Hindawi
Advances in Meteorology
Volume 2022, Article ID 3039797, 14 pages
https://doi.org/10.1155/2022/3039797

Research Article

LiDAR-Based Windshear Detection via Statistical Features

Jie Zhang,[1] **Pak Wai Chan**,[2] **and Michael K. Ng**[1]

[1]*Department of Mathematics, University of Hong Kong, Kowloon, Hong Kong*
[2]*Aviation Weather Services, Hong Kong Observatory, Kowloon, Hong Kong*

Correspondence should be addressed to Michael K. Ng; mng@maths.hku.hk

Received 11 October 2022; Revised 18 November 2022; Accepted 23 November 2022; Published 13 December 2022

Academic Editor: Francisco Molero

Windshear is a kind of microscale meteorological phenomenon which can cause danger to the landing and takeoff of aircrafts. Accurate windshear detection plays a crucial role in aviation safety. With the development of machine learning, several learning-based methods are proposed for windshear detection, i.e., windshear and non-windshear classification. To obtain accurate detection results, it is significant to extract features that can distinguish windshear and non-windshear properly from the obtained wind velocity data. In this paper, we mainly introduce two statistical indicators derived from the Doppler Light Detection and Ranging (LiDAR) observational wind velocity data by plan position illustrate (PPI) scans for windshear features construction. Besides the indicators directly derived from the wind velocity data, we also study the visual information from the corresponding conical images of wind velocity. Based on the proposed indicators, we construct three feature vectors for windshear and non-windshear classification. Inspired by the idea of multiple instance learning, the wind velocity data collected in the 4 minutes within the reported time spot are considered in the procedure of feature vector construction, which can reduce the possibility of windshear features missing. Both statistical methods and clustering methods are applied to evaluate the effectiveness of the proposed feature vectors. Numerical results show that the proposed feature vectors have good effect on windshear and non-windshear classification and can be used to provide more accurate windshear alerting to pilots in practice.

1. Introduction

Windshear refers to a sudden and sustained change of the headwind/tailwind encountered by aircrafts in direction and/or speed [1, 2]. A significant windshear occurring might change the lift force that the aircraft would experience and subsequently make the aircraft fly below or above the intended fight path. This might cause difficulties for the pilot to control the aircraft safely. Accurate windshear alerts can help pilots take timely and appropriate corrective actions to ensure the safety of the aircraft. Hence, windshear detection for airports are crucial for efficient and safe air traffic control (ATC) [3].

The occurrence of windshear strongly corresponds to convection, frontal systems, thermal instabilities, and microbursts [3]. Generally, windshear could occur in both rainy days and non-rainy days for different reasons [4]. In rainy weather conditions, microbursts and gust fronts associated with severe convection can result in windshear, which could be detected by anemometers and rain-detecting Doppler weather radar. In non-rainy weather conditions, windshear occurs mainly due to terrain effects, sea breezes, low-level jets, dry microbursts, etc. However, since there are few hydrometeors as reflectors for microwaves in clear air [2], it is difficult to obtain adequate wind velocity data by Doppler weather radar on non-rainy days. In addition, anemometers can only detect windshear within the lowest couple of hundred meters [4, 5]. Therefore, it is ineffective to capture dry windshear events at higher altitudes only by anemometers and Doppler weather radar during non-rainy weather conditions. Although wind profilers can be used as complementary tools to detect windshear to some extent, they are unstable for windshear detection along glide and takeoff paths [4].

With the development of Light Detection and Ranging (LiDAR) techniques [6, 7], a number of LiDAR-based windshear detection methods have been proposed, which are effective for dry windshear detection. Most of current

LiDAR windshear detection algorithms are based upon some indicators corresponding to the flight condition and mathematical statistics of wind velocity. In reference [1], Chan et al. introduced the operational LiDAR-based windshear detection system for the Hong Kong International Airport (HKIA) in which a glide path scan method is proposed to measure the headwind along the individual glide paths and a wind ramp detection method is used to detect windshear. Wind ramp refers to the change of wind speed along a certain distance (ramp length) [8, 9]. In practice, windshear ramps with different detected ramp lengths $(400, 800, \ldots, 6400$ m) will be detected. Some of them will be picked out by ramp prioritization based on a severity factor that corresponds to the ratio between wind velocity change and the inverse cube root of the related ramp length. If any one of chosen windshear ramps reaches or exceeds 14 knots (the predetermined operational threshold at HKIA), windshear alert will be generated. However, this method neglects the internal features of the ramp waveform, which may reflect the true feelings of the pilot. To address this issue, Li et al. [7] proposed a novel ramp algorithm based on a weighted smoothing approach and a secondary windshear recognition scheme, which can detect the flat areas for windshear ramp. In reference [10], Hon et al. introduced a "gentle ramp" removal method for the automatic windshear detection algorithm at HKIA in which the root-mean-square (RMS) of the original headwind profile from its running mean will be additionally checked for a selected windshear ramp whose intensity reaches or exceeds 14 knots and ramp length is greater than 3000 m (about 39 s for aircraft flight). Based on the consultation by aviation users, if the RMS measure of velocity fluctuations falls below 1.2 knots, the windshear alert would be withheld. Due to the transient and sporadic nature of terrain-disrupted airflow disturbances, it is difficult for the pilots to clearly differentiate between windshear and turbulence at times. In such cases, wind ramp-based methods may not work and it is advantageous to consider the gradient of headwind. F-factor is such an indicator whose first term is corresponding to the headwind gradient and second term is about the vertical acceleration term. Chan et al. investigated the performance of LiDAR-based F-factor for windshear detection as well as its statistics in reference [11]. However, they did not study the aircraft response to the headwind change and the vertical acceleration term of F-factor. In reference [12], Chan took the aircraft response into account by preprocessing the headwind profile using commercial flight simulator software and established a new threshold (−0.05) of F-factor that is less than the conventional threshold of F-factor for windshear detection (−0.105) based on the dataset collected at HKIA. In reference [13], the corresponding research is further in-depth. The center-averaging postprocessed F-factor was investigated besides the consideration of flight simulator-based F-factor. By selecting an optimal alerting threshold specific for each runway corridor, the successful alerting rate of F-factor could reach 86%. In addition to the wind ramp and F-factor-associated windshear detection methods, there are some other methods to detect windshear. For instance, Wu and Hon [14] applied the Fourier transformation to decompose the headwind profile and proposed to detect windshear by threshold based single/multiple channel method. A regional divergence algorithm based on LiDAR data was proposed by Li et al. in reference [15] for windshear detection at Lanzhou Zhongchuan International Airport which mainly concentrated on the regional wind change size along the runway.

In the past few decades, due to the development of machine learning, several machine learning methods for windshear detection by LiDAR data are proposed. For instance, Ma et al. [16] extracted the windshear features from the partial LiDAR scanning image by an invariant moment method and gray-gradient co-occurrence matrix. Then, the support vector machine (SVM) method, which can generate a line or hyperplane to separate data into two classes based on the input features, is used to validate the effectiveness of the extracted features. However, the accuracy of this method is too low (lower than 50%) to use in practice. In reference [17], Huang et al. proposed a statistical indicator based on the headwind profiles which measures the maximal difference in wind velocities along the range of the measurement LiDAR beam for different azimuth ranges. By employing a one-side normal distribution based decision rule, the threshold of this indicator is determined for windshear and non-windshear distinguishing. However, this method only considers the wind profile obtained at the nearest timestamp to the reported time spot, which may miss some windshear-related data.

In this paper, we will introduce two statistical indicators of windshear from LiDAR observational data. One is derived from the observed wind velocity data, and the other is constructed by the texture feature of the LiDAR plan position illustrate (PPI) scan images. Based upon the idea of multiple instance learning [18], we can construct several feature vectors based on the proposed indicators. This research is valuable in a number of aspects. First, the proposed indicators are based on the LiDAR data obtained from PPI scan whose requirements for elevation and azimuth angle change are lower than the frequently used glide path scan. Second, wind profiles collected in 4 minutes near the reported timestamp are considered, which could reduce the possibility of windshear features loss. Third, the application of the image texture extraction method can improve the identification efficiency of sporadic windshear. Moreover, numerical experiments based on real data collected at HKIA validate the effectiveness of the proposed indicators.

The outline of this paper is given as follows: in the second section, we will introduce the information of the LiDAR observational data we studied in this paper. The construction of statistical indicators and related feature vectors are presented in Section 3. In Section 4, numerical results and discussion will be shown. Finally, we will give some concluding remarks in Section 5.

2. LiDAR Observational Data

The Hong Kong International Airport (HKIA) is located in the northern part of Lantau Island, which is mountainous between 300 m and 900 m above sea level. Windshear

occurs frequently since the terrain near the airport is quite complex. In 2002, the Hong Kong Observatory devised a Doppler LiDAR system to HKIA for timely windshear detection and alerting [5, 19]. The wind velocity data collected in PPI mode with elevation angles of 3° (landing) and 6° (takeoff) are provided for windshear detection. 20–30 seconds are required to get one PPI scan. In Figure 1, we show two examples of the LiDAR radial velocity data of PPI scan, where the radius and the polar angle of the scan refer to the slant range and the azimuth angle of LiDAR beam, respectively. There are mainly two sectors that cover the airport approach/departure corridors including the touchdown zones, with the eastern sector (10°–150°) for tailwind of the aircraft and western sector (220°–340°) for headwind of the aircraft. Although the slant ranges of the LiDAR data are from 350 m to 10,000 m, the LiDAR data points with slant ranges over 4950 m are neglected since there are lots of missing Doppler velocities caused by the complex terrain. Hence, in this paper, we focus on the data collected from 10° to 150° and 220° to 340°, respectively, with slant ranges less than 4950 m.

3. Proposed Method

In this section, we will introduce the proposed indicators of windshear and the corresponding feature vectors. First, we will give a brief review of the physical property of windshear and its characteristics on LiDAR PPI scans. Then, we will introduce the proposed indicators and the corresponding feature vectors. Here, we use the LiDAR PPI data collected at HKIA as an example. It is interesting to note that the proposed methods can be widely applied to LiDAR PPI data obtained at any airport.

3.1. Properties of Windshear. Windshear can be caused by a wide variety of meteorological phenomena, such as winds blowing across terrain, sea breeze, microburst, gust front, and low-level jet. Generally, most windshear cases are terrain-induced. When the wind blows over rough terrain, wind speed and direction might change in the lee side of high obstacle, which may lead to alternating high-speed and low-speed air streams. Aircraft traversing through the alternating high-speed and low-speed air streams may experience a large headwind gain/loss that affects the aircraft lift. In addition to alternative high and low wind speed, the cross mountain flow can also cause some more complicated localized flows such as vortices or jumps, which could bring significant windshear to the aircraft. Due to the complicated characteristics of terrain-induced airflow disturbances, the corresponding windshear is transient and sporadic. This property can be properly illustrated by the LiDAR PPI scan of wind velocity for small-scaled areas that have reversed flow embedded in the background wind. Please refer to reference [1] and Figure 1 for more details. In this paper, we mainly derive the indicators based on the sustained change property of wind velocity and the transient and sporadic properties showed in the PPI scan images.

3.2. Proposed Indicators. Figure 1 shows two LiDAR PPI scans of Doppler velocities obtained at HKIA. There are mainly two sectors in each LiDAR scan which can cover the area around the whole corridors including the touch down zones. The sector with positive values denotes the Doppler velocities of wind blow away from the LiDAR which is actually the headwind encountered by aircraft, and the other sector denotes the Doppler velocities of tailwind. Here, we take both two sectors into account. To ensure the quality of velocity data, the data points obtained over a predetermined range value would be filtered out (4950 m for HKIA data). In references [20, 21], 2D wind field has been verified for windshear detection. However, it is generally derived quantity based on some assumptions of the nature of the flow, which may not be totally accurate in reality. On the other hand, using radial velocity, it is the direct measurement of the atmosphere and must be accurate. Hence, we mainly concentrate on the features corresponding to the radial velocity data obtained by LiDAR PPI scan in this paper. In addition to wind velocity, coherent Doppler LiDAR can retrieve multiple parameters, such as spectrum width, spectrum skewness, EDR, windshear intensity, and so on, which can also reflect the features of windshear. Nonetheless, future study may include 2D wind field and some other parameters retrieved by coherent Doppler LiDAR, to test the robustness of the algorithm and see if there is any improvement in windshear feature detection.

First, we will introduce several mathematical notations used in this paper.

For one LiDAR PPI scan, let $\mathbf{x}_i \in \mathbb{R}^n$ be the LiDAR data observed at azimuth angle θ_i. For simplicity, we assume that there are m_1 azimuth angles $(\theta_1^{(1)} < \theta_2^{(1)} < \cdots < \theta_{m_1}^{(1)})$ to be recorded in the tailwind sector and m_2 azimuth angles $(\theta_1^{(2)} < \theta_2^{(2)} < \cdots < \theta_{m_2}^{(2)})$ to be recorded in the headwind sector, and also, there are n range values to be recorded along fixed azimuth angle, i.e.,

$$\begin{aligned} \mathbf{x}_i^{(1)} &= \left[x_{i,1}^{(1)}, x_{i,2}^{(1)}, \ldots, x_{i,n}^{(1)}\right], 1 \le i \le m_1; \\ \mathbf{x}_i^{(2)} &= \left[x_{i,1}^{(2)}, x_{i,2}^{(2)}, \ldots, x_{i,n}^{(2)}\right], 1 \le i \le m_2. \end{aligned} \tag{1}$$

Note that the locations of such range values are not necessary to be uniform, and the distance between the LiDAR center and the observed value $x_{i,j}$ is equal to r_j.

3.2.1. Indicator Based on Wind Velocity Data. Since the main property of windshear is the change of wind speed and direction, we can evaluate the characteristics of windshear by the maximum variation of wind velocities along the LiDAR beam, which can reflect the wind velocity change in the investigated area. For fixed azimuth angle $\theta_i^{(1)}$ $(\theta_i^{(2)})$, the maximum variation of headwind (tailwind) velocities can be evaluated by indicator $\max_{1\le j\le n}(x_{i,j}^{(1)}) - \min_{1\le j\le n}(x_{i,j}^{(1)})$ $(\max_{1\le j\le n}(x_{i,j}^{(2)}) - \min_{1\le j\le n}(x_{i,j}^{(2)}))$. The higher the value is, the more likely windshear will occur. Note that we just take the difference of the velocity values instead of the difference of velocity magnitudes so that the sudden change of wind direction can also be evaluated by this indicator. More specifically, for the case where there are areas of reversed

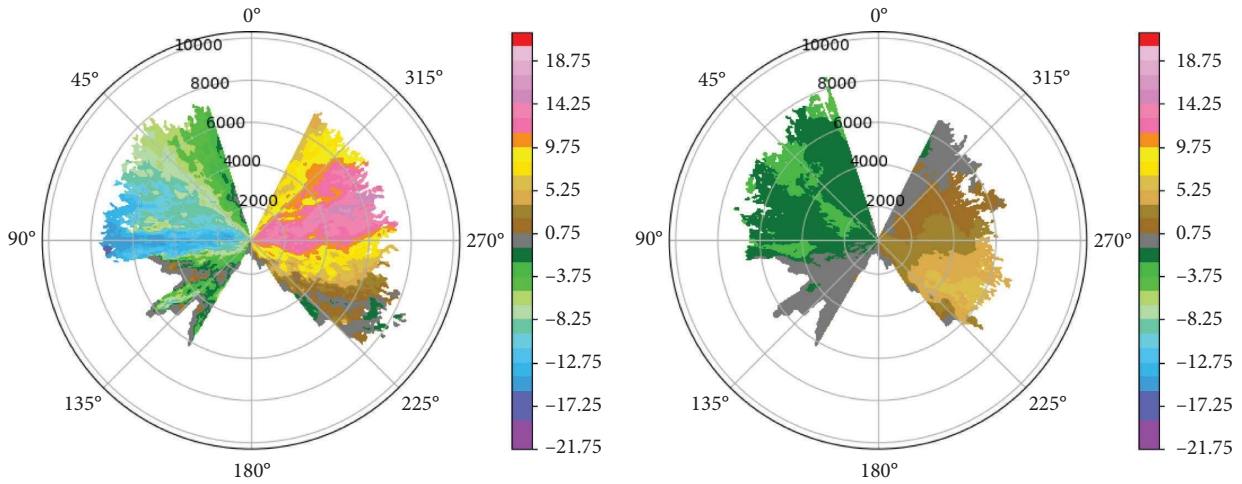

Figure 1: LiDAR radial velocity (knots) images collected at HKIA in PPI mode with 3° elevation and 350 − 10, 000 m slant ranges, where positive radial velocities (brown/yellow/orange/pink) are away from LiDAR and negative radial velocities (green/blue/purple) are toward the LiDAR. Each scan takes about 20–30 seconds. 400–500 beams are emitted to obtain one scan. Due to the complex terrain near HKIA, data with good quality are not available at a high slant range (blank sectors).

flow embedded in the background wind, the sign of Doppler wind velocity values would be different, which could make the difference between velocity values be large. Here, we extract the top k maximum variation values of each sector to represent the change of wind velocity in the corresponding sector (the determination of k will be illustrated in Section 3). Let $\mathbf{V}^{(1)} = \{\max_{1\le j\le n}(x_{i,j}^{(1)}) - \min_{1\le j\le n}(x_{i,j}^{(1)}), i = 1, \ldots, m_1\}$ denote the set of maximal velocity variation values obtained in the tailwind sector and $\mathbf{V}^{(2)} = \{\max_{1\le j\le n}(x_{i,j}^{(2)}) - \min_{1\le j\le n}(x_{i,j}^{(2)}), i = 1, \ldots, m_2\}$ denote the set of maximal velocity variation values obtained in the headwind sector, the associated indicator of each sector is calculated as follows:

$$\begin{aligned} \mathbf{f}_p^{(1)} &= \left[\mathbf{V}_1^{(1)}, \ldots, \mathbf{V}_k^{(1)}\right], \\ \mathbf{f}_p^{(2)} &= \left[\mathbf{V}_1^{(2)}, \ldots, \mathbf{V}_k^{(2)}\right], \end{aligned} \tag{2}$$

where $\mathbf{V}_i^{(1)}$ denotes the i-th maximum value in set $\mathbf{V}^{(1)}$ and $\mathbf{V}_i^{(2)}$ denotes the i-th maximum value in set $\mathbf{V}^{(2)}$. Since the windshear property mainly based on the great change of wind velocity, to find the largest change of wind profile, we select the one with larger L_2 norm in between $\mathbf{f}_p^{(1)}$ and $\mathbf{f}_p^{(2)}$ as the indicator $\mathbf{f}_p$ for the corresponding wind velocity data collected by one LiDAR PPI scan, i.e.,

$$\mathbf{f}_p = \begin{cases} \mathbf{f}_p^{(1)}, & \text{if } \left\|\mathbf{f}_p^{(1)}\right\|_2 \ge \left\|\mathbf{f}_p^{(2)}\right\|_2, \\ \mathbf{f}_p^{(2)}, & \text{otherwise.} \end{cases} \tag{3}$$

3.2.2. Indicator Based on the LiDAR PPI Scan Image of Wind Velocity. By the LiDAR PPI scan image of wind velocity (e.g., Figure 1), windshear can be detected visually based on the image texture. Generally, there would be more textures on the velocity images collected in windshear case. Hence, we propose to use the image texture extraction method to evaluate the property of windshear from the LiDAR PPI scan image of wind velocity.

Gray level co-occurrence matrix (GLCM) [22] is a kind of image texture feature extraction method, which is widely used in remote sensing, biometric issues, and pattern recognition. It mainly describes the relative frequencies of pixel pairs with gray-tone i and j, respectively, which can be separated by distance d under a specified angle occurring on the image. One can get several different co-occurrence matrices with different distances and angles. Each element p with position (i, j) in a co-occurrence matrix denotes the relative frequency where the pixel with grey level i is adjacent to a pixel with grey level j horizontally, vertically, or diagonally with a specified distance and angle.

$$p(i, j) = \frac{\#\{(p1, p2) \in I \mid p1 = i \,\&\, p2 = j\}}{\#I}, \tag{4}$$

where I denotes the image, $p1$, $p2$ denote two pixels, and $\#$ denotes count function. Here, we set the distance to be one and angles to be $0^\circ, 45^\circ, 90^\circ, 135^\circ$. Then, we can get four different co-occurrence matrices. There are several texture features, which are calculated by GLCM, such as energy, contrast, dissimilarity, entropy, correlation, homogeneity, and so on. Here, we calculate the contrast, dissimilarity, and correlation of the four co-occurrence matrices as the statistical indicator $\mathbf{f}_{im}$ for LiDAR PPI scans, whose equations are given as follows:

(i) Contrast:

$$\sum_{i,j=0}^{255} p(i, j) * (i - j)^2. \tag{5}$$

(ii) Dissimilarity:

$$\sum_{i,j=0}^{255} p(i, j) * |i - j|, \tag{6}$$

(iii) Correlation:

$$\sum_{i,j=0}^{255} p(i,j) * \left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right], \tag{7}$$

where μ_i, μ_j and σ_i, σ_j denote the mean values and standard deviation values, respectively.

3.3. Feature Vectors Based on the Proposed Indicators. From the proposed two indicators, we can readily get three feature vectors for windshear detection: $\mathbf{f}_1 = \mathbf{f}_p$, $\mathbf{f}_2 = \mathbf{f}_{im}$, and $\mathbf{f}_3 = [\mathbf{f}_p, \mathbf{f}_{im}]$.

Since the time stamps reported by pilots are not accurate, one cannot find the LiDAR scan where windshear occurs precisely. It is better for us to consider the wind velocity data obtained by LiDAR scans within four minutes near the windshear reported time stamp, which is also the same for non-windshear cases. Inspired by the idea of multiple instance learning in which several instances are arranged in sets (bags) and a label is provided for the entire bag, we consider treating the LiDAR scans collected in one time period as a bag and constructing one feature vector for this bag. Specifically, after extracting the feature vectors from the data collected by each single LiDAR scan in the corresponding period, we propose to select the one that can represent the most significant wind velocity variation (i.e., the one with the greatest L_2-norm) as the feature vector of the whole time period (bag).

3.3.1. Bag Feature Vector. Let $F = \{\mathbf{f}^1, \mathbf{f}^2, \ldots, \mathbf{f}^d\}$ be the set of feature vectors obtained in one time period, i.e., one bag. The feature vector of this bag is the one with the maximum L_2-norm: $\mathbf{f} = \mathbf{f}^j$ and $\|\mathbf{f}^j\|_2 \geq \|\mathbf{f}^i\|_2, \forall i = 1, 2, \ldots, d$.

4. Numerical Results and Discussion

In this section, we will do several numerical experiments to evaluate the effectiveness of the proposed feature vectors, i.e., their separability for windshear and non-windshear distinction. For comparison, the corresponding numerical results of feature vector proposed by Huang et al. [17] are also shown, which is also for windshear detection. First, we will introduce the dataset that we used for numerical experiments. Then, we will calculate the distances of the proposed feature vectors between the test windshear and non-windshear cases and then plot the histogram of these distances. Note that effective feature vectors that can be used for machine learning-based windshear detection methods should distinguish windshear cases and non-windshear cases properly by distance (dissimilarity), namely, the distance between features of same kind of cases should be small and the distance between the features of different kind of cases should be large. Finally, we apply several clustering methods to do data clustering based on the proposed feature vectors. Compared with the supervised classification methods, clustering methods only focus on the information given by the proposed feature vectors of the wind velocity data, which can adequately evaluate how good the proposed feature vectors are.

4.1. Dataset. The set of wind velocity data used in this study was collected at HKIA in the first three months from 2017 to 2019. All of them are preprocessed by the quality control method currently using at HKIA [1] to filter out the noisy velocities. Pilot reports are referred as the ground truth of windshear occurrence. The non-windshear cases are collected on days where there was no windshear. According to the pilot reports from 2017 to 2019, there were overall 369 windshear cases in the first three months. Correspondingly, 369 non-windshear cases are randomly selected on the days with no windshear under the suggestion from the Hong Kong Observatory. For each selected time stamp, LiDAR observational Doppler velocity data collected in 4 minutes with an elevation angle of 3° are used for feature vector construction. Note that there would be 3–6 LiDAR PPI scans collected in most of investigated periods, which makes it adequate for us to capture the statistical indicators proposed in this paper.

4.2. Histogram of Distances. In this part, we will show the histograms of distances between the proposed feature vectors. Three frequently used distance assessments are considered here. The equations of them are given as follows:

(i) Euclidean distance: $\|\mathbf{u} - \mathbf{v}\|_2$;

(ii) City block distance: $\|\mathbf{u} - \mathbf{v}\|_1$;

(iii) Bray-Curtis dissimilarity: $\sum_i |\mathbf{u}_i - \mathbf{v}_i| / \sum_i |\mathbf{u}_i + \mathbf{v}_i|$,

where $\mathbf{u}$ and $\mathbf{v}$ are vectors.

In each histogram, the corresponding distance values between arbitrary two feature vectors of the investigated cases are calculated and the frequencies of the corresponding distance values are shown by the heights of blue bar (y-label). There are three kinds of distance histograms, including the distance histogram of feature vectors between two windshear cases, the distance histogram of feature vectors between two non-windshear cases, and the distance histogram between one windshear case and one non-windshear case. For effective feature vectors, the frequencies of small distance values should be large in the histograms of the distances between feature vectors of two windshear/non-windshear and be small in the histograms of distances between feature vectors of one windshear case and one non-windshear case.

In the construction of indicator $\mathbf{f}_p$, there is a variable k which indicates the number of top maximum variation indicators we take into account. Here, we plot several distance histograms of this indicator with $k \in \{3, 5, 8, 10, 12, 15, 20\}$. Here, due to limit space, we only show the histogram with $k = 12$. The histograms for other cases are almost the same as it.

Figures 2–4 show the histograms of the distances corresponding to the feature vectors $\mathbf{f}_1 = \mathbf{f}_p$, $\mathbf{f}_2 = \mathbf{f}_{im}$, and $\mathbf{f}_3 = [\mathbf{f}_p, \mathbf{f}_{im}]$, respectively. Figure 5 shows the histogram of the distances corresponding to the feature vector proposed in

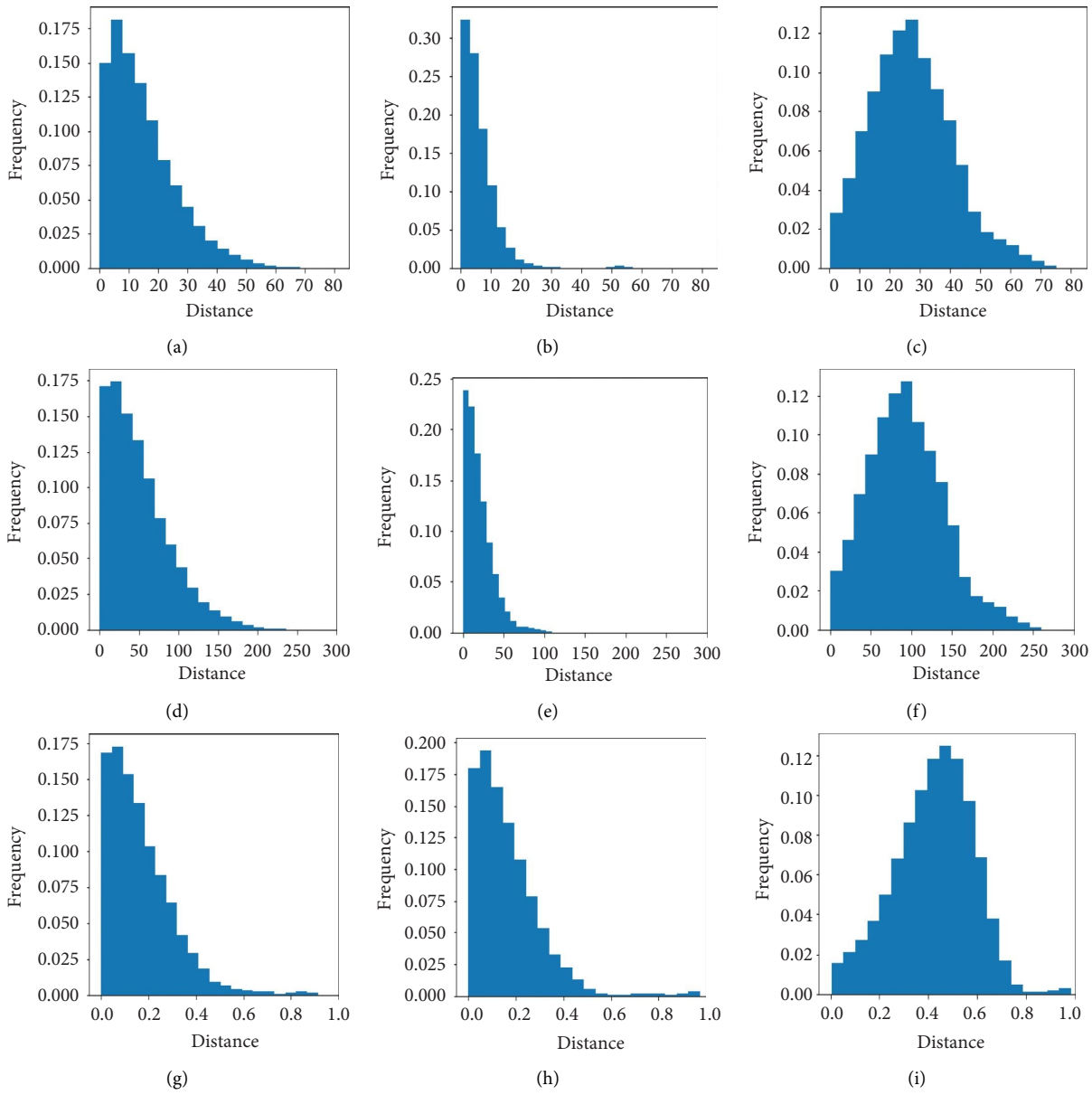

Figure 2: Histograms of different distances for feature vector $\mathbf{f}_1 = \mathbf{f}_p$ with $k = 12$, where x label denotes the distance between two feature vectors and y label denotes the frequency of the corresponding distance values in all the calculated distances. First column: distance histogram between windshear cases. Second column: distances between non-windshear cases. Third column: distances between windshear and non-windshear cases. Top row: histograms of the Euclidean distance. Second row: histograms of the city block distance. Bottom row: histograms of the Bray–Curtis distance.

reference [17] $\mathbf{f}_h$. From these results, one can readily find that most of distance values between feature vectors of two windshear/non-windshear cases are small and most of distance values between feature vectors of one windshear case and one non-windshear case are large. For the feature vector proposed in reference [17], this phenomenon is relatively trivial except for the result of the Bray–Curtis distance. Therefore, the proposed feature vectors are valid for learning-based windshear detection.

4.3. Clustering Results. To further evaluate the effectiveness of the proposed features, we apply four clustering methods to them, including the k-means clustering method, hierarchical clustering method, DBSCAN clustering method with the Euclidean distance, and DBSCAN clustering method with the city block distance. If the proposed features can be properly clustered into two categories where windshear cases or non-windshear cases are the majority, we can consider further using them in other learning-based

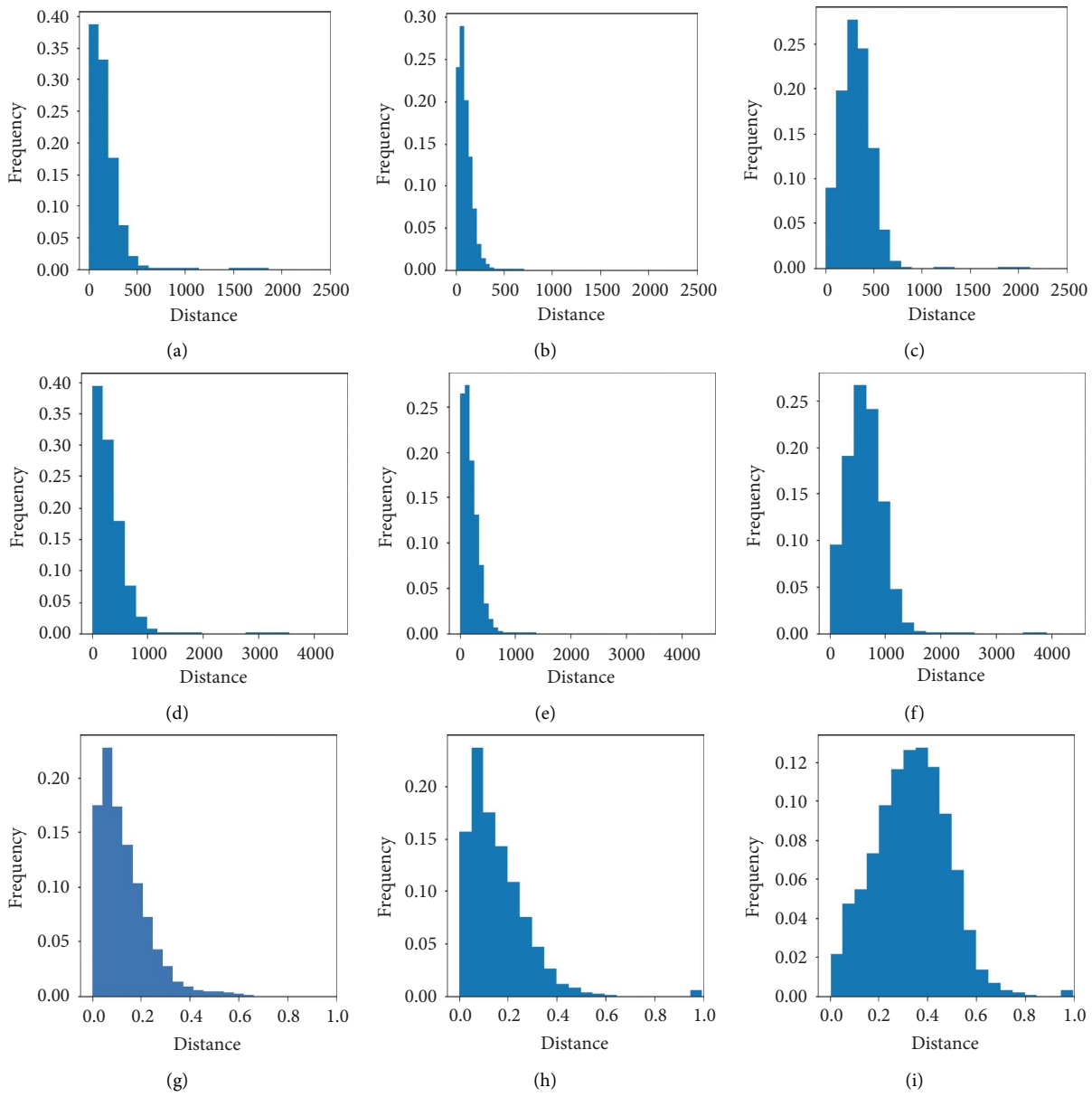

FIGURE 3: Histograms of different distances for proposed feature vector $\mathbf{f}_2 = \mathbf{f}_{im}$, where x label denotes the distance between two feature vectors and y label denotes the frequency of the corresponding distance values in all the calculated distances. First column: distances between windshear cases. Second column: distances between non-windshear cases. Third column: distances between windshear and non-windshear cases. Top row: histograms of the Euclidean distance. Second row: histograms of the city block distance. Bottom row: histograms of the Bray–Curtis distance.

windshear detection methods. Since we clearly know whether a feature vector corresponds to a windshear case or not, it is easy for us to calculate the accuracy of the clustering methods and evaluate the effects of the proposed feature vectors.

First, we need to determine the best value of k for the physical property based indicator $\mathbf{f}_p$. Here, we apply the clustering methods to feature vector $\mathbf{f}_1 = \mathbf{f}_p$ obtained with $k \in \{3, 5, 8, 10, 12, 15, 20\}$ and calculate the average accuracy. Figure 6 shows the average accuracy curve corresponding to different k. From Figure 6, we can find that the best average accuracy will be reached when $k = 12$. In the following section, we will set $k = 12$ for the indicator $\mathbf{f}_p$.

Tables 1–3 show the accuracy of these clustering methods for the proposed feature vectors. The clustering results for the feature vector proposed by Huang et al. are shown in Table 4 for comparison. To further show the clustering results intuitively, the scatter plots of them are shown in Figures 7–10.

From these results, we can readily find the following: (i) the physical property-based feature vector $\mathbf{f}_1$ has the best clustering results for most clustering methods except the k-means clustering method. (ii) The clustering accuracy of image-based feature vector $\mathbf{f}_2$ is greater than 91% for most clustering methods, but the hierarchical clustering method does not work well on this feature vector whose accuracy is only 74.12%. In

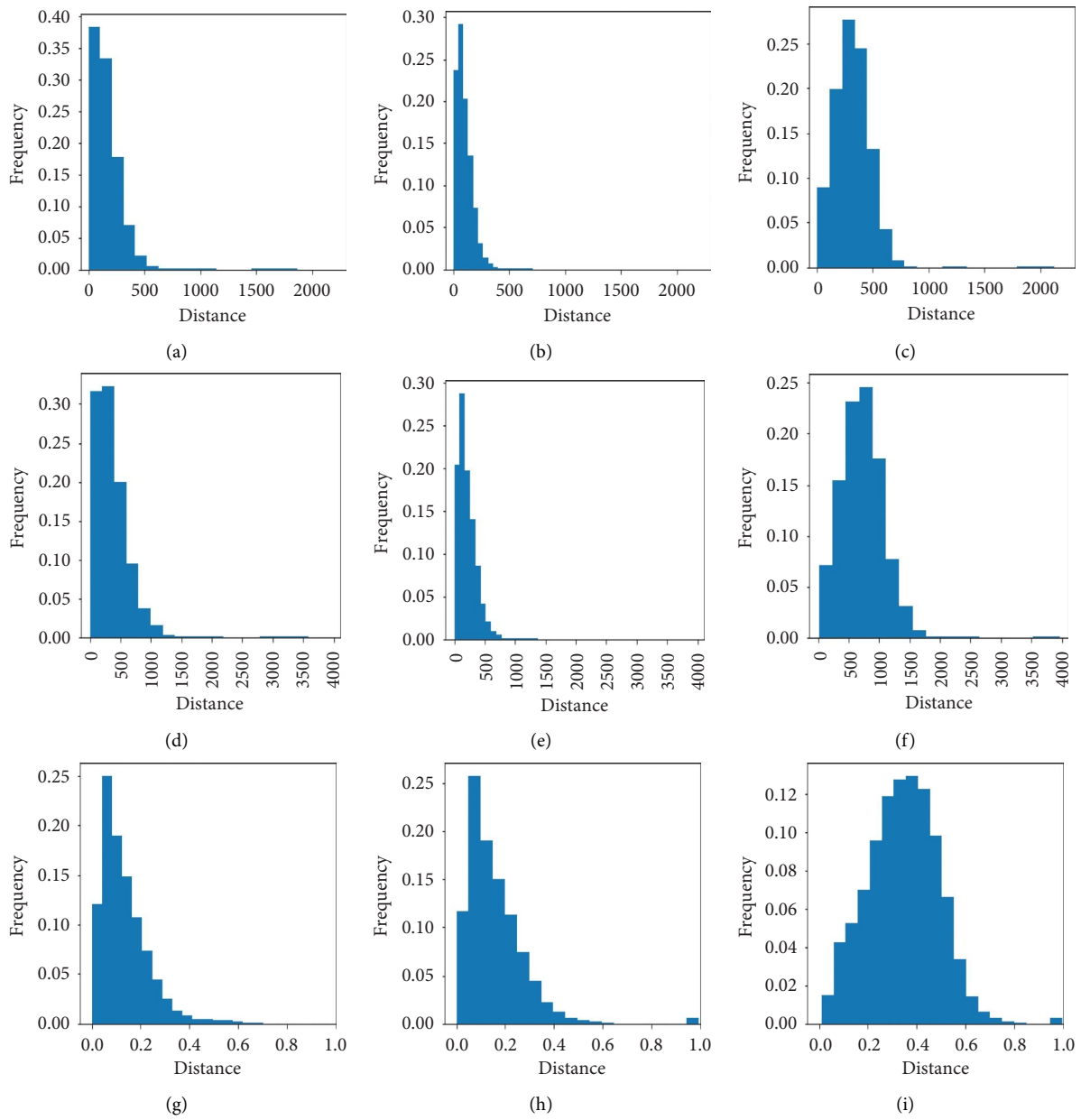

FIGURE 4: Histograms of different distances for proposed feature vector $\mathbf{f}_3 = [\mathbf{f}_p, \mathbf{f}_{im}]$, where x label denotes the distance between two feature vectors and y label denotes the frequency of the corresponding distance values in all the calculated distances. First column: distances between windshear cases. Second column: distances between non-windshear cases. Third column: distances between windshear and non-windshear cases. Top row: histograms of the Euclidean distance. Second row: histograms of the city block distance. Bottom row: histograms of the Bray–Curtis distance.

practice, the results of classification method based only on this feature vector might be bad. (iii) For the combining feature vector $\mathbf{f}_3$, the clustering accuracy are all greater than 90% which is quite robust for different clustering methods. (iv) Compared with the clustering results of feature vectors proposed by Huang et al., our proposed feature vectors can get much higher accuracy. (v) From the scattering plots of the clustering results, we can find that most of the points can be clustered correctly by the k-means clustering method since the densities of both blue and green points are almost the same. Both feature vector $\mathbf{f}_1$ and the combine feature vector $\mathbf{f}_3$ perform good in most of clustering experiments. Therefore, we can consider to apply feature vector $\mathbf{f}_1$ and the combined feature vector $\mathbf{f}_3$ in the learning-based windshear detection methods in the future.

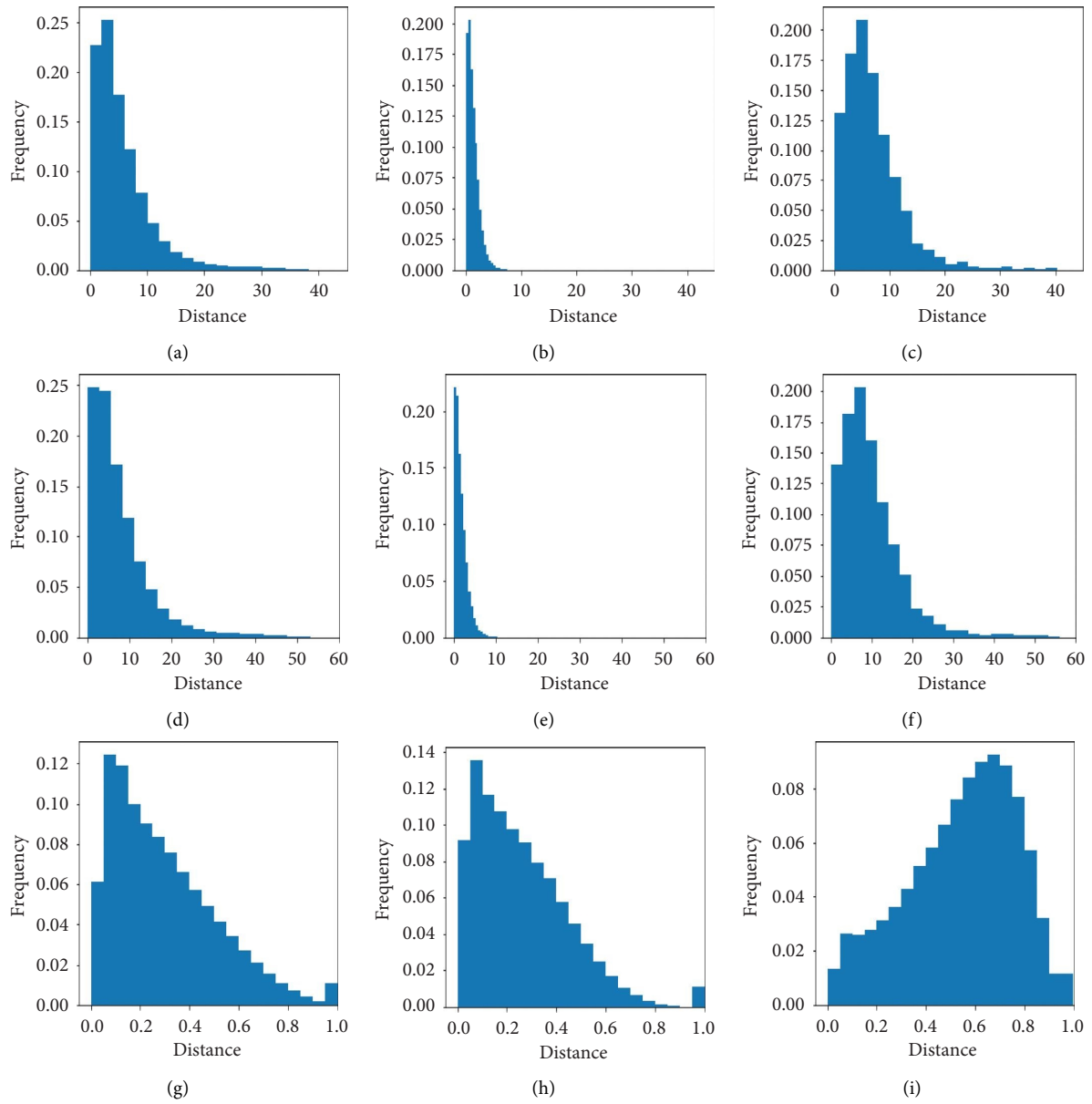

Figure 5: Histograms of different distances for the indicator proposed by Huang et al. $\mathbf{f}_h$, where x-label denotes the distance between two feature vectors, and y-label denotes the frequency of the corresponding distance values in all the calculated distances. First column: distances between windshear cases. Second column: distances between non-windshear cases. Third column: distances between windshear and non-windshear cases. Top row: histograms of the Euclidean distance. Second row: histograms of the city block distance. Bottom row: histograms of the Bray–Curtis distance.

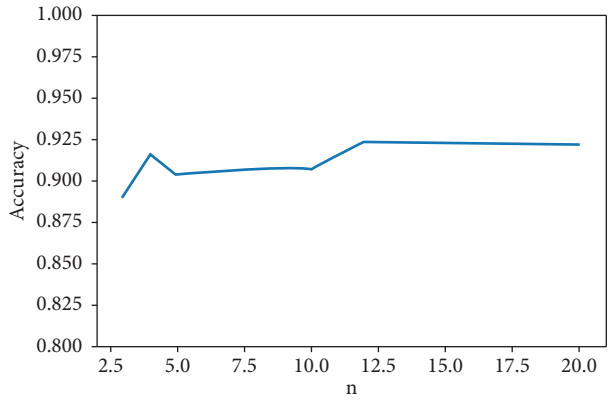

Figure 6: The average accuracy curve for feature vector $\mathbf{f}_1$ with $k \in \{3, 5, 8, 10, 12, 15, 20\}$.

Table 1: Clustering results of different methods for feature vector $\mathbf{f}_1 = \mathbf{f}_p$ ($k = 12$).

Clustering method	Clustering result	Ground truth windshear	Ground truth non-windshear	Accuracy (%)
K-means method	Windshear	292	3	89.16
	Non-windshear	77	366	
Hierarchical clustering	Windshear	339	17	93.63
	Non-windshear	30	352	
DBSCAN clustering (Euclidean distance)	Windshear	345	27	93.09
	Non-windshear	24	342	
DBSCAN clustering (city block distance)	Windshear	349	29	93.36
	Non-windshear	20	340	

Table 2: Clustering results of different methods for feature vector $\mathbf{f}_2 = \mathbf{f}_{im}$.

Clustering method	Clustering result	Ground truth windshear	Ground truth non-windshear	Accuracy (%)
K-means method	Windshear	313	9	91.19
	Non-windshear	56	360	
Hierarchical clustering	Windshear	190	1	74.12
	Non-windshear	179	368	
DBSCAN clustering (Euclidean distance)	Windshear	349	40	91.87
	Non-windshear	20	329	
DBSCAN clustering (city block distance)	Windshear	351	44	91.60
	Non-windshear	18	325	

Table 3: Clustering results of different methods for feature vector $\mathbf{f}_3 = [\mathbf{f}_p, \mathbf{f}_{im}]$.

Clustering method	Clustering result	Ground truth windshear	Ground truth non-windshear	Accuracy (%)
K-means method	Windshear	300	3	90.24
	Non-windshear	69	366	
Hierarchical clustering	Windshear	339	19	93.36
	Non-windshear	30	350	
DBSCAN clustering (Euclidean Distance)	Windshear	349	42	91.60
	Non-windshear	20	327	
DBSCAN clustering (city block distance)	Windshear	350	32	93.09
	Non-windshear	19	337	

Table 4: Clustering results of different methods for feature vector proposed by Huang et al.

Clustering method	Clustering result	Ground truth windshear	Ground truth non-windshear	Accuracy (%)
K-means method	Windshear	227	1	80.62
	Non-windshear	142	368	
Hierarchical clustering	Windshear	76	0	60.30
	Non-windshear	293	369	
DBSCAN clustering (Euclidean distance)	Windshear	316	24	89.57
	Non-windshear	53	345	
DBSCAN clustering (city block distance)	Windshear	312	16	90.11
	Non-windshear	57	353	

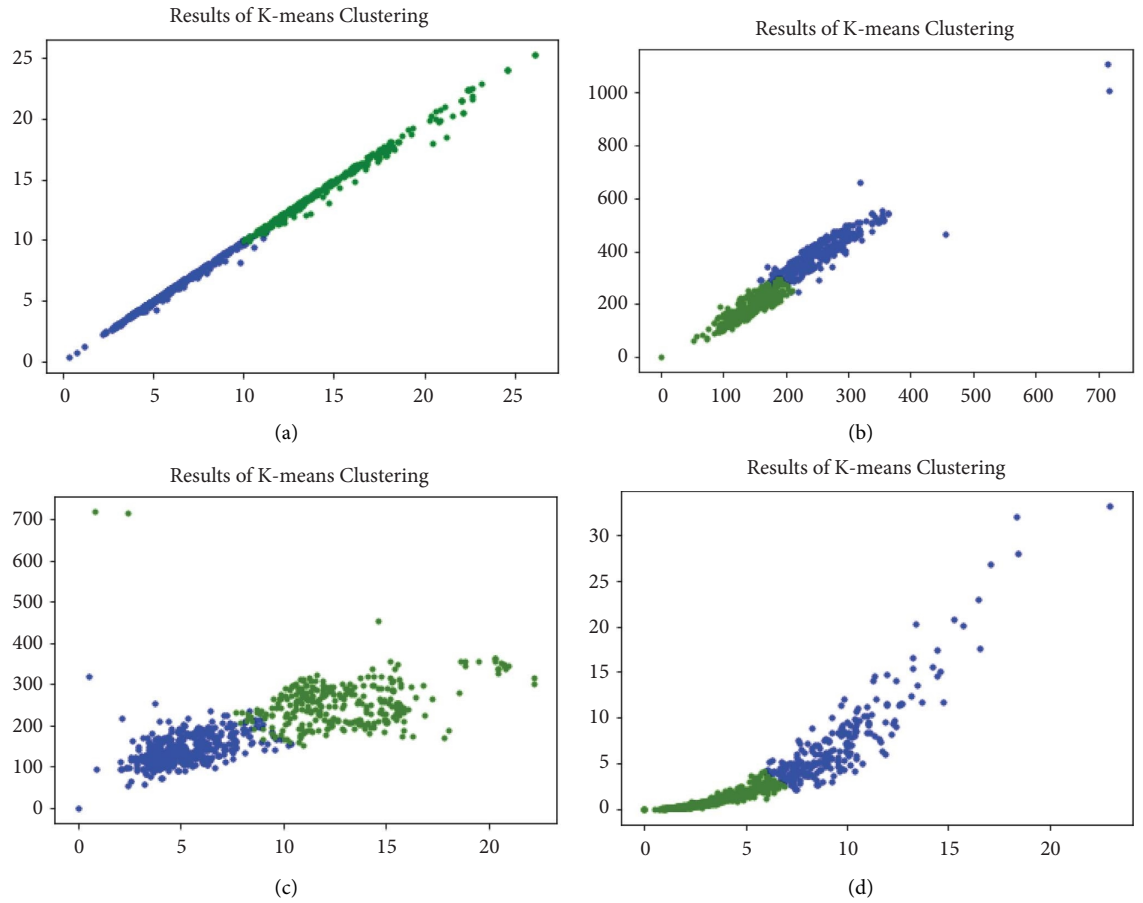

Figure 7: Clustering results of the k-means clustering method. (a) Feature vector f_1, (b) feature vector f_2, (c) feature vector f_3, and (d) feature vector f_h.

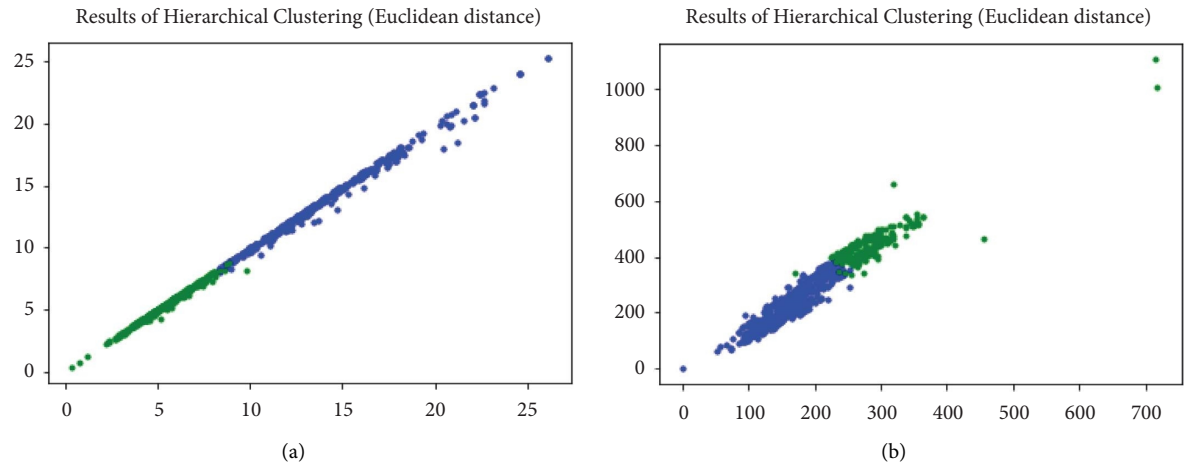

Figure 8: Continued.

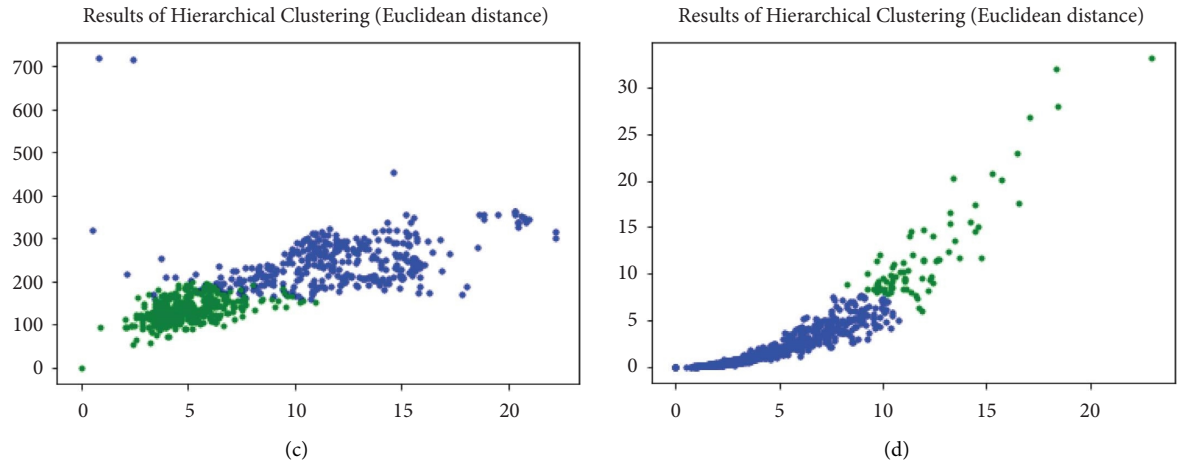

Figure 8: Clustering results of the hierarchical clustering method. (a) Feature vector f_1, (b) feature vector f_2, (c) feature vector f_3, and (d) feature vector f_h.

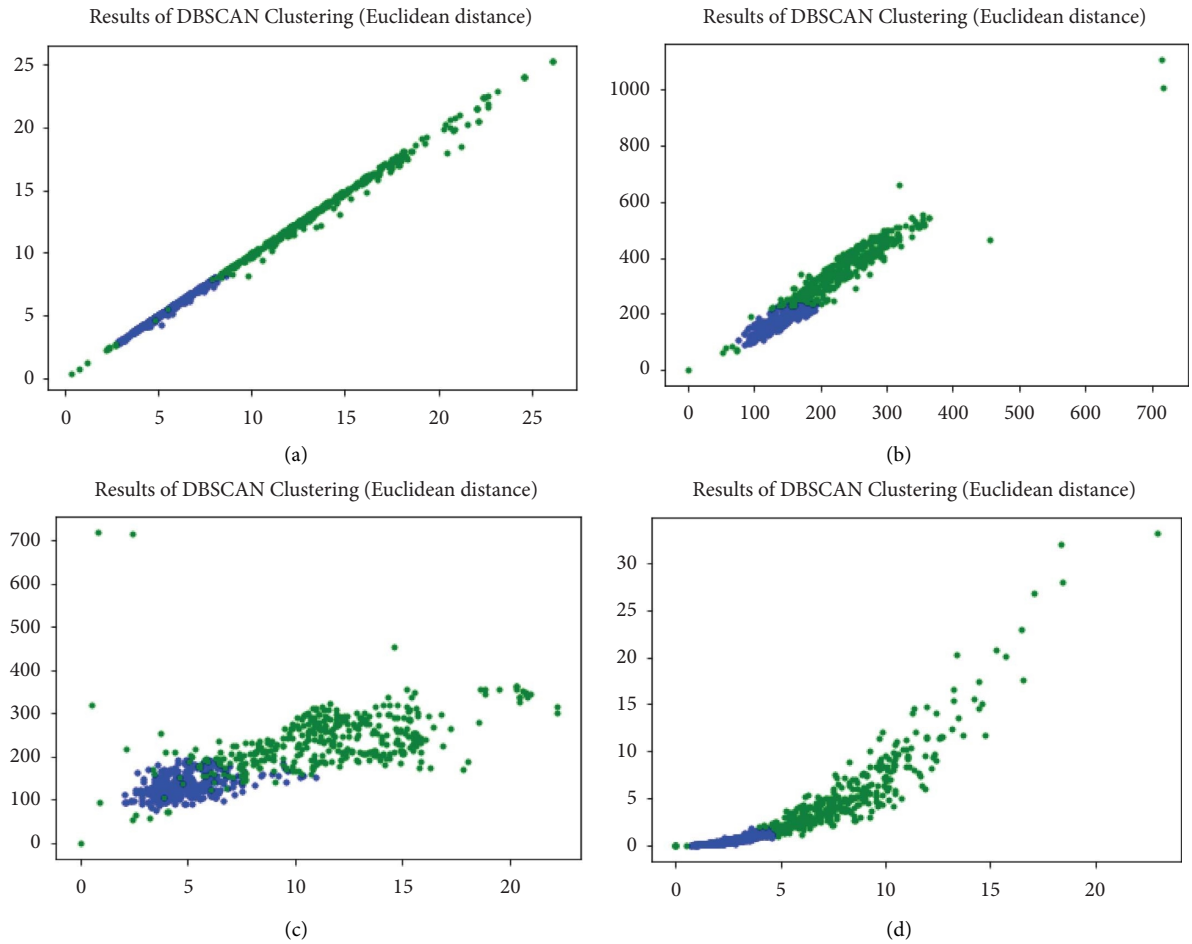

Figure 9: Clustering results of DBSCAN clustering (Euclidean distance). (a) Feature vector f_1, (b) feature vector f_2, (c) feature vector f_3, and (d) feature vector f_h.

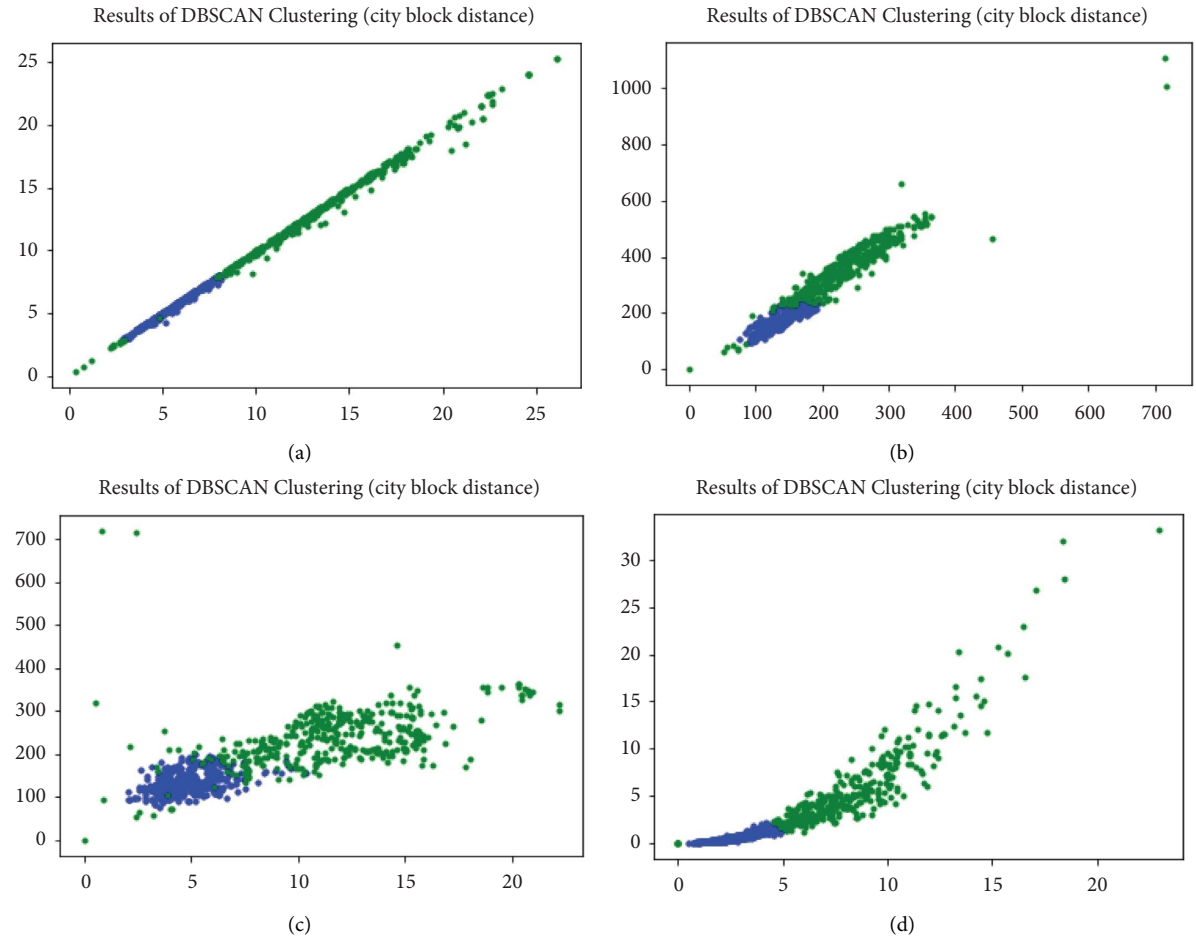

FIGURE 10: Clustering results of DBSCAN clustering (city block distance). (a) Feature vector f_1, (b) feature vector f_2, (c) feature vector f_3, and (d) feature vector f_h.

5. Conclusion

In this paper, we propose two statistical indicators of windshear from the LiDAR PPI scan observational wind velocity data for windshear features construction. Both indicators based on the physical properties of windshear and the image processing method are researched in this paper. To evaluate the effectiveness of the proposed indicators, we construct 3 feature vectors based on them. Different from the feature vectors used in previous learning-based windshear detection methods, which only takes the wind profiles obtained at the nearest time spot to the windshear reported timestamp, we consider the wind profiles collected within 4 minutes of the reported timestamp. Based upon the idea of multiple instance learning, a feature vector of the scans in one bag (time period) is constructed. The distance histograms for three frequently used distances are given to preliminarily validate the separability induced by the proposed feature vectors for windshear and non-windshear classification. Then, we apply four clustering methods to further evaluate the effects of the proposed feature vectors by the accuracy of these methods and the corresponding scatter plots. The corresponding numerical results of the feature vector proposed by Huang et al. are also shown in this paper for comparison. Numerical results show the following: (i) by the histograms of the distances between the test cases, we can readily find that most of distance values between feature vectors of two windshear/non-windshear cases are small and most of distance values between feature vectors of one windshear case and one non-windshear case are large, which indicate that one can distinguish windshear and non-windshear properly by distances of the proposed feature vectors. (ii) Feature vector based on the physical property performs better than other feature vectors for most clustering methods. (iii) One can get pretty high clustering accuracy by using the feature vector extracted by the image processing method except the hierarchical clustering method. (iv) The combined feature vector is quite stable to different clustering methods whose accuracy are all greater than 90%. (v) Our proposed feature vectors perform better than the feature vectors proposed by Huang et al. [17]. In the future, one can consider to use the proposed physical property based feature vector or the combined feature vector in other learning-based windshear and non-windshear classification to achieve better detection results. Moreover, we can also explore using some image processing methods

based on deep learning to extract the feature vector of windshear by the wind velocity LiDAR scans.

Data Availability

The request for data used to support the findings of this study could be addressed to the Hong Kong Observatory. The data provision would be considered on a case-by-case basis.

Conflicts of Interest

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

Acknowledgments

M. Ng's research was supported in part by Hong Kong Research Grant Council GRF 12300519, 17201020, 17300021 C1013-21GF, C7004-21GF, and Joint NSFC-RGC N-HKU76921.

References

[1] P. W. Chan, C. M. Shun, and K. C. Wu, "Operational lidar-based system for automatic windshear alerting at the Hong Kong international airport," *12th Conference on Aviation, Range, and Aerospace Meteorology*, vol. 6, 2006.

[2] P. W. Li, *Windshear–its detection and altering*, EP0125087A2 european patent office, Berlin, Vienna and Brussels, 2019.

[3] L. Thobois, J. P. Cariou, and I. Gultepe, "Review of lidar-based applications for aviation weather," *Pure and Applied Geophysics*, vol. 176, no. 5, pp. 1959–1976, 2019.

[4] International Civil Aviation Organization (Icao), *Manual on Low-Level Wind Shear*, ICAO, Quebec, Canada, 2005.

[5] C. M. Shun and P. W. Chan, "Applications of an infrared Doppler lidar in detection of wind shear," *Journal of Atmospheric and Oceanic Technology*, vol. 25, no. 5, pp. 637–655, 2008.

[6] A. Weipert, S. Kauczok, R. Hannesen, T. Ernsdorf, and B. Stiller, "Wind shear detection using radar and lidar at frankfurt and munich airports," in *Proceedings of the 8th European Conference on Radar in Meteorology and Hydrology*, pp. 1–5, Garmisch-Partenkirchen, Germany, September 2014.

[7] M. Li, J. Xu, X. Xiong, Y. Ma, and Y. Zhao, "A novel ramp method based on improved smoothing algorithm and second recognition for windshear detection using lidar," *Curr. Opt. Photon,* vol. 2, no. 1, pp. 7–14, 2018.

[8] J. G. Jones and A. Haynes, *A Peakspotter Program Applied to the Analysis of Increments in Turbulence Velocity*, RAE, Richmond, Virginia, 1984.

[9] A. A. Woodfield and J. F. Woods, *Worldwide Experience of Wind Shear during 1981-1982*, Royal Aircraft EstablIshment, Bedford, England, 1983.

[10] K. K. Hon and P. W. Chan, "Improving lidar windshear detection efficiency by removal of "gentle ramps"," *Atmosphere*, vol. 12, no. 11, p. 1539, 2021.

[11] P. W. Chan, K. K. Hon, and D. K. Shin, "Combined use of headwind ramps and gradients based on lidar data in the alerting of low-level windshear/turbulence," *Meteorologische Zeitschrift*, vol. 20, no. 6, pp. 661–670, 2011.

[12] P. W. Chan, "Application of lidar-based f-factor in windshear alerting," *Meteorologische Zeitschrift*, vol. 21, no. 2, pp. 193–204, 2012.

[13] Y. F. Lee and P. W. Chan, "Lidar-based f-factor for wind shear alerting: different smoothing algorithms and application to departing flights," *Meteorological Applications*, vol. 21, no. 1, pp. 86–93, 2014.

[14] T. C. Wu and K. K. Hon, "Application of spectral decomposition of lidar-based headwind profiles in windshear detection at the Hong Kong international airport," *Meteorologische Zeitschrift*, vol. 27, no. 1, pp. 33–42, 2018.

[15] L. Li, A. Shao, K. Zhang, N. Ding, and P. W. Chan, "Low-level wind shear characteristics and lidar-based alerting at lanzhou zhongchuan international airport, China," *Journal of Meteorological Research*, vol. 34, no. 3, pp. 633–645, 2020.

[16] Y. Ma, S. Li, and W. Lu, "Recognition of partial scanning low-level wind shear based on support vector machine," *Advances in Mechanical Engineering*, vol. 10, no. 1, Article ID 168781401775415, 2018.

[17] J. Huang, M. K. P. Ng, and P. W. Chan, "Wind shear prediction from light detection and ranging data using machine learning methods," *Atmosphere*, vol. 12, no. 5, p. 644, 2021.

[18] J. Amores, "Multiple instance classification: review, taxonomy and comparative study," *Artificial Intelligence*, vol. 201, pp. 81–105, 201.

[19] P. W. Chan and C. M. Shun, "Application of a ground-based Doppler LIDAR to automatic windshear alerting," *Journal of Atmospheric and Oceanic Technology*, vol. 25, no. 5, 2008.

[20] P. W. Chan and A. M. Shao, "Depiction of complex airflow near Hong Kong international airport using a Doppler lidar with a two-dimensional wind retrieval technique," *Meteorologische Zeitschrift*, vol. 16, no. 5, pp. 491–504, 2007.

[21] J. Yuan, L. Su, H. Xia et al., "Microburst, windshear, gust front, and vortex detection in mega airport using a single coherent Doppler wind lidar," *Remote Sensing*, vol. 14, no. 7, p. 1626, 2022.

[22] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, vol. 6, pp. 610–621, 1973.