Surface Defect Detection of Wet-Blue Leather Using Hyperspectral Imaging

SHIH-YU CHEN [ID], (Member, IEEE), YU-CHIH CHENG [ID], WEN-LONG YANG, AND MEI-YUN WANG

Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Yunlin 64002, Taiwan
Intelligence Recognition Industry Service Research Center, National Yunlin University of Science and Technology, Yunlin 64002, Taiwan

Corresponding author: Shih-Yu Chen (sychen@yuntech.edu.tw)

ABSTRACT Detection of surface defects on wet-blue leather is much more challenging than raw-hide leather. Since wet-blue leather turns blue and contains moisture after pre-treatment, it is a semi-product of the cowhide processing. At present, the defect detection of wet-blue leather is mostly carried out manually and is time-consuming and labor-intensive for professional inspectors. This paper is the first to use hyperspectral imaging (HSI) to implement the surface inspection of five wet-blue leather defects including brand masks, rotten grain, rupture, insect bites, and scratches in the pixel level detection. Hyperspectral Leather Defect Detection Algorithm (HLDDA) including Hyperspectral Target Detection (HTD) and Deep Learning (DL) techniques was proposed in this paper. In HTD, Weighted Background Suppression Constrained Energy Minimization (WBS-CEM) and WBS-Hierarchical CEM (WBS-hCEM) were developed in this paper by using weighting to suppress the background and enhance the contrast between the target and background. Experimental results showed that the overall performance of WBS was better than the original CEM. In the DL part, 1D-Convolutional Neural Network (CNN), 2D-Unet and 3D-UNet architectures were designed to segment defect areas. For various characteristics of defects, 1D-CNN emphasizes on defects with spectral features, 2D-Unet emphasizes on defects with spatial features, and 3D-Unet can simultaneously process spatial and spectral information in HSI. The experimental results verified that the proposed HLDDA could effectively quantify and estimate the size of the defect, thereby accelerating the leather inspection process by professional inspectors and develop an automated leather grading system towards Industry 4.0.

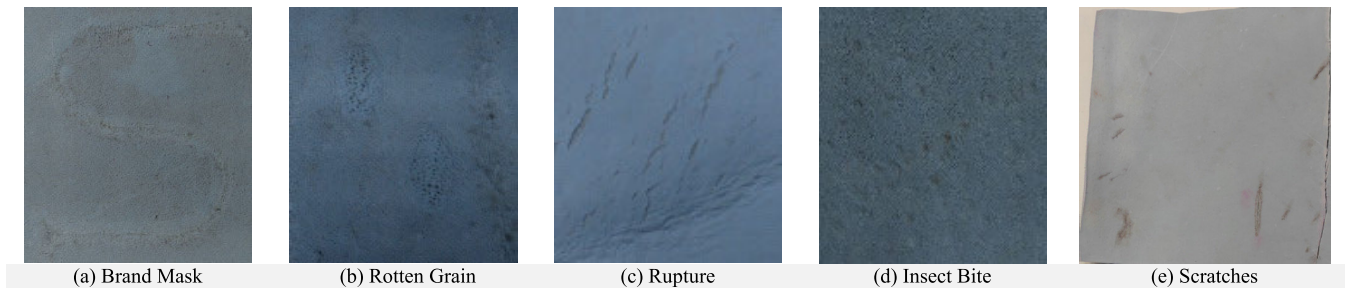**INDEX TERMS** Constrained energy minimization (CEM), deep learning (DL), hyperspectral image (HSI).

I. INTRODUCTION

The leather industry is one of the important traditional industries. The produced leather is mainly supplied for downstream leather goods factories, which use leather as the raw material to make various leather goods, including leather shoes, bags, suitcases, gloves, belts, and sofas. The foremost raw material of the leather industry is cowhide. The skin peeled from the freshly slaughtered cattle is generally known as fur or rawhide. It turns blue after deoiling, degreasing, unhairing, and chroming. It is often called wet-blue leather because it contains moisture and is the semi-product of the cowhide processing procedure. Some marks are left on the leather surface during the growth of cattle, such as brand masks, rotten grain, ruptures, insect bites, healed scars (closed/open), and

The associate editor coordinating the review of this manuscript and approving it for publication was Emre Koyuncu [ID].

scratches, as shown in Fig. 1. Therefore, the products made from natural leather tend to retain these marks, which in turn, affect the leather grade [1]. At present, the common grading standard is SATRA [2], which contains six grades according to the usable area ratio of leather with Grade 6 being considered as unusable. Grades 1-6 are listed in Table 1. Leather is inspected manually; the inspectors must visually inspect the leather and mark the defects with chalk. Manual inspection is likely to cause fatigue and misrecognition; thus, the judgment result must be verified and approved by other inspectors. Therefore, a rapid, comprehensive, and noninvasive inspection method for leather has become an important issue.

As the leather surface is blue after pretreatment, the defects are not obvious as compared to those in the raw-hide leather. The major challenge of detection and recognition of defects on wet blue leather is that the specimens provide very limited spatial information and they are generally difficult to be

| (a) Brand Mask | (b) Rotten Grain | (c) Rupture | (d) Insect Bite | (e) Scratches |

FIGURE 1. Defect appearance.

TABLE 1. SATRA grading.

Grade	1	2	3	4	5	6	unusable
Rate (%)	>96	91-95	86-90	81-85	76-80	71-75	<70

confirmed or detected with the naked eye using RGB images. Therefore, the image processing techniques in the traditional spatial domain may be inapplicable.

In recent years, the development of remote sensing instruments, hyperspectral image (HSI) has become an emerging technology and has been extensively used in the domains of geology [3], agriculture [4]–[7], global change [8], and national defense [9], with highly promising industrial potential [6], [10]–[13]. The hyperspectral sensor has 200 continuous spectral bands to enhance the spectral resolution. Hyperspectral sensors with very dense bands can be used for detecting, classifying, distinguishing, identifying, and quantifying micro-objects and substances. This study employed the spectral information of leather defects instead of the spatial information processing techniques to effectively identify leather defects.

This study proposed a novel method of using HSI to replace the traditional manual wet blue leather inspections. The method is known as the Hyperspectral Leather Defect Detection Algorithm (HLDDA) and combines Hyperspectral Target Detection (HTD) and Deep Learning (DL) techniques to locate and quantize the defective areas of leather. Since hyperspectral data volume is very large, a high data storage capacity is required and reducing data volume is also a topic worth exploring. In terms of the band selection, this study controlled the number of bands effectively at 10, so that the defect position can be detected rapidly with the spectral information of only one target object. The experimental results showed that the WBS technique of HTD effectively suppressed background interference and achieved better results. In addition, this study proposed 1D-CNN, 2D-Unet, and 3D-UNet architectures based on DL. The 1D-CNN performed analysis and quantization completely according to the signals of spectra, and 2D-Unet focused on the defects with spatial characteristics. As the HSI is a 3D image, the spectral and spatial information can be processed simultaneously

by 3D-UNet. The experimental results illustrated that the 3D-UNet has the best performance in detecting ruptures, rotten grains, and scratches defects, whereas the 1D-CNN has a better performance in detecting insect bites and 2D-Unet has the best result in detecting brand masks.

The HTD and DL are completely different concepts. The HTD separates the target from the background based on the concept of a matched filter and can detect the defect position rapidly. For DL, this study applied three different dimensions of convolution to extract the features and correlations from the spatial and spectral domain in HSI, thus, DL requires more prior information to train the neural network. HTD and DL both have their own merits and demerits. The experimental results of this paper can provide appropriate algorithms and spectral ranges, which can effectively quantify and estimate the size of various defects. Using HLDDA to detect defects can ensure the consistency of the inspection criteria, save time, and increase efficiency. The findings of this study can provide a reference for professional inspectors of image analysis, and for leather factories to develop an effective leather grading system. This technique will play an important role in the future development of leather grading towards Industry 4.0.

II. RELATED RESEARCH

Many previous studies have conducted experiments on defect classification for rawhide images. The shape features (e.g. length, width) and textural features (e.g. contrast, entropy) were selected according to the defect features and standard features of leather surface. Then the Feedforward Neural Network (FNN) [14] was combined with Decision Tree [15] to select the optimal attribute and classify the defects. Bong *et al.* [16] extracted the image features and located the defects. The extracted image features included color, shape, and textural features for Support Vector Machine (SVM) classification. Kwak *et al.* [17] used geometrical and statistical characteristics as characteristic sets in the defect classification process. A three-stage sequential decision-tree classifier was used to classify five kinds of defects. Liong *et al.* [18] used AlexNet classification and the U-Net segmentation method to detect three kinds of defect data. Each collected image was separated into 400 × 400 pixels and then classified. Liong *et al.* [19] used Mask

TABLE 2. Related literature comparison table.

Target	Spectrum Range	Leather Kind	Samples number Train / Test	Method	Accuracy	References
Knife Hole	RGB	Raw hide	Unknown / 38	Feedforward Neural Networks (FNN) Decision Tree	94.7%	[15]
Nicks			Unknown / 50		94%	
Imprint injury			Unknown / 53		92.5%	
Aberration			Unknown / 59		94.9%	
Scars	RGB	Raw hide	400 / 125	Image processing algorithms SVM	99.2%	[16]
Scratches			400 / 125		97.6%	
Pinholes			400 / 125		98.4%	
No defect			800 / 125		100%	
Lines	RGB	Raw hide	12 / 10	Thresholding Morphological processing Geometric and Statistical features Decision Tree classifier	100%	[17]
Holes			12 / 10		100%	
Stains			12 / 20		90%	
Wears			12 / 20		90%	
Knots			12 / 20		85%	
Black line	RGB	Raw hide	100 / 25	AlexNet U-Net	99.72%	[18]
Wrinkle			100 / 25		46.18%	
Black line & Wrinkle			200 / 50		99.84% 87.51%	
Tick bites	RGB	Raw hide	84 / 500	Mask R-CNN	70.35%	[19]
Open Cut	RGB	Wet Blue Leather	200 / 141	MLP	94.9%	[20]
Closed Cut			200 / 136		96.4%	
Fly Bite			200 / 174		96.4%	
Without Defect			600 / 118		98.2%	
Tick	RGB	Raw hide	2819	SVM	99.59%	[21]
Brand			3716			
Cut			2804			
Scabies			5383			
Normal	Grayscale	Raw hide	882	KNN MLP SVM	92.30%	[22]
Failure			992		88.45%	
Tick Marks, Brand, Cut, scabies	RGB	Raw hide (192) Wet Blue Leather (66)	400	SVM normalized Gaussian RBF KNN	90%~100%	[23]
Without Defect			400			
Thin spots Circular Spot Thin lines Strips Holes Patterns Irregulars	RGB	Wet Blue Leather	178	Image processing	-	[24]
Tick Marks Pin marks White Spot Flay cuts Cuts Scratch Marks Veins	RGB	Wet Blue Leather	Defect (609) Good (235)	Transfer Learning	69.89%~82.6%	[25]

R-CNN for automatic defect segmentation of tick bites. Each collected image was divided into 400×400 pixels, and a neural network was built using a sliding window. Villar *et al.* [20] used Multilayer Perceptron (MLP) to detect four defects in the wet blue leather. The first-order statistics and contrast characteristics were extracted from the Gray Scale image,
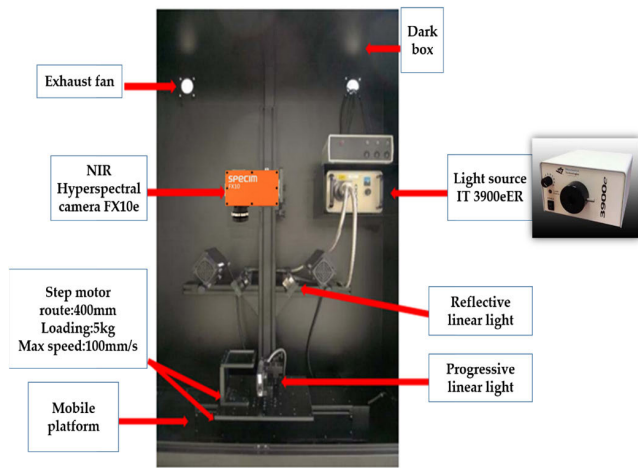
FIGURE 2. HSI collecting system.

Type	FX10	FX17
Wavelength range	400~1000nm	900~1700nm
Spectral resolution	5.5 nm	8 nm
Band number	224 Spectral Bands	224 Spectral Bands
Spatial resolution	1024 Spatial Pixels	640 Spatial Pixels

RGB, and HSV (Hue, Saturation, Lightness) channels for training. Viana *et al.* [21] used the Interaction Maps and the Grey Level Co-occurrence Matrices (GLCM) to extract the attributes of samples as textural features. Several color features were extracted. LIBSVM (a library for SVM) and SMO (sequential minimal optimization) were used for recognition. Pereira *et al.* [22] used GLCM, Local Binary Patterns (LBP), and Structural Co-occurrence Matrix (SCM) to extract features and trained k-nearest neighbors (KNN), Multilayer Perceptron (MLP), and SVM for detecting defects. Pistori *et al.* [23] used the software of the DTCOURO leather defect detection system to extract small samples from the defect and non-defect regions. The textural and color features were extracted from each sample, and the SVM and KNN were used to detect the defects in rawhide and wet blue leather. Yeh and Perng [24] used image processing techniques to detect the wet-blue leather defects and calculated the number of pixels in the unusable region. Aslam *et al.* [25] investigated various transfer learning strategies and integrated networks to improve the classification performance of leather defects by their Wet-blue Leather Image Dataset (WBLID). Although some deep learning-based methods have been applied to defect detection in recent studies [26]–[30], they are not workable in wet-blue leather defect detection, or may not be effective even if they can be applied to as the defects in the wet-blue leather.

Unlike most of the studies that used rawhide, this study focused on wet-blue leather, which defects become less apparent and difficult to recognize by naked eyes after the dyeing process. The traditional image processing techniques may fail to detect those defects. This paper is the first to use HSI to detect defects in the pixel level, and its contributions are as follows:

1) Most of the existing studies used the color, shape, and texture of rawhide leather defects as the benchmark for defect detection, and then used classifiers such as SVM, KNN, and MLP to classify the images of each defect. Evaluation results

used the number of images as a unit. This study used HSI to combine spectral and spatial information to achieve pixel level detection and segmentation on the wet-blue leather. Since the grading of leather is defined by the size of the available area, the proposed HLDDA can effectively quantify the area of the defect and is more accurate for automated leather grading, which is a great progress for the leather industry.

2) Past studies used deep learning to implement the image segmentation of rawhide leather, such as UNET [18], which could only detect the more obvious defects (e.g., black line and wrinkle), and Mask R-CNN [19], which only had an overall accuracy of 70% on image segmentation. The proposed method of this study could detect five common defects on the wet-blue leather and reached an accuracy of over 96% for each defect.

3) Different from other literature proposing one method to classify various defects, this study designed appropriate algorithms including matched filter based HTD, spectral based 1D-CNN, spatial based 2D-Unet, and spatial-spectral based 3D-UNet for various defects. The different characteristics were compared in HSI, which is a novel approach.

4) Different from other wet-blue leather literature using public data sets, this study created its own data set by collecting and establishing more than 20 HSI data sets with the assistance of the tannery, which can better reflect the actual defects handled by the leather factories.

III. MATERIALS AND METHODS
The hyperspectral sensors, leather defect types, and proposed algorithms are detailed as follows:

A. HYPERSPECTRAL SENSORS
The hyperspectral signal has a wider spectrum range and higher spectrum resolution [31]–[33], meaning that the hyperspectral signal has abundant information hidden in each pixel. This study used the push bloom hyperspectral sensors FX10 and FX17 of SPECIM, and the hyperspectral system supplied by Isuzu Optics Corp. The wavelength ranges are 400~1000 nm and 900~1700 nm. The number of bands is 224 and the system is controlled by the software provided by Isuzu Optics Corp. Table 3 tabulates the specification of sensors. The data of whiteboard (absolute reflection, reflectivity is 100%) and black cloth (absolute absorption, reflectivity is 0%) in the dark box were recorded automatically before

FIGURE 3. Inspecting defects by professional inspectors (Photo credit: SHAN BEEN JEOU INDUSTRIAL CO., LTD).

shooting. The sample was completely closed in the course of shooting and the linear light in the dark box was the only light source, as shown in Fig. 2. Finally, the sample reflectivity was calculated for experiments according to the collected light intensity.

B. LEATHER SAMPLES

The skin peeled off the slaughtered cattle is known as fur and it rots. To prevent it from rotting and losing hair, it is generally cured with salt. The processed fur with salt is called wet salted hide, which can be preserved for 3 to 5 months in general. The fur and wet salted hide are generally known as rawhide, which then turns blue after soaking, fleshing, limin, deliming, bating, pickling, and chroming. As it contains moisture, it is generally known as wet-blue leather. The defects on the surface after processing are less apparent than those on rawhide.

The skinner, Shan Been Jeou Ind. Co., Ltd., in Taichung, Taiwan provided nearly 20 sheets of wet-blue leather. Professional inspectors had identified the defects including brand mask, rotten grains, rupture, insect bites, and scratches, as shown in Fig. 3. For this study, the brand mask, rupture, and scratches were shot by the FX10 whereas the rotten grain and insect bite were shot by the FX17 camera. The collected images were cropped for the leather only (the original image contains the whiteboard and black cloth for calculating reflectivity). The dimensions of the cropped images are 350×750 pixels. The total number of pixels is $120,000 \times 280,000$.

1) BRAND MASK

Brand mask is a destructive sign on the cattle made by the livestock farm for the convenience of cattle management. Its feature is quite apparent but its color is similar to the leather.

2) ROTTEN GRAIN

Rotten grain is a bead-like defect formed on the rotten leather surface that resulted from defective preservation after the animal is slaughtered. There are micro craters in the defect, and the defect is not obvious in color images.

3) RUPTURE

Rupture is the destruction induced by folding the cowhide in preservation. It is apparent in color images, like broken skin.

4) INSECT BITE

Insect Bite is the scar resulted from a mosquito or parasite bite during the growing period of cattle, including the scars before healing and after healing. Its appearance is very small and unobvious in color images, but craters can be seen after a closer look, and are similar to the signal of normal leather in HSI.

5) SCRATCHES

Scratches are the marks induced by a slight cut in cattle. In general, there is a black region around it. It is obvious in color images, but the appearance of the cut is not obvious.

C. HYPERSPECTRAL TARGET DETECTION (HTD)

HTD can detect targets both actively and passively. Active target detection looks for certain target information, and detects specific targets. These targets of interest can be obtained by the supervised or unsupervised method. This study used active target detection to detect specific defects.

1) CONSTRAINED ENERGY MINIMIZATION (CEM)

The CEM [34]–[40] is an active target detection algorithm. Among the numerous existing target detection algorithms, CEM is stable and excellent in sub-pixel detection. During target detection, the CEM algorithm is given only one spectral signature as parameter 'd' (Desired Signature), without requiring other prior knowledge (e.g. multiple targets of interest or background). It means that the users can extract specific targets of interest and obtain the results of target detection when the other factors of the detection environment are unknown. Moreover, as many signals cannot be recognized or observed with the naked eyes, the CEM uses the correlation matrix **R** to suppress the background and uses the feature 'd' to match the customized FIR and highlight the target. This is to enhance the ability for detecting features and efficient execution. If there are n spectral signatures, all the spectral signatures are defined as $\{r_1, r_2, r_3, \ldots\ldots, r_{n-2}, r_{n-1}, r_n\}$. It is expressed as

$$\delta^{CEM} = \left(\mathbf{w}^{CEM}\right)^T \mathbf{r} = \left(\mathbf{d}^T \mathbf{R}_{LxL}^{-1} \mathbf{d}\right)^{-1} \left(\mathbf{R}_{LxL}^{-1} \mathbf{d}\right)^T \mathbf{r} \quad (1)$$

2) WEIGHTED BACKGROUND SUPPRESSION CEM (WBS-CEM)

The WBS-CEM [40] was proposed for the RGB image for the first time. Its main concept is to use different weights for each pixel in calculating correlation matrix **R**, and a new nonparametric correlation matrix is redefined for feature extraction. The method aims at the data with the non-Gaussian distribution. Since the distribution of data points is not Gaussian distribution, the average value of each class cannot represent the center of the integer. The WBS-CEM
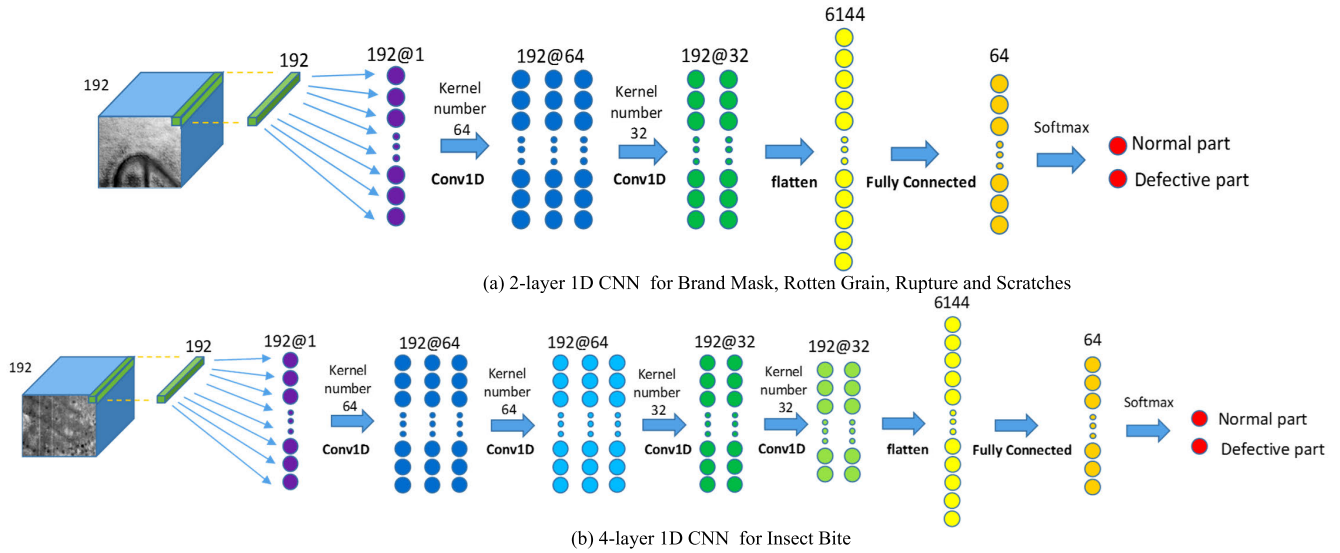
(a) 2-layer 1D CNN for Brand Mask, Rotten Grain, Rupture and Scratches

(b) 4-layer 1D CNN for Insect Bite

FIGURE 4. 1D-CNN network architecture.

uses w_i in calculating the correlation matrix for weight adjustment. The principle is to use the distance to the target point for weight adjustment. The shorter distance is closer to the target spectrum. The correlation matrix can be multiplied by the distance to reduce the influence of the spectral signature. It is similar to the target of interest on the correlation matrix, and the weighted correlation matrix is extended into a part of target detection algorithms. This is to enhance the accuracy of target detection algorithms.

The w_i computing modes can be Euclidean distance [41] or Spectral Information Divergence [42]

$$\mathbf{w_i} = \text{ED}\left(\mathbf{r_i}, \mathbf{d_i}\right) \tag{2}$$

$$\mathbf{w_i} = 1 - \text{SID}\left(\mathbf{r_i} || \mathbf{d_i}\right) \tag{3}$$

The correlation matrix R is adjusted by weight

$$\mathbf{R}^* = \frac{1}{N} \sum_{i=1}^{n} \mathbf{w_i} \mathbf{r_i} \mathbf{r_i}^{\text{T}} \tag{4}$$

The weight is substituted in the correlation matrix of CEM, expressed as:

$$\delta^{WBS-CEM} = \left(\mathbf{d}^{\text{T}} \mathbf{R}_{\text{LxL}}^{*-1} \mathbf{d}\right)^{-1} \left(\mathbf{R}_{\text{LxL}}^{*-1} \mathbf{d}\right)^{\text{T}} \mathbf{r} \tag{5}$$

3) HIERARCHICAL CEM (hCEM)
The CEM uses the FIR filter and calculates the correlation matrix to suppress the background. It does not perform well in some scenarios; hence, this paper uses hCEM [43]. The method performs the first calculation according to the traditional CEM method. $\mathbf{X}$ is the input image data including N spectral signatures. It is expressed as $\mathbf{X} = [x\_1, x\_2, \ldots, x\_N]$. The kth output y is expressed as

$$y = \left(\mathbf{R}^{-1}\mathbf{d}\right) / \left(\mathbf{d}^{\text{T}} \mathbf{R}^{-1}\mathbf{d}\right)\mathbf{X} \tag{6}$$

Afterward, the output result is multiplied by weight q to eliminate non-targets of interest. The weight q of the tth hyperspectral signal can be expressed as

$$q(t) = \begin{cases} 1 - e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{7}$$

To sum up, the $k + 1^{\text{th}}$ result can be expressed as

$$x_i^{k+1} = q * x_i^k \tag{8}$$

$k \Leftarrow k + 1$, and then the output of Eq. 8 is used in (6) and (8) and repeat until satisfaction. The final output is expressed as

$$\delta^{\text{hCEM}} = [y_1^k, y_2^k, \ldots, y_N^k] \tag{9}$$

4) WEIGHTED BACKGROUND SUPPRESSION HIERARCHICAL CEM (WBS-hCEM)
The method is derived from the combination of the concepts in the previous two sections. The hyperspectral signal is given an initial weight, expressed as Eq. 2 or 3. Then, the output result is multiplied by weight q, expressed as Eq. 7. The irrelevant information of interest is removed gradually by continuous iteration, expressed as

$$\delta^{WBS-hCEM} = \left(\mathbf{d}^{\text{T}} \mathbf{R}_{\text{LxL}}^{*-1}\mathbf{d}\right)^{-1} \left(\mathbf{R}_{\text{LxL}}^{*-1}\mathbf{d}\right)^{\text{T}} \mathbf{r} * q(t) \tag{10}$$

The next step is the same as hCEM in the previous section, iterated until satisfaction.

D. DEEP LEARNING (DL)
The present DL [44] is one of the major application domains of machine learning and has been widely used in Artificial Intelligence (AI). The earliest basic concept of neural networks was proposed by Chen *et al.* [40], and its network architecture has been used to simulate the image recognition method of humans to enhance the image recognition capability of the machine.
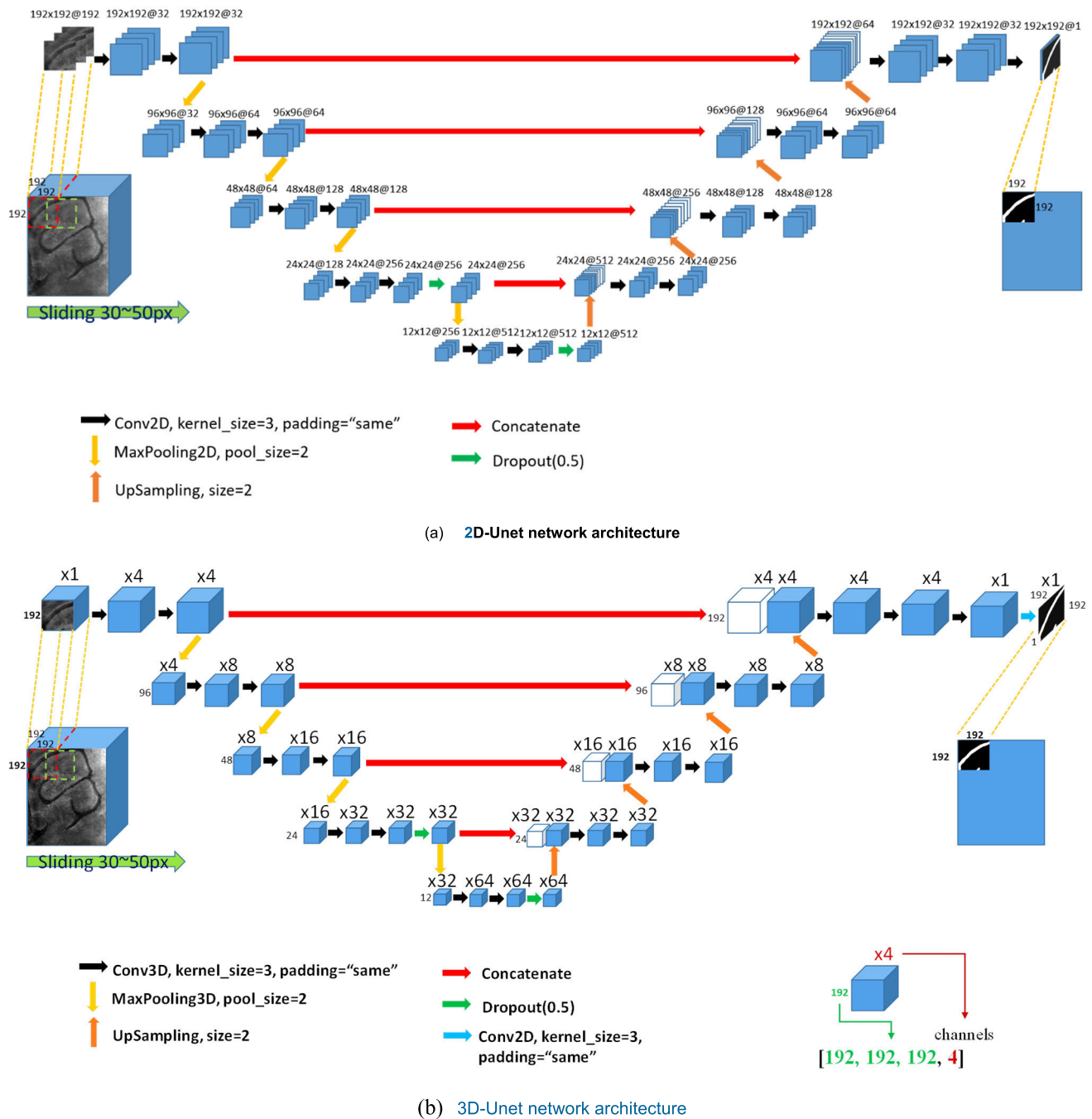
(a)   **2**D-Unet network architecture

(b)   3D-Unet network architecture

FIGURE 5. 2D and 3D Unet network architecture.

1) 1D-CNN

The CNN looks for the feature map of data through the convolutional layer. It is usually used to search for continuous coherent signals in the case of 1D. To obtain the features of interest from the shorter part of the entire data set, and when the positions of features in the segment are free of high correlation, the 1D CNN is very effective [41]. It is highly applicable to the data of sensors, such as time series analysis of audio signal or accelerometer data. As the hyperspectral signal has numerous and correlated bands, this paper attempts to find the features of leather defects through 1D-CNN. The network model architecture, as shown in Fig. 4, comprises the 2-layer and 4-layer 1D CNN are proposed in this paper. As shown, 2-layer 1D CNN is designed for brand mask, rotten grains, ruptures, and scratches; 4-layer 1D CNN is designed for insect bites. Since the spectral signature of insect bites is very similar to the normal leather, two more layers of 1D convolution are added to obtain more features.
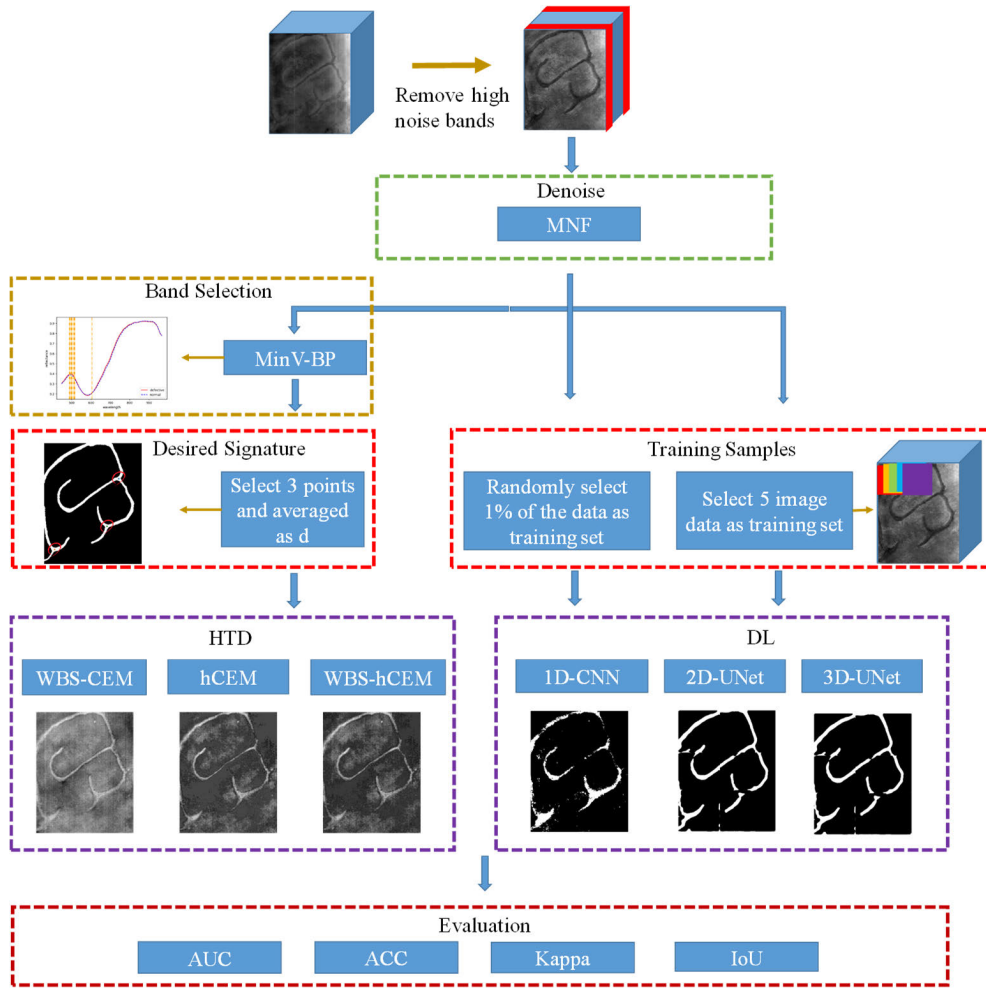
FIGURE 6. Hyperspectral Leather Defect Detection Algorithm (HLDDA) flowchart.

The input signal is the signal of each pixel in the leather image. After 1D convolution, the convolution kernel is 3×3, and the activation function is ReLu. The result of convolution is flattened into 1D through a fully connected layer and the number of outputs is 64. Finally, a fully connected layer with two outputs is connected and combined with the activation function Softmax for classification.

2) 2D-UNet

The architecture of 2D-UNet [47] is evolved from the Fully Convolutional Network (FCN) [48]. The concept of the residual block was imported by Milletari [49]. The residual block is well known in ResNet. The original design goal is to reduce the gradient vanishing and saturation problems induced by stacking a number of convolutional layers. The construction of UNet aims at prediction with a few data. The dimension reduction path in the left of the architecture diagram is called the contracting path, in the right is the expanding path. To obtain accurate information, there will be convolution of at least two successive layers before dimension reduction or dimension raising, known as successive convolution.

Another characteristic is to maintain high channel number of up-sampling so that relative position relation and detailed features can be fully combined and the quality of recognition is upgraded. The U-Net has been used in image segmentation, such as medical images [50], [51] and remote sensing [52]; it can be also extended and enhanced to the 3D-UNet models, which have been applied in medical applications [53]–[55]. The network architecture with parameters is shown in Fig. 5 (a). The convolution kernel of all the 2D convolutional layers of the model is [3, 3].

3) 3D-UNet

This paper considers the information content of the hyperspectral signal in spatial and spectral information and constructs the 3D-UNet architecture based on the 2D-UNet. The network model architecture is shown in Fig. 5(b). The shape of input data is set is [192, 192, 192]. It is set as 192 because the 3D-UNet requires four dimension reductions. Each dimension reduction reduces the information content of each dimension by half. There is 1/16 of input data remaining after four dimension reductions. To avoid
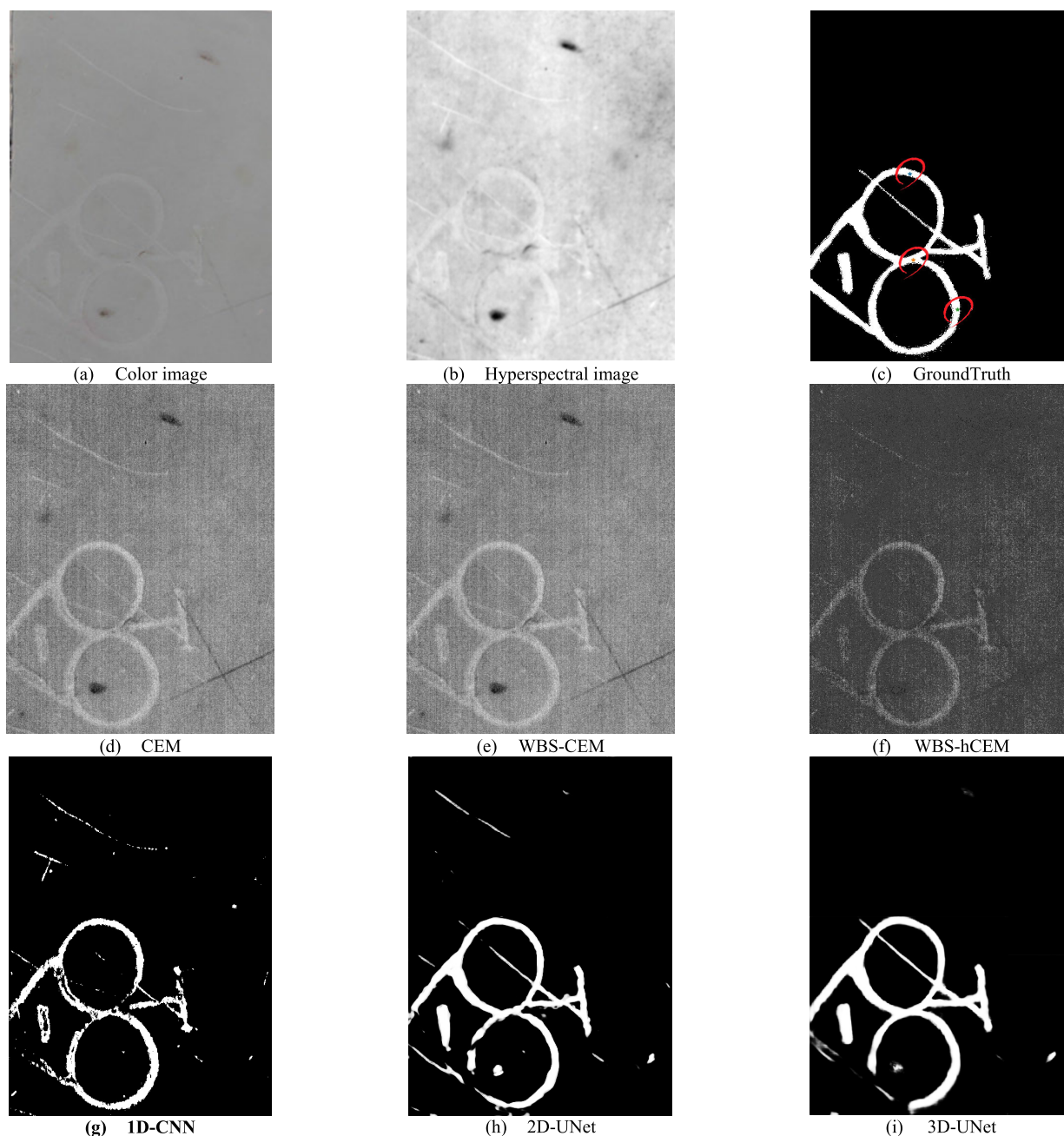
(a) Color image   (b) Hyperspectral image   (c) GroundTruth

(d) CEM   (e) WBS-CEM   (f) WBS-hCEM

(g) **1D-CNN**   (h) 2D-UNet   (i) 3D-UNet

FIGURE 7. Experimental results of brand mask.

zero-padding of the model, the 192 which is the maximum multiple of 16 within 200 is selected as the input size. The convolution kernel of all the 3D convolutional layers of the model is [3, 3, 3]. The activation function is Swish [56] and the output of the second to the last layer is 3D data. The last layer uses 2D convolution to eliminate the third dimension. The activation function is ReLu so that the output is a 2D image. It is noteworthy that the 3D-UNet has fewer samples than 1D-CNN. The loss function used in this model is balanced cross-entropy [57], the weight of the class (i.e. defect) with fewer samples is set as 0.85, and the other class (i.e. background) is set as 0.15, so as to effectively

suppress the behavior of the prior model, which is likely to identify the full sheet of leather as background. The loss function gives individual weight to each class, expressed as follows.

$$BCE\,(p, \hat{p}) = -\left(\beta p \log\left(\hat{p}\right) + (1 - \beta)\,(1 - p)\log(1 - \hat{p})\right) \tag{11}$$

E. BAND SELECTION (BS)

The concept of BS was discussed as early as in multispectral times. Mausel [34] used the concept of correlation coefficient of statistics to determine a subset from an 8-channel
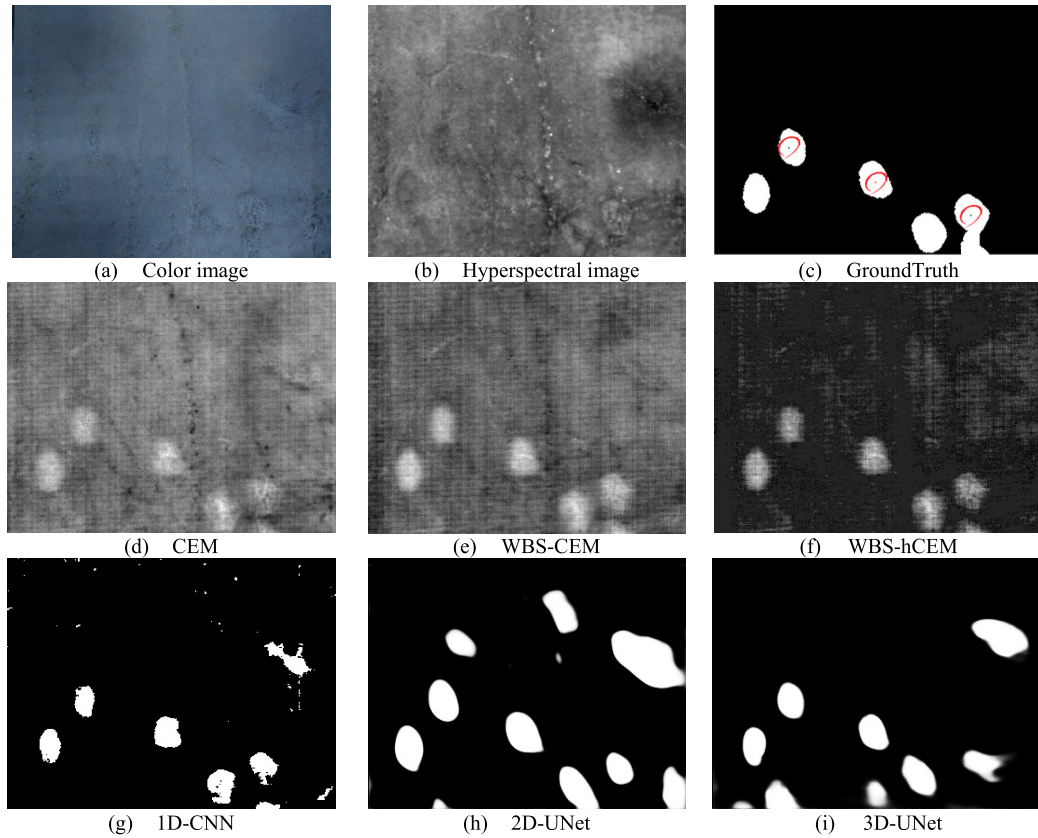
(a)　Color image

(b)　Hyperspectral image

(c)　GroundTruth

(d)　CEM

(e)　WBS-CEM

(f)　WBS-hCEM

(g)　1D-CNN

(h)　2D-UNet

(i)　3D-UNet

FIGURE 8. Experimental results of rotten grain.

multispectral image for classification. The classification rate was not affected. In recent years, the minimum similarity-based method is the most frequent BS method. Briefly, the newly selected band must have the minimum similarity and correlation with the selected band. The methods include correlation coefficient, Euclidean distance, Spectral Angle Mapper (SAM) proposed by Keshava [58], and Orthogonal Subspace Projection (OSP) [59]. This study used the target detection concept as the main band selection algorithm.

1) CONSTRAINED-TARGET BAND SELECTION (CTBS)

A new band selection method for multi-target detection is known as Constrained-Target Band Selection (CTBS) [60]. It is derived from the concept of CEM by constraining the target energy while minimizing the variance induced by background. Based on CEM, the variance resulting from the target of interest can be used for determining the frequency band priority and selecting the frequency band for the specific target.

If $\{b_l\}_{l=1}^{L}$ is a group of band images representing hyperspectral signal cube, where b_l is the l–th band. According to the obtained CEM error if Ω_{BS} is a group of bands selected and $\{r_{BS}^i\}_{i=1}^{N}$ is data set, the following equation can be defined as CTBS priority standard

$$\mathrm{V}\left(\Omega_{\mathrm{Bs}}\right) = \left(\mathbf{d}_{\Omega_{\mathrm{BS}}}^{\mathrm{T}} \mathbf{R}_{\Omega_{\mathrm{BS}}}^{-1} \mathbf{d}_{\Omega_{\mathrm{BS}}}\right)^{-1} \quad (12)$$

2) MINIMAL/MAXIMAL VARIANCE-BASED BAND PROCESSING (MIN/maxV-BP)

According to the optimization method of CEM, the priority score is processed by variance, if d is used for detecting the target of interest of d designated target. For each band b_l, the variance of the band can be calculated according to Eq. (12)

$$\mathrm{V}\left(b_l\right) = \left(\mathbf{d}_{b_l}^{T} \mathbf{R}_{b_l}^{-1} \mathbf{d}_{b_l}\right)^{-1} \quad (13)$$

As $\mathrm{V}\left(b_l\right)$ only uses the data sample vector designated by b_l, its value can be used in the priority score assigned to b_l. Therefore, the priorities of all bands can be sequenced according to the value of $\mathrm{V}\left(b_l\right)$. A smaller $\mathrm{V}\left(b_l\right)$ means that the b_l is more important. The band selection is known as Minimal Variance-based BP (MinV-BP).

The FIR filter of all band set Ω can be substituted, and $\Omega_{BS} = \Omega - b_l$, derived from Eq. (13) into

$$\mathrm{V}\left(\Omega - b_l\right) = \left(\mathbf{d}_{\Omega-\{b_l\}}^{T} \mathbf{R}_{\Omega-\{b_l\}}^{-1} \mathbf{d}_{\Omega-\{b_l\}}\right)^{-1} \quad (14)$$

The maximum variance is removed from Ω and $\mathrm{V}\left(\Omega - b_l\right)$ is used as a priority score. In other words, the higher the $\mathrm{V}\left(\Omega - b_l\right)$ is, the more important is the b_l band, known as Maximal Variance-based BP (MaxV-BP).

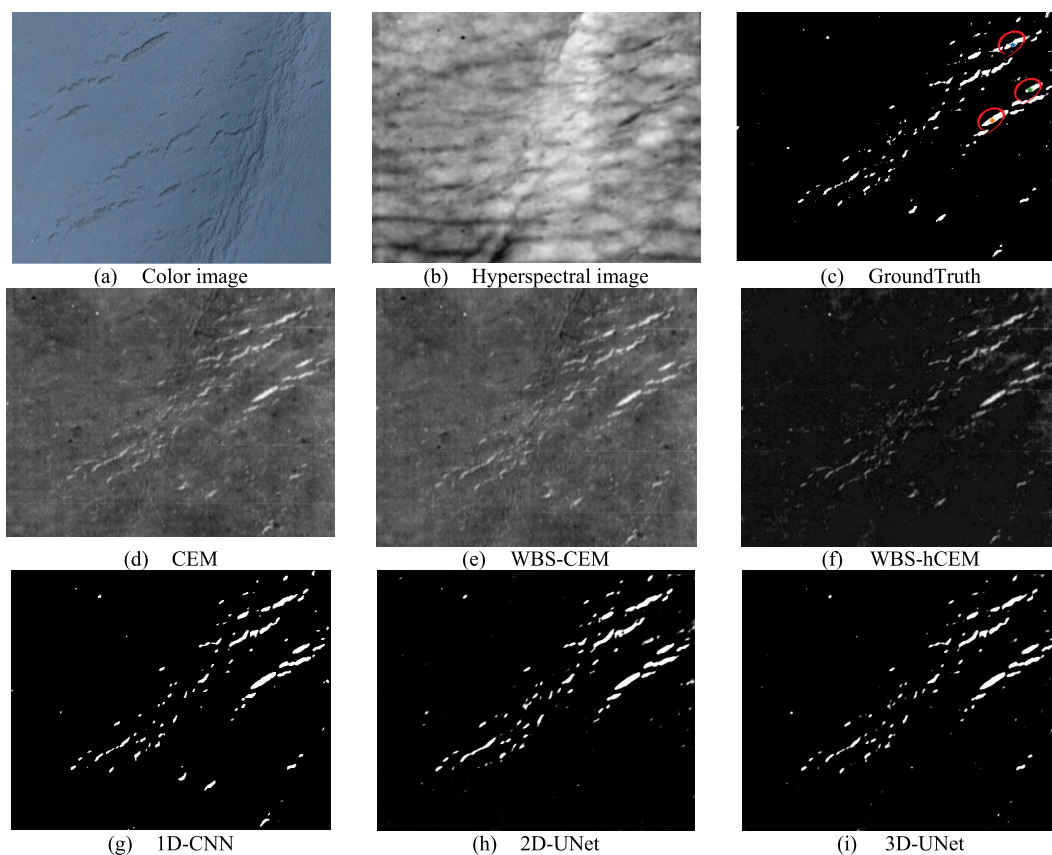| (a)　Color image	(b)　Hyperspectral image	(c)　GroundTruth
(d)　CEM	(e)　WBS-CEM	(f)　WBS-hCEM
(g)　1D-CNN	(h)　2D-UNet	(i)　3D-UNet

FIGURE 9. Experimental results of rupture.

TABLE 4. Computing time in second.

TIME (SEC.)	CEM	WBS-CEM	hCEM	WBS-hCEM	1D-CNN	2D-UNet	3D-UNet
TRAIN	X	X	X	X	15.23	250	4500
TEST	0.0059	0.0328	0.0325	0.0538	0.0057	0.1449	0.1915

F. HYPERSPECTRAL LEATHER DEFECT DETECTION ALGORITHM (HLDDA)

This paper combined the above methods to develop the Hyperspectral Leather Defect Detection Algorithm (HLDDA) including HTD and DL techniques. Considering the noise during imaging of the hyperspectral camera, HLDDA removed low signal/high noise bands: bands 1-20th and bands 211-224th. A total of 192 bands were used in the experiments. In addition, to remove noise more efficiently, the Minimum Noise Fraction Transform (MNF Transform) [61] was used for hyperspectral data. The method regards the Signal-to-Noise Ratio (SNR) as the indicator for evaluating signal quality. The noise interference can be reduced by MNF Transform, which makes the data cleaner. HLDDA does not use MNF Transform to directly reduce the bands because the clean band is not always suitable for various defects, as long as the noise of the original image is reduced and the suitable bands for various defects can be found according to the BS method.

The BS method is performed for the imported HSI to select 10 bands. It is worth noting that DL requires a lot of data information to achieve better results; in this case, it can skip BS and keep full bands HIS to obtain more features. In addition, the HTD requires a desired target as a prior knowledge; HLDDA uses the average of three target pixels selected from the ground truth as an input parameter. In the training process of DL, 1D-CNN randomly takes 1% of its data as the training set. The 3D-UNet uses a sliding window to split HSI to obtain five small images with defects as a training set, while the rest of the data is the testing set. Fig. 6 plots the flowchart of HLDDA.

IV. EXPERIMENTS

A. EVALUATION

As the number of samples is large and the quantities of two classes of samples are unequal, the Receiver operating characteristic (ROC) analysis is used. The area under curve (AUC), True Positive Rate (TPR), False Positive
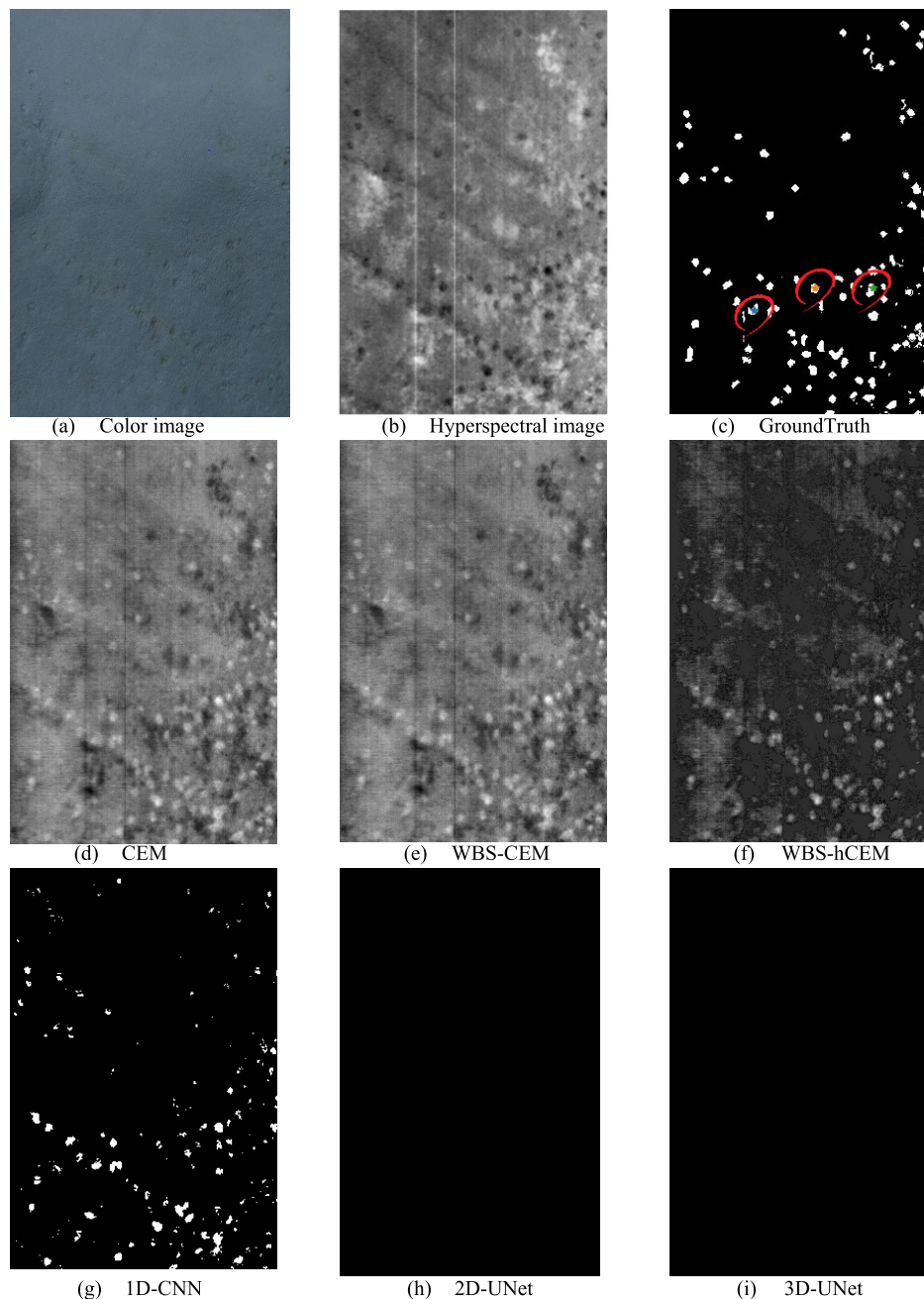
(a) Color image	(b) Hyperspectral image	(c) GroundTruth
(d) CEM	(e) WBS-CEM	(f) WBS-hCEM
(g) 1D-CNN	(h) 2D-UNet	(i) 3D-UNet

FIGURE 10. Experimental results of insect bite.

Rate (FPR), Accuracy (ACC), Kappa, and Intersection over Union (IoU) are also used for evaluation.

The TPR is the proportion of correctly identified samples to all the samples that are defective in fact (positive). The FPR is the proportion of misidentified samples to all the samples that are normal in fact (negative). The (FPR, TPR) coordinates of each threshold of the same model are drawn in the ROC space to form the ROC curve of a specific model. The area under the curve is AUC [62].

Cohen's kappa [63] is different from the ROC curve. The effect of the detector under different thresholds can be calculated. The kappa coefficient (κ) represents two errors (omission and commission) and the overall accuracy of the classifier. The confusion matrix expressed as Eq. (15) is used to calculate the performance index.

$$Confusion\ matrix(H) = \begin{bmatrix} H_{11} & \cdots & H_{1K} \\ \vdots & \ddots & \vdots \\ H_{K1} & \cdots & H_{KK} \end{bmatrix} \quad (15)$$
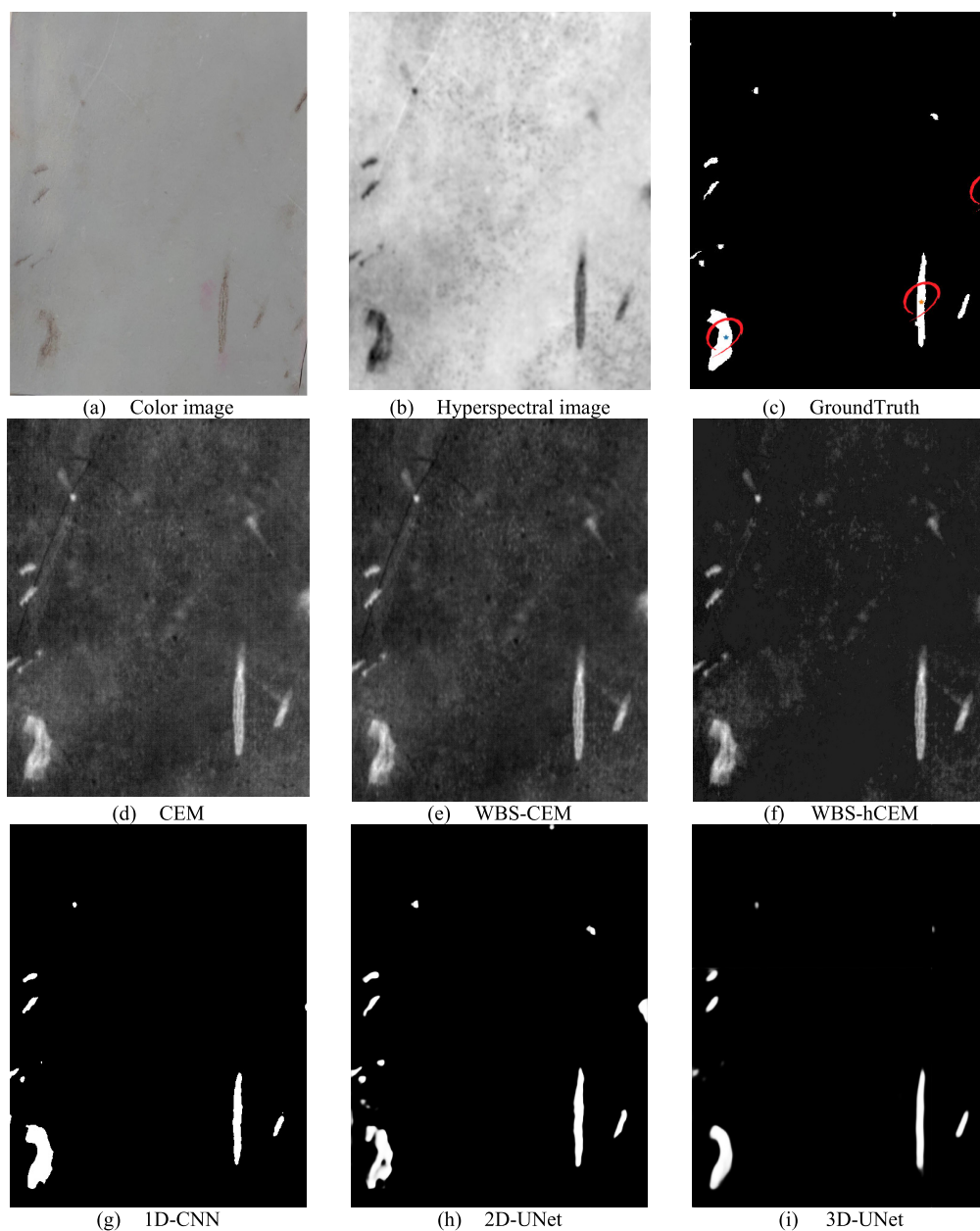
FIGURE 11. Experimental results of scratches.

The Intersection over Union (IoU) [64] is also known as the Jaccard similarity coefficient. It is the statistics for comparing the similarity and diversity of the sample set. The Jaccard index can measure the similarity of the finite sample set. It is defined as the ratio of intersection size to union size of two sets

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (16)$$

B. EXPERIMENTAL RESULTS

The experiments are performed in python 3.8, RTX-3080 10G, Inteli9-10900F@_2.8GHz, RAM 32GB. The collected data of different defects are tested.

Figs. 7-11 show the experimental results of HLDDA; (a) shows the color image of the defect; (b) shows the HSI of the defect in the No. 150 band; (c) shows the Ground Truth of the defect, where the white part is defect, the black part is non-defect, and three red circles indicate the positions of three desired signature for the parameter for HTD; (d)-(i) show the results of CEM, WBS-CEM, WBS-hCEM, 1D-CNN, 2D-UNet, and 3D-UNet, respectively. This study used the minV-BP method with a shorter computing time as a band selection method for selecting 10 bands. The band selection results are shown in Fig. 12. This wavelength range is an important criterion for customizing sensors in the future and the production cost is reduced. Four indexes including
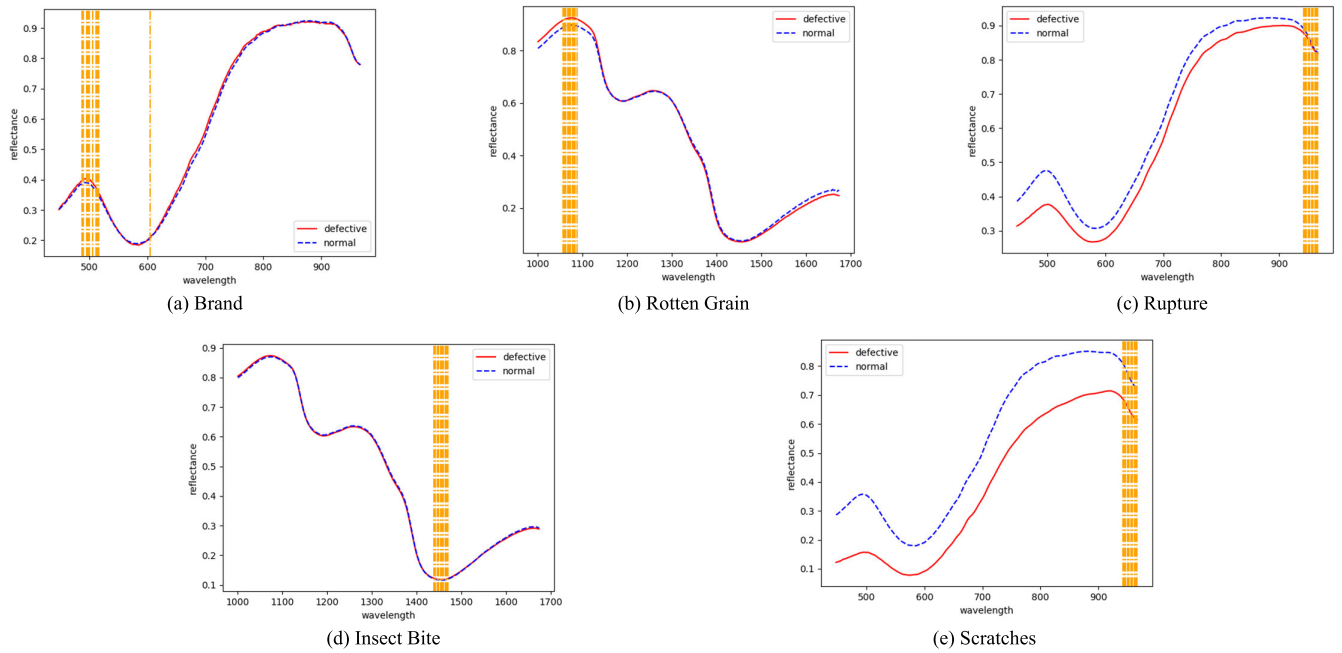
FIGURE 12. Results of band selection.

AUC, ACC, Kappa, and IoU were used for evaluation. The aforesaid five methods were compared. The result of HTD is a gray-level image, which must be binarized before classification. This study found the optimum threshold of maximization TPR+ (1-FPR). TPR, FPR, ACC, Kappa, and IoU were calculated according to the classification result. Finally, a comparison histogram of the detection results of each algorithm and overall performance is plotted in Fig. 13. There are at least three HSIs of each defect. The data of each result is the average of three HSIs.

According to the results in Fig.13(a), the 2D-UNet has the best performance in detecting brand mask because the brand mask is a defect with more spatial features. The 3D-UNet combining spatial information with spectral information is better than the other algorithms with only spectral information.

In detecting the rotten grain defect in Fig.13(b), the WBS-CEM has the best performance in AUC and TPR, and the 3D-UNet has the best performance in Kappa and IOU. Because the WBS- CEM uses the concept of matched filter, it has higher AUC and TPR, as well as a higher false alarm, leading to worse data of Kappa and IOU.

In detecting the rupture defect in Fig. 13(c), the WBS-CEM has a better effect in AUC, but 3D-UNet still has the best results in Kappa and IOU among other algorithms.

In detecting the insect bite in Fig. 13(d), the 2D and 3D-UNet cannot smoothly work. It is possibly because the insect bite defect is very small with very limited spatial information and its feature disappears after multiple convolutions. In this case, we can rely on spectral information based on 1D-CNN. According to the experimental results, 1D-CNN has the best result and the HTD may be disturbed by the

other defects, leading to an excessive false alarm. According to the spectral signature of insect bites, it is very close to the signal of normal leather, resulting in high difficulties in analysis.

The results on detecting scratches in Fig. 13(e) are universally good. The HTD results in better AUC and TPR. The DL has better FPR, ACC, Kappa, and IoU.

The results of overall performance are shown in Fig. 13(f). As seen, the proposed WBS-CEM and WBS-hCEM perform better than original CEM and hCEM in AUC and Acc. However, the overall performance of 3D-UNet in ACC, Kappa, and IoU is still better than other methods.

C. COMPUTING TIME
Table 4 lists the HTD and DL prediction time for reference. The average computing time of one image is given below. The image size is [390, 482]. It is worth noting that HTD does not require any training process, which is the major advantage. However, DL takes much more time than HTD including training and testing time. Especially, 3D-UNet spends longest time in training, in order to implement 3D convolution. HTD is the fastest using the original CEM, but the commuting time are all very close and fast.

D. DISCUSSION
According to the experimental results in the previous section, our proposed WBS-CEM and WBS-hCEM perform much better than the original CEM for different defects. The results prove that background suppression plays an important role in detection results. From the angle of vision, the WBS-hCEM has the best effect on background suppression, followed by WBS-CEM. The hCEM is an excellent algorithm, but can
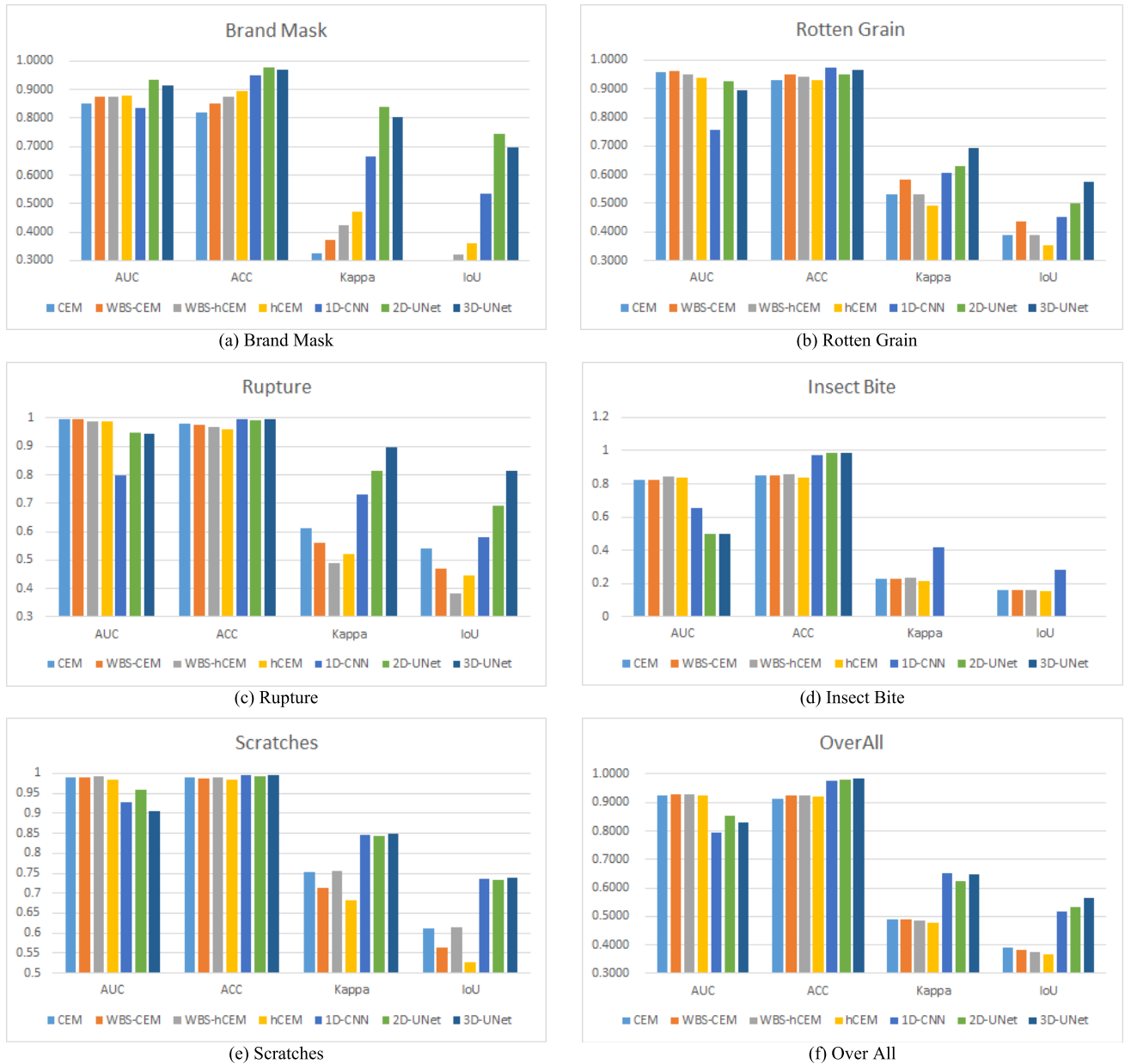
FIGURE 13. Comparison histogram of the detection results and overall performance.

also powerfully suppress the background to reduce the false alarm rate greatly. Additionally, some indistinct targets are suppressed as background. Excessive suppression results in worse evaluation indexes, such as AUC, than WBS-CEM. If the WBS-hCEM is selected, the result is generally between the former two, better or worse than the former two in some instances. In terms of DL, 1D-CNN has the best result in detecting insect bites since the defect of insect bites can only be recognized by spectral information; 2D-Unet takes advantage of spatial information so that it performs best in brand mask, which has more spatial features; the 3D-UNet considers spatial information and spectral information simultaneously. Therefore, it has the best performance in rotten

grain, rupture, and scratch defects. A disadvantage is that tiny defects cannot be detected since the features vanish during the convolution. This is the reason that the insect bite defect fails to be detected in 2D and 3D-Unet. However, the 3D-UNet generally has a better overall performance than 1D-CNN and 2D-Unet, but it requires more GPU memory. If the image is divided into small pieces, partial spatial characteristics will be lost and the detection results will be unacceptable.

To sum up, the HTD and DL perform signal analysis from different starting points. The HTD uses the concept of matched filter to separate the target object from the background. Our proposed WBS-CEM and WBS-hCEM use weighting to suppress background and enhance the contrast

between target and background. The advantage is that the target object can be located as long as one spectral signature is given as a parameter. On the contrary, the ID-CNN, 2D-Unet, and 3D-Unet analyze spatial and spectral signatures in the concept of a neural network. They must collect a lot of data for analysis. The ID-CNN mainly convolves spectral signature, whereas the 2D-Unet mainly convolves spatial features and 3D-Unet convolve spatial and spectral information simultaneously, but requires a longer computing time. The experimental results prove the above theory. Therefore, with limited resources and time, the HTD is preferred. On the contrary, if there are adequate resources and time, the DL is workable.

V. CONCLUSION

This paper analyzed five defects in wet-blue leather including brand mask, rotten grain, rupture, insect bite, and scratches by using HSI. Our proposed HLDDA includes HTD-based WBS-CEM and WBS-hCEM, as well as DL-based 1D-CNN, 2D-Unet, and 3D-Unet architectures to detect defects. The experimental results show that the HLDDA positively provides a new direction and feasibility of wet-blue leather inspection. Several contributions of this paper are considered to be significant and summarized in the following.

1) Prior studies mostly extract features from the spatial domain and then implement image classification on rawhide leather images. This paper is the first analytical study using HSI for wet-blue leather at the pixel level. The wet-blue leather after pre-processing is more challenging than general rawhide leather in defect detection.

2) Our proposed HLDDA uses the spectral and spatial information of leather and implements pixel level defect detection and segmentation to effectively quantize and estimate the size of defects so that it is more precise for leather grading.

3) We created our own data set by collecting more than 20 HSIs, almost 5.25 million spectral signatures with tannery's help. Our data set could be closer to the real situation of the factory.

4) In the HTD results, our proposed WBS-CEM and WBS-hCEM perform much better than the original CEM in different defects. The results prove that the background suppression technique plays an important role from the angle of vision, and can provide better visual contrast. In DL, we designed the spectral based 1D-CNN, spatial based 2D-Unet, and spatial-spectral based 3D-UNet for various defect features, which was never investigated in the past.

5) This paper provides an important wavelength range for recognizing different defects. This is an advantage in practical application and customizing hyper/multispectral sensors in the future.

6) This paper is the pilot study and guideline for HSI in the detection of wet-blue leather to design appropriate algorithms from the angle of HTD and DL. Our future

work is to present a better and novel network structure in the HTD or DL section. The short-term goal is to enable professional inspectors to diagnose defect positions faster. The long-term goal is to work with automated manipulators to develop intelligence leather grading towards Industry 4.0.

REFERENCES

[1] M. Aslam, T. M. Khan, S. S. Naqvi, G. Holmes, and R. Naffa, "On the application of automated machine vision for leather defect inspection and grading: A survey," *IEEE Access*, vol. 7, pp. 176065–176086, 2019.

[2] L. Lythgoe. *Effective Leather Grading*. Accessed: Aug. 2011. [Online]. Available: https://www.satra.com/bulletin/article.php?id=708

[3] R. J. Murphy, S. T. Monteiro, and S. Schneider, "Evaluating classification techniques for mapping vertical geology using field-based hyperspectral sensors," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3066–3080, Aug. 2012.

[4] C. B. Singh, D. S. Jayas, J. Paliwal, and N. D. G. White, "Detection of insect-damaged wheat kernels using near-infrared hyperspectral imaging," *J. Stored Products Res.*, vol. 45, no. 3, pp. 151–158, Jul. 2009.

[5] J. Qin, T. F. Burks, M. A. Ritenour, and W. G. Bonn, "Detection of citrus canker using hyperspectral reflectance imaging with spectral information divergence," *J. Food Eng.*, vol. 93, no. 2, pp. 183–191, 2009.

[6] D. P. Ariana, R. Lu, and D. E. Guyer, "Near-infrared hyperspectral reflectance imaging for detection of bruises on pickling cucumbers," *Comput. Electron. Agricult.*, vol. 53, no. 1, pp. 60–70, 2006.

[7] S.-Y. Chen, C.-Y. Chang, C.-S. Ou, and C.-T. Lien, "Detection of insect damage in green coffee beans using VIS-NIR hyperspectral imaging," *Remote Sens.*, vol. 12, no. 15, p. 2348, Jul. 2020.

[8] E. L. Hestir, V. E. Brando, M. Bresciani, C. Giardino, E. Matta, P. Villa, and A. G. Dekker, "Measuring freshwater aquatic ecosystems: The need for a hyperspectral global mapping satellite mission," *Remote Sens. Environ.*, vol. 167, pp. 181–195, Sep. 2015.

[9] K. C. Tiwari, M. K. Arora, and D. Singh, "An assessment of independent component analysis for detection of military targets from hyperspectral images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 13, no. 5, pp. 730–740, Oct. 2011.

[10] T. Vaarala, M. Aikio, and H. Keraenen, "Advanced prism-grating-prism imaging spectrograph in online industrial applications," *Proc. SPIE*, vol. 3101, Aug. 1997, pp. 322–330.

[11] E. Herrala, T. Hyva, O. Voutilainen, and J. Lammasniemi, "An optoelectronic sensor system for industrial multipoint and imaging spectrometry," *Sens. Actuators A, Phys.*, vol. 61, nos. 1–3, pp. 335–338, Jun. 1997.

[12] G. Candiani, N. Picone, L. Pompilio, M. Pepe, and M. Colledani, "Characterization of fine metal particles using hyperspectral imaging in automatic WEEE recycling systems," in *Proc. 7th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Jun. 2015, pp. 1–4.

[13] A. C. Karaca, A. Erturk, M. K. Gullu, M. Elmas, and S. Erturk, "Automatic waste sorting using shortwave infrared hyperspectral imaging system," in *Proc. 5th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Jun. 2013, pp. 1–4.

[14] P. Tahmasebi and A. Hezarkhani, "Application of a modular feedforward neural network for grade estimation," *Natural Resour. Res.*, vol. 20, no. 1, pp. 25–32, 2011.

[15] L. Jian, H. Wei, and H. Bin, "Research on inspection and classification of leather surface defects based on neural network and decision tree," in *Proc. Int. Conf. Comput. Design Appl.*, Jun. 2010, pp. V2-381–V2-384.

[16] H.-Q. Bong, Q.-B. Truong, H.-C. Nguyen, and M.-T. Nguyen, "Vision-based inspection system for leather surface defect detection and classification," in *Proc. 5th NAFOSTED Conf. Inf. Comput. Sci. (NICS)*, Nov. 2018, pp. 300–304.

[17] C. Kwak, J. A. Ventura, and K. Tofang-Sazi, "Automated defect inspection and classification of leather fabric," *Intell. Data Anal.*, vol. 5, no. 4, pp. 355–370, 2001.

[18] S.-T. Liong, D. Zheng, Y.-C. Huang, and Y. S. Gan, "Leather defect classification and segmentation using deep learning architecture," *Int. J. Comput. Integr. Manuf.*, vol. 33, nos. 10–11, pp. 1105–1117, Nov. 2020.

[19] S.-T. Liong, Y. S. Gan, Y.-C. Huang, C.-A. Yuan, and H.-C. Chang, "Automatic defect segmentation on leather with deep learning," 2019, *arXiv:1903.12139*. [Online]. Available: https://arxiv.org/abs/1903.12139

[20] P. Villar, M. Mora, and P. Gonzalez, "A new approach for wet blue leather defect segmentation," in *Proc. Iberoamer. Congr. Pattern Recognit.* Berlin, Germany: Springer, 2011, pp. 591–598.

[21] R. Viana, R. B. Rodrigues, M. A. Alvarez, and H. Pistori, "SVM with stochastic parameter selection for bovine leather defect classification," in *Proc. Pacific-Rim Symp. Image Video Technol.* Berlin, Germany: Springer, 2007, pp. 600–612.

[22] R. F. Pereira, C. M. S. Medeiros, and P. P. R. Filho, "Goat leather quality classification using computer vision and machine learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.

[23] H. Pistori, W. A. Paraguassu, P. S. Martins, M. P. Conti, M. A. Pereira, and M. A. Jacinto, "Defect detection in raw hide and wet blue leather," in *Computational Modelling of Objects Represented in Images*. Boca Raton, FL, USA: CRC Press, 2018, pp. 355–360.

[24] C. Yeh and D.-B. Perng, "Establishing a demerit count reference standard for the classification and grading of leather hides," *Int. J. Adv. Manuf. Technol.*, vol. 18, no. 10, pp. 731–738, 2001.

[25] M. Aslam, T. M. Khan, S. S. Naqvi, G. Holmes, and R. Naffa, "Ensemble convolutional neural networks with knowledge transfer for leather defect classification in industrial settings," *IEEE Access*, vol. 8, pp. 198600–198614, 2020.

[26] F.-C. Chen and M. R. Jahanshahi, "NB-CNN: Deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4392–4400, May 2018.

[27] J. Chen, Z. Liu, H. Wang, A. Nunez, and Z. Han, "Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 2, pp. 257–269, Dec. 2017.

[28] S. Martelli, L. Mazzei, C. Canali, P. Guardiani, S. Giunta, A. Ghiazza, I. Mondino, F. Cannella, V. Murino, and A. Del Bue, "Deep endoscope: Intelligent duct inspection for the avionic industry," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1701–1711, Apr. 2018.

[29] S. Lu, F. Jian, H. Zhang, J. Liu, and Z. Wu, "An estimation method of defect size from MFL image using visual transformation convolutional neural network," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 213–224, Jan. 2018.

[30] J. Sun, C. Li, X.-J. Wu, V. Palade, and W. Fang, "An effective method of weld defect detection and classification based on machine vision," *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6322–6333, Dec. 2019.

[31] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[32] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2013.

[33] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017.

[34] P. W. Mausel, W. J. Kramber, and J. K. Lee, "Optimum band selection for supervised classification of multispectral data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 1, pp. 55–60, Jan. 1990.

[35] C.-I. Chang, *Hyperspectral Data Processing: Algorithm Design and Analysis*. Hoboken, NJ, USA: Wiley, 2013.

[36] J. C. Harsanyi, "Detection and classification of subpixel spectral signatures in hyperspectral image sequences," Ph.D. dissertation, Univ. Maryland, Baltimore County, Baltimore, MD, USA, 1993.

[37] W. H. Farrand and J. C. Harsanyi, "Mapping the distribution of mine tailings in the Coeur d'Alene River Valley, Idaho, through the use of a constrained energy minimization technique," *Remote Sens. Environ.*, vol. 59, no. 1, pp. 64–76, Jan. 1997.

[38] C.-I Chang, "Target signature-constrained mixed pixel classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 5, pp. 1065–1081, May 2002.

[39] S.-Y. Chen, C. Lin, C.-H. Tai, and S.-J. Chuang, "Adaptive window-based constrained energy minimization for detection of newly grown tree leaves," *Remote Sens.*, vol. 10, no. 2, p. 96, Jan. 2018.

[40] S.-Y. Chen, C. Lin, S.-J. Chuang, and Z.-Y. Kao, "Weighted background suppression target detection using sparse image enhancement technique for newly grown tree leaves," *Remote Sens.*, vol. 11, no. 9, p. 1081, May 2019.

[41] P.-E. Danielsson, "Euclidean distance mapping," *Comput. Graph. Image Process.*, vol. 14, no. 3, pp. 227–248, 1980, doi: 10.1016/0146-664X(80)90054-4.

[42] C.-I. Chang, "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1927–1932, Aug. 2000.

[43] Z. Zou and Z. Shi, "Hierarchical suppression method for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 330–342, Jan. 2015.

[44] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[45] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[46] H. Zhang, L. Meng, X. Wei, X. Tang, X. Tang, X. Wang, B. Jin, and W. Yao, "1D-convolutional capsule network for hyperspectral image classification," 2019, *arXiv:1903.09834*. [Online]. Available: https://arxiv.org/abs/1903.09834

[47] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[48] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[49] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, May 2015, pp. 1520–1528.

[50] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 3–11.

[51] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*. [Online]. Available: https://arxiv.org/abs/1804.03999

[52] S.-Y. Chen, C. Lin, G.-J. Li, Y.-C. Hsu, and K.-H. Liu, "Hybrid deep learning models with sparse enhancement technique for detection of newly grown tree leaves," *Sensors*, vol. 21, no. 6, p. 2077, Mar. 2021.

[53] S. Qamar, H. Jin, R. Zheng, P. Ahmad, and M. Usama, "A variant form of 3D-UNet for infant brain segmentation," *Future Gener. Comput. Syst.*, vol. 108, pp. 613–623, Jul. 2020.

[54] A. G.-U. Juarez, R. Selvan, Z. Saghir, and M. de Bruijne, "A joint 3D UNet-graph neural network-based method for airway segmentation from chest CTs," in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Cham, Switzerland: Springer, 2019, pp. 583–591.

[55] U. Baid, S. Talbar, S. Rane, S. Gupta, M. H. Thakur, A. Moiyadi, S. Thakur, and A. Mahajan, "Deep learning radiomics algorithm for gliomas (DRAG) model: A novel approach using 3D UNet based deep convolutional neural network for predicting survival in gliomas," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2018, pp. 369–379.

[56] P. Ramachandran, B. Zoph, and Q. V. Le, "Swish: A self-gated activation function," 2017, *arXiv:1710.05941*. [Online]. Available: https://arxiv.org/abs/1710.05941

[57] S. Pan, W. Zhang, W. Zhang, L. Xu, G. Fan, J. Gong, B. Zhang, and H. Gu, "Diagnostic model of coronary microvascular disease combined with full convolution deep network with balanced cross-entropy cost function," *IEEE Access*, vol. 7, pp. 177997–178006, 2019.

[58] N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1552–1565, Jul. 2004.

[59] C. Yu, L.-C. Lee, C.-I Chang, B. Xue, M. Song, and J. Chen, "Band-specified virtual dimensionality for band selection: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2822–2832, May 2018.

[60] Y. Wang, L. Wang, C. Yu, E. Zhao, M. Song, C.-H. Wen, and C.-I. Chang, "Constrained-target band selection for multiple-target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6079–6103, Aug. 2019.

[61] X. Liu, L. Gao, B. Zhang, X. Zhang, and W. Luo, "An improved MNF transform algorithm on hyperspectral images with complex mixing ground objects," in *Proc. Congr. Image Signal Process.*, vol. 3, 2008, pp. 479–483.

[62] K. H. Zou, A. J. O'Malley, and L. Mauri, "Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models," *Circulation*, vol. 115, no. 5, pp. 654–657, Feb. 2007.

[63] M. L. McHugh, "Interrater reliability: The Kappa statistic," *Biochem. Med.*, vol. 22, no. 3, pp. 276–282, 2012.

[64] S. Kosub, "A note on the triangle inequality for the Jaccard distance," *Pattern Recognit. Lett.*, vol. 120, pp. 36–38, Apr. 2019.

SHIH-YU CHEN (Member, IEEE) received the B.S. degree in electrical engineering from Da-Yeh University, in 2005, the M.S. degree in electrical engineering from the National Chung Hsing University, in 2010, and the Ph.D. degree in electrical engineering from the University of Maryland, Baltimore County (UMBC), in 2014. He has been an Associate Professor with the Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Taiwan, since 2018. His research interests include multi/hyperspectral image processing, medical signal, image processing, and pattern recognition.

YU-CHIH CHENG received the M.S. degree in computer science and information engineering from the National Yunlin University of Science and Technology, Yunlin, Taiwan, in 2021.

His current research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning.

WEN-LONG YANG received the M.S. degree in computer science and information engineering from the National Yunlin University of Science and Technology, Yunlin, Taiwan, in 2020. His research interests include image processing, machine learning and deep learning for the development and application of algorithms of plasma sources, and fabrication of micro- or nanostructured surfaces.

MEI-YUN WANG is currently pursuing the master's degree with Yunlin National University of Science and Technology, Yunlin, Taiwan. Her current research interests include hyperspectral image processing and analysis, computer vision, and machine learning.

• • •