IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 14 June 2024, accepted 4 July 2024, date of publication 11 July 2024, date of current version 30 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3426619

RESEARCH ARTICLE

Estimating Turbulence Due to Low-Level Wind Shear in Airport Runway Zones Using TabNet-SHAP Framework

AFAQ KHATTAK[1], JIANPING ZHANG[2], PAK-WAI CHAN[3], FENG CHEN[1], ABDULRAZAK H. ALMALIKI[4], AND CAROLINE MONGINA MATARA[5]

[1]Key Laboratory of Infrastructure Durability and Operation Safety in Airfield of CAAC, Tongji University, Jiading, Shanghai 201804, China
[2]Second Research Institute, Civil Aviation Administration of China, Civil Unmanned Aircraft Traffic Management Key Laboratory of Sichuan Province, Beijing 100710, China
[3]Hong Kong Observatory, Kowloon, Hong Kong, China
[4]Department of Civil Engineering, College of Engineering, Taif University, Taif 21974, Saudi Arabia
[5]Department of Civil and Construction Engineering, University of Nairobi, Nairobi 00100, Kenya

Corresponding authors: Feng Chen (fengchen@tongji.edu.cn) and Afaq Khattak (khattak@tongji.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 52250410351, in part by the National Foreign Expert Project under Grant QN2022133001L, and in part by Xiaomi Young Talent Program.

ABSTRACT In this research, we present an advanced predictive framework designed to assess the turbulence induced by low-level wind shear near the runways at Hong Kong International Airport (HKIA), utilizing data from Pilot Reports (PIREP). This framework integrates the TabNet architecture with SHapley Additive exPlanations (SHAP), thereby enhancing both predictive accuracy and interpretability. Given the imbalance in the PIREP data, we implement various data augmentation techniques and employ Bayesian optimization to fine-tune the hyperparameters of the TabNet model. Our analysis revealed that the TabNet model, when applied to data balanced with the Support Vector Machine - Synthetic Minority Over-sampling Technique (SVM-SMOTE), demonstrated superior performance, evidenced by a Geometric Mean (G-Mean) of 0.74, a Matthews Correlation Coefficient (MCC) of 0.37, a Balanced Accuracy (BA) of 0.74, and an Area Under the Receiver Operating Characteristic (AU-ROC) of 0.739. SHAP analysis further enhanced the interpretability of the TabNet model by identifying key contributing factors, including the magnitude of wind shear and altitude, which significantly influence the likelihood of significant turbulence occurrence. Specifically, SHAP insights demonstrated the critical impact of wind shears between 15 and 25 knots and the pronounced effects of complex terrain and sea breezes on significant turbulence occurrence, predominantly at altitudes below 1200 feet. These findings not only demonstrate the efficacy of the TabNet-SHAP model in enhancing aviation safety through improved turbulence prediction but also provide actionable insights for operational and safety protocols at airports prone to low-level wind shear.

INDEX TERMS Aviation turbulence, civil aviation safety, SHAP, TabNet, wind shear.

I. INTRODUCTION

Low-level wind shear is recognized as a critical issue in aviation safety due to its ability to cause significant distress to passengers and crew, impair aircraft controllability, and potentially inflict structural damage on the air frame [1], [2], [3]. The International Civil Aviation Organization (ICAO) recognizes low-level wind shear, occurring below 1,600 feet and/or within 3 nautical miles from the runway's threshold, as a significant risk to aircraft during takeoff and landing phases. This phenomenon can induce substantial aviation turbulence, necessitating the abandonment of landing attempts by incoming flights.

The ICAO utilizes the eddy dissipation rate (EDR) metric, specifically its cube root, to quantify aviation turbulence intensity. EDR values in the range of 0.00 to

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Ni.

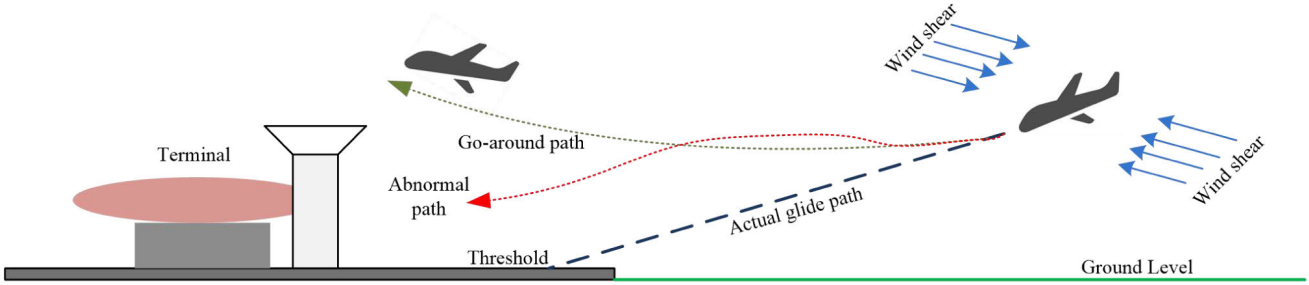


FIGURE 1. Effect of turbulence due to low-level wind shear events on approaching aircraft.

0.299 m^2s^{-3} signify low turbulence, while values between 0.30 to 0.49 m^2s^{-3} indicate moderate turbulence. EDR values exceeding 0.50 m^2s^{-3} are associated with severe turbulence [4]. Notably, significant turbulence encompasses both moderate and severe turbulence categories, representing considerable disturbances that can affect flight safety and operations. Figure 1 demonstrates potential anomalies during significant turbulence events caused by low-level wind shear, including aborted landings and deviations from the designated glide path during final approach.

Pilot Reports (PIREPs) are an essential and real-time resource for capturing observations of turbulence encountered during flight. These reports offer detailed accounts of the meteorological conditions experienced, encompassing information such as the date, time, geographical coordinates (latitude and longitude), and altitude, typically expressed as flight level [5]. Traditionally, PIREPs are communicated via radio to relevant ground stations for subsequent dissemination. In some instances, these observations may also be relayed via telephone communication post-landing [6].

The importance of evaluating significant turbulence caused by low-level wind shear is paramount within the aviation sector. Such assessments are critical for the formulation of vital mitigation strategies, which include the revision of aircraft operational procedures and the enhancement of pilot training programs. By integrating these strategies into operational practices, the aviation sector can proactively improve safety measures and optimize flight operations.

Understanding and accurately reporting significant turbulence through PIREPs enables the development of a comprehensive safety framework. This framework can lead to the advancement of flight safety protocols, ensuring that pilots are better prepared to handle adverse weather conditions. Additionally, these reports contribute to a larger database of meteorological information, aiding in the predictive modeling of turbulence and the continuous improvement of aircraft design and technology.

Advancements in the estimation of turbulence have markedly enhanced flight safety, especially in instances triggered by wind shear events. These improvements have been achieved through the deployment of in situ sensors, the application of remote turbulence sensors such as Doppler Light Detection and Ranging (LiDAR) systems, the incorporation of airborne sensors, and the use of radar technology [7], [8], [9], [10]. Furthermore, extensive computational efforts have led to the development of numerical simulations and algorithms that link radar detection data with turbulence intensity [11], [12], [13], [14] These computational tools are essential for the analysis of radar data, offering insights into the severity and nature of turbulence. This, in turn, allows pilots and air traffic controllers to make well-informed decisions and implement necessary safety precautions. Nonetheless, the accuracy of EDR values, estimated from Doppler LiDAR or radar data, is affected by data quality issues and the intrinsic limitations of the physical models used.

Recently, the adoption of Artificial Intelligence (AI) techniques has seen a notable increase across various sectors [15], [16], [17]. This surge in AI relevance is attributed to the growing volume and complexity of data, coupled with the enhanced availability and cost-effectiveness of computing power. These aspects of digital transformation facilitate the swift and autonomous generation of models that can process large and complex datasets with high efficiency and accuracy [18], [19]. Despite significant advancements in AI applications across various fields, its utilization in civil aviation safety is still relatively nascent. Table 1 provides a comprehensive review of the latest AI algorithms employed in aviation meteorology and operational safety.

The effective prediction of turbulence using AI models based on PIREPs data is hindered by two primary obstacles. The first challenge is the significant imbalance within the PIREP data, where there is a higher frequency of reports indicating low levels of turbulence compared to a limited number of reports on medium to severe (significant) turbulence events. The second challenge revolves around the lack of transparency in the existing AI models, often referred to as their ''black-box'' nature. This opacity makes it difficult to understand and trace the underlying logic behind the predictions made by these AI models.

To address the challenges of data imbalance and interpretability in predicting turbulence, we propose a three-phase TabNet-SHAP strategy. In the first phase, we balance the PIREPs data using various augmentation strategies such as Borderline-Synthetic Minority Over-sampling Technique (Borderline-SMOTE) [33], hybrid Support Vector

TABLE 1. Review of the latest artificial intelligence models in the field of aviation meteorology, operation and safety.

Key Factors	Detailed Description	AI Approach	Relevant Literature
Aviation Meteorological Aspects	Model to forecast intense low-level wind shear using Doppler LiDAR data of HKIA	Extreme Gradient Boosting (XGBoost)	[20]
	Application of AI technique artificial to improve real-time decision-making for high-impact weather	Random Forest (RF)	[21]
	Assessment of effects of turbulence on flights from Matsumoto Airport	Hybrid Principal Component Analysis (PCA) and K-Means	[22]
	Prediction of runway visual range at different airports in India	Hybrid Convolutional neural networks (CNN) and long-short-term memory (LSTM)	[23]
	Estimation of wind shear severity based on Doppler LiDAR and PIREPs data	BalanceCascade with SHAP	[24]
	Developing a Machine Learning Model to Identify Flight Turbulence Using LiDAR Data	Conditional Generative Adversarial Network (CGAN) and XGBoost	[25]
Aviation Operations and Safety Aspects	Integrated Machine Learning Models for Predicting Aviation Incident Risks	Hybrid Support Vector Machine(SVM) and DNN	[26]
	Using Bayesian Neural Networks for Predicting Flight Trajectories and Assessing Safety with FAA's System Wide Information Management Data	DNN and LSTM	[27]
	Flight delay prediction based on surveillance-broadcast (ADS-B) messages	LSTM	[28]
	Classification and prediction of the daily operations at Newark International Airport (EWR	Boosting Ensemble Learning Model	[29]
	Forecasting Critical Safety Metrics for Landings Using Data from Commercial Aircraft On-Board Recorders	RF	[30]
	Prediction and analysis of flight departure delays	SVM	[31]
	Estimation of aircraft estimated arrival time based on general flight information, trajectory data and weather data	RF and DNN	[32]

Machine- Synthetic Minority Over-sampling Technique (SVM-SMOTE) [34], Near Miss [35], Adaptive Synthetic (ADASYN) [36], and hybrid Synthetic Minority Over-sampling Technique - Edited Nearest Neighbour (SMOTE-ENN) [37]. This ensures a more suitable dataset for training the machine learning models. In the second phase, we employ the state-of-the-art TabNet model for effective classification and prediction of turbulence [38]. TabNet is chosen for its ability to handle tabular data and its strong performance in various machine learning tasks [39], [40]. The approach draws on the advantages of deep learning while aiming to overcome its shortcomings in processing tabular data [3], [41]. The Bayesian optimization is used to obtain the optimal hyperparameters of TabNet model [42]. In the final phase, we integrate Shapley Additive explanations (SHAP) analysis [43] with the TabNet model to enhance interpretability. SHAP analysis will allows us to understand the contribution of different factors in the prediction of turbulence, providing valuable insights into the underlying factors and patterns. By utilizing the Bayesian-tuned TabNet model and incorporating SHAP analysis, our research not only seeks to enhance the predictive accuracy of turbulence prediction but also to offer clear, interpretable explanations for the predictions of TabNet model. Figure 2 depicts the three-phase framework used for predicting and interpreting turbulence caused by wind shear events.

This study outlines its key contributions as follows:

- The implementation of a deep learning classifier, TabNet, for predicting and classifying turbulence, utilizing

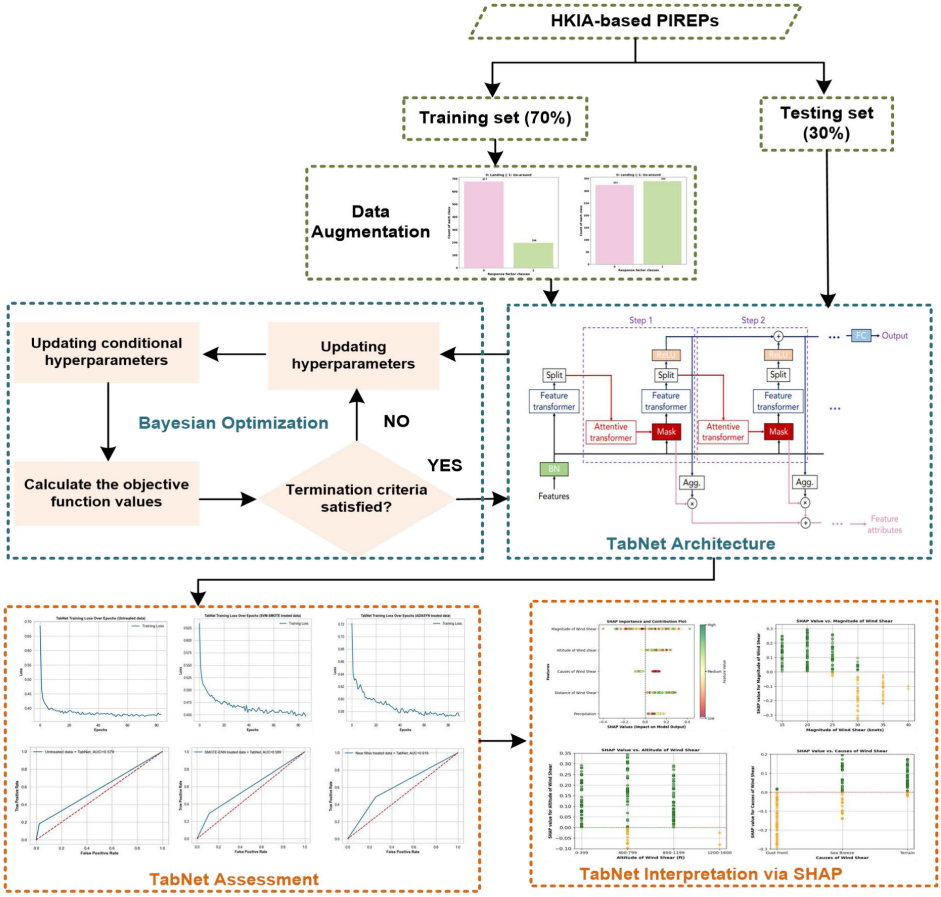


FIGURE 2. Proposed TabNet-SHAP strategy for predicting and interpreting turbulence due to wind shear in the airport runway zones.

PIREPs data gathered from HKIA. Through the adoption of various data augmentation strategies, coupled with the application of Bayesian Optimization, the TabNet model's learning capabilities have been notably enhanced by fine-tuning its hyperparameters.

- Furthermore, the interpretability of the TabNet model's classification outcomes on turbulence has been augmented using SHAP analysis, providing insights into contributing factors. This includes analysis of important individual factors and their interactions.

The following sections of the paper are organized as follows: Section II illustrates detailed information about the study location, HKIA-based PIREPs data, and provides an overview of the TabNet architecture, Bayesian optimization, SHAP analysis, and performance metrics. Section III shows the results of the study and their implications. Finally, Section IV provides conclusion, summarizing the findings and their significance in the realm of civil aviation safety.

II. DATA AND METHOD

A. STUDY LOCATION

In this research, we assess the levels of turbulence near the runways at Hong Kong International Airport (HKIA), which is the primary air transport gateway for Hong Kong. It is situated on Lantau Island, an artificially expanded island located near the subtropical coast of mainland China, as shown in Figure 3. This strategic positioning allows HKIA to serve as a significant hub for international air traffic, capitalizing on its proximity to the economic powerhouses of the Asia-Pacific region [44]. The airport's location on Lantau Island, coupled with its advanced infrastructure, has been instrumental in its development into one of the world's busiest and most efficient airports, facilitating passenger and cargo flights globally as shown in Figure 4 [45]. However, this geographical area is frequently subjected to convective weather patterns, characterized by tropical cyclones and the influence of the southwest monsoon system [46], [47]. These climatic conditions are well-known for their role in producing, wind shear, thunderstorms and substantial rainfall, frequently leading to interruptions in flight schedules and delays at HKIA [48]. The propensity for such weather phenomena in this region often precedes challenges in maintaining smooth aviation operations.

Extensive studies in the past have highlighted the vulnerability of HKIA to a range of adverse meteorological events [49]. The airport's exposure to extreme weather conditions, including typhoons, substantial rainfall episodes, wind shear phenomena with speeds varying between 14 to

FIGURE 3. Lantau island to the south of hong kong international airport.

FIGURE 4. Structures around and within hong kong international airport.

30 knots, notable instances of turbulence, and intense crosswinds impacting runway operations, has been well reported [50], [51], [52]. These elements underscore the critical need for ongoing research into mitigating the impacts of severe weather on aviation safety and operational efficiency at HKIA.

B. PIREPs DATA DESCRIPTION AND PRE-PROCESSING

The Hong Kong Observatory (HKO) regularly gathers PIREPs to document instances of low-level wind shear and turbulence experienced during aircraft takeoff or landing at HKIA. Consequently, the data used in this study were specifically sourced from PIREPs obtained at HKIA. These reports are crucial for pilots to communicate real-time weather conditions that could pose hazards. The HKIA-based PIREPs used in our study provide extensive details on the prevalence and incidents of wind shear, as well as observations on turbulence caused by these wind shear events. Our analysis of these reports revealed a consistent pattern of concurrent reports of turbulence and wind shear events. Pilots use a standardized numerical system to classify the intensity of the turbulence encountered during flight, as shown in Table 2.

This approach ensures a consistent and measurable system for reporting turbulence, enabling a more systematic analysis and understanding of turbulence events around HKIA runways. These reports provide detailed insights into the intensity, origins, and frequency of wind shear occurrences at various altitudes above the runway. As shown in Figure 5, these reports carefully document wind shear events at specific horizontal distances from the runway threshold, focusing on occurrences at 1 nautical mile (1-MF/MD) and 2 nautical miles (2-MF/MD) from both the approach and departure ends of the runway. This is crucial for analyzing the spatial distribution of wind shear incidents and their potential impact on aviation safety during different phases of flight, such as landing and takeoff.

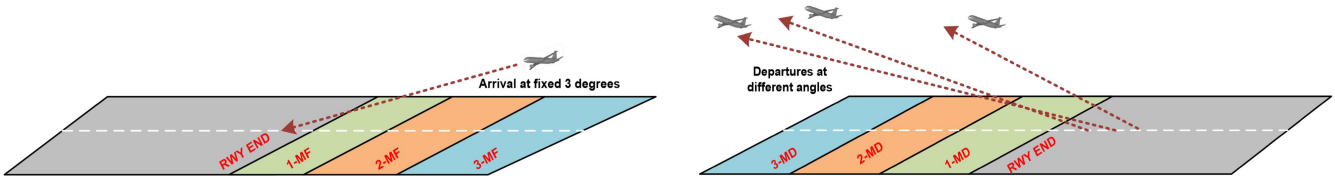
TABLE 2. Numerical values assigned to the different levels of turbulence due to wind shear.

Turbulence Level	Assigned Numerical values
Null	0
Light	1
Light to moderate	1.5
Moderate	2
Moderate to severe	2.5
Severe	3

Additionally, PIREPs indicate causes of wind shear, such as gust fronts or sea breezes. It is important to note that, without specific terminology for pilots to identify building-related wind shear events, their identification must follow established criteria. Firstly, these events should be reported under clear skies, free of precipitation or thunderstorms at HKIA. Secondly, the focus is on incidents reported within 2 nautical miles of touchdown. Table 3 presents a subset of data derived from HKIA-based PIREPs, providing a glimpse into the extensive dataset available for this study.

In most HKIA-based PIREPs, the notation ''Null'' frequently indicates the absence of turbulence during takeoff or landing phases. Instances of light or light-to-moderate turbulence, while occasionally recorded, are typically managed by pilots through established training protocols. Conversely, occurrences of moderate-to-severe or severe turbulence are less common but raise significant safety concerns. To address this, we propose a binary classification framework where observations categorized as null, light, or light-to-moderate turbulence are designated as the negative class (0), labeled as Insignificant Turbulence (IST). Conversely, reports indicating moderate, moderate-to-severe, and severe turbulence are classified as the positive class (1), termed Significant Turbulence (ST), underscoring their critical importance to aviation safety analysis, as shown in (1).

$$\text{Turbulence} = \begin{cases} 1 : \text{ST} \\ \text{Moderate/ mod erate} - \text{to} - \text{severe/severe} \\ 0 : \text{IST} \\ \text{Null/low/low} - \text{to} - \text{moderate} \end{cases} \tag{1}$$



(a) (b)

FIGURE 5. Occurrence location of wind shear events; (a) Occurrence at finals/arrival, (b) Occurrence at departure.

TABLE 3. Sample data gathered from pilot reports at hong kong international airport.

Occurrence date	Occurrence time	Assigned runway	Flight	Aircraft type	magnitude of wind shear (knots)	Wind shear distance from runway (MF/MD)	Turbulence	altitude of wind shear (ft)	Rain	Aborted Landing	causes of wind shear
2007-06-21	6:02 AM	07RD	KA992	A333	20	1	1.5	250	No	No	Terrain
2007-11-29	7:13 PM	25CA	SQ872	B772	-15	2	1.5	1000	No	Yes	Terrain
2008-04-19	1:33 PM	25CA	MH6064	B742	20	0	3	1100	Yes	Yes	Sea breeze
2011-02-15	11:11 PM	07RA	CX156	A333	-15	1	2	1000	No	No	Gust front
2012-01-17	10:41 PM	07RA	CX234	B777	-15	2	2.5	750	No	Yes	Sea breeze
2016-08-03	2:36 AM	07CA	HX497	A320	35	2	3	450	No	No	Terrain
2017-05-23	1:38 PM	25LD	HX126	A332	-20	2	1.5	1000	No	No	Gust front
2018-09-21	4:52 PM	07LA	CX543	B744	-15	1	2.5	700	No	No	Terrain
2020-07-09	8:05 AM	25CA	VS206	B789	30	3	2	900	No	No	Sea breeze
2022-03-19	12:16 PM	07CA	EK9294	B77L	15	1	1.5	1250	No	No	Gust front

In the HKIA-based PIREPs dataset, wind shear-related factors are categorized and encoded to facilitate analysis and interpretation, as illustrated in Table 4. This encoding method uses both label coding for discrete factors and direct representation for continuous factors, resulting in a structured and interpretable dataset. For instance, the magnitude of wind shear is represented as a continuous variable, quantifying the intensity of wind shear events in knots. This direct representation offers a precise measurement of wind shear strength without requiring additional encoding.

Wind shear distance from runway (MF/MD) is treated as a discrete factor and encoded to denote the event's distance from the runway. The encoding values are '0' for events occurring right at the runway (RWY), '1' for those 1-MF or 1-MD away, '2' for those 2-MF or 2-MD away, and '3' for those 3-MF or 3-MD away. This method categorizes the proximity of wind shear events relative to both the landing and takeoff phases. The altitude of wind shear (ft) is encoded as a discrete factor segmented into four categories based on altitude in feet: '0' for 0-399 ft, '1' for 400-799 ft, '2' for 800-1199 ft, and '3' for 1200-1600 ft. The causes of wind shear are also encoded as discrete values to identify their origin: '0' for terrain, '1' for sea breeze, and '2' for gust front. Additionally, the rain factor indicates whether rain was present ('1') or absent ('0') at the time of the wind shear event.

C. THEORETICAL OVERVIEW OF TabNet ARCHITECTURE

TabNet is a pioneering deep learning model tailored for tabular data, combining the interpretability of decision trees with the power and flexibility of neural networks. It stands out for its strategic use of a learnable attention mechanism to dynamically select features, making it adept at handling

TABLE 4. Label encoding of different factors extracted from HKIA-based PIREPs for the analysis of turbulence due to wind shear.

Factors from PIREPs	Data Type	Description and coding
Magnitude of wind shear (knots)	Continuous	-
Wind shear distance from runway (MF/MD)	Discrete	0: RWY; 1: 1 nautical miles; 2: 2 nautical miles; 3: 3 nautical miles
Altitude of wind shear (ft)	Continuous	0: 0-399 ft; 1: 400-799ft; 2: 800-1199ft, 3: 1200 - 1600ft
Causes of wind shear	Discrete	0: Terrain; 1: Sea Breeze; 2: Gust front
Rain	Discrete	0: No, 1: Yes

high-dimensional data and complex interactions between features. The architecture of TabNet is designed to process data through a series of decision steps, allowing for sequential attention to different subsets of features at each step as shown in Figure 6. This structure enables the model to focus on the most relevant features for making predictions, akin to how a human analyst might approach a data set by concentrating on different factors sequentially.

The TabNet introduced involves several key components such as Attentive Transformer, Feature Transformer, Decision step and Output layer.The Attentive Transformer is responsible for selecting the features to be used at each decision step. It applies a mask to the input features, determining which ones are relevant based on the current state of the model and the data as shown by (2).

$$M = Sparemax\left(FC\left(BN\left(D_{last}\right)\right)\right) \tag{2}$$

where M is the mask generated for feature selection, FC denotes a fully connected layer, BN represents batch normalization applied to the decision output from the previous step D_{last}, and Sparsemax is a sparsity-inducing function that ensures only a subset of features is selected.

Once the Attentive Transformer selects features, the Feature Transformer generates a more abstract representation of these features. It is crucial for capturing complex relationships and interactions between the selected features as shown by (3).

$$H = \mathrm{ReLU}\left(FC\left(BN\left(X \circ M\right)\right)\right) \tag{3}$$

Here, H represents the transformed features, X is the input feature set, $^\circ$ denotes element-wise multiplication with the mask M, ensuring only selected features are processed. TabNet's architecture includes multiple decision steps, each consisting of an Attentive Transformer and a Feature Transformer. These steps iteratively refine the model's focus and representation of the data, contributing piece-wise to the final prediction, which is denoted by (4).

$$D_i = \mathrm{ReLU}\left(FC\left(BN\left(H_{i-1}\right)\right)\right) \tag{4}$$

where, D_i is the decision output at step i, with H_{i-1} being the output of the Feature Transformer from the previous step. This recursive process allows for the sequential refinement of information. After passing through the designated number of decision steps, the outputs are aggregated to form the final model prediction. This aggregation captures the comprehensive analysis performed by the model across all steps as shown by (5).

$$Y = \sum\nolimits_i FC\left(D_i\right) \tag{5}$$

where, Y is the final prediction, aggregating the contributions (D_i) from each decision step through fully connected layers (FC).

D. BAYESIAN OPTIMIZATION FOR TabNet HYPERPARAMETER TUNING

Bayesian optimization is a strategic process for hyperparameter tuning, particularly suited for models like TabNet, which are complex and sensitive to hyperparameter settings. The essence of Bayesian optimization lies in its iterative approach to model the objective function (e.g., validation loss or accuracy) using a probabilistic model. The steps outlined below demonstrate the process of using Bayesian optimization to tune the hyperparameters of the TabNet model.

1) DEFINE THE HYPERPARAMETER SPACE

To begin, it is necessary to identify and specify the range for the TabNet hyperparameters, as outlined in Table 5, due to their specific role in the modeling process. This initial step basically defines the search area for the Bayesian optimization process.

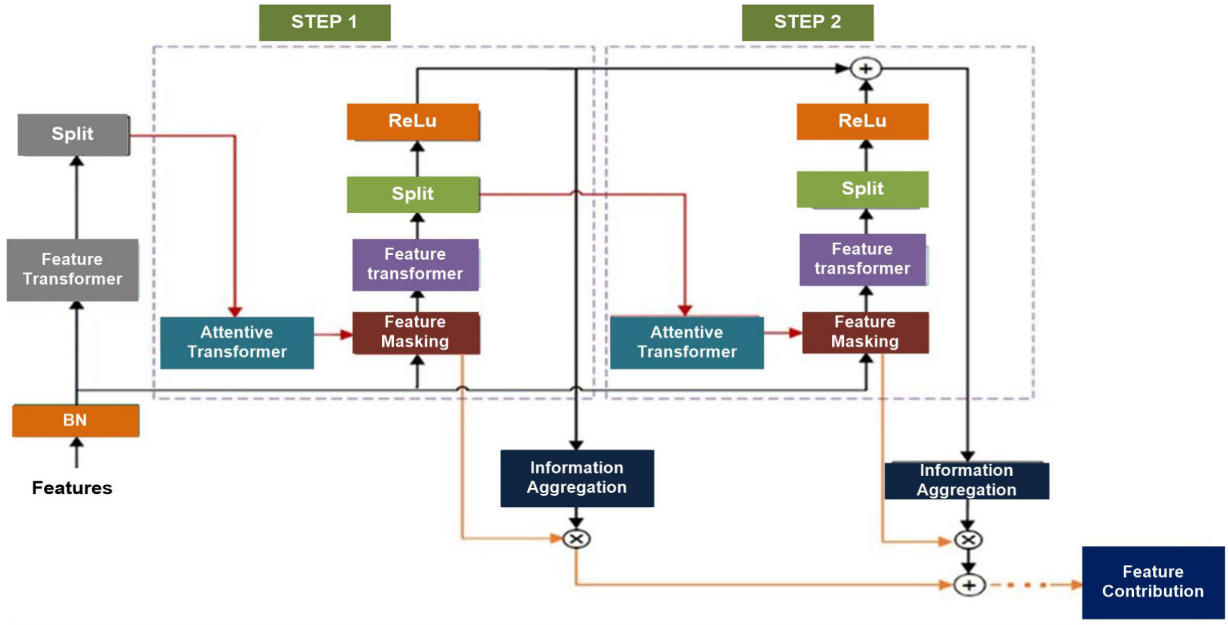


FIGURE 6. TabNet architecture.

TABLE 5. Description of hyperparameter of the TabNet model.

Hyperparameter	Designation	Function
Optimizer	optimizer	Dictates the optimization algorithm used to minimize the model's loss function during training. Common optimizer include Adam, SGD (Stochastic Gradient Descent), and RMSprop
Number of Units in Attentive Transformer	n_a	Specifies the number of units in the Attentive Transformer, which is part of the TabNet architecture responsible for selecting relevant features for making predictions at each decision step
Sparsity Regularization Coefficient	lambda_sparse	Controls the sparsity of the learned attention masks in the Attentive Transformer.
Feature Reusage Regularization Coefficient	gamma	Encourages the reuse of features at different decision steps within the model
Number of Shared Blocks	n_shared	Determines the number of shared fully connected feature transformer blocks at the bottom of the architecture.
Number of Decision Steps	n_steps	The number of decision steps in the model, where each step involves feature selection and transformation for making a part of the final prediction.
Learning rate	learning_rate	A crucial hyperparameter for most gradient-based optimization algorithms, affecting the size of the steps taken during optimization
Mask Type	mask_type	Specifies the type of masking function used in the Attentive Transformer for selecting features. Options include sparsemax and entmax

2) CHOOSE A SURROGATE MODEL

Bayesian optimization requires a surrogate model to approximate the objective function's landscape based on available data points. Gaussian Processes (GPs) are commonly used due to their ability to model the uncertainty of predictions. The prediction at any point x can be described using the Gaussian Process as (6).

$$f(x) = \mathrm{GP}\left(m(x), k\left(x, x'\right)\right) \tag{6}$$

where $m(x)$ denotes the mean function (often assumed to be zero), and the parameter $k\left(x, x'\right)$ represents the covariance function between two points x and x', expressing the expectation that points closer in the hyperparameter space will yield similar performance.

3) SELECT AN ACQUISITION FUNCTION

The acquisition function guides where to sample next by trading off exploration and exploitation. Expected Improvement (EI) is a better choice due to its effectiveness in many

TABLE 6. Description of performance measure.

Metric	Definition	Expression
Precision	It measures the proportion of true positive results in all positive predictions made by the model.	$\frac{t_p}{t_p+f_p}$
Recall	Also known as sensitivity. It assesses the model's ability to correctly identify all actual positives from the data.	$\frac{t_p}{t_p+f_n}$
F1-Score	It provides a harmonic mean of precision and recall, offering a balance between the two metrics for a comprehensive performance measure.	$2\times\frac{\left(\frac{t_p}{t_p+f_p}\right)\times\left(\frac{t_p}{t_p+f_n}\right)}{\left(\frac{t_p}{t_p+f_p}\right)+\left(\frac{t_p}{t_p+f_n}\right)}$
MCC	It is a robust metric that gives a high quality measurement of how well the classification model performs across all four quadrants of the confusion matrix (true positives, false negatives, true negatives, and false positives).	$\frac{t_p\times t_n - f_p\times f_n}{\sqrt{\left(t_p+f_p\right)\times\left(t_p+f_n\right)\times\left(t_n+f_p\right)\times\left(t_n+f_n\right)}}$
G-Mean	It is the geometric mean of recall and specificity, offering insight into the balance between sensitivity and specificity.	$\sqrt{\left(\frac{t_p}{t_p+f_n}\right)\times\left(\frac{t_n}{f_p+t_n}\right)}$
BA	It calculates the average of recall obtained on each class, thereby treating all classes equally, regardless of their size.	$\frac{1}{2}\times\left(\frac{t_p}{t_p+f_n}+\frac{t_n}{t_n+f_p}\right)$

Note: t_p : true positive, t_n : true negative, f_p : false positive, f_n : false negative

TABLE 7. Statistical descriptions of factors obtained from hkia-based pireps for turbulence due to wind shear.

Descriptions	Descriptive statistics			
	Mean	SD	Minimum	Maximum
Magnitude of wind shear (knots)	18.364	4.712	15	45
Wind shear distance from runway (MF)	1.182	1.053	0	3
Altitude of wind shear (ft)	0.491	0.784	0	3
Causes of wind shear	0.469	0.694	0	2
Rain	0.225	0.417	0	1

scenarios. Given a set of observed points, the EI at a new point x is calculated as (7).

$$\mathrm{EI}(x) = E\left[\max\left(0, f(x) - f\left(x^+\right)\right)\right] \tag{7}$$

where $f\left(x^+\right)$ denotes the objective function value of the optimal sample observed so far. EI measures the expected increase in the objective function over the current best value.

4) INITIALIZE THE PROCESS AND UPDATE THE SURROGATE MODEL

Start by selecting a few points in the hyperparameter space, either randomly or based on prior knowledge, and evaluate the TabNet model's performance at these points. These initial evaluations will form the basis for the surrogate model's first approximation. Use the outcomes of the evaluations (i.e., the performance of the TabNet model under the selected hyperparameters) to update the surrogate model.

5) OPTIMIZATION LOOP

Iteratively perform the following until a termination criteria is satisfied such as convergence to a satisfactory performance level or maximum iterations.

6) CHOOSE THE BEST HYPERPARAMETERS

After completing the optimization loop, select the set of hyperparameters associated with the best performance observed during the process.

E. SHAP ANALYSIS FOR TabNet INTERPRETATION

SHAP analysis is a method employed to enhance the transparency of predictions made by machine learning models. It assigns importance values to individual features, illuminating their role in the model's output [53]. Drawing from cooperative game theory, SHAP utilizes Shapley values, which quantify each feature's contribution to the prediction. These values are calculated using principles that apportion the overall payoff of a coalition game among players, where each player's payout is proportional to their marginal contribution. In the context of machine learning, each feature of a dataset acts like a player in the game, and the SHAP value measures the impact of including each feature in the model. This distribution is formalized mathematically in (8), which outlines how the contributions of individual features to the predictive accuracy are calculated, thereby providing insights into the workings of the model.

$$\psi_j(g,z) = \sum_{T\subseteq M\setminus\{j\}} \frac{|T|!\,(|M|-|T|-1)!}{|M|!} \times \left[g_z(T\cup\{j\}) - g_z(T)\right] \tag{8}$$

where;

- ψ_j denotes the SHAP value of the feature j,
- g represent the model, which is TabNet in our case,
- z is the specific unit for the predictions to be explained,
- M is total set of the features,
- T is the subset if M excluding feature j,
- $|T|$ and $|M|$ denotes the cardinality of sets T and M, respectively,
- $g_z(T)$ is the TabNet's model prediction with feature in set T,
- $g_z(T\cup\{j\})$ is the TabNet's model prediction with the features in T plus feature j.

This equation computes the marginal contribution of feature j by considering all possible combinations of features and averaging the change in the prediction that feature j contributes. For practical application, especially in complex models such as TabNet and big datasets, direct computation using (8) an be computationally intensive. Various approximations and algorithms, such as the Kernel SHAP method [54], have

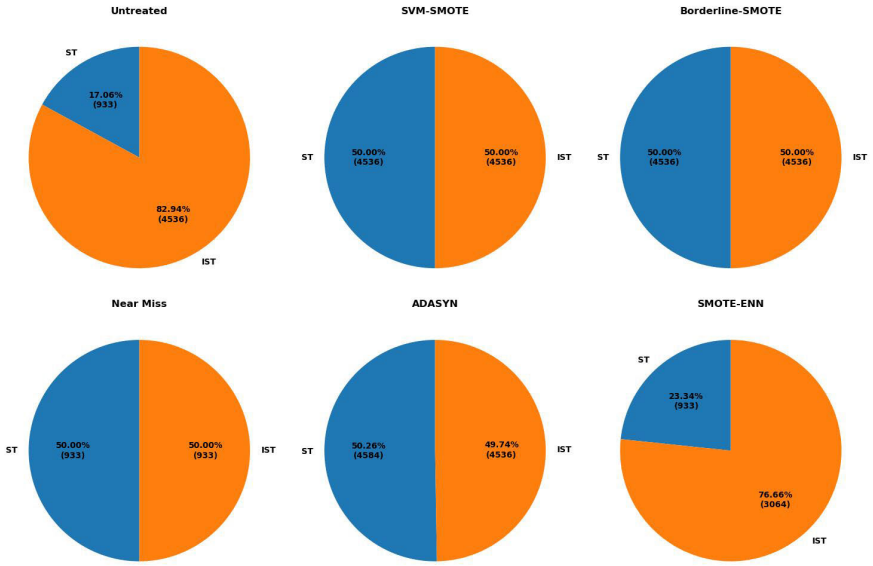


FIGURE 7. Data balancing by using various data augmentation strategies.

been developed to efficiently compute SHAP values. The Kernel SHAP method applies a weighted linear regression to estimate SHAP values, simplifying the computation as shown in (9).

$$\phi = \left(X^T W X\right)^{-1} X^T W y \tag{9}$$

where:

- X is a matrix of binary indicators representing the presence or absence of features in subsets,
- W is a diagonal matrix of weights for each subset,
- y is a vector of TabNet model outputs for each subset.

F. PERFORMANCE MEASURES

The evaluation of TabNet's model effectiveness, both in prediction and classification tasks, is conducted through the analysis of six key performance metrics: precision, recall, F1-Score, Matthews Correlation Coefficient (MCC), Geometric mean (G-Mean), Balanced Accuracy (BA) and the receiver operating characteristic (ROC) curve. Table 6 provides the detailed descriptions of each metrics.

In addition, ROC curve can be employed, which is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. It plots the true positive rate (Recall) against the false positive rate (1 - Specificity) at various threshold settings. The area under the ROC curve (AU-ROC) provides a single measure of overall model performance across all classification thresholds.

III. RESULTS AND DISCUSSION

For developing the turbulence model due to wind shear based on TabNet, the challenge of imbalanced data within the HKIA-based PIREPs dataset was first addressed with an initial phase of data cleansing to correct any inconsistencies or missing values. After this preprocessing step, the dataset was divided into training and testing subsets. A randomized method was used, allocating 70% of the data for training and reserving the remaining 30% for testing. To tackle the issue of dataset imbalance, several resampling techniques were applied to the training set, including Borderline-SMOTE, SVM-SMOTE, Near Miss, ADASYN, and SMOTE-ENN. These techniques created a balanced class distribution, enhancing the TabNet model's ability to effectively learn from less represented classes.

Following data balancing, Bayesian optimization was applied to both the treated and untreated segments of the training dataset to fine-tune the TabNet model's hyperparameters through a systematic and iterative approach. This optimization used a 10-fold cross-validation strategy, ensuring robustness and generalizability in TabNet model training. Afterwards, the 30% portion of the dataset set aside for testing was used to assess the TabNet model's predictive accuracy, allowing for a comprehensive evaluation of its performance across various metrics. To further clarify the TabNet model's decision-making process, SHAP analysis was conducted. This interpretative technique provided both global and local perspectives on the TabNet's predictions, revealing the influence of individual features on the model's output. Through SHAP analysis, insightful and interpretable explanations were derived, enhancing the transparency of the model's decision-making processes.

From January 1, 2007, to July 2023, HKIA-based PIREPs reported 6,838 instances of wind shear on both outbound and inbound flights. Among these, 1,169 instances of ST due to wind shear were detected, and 5,668 instances of IST were recorded. This compilation aims to show the complex

TABLE 8. Best hyperparameters for the TabNet model across various data augmentation strategies.

Hyperparameters of TabNet	Data Augmentation Strategies					
	Untreated Data	SVM-SMOTE	Borderline-SMOTE	Near Miss	ADASYN	SMOTE-ENN
optimizer	Adam	Adam	Adam	Adam	Adam	Adam
n_a	52	50	58	62	60	55
lambda_sparse	4.70E-4	4.30E-5	4.60E-4	3.80E-3	4.50E-4	4.10E-5
gamma	1.20	1.80	1.80	1.40	1.20	1.70
n_shared	2	3	2	3	3	2
n_steps	3	3	3	4	4	3
mask_type	sparsemax	sparsemax	sparsemax	sparsemax	sparsemax	sparsemax
learning_rate	0.064	0.034	0.156	0.085	0.141	0.116

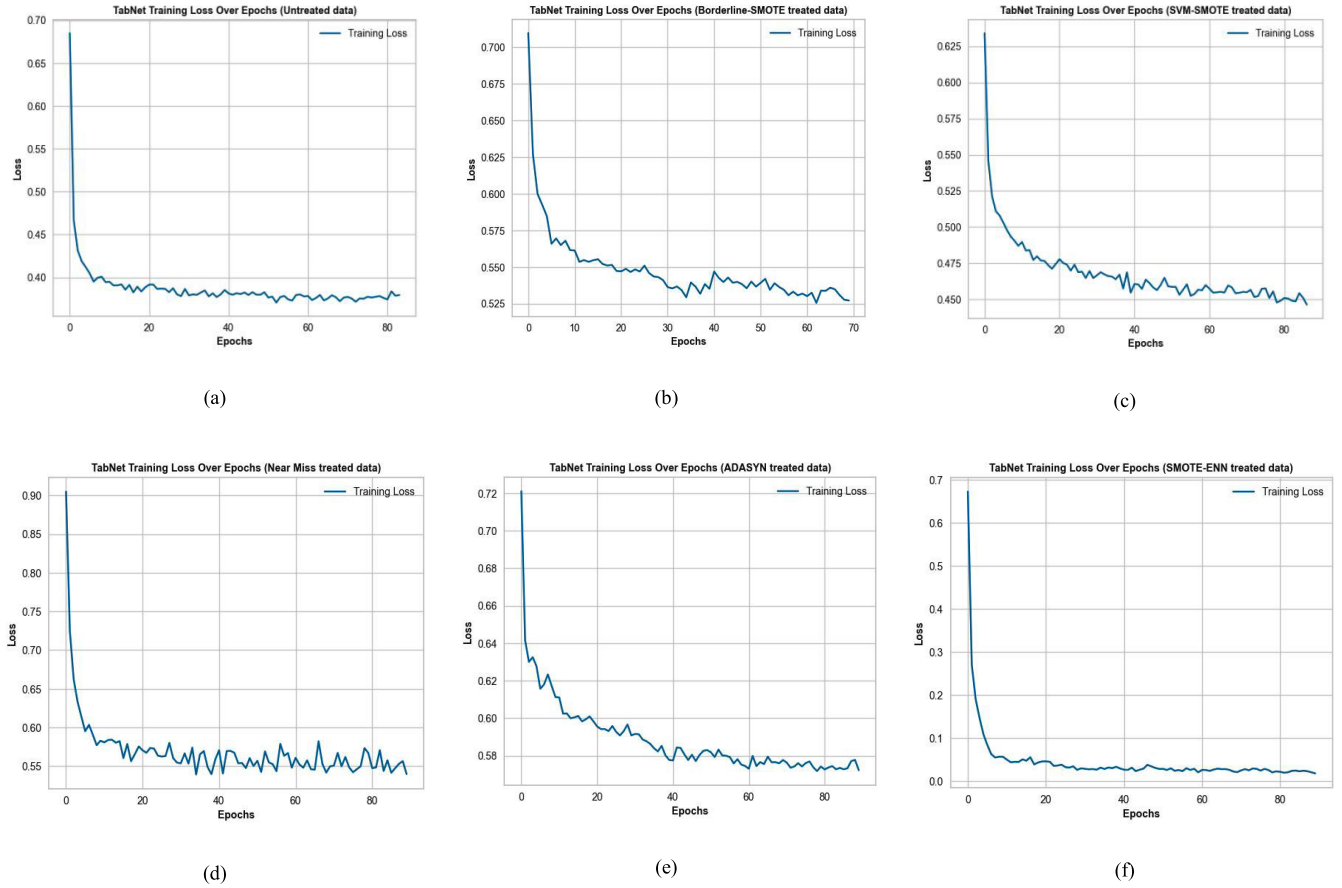


FIGURE 8. Training loss trends across epoch sizes for the TabNet Model with Different Data Augmentation Strategies: (a) TabNet using original untreated data; (b) TabNet with Borderline-SMOTE treatment; (c) TabNet with SVM-SMOTE treatment; (d) TabNet with Near Miss treatment; (e) TabNet with ADASYN treatment; (f) TabNet with SMOTE-ENN treatment.

nature of wind shear and the resulting turbulence, providing a comprehensive perspective on the frequency, distribution, and severity of these events over the 16-year observation period. The detailed analysis serves as a cornerstone for advancing our understanding of turbulence, particularly in the context of wind shear, thereby contributing to the broader body of knowledge within aviation safety research. Table 7 provides a detailed overview of descriptive statistics for factors derived from HKIA-based PIREPs.

A. TREATMENT OF IMBALANCED HKIA-BASED PIREPs DATA

Figure 7 presents a series of pie charts comparing the class distributions within the HKIA-based PIREPs dataset before and after applying various data augmentation techniques, with classes labeled as IST and ST. The original, untreated dataset reveals a significant imbalance, with IST comprising 82.94% (4536 instances) and ST at 17.06% (933 instances). After applying SVM-SMOTE, the class distribution becomes

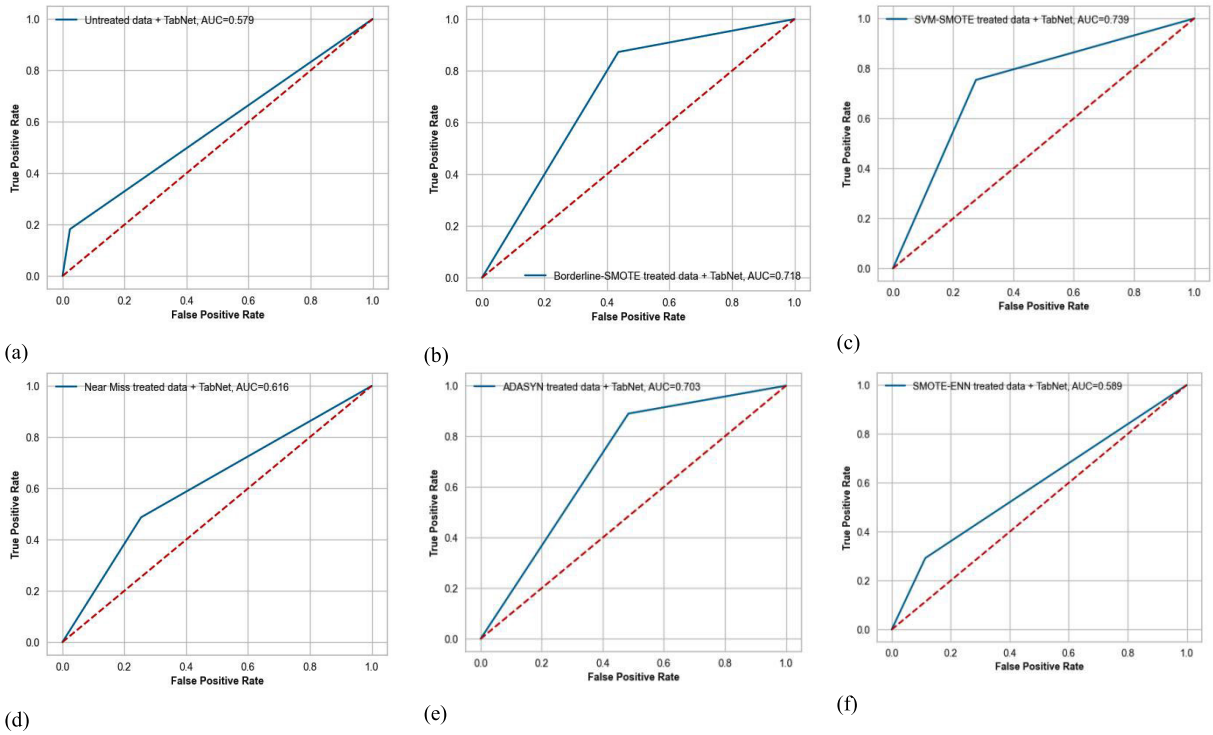


FIGURE 9. ROC curves and AU-ROC values for TabNet models under various data augmentation strategies: (a) TabNet using original, untreated data; (b) TabNet with Borderline-SMOTE modification; (c) TabNet with SVM-SMOTE modification; (d) TabNet with Near Miss modification; (e) TabNet with ADASYN modification; (f) TabNet with SMOTE-ENN modification.

TABLE 9. Comparative analysis of confusion matrix results for TabNet versus other models using various data augmentation strategies.

Data Treatment Strategy	Models	True Negative (TN)	False Positive (FP)	False Negative (FN)	True Positive (TP)
Untreated Data	TabNet	1105	27	193	43
	W&DN	1154	37	141	36
	XGBoost	1138	33	158	39
SVM-SMOTE	TabNet	820	312	58	178
	W&DN	945	200	99	124
	XGBoost	901	256	58	153
Borderline-SMOTE	TabNet	638	494	30	206
	W&DN	890	255	85	138
	XGBoost	860	310	36	162
Near Miss	TabNet	844	288	121	115
	W&DN	837	308	105	118
	XGBoost	830	299	123	116
ADASYN	TabNet	584	548	26	210
	W&DN	895	250	84	139
	XGBoost	634	437	67	185
SMOTE-ENN	TabNet	1003	129	167	69
	W&D	1105	40	176	47
	XGBoost	1045	59	206	58

perfectly balanced, with each class representing 50% of the data, totaling 4536 instances each. Similarly, Borderline-SMOTE achieves an exact 50-50 balance between IST and ST, with 4536 instances in each class.

TABLE 10. Evaluation of TabNet performance across various data augmentation techniques.

TabNet model with with data augmentation strategies	Class	Precision	Recall	F1-Score	BA	MCC	G-Mean
TabNet model using original untreated data	IST	0.85	0.98	0.91	0.57	0.27	0.42
	ST	0.61	0.18	0.28			
	Average	0.81	0.84	0.80			
TabNet model using Borderline-SMOTE augmented data	IST	0.96	0.56	0.71	0.71	0.33	0.70
	ST	0.29	0.87	0.44			
	Average	0.84	0.62	0.66			
TabNet model using SVM-SMOTE augmented data	IST	0.93	0.72	0.82	0.74	0.37	0.74
	ST	0.36	0.75	0.49			
	Average	0.84	0.73	0.76			
TabNet model using Near Miss augmented data	IST	0.87	0.75	0.80	0.61	0.19	0.60
	ST	0.29	0.49	0.36			
	Average	0.77	0.70	0.73			
TabNet model using ADASYN augmented data	IST	0.96	0.52	0.67	0.70	0.30	0.68
	ST	0.28	0.89	0.42			
	Average	0.84	0.58	0.63			
TabNet model using SMOTE-ENN augmented data	IST	0.97	0.55	0.63	0.65	0.22	0.63
	ST	0.27	0.85	0.40			
	Average	0.62	0.70	0.515			

The Near Miss technique also balances the dataset, resulting in an even split, with both IST and ST constituting 50% of the data, each with 933 instances. The ADASYN technique results in a slight variation from perfect balance, with IST at 49.74% (4536 instances) and ST marginally higher at 50.26% (4584 instances). The hybrid SMOTE-ENN shows a more considerable disparity in class distribution, with IST accounting for a larger portion at 82.83% (3064 instances) compared to ST at 17.17% (635 instances).

B. TUNING TabNet HYPERPARAMETERS USING BAYESIAN OPTIMIZATION

This study demonstrates the application of Bayesian Optimization to fine-tune the hyperparameters of the TabNet model, with the aim of enhancing predictive and classification performance in turbulence detection. By employing various data augmentation strategies to prepare balanced training datasets, the study identifies optimal hyperparameters for TabNet across different treated datasets.

Utilizing Bayesian Optimization with a Gaussian Process, the study adjusts TabNet's hyperparameters within a specific search space, aiming to maximize the AU-ROC curve area. The optimal hyperparameter configurations for TabNet across data augmentation strategies are provided in Table 8, specifying important hyperparameters such as optimizer, n_a, lambda_sparse, gamma, n_shared, n_steps, mask_type, and learning_rate. Additionally, Figure 8 illustrates the relationship between loss and epoch size for the TabNet model under different data augmentation strategies.

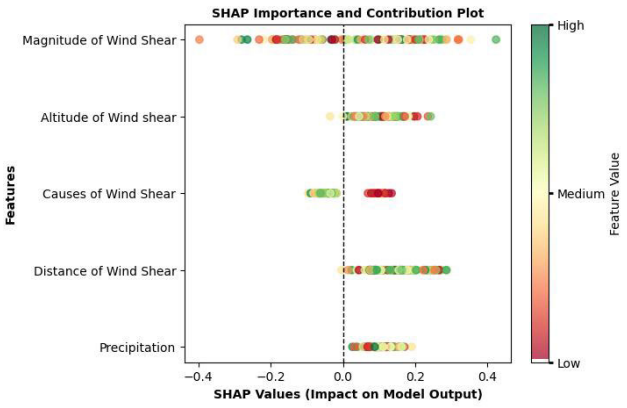


FIGURE 10. SHAP importance and contribution plot for the turbulence due to wind shear.

C. PERFORMANCE PREDICTION AND COMPARISON

The evaluation of predictive performance using the testing dataset begins with collecting classification metrics, such as true positives, true negatives, false positives, and false negatives, from the confusion matrices of TabNet and competing models. This evaluation involves both actual and augmented data, as shown in Table 9. Key evaluation metrics, including the G-Mean, MCC, and BA, were calculated to enable a comprehensive comparison between models. These metrics are crucial for addressing the challenges associated with imbalanced classification tasks, aiming for higher values. The

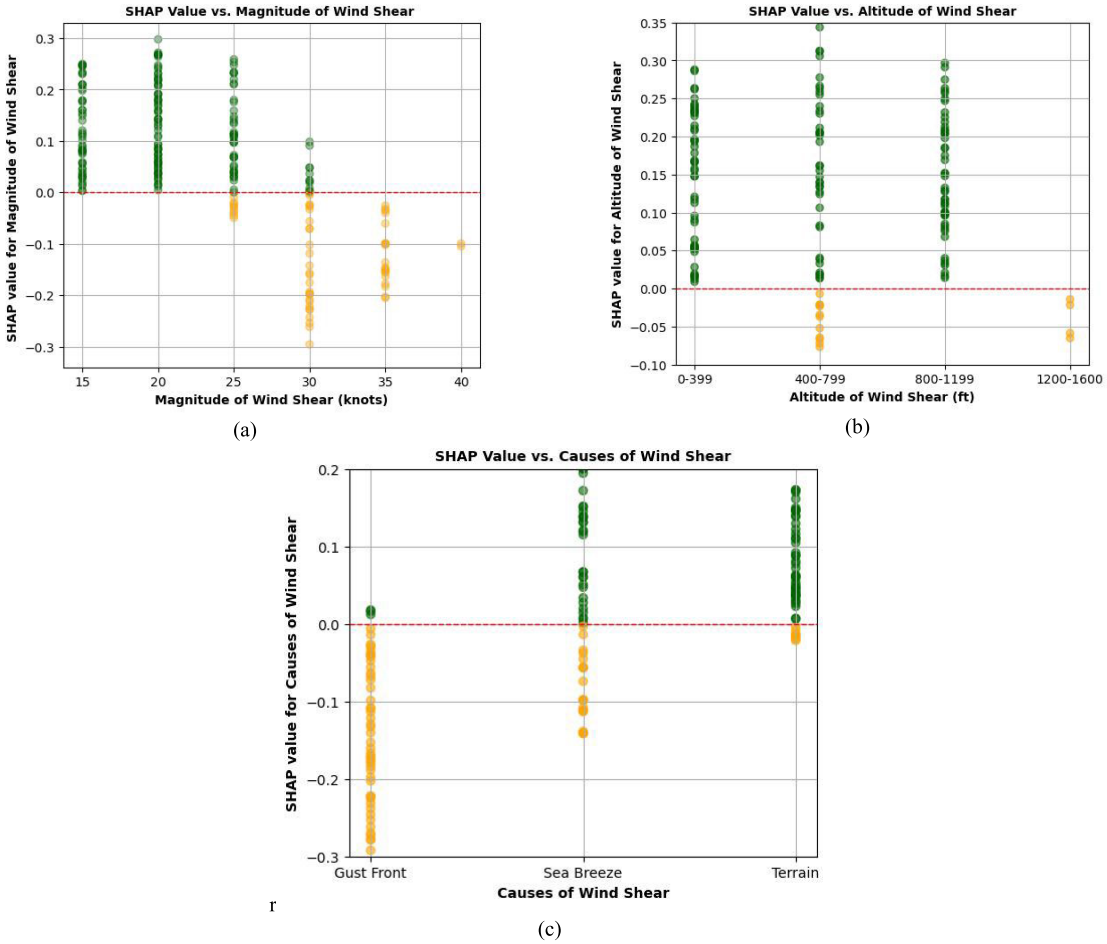


FIGURE 11. Partial SHAP dependence plots (a) SHAP dependence plot for magnitude of wind shear; (b) SHAP dependence plot for altitude of wind shear; (c) SHAP dependence plot for causes of wind shear.

performance metrics are presented in Table 10, with the ROC curves displayed in Figure 9.

When the TabNet model was applied to untreated PIREPs data from HKIA, it achieved a G-Mean of 0.57, an MCC of 0.27, a BA of 0.42, and an AU-ROC of 0.579. An examination of various data augmentation strategies revealed that using the TabNet model with SVM-SMOTE-treated data significantly improved performance outcomes, achieving the highest recorded G-Mean of 0.74, an MCC of 0.37, a BA of 0.74, and an AU-ROC of 0.739, thus outperforming other data augmentation strategies.

D. INTERPRETATION BY SHAP ANALYSIS

The utilization of SHAP analysis for post-hoc interpretation plays a crucial role. The SHAP importance and contribution, often depicted through a SHAP dot plot, offers an insightful visual exploration of the relationships and influence of various factors on model predictions. This plot aggregates individual data points vertically to display the distribution of events, where each point signifies a specific incident. The utilized color gradient, ranging from red to green, denotes the intensity of each factor's value i.e., green for higher values and red for lower values. The SHAP values, shown along the horizontal axis, indicate the influence of each factor: positive values represent a factor's positive impact on predicting outcomes aligned with the optimal model and an increased probability of an event occurring, while negative values denote a negative impact and a reduced probability of occurrence. Consequently, in the scope of this study pertaining to aviation safety, the SHAP contribution plot emerges as an essential tool, elucidating both the significance and the directional influence of different factors on the predictive accuracy of the TabNet model.

Figure 10 illustrates that the optimal TabNet model, enhanced by SVM-SMOTE data treatment and SHAP analysis, identified three primary factors influencing the likelihood of ST: the magnitude of wind shear, the altitude at which wind shear occurs, and the causes of wind shear. The plot also revealed that when these factors exhibit low to medium values, their impact on the probability of encountering ST is significant.

Through the assessment of partial SHAP dependence plots, insights into how specific values of key factors such as wind shear magnitude, altitude, and causes influence the likelihood

of ST can be gained. These plots, shown in Figure 11, illustrate the relationship between variations in these factors and their SHAP values, which quantify each factor's contribution to predicting turbulence.

Figure 11a shows that ST is most commonly associated with wind shear magnitudes between 15 and 25 knots. Interestingly, wind shear strengths beyond 25 knots do not significantly increase the likelihood of encountering ST. Figure 11b demonstrates that ST is predominantly encountered when wind shear occurs at altitudes below 1200 feet. This finding indicates that the lower atmospheric layer, often characterized by more variable and unstable airflow closer to the ground and various obstacles, is a critical zone for the development of ST. This altitude range is crucial during takeoff and landing phases, making understanding wind shear effects here essential for aviation safety.

Figure 11c highlights that wind shear caused by complex topography near airports and sea breezes has a more pronounced effect on ST occurrence. Complex topography can lead to erratic air movements as winds navigate around and over obstacles, while sea breezes, resulting from temperature differences between land and sea, create conditions conducive to turbulence. In contrast, gust fronts, although potentially disruptive, do not significantly contribute to severe turbulence according to the analysis. This might be because gust fronts, capable of producing sharp changes in wind speed and direction, may not always occur under conditions or in locations that typically lead to severe turbulence, or their effects are more predictable and hence better managed.

IV. CONCLUSION AND RECOMMENDATION

This research has demonstrated the effectiveness of the novel TabNet-SHAP approach in predicting and classifying turbulence caused by low-level wind shear in the runway zone at HKIA, using data from PIREPs. By addressing the challenge of imbalanced data, the study has significantly enhanced the predictive accuracy of the TabNet model, making notable advancements in aviation safety research. The application of Bayesian optimization and SHAP analysis has not only optimized the TabNet model's performance but also provided profound insights into the influence of key factors. Based on these findings, the following conclusions can be drawn:

1. The TabNet model based on untreated PIREPs data exhibited lower performance, with a G-Mean of 0.57, MCC of 0.27, BA of 0.42, and AU-ROC of 0.579.
2. The TabNet model using SVM-SMOTE-treated data significantly outperformed other models, with a G-Mean of 0.74, MCC of 0.37, BA of 0.74, and AU-ROC of 0.739.
3. The second-best model performance was achieved by the TabNet model using Borderline-SMOTE-treated data, with a G-Mean value of 0.70, an MCC value of 0.33, and a BA value of 0.71.
4. SHAP analysis provided post-hoc interpretation, offering insights into how various factors influence model predictions. It utilized SHAP dot plots for a visual exploration of the relationships between factors and their impact on predictions.
5. Outcomes of SHAP Analysis identified magnitude of wind shear, altitude of wind shear, and causes of wind shear as primary factors influencing the likelihood of ST. It has been demonstrated that low to medium values of these factors have a significant impact on ST probability, highlighting specific insights such as ST association with magnitude of wind shears between 15 and 25 knots, predominance of ST at altitudes below 1200 feet, and the pronounced effect of complex topography and sea breezes.

To enhance the scope and impact of this study on global aviation safety, future work could explore a number of of directions. Incorporating datasets from diverse geographical locations and a broader range of environmental factors would improve the model's predictive accuracy and applicability. Additionally, the insights gained could inform the development of updated safety protocols and training programs, and influence public policy and aviation regulations. These efforts would collectively advance the understanding and management of turbulence, leading to safer flight operations worldwide.

ACKNOWLEDGMENT

The authors would like to express their gratitude to their colleagues at Hong Kong Observatory (HKO) of Hong Kong International Airport for provision PIREPs and their guidance in analysis.

REFERENCES

[1] A. O'Connor and D. Kearney, "Evaluate the effect of turbulence on aircraft during landing and take-off phases," *Int. J. Aviation, Aeronaut., Aerosp.*, vol. 5, no. 4, p. 10, 2018.

[2] D. Augustin and M. B. Cesar, "Analysis of wind shear variability and its effects on the flights activities at the Garoua airport," *J. Extreme Events*, vol. 9, no. 1, Mar. 2022, Art. no. 2250006.

[3] A. Rafe and P. A. Singleton, "Exploring factors affecting pedestrian crash severity using TabNet: A deep learning approach," 2023, *arXiv:2312.00066*.

[4] R. D. Sharman and J. M. Pearson, "Prediction of energy dissipation rates for aviation turbulence. Part I: Forecasting nonconvective turbulence," *J. Appl. Meteorol. Climatol.*, vol. 56, no. 2, pp. 317–337, Feb. 2017.

[5] S. Chen, H. Kopald, B. Avjian, and M. Fronzak, "Automatic pilot report extraction from radio communications," in *Proc. IEEE/AIAA 41st Digit. Avionics Syst. Conf. (DASC)*, Sep. 2022, pp. 1–8.

[6] A. Khattak, P.-W. Chan, F. Chen, and H. Peng, "Prediction and interpretation of low-level wind shear criticality based on its altitude above runway level: Application of Bayesian optimization–ensemble learning classifiers and Shapley additive exPlanations," *Atmosphere*, vol. 13, no. 12, p. 2102, Dec. 2022.

[7] P. D. Williams, "Increased light, moderate, and severe clear-air turbulence in response to climate change," *Adv. Atmos. Sci.*, vol. 34, no. 5, pp. 576–586, May 2017.

[8] T. P. Lane, R. D. Sharman, S. B. Trier, R. G. Fovell, and J. K. Williams, "Recent advances in the understanding of near-cloud turbulence," *Bull. Amer. Meteorological Soc.*, vol. 93, no. 4, pp. 499–515, Apr. 2012.

[9] F. Barbaresco, L. Thobois, and A. Dolfi-Bouteyre, "Monitoring wind, turbulence and aircraft wake vortices by high resolution RADAR and LiDAR remote sensors in all weather conditions," in *Proc. URSI Sci. Days*, 2015, pp. 81–110.

[10] Z. Liu, J. F. Barlow, P.-W. Chan, J. C. H. Fung, Y. Li, C. Ren, H. W. L. Mak, and E. Ng, "A review of progress and applications of pulsed Doppler wind LiDARs," *Remote Sens.*, vol. 11, no. 21, p. 2522, Oct. 2019.

[11] J. K. Williams and G. Meymaris, "Remote turbulence detection using ground-based Doppler weather radar," in *Aviation Turbulence: Processes, Detection, Prediction*. Cham, Switzerland: Springer, 2016, pp. 149–177.
[12] J. K. Williams, "Using random forests to diagnose aviation turbulence," *Mach. Learn.*, vol. 95, no. 1, pp. 51–70, Apr. 2014.
[13] J. Cai, "Advanced aviation weather radar data processing and real-time implementations," Tech. Rep., 2017.
[14] W. K. Zhao, S. J. Zhao, Y. L. Shan, and X. J. Sun, "Numerical simulation for wind shear detection with a glide path scanning algorithm based on wind LiDAR," *IEEE Sensors J.*, vol. 21, no. 18, pp. 20248–20257, Sep. 2021.
[15] W. Lyu and J. Liu, "Artificial intelligence and emerging digital technologies in the energy sector," *Appl. Energy*, vol. 303, Dec. 2021, Art. no. 117615.
[16] Z. Jan, F. Ahamed, W. Mayer, N. Patel, G. Grossmann, M. Stumptner, and A. Kuusk, "Artificial intelligence for Industry 4.0: Systematic review of applications, challenges, and opportunities," *Expert Syst. Appl.*, vol. 216, Apr. 2023, Art. no. 119456.
[17] S. O. Abioye, L. O. Oyedele, L. Akanbi, A. Ajayi, J. M. D. Delgado, M. Bilal, O. O. Akinade, and A. Ahmed, "Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges," *J. Building Eng.*, vol. 44, Dec. 2021, Art. no. 103299.
[18] Z. Ahmed, K. Mohamed, S. Zeeshan, and X. Dong, "Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine," *Database*, vol. 2020, Jan. 2020, Art. no. baaa010.
[19] M. Elahi, S. O. Afolaranmi, J. L. M. Lastra, and J. A. P. Garcia, "A comprehensive literature review of the applications of AI techniques through the lifecycle of industrial equipment," *Discover Artif. Intell.*, vol. 3, no. 1, p. 43, Dec. 2023.
[20] A. Khattak, P.-W. Chan, F. Chen, and H. Peng, "Time-series prediction of intense wind shear using machine learning algorithms: A case study of Hong Kong international airport," *Atmosphere*, vol. 14, no. 2, p. 268, Jan. 2023.
[21] A. McGovern, K. L. Elmore, D. J. Gagne, S. E. Haupt, C. D. Karstens, R. Lagerquist, T. Smith, and J. K. Williams, "Using artificial intelligence to improve real-time decision-making for high-impact weather," *Bull. Amer. Meteorological Soc.*, vol. 98, no. 10, pp. 2073–2090, Oct. 2017.
[22] S. Mizuno, H. Ohba, and K. Ito, "Machine learning-based turbulence-risk prediction method for the safe operation of aircrafts," *J. Big Data*, vol. 9, no. 1, p. 29, Dec. 2022.
[23] A. Shankar and B. C. Sahana, "Efficient prediction of runway visual range by using a hybrid CNN-LSTM network architecture for aviation services," *Theor. Appl. Climatol.*, vol. 155, no. 3, pp. 2215–2232, Mar. 2024.
[24] A. Khattak, P. Chan, F. Chen, and H. Peng, "Interpretable ensemble imbalance learning strategies for the risk assessment of severe-low-level wind shear based on LiDAR and PIREPs," *Risk Anal.*, vol. 44, no. 5, pp. 1084–1102, May 2024.
[25] Z. Zhuang, H. Zhang, P.-W. Chan, H. Tai, and Z. Deng, "A machine learning-based model for flight turbulence identification using LiDAR data," *Atmosphere*, vol. 14, no. 5, p. 797, Apr. 2023.
[26] X. Zhang and S. Mahadevan, "Ensemble machine learning models for aviation incident risk prediction," *Decis. Support Syst.*, vol. 116, pp. 48–63, Jan. 2019.
[27] X. Zhang and S. Mahadevan, "Bayesian neural networks for flight trajectory prediction and safety assessment," *Decis. Support Syst.*, vol. 131, Apr. 2020, Art. no. 113246.
[28] G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou, and D. Zhao, "Flight delay prediction based on aviation big data and machine learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 140–150, Jan. 2020.
[29] E. Mangortey, T. G. Puranik, O. J. Pinon-Fischer, and D. N. Mavris, "Classification, analysis, and prediction of the daily operations of airports using machine learning," in *Proc. AIAA Scitech Forum*, Jan. 2020, p. 1196.
[30] T. G. Puranik, N. Rodriguez, and D. N. Mavris, "Towards online prediction of safety-critical landing metrics in aviation using supervised machine learning," *Transp. Res. C, Emerg. Technol.*, vol. 120, Nov. 2020, Art. no. 102819.
[31] E. Esmaeilzadeh and S. Mokhtarimousavi, "Machine learning approach for flight departure delay prediction and analysis," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2674, no. 8, pp. 145–159, Aug. 2020.
[32] O. Basturk and C. Cetek, "Prediction of aircraft estimated time of arrival using machine learning methods," *Aeronaut. J.*, vol. 125, no. 1289, pp. 1245–1259, Jul. 2021.
[33] H. Han, W. Y. Wang, and B. H. Mao, "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," in *Proc. Int. Conf. Intell. Comput.* Cham, Switzerland: Springer, Aug. 2005, pp. 878–887.
[34] H. M. Nguyen, E. W. Cooper, and K. Kamei, "Borderline over-sampling for imbalanced data classification," *Int. J. Knowl. Eng. Soft Data Paradigms*, vol. 3, no. 1, p. 4, 2011.
[35] I. Mani and I. Zhang, "kNN approach to unbalanced data distributions: A case study involving information extraction," in *Proc. workshop Learn. Imbalanced Datasets*, 2003, no. 1, pp. 1–7.
[36] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 1322–1328.
[37] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explorations Newslett.*, vol. 6, no. 1, pp. 20–29, Jun. 2004.
[38] S. Ö. Arik and T. Pfister, "TabNet: Attentive interpretable tabular learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 2021, no. 8, pp. 6679–6687.
[39] J. Yan, T. Xu, Y. Yu, and H. Xu, "Rainfall forecast model based on the TabNet model," *Water*, vol. 13, no. 9, p. 1272, Apr. 2021.
[40] E. Borghini and C. Giannetti, "Short term load forecasting using TabNet: A comparative study with traditional State-of-the-Art regression models," *Eng. Proc.*, vol. 5, no. 1, p. 6, 2021.
[41] Y. Chen, H. Li, H. Dou, H. Wen, and Y. Dong, "Prediction and visual analysis of food safety risk based on TabNet-GRA," *Foods*, vol. 12, no. 16, p. 3113, Aug. 2023.
[42] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S.-H. Deng, "Hyperparameter optimization for machine learning models based on Bayesian optimization," *J. Electron. Sci. Technol.*, vol. 17, pp. 26–40, Mar. 2019.
[43] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
[44] S. W. Chiu and K. Y. Siu, "Hong Kong as an international hub: The rise of Hong Kong in the modern world-system," in *Hong Kong Society: High-Definition Stories Beyond the Spectacle of East-Meets-West*. Singapore: Springer, 2022, pp. 39–70.
[45] P. Asen, V. Bobek, and T. Horvat, "The efficiency of ports and airports in emerging markets measured by key performance indicators: The case of Hong Kong and Durban," *Management*, vol. 16, no. 2, pp. 75–81, 2021.
[46] Y. Zheng, W. Li, C. Fang, B. Feng, Q. Zhong, and D. Zhang, "Investigating the impact of weather conditions on urban heat island development in the subtropical city of Hong Kong," *Atmosphere*, vol. 14, no. 2, p. 257, Jan. 2023.
[47] K. Hon and P. W. Chan, "A decade (2011–2020) of tropical cyclone reconnaissance flights over the South China Sea," *Weather*, vol. 77, no. 9, pp. 308–314, Sep. 2022.
[48] G. Lui, R. Liem, and K. Hon, "Towards understanding the impact of convective weather on aircraft arrival traffic at the Hong Kong international airport," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 569, no. 1, Sep. 2020, Art. no. 012067.
[49] A. Khattak, P.-W. Chan, F. Chen, H. Peng, and C. M. Matara, "Missed approach, a safety-critical go-around procedure in aviation: Prediction based on machine learning-ensemble imbalance learning," *Adv. Meteorol.*, vol. 2023, pp. 1–24, Jul. 2023.
[50] K.-K. Hon, "Predicting low-level wind shear using 200-m-resolution NWP at the Hong Kong international airport," *J. Appl. Meteorol. Climatol.*, vol. 59, no. 2, pp. 193–206, Feb. 2020.
[51] P. W. Chan and Q. S. Li, "Some observations of low level wind shear at the Hong Kong international airport in association with tropical cyclones," *Meteorological Appl.*, vol. 27, no. 2, p. e1898, Mar. 2020.
[52] P. W. Chan, "Case study of a special event of low-level windshear and turbulence at the Hong Kong international airport," *Atmos. Sci. Lett.*, vol. 24, no. 4, p. e1143, Apr. 2023.
[53] D.-C. Feng, W.-J. Wang, S. Mangalathu, and E. Taciroglu, "Interpretable XGBoost-SHAP machine-learning model for shear strength prediction of squat RC walls," *J. Struct. Eng.*, vol. 147, no. 11, Nov. 2021, Art. no. 04021173.
[54] K. Aas, M. Jullum, and A. Løland, "Explaining individual predictions when features are dependent: More accurate approximations to Shapley values," *Artif. Intell.*, vol. 298, Sep. 2021, Art. no. 103502.

AFAQ KHATTAK received the B.S. degree in civil engineering, the M.S. degree in transportation engineering, and the Ph.D. degree in traffic engineering. Currently, he is a Postdoctoral Researcher with Tongji University, Shanghai, China. Over the years, he has contributed significantly to the field through several SCI and EI-indexed publications in the domain of transportation engineering. His research interest includes the application of artificial intelligence in transportation systems.

JIANPING ZHANG received the Ph.D. degree. He is currently a prominent figure with the Civil Aviation Administration of China. He specializes in civil aviation air traffic management and human factor engineering. His work involves enhancing the efficiency and safety of air traffic systems through the integration of advanced technologies, such as automation and artificial intelligence. Additionally, he focuses on human factor engineering, aiming to improve the ergonomic design of aviation systems to optimize performance and safety.

PAK-WAI CHAN received the Ph.D. degree from the City University of Hong Kong, Hong Kong, in 2023, with a study on wind engineering and low level wind shear for aviation applications. He is currently the Director of Hong Kong Observatory. He is a Visiting Professor of a number of universities in mainland China and an Adjunct Associate Professor with The University of Hong Kong. He had been with the Hong Kong International Airport for more than 20 years, with research and operational efforts in airport meteorological instrumentation, low level wind shear and turbulence alerting, and high resolution numerical weather prediction.

FENG CHEN received the Ph.D. degree. He was a Postdoctoral Fellow with the Colorado State University. From 2007 to 2012, he participated in eight USA projects on traffic safety, vehicle dynamic model, and reliability analysis. Since 2012, he has been the PI of 23 research grants funded by various national and provincial agencies, including NSFC and Hong Kong Zhuhai Macao Bridge project, in the field of traffic safety and driving simulator, and committed to studies in the areas of traffic safety, intelligent transport system, and human factor.

ABDULRAZAK H. ALMALIKI received the B.Sc. degree (Hons.) in abbreviation from Civil Engineering, College of Engineering, Umm Alqura University, in 2003, and the M.Sc. degree in construction engineering and management and the Ph.D. degree in water resources engineering from King Abdulaziz University, in 2007 and 2015, respectively. His major field is in water resources and environmental engineering. He has been the Head of the Water Resources and Environmental Engineering Group, Taif University, since 2017. He is currently an Associate Professor of civil engineering with the College of Engineering, Taif University.

CAROLINE MONGINA MATARA received the M.S. degree in transportation from the School of Transportation and Logistics, Southwest Jiaotong University. She is currently with the Department of Civil and Construction Engineering, University of Nairobi, Kenya. She is doing research in transportation engineering, safety engineering, and civil engineering.