

Research Article

An Incremental Self-Adaptive Wood Species Classification Prototype System

Peng Zhao , Zhen-Yu Li , and Yue Li

Information and Computer Engineering College, Northeast Forestry University, Harbin City 150040, China

Correspondence should be addressed to Peng Zhao; bit_zhao@aliyun.com

Received 17 August 2019; Accepted 30 October 2019; Published 15 November 2019

Academic Editor: Hicham Fenniri

Copyright © 2019 Peng Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The present wood species classification systems can usually process the limited wood species quantity. We propose a novel incremental self-adaptive wood species classification system to solve the above-mentioned issue. The visible/near-infrared (VIS/NIR) spectrometer is used to pick up the spectral curves of wood samples for the subsequent wood species classification. First, when new wood samples of unknown wood species are added, they are classified as an unknown category by our one-class classifier, Support Vector Data Description (SVDD), while the existent wood species are classified as a known category by the SVDD. Second, the wood samples of known species are sent into the BP neural network for subsequent wood species classification. Third, the new wood samples of unknown species are sent into the Clustering by Fast Search and Find of Density Peaks (CFSFDP) algorithm for the unsupervised clustering, and the clustering result is evaluated by the internal and external norms. Last, if one cluster of one unknown species has an adequate amount of wood samples, these wood samples are removed and identified by human experts or other schemes to ensure to get the correct wood species name. Then, these wood samples are considered as a new known species and are sent into the classifiers, SVDD and BP neural network, to train them again. Experiments on 13 wood species prove the effectiveness of our prototype system with an overall classification accuracy of above 95%.

1. Introduction

Wood species recognition has been investigated for some years since different wood species have different physical and chemical properties with a different price. Many wood species classification systems are used for automatic processing by use of sensors and computers. In terms of the used sensors, spectrum-based [1–3], chemometric-based [4], and image-based systems [5–7] are usually investigated. The wood spectral analysis scheme usually deals with the 1D spectral reflectance ratio curves for wood species recognition, which has a low computational complexity and thus a high processing speed.

However, most present wood species classification systems usually process the limited wood species quantity. For example, Yusof et al. proposed a kernel-genetic nonlinear feature selection scheme for 52 tropical wood species recognition [8]. Ibrahim et al. proposed a fuzzy preclassifier to classify 48 tropical wood species into 4 broad categories, and

then they used a Support Vector Machine (SVM) classifier in each broad category to further determine wood species of the detected wood sample [9]. They pointed out that one advantage is that when a new wood species is added into the system, only the SVM classifier of one broad category requires to be retrained instead of the whole system. But this two-level classification system may not recognize the unknown new wood species automatically [9].

In this paper, we propose a novel prototype system which can process the incremental wood species quantity. Our prototype system is illustrated in Figure 1, which consists of 3 parts and one renewal procedure mostly. The 1st part is a one-class classifier SVDD, which can classify wood samples into two categories (i.e., the known wood species which are included in the wood training dataset, and the unknown species which are not included in the training dataset). The 2nd part is a BP neural network, which can classify the known wood species samples into a specific species. The 3rd part is an unsupervised clustering analyzer

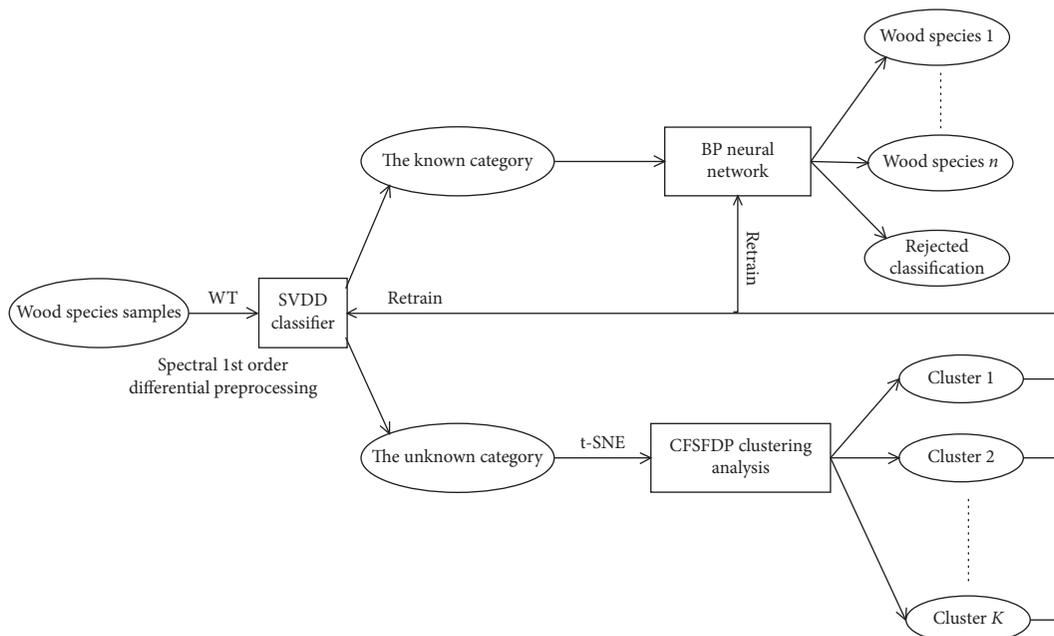


FIGURE 1: Our prototype system's structure graph.

TABLE 1: Wood species samples.

Category	English name	Latin name	Wood sample quantity	
			Training dataset	Testing dataset
The known category	Bubinga	<i>Guibourtia tessmannii</i>	60	20
	Merbau	<i>Intsia bijuga</i>	60	20
	Burma teak	<i>Tectona grandis L.F.</i>	60	20
	African padauk	<i>Pterocarpus soyauxii</i>	60	20
	Casla	<i>Terminalia neotaliala Capuron</i>	60	20
The unknown category	Birch	<i>Betula platyphylla</i>		55
	Mongolian pine	<i>Pinus sylvestris</i>		55
	White pine	<i>Picea jezoensis</i>		55
	Larch	<i>Larix gmelinii</i>		55
	Pometia	<i>Pometia tomentosa</i>		55
	Indian teak	<i>Tectona grandis L.F.</i>		55
	Siam teak	<i>Tectona grandis L.F.</i>		55
	Thailand teak	<i>Tectona grandis L.F.</i>		55

which can classify the unknown wood species samples into several clusters by the use of CFSFDP algorithm. Moreover, an updating procedure is performed to increase the wood species quantity in the training dataset. If one cluster of one unknown species has an adequate amount of wood samples, these wood samples are removed from clusters and identified by human experts or other schemes to ensure to get the correct wood species name. Then, these wood samples are considered as a new known species and are sent into the classifiers, SVDD and BP neural network, to retrain them. In this way, our system can recognize more and more wood species.

2. Materials and Methods

2.1. Wood Samples. Our wood samples consist of 13 wood species, as illustrated in Table 1. Every wood sample is

processed into a 2 cm × 2 cm × 3 cm block. The known wood species category consists of 5 species, while the unknown category consists of 8 species. The training dataset in the known category is used to train the one-class classifier, SVDD, while the testing dataset in the known category and all wood samples in the unknown category are used to test SVDD. Moreover, the unknown category is used for unsupervised clustering analysis.

The intraclass spectral variations of wood samples may be caused by the wood age, cutting position, and geographic place. For each wood species, our wood samples are taken from different individual trees at random position to include these intraclass differences. In this way, the reliability of our proposed scheme can be ensured.

2.2. Spectral Data Acquisition. The spectrometer used here is an Ocean Optics USB2000 + microfiber spectrometer

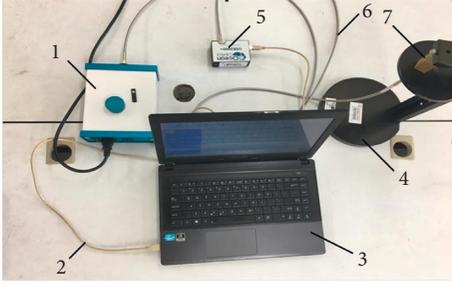


FIGURE 2: Spectral data acquisition system's structure graph: (1) halogen lamp; (2) universal serial bus; (3) computer; (4) holder; (5) spectrometer; (6) optic fiber; (7) test sample.

of Ocean America. The visible/near-infrared (VIS/NIR) spectra of wood samples are picked up by this spectrometer system which consists of a halogen lamp, a universal serial bus, a computer, a customized holder, a fiber spectrometer (Ocean Optics Inc, USA), as well as a piece of optics fiber, as illustrated in Figure 2. The spectral acquisition is performed in a controlled environment at a temperature of 20°C and in diffuse reflectance. The spectral band covers a range of 340~1027 nm with a spectral resolution of 0.3 nm approximately. Spectrometer parameter setting, spectral data acquisition, and storing are implemented by the Spectrasuite standard software. The integration time is 300 ms, and the average number of every point measurements is 900. Before spectral acquisition, reference spectra and dark spectra are measured and stored. The reference spectra are acquired with a standard white panel of known 100% reflectance, and the dark spectra are acquired by blocking the radian. The wood sample to be measured is closely fixed on the holder, and the reflectance spectra of each wood sample are acquired. For each wood sample, four diffuse reflectance spectrums are measured randomly at different locations across the wood sample's cross section, and the mean spectrum is computed and stored for wood spectral analysis and species classification.

Considering the spectrometer's price, we choose the VIS/NIR spectrometer for spectral acquisition. As we all know, the spectral data in visible band may be variable because the wood sample's color may change with the environmental variation. Therefore, we must keep the environment stable during the spectral acquisition and wood sample's preservations. In addition, 780~900 nm is a near-infrared band, which can characterize the different wood properties to some extents.

2.3. Spectral Dimension Reduction. For each spectral curve, the VIS/NIR band includes thousands of spectral data so that the dimension of acquired spectral curve is very large. In fact, each spectral curve contains redundant information, which will reduce the wood species classification speed and accuracy. In this paper, Principal Component Analysis

(PCA) and Wavelet Transform (WT) are used and compared for spectral dimension reduction in our one-class classifier, SVDD.

PCA is a transformation scheme in multivariate statistics, which is one of the most commonly used spectral feature extraction schemes. WT is an efficient data compression tool that decomposes the signal into a series of wavelet functions with time-frequency analysis capabilities. WT can compress the spectral data, filter the spectral noises, and extract useful spectral features. The spectra are divided into low-frequency and high-frequency parts. The low-frequency coefficient contains most useful information, while the high-frequency coefficient contains redundant information such as noises. Therefore, we replace the original spectral data with low-frequency wavelet coefficients.

Moreover, in our unsupervised clustering stage, the t-distributed stochastic neighbor embedding (t-SNE) algorithm is used for the secondary spectral dimension reduction. This algorithm is a nonlinear dimension reduction which is an improved version of the original SNE algorithm [10]. Here, the symbol "t" in the t-SNE algorithm represents the t distribution, and the freedom of t distribution is 1 in this paper.

2.4. One-Class Classifier: SVDD. In the SVDD, all objects in known category are considered as an ensemble to construct an optimal hypersphere which can almost consist of all objects $X = \{x_i\}_{i=1}^n$ in known category so that all objects in the known category are separated from all objects in the unknown category. The optimal hypersphere construction can be transformed to the following optimization issue [11]:

$$\begin{cases} \min_{R,a,\xi_i} \left[R^2 + C \sum_{i=1}^n \xi_i \right], \\ \text{s.t.} \quad \|x_i - a\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0, \end{cases} \quad (1)$$

where a and R represent the hypersphere's center and radius, and the slack factor ξ_i is introduced to increase the classification robustness with the penalty factor C . This optimization issue can be further transformed to the following Lagrange extreme issue:

$$L = R^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i (R^2 \xi_i - \|x_i - a\|^2) - \sum_{i=1}^n \beta_i \xi_i, \quad (2)$$

where α_i and β_i are Lagrange coefficients and $\alpha_i \geq 0$ and $\beta_i \geq 0$, $i = 1, 2, \dots, n$. By minimizing equation (2), we can obtain the following dual equation:

$$\begin{cases} L = \sum_{i=1}^n \alpha_i (x_i, x_i) - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j (x_i, x_j) \\ \text{s.t.} \quad 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^n \alpha_i = 1. \end{cases} \quad (3)$$

When the input objects are not linearly divisible, a nonlinear mapping $\Phi : X \rightarrow F$ can be used to map the

TABLE 2: Class label and code for 5 wood species in known category.

Class label	Wood species	Code
1	Bubinga	[1 0 0 0 0]
2	Merbau	[0 1 0 0 0]
3	Burma teak	[0 0 1 0 0]
4	African padauk	[0 0 0 1 0]
5	Casla	[0 0 0 0 1]

original low-dimensional space into the high-dimensional space. In this case, the inner product operation (x_i, x_j) in equation (3) can be substituted by the kernel function $K(x_i, x_j)$:

$$\begin{cases} L(R, a, \xi, \alpha, \beta) = \sum_{i=1}^n \alpha_i (x_i, x_i) - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(x_i, x_j), \\ \text{s.t. } 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^n \alpha_i = 1. \end{cases} \quad (4)$$

The standard quadratic programming algorithm can be used to solve the optimal solution α_i^* . Suppose y is a support vector whose corresponding specimen lies on the surface of the hypersphere, then its center and radius can be represented as follows:

$$\begin{cases} a = \sum_{i=1}^n \alpha_i^* \varphi(x_i), \\ R^2 = K(y, y) - 2 \sum_{i=1}^n \alpha_i^* (y, x_i) + \sum_{i=1}^n \sum_{j=1}^n \alpha_i^* \alpha_j^* K(x_i, x_j). \end{cases} \quad (5)$$

For a new testing specimen z , the distance between z and the hypersphere's center can be represented as follows:

$$D^2 = K(z, z) - 2 \sum_{i=1}^n \alpha_i^* K(z, x_i) + \sum_{i=1}^n \sum_{j=1}^n \alpha_i^* \alpha_j^* K(x_i, x_j). \quad (6)$$

If $D^2 \leq R^2$, this specimen z is considered to belong to the known category; otherwise, it belongs to the unknown category.

2.5. Wood Species Recognition for the Known Category. In this study, there are 5 wood species in the known category, as illustrated in Table 1. The BP neural network is used for the species classification of these 5 wood species. These 5 species are marked with different class labels "1, 2, ..., 5." The network's input is the spectral data after spectral dimension reduction processing, and its expected output for 5 wood species is the unit matrix code, as illustrated in Table 2. The predictive value of each node P_j ($j = 1, 2, \dots, 5$) is transformed into the binary code Q_j with the following rules. The maximum predictive value is transformed into 1, while the other 4 predictive values are transformed into 0. Then, the sum of squares of errors (SSE) is calculated for each sample's predictive values and binary values:

$$\text{SSE}(i) = \sum_{j=1}^l (P_j^{(i)} - Q_j^{(i)})^2, \quad i = 1, 2, \dots, N, \quad (7)$$

here, N is the wood sample number, and $l = 5$. After the network's training is over, a threshold T should be set as follows. If $\text{SSE}(i) \leq T$, this sample is considered as in the known category; otherwise, it is in the unknown category.

2.6. Unsupervised Clustering Analysis for the Unknown Category

2.6.1. CFSFDP. In our clustering analysis, an accurate CFSFDP algorithm [12] is used. In this algorithm, the clustering centers are firstly determined. The clustering centers have relatively high local density, and they are relatively far from those points with higher local density. Define a dataset $S = \{x_i\}_{i=1}^n$, and we need to compute the local density ρ_i and distance δ_i for every sample X_i . As for local density ρ_i , two computation equations can be chosen as follows:

$$\text{cutoff kernel: } \rho_i = \sum_j x(d_{ij} - d_c), \quad (8)$$

$$\text{Gaussian kernel: } \rho_i = \sum_j e^{-(d_{ij}/d_c)}. \quad (9)$$

Here, d_{ij} is the distance between samples i and j ; d_c is the cutoff distance which is determined by human experts; and $\chi(x)$ is a 0-1 function as follows. Therefore, the ρ_i calculated by equation (8) is a discrete value, while that by equation (9) is a continuous value. We choose Gaussian kernel to compute the local density.

$$\chi(x) = \begin{cases} 0, & x \geq 0, \\ 1, & x < 0, \end{cases} \quad (10)$$

The distance δ_i is the distance between sample X_i and its nearest sample X_j among those samples with higher local density.

$$\delta_i = \min_{j: \rho_j > \rho_i} (d_{ij}). \quad (11)$$

Therefore, one sample may be a clustering center when it has a relatively large ρ_i and δ_i . Therefore, the possibility of one sample being a clustering center can be computed by γ_i .

$$\gamma_i = \rho_i * \delta_i. \quad (12)$$

Once the clustering centers are determined, one sample is classified into the cluster which consists of the nearest neighbor of this sample with a higher local density.

2.6.2. Evaluation Criterion

- (1) The performance of one-class classifier SVDD is evaluated by the classification precision, recall, and F value as follows:

$$\left\{ \begin{array}{l}
 \text{precision} = \frac{\text{the number of correctly classified samples in the known category and in the unknown category}}{\text{the overall sample number}}, \\
 \text{recall} = \frac{\text{the number of correctly classified samples in the known category}}{\text{the number of samples in the known category}}, \\
 F = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}.
 \end{array} \right. \quad (13)$$

- (2) In the CFSFDP algorithm [12], the clustering centers require to be chosen manually in the decision chart by human experts. In order to fulfill the automatic selection of clustering centers, we make some improvements on this algorithm by the use of internal criteria to evaluate the clustering results. As we know, the goal of clustering analysis is to increase the intraclass similarity and the interclass difference as much as possible. This goal is considered generally by internal criteria which take into account the self-feature of processed dataset. In this work, the CH and DB criteria [13, 14] are used here.

$$\text{CH} = \frac{N - k}{k - 1} \times \frac{\text{BGSS}}{\text{WGSS}} \quad (14)$$

Here, N is the sample number, k is the cluster number, BGSS is the trace of interclass divergence matrix, and WGSS is the trace of intraclass divergence matrix. Therefore, a high CH value indicates a good clustering result.

$$\text{DB}(k) = \frac{1}{k} \sum_{i=1}^k \max_{j=1 \sim k, j \neq i} \left(\frac{W_i + W_j}{C_{ij}} \right). \quad (15)$$

Here, C_{ij} is the distance between the center of cluster C_i and that of clusters C_j and W_i represent the mean distance from all its samples to its cluster center for the cluster C_i . Therefore, a low DB value indicates a good clustering result.

- (3) The clustering correct ratio (CCR) is used to evaluate the clustering accuracy.

$$\text{CCR}(\%) = \frac{\text{correct clustering times}}{\text{overall clustering times}} \quad (16)$$

- (4) When the cluster number is determined, external criteria are used for clustering evaluation by use of each sample's true class. Since the calculated clusters may be different from the true clusters, we suppose the true clusters of dataset with N samples as $P = \{P_1, P_2, \dots, P_k\}$ and the calculated clusters as $C = \{C_1, C_2, \dots, C_m\}$. By the pairwise combination, we can get the M sample pairs ($M = C_N^2$). Then, the sample pair's clustering is compared in P and C . First, we define "a" to represent the number of sample pairs that belongs to the same cluster in both

P and C . Second, we define "b" to represent the number of sample pairs that belongs to the same cluster only in P . Third, "c" represents the number of sample pairs that belongs to the same cluster only in C . Last, "d" represents the number of sample pairs that belongs to different clusters in both P and C . Finally, the 4 external criteria (i.e., Rand, Adjusted Rand, Jaccard, and Fowlkes–Mallows) [15] are calculated as follows.

$$\text{Rand} = \frac{a + b}{M}, \quad (17)$$

$$\text{AR} = \frac{2(Ma - (a + b)(a + c))}{M(2a + b - c) - 2(a + b)(a + c)}, \quad (18)$$

$$J = \frac{a}{a + b + c}, \quad (19)$$

$$\text{FM} = \sqrt{\frac{a}{a + b} \frac{a}{a + c}} \quad (20)$$

2.7. Incremental Updating Procedure. In this section, an updating procedure is performed to increase the wood species quantity in the training dataset so that our system can recognize more and more wood species. If one cluster of an unknown species in the unknown category has an adequate amount of wood samples (i.e., approximately 50 samples per wood species), these wood samples are removed from clusters and identified by human experts or other schemes to ensure to get the correct wood species name. Then, these wood samples are considered as a new known species and are sent into the classifiers, SVDD and BP neural network, to retrain them. By this way, our system can classify this new wood species as in the known category with SVDD and then recognize this wood species correctly with the BP neural network.

Moreover, two issues should be considered. First, in the one-class classification by the use of SVDD, some wood samples in the known category may be falsely classified as in the unknown category. These wood samples will be put into some clusters later in the subsequent unsupervised clustering processing. Therefore, if one cluster of an unknown species is removed from clusters and identified by human experts or other schemes to ensure to get the correct wood species name, all wood samples of this cluster must be processed one by one to find and discard those samples

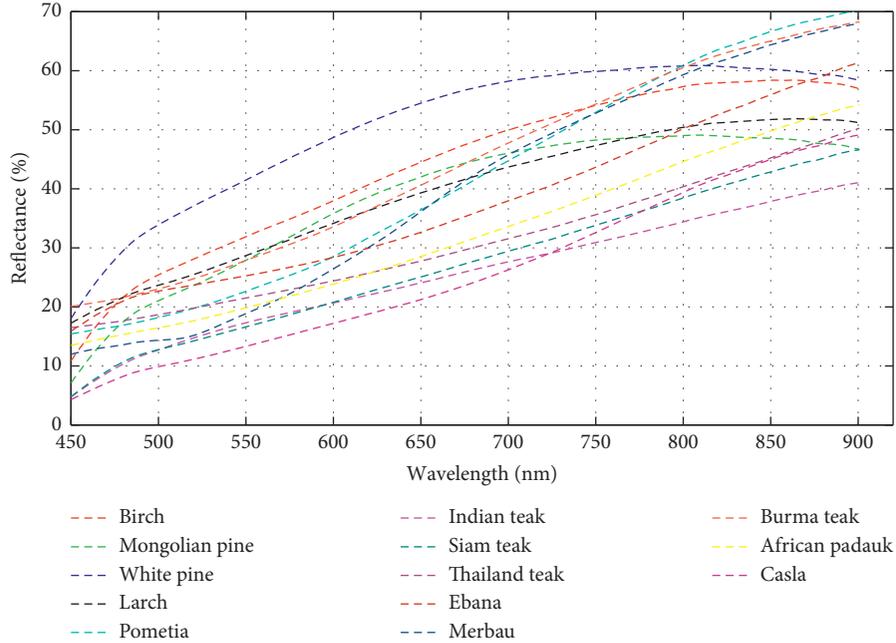


FIGURE 3: Spectral curve graph of our 13 wood species.

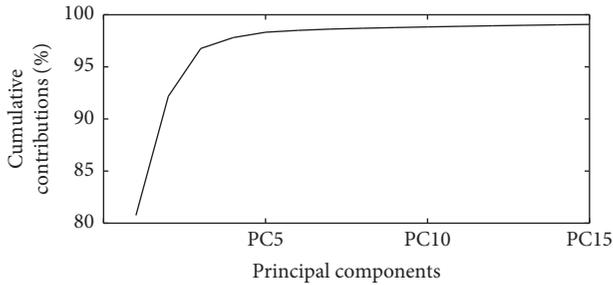


FIGURE 4: The cumulative contribution ratio graph of PCA.

falsely classified by SVDD possibly. Second, in the one-class classification, some wood samples in the unknown category may be falsely classified as in the known category on the contrary. These wood samples will be sent into the BP neural network for further species recognition. The ideal classification output should be “rejected classification.”

3. Results and Discussions

3.1. Spectral Preprocessing. The acquired spectral curve has noises in the beginning and terminal bands due to the system error. Therefore, these noisy bands should be deleted, and the spectral curve in the band of 450 nm~900 nm is remained for the model training and testing. Figure 3 illustrates spectral curves of our 13 wood species. In this section, the spectral first-order differential operation is used to overcome the random noises, baseline drift, and light-scattering effect.

3.2. SVDD Classification. The WT and PCA are used and compared for spectral dimension reduction. As for WT, the sym4 wavelet basis is used, and the preprocessed spectral

TABLE 3: One-class classification of SVDD with optimal parameters and WT/PCA.

Methods	Precision	Recall	F	C	σ
WT-SVDD	0.9963	0.98	0.9881	0.17	18
PCA-SVDD	0.9241	0.97	0.9465	0.06	18

curve is decomposed into 5 levels. Therefore, the original 1328D spectral curve is reduced to 48D, and the low frequency wavelet coefficient is remained and sent into the SVDD classifier. As for PCA, the cumulative contribution ratio (%) reaches to 99% for the first 15 principal components, which can represent the original spectral curve’s basic information. Therefore, the first 15 principal components are remained for SVDD classification, as illustrated in Figure 4.

In our one-class classification experiment, the Gaussian kernel function is used in $K(x_i, x_j)$, and the hypersphere’s boundary shape is influenced by the penalty factor C and Gaussian kernel parameter σ . When σ is fixed, the number of support vectors is influenced by C , but to some extents, the influence of C on the hypersphere’s boundary shape is relatively small. When C is fixed, a small σ will produce a tight boundary shape and thus an overfit, while a large σ will produce a slack boundary shape and thus an underfit. Therefore, the optimal C and σ should be acquired before the SVDD classification. Table 3 illustrates the SVDD classification results for the testing dataset of the known category and all datasets of the unknown category by use of the optimal C and σ . We can see that WT’s performance is superior to PCA.

3.3. Wood Species Classification in Known Category. The BP neural network’s structure is as follows. Its input layer has 48

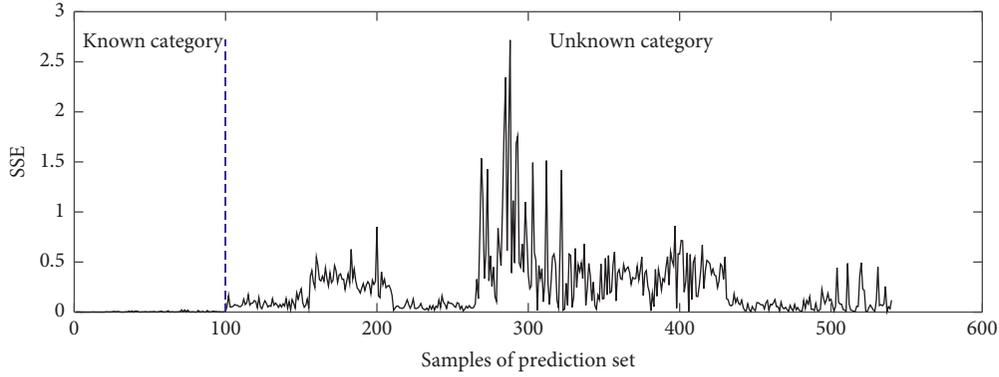
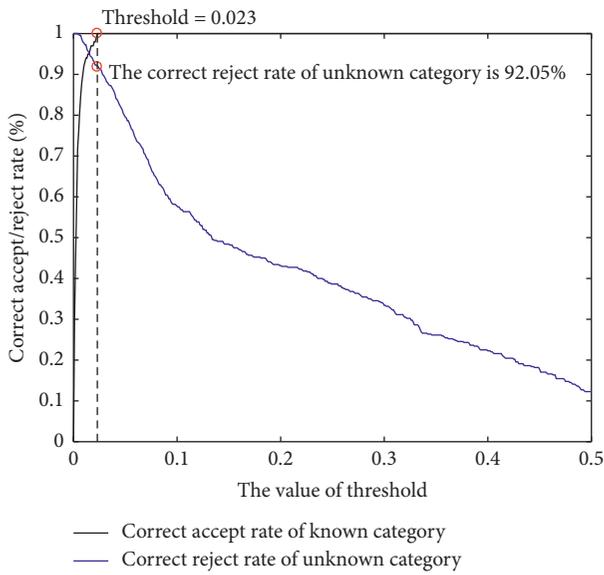


FIGURE 5: SSE comparison for the known category and unknown category.

FIGURE 6: The relationship between threshold T and correct accept/reject rate.

nodes, while the output layer has 5 nodes. The hidden layer has 27 nodes that are determined by multiple tests. The delivery function between the input layer and hidden layer is selected as “tansig,” while that between the hidden layer and output layer is “purelin.” The iteration times are 500, with an error target of 0.0002 and a learning rate of 0.01. After network training, this neural network is used for wood species recognition for samples in the known category which may include some samples falsely classified as in the known category by SVDD possibly.

To test neural network’s classification performance, the testing dataset in the known category ($20 \times 5 = 100$ samples) and all samples in the unknown category ($55 \times 8 = 440$ samples) are sent into the network. Figure 5 illustrates SSE for all wood samples, where we can see that SSE of the testing dataset in the known category is much smaller than that of all samples in the unknown category. Therefore, an appropriate threshold T can be used to separate the known category from the unknown category. Figure 6 illustrates the relationship between threshold T and correct accept/reject rate. We can see

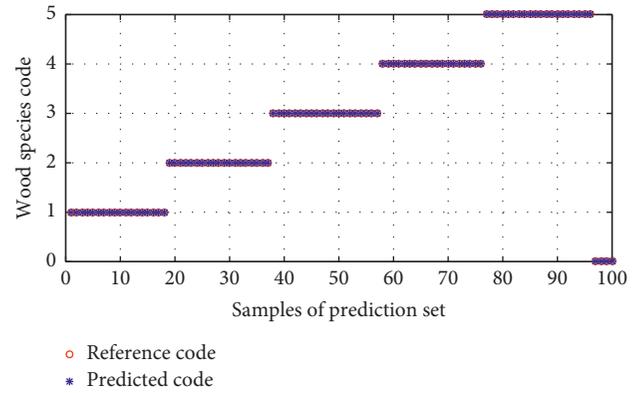


FIGURE 7: Classification result of the BP neural network.

that when $T=0.023$, the correct accept rate for the known category is 100%, while the correct reject rate for the unknown category is 92.05%. However, the correct classification rate is less than 74% (i.e., $100 \div (100 + 35) = 74\%$).

In our practical system, there are 100 samples in testing datasets of 5 wood species in the known category (Table 1). In the SVDD classification test, 4 samples in the known category are falsely classified as in the unknown category, and again 4 samples in the unknown category are falsely classified as in the known category. The BP network’s classification result is illustrated in Figure 7 with $T=0.023$, where we can see that the correct classification rate for the known category is 96% (i.e., all these 96 samples are classified as correct wood species), and the remaining 4 samples are correctly rejected so that the correct reject rate for the unknown category is 100%. Therefore, in view of the correct classification rate for the known category and the correct reject rate for the unknown category, the performance of our SVDD plus BP network is much better than that of the BP network alone.

3.4. Unsupervised Clustering in Unknown Category. First, the t-SNE algorithm is used for the spectral secondary dimension reduction after WT. The dimension reduction results of t-SNE, PCA, Sammon Mapping (SM) [16], Isometric Feature Mapping (IFM) [17], Laplacian Eigenmapping

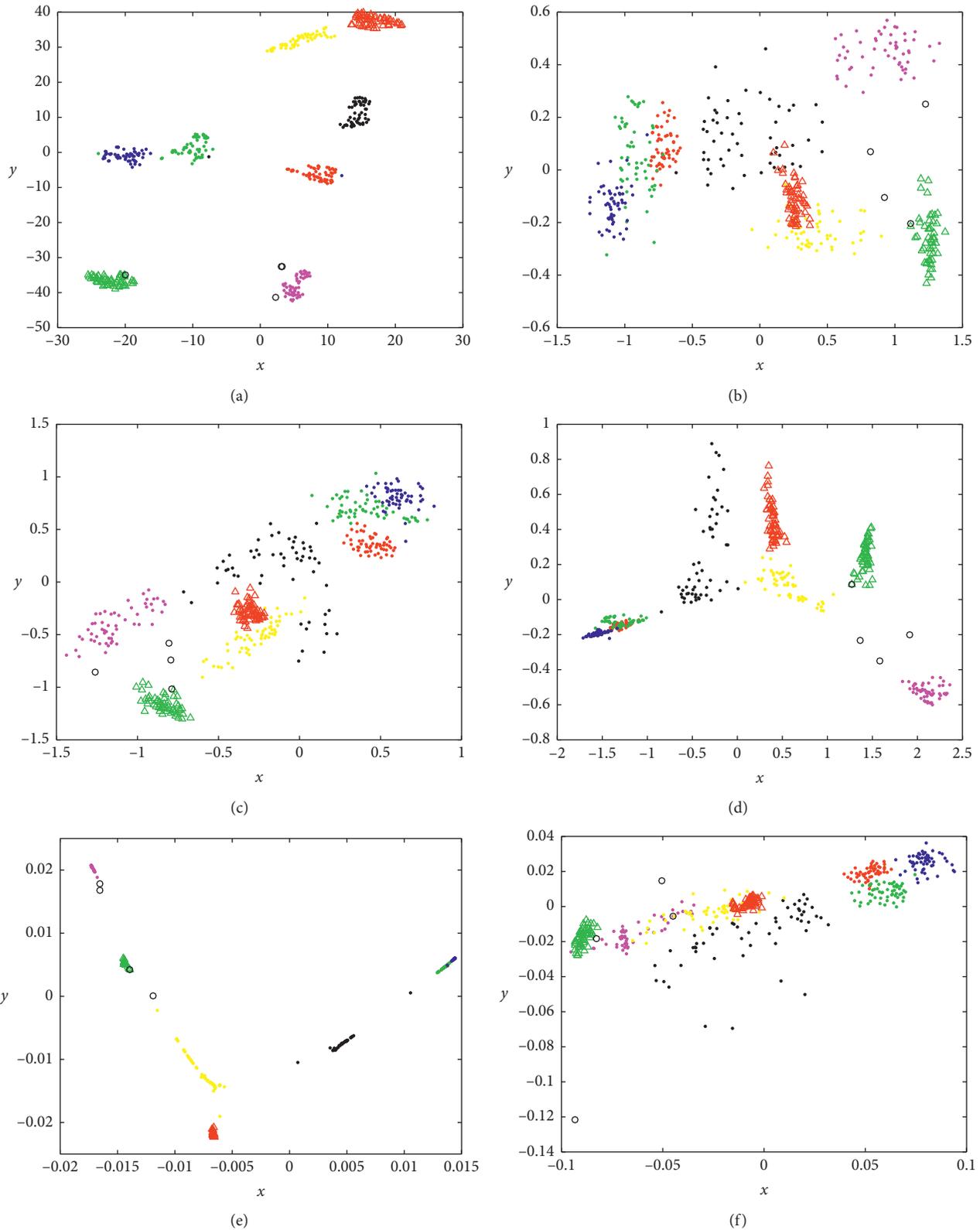


FIGURE 8: Comparison of secondary dimension reduction results: (a) t-SNE; (b) PCA; (c) SM; (d) IFM; (e) LE; and (f) LDA.

(LE) [18], and Linear Discriminator Analysis (LDA) [19] are compared in Figure 8. In Figure 8, x -axis and y -axis refer to the first two dimensions after dimension reduction, and we

can see that the performance of the t-SNE algorithm is the best in view of the intraclass similarity and the interclass difference.

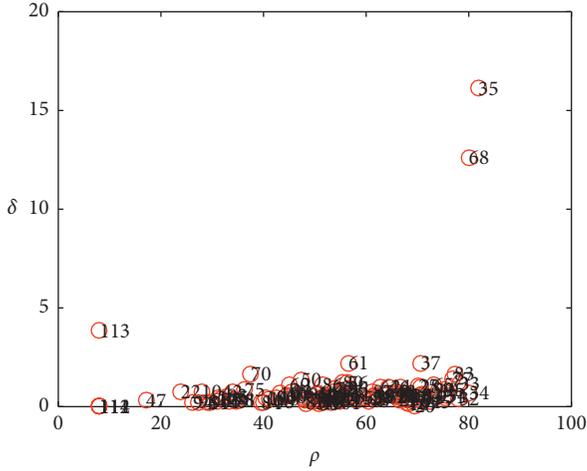
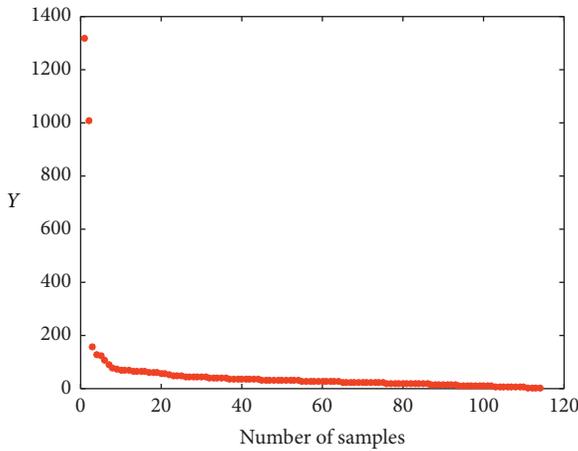
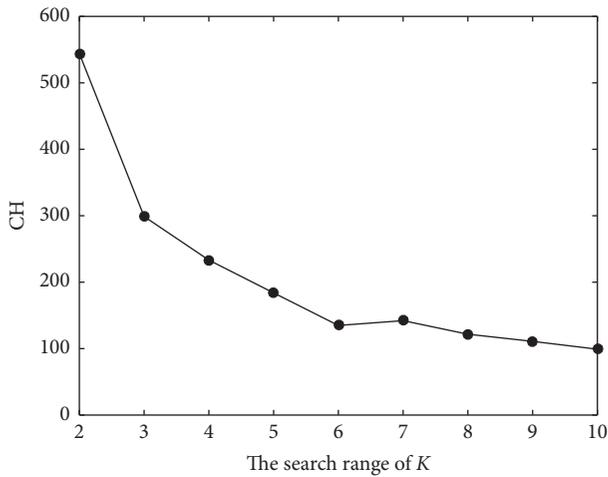
FIGURE 9: Decision chart for parameters ρ_i and δ_i .FIGURE 10: γ_i of each wood sample calculated and sorted in descending order.

FIGURE 11: CH value for different clustering numbers.

Second, the CFSFDP algorithm is used for clustering of 2 wood species *Betula platyphylla* and *Picea jezoensis*. The cutoff distance is set as $d_c = 4$, and the parameters ρ_i

TABLE 4: CFSFDP clustering analysis with CH criterion.

True wood species number	Different values of clustering number K									CCR (%)
	2	3	4	5	6	7	8	9	10	
2	50	0	0	0	0	0	0	0	0	100
3	0	50	0	0	0	0	0	0	0	100
4	0	0	50	0	0	0	0	0	0	100
5	0	0	0	50	0	0	0	0	0	100
6	0	0	0	0	50	0	0	0	0	100
7	0	0	0	0	0	49	1	0	0	98
8	0	0	0	0	0	0	49	1	0	98

TABLE 5: CFSFDP clustering analysis with DB criterion.

True wood species number	Different values of clustering number K									CCR (%)
	2	3	4	5	6	7	8	9	10	
2	50	0	0	0	0	0	0	0	0	100
3	0	50	0	0	0	0	0	0	0	100
4	0	0	50	0	0	0	0	0	0	100
5	0	0	0	50	0	0	0	0	0	100
6	0	0	0	1	49	0	0	0	0	98
7	0	0	1	7	24	18	0	0	0	36
8	0	0	0	0	3	10	37	0	0	74

and δ_i of each sample are calculated, as illustrated in decision chart Figure 9. We can see that these two parameters of no. 35 and no. 68 samples are relatively large, and these two samples are more likely to be clustering centers. Then the γ_i of each sample is calculated and sorted in descending order, as illustrated in Figure 10. In order to fulfill the automatic selection of clustering centers, the clustering number K is set as an interval $K \in [2, 10]$. The first K larger $\gamma_{i1} - \gamma_{ik}$ correspond to K samples which are set as clustering centers. The remaining samples are classified into these K clusters, and the CH or DB internal criterion can be computed to evaluate the clustering effect. Figure 11 illustrates the CH value for each $K \in [2, 10]$, and we can see that the CH value is the largest when $K = 2$. Therefore, the best clustering number is $K = 2$.

Third, the remaining 6 wood species are added to our system one by one, and the clustering analysis is performed again. Since the t-SNE algorithm is based on uncertainty in essence, we fulfill the clustering experiments for 50 times when one wood species is added. The times of correct clustering are recorded and CCR is computed, as illustrated in Tables 4 and 5.

Last, the clustering results are also evaluated by some external criteria such as Rand, Adjusted Rand, Jaccard, Fowlkes–Mallows, whose value interval is $[-1, 1]$. The upper-bound 1 indicates that the clustering result is the best, while the lower-bound -1 is the worst. Therefore, the more these four criteria approach to 1, the more accurate the clustering result is. Table 6 illustrates these four criteria which are all above 0.8, indicating that the CFSFDP algorithm's clustering effect is good.

TABLE 6: CFSFDP clustering analysis with external criterion.

Clustering algorithm	External criterion	Different values of clustering number K						
		2	3	4	5	6	7	8
CFSFDP	Rand	0.9580	0.9504	0.9681	0.9793	0.9854	0.9887	0.9912
	Adjusted Rand	0.9160	0.8861	0.9128	0.9337	0.9462	0.9527	0.9587
	Jaccard	0.9155	0.8564	0.8760	0.8985	0.9138	0.9218	0.9301
	Fowlkes–Mallows	0.9568	0.9231	0.9341	0.9467	0.9551	0.9594	0.9638

4. Conclusions

In this paper, we propose a novel incremental self-adaptive wood species classification prototype system. Our prototype system can recognize the incremental wood species quantity. The visible/near-infrared spectrometer is used to pick up the spectral curves of wood samples for the subsequent wood species classification. First, when new wood samples of unknown wood species are added, they are classified as an unknown category by our one-class classifier, SVDD, while the existent wood species are classified as a known category by the SVDD. Second, the wood samples of known species are sent into the BP neural network for the subsequent wood species classification. A few wood samples of the unknown category falsely classified by SVDD as the known category may be processed by the BP network as “rejected classification.” Third, the new wood samples of unknown species are sent into the CFSFDP algorithm for the unsupervised clustering, and the clustering result is evaluated by the internal and external norms. An improvement on this CFSFDP algorithm is proposed to fulfill the automatic determination of clustering number by use of CH or DB internal criterion. Last, if one cluster of one unknown species has an adequate amount of wood samples, these wood samples are removed and identified by human experts or other schemes to ensure to get the correct wood species name. All wood samples of this cluster must be processed one by one to find and discard those samples falsely classified by SVDD possibly. Then, these wood samples are considered as a new known species and are sent into the classifiers, SVDD and BP neural network, to train them again. This procedure is repeated so that our system can recognize more and more wood species.

Among wood properties such as wood species, wood density, surface color, surface roughness, surface defects, wood strength, moisture content, and so on, wood species is the most important property since other wood properties are usually related to wood species. Many sensors and measurement schemes have been used to detect these wood properties [20–25]. For instance, Salim et al. measured the bonding strength of three wood species pine, oak, and nyatoh [25]. With regard to wood species recognition, all the present wood species classification systems can usually process the limited wood species quantity. Therefore, we propose a novel incremental self-adaptive wood species classification prototype system to solve the above-mentioned issue. Our prototype system can process the incremental wood species quantity. Moreover, our experimental system

is simple and cheap, since the spectrometer used here is an Ocean Optics USB2000 + microfiber spectrometer whose price is only approximately USD 2000. This spectrometer is small and light, which can be used for embedded system development by using Linux operation system and C language programming in the future, but the present prototype system software is realized by using Matlab programming. The research content of this paper lays a theoretical foundation for the development of a portable online practical system that can process more and more wood species.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China with grant no. 31670717, the Fundamental Research Funds for the Central University with grant no. 2572017EB09, and the Heilongjiang Province Natural Science Foundation with grant no. C2016011.

References

- [1] B. K. Lavine, C. E. Davidson, A. J. Moores, and P. R. Griffiths, “Raman spectroscopy and genetic algorithms for the classification of wood types,” *Applied Spectroscopy*, vol. 55, no. 8, pp. 960–966, 2001.
- [2] V. Piuri and F. Scotti, “Design of an automatic wood types classification system by using fluorescence spectra,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 3, pp. 358–366, 2010.
- [3] A. K. Moore and N. L. Owen, “Infrared spectroscopic studies of solid wood,” *Applied Spectroscopy Reviews*, vol. 36, no. 1, pp. 65–86, 2001.
- [4] I. Miranda, J. Gominho, S. Ferreira-Dias, and H. Pereira, “Pattern recognition as a tool to discriminate softwood and hardwood bark fractions with different particle size,” *Wood Science and Technology*, vol. 48, no. 6, pp. 1197–1211, 2014.
- [5] M. Khalid, L. Y. L. Eileen, R. Yusof et al., “Design of an intelligent wood species recognition system,” *International Journal of Simulation: Systems, Science & Technology*, vol. 9, no. 3, pp. 9–19, 2008.
- [6] F. Ruffinatto, A. Crivellaro, and A. C. Wiedenhoeft, “Review of macroscopic features for hardwood and softwood

- identification and a proposal for a new character list," *IAWA Journal*, vol. 36, no. 2, pp. 208–241, 2015.
- [7] U. Khairuddin, R. Yusof, M. Khalid et al., "Optimized feature selection for improved tropical wood species recognition system," *ICIC Express Letters B Applications*, vol. 2, no. 2, pp. 441–446, 2011.
- [8] R. Yusof, M. Khalid, and A. S. M. Khairuddin, "Application of kernel-genetic algorithm as nonlinear feature selection in tropical wood species recognition system," *Computers and Electronics in Agriculture*, vol. 93, no. 1, pp. 68–77, 2013.
- [9] I. Ibrahim, A. S. M. Khairuddin, M. S. A. Talip et al., "Tree species recognition system based on macroscopic image analysis," *Wood Science and Technology*, vol. 51, no. 3, pp. 431–444, 2017.
- [10] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.
- [11] D. M. J. Tax and R. P. W. Duin, "Support vector domain description," *Pattern Recognition Letters*, vol. 20, no. 11–13, pp. 1191–1199, 1999.
- [12] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.
- [13] T. Calinski and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics–Theory and Methods*, vol. 3, no. 1, pp. 1–27, 1974.
- [14] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979.
- [15] L. S. Hyun, J. Y. Seon, K. J. Yun et al., "A new clustering validity index for arbitrary shape of clusters," *Pattern Recognition Letters*, vol. 112, no. 2, pp. 263–269, 2018.
- [16] J. W. Sammon, "A nonlinear mapping for data structure analysis," *IEEE Transactions on Computers*, vol. C-18, no. 5, pp. 401–409, 1969.
- [17] B. J. Tenenbaum, V. Silva, and C. J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 22, pp. 2319–2323, 2000.
- [18] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimension reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [19] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.
- [20] V. Bombardier, E. Schmitt, and P. Charpentier, "A fuzzy sensor for color matching vision system," *Measurement*, vol. 42, no. 2, pp. 189–201, 2009.
- [21] A. G. Christy, T. J. Senden, and P. D. Evans, "Automated measurement of checks at wood surfaces," *Measurement*, vol. 37, no. 2, pp. 109–118, 2005.
- [22] S. Tiryaki and C. Hamzaçebi, "Predicting modulus of rupture (MOR) and modulus of elasticity (MOE) of heat treated woods by artificial neural networks," *Measurement*, vol. 49, pp. 266–274, 2014.
- [23] E. D. Tomak, D. Ustaomer, S. Yildiz, and E. Pesman, "Changes in surface and mechanical properties of heat treated wood during natural weathering," *Measurement*, vol. 53, pp. 30–39, 2014.
- [24] Z. W. Zhong, S. Hiziroglu, and C. T. M. Chan, "Measurement of the surface roughness of wood based materials used in furniture manufacture," *Measurement*, vol. 46, no. 4, pp. 1482–1487, 2013.
- [25] H. Salim, Z. W. Zhong, and W. K. Ong, "Evaluating of bonding strength of pine, oak, and nyatoh wood species related to their surface roughness," *Measurement*, vol. 49, pp. 397–400, 2014.



Hindawi

Submit your manuscripts at
www.hindawi.com

