. The data utilized for this purpose were obtained from the Doppler LiDAR system located at HKIA. Various factors were considered, such as the approach and departure runways of

HKIA, the season of the year, and the locations where wind shear hits occurred. Various resampling strategies were employed during the pre-processing stage to address the imbalanced data, which was caused by a limited number of severe wind shear events. The proposed model trained in conjunction with Bayesian optimization provided a level of accuracy comparable to that of other deep learning models, including FCN and its machine learning counterpart, XGBoost, while being explicable and innately comprehensible. In addition, the LIME strategy was also employed to assess different factors from a local perspective. From this research study, the following conclusions can be drawn:

- Based on the testing dataset, the finely tuned TabNet model using balanced data outperformed other
- The performance of BO-TabNet using data treated with different resampling strategies differed slightly but was comparable. There was a significant difference in the results when compared to the untreated data.
- The BO-TabNet model with SVM-SMOTE-treated data resulted in a higher MCC (0.36), G-Mean (0.73), and AU-ROC (0.73) as compared to the other models.
- When the SMOTE-ENN strategy was used to treat the imbalanced HKIA-based Doppler LiDAR data, the BO-TabNet model had the worst MCC (0.18), G-Mean (63), and AU-ROC (0.64) performance. However, the results were still better than the BO-TabNet using untreated data. The BO-TabNet using untreated data resulted in MCC (0.12), G-Mean (0.49), and AU-ROC (0.62).
- The BO-TabNet model also demonstrated efficacy in interpretation from a global. In terms of factor importance, summer was the most influential factor contributing to the severe wind shear occurrence. One nautical mile (1 MD) from the departure end of the runway was another influential factor. Similarly, 25LD came out to be a highly susceptible runway to the occurrence of severe wind shear.

The potential application of the proposed deep learning BO-TabNet framework lies in its ability to conduct a comprehensive evaluation of wind shear severity in the vicinity of airport runways. Undoubtedly, this resource holds significant value for individuals engaged in the domain of civil aviation. By predicting wind shear near runways, aviation stakeholders can proactively manage and mitigate the associated risks, leading to safer and more efficient operations. The potential areas for future investigation may encompass the impact of turbulence on flight operations as well as the psychological aspects influencing pilots, among other relevant factors. In addition to Doppler LiDAR data, Pilot Report (PIREPs) data will be taken into account to consider the altitude of occurrence of wind shear in the vicinity of the runway.

Acknowledgements

The present study was financially supported by the Research Fund for International Young Scientists of the National Natural Science Foundation of China (Grant No. 52250410351), the National Foreign Expert Project (Grant No. QN2022133001L), and the National Natural Science Foundation of China (U1733113). The authors express gratitude to the Hong Kong Observatory of Hong Kong International Airport for their provision of Doppler LiDAR data, which were utilized in this study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the National Natural Science Foundation of China [52250410351].

Data availability statement

The data that support the findings of this study are available on request from the corresponding author, Feng Chen.

References

Airport Council International. 2017. *World airport traffic forecast 2017–2040 airport council international*. Montréal, QC, USA: Airport Council International.

Alahmari, F. 2020. A comparison of resampling techniques for medical data using machine learning. *Journal of Information & Knowledge Management* 19 (1):2040016. doi:10.1142/S021964922040016X.

Alkhamisi, A. O., and R. Mehmood. 2020, March. An ensemble machine and deep learning model for risk prediction in aviation systems. In *2020 6th Conference on Data Science and Machine Learning Applications (CDMA)*, 54–59. IEEE.

Arik, S. Ö., and T. Pfister. 2021, May. Tabnet: Attentive interpretable tabular learning. *Proceedings of the AAAI Conference on Artificial Intelligence* 35(8):6679–87. doi:10.1609/aaai.v35i8.16826.

Batista, G. E., R. C. Prati, and M. C. Monard. 2005, September. Balancing strategies and class overlapping. International symposium on intelligent data analysis, 24–35. Berlin Heidelberg: Berlin, Heidelberg: Springer.

Ben-Cohen, A., I. Diamant, E. Klang, M. Amitai, and H. Greenspan. 2016. Fully convolutional network for liver segmentation and lesions detection. Deep Learning and Data Labeling for Medical Applications: First International Workshop, LABELS 2016, and Second International Workshop, DLMIA 2016, Held in Conjunction with MICCAI 2016, Proceedings 1, 77–85. Athens, Greece: Springer International Publishing. October 21.

Bergstra, J., R. Bardenet, Y. Bengio, and B. Kégl. 2011. Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems* 24.

Bergstra, J., and Y. Bengio. 2012. Random search for hyper-parameter optimization. *Journal of Machine Learning Research* 13 (2).

Borsky, S., and C. Unterberger. 2019. Bad weather and flight delays: The impact of sudden and slow onset weather events. *Economics of Transportation* 18:10–26. doi:10.1016/j.ecotra.2019.02.002.

Cankaya, B., K. Topuz, D. Delen, and A. Glassman. 2023. Evidence-Based Managerial decision-making with machine learning: The case of bayesian inference in aviation incidents. *Omega* 120:102906. doi:10.1016/j.omega.2023.102906.

Chan, P. W. 2006, January. 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*, Atlanta, GA.

Chan, P. W. 2012. A significant wind shear event leading to aircraft diversion at the Hong Kong international airport. *Meteorological Applications* 19 (1):10–16. doi:10.1002/met.242.

Chan, P. W. 2017. Severe wind shear at Hong Kong International airport: Climatology and case studies. *Meteorological Applications* 24 (3):397–403. doi:10.1002/met.1637.

Chang, L., G. Xing, H. Yin, L. Fan, R. Zhang, N. Zhao, F. Huang, and J. Ma. 2023. Landslide susceptibility evaluation and interpretability analysis of typical loess areas based on deep learning. *Natural Hazards Research* 3 (2):155–169. doi:10.1016/j.nhres.2023.02.005.

Chan, P. W., and Q. S. Li. 2020. Some observations of low level wind shear at the Hong Kong International Airport in association with tropical cyclones. *Meteorological Applications* 27 (2):e1898. doi:10.1002/met.1898.

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. SMOTE: Synthetic minority over-sampling technique. *The Journal of Artificial Intelligence Research* 16:321–357. doi:10.1613/jair.953.

Chen, T., and C. Guestrin. 2016, August. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, San Francisco, USA, 785–94.

Chen, F., H. Peng, P. W. Chan, X. Ma, and X. Zeng. 2020. Assessing the risk of windshear occurrence at HKIA using rare-event logistic regression. *Meteorological Applications* 27 (6): e1962. doi:10.1002/met.1962.

Choi, S., Y. J. Kim, S. Briceno, and D. Mavris. 2016, September. Prediction of weather-induced airline delays based on machine learning algorithms. In *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, Sacramento, USA, 1–6. IEEE.

Cho, H., Y. Kim, E. Lee, D. Choi, Y. Lee, and W. Rhee. 2020. Basic enhancement strategies when using Bayesian optimization for hyperparameter tuning of deep neural networks. *Institute of Electrical and Electronics Engineers Access* 8:52588–52608. doi:10.1109/ACCESS.2020.2981072.

Coburn, J., J. Arnheim, and S. C. Pryor. 2022. Short-Term forecasting of wind gusts at airports across CONUS using machine learning. *Earth & Space Science* 9 (12):e2022EA002486. doi:10.1029/2022EA002486.

Dhief, I., Z. Wang, M. Liang, S. Alam, M. Schultz, and D. Delahaye. 2020, September. Predicting aircraft landing time in extended-TMA using machine learning methods. In *Proceedings of the 9th International Conference for Research in Air Transportation (ICRAT)*, Tampa, FL, USA, 23–26.

Frazier, P. I. 2018. A tutorial on Bayesian optimization. arXiv preprint arXiv:1807.02811.

Gultepe, I., R. Sharman, P. D. Williams, B. Zhou, G. Ellrod, P. Minnis, S. Trier, S. Griffin, S. Yum, B. Gharabaghi, et al. 2019. A review of high impact weather for aviation meteorology. *Pure & Applied Geophysics* 176 (5):1869–1921. doi:10.1007/s00024-019-02168-6.

Han, H., W. Y. Wang, and B. H. Mao. 2005, August. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In *International conference on intelligent computing*, Berlin, Heidelberg, 878–87. Springer.

Hon, K. K., and P. W. Chan. 2022. Historical analysis (2001–2019) of low-level wind shear at the Hong Kong International Airport. *Meteorological Applications* 29 (2):e2063. doi:10.1002/met.2063.

Khattak, A., P. W. Chan, F. Chen, and H. Peng. 2022. Prediction of a Pilot's invisible foe: The severe low-level wind shear. *Atmosphere* 14 (1):37. doi:10.3390/atmos14010037.

Khattak, A., P. W. Chan, F. Chen, and H. Peng. 2023a. Assessing wind field characteristics along the airport runway glide slope: An explainable boosting machine-assisted wind tunnel study. *Scientific Reports* 13 (1):10939. doi:10.1038/s41598-023-36495-5.

Khattak, A., P. W. Chan, F. Chen, and H. Peng. 2023b. Time-series prediction of intense wind shear using machine learning algorithms: A case study of Hong Kong International Airport. *Atmosphere* 14 (2):268. doi:10.3390/atmos14020268.

Khattak, A., P. W. Chan, F. Chen, H. Peng, and C. Mongina Matara. 2023. Missed approach, a safety-critical go-around procedure in aviation: Prediction based on machine learning-ensemble imbalance learning. *Advances in Meteorology* 2023:1–24. doi:10.1155/2023/9119521.

Leung, M. Y., W. Zhou, C. M. Shun, and P. W. Chan. 2018. Large-scale circulation control of the occurrence of low-level turbulence at Hong Kong International Airport. *Advances in Atmospheric Sciences* 35 (4):435–44. doi:10.1007/s00376-017-7118-y.

Liu, Z. 2023, February. A new porosity prediction method based on Deep Learning of TabNet Algorithm. In *2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, Changchun, China, 1681–85. IEEE.

Liu, J. N., K. M. Kwong, and P. W. Chan. 2012. Chaotic oscillatory-based neural network for wind shear and turbulence forecast with LiDAR data. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42 (6):1412–23. doi:10.1109/TSMCC.2012.2188284.

Lundberg, S. M., and S. I. Lee. 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* 30.

Luo, M., M. Schultz, H. Fricke, B. Desart, F. Herrema, and R. B. Montes. 2021, October. Agent-based simulation for aircraft stand operations to predict ground time using machine learning. In *2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC)*, San Antonio, Texas, USA, 1–8. IEEE.

McDonnell, K., F. Murphy, B. Sheehan, L. Masello, and G. Castignani. 2023. Deep learning in insurance: Accuracy and model interpretability using TabNet. *Expert Systems with Applications* 217:119543. doi:10.1016/j.eswa.2023.119543.

Mizuno, S., H. Ohba, and K. Ito. 2022. Machine learning-based turbulence-risk prediction method for the safe operation of aircrafts. *Journal of Big Data* 9 (1):29. doi:10.1186/s40537-022-00584-5.

Molnar. C., 2020. Interpretable machine learning. Lulu. com.

Nechaj, P., L. Gaál, J. Bartok, O. Vorobyeva, M. Gera, M. Kelemen, and V. Polishchuk. 2019. Monitoring of low-level wind shear by ground-based 3D lidar for increased flight safety, protection of human lives and health. *International Journal of Environmental Research and Public Health* 16 (22):4584. doi:10.3390/ijerph16224584.

Palatnik de Sousa, I., M. Maria Bernardes Rebuzzi Vellasco, and E. Costa da Silva. 2019. Local interpretable model-agnostic explanations for classification of lymph node metastases. *Sensors* 19 (13):2969. doi:10.3390/s19132969.

Pouyanfar, S., S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M. L. Shyu, S. C. Chen, and S. S. Iyengar. 2018. A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)* 51 (5):1–36. doi:10.1145/3234150.

Qu, J., T. Zhao, M. Ye, J. Li, and C. Liu. 2020. Flight delay prediction using deep convolutional neural network based on fusion of meteorological data. *Neural Processing Letters* 52 (2):1461–1484. doi:10.1007/s11063-020-10318-4.

Ryan, M., A. H. Saputro, and A. Sopaheluwakan. 2023. Review of low-level wind shear: Detection and prediction. In *AIP Conference Proceedings*, Palu, Indonesia, vol. 2719. AIP Publishing.

Schultz, M., and S. Reitmann. 2019. Machine learning approach to predict aircraft boarding. *Transportation Research Part C: Emerging Technologies* 98:391–408. doi:10.1016/j.trc.2018.09.007.

Shun, C. M., and P. W. Chan. 2008. Applications of an infrared Doppler lidar in detection of wind shear. *Journal of Atmospheric and Oceanic Technology* 25 (5):637–655. doi:10.1175/2007JTECHA1057.1.

Snoek, J., H. Larochelle, and R. P. Adams. 2012. Practical bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems* 25.

Stocker, J., K. Johnson, R. Jackson, S. Smith, D. Connolly, D. Carruthers, and P. W. Chan. 2022. Hong Kong Airport wind shear now-casting system development and evaluation. *Atmosphere* 13 (12):2094. doi:10.3390/atmos13122094.

Szeto, K. C., and P. W. Chan, 2006, January. High resolution numerical modelling of windshear episodes at the Hong Kong International Airport. In *12th Conference on Aviation, Range, and Aerospace Meteorology*.

Thobois, L., J. P. Cariou, and I. Gultepe. 2019. Review of lidar-based applications for aviation weather. *Pure & Applied Geophysics* 176 (5):1959–1976. doi:10.1007/s00024-018-2058-8.

Wicaksono, A. S., and A. A. Supianto. 2018. Hyper parameter optimization using genetic algorithm on machine learning methods for online news popularity prediction. *International Journal of Advanced Computer Science & Applications* 9 (12). doi:10.14569/IJACSA.2018.091238.

Yang, L., P. Li, R. Xue, X. Ma, X. Li, and Z. Wang. 2018, July. Intelligent classification model for railway signal equipment fault based on SMOTE and ensemble learning. *IOP Conference Series: Materials Science & Engineering* 383 (1):012042. IOP Publishing. doi:10.1088/1757-899X/383/1/012042.

Yan, J., T. Xu, Y. Yu, and H. Xu. 2021. Rainfall forecast model based on the tabnet model. *Water* 13 (9):1272. doi:10.3390/w13091272.

Zhang, H., S. Wu, Q. Wang, B. Liu, B. Yin, and X. Zhai. 2019. Airport low-level wind shear lidar observation at Beijing Capital International airport. *Infrared Physics & Technology* 96:113–122. doi:10.1016/j.infrared.2018.07.033.