

Article

Prediction of a Pilot's Invisible Foe: The Severe Low-Level Wind Shear

Afaq Khattak ^{1,*}, Pak-Wai Chan ², Feng Chen ^{1,*} and Haorong Peng ³ 

¹ The Key Laboratory of Infrastructure Durability and Operation Safety in Airfield of CAAC, Tongji University, 4800 Cao'an Road, Jiading, Shanghai 201804, China

² Hong Kong Observatory, 134A Nathan Road, Kowloon, Hong Kong, China

³ Shanghai Research Center for Smart Mobility and Road Safety, Shanghai 200092, China

* Correspondence: khattak@tongji.edu.cn (A.K.); fengcheng@tongji.edu.cn (F.C.)

Abstract: Severe low-level wind shear (S-LLWS) in the vicinity of airport runways (25 knots or more) is a growing concern for the safety of civil aviation. By comprehending the causes of S-LLWS events, aviation safety can be enhanced. S-LLWS is a rare occurrence, but it is hazardous for approaching and departing aircraft. This study introduced the self-paced ensemble (SPE) framework and Shapley additive explanations (SHAP) interpretation system for the classification, prediction, and interpretation of LLWS severity. Doppler LiDAR- and PIREPs-based LLWS data from Hong Kong International Airport were obtained, trained, and evaluated to predict LLWS severity. The SPE framework was also compared to state-of-the-art tree-based models, including light gradient boosting machine, adaptive boosting, and classification and regression tree models. The SPE does not require prior data treatment; however, SMOTE-ENN was utilized to treat highly imbalanced LLWS training data for tree-based models. In terms of prediction performance, the SPE framework outperforms all tree-based models. Using SHAP analysis, the SPE was interpreted. It was determined that “runway 25LD”, “mean hourly temperature”, and “mean wind speed” were the most significant contributors to the occurrence of S-LLWS. The most optimistic projections for the occurrence of S-LLWS events at runway 25LD were during periods of low-to-moderate temperatures and relatively medium-to-high wind speeds. Similarly, the majority of S-LLWS events took place on the runway. Without the need for data augmentation during preprocessing, the SPE framework coupled with the SHAP interpretation system could be utilized effectively for the prediction and interpretation of LLWS severity. This study is an invaluable resource for aviation policymakers and air traffic safety analysts.

Keywords: civil aviation safety; low-level wind shear; pilot reports; machine learning; self-paced ensemble; Shapley additive explanations



Citation: Khattak, A.; Chan, P.-W.; Chen, F.; Peng, H. Prediction of a Pilot's Invisible Foe: The Severe Low-Level Wind Shear. *Atmosphere* **2023**, *14*, 37. <https://doi.org/10.3390/atmos14010037>

Academic Editors: Duanyang Liu, Hongbin Wang and Shoupeng Zhu

Received: 4 December 2022

Revised: 21 December 2022

Accepted: 21 December 2022

Published: 25 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Airline operations are profoundly impacted by weather conditions. Major causes of flight cancellations, delays, and even fatal crashes [1–3] can be traced back to this concern. Wind shear refers to an abrupt shift in the wind's speed or direction in the atmosphere. Particularly during landing and takeoff, aircraft are impacted by low-level wind shear (LLWS), which is present in a lower layer at 1600 feet above ground level (AGL). LLWS is defined by the International Civil Aviation Organization (ICAO) [4] as a 15-knot-or-greater change in wind direction at or below 1600 feet above ground level. It affects the aircraft's lift, and the resulting course deviation could endanger planes taking off or landing [5,6].

Many LLWS events with a magnitude of 25 knots or higher have been registered at airports around the world. Because S-LLWS may have a stronger impact on aircraft operations, timely warnings are crucial. Hong Kong International Airport is one of the most at-risk airports for LLWS (HKIA). It is located in Lantau Island's northern region, which is mountainous, with peaks reaching over 900 m and valleys dropping to 300 m. Lowering

the adverse effects of S-LLWS on airport safety and productivity is vital. A reliable LLWS severity prediction approach is crucial for achieving the goal of providing precise and effective wind hazard alerts and ensuring the safety of civil aviation. The development of models for predicting the severity of LLWS closer to airport runways, however, remains among the most challenging areas of research in civil aviation today.

Due to the fact that wind shear exhibits characteristics of meso- and micro-scale meteorological phenomena, such as abrupt changes in speed and direction and a small temporal–spatial scale, predicting wind shear is a difficult endeavor. LLWS events occur in both rainy and non-rainy weather and include phenomena such as frontal gusts and microbursts associated with severe convection, dry microbursts, low-level jets, sea breezes, complex terrain effects, etc [7]. To ensure the safety of civil aircraft, various technologies, including anemometers, terminal Doppler weather radar (TDWR), and Doppler light detection and range (DLDR), have been installed at major airports around the world to detect LLWS (LIDAR). Few airports, including those in Japan, Malaysia, the United States, Germany, France, South Korea, Singapore, and Hong Kong, have LLWS alerting technologies due to high instrument and maintenance costs, a lack of relevant research, and unique local environments [8]. The anemometer-based LLWS alert system and the TDWR have been developed since the 1970s. Their effectiveness for detecting and warning of LLWS in rainy conditions has been demonstrated [9]. The complementary TDWR is also capable of detecting LLWS caused by terrain. However, these technologies are incapable of capturing LLWS events in non-rainy weather [10,11] and are unsuitable for detecting LLWS along the glide path.

Doppler LIDAR [8], a relatively new remote sensing technology, offers a promising alternative for detecting LLWS when the weather is clear. Similarly, certain LLWS events are terrain-induced LLWS phenomena caused by the complex terrain surrounding an airport. Doppler LIDAR technology, which does not depend on humid conditions for detecting LLWS and captures LLWS due to complex terrain near airports, has been developed to address these scenarios. Hong Kong International Airport [12], Nice Côte d’Azur Airport in France [13], Tokyo Haneda International Airport in Japan [14] and Beijing Capital International Airport in China [15] are equipped with the Doppler LIDAR system. It has been added to the TDWR as an augmentation in order to detect and warn of LLWS, even in clear skies. However, the development of a model to predict the severity of LLWS based on Doppler LIDAR observations remains a challenging task that must be addressed. Similarly, all of these LLWS alerting technologies (based on remote sensing and/or on-site measurements) have been proven effective and operational. These technologies send notifications or alerts when LLWS events are detected or observed. However, these hardware-based technologies are incapable of predicting the occurrence of LLWS events and assessing the risk factors that contribute to their occurrence [16].

In the past, numerous numerical modeling techniques, including large-eddy simulations (LES) [17], computational fluid dynamics (CFD) [18] and numerical weather prediction (NWP) [12] have been employed to attempt to predict or simulate wind shear conditions. In general, these studies focused on single or isolated occurrences of reported wind shear events and were conducted on a case-by-case basis. There are insufficient systematic, long-term evaluations of the ability of numerical models to predict the occurrence of LLWS events. These days, machine learning algorithms have gained significant ground. It has become one of the most widely used and beneficial tools in transportation research such as road safety, transportation planning, and pavement analysis [19–22]. However, there is a significant gap in the application of machine learning algorithm in the aviation safety domain, particularly in the prediction and classification of LLWS severity. In this research, efficiently predicting S-LLWS is of interest to us. However, in the data from LiDAR and pilot reports (PIREPs), the S-LLWS class is typically much smaller than the non-severe low-level wind shear (NS-LLWS) class. This creates a data imbalance issue and requires data balancing prior to training and evaluation. Therefore, in contrast to hardware-based technologies and numerical simulation modeling, which efficiently predict LLWS severity

while simultaneously dealing with the class imbalance issue, we propose the self-paced ensemble (SPE) framework [23]. This is an ensemble imbalance learning model for dealing with highly imbalanced data. It aims to produce a robust ensemble by the self-paced harmonizing of data hardness via the undersampling method that has been developed. This framework, despite being computationally efficient, has resulted in robust performances in the presence of extremely skewed distributions.

Although machine learning models are efficient in prediction, they do not explicitly demonstrate the relations between input and output factors due to their black-box nature. The interpretation of the model is equally important for appropriately assessing the model's performance. Previously, the machine learning model's results were interpreted using the variable importance analysis methods such as permutation-based importance scores. The variable importance analysis, however, can only provide a ranking of the variables' importance and is unable to explain how each variable individually influences the prediction of the model. Shapley additive explanations (SHAP) analysis, based on the concept of game theory [24], has been utilized in recent studies to quantify each factor's effect on the outcome [25,26]. In this research, we have also employed SHAP analysis in conjunction with SPE framework, for the assessment of the relative importance of various factors as well as their contributions.

The rest of this paper is organized as follows: The following sections constitute the research methodology, which provides the data description, the details of the proposed SPE framework, a Bayesian optimization strategy for hyperparameter tuning, and the description of the SHAP interpretation system. These are then followed by Section 3, discussing the SPE framework and comparison with other machine learning models, as well as SHAP analysis. Finally, Section 4 summarizes the conclusions and makes additional research recommendations.

2. Materials and Methods

Initially, the LLWS data consisted of LiDAR data and pilot flight reports (PIREPs) obtained from the Hong Kong Observatory (HKO) at HKIA. The details of data extraction from LiDAR and PIREPs are provided in the subsequent section. The extracted data were merged together and preprocessed to separate training-validation and testing datasets into 70% and 30%, respectively. The training dataset was used to develop an SPE framework and tree-based machine learning models, including light gradient boosting machine (LGBM), adaptive boosting (AdaBoost), and classification and regression tree (CART), and the testing dataset was used to evaluate the performance of the developed model. The SPE framework is an ensemble imbalance learning system, which does not require data balancing during the preprocessing phase. In contrast, data balancing was required for the tree-based machine learning models prior to training and validation, which were used to compare the results with the SPE framework. For data balancing, a hybrid synthetic minority oversampling technique—edited nearest neighbor (SMOTE-ENN) treatment was applied to the LLWS training dataset. A portion of the training-validation data were also used to tune model hyperparameters. A Bayesian optimization approach was utilized for the hyperparameter tuning. Afterwards, the SHAP interpretation system was used to evaluate the significance and contribution of various risk factors that generate S-LLWS in the vicinity of airport runways. In addition, factor interaction analysis by SHAP was also conducted. Figure 1 depicts the entire operational paradigm described in this study.

2.1. Study Location

The Hong Kong International Airport (HKIA) is situated on an artificial island called Lantau, which is surrounded on three sides by water. To the south, there are mountains that rise to more than 900 m above sea level. The complex land-sea contrast and intricate orography of HKIA have been the subject of numerous observational and modeling studies, all of which have identified that they are favorable conditions for the occurrence of LLWS [27,28]. As seen in Figure 2, the mountainous area to the south of HKIA amplifies LLWS, disrupting airflow

and causing mountain waves, gap effluents, and other disruptions along the HKIA flight paths. The north runway and the south runway are the two runways at HKIA. The directions they orient are 070° and 250° . Eight different arrangements are possible because each runway can be used for takeoffs and landings in either direction. For instance, runway ‘07LA’ stands for a landing (‘A’ stands for arrival) using the left runway (hence ‘L’) and a heading angle of 070° . This depicts a plane landing on HKIA’s north runway from the west. The same goes for an aircraft taking off from the south runway and heading west—runway 25LD.

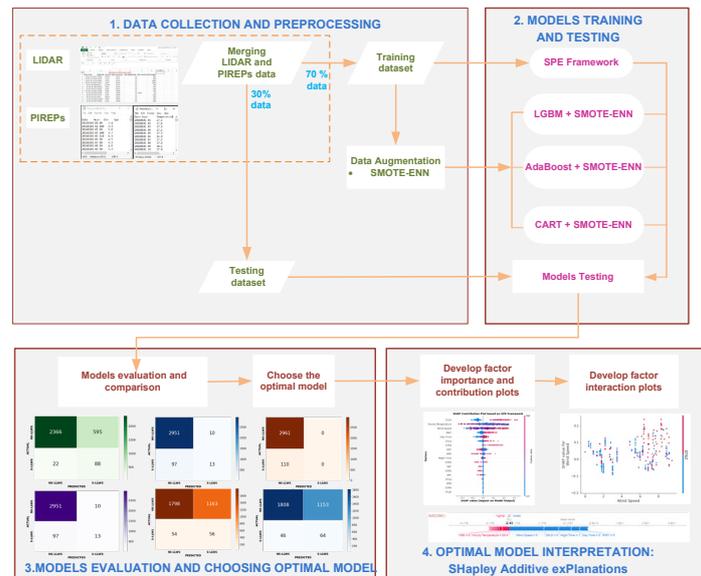


Figure 1. Framework for the prediction and interpretation of LLWS severity in the vicinity of runways.



Figure 2. Hong Kong International airport and surrounding terrain.

2.2. Instrument and Data

In this section, the Doppler LiDAR of HKIA and the pilot flight reports (PIREPs) of HKIA inbound and outbound flights, are thoroughly discussed.

2.2.1. Doppler LiDAR at HKIA

In this study, LLWS data gathered from the $2 \times$ long-range Doppler LiDAR at HKIA were analysed. LiDAR operates at an infrared wavelength of approximately 1.5 microns; 100 m is the radial resolution or physical range gate. Maximum radial velocity is roughly 40 m per second. Typically, under ideal weather conditions and in the absence of obstructions such as low clouds, an observation range of 10 or 15 kilometres is achievable. In addition to the standard fixed-elevation scans (plan-position indicator), each LiDAR is

configured to conduct “glide-path” scans along take-off and landing flight paths. Coordinating the elevation and azimuth movements of the laser scanner head accomplishes this. Typically, the four possible configurations of the north runway (07LA, 25RA, 07LD, and 25RD) are covered by the north LiDAR, including arrivals (A) and departures (D) directions, and the four possible configurations of the south runway (07RA, 25LA, 07RD and 25LD) are covered by the south LiDAR, including arrivals (A) and departures (D) directions towards the west and the east. The headwind component along each runway configuration (labelled “corridor”) can be derived from the “glide-path” scans’ radial velocity data. Typically, the scan revisit time for each corridor is roughly one minute, indicating that the temporal resolution or update frequency of the headwind profiles is also roughly one minute.

The LiDAR at HKIA usually operates by a “GLYGA LLWS alerting algorithm” [7]. For each runway corridor, GLYGA receives as input the profile of headwind components gridded with a 100 m interval. The headwind profiles typically extend up to 4–5 NM from the respective runway endpoint, based on scanning range and prevalent atmospheric conditions at the time. Then, a ramp identification procedure is used to identify sudden, consistent changes in the headwind. This is based on the “Peak Spotter” algorithm [29]. First, a profile of velocity increment is quantified by adjusting adjacent data points from the profile of quality-controlled headwind. Next, LLWS “ramps” are identified by sequentially recognizing the velocity increment (i.e., headwind change) within length windows of 400, 800, 1600, and 6400 m. The collection of such “ramps”, identified within a single headwind profile, is then ranked using a severity factor [30] that scales with the headwind increment and the inverse cube root of the ramp length. The ramp with the highest severity factor is then used to release an automatic alert when intensity exceeds a predetermined threshold (15 knots) at HKIA.

Mathematically, the quality-controlled headwind profile can be represented as $v(x_k)$, where v is the headwind component at the x_k position, which is the k th data point or range gate along the corresponding glide path. The velocity increment at location x_k can be expressed as $\Delta v(x_k, \lambda) = v(x_k) - v(x_k + \lambda)$ for a given length window (or ramp length), λ . (For a detailed explanation of the ramp identification process at HKIA, please see [7].) The resultant identified ramps, which correspond to a collection of data pairs $(\Delta v, \lambda)$, are then ranked by the severity factor \mathcal{F}_s , which is computed using Equation (1).

$$\mathcal{F}_s = \left(\frac{\Delta v}{\lambda} \right)^3 / \Phi_{app} \quad (1)$$

where Φ_{app} denotes the aircrafts’ approach speed, which is taken as constant. The \mathcal{F}_s depends primarily on $\Delta v/\lambda$.

2.2.2. HKIA-Based PIREPs

Pilot flight reports (PIREP) of LLWS are an established source for confirming LLWS alerts at HKIA. A PIREP is an abbreviation for pilot reports used in the aviation sector. It is a report that pilots who encounter hazardous weather conditions send to air traffic controllers. Typical PIREPs cover the flight’s en route phase and include information on turbulence and aircraft icing. However, the HKIA wind shear PIREPs contain information regarding the timing, location (to the nearest nautical mile), altitude (to the nearest 50 or 100 feet), and velocity (to the nearest 5 knots) of an LLWS event. Pilots can report LLWS events to the air traffic controller at HKIA in two standard ways: by submitting a report form after landing or departure, or by using an on-board radio communication.

2.3. Data Processing

As discussed early, the occurrence of S-LLWS is a substantial risk to inbound and outbound flight safety. Therefore, in order to predict the S-LLWS events, in this study, the occurrence of LLWS severity is defined by the threshold, as shown in Equation (2).

$$LLWS \text{ Severity} = \begin{cases} 1 : S - LLWS, & LLWS \geq 25 \text{ knots} \\ 0 : NS - LLWS, & LLWS 15 - 24 \text{ knots} \end{cases} \quad (2)$$

The original wind shear dataset contains nominal and continuous factors as well as a single target factor, LLWS severity. S-LLWS events represent all LLWS with a magnitude of equal to or greater than 25 knots and are coded as 1, whereas NS-LLWS events have a magnitude between 15 to 24 knots and are coded as 0. S-LLWS events are far less in number than NS-LLWS, but they are an important class for aviation safety. Any i th event in the original dataset can be represented as $(X_i, y_i) = (C_i, N_i, y_i)$, where C_i is the continuous factor, N_i is the nominal factor and y_i is the target factor. As indicated in Table 1, the nominal factors N of the dataset are one-hot encoded. Each nominal value in the dataset is translated into a new column, and the column is assigned a 0 or 1 value. The number of columns is equal to the number of nominal values. For example, an eight-column matrix is created from a nominal factor “Runway” with eight different values (07LA, 07LD, 07RA, 07RD, 25LA, 25LD, 25RA, 25RD). The continuous features of the datasets, on the other hand, are normalized as stated in Equation (3).

$$C_{i,j}^{norm} = \frac{C_{i,j}^{orig} - \min C_j}{\max C_j - \min C_j} \quad (3)$$

where $C_{i,j}^{norm}$ represents the j th normalized continuous factors of the i th instance of the data. $C_{i,j}^{orig}$ represents the original j th continuous factors in the i th instance of the data. The $\min C_j$ and $\max C_j$ represent the minimum and maximum of the j th continuous factor in the original wind shear dataset, respectively.

Table 1. One-hot encoding of categorical factors for the modeling.

| Factor | Codes and Description |
|------------------------|---|
| LLWS Severity | 1: If LLWS magnitude is equal to greater than 25 knots, 0: ‘Otherwise’ |
| Runways | |
| 07LA | 1: If a wind shear event is reported at Runway 07LA, 0: ‘Otherwise’ |
| 07LD | 1: If a wind shear event is reported at Runway 07LD, 0: ‘Otherwise’ |
| 07RA | 1: If a wind shear event is reported at Runway 07RA, 0: ‘Otherwise’ |
| 25RD | 1: If a wind shear event is reported at Runway 25RD, 0: ‘Otherwise’ |
| 25LA | 1: If a wind shear event is reported at Runway 25LA, 0: ‘Otherwise’ |
| 25LD | 1: If a wind shear event is reported at Runway 25LD, 0: ‘Otherwise’ |
| 25RA | 1: If a wind shear event is reported at Runway 25RA, 0: ‘Otherwise’ |
| 25RD | 1: If a wind shear event is reported at Runway 25RD, 0: ‘Otherwise’ |
| Location of Occurrence | |
| 1MD | 1: If a wind shear event is reported at 1MD from Runway, 0: ‘Otherwise’ |
| 1MF | 1: If a wind shear event is reported at 1MF from Runway, 0: ‘Otherwise’ |
| 2MD | 1: If a wind shear event is reported at 2MD from Runway, 0: ‘Otherwise’ |
| 2MF | 1: If a wind shear event is reported at 2MF from Runway, 0: ‘Otherwise’ |
| 3MF | 1: If a wind shear event is reported at 3MF from Runway, 0: ‘Otherwise’ |
| RWY | 1: If a wind shear event is reported at Runway, 0: ‘Otherwise’ |
| Time of the Day | |
| Day Time | 1: If a wind shear event is reported during daytime, 0: ‘Otherwise’ |
| Night Time | 1: If a wind shear event is reported during nighttime, 0: ‘Otherwise’ |

Finally, there are 18 factors in the standardized wind shear dataset. The standardized original wind shear dataset consists of the normalized continuous factors (2 factors including hourly temperature and wind speed) as well as one-hot encoded nominal factors (16 factors).

2.4. Self-Paced Ensemble Framework

We propose a newly developed SPE framework, which is an ensemble-based imbalance learning framework, to develop a classification and prediction model for S-LLWS using untreated data from LIDAR and PIREPs. Before employing the SPE framework, we present the concepts of hardness harmonization and a self-paced factor.

2.4.1. Hardness Harmonization

All majority class samples are divided into k bins, where k a hyperparameter, based on their hardness values. Each k th bin represents a particular level of hardness. Then, majority class instances are undersampled to create a balanced dataset by maintaining the same total hardness contribution in each bin. Such a method is referred to as “harmonize” in the literature of gradient-based optimization. A similar strategy has been adopted here to harmonize the hardness in the initial iteration. However, hardness harmonization is not utilized in every iteration. The primary reason for this is that the number of trivial samples increases during the training process, as the ensemble classifier gradually conforms to the training set. Consequently, merely harmonizing the hardness contribution leaves a large number of trivial samples. Later iterations of the learning procedure are significantly slowed down by these less informative examples. In lieu of this, “self-paced factors” have been introduced to perform the self-paced harmonization of undersampling.

2.4.2. Self-Paced Factor

In particular, after harmonizing the hardness contribution of each bin, the sample probability of bins with a large population is gradually decreased. The rate of decrease is determined by a self-paced factor (σ). When σ is large, more focus is on the harder samples as opposed to the simple hardness contribution harmonization. In the initial few iterations, the framework focuses primarily on informative borderline samples, and so outliers and noise have little impact on the model’s ability to generalize. In later iterations where σ is very large, the framework retains a reasonable proportion of trivial (high confidence) samples as the “skeleton”, thereby preventing over-fitting. The detail of SPE framework is shown in Algorithm 1. It is pertinent to mention that the hardness value in each iteration (line 5–6) is updated in order to select those data samples that were most beneficial for the current ensemble. The tangent function (line 8) has been used to control the growth of the self-paced factor. Thus, the self-paced factor is equal to zero in the first iteration and to infinity in the last iteration.

Algorithm 1: Self-Paced Ensemble (SPE) Framework.

```

1  Input: Hardness function ( $h$ ), training dataset  $D = \{(x_k, y_k)\}_1^n$ , number of bins ( $k$ ), base classifier ( $\zeta$ )
   and number of base classifiers ( $\partial$ )
2  Initialize:  $P \leftarrow$  minority class in training dataset  $D$ ,  $N \leftarrow$  majority class in training dataset  $D$ ,
   Train classifier  $\zeta_0$  by using random undersampling of subsets of majority class  $N_0'$  and  $P$  such that
3  where  $|N_0'| = |P|$ 
4  for  $i = 1$  to  $n$  do
5      Ensemble  $F_i(x) = \frac{1}{i} \sum_{j=0}^{i-1} \zeta_j(x)$ 
6      Separate majority class dataset into  $k$  bins with respect to  $h(x, y, F_i)$ :  $(b_1, b_2, \dots, b_\zeta)$ 
7      In the  $l$ th bin, the average hardness contribution can be computed as
       $h_l = \sum_{s \in b_l} h(x_s, y_s, F_i) / |b_l|, \forall l = 1, 2, \dots, k$ 
8      The self-paced factor is updated as  $\sigma = \tan\left(\frac{i\pi}{2\partial}\right)$ 
9      The  $l$ th bin, non-normalized sampling weight:  $p_l = \frac{1}{h_l + \sigma}, \forall l = 1, 2, \dots, k$ 
10     Undersampling from the  $l$ th bin with  $\frac{p_l}{\sum_m p_m} |P|$  samples
11     Using newly undersampled data subset, train  $\zeta_i$ 
12 End
13 Return Final robust ensemble  $F(x) = \frac{1}{\partial} \sum_{m=1}^{\partial} \zeta_m(x)$ 

```

2.5. Bayesian Optimization for Hyperparameter Tuning

In this study, a Bayesian optimization strategy [31] is employed alongside SPE models and a tree-based machine learning model to determine the optimal hyperparameters. The Bayesian optimization built a probability model for determining the value, which automatically reduces the objective function based on the objective’s prior estimation result. Figure 3 is a flowchart of a hybrid Bayesian optimization machine learning approach. Additionally, provided below is the detailed procedure.

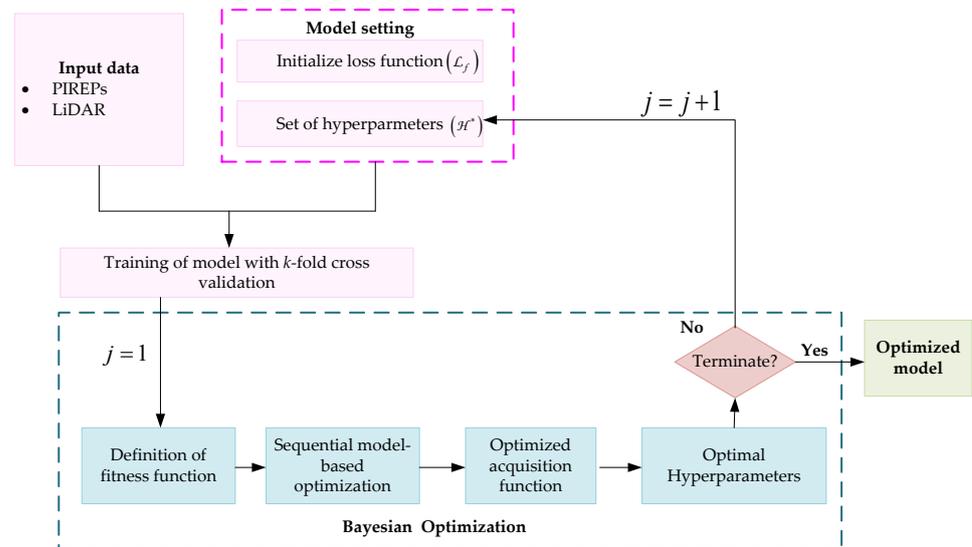


Figure 3. Bayesian approach for hyperparameters tuning.

2.5.1. Initialization

This step involves randomly initializing the appropriate hyperparameter settings (Equation (4)), which can be used to train both the SPE model and machine learning models based on k-fold cross validation. The loss function (L_f) is additionally initialized.

$$H = \begin{pmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,l} \\ h_{2,1} & h_{2,2} & \dots & h_{2,l} \\ h_{3,1} & h_{3,2} & \dots & h_{3,l} \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ h_{m,1} & h_{m,2} & \dots & h_{m,l} \end{pmatrix} \quad (4)$$

2.5.2. Fitness Function

From the initialized values, the solution’s random number is generated. Based on the following Equation (5), the fitness function can be used to minimize the objective function.

$$\text{fitness function} \left(\frac{L}{H} \right) = \begin{cases} D(H) & L < L^* \\ G(H) & L \geq L^* \end{cases} \quad (5)$$

where L denotes the loss value, $D(H)$ denotes the density estimation, which is based on the loss value during the observations, $G(H)$ is produced by the leftover observations value of loss, and L^* represents the particular quantiles.

2.5.3. Sequential Model-Based Optimization

For fine-tuning the hyperparameters of SPE and tree-based models, sequential model-based optimization is one of the succinct forms of Bayesian optimization. Sequential model-based optimization operates by finding the optimal hyperparameter setting, H^* ,

by building the Gaussian process, Θ_z , with sampled points which can be obtained by the following Equation (6).

$$H^* = \operatorname{argmin}_{\Theta_{z-1}}(H) \tag{6}$$

Equation (7) can be used to calculate the loss value under ideal hyperparameter settings.

$$L = L_f(H^*) \tag{7}$$

The corresponding L and the H^* settings are stored in the corresponding trails, which can be represented as Ω . These corresponding trails (Ω) are used for parameter settings and evaluations purposes. The Ω update can be determined with the help of following Equation (8).

$$\Omega = \Omega \cup (H^*, L) \tag{8}$$

Finally, build the Gaussian process model, Θ_z , based on updated Ω .

2.5.4. Acquisition Function

The next iteration of the search process is computed using the acquisition function of Bayesian optimization. The maximization of G-Mean, which is the expected improvement in this study, is chosen as an acceptable performance criterion for the SPE model and tree-based machine learning models. Equation (9) can be used to achieve the improvement.

$$D(H) = \max\{L_{\min} - L(H), 0\} \tag{9}$$

2.5.5. Termination

When the termination criteria are satisfied in this step, the best hyperparameters for the SPE model and tree-based machine learning models are obtained.

2.6. Evaluation of Models

In case of binary classification problem, one class is the majority (the negative) and its sample size is represented by n_1 ; the other class is the minority (the positive) and its sample size is represented by n_2 . Let n represent the total size of training dataset, $n = n_1 + n_2$. A binary classifier predicts whether each instance is positive or negative. Therefore, it generates outcomes of four types: true positive (t_p), false positive (f_p), true negative (t_n), and false negative (f_n) (see confusion matrix Figure 4). The confusion matrix provides an in-depth examination of the model's performance when predictions are made for each class. The precision and recall are two exceptionally vital model evaluation metrics. The precision is obtained as the ratio of total number of true positives to the total number of true positives and false positives, whereas recall is the ratio of total number of true positives to the total number of true positives and false negatives. Both precision and recall can be computed from the confusion matrix, as shown by Equations (10) and (11), respectively.

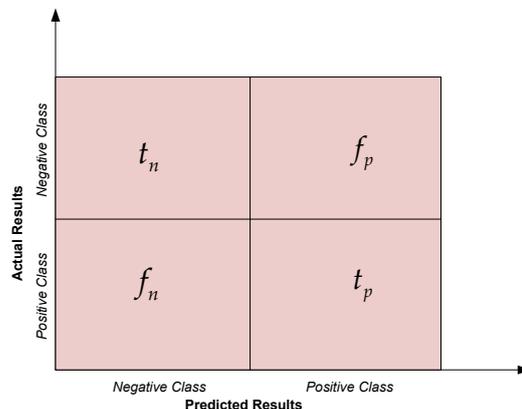


Figure 4. Confusion matrix plot.

However, in ensemble imbalance learning, the imbalanced datasets pose a large challenge to the use of proper metrics for the evaluation of the accuracy in the classification outcomes [32]. The geometric mean (G-Mean), and Matthews’ correlation coefficient (MCC) have been used in various studies instead of classification accuracy or F1-score. MCC values should range between -1 and 1 . Values closer to $+1$ represent improved performance. Both MCC and G-Mean have been generally regarded as a balanced measure which can be used even if the classes are of very different sizes. The expressions for the computation of MCC and G-Mean from confusion matrix are given by Equations (12) and (13).

$$Recall = \frac{t_p}{t_p + f_n} \tag{10}$$

$$Precision = \frac{t_p}{t_p + f_p} \tag{11}$$

$$G - Mean = \sqrt{\left(\frac{t_p}{t_p + f_n}\right) \left(\frac{t_n}{f_p + t_n}\right)} \tag{12}$$

$$MCC = \frac{t_p \times t_n - f_p \times f_n}{\sqrt{(t_p + f_p)(t_p + f)(t_n + f_p)(t_n + f_n)}} \tag{13}$$

2.7. Interpretation of Model by Shapley Additive Explanations (SHAP)

The SHAP analysis relies on a game-theoretical approach to explain the outputs of the ensemble machine learning classifiers. Since machine learning models are “black boxes”, the core ideas behind the SHAP analysis involve interpreting these models from both a global and local perspective. The SHAP values, which correspond to the value assigned to each factor in the computation of a machine learning prediction, are estimated. The contribution of each factor is determined and displayed as a Shapley value using Equation (14).

$$\varphi_i = \sum_{\Upsilon \in \Pi \setminus \{i\}} \frac{\Upsilon!(n - |\Upsilon| - 1)!}{n!} [E(\Upsilon \cup \{i\}) - E(\Upsilon)] \tag{14}$$

where φ_i illustrates the i th factor contribution, Π the set of all factors, Υ is the subset of decision factor, $E(\Upsilon_i)$ and $E(\Upsilon)$ illustrate the best model results with and without i th factors, respectively. SHAP analysis yields the outputs of machine learning models through an additive factors imputation strategy, wherein the output model is defined as a linear sum of the input factors (Equation (15)).

$$g(\Psi') = \mu_0 + \sum_{i=1}^{\Lambda} \mu_i \Psi' \quad \Psi' \in \{0, 1\}^{\Lambda} \tag{15}$$

It is equal to 1 in cases when a factor is observed, otherwise it is 0. It illustrates that the amount of all input factors, μ_0 , represents an outcome without factors (i.e., base value), whereas μ_i shows the Shapley value of factor i th.

3. Results and Discussion

In order to predict the severity of LLWS, this study used an effective and cutting-edge SPE framework along with tree-based machine learning models. Python 3.6.6, a free and open-source programming language, was used in this context. For model training, hyperparameter tuning, performance evaluation, and interpretation, we used the Scikit-learn, sklearn.metrics, HyperOpt, and Shap libraries, as well as Python’s sklearn.metrics, imbeans, and sklearn.ensemble. Figure 5 shows how LLWS events are distributed in relation to runway orientation, location of occurrence, and time of day. The box plot of the

hourly temperature and wind speed is shown in Figure 6. On the training set for tree-based models, the SMOTE-ENN treatment strategy was used. The training–validation dataset contained 257 instances of S-LLWS and 6908 instances of NS-LLWS prior to treatment. The NS-LLWS instances changed into 6518 and 3069 S-LLWS instances, respectively, after the SMOTE-ENN treatment. The performance evaluation was conducted using a testing dataset, and comparisons were made. The best model is then utilized for SHAP analysis.

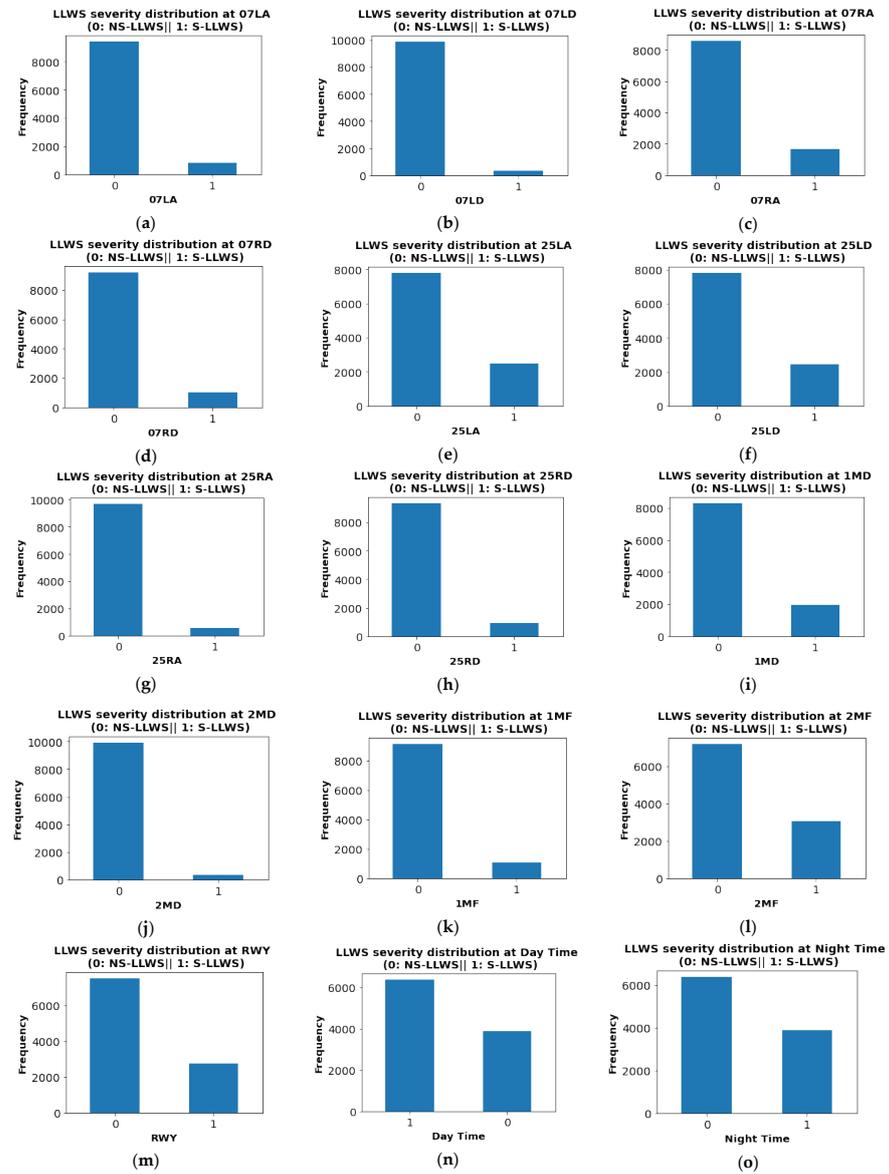


Figure 5. LLWS events distribution: (a) Frequency of S-LLWS and NS-LLWS at Runway 07LA, (b) Frequency of S-LLWS and NS-LLWS at Runway 07LD, (c) Frequency of S-LLWS and NS-LLWS at Runway 07RA, (d) Frequency of S-LLWS and NS-LLWS at Runway 07RD, (e) Frequency of S-LLWS and NS-LLWS at Runway 25LA, (f) Frequency of S-LLWS and NS-LLWS at Runway 25LD, (g) Frequency of S-LLWS and NS-LLWS at Runway 25RA, (h) Frequency of S-LLWS and NS-LLWS at Runway 25RD, (i) Frequency of S-LLWS and NS-LLWS at 1MD from Runway, (j) Frequency of S-LLWS and NS-LLWS at 2MD from Runway, (k) Frequency of S-LLWS and NS-LLWS at 1MF from Runway, (l) Frequency of S-LLWS and NS-LLWS at 2MF from Runway, (m) Frequency of S-LLWS and NS-LLWS at RWY, (n) Frequency of S-LLWS and NS-LLWS during day time, (o) Frequency of S-LLWS and NS-LLWS during night time.

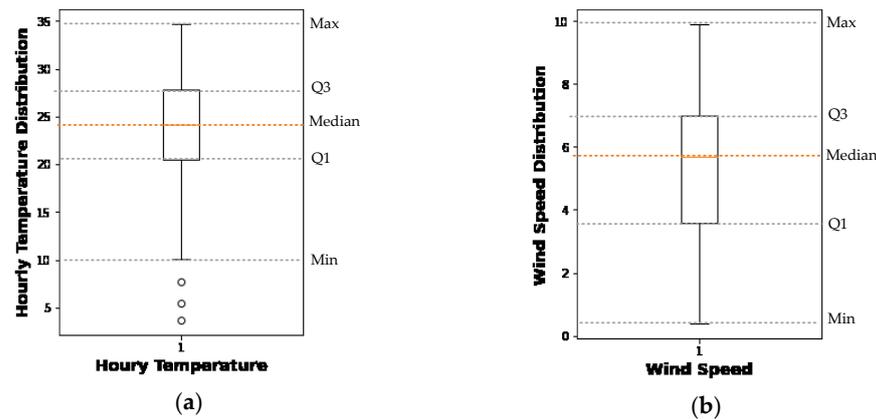


Figure 6. Box plot: (a) hourly temperature distribution (b) wind speed distribution.

3.1. Hyperparameter Tuning Using Bayesian Optimization

We used a Bayesian optimization technique that maximized the G-Mean metric to identify the optimal hyperparameters. It is important to note that the SPE framework did not require any prior data treatment, and so imbalanced data were used as input. For tree-based models, both untreated and SMOTE-ENN-treated data were used in the hyperparameter tuning process. Table 2 shows the hyperparameters along with their ranges and optimal values.

Table 2. Machine learning models hyperparameter tuning.

| Treatment | Strategy | Hyperparameters | Range | Optimal Values |
|--------------|-------------------------|-------------------------|--------------|----------------|
| No treatment | SPE | <i>n_estimators</i> | [500, 3000] | 833 |
| | | <i>max_depth</i> | [0, 10] | 7 |
| | | <i>learning_rate</i> | [0.001, 0.1] | 0.077 |
| | LGBM | <i>n_estimators</i> | [500, 3000] | 2099 |
| | | <i>learning_rate</i> | [0.001, 0.1] | 0.043 |
| | | <i>max_depth</i> | [0, 10] | 5 |
| | | <i>lambda_l1</i> | [0.001, 5] | 0.39 |
| | | <i>lambda_l2</i> | [0.001, 5] | 0.22 |
| | AdaBoost | <i>n_estimators</i> | [500–3000] | 1873 |
| | | <i>Learning_rate</i> | [0.01, 1] | 0.056 |
| CART | <i>min_samples_leaf</i> | [0.05, 0.1] | 0.04 | |
| | <i>max_depth</i> | [0, 10] | 8 | |
| SMOTE-ENN | LGBM | <i>learning_rate</i> | [0.001, 0.1] | 0.079 |
| | | <i>n_estimators</i> | [500, 3000] | 2371 |
| | | <i>max_depth</i> | [0, 10] | 4 |
| | | <i>lambda_l1</i> | [0.001, 0.1] | 0.57 |
| | | <i>lambda_l2</i> | [0.001, 0.1] | 0.41 |
| | AdaBoost | <i>n_estimators</i> | [500, 3000] | 3110 |
| | | <i>learning_rate</i> | [0.001, 0.1] | 0.093 |
| | CART | <i>min_samples_leaf</i> | [0.05, 0.1] | 0.03 |
| | | <i>max_depth</i> | [0, 10] | 8 |

3.2. Models Performance Assessment and Comparison

The terms S-LLWS and NS-LLWS events were used in this study to designate positive and negative classes of LLWS, respectively. Different performance measures that were derived from the confusion matrix were used to evaluate each model (Figure 7). The recall value and precision values in Table 3 show how well the classifier performed in correctly classifying S-LLWS cases and NS-LLWS cases, respectively. All models were

observed to be able to classify NS-LLWS events with high accuracy—more than 95.02%. Given the large number of NS-LLWS cases in the LLWS data, this was expected. The SPE framework utilizing testing data had an 80.13% recall value, compared to all others, each of which had a recall value of between 0.00% and 62.43% regarding the recall values for correctly classifying S-LLWS cases. Figure 7 demonstrates that 88 of the 110 S-LLWS cases in the testing dataset were correctly classified by the SPE framework. After that, CART with SMOTE-ENN-treated data were used, correctly classifying 68 out of 110 S-LLWS while incorrectly classifying the remaining 42. S-LLWS by SPE had a relative classification accuracy rate of 29.41% higher than CART with SMOTE-ENN-treated data. The AdaBoost model with no data treatment did the worst job of correctly classifying S-LLWS. The 110 S-LLWS cases were incorrectly classified in none of them.

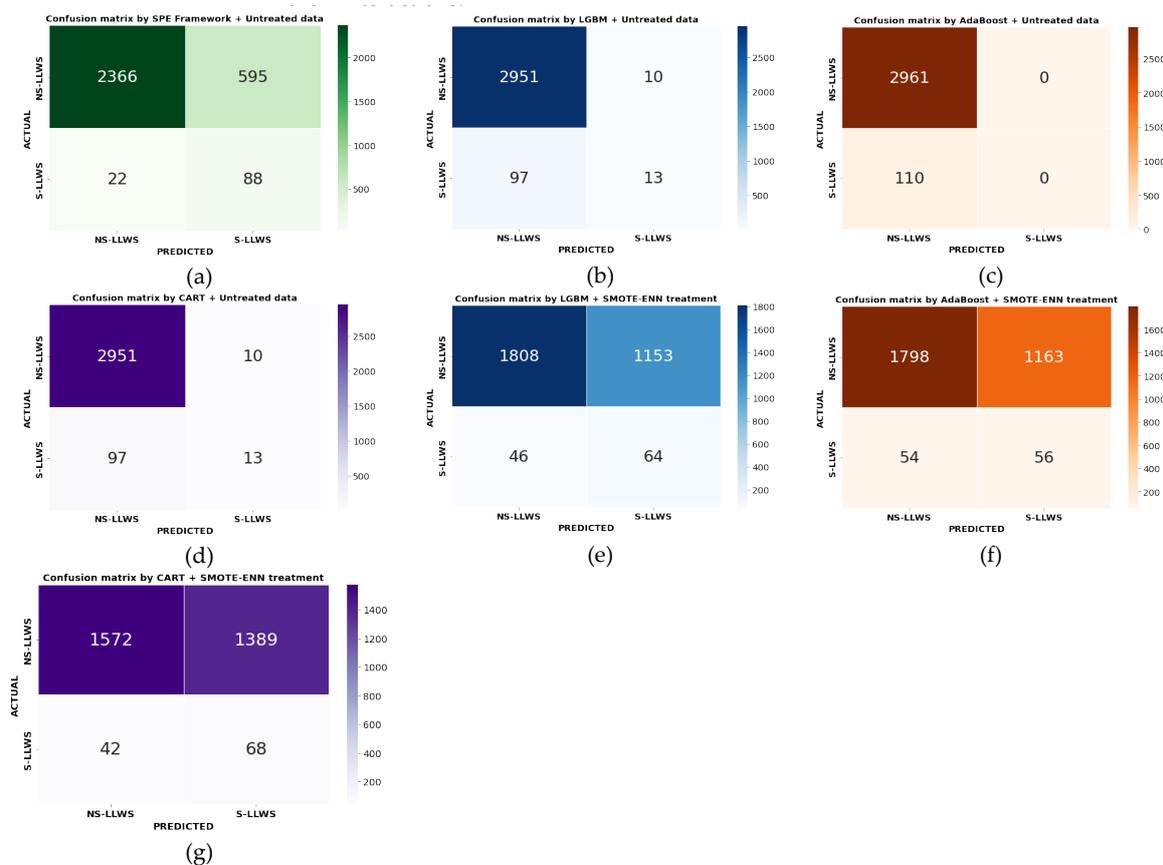


Figure 7. Confusion Matrix: (a) SPE framework, (b) LGBM without data treatment, (c) AdaBoost without data treatment, (d) CART without data treatment, (e) LGBM with SMOTE-ENN data treatment, (f) AdaBoost with SMOTE-ENN data treatment, (g) CART with SMOTE-ENN data treatment.

In addition, we have utilized G-Mean and MCC methods in our study. On the testing dataset, the SPE framework demonstrated a higher G-Mean than all other models with treated and untreated data. G-Mean was 0.82 for the SPE framework and 0.59 for LGBM with SMOTE-ENN-treated data. AdaBoost displays the lowest G-Mean value of 0.50 with no treated data. The G-Mean value of the SPE framework was 39.98% greater than that of the LGBM with SMOTE-ENN-treated data. Likewise, comparing MCC values, the SPE framework also outperformed LGBM, AdaBoost, and CART models, with an MCC value of 0.27 indicating the best performance, followed by 0.24 for LGBM. Using G-Mean and MCC as balanced measures of performance, the SPE framework utilizing imbalanced data outperformed the tree-based model SMOTE-ENN that was applied to the balanced data. Consequently, it could be regarded as the optimal model for the interpretation provided by the SHAP analysis, such as the relative importance of factors, their contributions, and their interactions.

Table 3. Performance measure of machine learning models.

| Treatment | Model | Class | Precision | Recall | G-Mean | MCC |
|--------------|----------|---------|-----------|--------|--------|------|
| No treatment | SPE | NS-LLWS | 0.99 | 0.80 | 0.82 | 0.27 |
| | | S-LLWS | 0.13 | 0.80 | | |
| | | Average | 0.56 | 0.80 | | |
| | LGBM | NS-LLWS | 0.97 | 1.00 | 0.55 | 0.24 |
| | | S-LLWS | 0.57 | 0.12 | | |
| | | Average | 0.77 | 0.56 | | |
| | AdaBoost | NS-LLWS | 0.96 | 1.00 | 0.50 | 0.00 |
| | | S-LLWS | 0.00 | 0.00 | | |
| | | Average | 0.48 | 0.56 | | |
| SMOTE-ENN | CART | NS-LLWS | 0.97 | 1.00 | 0.55 | 0.23 |
| | | S-LLWS | 0.57 | 0.12 | | |
| | | Average | 0.77 | 0.56 | | |
| | LGBM | NS-LLWS | 0.98 | 0.61 | 0.59 | 0.07 |
| | | S-LLWS | 0.05 | 0.58 | | |
| | | Average | 0.51 | 0.60 | | |
| | AdaBoost | NS-LLWS | 0.97 | 0.61 | 0.58 | 0.04 |
| | | S-LLWS | 0.05 | 0.51 | | |
| | | Average | 0.51 | 0.56 | | |
| CART | NS-LLWS | 0.97 | 0.53 | 0.57 | 0.05 | |
| | S-LLWS | 0.05 | 0.62 | | | |
| | Average | 0.51 | 0.57 | | | |

3.3. Self-Paced Ensemble Framework Interpretation by SHAP

3.3.1. Global Factor Interpretation

Numerous techniques can be employed to determine the relative significance of factors. However, factor contribution is distinct from factor significance. The contribution of a factor indicates which factor has the greatest influence on a model’s performance. In addition to identifying relevant factors, the factor contributions provide a rational explanation for the observed results. This study investigated the significance of each factor and its contribution using SHAP analysis. Figure 8a depicts, initially, the factor importance of the input factors, indicating the overarching influence of the factors on the predictions. It is the mean of the absolute Shapley values for the entire training dataset. The average absolute SHAP value of 0.185 indicates that, of all the features, “Runway 25LD” is the most vulnerable to S-LLWS occurrences. The average absolute SHAP values for “hourly temperature” and “wind speed” are 0.145 and 0.135, respectively, making them the second and third most influential factors.

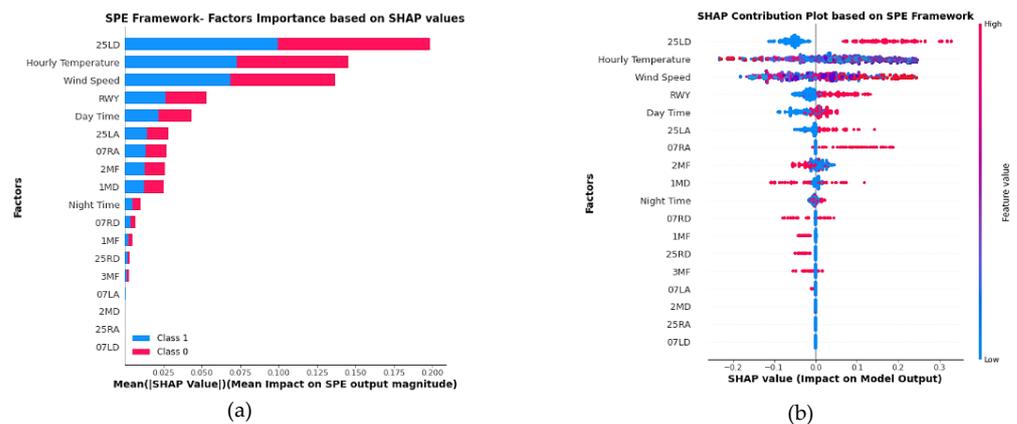


Figure 8. Global Factor Interpretation; (a) Factor Importance plot; (b) Factor Contribution plot.

Figure 8b is a SHAP contribution plot of the factors, illustrating the distribution of SHAP values for each factor and the corresponding impact patterns. It is also known as the SHAP bee swarm plot. The horizontal axis of this plot represents the SHAP value, while the vertical axis contains all of the factors in the LLWS dataset. Each point on the plot represents a single SHAP value for a given prediction. Red indicates a higher value for a factor, while blue indicates a lower value. Based on the distribution of the red and blue dots, we can derive a general sense of the impact of factors' directionality. Some valuable insights can be drawn from the plots for the top three factors.

The runway 25LD factor, denoted by red dots, is coded as 1. All the red dots fall to the right of the vertical reference line, indicating the likelihood of the occurrence of S-LLWS over runway 25LD. Blue dots fall to the left of the vertical line, indicating the occurrence of NS-LLWS over other runways of HKIA. The previous studies [33,34] indicated that hourly prevailing wind directions such as east, south-east, south, and south-west were found to cause a higher risk of S-LLWS. This indicates that at 25LD, an S-LLWS event could be more likely to happen under the easterly, southeasterly, southerly, and southwesterly directions.

In the case of the hourly temperature factor, most of the purple dots fall to the right of the vertical line, while most of the blue dots and red dots fall to the left of the vertical line. This illustrates that S-LLWS is most likely to occur at low-to-moderate hourly temperatures, while a few high temperatures are more likely to cause the occurrence of NS-LLWS. The reason for this might be a temperature inversion [35–37], which is an alteration in the troposphere's typical temperature lapse rate, i.e., the reduction in temperature with altitude. On chilly, clear nights, this phenomenon typically occurs close to the ground, where the air immediately above the ground rapidly cools and becomes much colder than the layer of air higher up. As a result, the densely packed lower-level cold air is trapped by the layer of warm air. This may result in severe turbulence and, subsequently, S-LLWS.

Moderate-to-high values of wind speed mostly caused the occurrence of S-LLWS and vice versa. The findings are also consistent with previous HKIA research [33,38–41]. As for the occurrence of LLWS, however, wind speed variation is more significant than mean wind speed. Due to the fact that the average duration of an LLWS event confronted by an aircraft is somewhere between a few seconds and several minutes, the hourly mean wind speed cannot offer an accurate indication of LLWS. Therefore, more sophisticated data about wind conditions, such as a 1 min mean turbulence intensity, is necessary to enhance the performance of the models.

3.3.2. Local Factor Interpretation

Figure 9 depicts the SHAP explanatory force chart for two instances, randomly selected from the actual estimate results. The base value (0.656) on the plot represents the mean optimal SPE framework model estimations for the training dataset. The NS-LLWS condition occurs when the SPE framework output value is less than the base value. The S-LLWS condition occurs when the output value of the SPE Framework exceeds the base value. The blue arrows represent the magnitude of the influence of an input factor on the probability of NS-LLWS events. The influence of input factors on the occurrence of S-LLWS is highlighted by red arrows. The amount of space a factor occupies on each arrow demonstrates the size of its effect.

Consider two LLWS severity cases, one from S-LLWS and the other from NS-LLWS, which were correctly classified with estimated values of 1.03 and 0.52, respectively (see Figure 9). The value for S-LLWS is greater than the base value (0.656). Similarly, the value for NS-LLWS is less than 0.656. Figure 9a depicts an S-LLWS event that occurred when runway 25LD = 1, wind speed = 2.2 m/s, and hourly temperature = 17.9C. This is shown by the red arrows pointing to the right. The size of the "Runway 25LD" arrow is larger than the "Wind Speed" and "Hourly Temperature" arrows. This shows that "Runway 25LD" is a stronger predictor of S-LLWS in this case than "Wind Speed" and "Hourly Temperature." In contrast, for the same instance, "Day Time = 0", as represented by the blue arrow pointing to the left, indicates nighttime and depicts the likelihood of the occurrence of NS-LLWS.

Similarly, in Figure 9b, for another instance correctly classified as NS-LLWS, “1MD = 1”, “Wind Speed = 6.9 m/s”, and “pointing” the blue arrows, pointing to the left, are more likely to result in the occurrence of NS-LLWS. It demonstrates that, 1 nautical mile away from the end of the runway, an NS-LLWS event occurred.

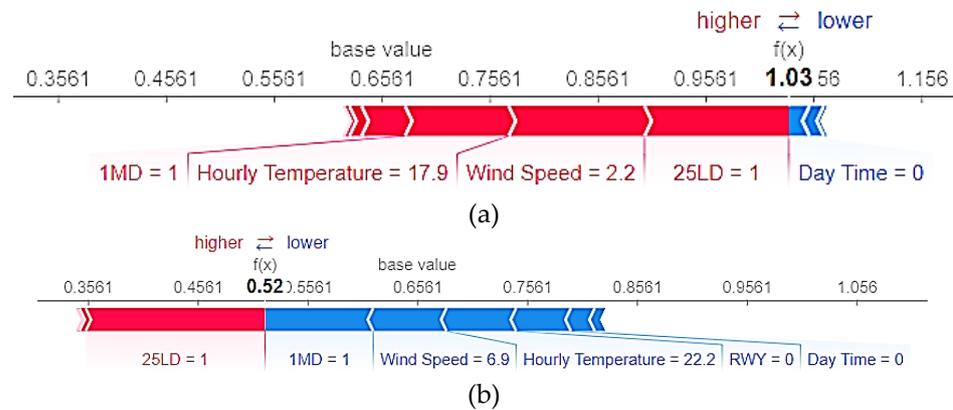


Figure 9. SHAP Explanatory Force Pot: (a) Plot for an instance value equals to 1.03; (b) Plot for an instance value equals to 0.52.

3.3.3. Factor Interaction Analysis

The SHAP interaction plots are examined to identify how the input factors, used to evaluate the SPE framework, interact with one another in terms of their contributions (see Figure 10). The interaction analysis of the top four influential factors, i.e., runway 25LD, hourly temperature, wind speed, and RWY (horizontal location of LLWS occurrence), is provided. Other factors’ interactions, however, could be examined as well. The red and blue scatter plots in Figure 10a depict the variability in the runway 25LD and 25LD SHAP values. When the hourly temperature is low to moderate, the SHAP value for runway 25LD is higher. This means that most of the S-LLWS occurs in the vicinity of runway 25LD when the hourly temperature ranges from low to moderate. The temperature inversion on Hong Kong’s Lantau Island could also be contributing to this scenario.

Figure 10b depicts the distribution of wind speed at runway 25LD. Wind speed points greater than 5 m/s have a higher SHAP value, indicating the likelihood of an S-LLWS event. Figure 10c illustrates that most of the S-LLWS events occurred “on the runway.” The PIREPs reported S-LLWS when aircraft were making their final approach or just when they became airborne after takeoff.

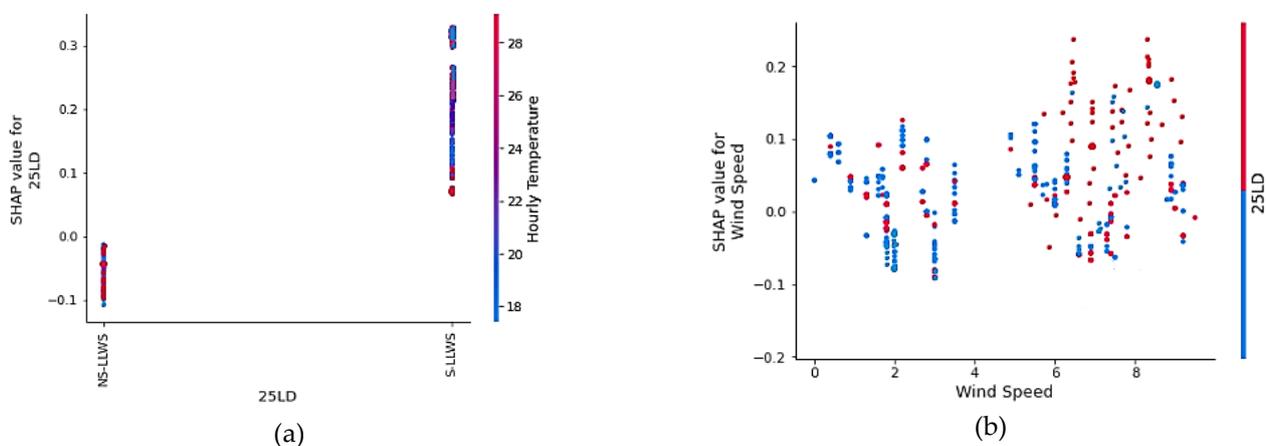


Figure 10. Cont.

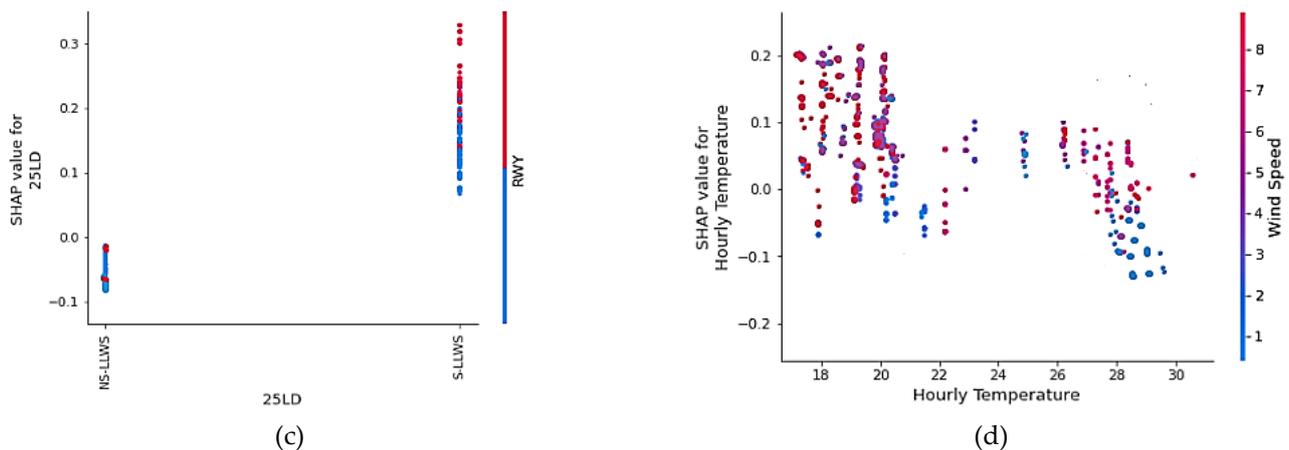


Figure 10. SHAP Interaction Plots: (a) Interaction of Runway 25LD and Hourly Temperature; (b) Interaction of Wind Speed and Runway 25LD; (c) Interaction of Runway 25LD and RWY (location of LLWS occurrence); (d) Interaction of Hourly Temperature and Wind Speed.

Figure 10d shows that the optimum conditions for the occurrence of S-LLWS were lower than average hourly temperatures and medium-to-high wind speeds. The points representing that scenario fall to the left of the plot and above the SHAP 0.00 reference line. However, to obtain a clear threshold, it may be necessary to know the altitude at which LLWS happen in addition to the parameters that are already known.

4. Conclusions and Recommendations

In this research, a novel SPE framework for the prediction and imbalance classification of LLWS severity has been proposed and compared with tree-based machine learning models, using both treated and untreated HKIA-based LLWS data from LiDAR and PIREPs. The SHAP interpretation system was also used to identify key risk factors and quantify their effects on the occurrence of S-LLWS. In this study, the SPE framework was trained and evaluated using untreated data, whereas both untreated and treated data were used to train the LGBM, AdaBoost, and CART machine learning models. The SMOTE-ENN technique was used as a treatment technique for highly imbalanced LLWS data. In terms of precision, recall, G-Mean, and MCC, the experimental results demonstrated that the SPE framework, based on the untreated data, outperforms all other tree-based models. The newly introduced SPE framework model offers a viable option for modeling and predicting LLWS severity based on imbalanced LLWS data.

Machine learning models, on the other hand, are regularly chastised for their lack of transparency and interpretability. Despite the fact that machine learning models are more adaptable and efficient than statistical approaches, their widespread recognition in the engineering domain continues to be a challenge. To tackle the SPE framework's interpretability issue, the SHAP interpretation system was used to evaluate the SPE's output in order to identify major risk factors and assess their impact on the severity of the LLWS. The results of the SHAP interpretation system can be used to rank the risk factor's overall significance. It can also be used to look into the individual and interaction effects of risk factors (for instance, how specific effects alter in response to changes in the risk factor's value). The analysis revealed that runway 25LD, hourly temperature, wind speed, and RWY (location of LLWS occurrence) were the top four most significant factors in predicting LLWS severity. The optimistic projections for the occurrence of S-LLWS events were low-to-medium temperatures at runway 25LD with relatively moderate-to-high wind speeds. Likewise, most of the S-LLWS events happen on the runway.

This research outlines a strategy that can be used to conduct a large-scale analysis of LLWS in aviation and serves as a useful tool for aviation policymakers and air traffic safety researchers. This paper discussed the SPE framework using highly imbalanced

LLWS data and the SHAP interpretation system. Additional research could be conducted by combining a number of other machine learning techniques with a number of additional risk factors, including monthly variation, location of occurrence of LLWS above ground level, etc. Future research could be expanded by employing additional techniques for augmenting data to deal with highly imbalanced LLWS data.

Author Contributions: Conceptualization, A.K.; Data curation, P.-W.C.; Formal analysis, A.K.; Funding acquisition, P.-W.C. and F.C.; Investigation, H.P.; Methodology, A.K., P.-W.C. and H.P.; Project administration, F.C.; Resources, P.-W.C.; Supervision, F.C.; Validation, F.C. and H.P.; Visualization, A.K.; Writing—original draft, A.K.; Writing—review and editing, H.P. All authors have read and agreed to the published version of the manuscript.

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

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We are thankful to the Hong Kong Observatory at Hong Kong International Airport for providing us PIREPs and LiDAR data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Borsky, S.; Unterberger, C. Bad weather and flight delays: The impact of sudden and slow onset weather events. *Econ. Transp.* **2019**, *18*, 10–26. [[CrossRef](#)]
2. Choi, S.; Kim, Y.J.; Briceno, S.; Mavris, D. Prediction of weather-induced airline delays based on machine learning algorithms. In Proceedings of the 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), Sacramento, CA, USA, 25–29 September 2016; IEEE: New York, NY, USA; pp. 1–6.
3. Gultepe, I.; Sharman, R.; Williams, P.D.; Zhou, B.; Ellrod, G.; Minnis, P.; Trier, S.; Griffin, S.; Yum, S.; Gharabaghi, B.; et al. A review of high impact weather for aviation meteorology. *Pure Appl. Geophys.* **2019**, *176*, 1869–1921. [[CrossRef](#)]
4. Airport Council International. *World Airport Traffic Forecast 2017–2040 Airport Council International, Montréal*. 2017. Available online: https://issuu.com/aciworld/docs/aci_annualreport2017 (accessed on 21 February 2022).
5. Fichtl, G.H.; Camp, D.W.; Frost, W. Sources of low-level wind shear around airports. *J. Aircr.* **1977**, *14*, 5–14. [[CrossRef](#)]
6. Stratton, D.A.; Stengel, R.F. Probabilistic reasoning for intelligent wind shear avoidance. *J. Guid. Control. Dyn.* **1992**, *15*, 247–254. [[CrossRef](#)]
7. Shun, C.; Chan, P. Applications of an infrared Doppler LiDAR in detection of wind shear. *J. Atmos. Ocean. Technol.* **2008**, *25*, 637–655. [[CrossRef](#)]
8. Thobois, L.; Cariou, J.P.; Gultepe, I. Review of LiDAR-based applications for aviation weather. *Pure Appl. Geophys.* **2019**, *176*, 1959–1976. [[CrossRef](#)]
9. Hallowell, R.G.; Cho, J.Y. Wind-shear system cost-benefit analysis. *Linc. Lab. J.* **2010**, *18*, 47–68.
10. Lau, S.Y.; Shun, C.M. Terrain-induced wind shear during the passage of Typhoon Utor near Hong Kong in July 2001. In Proceedings of the Tenth Conference on Mountain Meteorology and MAP Meeting, Park City, UT, USA, 16–21 June 2002; pp. 433–436, in preprints.
11. Hon, K.K.; Chan, P.W. Improving LiDAR Windshear Detection Efficiency by Removal of “Gentle Ramps”. *Atmosphere* **2021**, *12*, 1539. [[CrossRef](#)]
12. Chan, P.W.; Hon, K.K. Performance of super high resolution numerical weather prediction model in forecasting terrain-disrupted airflow at the Hong Kong International Airport: Case studies. *Meteorol. Appl.* **2016**, *23*, 101–114. [[CrossRef](#)]
13. Boilley, A.; Mahfouf, J.F. Wind shear over the Nice Côte d’Azur airport: Case Study. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 2223–2238. [[CrossRef](#)]
14. Matayoshi, N.; Iijima, T.; Yamamoto, K.; Fujita, E. Development of Airport Low-level Wind Information (ALWIN). In Proceedings of the 16th AIAA Aviation Technology, Integration, and Operations Conference, Washington, DC, USA, 13–17 June 2016; p. 4362.
15. Zhang, H.; Wu, S.; Wang, Q.; Liu, B.; Yin, B.; Zhai, X. Airport low-level wind shear LiDAR observation at Beijing Capital International Airport. *Infrared Phys. Technol.* **2019**, *96*, 113–122. [[CrossRef](#)]
16. Hon, K.-K. Predicting low-level wind shear using 200-m-resolution NWP at the Hong Kong International Airport. *J. Appl. Meteorol. Climatol.* **2020**, *59*, 193–206. [[CrossRef](#)]

17. Keck, R.E.; Mikkelsen, R.; Troldborg, N.; de Maré, M.; Hansen, K.S. Synthetic atmospheric turbulence and wind shear in large eddy simulations of wind turbine wakes. *Wind Energy* **2014**, *17*, 1247–1267. [[CrossRef](#)]
18. Lei, L.; Chan, P.W.; Li-Jie, Z.; Hui, M. Numerical simulation of terrain-induced vortex/wave shedding at the Hong Kong International Airport. *Meteorol. Z.* **2013**, *22*, 317–327. [[CrossRef](#)]
19. Jiang, L.; Xie, Y.; Wen, X.; Ren, T. Modeling highly imbalanced crash severity data by ensemble methods and global sensitivity analysis. *J. Transp. Saf. Secur.* **2022**, *14*, 562–584. [[CrossRef](#)]
20. Guo, R.; Fu, D.; Sollazzo, G. An ensemble learning model for asphalt pavement performance prediction based on gradient boosting decision tree. *Int. J. Pavement Eng.* **2021**, *23*, 3633–3646. [[CrossRef](#)]
21. Feng, D.C.; Wang, W.J.; Mangalathu, S.; Taciroglu, E. Interpretable XGBoost-SHAP machine-learning model for shear strength prediction of squat RC walls. *J. Struct. Eng.* **2021**, *147*, 04021173. [[CrossRef](#)]
22. Zhang, S.; Khattak, A.; Matara, C.M.; Hussain, A.; Farooq, A. Hybrid feature selection-based machine learning Classification system for the prediction of injury severity in single and multiple-vehicle accidents. *PLoS ONE* **2022**, *17*, e0262941. [[CrossRef](#)]
23. Khattak, A.; Almujibah, H.; Elamary, A.; Matara, C.M. Interpretable Dynamic Ensemble Selection Approach for the Prediction of Road Traffic Injury Severity: A Case Study of Pakistan's National Highway N-5. *Sustainability* **2022**, *14*, 12340. [[CrossRef](#)]
24. Liu, Z.; Cao, W.; Gao, Z.; Bian, J.; Chen, H.; Chang, Y.; Liu, T.Y. Self-paced ensemble for highly imbalanced massive data classification. In Proceedings of the 2020 IEEE 36th International Conference on Data Engineering (ICDE), Dallas, TX, USA, 20–24 April 2020; IEEE: New York, NY, USA; pp. 841–852.
25. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
26. Mangalathu, S.; Hwang, S.H.; Jeon, J.S. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Eng. Struct.* **2020**, *219*, 110927. [[CrossRef](#)]
27. Dong, S.; Khattak, A.; Ullah, I.; Zhou, J.; Hussain, A. Predicting and analyzing road traffic injury severity using boosting-based ensemble learning models with SHapley Additive exPlanations. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2925. [[CrossRef](#)] [[PubMed](#)]
28. Szeto, K.C.; Chan, P.W. High resolution numerical modelling of wind shear episodes at the Hong Kong International Airport. In Proceedings of the 12th Conference on Aviation, Range, and Aerospace Meteorology, Atlanta, GA, USA, 29 January–2 February 2006.
29. Hon, K.K.; Chan, P.W. Historical analysis (2001–2019) of low-level wind shear at the Hong Kong International Airport. *Meteorol. Appl.* **2022**, *29*, e2063. [[CrossRef](#)]
30. Jones, J.G.; Haynes, A. *A Peakspotter Program Applied to the Analysis of Increments in Turbulence Velocity*; RAE: Bedford, VA, USA, 1984.
31. Woodfield, A.A.; Woods, J.F. *Worldwide Experience of Wind Shear during 1981–1982*; Royal Aircraft Establishment: Bedford, VA, USA, 1983.
32. Wu, J.; Chen, X.Y.; Zhang, H.; Ziong, I.-D.; Lei, H.; Deng, S.-H. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. Electron. Sci. Technol.* **2019**, *17*, 26–40.
33. Zhu, Q. On the performance of Matthews correlation coefficient (MCC) for imbalanced dataset. *Pattern Recognit. Lett.* **2020**, *136*, 71–80. [[CrossRef](#)]
34. Chen, F.; Peng, H.; Chan, P.W.; Ma, X.; Zeng, X. Assessing the risk of windshear occurrence at HKIA using rare-event logistic regression. *Meteorol. Appl.* **2020**, *27*, e1962. [[CrossRef](#)]
35. Chan, P.W. Case study of a special event of low-level windshear and turbulence at the Hong Kong International Airport. *Atmos. Sci. Lett.* **2022**, e1143. [[CrossRef](#)]
36. Stocker, J.; Johnson, K.; Forsyth, E.; Smith, S.; Gray, S.; Carruthers, D.; Chan, P.W. Derivation of High-Resolution Meteorological Parameters for Use in Airport Wind Shear Now-Casting Applications. *Atmosphere* **2022**, *13*, 328. [[CrossRef](#)]
37. Gernowo, R.; Subagio, A.; Adi, K.; Widodo, A.P.; Widodo, C.E.; Putranto, A.B. Atmospheric dynamics and early warning system low level windshear for airport runway hazard mitigations. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2021; Volume 1943, p. 012029.
38. Leonidov, V.I.; Semenets, V.V. Analysis of methods for wind shear detection in area of airports by data of atmosphere acoustic sounding. *Telecommun. Radio Eng.* **2018**, *77*, 363–372. [[CrossRef](#)]
39. Chen, F.; Peng, H.; Chan, P.W.; Zeng, X. Low-level wind effects on the glide paths of the North Runway of HKIA: A wind tunnel study. *Build. Environ.* **2019**, *164*, 106337. [[CrossRef](#)]
40. Chen, F.; Peng, H.; Chan, P.W.; Zeng, X. Wind tunnel testing of the effect of terrain on the wind characteristics of airport glide paths. *J. Wind Eng. Ind. Aerodyn.* **2020**, *203*, 104253. [[CrossRef](#)]
41. Chen, F.; Peng, H.; Chan, P.W.; Huang, Y.; Hon, K.K. Identification and analysis of terrain-induced low-level wind shear at Hong Kong International Airport based on WRF–LES combining method. *Meteorol. Atmos. Phys.* **2022**, *134*, 60. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.