

Article

Rapeseed Seed Coat Color Classification Based on the Visibility Graph Algorithm and Hyperspectral Technique

Chaojun Zou ^{1,2}, Xinghui Zhu ^{1,*}, Fang Wang ^{3,*}, Jinran Wu ⁴ and You-Gan Wang ⁵

¹ College of Information and Intelligent Science and Technology, Hunan Agricultural University, Changsha 410125, China; zouchaojun@stu.hunau.edu.cn

² School of Economics and Management, Changsha Normal University, Changsha 410125, China

³ Key Laboratory of Intelligent Computing and Information Processing of Ministry of Education and Hunan Key Laboratory for Computation and Simulation in Science and Engineering, Xiangtan University, Xiangtan 411105, China

⁴ Institute for Positive Psychology and Education, Australian Catholic University, Banyo, Brisbane 4014, Australia; ryan.wu@acu.edu.au

⁵ School of Mathematics and Physics, The University of Queensland, St Lucia, Brisbane 4067, Australia; you-gan.wang@uq.edu.au

* Correspondence: zhuxh@hunau.edu.cn (X.Z.); fwang4@xtu.edu.cn (F.W.)

Abstract: Information technology and statistical modeling have made significant contributions to smart agriculture. Machine vision and hyperspectral technologies, with their non-destructive and real-time capabilities, have been extensively utilized in the non-destructive diagnosis and quality monitoring of crops and seeds, becoming essential tools in traditional agriculture. This work applies these techniques to address the color classification of rapeseed, which is of great significance in the field of rapeseed growth diagnosis research. To bridge the gap between machine vision and hyperspectral technology, a framework is developed that includes seed color calibration, spectral feature extraction and fusion, and the recognition modeling of three seed colors using four machine learning methods. Three categories of rapeseed coat colors are calibrated based on visual perception and vector-square distance methods. A fast-weighted visibility graph method is employed to map the spectral reflectance sequences to complex networks, and five global network attributes are extracted to fuse the full-band reflectance as model input. The experimental results demonstrate that the classification recognition rate of the fused feature reaches 0.943 under the XGBoost model, confirming the effectiveness of the network features as a complement to the spectral reflectance. The high recognition accuracy and simple operation process of the framework support the further application of hyperspectral technology to analyze the quality of rapeseed.

Keywords: hyperspectral reflectance; visibility graph algorithm; vector-square distance; color classification; machine learning



Citation: Zou, C.; Zhu, X.; Wang, F.; Wu, J.; Wang, Y.-G. Rapeseed Seed Coat Color Classification Based on the Visibility Graph Algorithm and Hyperspectral Technique. *Agronomy* **2024**, *14*, 941. <https://doi.org/10.3390/agronomy14050941>

Academic Editor: Giovanni Cabassi

Received: 27 March 2024

Revised: 22 April 2024

Accepted: 26 April 2024

Published: 30 April 2024



Copyright: © 2024 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

Rape (*Brassica napus* L.) is a significant and widely grown oilseed crop whose seeds, known as rapeseed, are rich in fatty acids and proteins and serve as the raw material not only of edible vegetable oils but also of premium-grade proteins for animal feed. In general, yellow rapeseed has a thinner seed coat, a lower percentage of hulls, and a larger embryo compared to other colors. Research indicates that rapeseeds with a more yellowish and lighter color tend to have higher oil content, as well as higher levels of oleic and linoleic acids, while also exhibiting higher protein content and lower fiber content [1,2]. Since the color of the rapeseed coat is controlled by its corresponding gene, it has become an important trait for breeding. Yellow-seeded *Brassica napus* is a hybridization of other varieties [3,4]. Therefore, it is of great value economically and very meaningful for breeding to identify rapeseed coat color and screen lighter yellow seeds.

Some researchers have explored various methods for the color recognition of rapeseed seeds. Usually, there are several types: (1) Artificial visual recognition. The classification of rapeseed seed coat color through manual observation or a magnifying glass. (2) The use of instruments, including colorimeters, color analyzers, and other instruments to measure and analyze colors [1]. (3) Spectral analysis methods. Using equipment such as spectrophotometers, near-infrared spectroscopy, and Fourier transform infrared photoacoustic spectroscopy to obtain spectral reflectance information, based on the principle of color formation and the relationship between spectra, spectral data are used to analyze or identify colors [5–7]. (4) Digital image analysis methods. Obtain images using a digital camera or scanner, and use computer image processing software or other specialized software for recognition [8,9]. In addition, yellow seeds are identified with the help of chemical solutions [10]. Among them, visual methods are simple and convenient but rely on experience, lack unified standards, and have significant differences. The objective results with a certain degree of accuracy using optical and spectral analysis techniques have low accuracy in color segmentation due to resolution reasons. The popularity of digital devices has made the digital image analysis of rapeseed color an efficient and non-destructive method. Hyperspectral imaging technology can simultaneously obtain spectral information from multiple bands, providing richer spectral information (including images and data) and characterizing the inherent features of substances. Due to its high resolution and comprehensive spectrum, it has achieved successful applications in many fields such as agriculture, geology, and environment [11,12], for example, the identification of rapeseed varieties and the prediction of oleic acid, protein, and other components by hyperspectral [13,14].

Different species have specific reflectance absorption characteristics at different wavelengths. In traditional hyperspectral analysis techniques, spectral indices derived from the combination of spectral values in one or more bands are often used as research variables. In studying plant traits, some spectral vegetation indices have been proposed, such as NDVI, RVI, and DVI, and have been widely applied in remote sensing fields such as soil and water [15]. At the same time, some scholars have constructed spectral indices related to color, including the anthocyanin index (ARI) [16], carotenoid reflectance index (CRI) [17,18], red to green ratio index (RGRI) [19], chlorophyll absorption index (CARI) [20], photochemical vegetation index (PRI) [21,22], and improved chlorophyll absorption ratio index (MCARI) [23,24]. Although these studies have achieved certain results, extracting single or multiple spectral indices as parameter variables can result in the loss of some information from hyperspectral data, thereby limiting their effectiveness in identifying or inverting species varieties or traits. High-dimensional spectra can be viewed as sequences or curves, whose features will be hidden in their topological structures [25]. A complex network is a powerful tool for studying sequence topology relationships, by mining the inherent properties between objects and the entire system. Given the advantages of this technology in mining spectral reflectance sequences, in this work, we attempt to explore the topological feature of hyperspectral reflectance from the network angle and build a bridge between RGB-based machine vision systems and hyperspectral signals to challenge the rapeseed coat classification, which is a difficult task in traditional rapeseed growth diagnose, as mentioned above.

In practice, our goal is to achieve the following objectives:

- Using vector and angular distance formulations for color difference calculations in the RGB color space to achieve color calibration.
- Constructing complex networks of spectral reflectance by fast-weighted visibility graph algorithm.
- Establishing an intelligent model for rapeseed seed coat color classification by combining hyperspectral technology and machine vision systems.

By utilizing complex networks and machine learning methods, the precise automatic recognition of rapeseed seed coat color can be achieved, which provides a scientific basis for further exploring the close relationship between rapeseed color and other physical

properties, as well as oil content, fatty acid composition, breeding research, and industrial quality optimization.

The rest of the paper is structured as follows: in Section 2, we describe the material preparation, data acquisition, and methodology. Section 3 presents the experimental results. We give discussions and conclusions in Sections 4 and 5, respectively.

2. Materials and Methods

2.1. Data Acquisition and Preprocessing

Two original varieties of *Brassica napus* seeds are employed in our study, namely, Xiangyou 708 and Xiangyou 710, of high oleic acid rapeseed. They were cultivated in the paddy fields of Yunyuan Experimental Base (28°23' N, 112°93' E) at Hunan Agricultural University, Changsha City, Hunan Province in September 2020 and harvested in April 2021. The two varieties, having been cultivated for numerous years, which are widely promoted in the Hunan province of China, serve as representative examples in our study. The SOC710 portable hyperspectral imager (Spectral range: 380~1091 nm; Resolution: 4.9458 nm; Spectral channels: 128; Sensor material: CCD; Spatial pixels: 520; Spatial resolution (Avg. RMS spot radius): <40 microns; Manufacturer: Surface Optics Corporation of the United States) and its darkroom system are used to collect the rapeseed hyperspectral data. The sample collection process strictly follows the spectral collection specifications. Rapeseeds are placed flat (see Figure 1) in a transparent petri dish in a dark room with a light source. The hyperspectral imager is placed in a small task box 370 mm vertically above the rapeseed samples. On each seed's surface, the SOC710 hyperspectral imager is used to measure the spectral reflectance five times randomly in some regions of interest (ROIs) and followed by an arithmetic averaging operation, which is as the reflectance of the sample. Meanwhile, the image color information of the three primary colors (red, green, and blue) of the samples is collected in the same ROIs. We finally obtained 282 samples in all, including 282 hyperspectral reflectance data and the corresponding RGB values.

Since the noise is always hidden at the end of bands, we remove the two ends of bands to obtain the reflectance from 400 nm to 1000 nm. Then, we perform linear interpolation with a resampling interval of 1 nm, and thus 601 reflectance values are obtained for each sample. The reflectance values characterized the physical (e.g., seed coat color) and chemical (e.g., content of components such as fatty acids, proteins) traits of the different samples of rapeseeds. The spectral curves are smoothed and filtered using the convolutional (Savitzky–Golay) smoothing method with noise reduction. Finally, the spectral data are normalized.



Figure 1. Rapeseed samples.

2.2. Framework of the Classification Model

The traits of rapeseed offspring are often reflected in their seed coat color; therefore, identifying and classifying seed coat color will help screen high-quality offspring for breeding and cultivation. For example, the genetic improvement of yellow-seeded rapeseed has garnered more attention than other color variations within the spectrum of rapeseed due to its favorable characteristics such as reduced lignin content, elevated oil content, and increased protein content.

As stated above, in traditional agriculture, relying on human eye recognition is inefficient and has a high error rate, which motivates us to develop an intelligent approach to fulfill this task. In this work, modern tools, machine vision technology, and hyperspectral technology are employed for our consideration. We aim to establish a bridge between them. Our model framework is shown in Figure 2. In the following, we present the color

calibration process by machine vision system (Section 2.3). Then, we introduce a tool that prevails in the field of time series analysis, the visibility graph algorithm, which is used to extract hyperspectral intrinsic features (Section 2.4). Next, the four classifiers are introduced briefly in Section 2.5, which will be used to execute the classification task. Finally, four indicators are given in the Section 2.6 to evaluate the performance of the models.

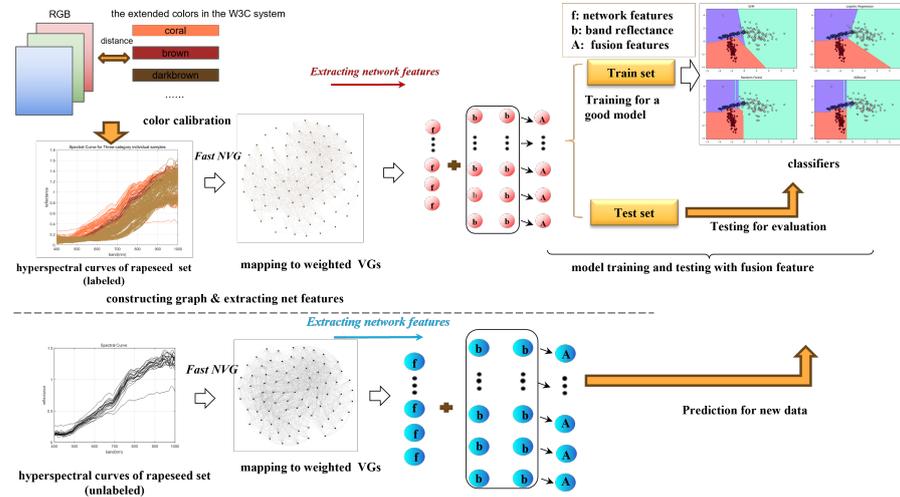


Figure 2. Framework of recognition model based on the hyperspectral technology and complex network analysis. The abbreviation “VG” means visible graph algorithm, and “W3C” refers to World Wide Web Consortium.

2.3. Color Calibration

Color calibration is necessary for the task of seed color separation. It is often performed by colorimeters, color cards, artificial vision, and machine vision systems. Ensuring definite standards and processes is key to accurately comparing the color of rapeseed from different batches. The RGB color space (which refers to red, green, and blue, respectively) is the most commonly used color system in machine vision. In what follows, we introduce vector-angular distance to fulfill the task of color calibration.

2.3.1. Vector-Angular Distance (VAD)

The Vector-Angular Distance (VAD) color difference equation is a method used to quantify the difference between two colors in the RGB color space [26–28]. The VAD equation [29] takes into account both the magnitude and the angular information of the color vectors, as defined by

$$dist = \sqrt{\frac{S_r^2 \cdot w_r \cdot (r_1 - r_2)^2 + S_g^2 \cdot w_g \cdot (g_1 - g_2)^2 + S_b^2 \cdot w_b \cdot (b_1 - b_2)^2}{(w_r + w_g + w_b) \cdot 255^2} + \theta^2 \cdot S_\theta \cdot S_{ratio}}, \quad (1)$$

where w_r , w_g , and w_b are the weights of the human eye’s sensitivity to changes in the red, green, and blue components, respectively, with the range [0,1] (in this work, we took the equality weight for them). S_i , $i = r, g, b$ denotes the importance of the R, G, and B color channels, respectively, determined by

$$S_i = \min\left(\frac{3(i_1 + i_2)}{r_1 + r_2 + g_1 + g_2 + b_1 + b_2}, 1\right), \quad i = r, g, b. \quad (2)$$

Subscripts 1 and 2 denote two different colors. θ is the normalized angle between two color vectors in RGB space, given by

$$\theta = \frac{2}{\pi} \cdot \arccos \left(\frac{r_1 \cdot r_2 + g_1 \cdot g_2 + b_1 \cdot b_2}{\sqrt{(r_1^2 + g_1^2 + b_1^2) \cdot (r_2^2 + g_2^2 + b_2^2)}} \right). \quad (3)$$

S_θ is the contribution of the three colors' difference to the vector angle of two colors to be compared, which is adjusted by θ under the importance of each channel

$$S_\theta = S_{\theta_r} + S_{\theta_g} + S_{\theta_b}, \quad (4)$$

and

$$S_{\theta_i} = \frac{\frac{|i_1 - i_2|}{i_1 + i_2}}{\frac{|r_1 - r_2|}{r_1 + r_2} + \frac{|g_1 - g_2|}{g_1 + g_2} + \frac{|b_1 - b_2|}{b_1 + b_2}} \cdot S_\theta^2, \quad i = r, g, b. \quad (5)$$

and

$$S_{ratio} = \frac{\max(r_1, r_2, g_1, g_2, b_1, b_2)}{255} \quad (6)$$

According to Equations (1)–(6), the RGB values are dynamically adjusted to compensate for the inhomogeneity of the RGB system. Thus, the distance between each pair of colors composed of RGB is obtained.

2.3.2. Rapeseed Seed Color Calibration

Comparing the rapeseed seed colors involved in our experimental as shown in Figure 1 with the standard and extended colors in the W3C (https://www.w3.org/wiki/CSS3/Color/Extended_color_keywords, accessed on 19 October 2023), referring to the world wide web consortium, an international organization tasked with developing web technology standards, 14 colors in the families of yellow, orange, and brown, namely, dark gray, red, brown, dark brown, tan, maroon, orange, dark orange, coral orange-red, green yellow, yellow, light yellow, and golden, are selected as benchmarks for the calibration of seed color, as shown in Figure 3. According to the VAD method presented in Section 2.3.1, the distance between the RGB values of the 14 colors and those of our 282 samples is calculated. The category with the closest distance is taken as the color category of the samples. It shows that the 282 samples are categorized into three color categories of coral (75 samples, denoted as Class 0), brown (20 samples, denoted as Class 1), and dark brown (187 samples, denoted as Class 2), whose corresponding HSV systems are (0.0448, 0.6863, 1), (0.0000, 0.7455, 0.6471), (0.0833, 0.6733, 0.3961), respectively. The HSV color model is a color space, also known as the hexane model, where H stands for Hue, which describes the position of color in the color spectrum; S stands for saturation, which describes the vividness of color; and V stands for Value, which is the luminance value of the color, with higher values indicating lighter colors and lower values referring to darker colors. From the corresponding HSV values, it can be concluded that the coral category of rapeseed is the lightest in color, followed by the brown category, and the darkest in color is the dark brown category of rapeseed; thus, the coral category of rapeseed owns a higher oil content and is a target for breeding and industrial screening. In this respect, the identification of seed color plays an important role in predicting oil content and breeding. Because the different methods of light absorption and scattering will lead to different colors on the seed's surface, spectral reflectance is an ideal measure used to characterize the color of the seed. To verify this visually, we lay the above three categories of samples into a three-dimensional space constructed by the top three principal components with the highest contribution of spectral reflectance, as shown in Figure 4. The samples of the three categories are separated basically, which prompts us to further study the identification model of the seed color separation by using the hyperspectral feature. We also extract nine color indices, as listed in Table 1, which are highly related to the visual inspection of the seed coat color.

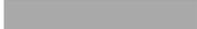
| Color card | Color names | RGB values |
|---|-------------|-----------------|
|  | darkgray | (169, 169, 169) |
|  | red | (255, 0, 0) |
|  | brown | (164, 42, 42) |
|  | darkbrown* | (101, 67, 33) |
|  | tan | (210, 180, 140) |
|  | maroon | (128, 0, 0) |
|  | orange | (255, 165, 0) |
|  | darkorange | (255, 140, 0) |
|  | coral | (255, 127, 80) |
|  | orangered | (255, 69, 0) |
|  | greenyellow | (173, 255, 47) |
|  | yellow | (255, 255, 0) |
|  | lightyellow | (255, 255, 224) |
|  | gold | (255, 215, 0) |

Figure 3. Color card with RGB values. Tips: * according to the rule of W3C css-system (<https://www.w3.org/TR/css-color-3/#css-system>, accessed on 19 October 2023), authors can customize the color values according to the environmental needs, we add a new color card, namely, darkbrown, RGB (101, 67, 33), with reference to color-hex (<https://www.color-hex.com/color/654321>, accessed on 19 October 2023), that is closer to the real rapeseed seed coat color to the color card family.

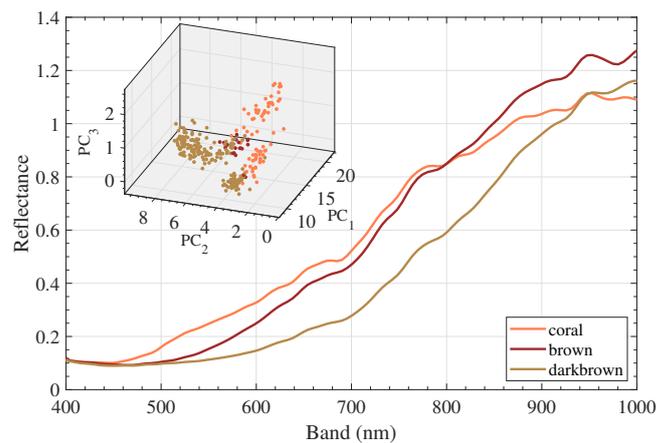


Figure 4. Main panel is the average spectral reflectance of rapeseed samples of the three categories; insert is the scatter plot in a 3-dimensional space constructed by the three principal components of the spectral reflectance.

Table 1. Nine selected color indices.

| Name | Abbreviation | Formula | Source |
|--|--------------|---|-----------|
| 1 chlorophyll absorption ratio index | CARI | $(R_{750} - R_{705}) / (R_{750} + R_{705})$ | Ref. [17] |
| 2 photochemical reflectance index | PRI | $(R_{531} - R_{570}) / (R_{531} + R_{570})$ | Ref. [22] |
| 3 red ratio spectral index | RRSI | R_{705} / R_{910} | Ref. [30] |
| 4 green ratio spectral index | GRSI | R_{582} / R_{910} | Ref. [30] |
| 5 red difference spectral index | RDSI | $R_{418} - R_{610}$ | Ref. [30] |
| 6 green difference spectral index | GDSI | $R_{421} - R_{536}$ | Ref. [30] |
| 7 red normalized difference spectral index | RNDSI | $(R_{646} - R_{995}) / (R_{646} + R_{995})$ | Ref. [30] |
| 8 green normalized difference spectral index | GNDSI | $(R_{568} - R_{988}) / (R_{568} + R_{988})$ | Ref. [30] |
| 9 blue normalized difference spectral index | RNDSI | $(R_{531} - R_{571}) / (R_{531} + R_{571})$ | Ref. [30] |

2.4. Visibility Graph Algorithm

Spectral reflectance is the ratio of the reflected flux of a ground object in a certain wavelength band to the incident flux in that band, which contains two aspects of information, namely, the data value and the order of bands. The spectral data value is of importance for characterizing the characteristics of ground objects. However, it only represents the absolute reflection of light from different wavebands on an object, but cannot reflect the difference in their reflectance at different bands. We need a tool to effectively capture the difference in the reflectance among every band associated with their order. Here, we use an idea of the visibility graph (VG) algorithm [31] to transform each spectral reflectance series into a network firstly, where the network nodes are the bands and the link among them refers to the relationship of the reflectance according to visibility rules. It has the advantage of retaining the information on the reflectance values and the order of bands. Then, five network global attributes are extracted to be used for the complements to the full-band reflectance data value. Finally, the full-band reflectance and the net attributes are fused to be used as features and input into the intelligent models for seed color identification.

2.4.1. Introduction of Visibility Graph Algorithm

The VG algorithm was proposed to analyze time series from a higher dimensional view. With the merit of operability without any assumptions, the algorithm was widely applied in various fields [32–36]. The process of VG consists of two parts: firstly, the reflectance values are arranged on the axes as a bar chart in sequential order; then, a line is drawn at the top of the bar chart according to the visibility criterion. The visibility criterion is defined below:

$$x_k < x_j + (x_i - x_j) \cdot \frac{t_k - t_j}{t_i - t_j}, \quad \forall k \in (i, j), \quad (7)$$

where t and x denote the horizontal and vertical coordinates, respectively. (t_i, x_i) and (t_j, x_j) are two points, and (t_k, x_k) is an arbitrary point between them. According to this criterion, it determines whether there is a link between two given bands and thus yields a so-called VG network. Using this criterion on the hyperspectral data can reflect the relative values of the spectral reflectance between any two bands. To capture more accurate information on the relative changes in reflectance between two bands, in this work, we consider a weight edge connection based on the original criterion. The edge weight is defined as the angle between the line connecting the reflectance of two visible bands and the horizontal direction,

$$w_{ij} = \frac{1}{2} \arctan\left(\frac{x_i - x_j}{t_i - t_j}\right), \quad (8)$$

where x_i and x_j are the two points that satisfy Equation (7). Accordingly, the weighted VG networks can be constructed by using the weighted links.

The idea of VG is simple, and the corresponding VG network possesses the proprieties of connectivity, anisotropy, and affine invariance, which makes it inherit the intrinsic structural characteristics of the original series. However, when examining the visibility between every pair of points, it is necessary to judge the value of each point between them by Equation (7), which leads to high time complexity. Therefore, in this work, we use a fast conversion algorithm based on the partition strategy introduced by [37]. The main idea is: firstly, divide the original bands into left and right segments at maximum value; then, starting from the maximum point, examine the visibility by comparing the angle between adjacent two bands (given by Equation (7)) and traverse each point of the left and right bands, respectively, meanwhile, the weight is calculated by Equation (8); and lastly, continue the above connection process recursively until the band on both sides is no longer divisible. Accordingly, by mapping every band into network nodes and using the fast-weighted VG algorithm, 282 spectral-weighted VG networks can be obtained.

2.4.2. Network Global Attributes

Complex networks are characterized by simplicity intuition, and structural robustness [38,39]. For each hyperspectral reflectance signal, its VG network $G = (N, L, W)$, N is the node set of all nodes with the number n ; L is the set of all links with the number l ; and W is the set of weight values of all links, with the element w_{ij} given by Equation (8). (i, j) is a link between nodes i and j with the weight w_{ij} , $(i, j \in N)$. The interpretation of the spectral data can be achieved by extracting its corresponding VG network topological properties. To this end, five global network attributes are employed for our consideration.

(1) Clustering coefficient

The clustering coefficient characterizes the connectivity of the neighboring nodes of a node. The global clustering coefficient can be defined as the average of the clustering coefficients of all nodes, or it can be expressed as the ratio of three times the number of all triangles in the network to the ratio of all connected triples in the network. The two definitions are formally different and not exactly equivalent, and the second definition is used in this study, given by

$$\langle CC \rangle = \frac{1}{n} \sum_{i=1}^n \frac{\sum_{j \neq k} w_{ij} \cdot w_{ik} \cdot w_{jk}}{\max(w_{ij}) \sum_{j \neq k} w_{ij} \cdot w_{ik}}. \quad (9)$$

(2) Assortativity coefficient

The assortativity coefficient measures the degree of similarity or interconnectedness of nodes in a network and provides a quantitative way to analyze the network's similarity and connectivity patterns [38]. Positive assortativity coefficients indicate that nodes tend to interconnect with other nodes with similar or the same strength, which means that nodes with similar properties are more likely to form connections. The weighted assortativity coefficient was developed by [40], calculated by

$$\gamma = \frac{l^{-1} \sum_{(i,j) \in L} w_{ij} \cdot k_i^w \cdot k_j^w - [l^{-1} \sum_{(i,j) \in L} w_{ij} \cdot (k_i^w + k_j^w) / 2]^2}{l^{-1} \sum_{(i,j) \in L} w_{ij} \cdot (k_i^{w2} + k_j^{w2}) / 2 - [l^{-1} \sum_{(i,j) \in L} w_{ij} \cdot (k_i^w + k_j^w) / 2]^2}, \quad (10)$$

where k_i^w is the weighted degree of node i , calculated as $k_i^w = \sum_{j \in N} w_{ij}$.

(3) Global efficiency

Global efficiency describes the speed and efficiency of information transfer between any two nodes in a network. Global efficiency can be calculated by the following steps: for each pair of nodes in the network, calculate its shortest path length (shortest path length). The shortest path length is the sum of the minimum number of edges or weights required for a connection between two nodes. The reciprocal of all the shortest path lengths is added and divided by the total number of nodes in the network minus 1. Here, the reciprocal indicates that the shorter the distance of the path, the more efficient it is. The final result obtained is the global efficiency. The weighted efficiency (denoted by Eff) is computed using an auxiliary connection-length matrix related to the edge weight [38], determined by

$$Eff = \frac{1}{n(n-1)} \sum_{i \in N} \sum_{j \in N, j \neq i} (d_{ij}^w)^{-1}, \quad (11)$$

where d_{ij}^w is the shortest weighted path length between nodes i and j computed by Dijkstra's algorithm.

(4) Structural entropy

In information theory, entropy is used to measure the uncertainty of information. Structural entropy, on the other hand, is used to measure the complexity of the structure of a system. Network structural entropy is a measure of the structural complexity of a network [41,42]. A higher structural entropy of a network indicates a more complex structure, implying a more even or diverse distribution of node degree values within the network. Conversely, a lower structural entropy suggests a simpler structure, with a

more concentrated or uneven distribution of node degrees. Calculating network structural entropy is a common method used to assess network complexity; it is based on the entropy of node degree distribution [43,44], which is determined by

$$Ent = - \sum_{i=1}^n I_i \cdot \ln I_i,$$

where I_i is the importance of node i , defined by $I_i = k_i / \sum_{i \in N} k_i$, and k_i is the degree of node i . In calculating the structural entropy of the network, we excluded nodes with degree 0.

The two extreme values are $Ent_{\max} = \ln n$ when all nodes in the network have the same degree, where $I_i = 1/n$ while $Ent_{\min} = [\ln 4(n-1)]/2$, when the network has a highly skewed distribution of node degrees, with some nodes having very high degrees and others having very low degrees, where $I_1 = 1/2$ and $I_j = 1/[2(n-1)]$ for $j > 1$. To mitigate the influence of the network size n on the Ent , we can normalize Ent using the following approach to ensure that $0 \leq NEnt \leq 1$:

$$NEnt = \frac{Ent - Ent_{\min}}{Ent_{\max} - Ent_{\min}} = \frac{-2 \sum_{i=1}^n I_i \cdot \ln I_i - \ln 4(n-1)}{2 \ln n - \ln 4(n-1)} \quad (12)$$

(5) Graph density

Graph density [38] is the ratio between the number of actual edges and the number of possible edges in a network and measures how closely the nodes in the network interact with each other and the connectivity of the graph. The graph density ranges from 0 to 1. When the graph density is close to 1, it means that the nodes in the network are very tightly connected and are close to being fully connected. When the graph density is close to 0, it means that the nodes in the network are very sparsely connected, and the correlation between the nodes is weak. The graph density only considers the number of edges and does not consider the edge weights, which are calculated as follows:

$$\rho = \frac{2l}{n(n-1)}. \quad (13)$$

We briefly conclude how the VG network works for the hyperspectral feature extraction as follows: Firstly, each hyperspectral reflectance curve is mapped into the weighted network according to the fast-weighted VG algorithm. Each network can be represented by a weighted adjacency matrix. Then, the five network attributes are calculated by using Equations (9)–(13) based on the weighted adjacency matrix, which is used to characterize the topological structure of each reflectance curve. As mentioned above, the VG network inherits the intrinsic structural features of the original sequence, reflecting the relationship between the reflectance of different bands. It can be considered complementary to the full-band reflectance values. It is important to note that the network attributes cannot replace the reflectance in our model since the VG network properties reflect the relationship between the reflectances (relative difference) among every band rather than the reflectance information itself. Next, the five attributes are combined as the network feature. The fusion feature of the full-band reflectance values and the network feature is input into the intelligent models (see the next subsection) to implement the rapeseed color classification task, which is expected to achieve higher accuracy.

2.5. Classifiers

As a central task in supervised learning, classification has received widespread attention in various scientific fields. Up to now, lots of excellent classification models have been proposed. In this work, four classifiers, namely, Logistic Regression (Logit-R), Support Vector Machines (SVM), Random Forests (RF), and Extreme Gradient Boosting (XGBoost, we use xgboost with version number 1.7.6 in Python with version number 3.8.13), are employed for rapeseed color classification. Logit-R is a classical generalized linear regression

analysis model, which was introduced in the early 20th century. SVM was generated in the 1960s, which performs well in high-dimensional spaces and is especially suitable for the case of a few samples. RF was developed in 1995, contains multiple decision trees, and has good immunity to outliers. As one of the most outstanding models in machine learning in the past decade, XGBoost is a significant promotion of the gradient-boosting decision tree. Next, we introduce them briefly.

2.5.1. Logit-R

Logit-R is a generalized linear regression model. It first estimates the probability of an event occurring based on a given dataset of independent variables and then uses a step function or sigmoid function to achieve the purpose of sample classification. With this simple and easy-to-operate idea, Logit-R has become one of the frequently employed techniques for classification in the field of machine learning. It was originally used for binary classification tasks. Some meaningful extensions, such as multinomial Logit-R, are used to handle multi-classification tasks. Meanwhile, regularized versions, like L_1 or L_2 regularization, help to prevent overfitting by penalizing large coefficients. However, Logit-R cannot solve linear inseparable problems and nonlinear problems since the essence of this method is to classify by generating a straight line or a hyperplane.

2.5.2. SVM

SVM is a popular statistical theory-based learning method. The core idea is to find a segmentation hyperplane so that the points on both sides are as far away as possible from the hyperplane, i.e., the maximum spacing hyperplane. Mapping data to higher dimensional spaces using kernel functions makes the samples easier to separate in a new space; in the meantime, it avoids the computation of nonlinear surface segmentation in the original input space. A linearly separable binary classification hyperplane can be expressed as $\omega \cdot x + b = 0$ with a constraint of $y(\omega \cdot x + b) \geq 1$, whose optimal solution can be obtained via the Lagrange function. Additionally, by introducing a regularization parameter, it is easy to control the penalty for classification errors. In the method of SVM, the penalty coefficient C is a critical parameter. The smaller values of C will allow some sample classification errors but will produce larger intervals. Conversely, the larger C will strictly penalize classification errors and may lead to more compact but complex decision boundaries. SVM is an efficient method for representing complex functional dependencies in high-dimensional spaces [45–47]. SVM can be used to process the data with nonlinear features; however, its classification results mainly depend on the selection of kernel functions. Therefore, parameter tuning is important when applying support vector machine modeling for classification.

2.5.3. RF

RF is a parallel ensemble learning algorithm through random sampling and selecting features, which is based on bootstrap aggregation. It constructs multiple decision trees and integrates them to improve the model performance and generalization ability. With these advantages, it has a good ability to process high-dimensional and large-scale datasets with good immunity to noisy data. In addition, overfitting will not occur even without the need for complex parameter tuning in real data modeling.

2.5.4. XGBoost

XGBoost is a class of integrated algorithms based on gradient-boosted trees proposed by Chen [48]. Gradient boosting is an integrated learning method that iteratively adds multiple base learners (decision trees) by learning on the residuals of the previous round of base learners, thereby reducing the prediction error and improving the overall performance. With the base learner of classification and regression tree (CART), XGBoost is a scalable machine learning system that can be used to solve classification and regression problems. It predicts the output by using the K additivity function consisting of multiple weak classifiers:

$\hat{y} = \sum_{k=1}^K f_k(x_i)$, $f_k \in F$, where F is the ensemble space of CART. To learn the set of functions, a regularization objective is typically included to minimize the objective function [49]: $L(\phi) = \sum_i \tau(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$, where $\Omega(f) = \gamma T + \frac{1}{2} \lambda |\omega|^2$. By using gradient boosting methods and iterative strategy, XGBoost adopts parallel and distributed computation.

Moreover, the XGBoost makes full use of hardware acceleration and multi-threaded processing to accelerate the training process, which greatly improves computational efficiency; therefore, it can be used to solve large-scale problems. Moreover, XGBoost can handle advanced, complex feature spaces and, through techniques such as weight adjustment and sampling strategies, can handle data that are highly unbalanced in class distribution to make accurate predictions [48,50–54].

The XGBClassifier function of the XGBoost module in the scikit-learn extension library has three types of arguments: regular parameters, model parameters, and learning parameters. *n_estimators* and *learning_rate* are some of the most important parameters in the model parameters and learning parameters, respectively. Among them, *n_estimators* is the number of decision trees used in constructing the gradient boosting model. A larger value of *n_estimators* means that the model has more decision trees and therefore higher complexity and learning capacity. Increasing *n_estimators* can improve model performance, especially when working with complex problems or large datasets. However, if *n_estimators* is set to be too large, the model may overfit the training data, resulting in a decrease in generalization performance on new data and an increase in training time. *learning_rate* is the step size that controls how much each base learner (decision tree) learns about the residuals during an iteration. A smaller value of *learning_rate* means that the model weights are updated less at each iteration, so training will be slower. The value is a floating point number between 0 and 1, with a default value of 0.3. A smaller *learning_rate* improves the stability and generalization of the model and reduces the risk of overfitting. However, training time may increase with a smaller *learning_rate*. Optimization techniques can be used to find the best combination of hyperparameters to obtain the best model performance.

2.6. Performance Evaluation

Accuracy indicates the proportion of samples correctly classified by the model for the entire test dataset. The average accuracy (*Acc*) is the most widely employed performance measure for classification models. For the unbalanced classification data, however, there is a flaw in this metric that cannot truly reflect the classification level of the model. Naturally, several metrics such as precision (*Pre*), recall (*Rec*), and F_1 score are designed to supplement the average accuracy rate. In this work, we focus on the 3-categorization rapeseed seed color classification task. The above-evaluating indicators are calculated separately for each category. The category of interest is regarded as the positive class, and all others are negative classes. Four kinds of predictions are defined for a certain class on the test set: (1) *TP*: refers to the number of samples in the positive class that the model accurately predicts; (2) *FN*: refers to cases where the model misclassifies a sample that is a positive class as a negative class; (3) *FP*: refers to the number of positive classes is mistakenly predicted as the negative class; and (4) *TN*: represents the number of samples in the negative class that the model correctly predicts. The confusion matrix is shown in Figure 5. Accordingly, *Acc*, *Pre*, *Rec*, and F_1 are given by Equation (14). The K -fold cross-validation is used to evaluate the performance of models. The calculation process is then repeated 100 times to mitigate the influence of randomness.

$$\begin{cases} Acc = \frac{TP + TN}{TP + FN + FP + TN} \\ Pre = \frac{TP}{TP + FP} \\ Rec = \frac{TP}{TP + FN} \\ F_1 = \frac{2TP}{TP + FP + FN} \end{cases} \quad (14)$$

| | | predicted | | | | | predicted | | |
|----------|---------|-----------|---------|---------|----------|---------|-----------|-----------|-----------|
| | | class 1 | class 2 | class 3 | | | class 1 | class 2 | class 3 |
| observed | class 1 | a11 | a12 | a13 | observed | class 1 | <i>TP</i> | <i>FN</i> | |
| | class 2 | a21 | a22 | a23 | | class 2 | <i>TN</i> | | |
| | class 3 | a31 | a32 | a33 | | class 3 | | | <i>FP</i> |

Figure 5. Confusion matrix of 3-categorization (**left**), where a11, a22, and a33 denote the correct categorization for each class, respectively. Taking class 1 as an example of interest, TP, TN, FP, and FN are shown on the (**right**) panel.

3. Results

According to the framework shown in Figure 2, each sample in the hyperspectral dataset is transferred into a network (graph) structure by using the VG algorithm. For each network, the five network attributes are extracted, whose basic statistics are shown in Table 2. The numerical differences in different categories help to classify them. Following this, we enter the experimental process of the classification task. Three aspects of comparison experiments, namely, comparison among classifiers, comparison among features, and impact of model parameters, are considered in this process. The results are reported in each subsection.

Table 2. Statistical description of the global network attributes of the three categories samples.

| Attribute | Categories | Mean | Std | Min | 25% | 50% | 75% | Max |
|----------------------|------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| $\langle CC \rangle$ | coral | 3.60×10^{-3} | 8.18×10^{-4} | 2.11×10^{-3} | 3.08×10^{-3} | 3.50×10^{-3} | 4.19×10^{-3} | 5.66×10^{-3} |
| | brown | 3.78×10^{-3} | 6.11×10^{-4} | 2.58×10^{-3} | 3.30×10^{-3} | 3.86×10^{-3} | 4.17×10^{-3} | 5.21×10^{-3} |
| | dark brown | 3.71×10^{-3} | 8.35×10^{-4} | 2.12×10^{-3} | 3.18×10^{-3} | 3.54×10^{-3} | 4.06×10^{-3} | 6.71×10^{-3} |
| $NEnt$ | coral | 0.89 | 0.01 | 0.86 | 0.88 | 0.89 | 0.89 | 0.93 |
| | brown | 0.9 | 0.01 | 0.88 | 0.9 | 0.9 | 0.91 | 0.92 |
| | dark brown | 0.93 | 0.03 | 0.87 | 0.91 | 0.93 | 0.95 | 0.98 |
| γ | coral | 0.19 | 0.11 | -0.03 | 0.1 | 0.19 | 0.27 | 0.46 |
| | brown | 0.19 | 0.13 | -0.16 | 0.14 | 0.19 | 0.27 | 0.45 |
| | dark brown | 0.09 | 0.15 | -0.36 | 0.02 | 0.12 | 0.19 | 0.38 |
| Eff | coral | 2.71×10^{-3} | 7.75×10^{-4} | 1.40×10^{-3} | 2.25×10^{-3} | 2.61×10^{-3} | 3.02×10^{-3} | 4.75×10^{-3} |
| | brown | 3.54×10^{-3} | 6.54×10^{-4} | 2.49×10^{-3} | 3.18×10^{-3} | 3.47×10^{-3} | 3.83×10^{-3} | 5.44×10^{-3} |
| | dark brown | 4.32×10^{-3} | 1.27×10^{-3} | 2.00×10^{-3} | 3.43×10^{-3} | 4.31×10^{-3} | 5.00×10^{-3} | 9.01×10^{-3} |
| ρ | coral | 0.16 | 0.04 | 0.09 | 0.13 | 0.15 | 0.18 | 0.27 |
| | brown | 0.25 | 0.03 | 0.19 | 0.24 | 0.25 | 0.27 | 0.31 |
| | dark brown | 0.36 | 0.10 | 0.14 | 0.28 | 0.37 | 0.44 | 0.55 |

3.1. Performance Comparison among Four Classifiers

With the different working principles, the performance of the classifiers is influenced by the features as well as the datasets. To comprehensively examine the validity of the proposed framework of the rapeseed color classification, the four classifiers mentioned in Section 2.5 are employed in our dataset. We input the fusion features composed of the full-band reflectance and the five network attributes into the classifier of RF, SVM, Logit-R, and XGBoost, respectively. The K -fold cross-validation with $K = 2 \sim 10$ is executed on the four models. The model performance presented by Acc is shown in Figure 6. The standard deviation of the Acc calculated from 10 repetitions is also shown on each bar. Clearly, although there is different model performance obtained from the four considered classifiers, the fact of the Acc with more than 0.9 shows our model is workable. The best performance comes from the XGBoost model, with its highest Acc reaching at 0.9315 under 9-fold cross-validation. This is not surprising since this popular method can adjust weights on imbalanced samples adaptively, which applies to our data exactly (in our dataset, the numbers of the three color categories are 75, 20, and 187). By contrast, the performance

of the SVM classifier is the worst. One of the reasons is just the impact of imbalanced samples in our dataset.

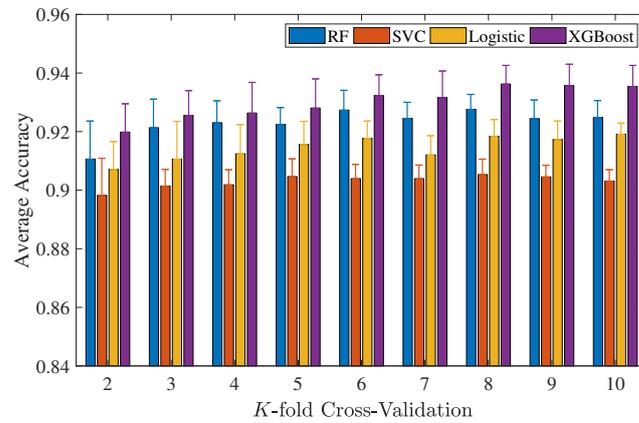


Figure 6. Comparison of the average *Acc* among four classifiers of RF, SVM, Logit-R, and XGBoost concerning *K*-fold cross-validation. The error bar represents the standard deviation calculated from 10 repetitions.

3.2. Performance Comparison among Features

In this subsection, we focus on the comparison of the model performance among different types of hyperspectral features. Besides the network attributes and full-band reflectance, we also investigate the model result from the fusion feature of nine color indices, as listed in Table 1, which is the full-band reflectance. Therefore, we have six types of features, namely, spectral reflectance, color index, network attributes, the fusion feature of color index and reflectance, the fusion feature of network attributes and reflectance, and the fusion feature of color index and network attributes. We perform the same experimental process on them. At this time, only the XGBoost classifier is employed for this purpose due to its best performance. The result is presented in Figure 7. As expected, the fusion feature displays better model performance than the full-band reflectance with every *K*-fold cross-validation. Although the only network attributes produce unsatisfactory results, they help to upgrade the model performance via the feature fusion with the spectral reflectance, demonstrating that the network feature is an important supplement for the original spectral reflectance. We will discuss this detail in Section 4. After fusing the two types of features, the *Acc* is improved obviously under all folds of cross-validation, with the highest upgrade rate being 1.33%.

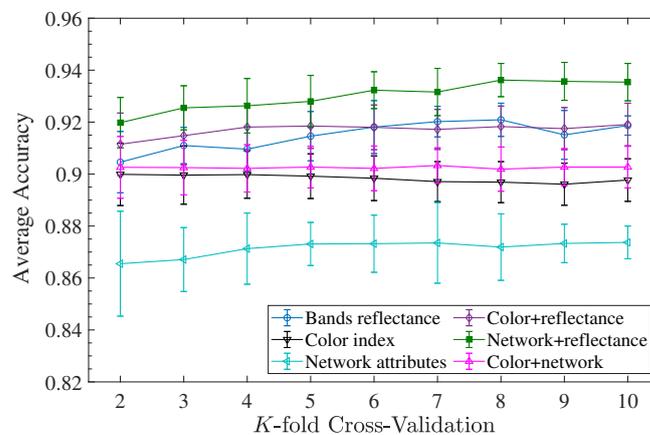


Figure 7. Comparison of the average *Acc* between six types of features with *K*-fold cross-validation by using XGBoost. The error bar represents the standard deviation calculated from 10 repetitions.

3.3. Performance of Parameter Tuning

As is well as known, hyper-parameter selection is critical in machine learning, which affects the model performance greatly. In this subsection, we try to achieve optimal model performance by parameter tuning. Generally, grid search is a universal method for exploring model hyper-parameters; however, it has huge time complexity. To efficiently address the hyperparameter tuning, the sparrow algorithm [55] was proposed, which is a metaheuristic bio-inspired optimization algorithm inspired by the foraging and evasion behaviors of sparrows. In the sparrow algorithm, individuals are divided into two categories: producers and foragers. Producers behave like resource seekers, actively searching for potential solutions in the solution space, akin to foraging. Foragers are responsible for collecting these solutions through interactions with producers, similar to how sparrows obtain food from their peers. The sparrow algorithm utilizes random optimization techniques to perform global and local searches in the solution space for model parameter optimization. In what follows, we apply the sparrow algorithm to find the optimal parameter in the XGBoost model. The two optimal parameters of $n_estimators$ and $learning_rate$ of the XGBoost model are autonomously searched, as listed in Table 3 for K -fold cross-validation ($K = 2\sim 10$), which is expected to bring better model performance. Figure 8 illustrates the results obtained from the fusion feature of network and bands before and after parameter tuning under 2~10-fold cross-validation. It is seen that the Acc has improved slightly under all folds cross-validation after tuning with the maximum upgrade rate of 1%, as shown in the insert plot. Quantitatively, the average classification result under five-fold cross-validation is listed in Table 4, where the three evaluative indicators defined by Equation (14) are shown on the left side and the confusion matrix is shown on the right side.

Table 3. Optimal parameters of the XGBoost model for each fold of K -fold Cross-validation ($K = 2$ to 10) Using the Sparrow Algorithm.

| K | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $n_estimators$ | 1500 | 1500 | 10 | 1500 | 620 | 77 | 1500 | 1500 | 1443 |
| $learning_rate$ | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.01 | 0.3 | 0.3 | 0.3 |
| fitness value | 0.397 | 0.159 | 0.162 | 0.144 | 0.106 | 0.102 | 0.095 | 0.063 | 0.073 |

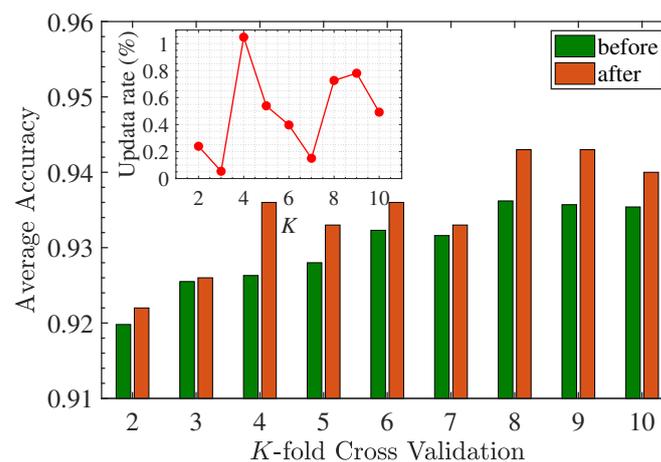


Figure 8. Comparison of the average Acc calculated from the fusion feature of network and spectral reflectance before and after parameter tuning.

Table 4. The average classification result under five-fold cross-validation. The left one shows the results of precision, recall, and F_1 -score. The right one shows the corresponding confusion matrix.

| | <i>Pre</i> | <i>Rec</i> | <i>F₁-Score</i> | <i>Support</i> | | 0 | 1 | 2 |
|------------|------------|------------|----------------------------|----------------|---|----------|----------|----------|
| 0 | 0.92 | 0.95 | 0.93 | 75 | 0 | 71 | 2 | 2 |
| 1 | 0.75 | 0.70 | 0.67 | 20 | 1 | 2 | 12 | 6 |
| 2 | 0.96 | 0.97 | 0.96 | 187 | 2 | 4 | 2 | 181 |
| <i>Acc</i> | | | 0.94 | 282 | | | | |

3.4. Misclassification Sample Analysis

In this subsection, we focus on the misclassified samples, which may help us better understand the rapeseed classification mechanism. According to the confusion matrix shown on the right side of Table 4, we record the 18 misclassified samples in Table 5, where ‘Center label’ refers to the class label that has the shortest average Euclidean distance between the misclassified sample under consideration and the classes in the training set, and ‘Closest label’ refers to the class label of the nearest neighbor (denoted as Closest ID, see the second column on the right of the table) of the misclassified sample being considered. Three situations can be concluded: namely, the sample is misclassified to the class with the center label (Case I), the sample is misclassified to the class with the closest label (Case II), and the other case (case III).

Table 5. The three situations of the 18 misclassified samples.

| | Misclassified ID | True Label | Pred Label | Center Label | Closest ID | Closest Label |
|----------|-------------------------|-------------------|-------------------|---------------------|-------------------|----------------------|
| Case I | 30 | 0 | 1 | 1 | 37 | 2 |
| | 41 | 0 | 1 | 1 | 121 | 2 |
| | 66 | 0 | 2 | 2 | 136 | 2 |
| | 73 | 0 | 2 | 2 | 81 | 2 |
| | 71 | 1 | 2 | 2 | 77 | 1 |
| | 77 | 1 | 2 | 2 | 179 | 2 |
| | 153 | 1 | 2 | 2 | 168 | 2 |
| | 121 | 2 | 1 | 1 | 41 | 0 |
| | 122 | 2 | 1 | 1 | 109 | 2 |
| Case II | 84 | 1 | 2 | 1 | 124 | 2 |
| | 93 | 1 | 2 | 1 | 77 | 2 |
| | 98 | 1 | 2 | 1 | 102 | 2 |
| | 37 | 2 | 0 | 1 | 30 | 0 |
| | 54 | 2 | 0 | 2 | 47 | 0 |
| | 131 | 2 | 0 | 1 | 94 | 0 |
| | 134 | 2 | 0 | 2 | 55 | 0 |
| Case III | 86 | 1 | 0 | 1 | 99 | 1 |
| | 100 | 1 | 0 | 1 | 36 | 1 |

For case I, there are nine misclassified samples (#66, #73, . . . , #121), whose predicted label is the class with the center label, suggesting that they are closer to these classes. We show the samples with index #73 and #77 as an example in Figure 9a. Both of them are closest to class 2 from the spectral reflectance curve; thus they were misclassified to class 2. For case II, there are seven misclassified samples (#84, #93, . . . , #134), whose predicted label is the class with the closest label according to their nearest neighbor’s label. The sample with index #84 is exactly this case, as shown in Figure 9b, whose nearest neighbor is #124 with label 2; hence, it was misclassified to class 2. In both figures, *Dis* means Euclidean distance. Now, let us pay attention to samples #100 and #86, which do not belong to the above two cases. They are misclassified to Class 0; however, the closest sample to them is not in Class 0. This forces us to search for new potential evidence. According to the decision-making process of the XGBoost model, which makes bifurcation decisions based on the importance of features, say, the distribution of the reflectance may affect the classification. To this end,

Jensen – Shannon divergence (denoted as D_{J-S}) is employed for our consideration, which is always used to appraise the difference between two distributions. Figure 9c and 9d illustrate the probability density function (PDF) of the reflectance of the samples #100 and #86, respectively. We calculate the D_{J-S} between one of them and the average reflectance distribution of training samples for each class, as also shown in the legend of the figures. For sample #100, the value of D_{J-S} between its PDF and the PDF of the averaged reflectance of class 0 is the smallest, indicating that the sample #100 may be classified as class 0. For the sample #86, although the class with the smallest D_{J-S} still comes from its true label class (Class 1), its nearest neighbor (in the sense of minimum divergence) is the training sample #35, which belongs to Class 0. Therefore, sample #86 may be misclassified to Class 0.

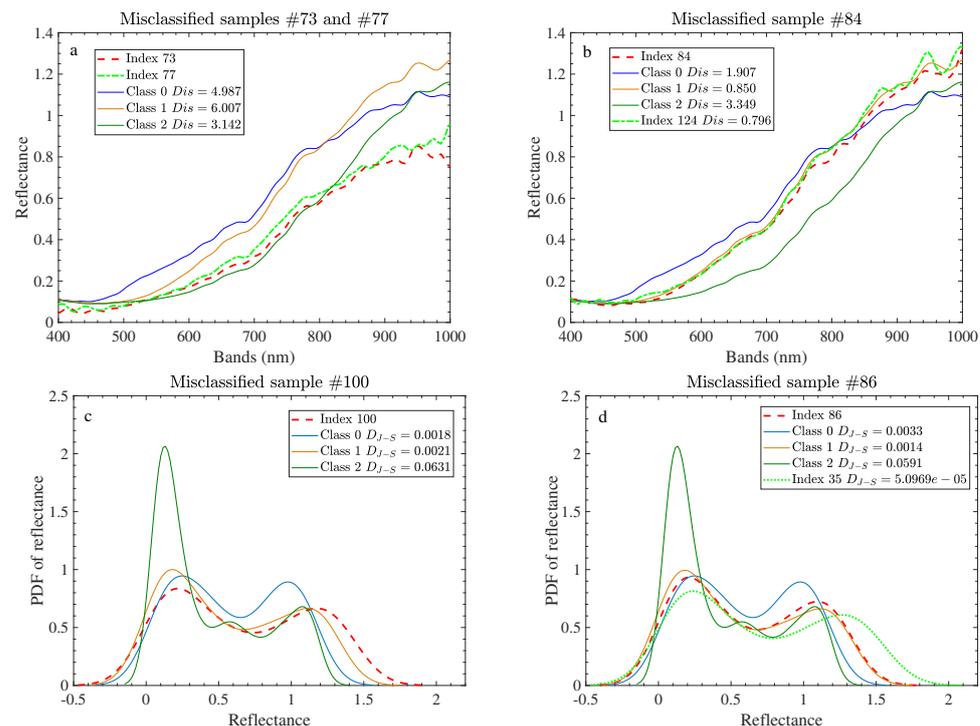


Figure 9. Three cases of misclassification. (a) is an example of misclassified samples #73 and #77, which belongs to case I; (b) is an example of a misclassified sample #84, which belongs to case II; (c) and (d) are examples of misclassified samples #100 and #86, respectively, which belong to the case III. Dis in upper panels means Euclidean distance, and D_{J-S} in bottom panels indicates Jensen – Shannon divergence.

4. Discussions

Hyperspectral technology, as a crucial near-earth remote sensing technique, boasts advantages such as non-destructiveness and efficiency, making it suitable for various crop growth monitoring tasks. Utilizing hyperspectral sensors enables the acquisition of spectral images and data across hundreds of bands. Different objects or surface materials exhibit distinct spectral reflectance in different bands, forming a unique spectral fingerprint. Leveraging spectral fingerprints allows for the effective differentiation of species types or prediction of surface properties and chemical compositions [13,56–58]. The complex network is a powerful tool used to understand the topological and dynamic characteristics of complex systems [34,38,41,59]. By applying complex network analysis methods, we can delve into the structural characteristics of spectral fingerprints and unearth special information embedded within them.

In this work, a visibility graph algorithm was adopted to extract global network features of spectral reflectance, revealing non-linear structural relationships within spectral fingerprints. Through integrating original spectral reflectance with the network features

and employing a machine learning classification model for rapeseed seed coat color identification, we observed a significant performance improvement compared to methods using only original reflectance. We also put the fusion features consisting of the original reflectance and spectral color index in the same model. The result shows that the recognition accuracy is not better than using only the original spectral reflectance, which indicates that the spectral color index cannot form a benign complementarity for the original spectral reflectance, while the network features can do so [60]. One potential reason may lie in the design of VG algorithms, i.e., the VG can capture the difference in the reflectance among every band associated with their order, which compensates for the differential information in variable bands of the original spectral reflectance. In addition, the fast-weighted VG algorithm enhances this difference numerically and improves the efficiency of network construction.

Machine learning, which has played an important role in the growth diagnosis of modern smart agriculture, helps train intelligent models for our rapeseed color classification. As a popular machine learning technology, XGBoost, renowned for its gradient boosting framework, weighted learning strategies, and parallel processing, typically outperforms traditional machine learning methods in terms of both performance and computational efficiency [48,50]. In this work, the XGBoost model is also confirmed to bring the best performance to the rapeseed color classification task.

Since this work involves two key issues in machine learning (feature extraction and model parameter selection), we will now discuss the sensitivity of these two issues to the results. On the one hand, the parameter tuning, as mentioned above, is critical to the machine learning model. In Section 3.3, we utilize the sparrow algorithm to find the optimal parameter combination of the XGBoost model for K -fold cross-validation ($K = 2$ to 10). The optimal parameter combinations of $\{n_estimators, learning_rate\}$ are quite different from those obtained by the sparrow algorithm as listed in Table 3. Here, we reconsider the parameter selection for the XGBoost model. At this time, we use a global search strategy to search for the optimal parameters of $n_estimators$ and $learning_rate$ under five-fold cross-validation. The $n_estimators$ and $learning_rate$ are set from 5 to 1630 and from 0.005 to 0.4, respectively, with equidistant 26 and 25 values. The Acc concerning the two parameters is shown in Figure 10. It is explained that the sparrow algorithm is a fast optimization method that may fall into local optima. As seen from Figure 10, there are five sets of parameter combinations (see the red triangles) that enable the Acc to reach its maximum value of 0.9397; interestingly, the corresponding optimal $learning_rate$ is at a very small level, i.e., 0.005~0.01, indicating that the trained model is relatively complex. By contrast, the optimal $n_estimators$ lacks an obvious pattern (the five optimal selections are 70, 265, 720, 785, and 1110), suggesting that the contribution of $n_estimators$ to the model is not as significant as that of $learning_rate$.

On the other hand, to investigate the performance of the five network attributes, a sensitive analysis is conducted for them. We employ the concept of information gain [61] to measure the importance of features. In practice, for every feature, calculate the information gain (denoted by $Gain$), defined by

$$Gain = H(D) - \sum_{v=1}^V \frac{N_{D_v}}{N_D} H(D_v), \quad (15)$$

where $Gain$ is the information gain of each feature, referring to the feature importance, N_D and N_{D_v} denote the number of samples in the dataset and feature subset, and $H(D)$ is the Shannon entropy, defined as

$$H(D) = - \sum_{i=1}^c p_i \log p_i,$$

where p_i is the ratio of the class i in the dataset D produced by dividing the dataset D into different subsets, and then we summarize these information gains. In general, the greater the $Gain$, the greater the contribution of the studied feature to classification,

and the higher the importance of the feature. We use this notion to calculate the importance of the five network attributes, which are illustrated in Figure 11a according to the model performance based on the fusion feature of the full-band reflectance and the network features. It illustrates that the most important network attribute is the graph density ρ , whose importance is higher than other attributes significantly. In addition, we calculate the *Acc* of the model under the fusion feature of full-band reflectance and the different number of the optimal network attributes (according to the importance sort shown in Figure 11a), as shown in Figure 11b. The result is obtained under a five-fold cross-validation. It shows that there are no significant differences between the five *Acc*, say, when we use only one network attribute (i.e., the graph density ρ), the *Acc* reaches 0.9325, which is slightly lower than 0.9397 in the case that we use all the five network attributes. The result is calculated under the optimal parameter combination discussed above. In this regard, the graph density is the only key feature, which can be suggested to execute the rapeseed coat color classification task.

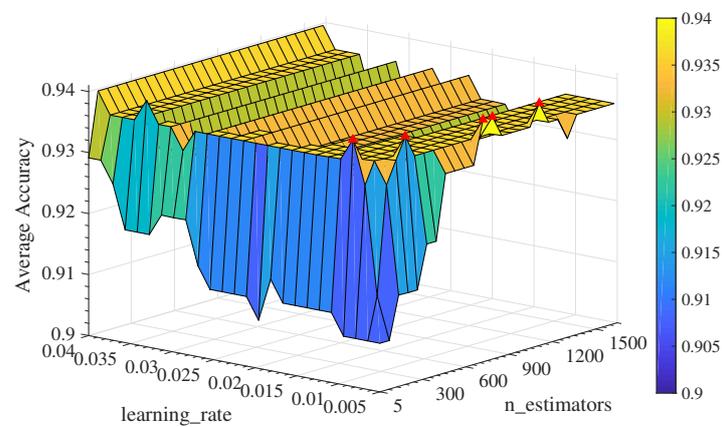


Figure 10. Parameter sensitivity analysis of XGBoost model. The best *Acc* is 0.9397; there are five corresponding optimal parameter combinations of $\{n_estimators, learning_rate\}$, as presented by the red triangle.

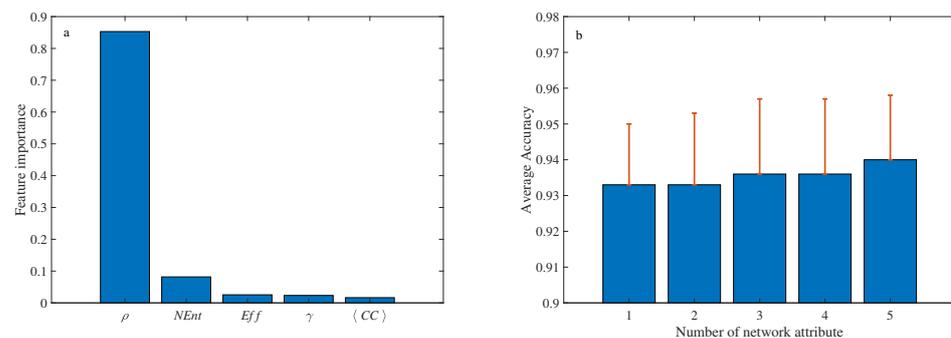


Figure 11. Network feature sensitivity analysis. (a) is the importance rate of the five network attributes, which is according to the model performance based on the fusion feature of the full-band reflectance and the network features. (b) is the *Acc* under the fusion feature of full-band reflectance and the different number of optimal network attributes with five-fold cross-validation.

Misclassification sample analysis is necessary for the classification task, especially when using machine learning models. By carefully analyzing the misclassification samples based on the confusion matrix, we can conclude there are three cases of misclassification. The mainstream cases are that the samples are misclassified to the class with the center label and the closest label, both in the sense of minimum Euclidean distance. The other two samples are exceptions, which are misclassified to the class with their closest class in the sense of minimum divergence.

One limitation of this study is the restricted sample size and color range analyzed. The experimental data are derived from two *Brassica napus* varieties grown in a single field trial season (2020–2021) in the Changsha City region. Due to the relatively small sample size, only 14 main colors representing each color family from the W3C system were selected as baseline color values. It is important to note that our integration of hyperspectral technology and machine vision for rapeseed seed coat color identification represents an initial effort in this field. This study contributes to a better understanding of the relationship between hyperspectral signals and image color information of crops, which are two critical technologies in smart agriculture. Furthermore, this work offers innovative insights into hyperspectral data mining from the perspective of complex networks, providing a necessary complement to the original reflectance data.

5. Conclusions

Machine vision technology has been widely applied in the growth monitoring and diagnosis of agricultural informatization, serving as a crucial auxiliary tool in traditional agriculture. Over the past decade, hyperspectral technology, as a novel information technology, has found extensive applications in both military and civilian sectors. In this work, we have proposed a framework that combines machine vision and hyperspectral technology with machine learning methods to achieve intelligent seed classification tasks. This approach provides a non-destructive and highly effective alternative to traditional labor-intensive and laboratory techniques for identifying the color of rapeseed seeds. With the utilization of inherent rich spectral and image data from rapeseed samples, this technology holds significant potential in the field of agriculture.

The proposed framework establishes a bridge between machine vision and hyperspectral technology. Specifically, our focus is on the color classification task of rapeseed seeds. Initially, three color categories are calibrated using visual perception and the vector angle distance method: 75 samples for coral color, 20 samples for brown, and 187 samples for dark brown. Subsequently, the weighted VG algorithm is used to map each of the hyperspectral samples into a network. Five network attributes are extracted and fused with the full-band spectral reflectance, which is input into the XGBoost model to achieve rapeseed color recognition. The experimental results show that the best *Acc* of the fused feature reaches 0.932 (under 8-fold cross-validation), which is higher than the results from the original full-band reflectance ($Acc = 0.922$) and other combination features significantly. The average classification recognition rate can be up to 0.943 through tuning parameters by the sparrow algorithm.

Theoretically, hyperspectral technology covers a wide range of bands, including visible light and near-infrared, with the advantages of high resolution and spectral integration, offering the potential for higher accuracy. However, due to the high time and cost associated with acquiring various sample data, the experimental sample data in this study are relatively limited. When the sample size is sufficient and relatively balanced, theoretically, a more accurate identification of a greater range of colors can be achieved by selecting more calibrated colors. Hyperspectral analysis is a key technology and method for achieving high-throughput analysis of crops [1,58,62–65].

This study proposes a novel framework that employs hyperspectral reflectance data and computer vision techniques to classify the color of rapeseed coats. By establishing a correlation between spectral characteristics and visual color perception, we introduce an intelligent model that offers a remarkably efficient alternative to conventional empirical and laboratory-based methods for determining the color of rapeseed seeds. Leveraging the inherent spectral and image information contained within rapeseed samples, this technology holds promising applications across various domains within the agricultural sector. The use of hyperspectral technology for identifying rapeseed seed coat colors has become part of research efforts, including identifying rapeseed varieties and quality, predicting rapeseed yield, and analyzing oil content. We aim to extend this research to address other agricultural challenges. One key objective is to establish a connection between

rapeseed of different colors and their genetic information, enabling the classification of rapeseed quality based on color. This approach would provide support for traditional laboratories in identifying rapeseed quality through biological and chemical methods, complementing their existing techniques. Ultimately, the researchers plan to leverage the established connection between rapeseed color and genetic information to grade rapeseed quality. This grading system would assist breeding scientists in cultivating high-quality rapeseed varieties, contributing to the development of superior rapeseed cultivars.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/agronomy14050941/s1>, Table S1: The hyperspectral reflectance and the corresponding RGB data.

Author Contributions: C.Z.: data curation, formal analysis, investigation, validation, writing—original draft, and writing—review and editing; X.Z.: conceptualization, project administration; F.W.: conceptualization, funding acquisition, methodology, and writing—review and editing; J.W.: software, validation, and writing—review and editing; Y.-G.W.: software, funding acquisition, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported partially by the Key Research Project of the Department of Education of Hunan Province (CN) (Grant No. 22A0135), the “Chunhui” Program Collaborative Scientific Research Project (202202004), the Research Project of the Department of Education of Hunan Province (CN) (Grant No. KN2022026), the Research Project of Educational Science Planning in Hunan Province (CN) (Grant No. 19JYKX007), and the Australian Research Council project (DP160104292).

Data Availability Statement: The data used in this work are openly available in the Supplementary Material.

Acknowledgments: The authors wish to thank the anonymous reviewers and the handling editor for their constructive comments and suggestions, which led to a great improvement in the presentation of this work.

Conflicts of Interest: The authors declare that they have no competing interests.

References

- Łopatyńska, A.; Wolko, J.; Bocianowski, J.; Cyplik, A.; Gacek, K. Statistical Multivariate Methods for the Selection of High-Yielding Rapeseed Lines with Varied Seed Coat Color. *Agriculture* **2023**, *13*, 992. [\[CrossRef\]](#)
- Rahman, M.; McVetty, P. A review of Brassica seed color. *Can. J. Plant Sci.* **2011**, *91*, 437–446. [\[CrossRef\]](#)
- Chen, B.; Heneen, W. Inheritance of seed colour in *Brassica campestris* L. and breeding for yellow-seeded *B. napus* L. *Euphytica* **1992**, *59*, 157–163. [\[CrossRef\]](#)
- Rahman, M. Production of yellow-seeded Brassica napus through interspecific crosses. *Plant Breed.* **2001**, *120*, 463–472. [\[CrossRef\]](#)
- Michalski, K. Seed color assessment in rapeseed seeds using Color and Near Infrared Reflectance Spectrometers. *Rośl. Oleist* **2009**, *30*, 119–132.
- Van Deynze, A.; Pauls, K. Seed colour assessment in Brassica napus using a Near Infrared Reflectance spectrometer adapted for visible light measurements. *Euphytica* **1994**, *76*, 45–51. [\[CrossRef\]](#)
- Velasco, L.; Fernández-Martínez, J.M.; De Haro, A. An efficient method for screening seed colour in Ethiopian mustard using visible reflectance spectroscopy and multivariate analysis. *Euphytica* **1996**, *90*, 359–363. [\[CrossRef\]](#)
- Tańska, M.; Ambrosewicz-Walacik, M.; Jankowski, K.; Rotkiewicz, D. Possibility use of digital image analysis for the estimation of the rapeseed maturity stage. *Int. J. Food Prop.* **2017**, *20*, S2379–S2394. [\[CrossRef\]](#)
- Tańska, M.; Rotkiewicz, D.; Kozirok, W.; Konopka, I. Measurement of the geometrical features and surface color of rapeseeds using digital image analysis. *Food Res. Int.* **2005**, *38*, 741–750. [\[CrossRef\]](#)
- Lu, Y.; Liu, X.; Liu, S.; Yue, Y.; Guan, C.; Liu, Z. A simple and rapid procedure for identification of seed coat colour at the early developmental stage of Brassica juncea and Brassica napus seeds. *Plant Breed.* **2012**, *131*, 176–179. [\[CrossRef\]](#)
- Pande, C.B.; Moharir, K.N. Application of hyperspectral remote sensing role in precision farming and sustainable agriculture under climate change: A review. In *Climate Change Impacts on Natural Resources, Ecosystems and Agricultural Systems*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 503–520.
- Moharram, M.A.; Sundaram, D.M. Land Use and Land Cover Classification with Hyperspectral Data: A comprehensive review of methods, challenges and future directions. *Neurocomputing* **2023**, *536*, 90–113. [\[CrossRef\]](#)
- Liu, F.; Wang, F.; Wang, X.; Liao, G.; Zhang, Z.; Yang, Y.; Jiao, Y. Rapeseed variety recognition based on hyperspectral feature fusion. *Agronomy* **2022**, *12*, 2350. [\[CrossRef\]](#)

14. Rybacki, P.; Niemann, J.; Bahcevandziev, K.; Durczak, K. Convolutional neural network model for variety classification and seed quality assessment of winter rapeseed. *Sensors* **2023**, *23*, 2486. [[CrossRef](#)] [[PubMed](#)]
15. Field, C.; Gamon, J.; Peñuelas, J. Remote Sensing of Terrestrial Photosynthesis. In *Ecophysiology of Photosynthesis*; Springer: Berlin/Heidelberg, Germany, 1995; pp. 511–527.
16. Gitelson, A.A.; Merzlyak, M.N.; Chivkunova, O.B. Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochem. Photobiol.* **2001**, *74*, 38–45. [[CrossRef](#)]
17. Gitelson, A.A.; Zur, Y.; Chivkunova, O.B.; Merzlyak, M.N. Assessing carotenoid content in plant leaves with reflectance spectroscopy. *Photochem. Photobiol.* **2002**, *75*, 272–281. [[CrossRef](#)] [[PubMed](#)]
18. Wang, J.; Xian, X.; Xu, X.; Qu, C.; Lu, K.; Li, J.; Liu, L. Genome-wide association mapping of seed coat color in *Brassica napus*. *J. Agric. Food Chem.* **2017**, *65*, 5229–5237. [[CrossRef](#)] [[PubMed](#)]
19. Gamon, J.; Surfus, J. Assessing leaf pigment content and activity with a reflectometer. *New Phytol.* **1999**, *143*, 105–117. [[CrossRef](#)]
20. Kim, M.S.; Daughtry, C.; Chappelle, E.; McMurtrey, J.; Walthall, C. The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation (A par). In Proceedings of the 6th International Symposium on Physical Measurements and Signatures in Remote Sensing, CNES, Val d'Isère, France, 17–21 January 1994.
21. Gamon, J.; Penuelas, J.; Field, C. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sens. Environ.* **1992**, *41*, 35–44. [[CrossRef](#)]
22. Gamon, J.; Serrano, L.; Surfus, J. The photochemical reflectance index: An optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. *Oecologia* **1997**, *112*, 492–501. [[CrossRef](#)]
23. Daughtry, C.S.; Walthall, C.; Kim, M.; De Colstoun, E.B.; McMurtrey Iii, J. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sens. Environ.* **2000**, *74*, 229–239. [[CrossRef](#)]
24. Matoušková, E.; Kovářová, K.; Cihla, M.; Hodač, J. Monitoring biological degradation of historical stone using hyperspectral imaging. *Eur. J. Remote Sens.* **2023**, 2220565. [[CrossRef](#)]
25. Deng, Y.J.; Yang, M.L.; Li, H.C.; Long, C.F.; Fang, K.; Du, Q. Feature Dimensionality Reduction with L₂, p-Norm-Based Robust Embedding Regression for Classification of Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5509314. [[CrossRef](#)]
26. Zhao, Y.; Kwon, K.C.; Piao, Y.L.; Jeon, S.H.; Kim, N. Depth-layer weighted prediction method for a full-color polygon-based holographic system with real objects. *Opt. Lett.* **2017**, *42*, 2599–2602. [[CrossRef](#)] [[PubMed](#)]
27. Sağlam, A.; Baykan, N.A. A new color distance measure formulated from the cooperation of the Euclidean and the vector angular differences for lidar point cloud segmentation. *Int. J. Eng. Geosci.* **2021**, *6*, 117–124. [[CrossRef](#)]
28. Long, X.; Sun, J. Image segmentation based on the minimum spanning tree with a novel weight. *Optik* **2020**, *221*, 165308. [[CrossRef](#)]
29. Yang, Z.; Wang, Y.; Yang, Z. Vector-angular distance color difference formula in RGB color space. *Comput. Eng. Appl.* **2010**, *46*, 154–156.
30. Liu, F. Identification of Rapeseed Variety and Modeling of Fatty Acid Content Using Hyperspectral Features Fusion. Ph.D. Thesis, Hunan Agricultural University, Changsha, China, 2021.
31. Lacasa, L.; Luque, B.; Ballesteros, F.; Luque, J.; Nuno, J.C. From time series to complex networks: The visibility graph. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4972–4975. [[CrossRef](#)]
32. Ahmadlou, M.; Adeli, H. Visibility graph similarity: A new measure of generalized synchronization in coupled dynamic systems. *Phys. D Nonlinear Phenom.* **2012**, *241*, 326–332. [[CrossRef](#)]
33. Lacasa, L.; Luque, B.; Luque, J.; Nuno, J.C. The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion. *Europhys. Lett.* **2009**, *86*, 30001. [[CrossRef](#)]
34. Zhang, R.; Ashuri, B.; Shyr, Y.; Deng, Y. Forecasting construction cost index based on visibility graph: A network approach. *Phys. A Stat. Mech. Appl.* **2018**, *493*, 239–252. [[CrossRef](#)]
35. Silva, V.F.; Silva, M.E.; Ribeiro, P.; Silva, F. Time series analysis via network science: Concepts and algorithms. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2021**, *11*, e1404. [[CrossRef](#)]
36. Wen, T.; Chen, H.; Cheong, K.H. Visibility graph for time series prediction and image classification: A review. *Nonlinear Dyn.* **2022**, *110*, 2979–2999. [[CrossRef](#)] [[PubMed](#)]
37. Lan, X.; Mo, H.; Chen, S.; Liu, Q.; Deng, Y. Fast transformation from time series to visibility graphs. *Chaos Interdiscip. J. Nonlinear Sci.* **2015**, *25*, 083105. [[CrossRef](#)] [[PubMed](#)]
38. Silva, T.C.; Zhao, L. *Machine Learning in Complex Networks*; Springer: Berlin/Heidelberg, Germany, 2016.
39. Tian, M.; Feng, J.; Rivard, B.; Zhao, C. A method to compute the n-dimensional solid spectral angle between vectors and its use for band selection in hyperspectral data. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *50*, 141–149. [[CrossRef](#)]
40. Leung, C.; Chau, H. Weighted assortative and disassortative networks model. *Phys. A Stat. Mech. Appl.* **2007**, *378*, 591–602. [[CrossRef](#)]
41. Almog, A.; Shmueli, E. Structural entropy: Monitoring correlation-based networks over time with application to financial markets. *Sci. Rep.* **2019**, *9*, 10832. [[CrossRef](#)]
42. Girvan, M.; Newman, M.E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826. [[CrossRef](#)]
43. Zhang, Z.; Wang, F.; Shen, L.; Xie, Q. Multiscale time-lagged correlation networks for detecting air pollution interaction. *Phys. A Stat. Mech. Appl.* **2022**, *602*, 127627. [[CrossRef](#)]

44. Cai, M.; Du, H.; Feldman, M.W. A new network structure entropy based on maximum flow. *Acta Phys. Sin.* **2014**, *63*, 102–112.
45. Kok, Z.H.; Shariff, A.R.M.; Alfatni, M.S.M.; Khairunniza-Bejo, S. Support vector machine in precision agriculture: A review. *Comput. Electron. Agric.* **2021**, *191*, 106546. [[CrossRef](#)]
46. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1999.
47. Nasien, D.; Yuhaniz, S.S.; Haron, H. Statistical learning theory and support vector machines. In Proceedings of the 2010 Second International Conference on Computer Research and Development, Kuala Lumpur, Malaysia, 7–10 May 2010; pp. 760–764.
48. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
49. Chung, C.C.; Su, E.C.Y.; Chen, J.H.; Chen, Y.T.; Kuo, C.Y. XGBoost-based simple three-item model accurately predicts outcomes of acute ischemic stroke. *Diagnostics* **2023**, *13*, 842. [[CrossRef](#)] [[PubMed](#)]
50. Gertz, M.; Große-Butenuth, K.; Junge, W.; Maassen-Francke, B.; Renner, C.; Sparenberg, H.; Krieter, J. Using the XGBoost algorithm to classify neck and leg activity sensor data using on-farm health recordings for locomotor-associated diseases. *Comput. Electron. Agric.* **2020**, *173*, 105404. [[CrossRef](#)]
51. Lundberg, S.M.; Erion, G.; Chen, H.; De Grave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [[CrossRef](#)]
52. Obsie, E.Y.; Qu, H.C.; Drummond, F. Wild blueberry yield prediction using a combination of computer simulation and machine learning algorithms. *Comput. Electron. Agric.* **2020**, *178*, 105778. [[CrossRef](#)]
53. Gomes, W.P.C.; Gonçalves, L.; da Silva, C.B.; Melchert, W.R. Application of multispectral imaging combined with machine learning models to discriminate special and traditional green coffee. *Comput. Electron. Agric.* **2022**, *198*, 107097. [[CrossRef](#)]
54. Babajide Mustapha, I.; Saeed, F. Bioactive molecule prediction using extreme gradient boosting. *Molecules* **2016**, *21*, 983. [[CrossRef](#)] [[PubMed](#)]
55. Xue, J.; Shen, B. A novel swarm intelligence optimization approach: Sparrow search algorithm. *Syst. Sci. Control. Eng.* **2020**, *8*, 22–34. [[CrossRef](#)]
56. Jiang, S.; Wang, F.; Shen, L.; Liao, G. Local detrended fluctuation analysis for spectral red-edge parameters extraction. *Nonlinear Dyn.* **2018**, *93*, 995–1008. [[CrossRef](#)]
57. Jiang, S.; Wang, F.; Shen, L.; Liao, G.; Wang, L. Extracting sensitive spectrum bands of rapeseed using multiscale multifractal detrended fluctuation analysis. *J. Appl. Phys.* **2017**, *121*, 104702. [[CrossRef](#)]
58. Liu, F.; Wang, F.; Liao, G.; Lu, X.; Yang, J. Prediction of oleic acid content of rapeseed using hyperspectral technique. *Appl. Sci.* **2021**, *11*, 5726. [[CrossRef](#)]
59. Holme, P.; Park, S.M.; Kim, B.J.; Edling, C.R. Korean university life in a network perspective: Dynamics of a large affiliation network. *Phys. A Stat. Mech. Appl.* **2007**, *373*, 821–830. [[CrossRef](#)]
60. Saire, J.C.; Zhao, L. Complex Network-Based Data Classification Using Minimum Spanning Tree Metric and Optimization. In Proceedings of the 2023 International Joint Conference on Neural Networks (IJCNN), Gold Coast, Australia, 18–23 June 2023; pp. 1–7.
61. Chen, J.; Luo, D.L.; Mu, F.X. An improved ID3 decision tree algorithm. In Proceedings of the 2009 4th International Conference on Computer Science & Education, Nanning, China, 25–28 July 2009; pp. 127–130.
62. Kumari, P.; Gangwar, H.; Kumar, V.; Jaiswal, V.; Gahlaut, V. Crop Phenomics and High-Throughput Phenotyping. In *Digital Agriculture: A Solution for Sustainable Food and Nutritional Security*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 391–423.
63. Rehman, T.U.; Ma, D.; Wang, L.; Zhang, L.; Jin, J. Predictive spectral analysis using an end-to-end deep model from hyperspectral images for high-throughput plant phenotyping. *Comput. Electron. Agric.* **2020**, *177*, 105713. [[CrossRef](#)]
64. Bohorquez, J.C.S.; Reyes, J.M.R.; Alama, W.I.; Alama, C.C. New Hyperspectral Index for Determining the State of Fermentation in the Non-Destructive Analysis for Organic Cocoa Violet. *IEEE Lat. Am. Trans.* **2018**, *16*, 2435–2440. [[CrossRef](#)]
65. Dutta, A.; Tyagi, R.; Chattopadhyay, A.; Chatterjee, D.; Sarkar, A.; Lall, B.; Sharma, S. Early detection of wilt in *Cajanus cajan* using satellite hyperspectral images: Development and validation of disease-specific spectral index with integrated methodology. *Comput. Electron. Agric.* **2024**, *219*, 108784. [[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.