A NONINVASIVE METHOD TO DETECT DIABETES MELLITUS AND LUNG CANCER USING THE STACKED SPARSE AUTOENCODER

Qi Zhang[*], *Jianhang Zhou*[*], *Bob Zhang*[*]

[*] PAMI Research Group, Department of Computer and Information Science, University of Macau, Taipa, Macau, China

ABSTRACT

Diabetes mellitus and lung cancer are two of the most common fatal diseases in the world, causing considerable deaths every year. However, it is not easy to detect diabetes mellitus and lung cancer efficiently--needing professional medical instruments such as a CT and a qualified individual to perform the Fasting Plasma Glucose test. Considering the risks and various inconveniences with conventional diagnosis methods, noninvasive approaches based on computerized analysis are desired. The aim of this paper is to distinguish patients with diabetes mellitus, lung cancer from healthy people simultaneously by analyzing facial images through the stacked sparse autoencoder. Experimental results on a dataset containing 450 healthy samples, 284 diabetes and 175 lung cancer patients produced the F1-score of 93.57%, 97.54%, 81.56% for detecting healthy, diabetes and lung cancer, respectively, validating the effectiveness of our proposed method.

Index Terms— Diabetes mellitus, lung cancer, facial image, medical biometrics, stacked sparse autoencoder.

1. INTRODUCTION

Diabetes mellitus is a severe metabolic disease, often leading to muscle weakness, cardiovascular disease, stroke, retinopathy, etc. [1]. According to the International Diabetes Federation, the number of diabetic patients worldwide was 124 million in 1997 and 422 million in 2014, respectively, where it will increase to 642 million in 2040 [2]. An approximate 1.6 million deaths were caused by diabetes mellitus directly in 2016, which is the 7th leading cause of death in the USA [3]. Lung cancer is a fatally malignant tumor engendered by irrepressible growth of the lung tissue cells [4]. According to the World Health Organization (WHO), nearly 1.8 million people in the world suffered from lung cancer causing 1.6 million deaths in 2012, ranking first in the cause of death from cancer in males and second in females [5].

The most common way to detect diabetes at present is using a Fasting Plasma Glucose test (FPG) [6]. Typically, the FPG needs blood samples to analyze glucose by piercing a patient's fingertip, which causes some pain for the examinee. Besides this, the FPG also has strict requirements for the testing time, i.e., examinee have to wait ~ 12 hours after eating, before taking this test. For lung cancer, the conventional diagnosis method is to apply medical imaging instruments such as chest X-ray or computed tomography (CT) [7]. However, these methods emit harmful radiation to the human body and may cause some injuries in some patients [8].

Considering these issues in detecting diabetes and lung cancer using traditional medical instruments, it is necessary and urgent to develop noninvasive and efficient approaches to detect these two diseases. Recently, many researchers have proposed and implemented noninvasive methods by extracting color features from facial images to detect various diseases such as diabetes, heart disease, and fatty liver [9, 10, 11, 12]. The idea of these works originated from medical biometrics [13], as distinguishable and unique features can be extracted from a specific disease [14]. The motivation of analyzing a facial image is inspired by traditional medicines, which believe various facial regions can indicate the different statuses of our organs [15]. Although optimistic results were achieved by analyzing color features from facial images [9], to the best of our knowledge, there are few works to investigate the detection of two or more diseases simultaneously using a facial image. The reason is that the conventional color feature extraction method is not effective at retrieving comprehensively effective features or latent information in a facial image, leading to poor results when detecting multiple diseases together.

Nowadays deep learning techniques have become popular and successful in numerous image processing areas such as image denoising and image feature extraction, producing better performances compared to traditional methods [16]. One of the deep learning methods called stacked sparse autoencoder, has been widely applied in dimension reduction and principle feature extraction at present [17]. Therefore, the aim of this study is to detect diabetes and lung cancer simultaneously based on facial images using stacked sparse autoencoder to extract high-level features, which are then subsequently exported to a classifier, categorizing two diseases from healthy samples, respectively.

This work was supported by the University of Macau (MYRG2019-00006-FST).

2. MATERIALS AND METHODS

2.1. Facial image dataset

To capture effective facial images used in this study, a particular photographing device with color correction was applied to reduce the interference by the environment, position, illumination, etc. More details about this photographing device can be found in [18, 19, 20]. The candidate lays his/her chin on a chin rest of this device, allowing their facial image to be taken efficiently (Fig. 1). The original captured facial image size is 768*576 pixels. In order to further decrease the influence of facial organs (e.g., eyes, mouth, etc.) for diagnosis, four facial key blocks [9], i.e., Forehead Block (FHB), Right Cheek Block (RCB), Left Check Block (LCB), and Nose Bridge Block (NBB) with 64*64 pixels were extracted from 1 facial image (Fig. 2).
This dataset contains a total of 909 facial images, including 450 healthy samples and 175 lung cancer patients obtained from the Guangdong Provincial TCM Hospital in 2011 and 2016 as well as 284 diabetes patients captured from Hong Kong Foundation for Research and Development in Diabetes, Prince of Wales Hospital in 2012. Diabetes and lung cancer were diagnosed using the traditional approaches mentioned in Section 1, while individuals were determined to be healthy based on a blood test (another invasive approach) and other commonly used tests. Please note all patients collected were confirmed not have these two diseases simultaneously in this study. All related data was collected in compliance with the ethical standards expressed in the Declaration of Helsinki.

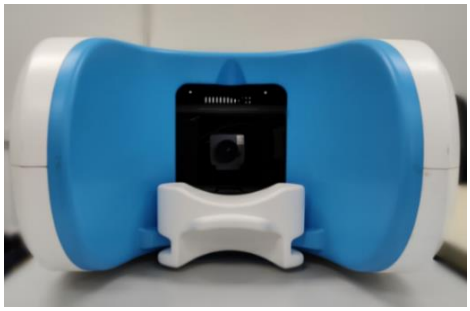

Fig. 1. The special photographing device to capture a facial image.

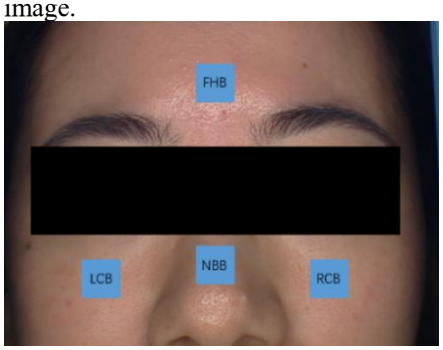

Fig. 2. Four facial key blocks generated from the captured facial image.

In order to compare with our previous color feature extraction method for detecting diseases, the corresponding color features of each candidate's facial key blocks were generated by analyzing the color gamut [9]. For more information about color feature extraction, please refer to our previous work [18].

2.2. Sparse autoencoder

The original sparse autoencoder is an unsupervised learning method, often used to learn effective features from the input sample [21]. The architecture of the sparse autoencoder for extracting effective features is shown in Fig. 3.

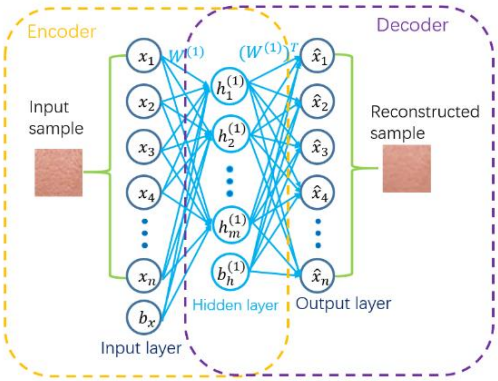

Fig. 3. The architecture of the basic sparse autoencoder, containing encoder and decoder for high-level feature extraction from facial key blocks. The extracted high-level features are in the hidden layer.

Typically, the encoder in the sparse autoencoder contains the input layer and hidden layer, that can transform input sample x into a corresponding representative matrix h, where the subsequent hidden layer can be regarded as a new representative feature of the input sample. The output layer in a decoder aims to reconstruct an approximated sample $\hat{x}$ of the input sample from the hidden representative feature h. The loss can be calculated and minimized between input sample x and its reconstructed sample $\hat{x}$, which is used to find the optimal parameters for training this autoencoder. Basically, the loss would be decreased through the back-propagation algorithm by learning the encoder and decoder neural networks, which generates corresponding weights W and biases b. The cost function of a general sparse autoencoder can be expressed as:

$$\mathcal{L}_{SAE} = \frac{1}{N}\sum_{k=1}^{N}(L(x(k), d_{\hat{\theta}}(e_{\hat{\theta}}(x(k))))) + \alpha\sum_{j=1}^{m} KL(\rho||\hat{\rho}_j) + \beta\|W\|_2^2 \quad (1)$$

The first portion indicates the average sum-of-squares error calculated from the difference between the input sample $x(k)$ and reconstructed sample $\hat{x}(k)$. Here, the m and index j in the second portion represents the number of units in the hidden layer, summing all hidden units in this network. The $KL(\rho||\hat{\rho}_j)$ is the Kullback-Leibler (KL) divergence between the desired activations ρ, and $\hat{\rho}_j$, i.e., the average activation

of the hidden unit *j*. Here, the KL divergence can be defined as:

$$KL(\rho||\hat{\rho}_j) = \rho log\frac{\rho}{\hat{\rho}_j} + (1-\rho)log\frac{1-\rho}{1-\hat{\rho}_j} \quad (2)$$

where $\hat{\rho}_j$ can be expressed as:

$$\hat{\rho}_j = \frac{1}{n}\sum_i [a_j(x^{(i)})] \quad (3)$$

The $x^{(i)}$ represents the *n*th training sample.

The third portion represents a weight decay function, which attempts to reduce the magnitude of the weight, and avoid the overfitting problem. This weight decay portion can be defined as:

$$\|W\|_2^2 = tr(W^T W) = \sum_{l=1}^{n_l}\sum_i^{S_{l-1}}\sum_j^{S_l}(w_{i,j}^{(l)})^2 \quad (4)$$

where n_l represents the number of layers, while S_{l-1} and S_l indicate the number of neurons in layer *l*-1 and *l* respectively. The $w_{i,j}^{(l)}$ is used to connect the *i*th neuron of layer *l*-1 and *j*th neuron of layer *l*.

2.3. Stacked sparse autoencoder

The stacked sparse autoencoder is derived from a simple sparse autoencoder, containing multiple layers of basic sparse autoencoders (Fig. 4). The output of the encoder in each basic sparse autoencoder is regarded as the input to the subsequent encoder of the next sparse autoencoder, where the constructed multiple encoders are successively formed [22]. Compared to the basic sparse autoencoder, the stacked sparse autoencoder is more powerful at extracting high-level features from the input samples [23].

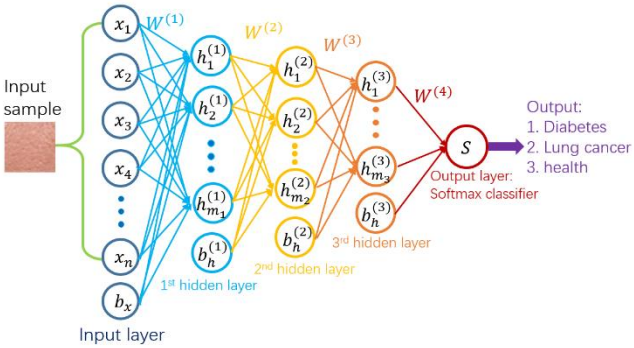

Fig. 4. The architecture of the proposed stacked sparse autoencoder for classifying diabetes, lung cancer, and healthy individuals in this study.

To train the stacked sparse autoencoder in this study, the optimal parameters $\theta = (W, b_h, b_x)$ should be found simultaneously through minimizing the loss between the input sample and the corresponding reconstructed sample. The stacked sparse autoencoder generates the first new feature representation i.e., $h^{(1)}$ for the input facial key blocks through the first hidden layer, which can be expressed as:

$$f: R^{d_x} \rightarrow R^{d_{h^{(1)}}} \quad (5)$$

$$h^{(2)} = f(x) \in R^{d_{h^{(1)}}} \quad (6)$$

where the function *f* aims to convert the input sample to a new representative feature. The second new feature representation can be yielded in the same way, i.e., $h^{(1)}$ is regarded as the input sample through the second hidden layer, before generating $h^{(2)}$. The third new representative feature $h^{(3)}$ is also produced from the third hidden layer similarly.

It can be noticed that each candidate facial key blocks *x(k)* are finally expressed by a high-level representative feature $h^{(3)}$ through the third hidden layer in Fig. 4. Thus, all training facial key blocks can be expressed as:

$$\{h^{(3)}(k), y(k)\}_{k=1}^{N} \quad (7)$$

where *N* is the number of training candidates' facial key blocks, while $h^{(3)}(k), y(k)$ represent high-level features and the corresponding labels, respectively. The three classes: healthy, diabetes and lung cancer are classified in this study. Therefore, $y(k) \in \{0, 1, 2\}$, where 0, 1, and 2 represent the three classes respectively. Please note that the stacked sparse autoencoder learnt here is an unsupervised learning approach, as the information of label $y(k)$ is not used. The corresponding high-level features of the input facial key blocks, as well as their labels are fed to the output layer i.e., softmax classifier.

The softmax classifier [21] is a supervised method producing logistic regression as:

$$f_{W^4}(z) = \frac{1}{1+exp\,(-W^{(4)^T} z)} \quad (8)$$

where $f_{W^4}(\cdot)$ signifies a sigmoid function with parameters W^4. The input z represents the high-level representative feature through the third hidden layer of the stacked sparse autoencoder. The parameters W^4 used in the softmax classifier is trained to minimize the cost function with the training sets in eq. (7).

The parameters θ in the stacked sparse autoencoder and W^4 in the softmax classifier can be determined after training. Then, this proposed model can be applied for detecting diabetes and lung cancer from healthy examinees simultaneously.

3. EXPERIMENTAL RESULTS

3.1. Experimental settings

Based on Section 2.1, four facial key blocks (FHB, LCB, RCB, and NBB) were applied to represent the whole facial image. All blocks are combined together (FHB+LCB+RCB+NBB) and evaluated by our proposed methodology. Besides this, four classifiers: *k*-NN, Collaborative Representation Classifier (CRC), Linear Discriminant Analysis (LDA), and Random Forest [24, 25, 26] using our previously conventional color features extracted from the facial key blocks were also analyzed in this study in order to compare with the proposed model.

Half of the samples in each class in our dataset were applied for training randomly, while the rest of the data were used for testing. Each single block was resized to 32*32, before each resized single facial key block was combined together to construct a new input sample (64*64) for our model. The first,

second, and third hidden layer contains 4000, 1000, and 400 hidden units, respectively, where the output is the label of the three classes through the softmax layer (refer to Fig. 4). To avoid the imbalance problem caused by our collected dataset [27], all experiments were run 30 times randomly, quantified by the Precision, Recall (Sensitivity), Specificity, Accuracy, and F-1 score [28, 29] for each class, respectively. These five metrics were calculated as the following:

$$Presion = \frac{TP}{TP+FP} \quad (9)$$

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

$$Specificity = \frac{TN}{TN+FP} \quad (11)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

$$F1-score = \frac{2TP}{2TP+FN+FP} = \frac{2 \cdot Precision \cdot Recall}{Precision+Recall} \quad (13)$$

where TP is expressed as the number of samples in each class identified correctly (true positive), FP, FN and TN refer to false positive, false negative and true negative errors, respectively.

3.2. Experimental results

Table 1 indicates the corresponding results of Precision by using our proposed model along with the results of the four classifiers applying the color features (refer to Section 3.1). Tables 2 and 3 present the results of Recall and Specificity produced by our proposed model along with the results of the comparison classifiers using the color features extracted from the four facial blocks. Finally, Tables 4 and 5 illustrate the Accuracy and F1-score following the same format as Tables 1 and 2.

Table 1. The Precision (precision ± standard deviation) with *k*-NN, CRC, LDA, random forest and our proposed stacked autoencoder for detecting healthy, diabetes, and lung cancer, respectively.

Class	*k*-NN	CRC	LDA	Random Forest	Our proposed model
Healthy	42.45 ± 6.96%	80.00 ± 5.12%	84.72 ± 4.83%	83.58 ± 2.52%	**93.36±1.10%**
Diabetes	60.94 ± 5.63%	80.19 ± 7.18%	75.09 ± 8.14%	74.15 ± 4.54%	**97.20±1.35%**
Lung cancer	33.96 ± 9.82%	58.49 ± 7.80%	61.70 ± 8.14%	57.17 ± 4.54%	**82.56±4.50%**

Table 2. The Recall (recall ± standard deviation) with *k*-NN, CRC, LDA, random forest and our proposed stacked autoencoder for detecting healthy, diabetes, and lung cancer, respectively.

Class	*k*-NN	CRC	LDA	Random Forest	Our proposed model
Healthy	54.08 ± 3.20%	68.85 ± 4.45%	73.19 ± 5.93%	62.02 ± 3.63%	**93.78±1.15%**
Diabetes	49.80 ± 4.44%	81.11 ± 3.02%	78.53 ± 2.99%	82.16 ± 5.27%	**97.89±1.82%**
Lung cancer	39.33 ± 5.79%	69.79 ± 5.05%	69.81 ± 4.57%	74.05 ± 3.63%	**80.68±3.18%**

Table 3. The Specificity (specificity ± standard deviation) with *k*-NN, CRC, LDA, random forest and our proposed stacked autoencoder for detecting healthy, diabetes, and lung cancer, respectively.

Class	*k*-NN	CRC	LDA	Random Forest	Our proposed model
Healthy	72.09 ± 1.84%	88.75 ± 2.02%	91.63 ± 2.29%	90.89 ± 3.13%	**93.48±2.38%**
Diabetes	80.07 ± 2.42%	90.06 ± 2.59%	86.17 ± 2.73%	85.23 ± 3.02%	**98.72±0.62%**
Lung cancer	68.26 ± 1.86%	81.81 ± 1.80%	82.68 ± 1.52%	82.56 ± 2.04%	**95.91±1.18%**

Table 4. The Accuracy (accuracy ± standard deviation) with *k*-NN, CRC, LDA, random forest and our proposed stacked autoencoder for detecting healthy, diabetes, and lung cancer, respectively.

Class	*k*-NN	CRC	LDA	Random Forest	Our proposed model
Healthy	69.43 ± 2.77%	81.38 ± 3.66%	82.70 ± 3.26%	76.67 ± 3.61%	**93.63±1.03%**
Diabetes	65.22 ± 1.76%	87.48 ± 1.93%	87.23 ± 2.14%	86.92 ± 2.87%	**98.46±0.82%**
Lung cancer	58.55 ± 3.59%	77.55 ± 2.78%	78.11 ± 2.33%	77.17 ± 2.57%	**92.97±1.35%**

Table 5. The F1-score (F1-score ± standard deviation) with *k*-NN, CRC, LDA, random forest and our proposed stacked autoencoder for detecting healthy, diabetes, and lung cancer, respectively.

Class	*k*-NN	CRC	LDA	Random Forest	Our proposed model
Healthy	60.16 ± 3.03%	73.81 ± 2.91%	73.83 ± 5.01%	67.42 ± 2.74%	**93.57±1.05%**
Diabetes	46.61 ± 4.61%	80.48 ± 3.25%	81.38 ± 2.19%	82.72 ± 1.97%	**97.54±1.34%**
Lung cancer	36.22 ± 5.73%	63.33 ± 5.70%	65.34 ± 3.92%	64.32 ± 3.06%	**81.56±3.44%**

From these tables (Tables 1 to 5) it can be noted that our proposed method is superior to the other classifiers using color features extracted from the facial key blocks in all measured metrics for classifying the three classes. Besides the results in the five tables, the final average accuracy of classifying the three classes by applying our proposed model is 92.5%. This compares to 48.28% for *k*-NN, 71.93% using CRC, 73.02% employing LDA, and Random Forest producing 72.14% all applying the original color features (refer to Section 2.1).

4. CONCLUSION

Given the severity and invasiveness of detecting diabetes mellitus and lung cancer, this paper proposed a model based on autoencoder to classify diabetes mellitus and lung cancer from healthy candidates simultaneously. The input facial key blocks of the candidates were processed by a four-layer stacked sparse autoencoder to generate high-level features, which were subsequently classified using a softmax classifier. Compared to the previous color extraction approach, our proposed method can extract high-level features from the facial key blocks (images directly) successfully, achieving the highest results using five metrics on a dataset consisting of 450 healthy samples, 284 diabetes and 175 lung cancer patients, respectively.

5. REFERENCES

[1] Centers for Disease Control and Prevention., "National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011," *Atlanta, GA: US department of health and human services, centers for disease control and prevention,* vol. 201, no. 1, pp. 2568-2569, 2011.

[2] D. Atlas, "International diabetes federation," *IDF Diabetes Atlas, 7th edn. Brussels, Belgium: International Diabetes Federation,* 2015.

[3] World Health Organization, "Diabetes". [online] Available at: https://www.who.int/news-room/fact-sheets/detail/diabetes [Accessed 30 October 2018].

[4] D. S. Ettinger et al., "NCCN guidelines insights: non–small cell lung cancer, version 4.2016," *Journal of the National Comprehensive Cancer Network,* vol. 14, no. 3, pp. 255-264, 2016.

[5] World Health Organization, & World Health Organization. Management of Substance Abuse Unit., *Global status report on alcohol and health, 2014.* World Health Organization, 2014.

[6] A. Tirosh et al., "Normal fasting plasma glucose levels and type 2 diabetes in young men," *New England Journal of Medicine,* vol. 353, no. 14, pp. 1454-1462, 2005.

[7] P. B. Bach et al., "Computed tomography screening and lung cancer outcomes," *Jama,* vol. 297, no. 9, pp. 953-961, 2007.

[8] D. J. Brenner and E. J. Hall, "Computed tomography—an increasing source of radiation exposure," *New England Journal of Medicine,* vol. 357, no. 22, pp. 2277-2284, 2007.

[9] B. Zhang and D. Zhang, "Noninvasive diabetes mellitus detection using facial block color with a sparse representation classifier," *IEEE transactions on biomedical engineering,* vol. 61, no. 4, pp. 1027-1033, 2013.

[10] T. Shu, B. Zhang, and Y. Y. Tang, "Effective heart disease detection based on quantitative computerized traditional chinese medicine using representation based classifiers," *Evidence-Based Complementary and Alternative Medicine,* vol. 2017, 2017.

[11] J. Li, B. Zhang, and D. Zhang, "Joint discriminative and collaborative representation for fatty liver disease diagnosis," *Expert Systems with Applications,* vol. 89, pp. 31-40, 2017.

[12] P. Zhang and B. Zhang, "A study of diabetes mellitus detection using sparse representation algorithms with facial block color features," in *2016 IEEE International Conference on Signal and Image Processing (ICSIP),* 2016: IEEE, pp. 563-567.

[13] D. Zhang and W. Zuo, *Medical biometrics: Computerized TCM data analysis.* World Scientific, 2016.

[14] Healthwise Staff, " Symptoms of High Blood Sugar". [Online]. Avialble at: https://www.uwhealth.org/health/topic/special/symptoms-of-highblood-sugar/aa21178.html, Accessed on: Jul. 25, 2018.

[15] B. Zhu and H. Wang, *Diagnostics of traditional Chinese medicine.* Singing Dragon, 2010.

[16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning.* MIT press, 2016.

[17] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Advances in neural information processing systems,* 2007, pp. 153-160.

[18] X. Wang and D. Zhang, "An optimized tongue image color correction scheme," *IEEE Transactions on information technology in biomedicine,* vol. 14, no. 6, pp. 1355-1364, 2010.

[19] T. Shu, B. Zhang, and Y. Y. Tang, "An improved noninvasive method to detect Diabetes Mellitus using the Probabilistic Collaborative Representation based Classifier," *Information Sciences,* vol. 467, pp. 477-488, 2018.

[20] T. Shu, B. Zhang, and Y. Tang, "Novel noninvasive brain disease detection system using a facial image sensor," *Sensors*, vol. 17, no. 12, p. 2843, 2017.

[21] A. Ng, "Sparse autoencoder," *CS294A Lecture notes,* vol. 72, no. 2011, pp. 1-19, 2011.

[22] J. Xu et al., "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE transactions on medical imaging,* vol. 35, no. 1, pp. 119-130, 2015.

[23] C. Tao, H. Pan, Y. Li and Z. Zou, "Unsupervised spectral–spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification," *IEEE Geoscience and remote sensing letters,* vol. 12, no. 12, pp. 2438-2442, 2015.

[24] P. Zhu, L. Zhang, Q. Hu, and S. C. Shiu, "Multi-scale patch based collaborative representation for face recognition with margin distribution optimization," in *European Conference on Computer Vision,* 2012: Springer, pp. 822-835.

[25] C. M. Bishop, *Pattern recognition and machine learning.* springer, 2006.

[26] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition,* 1995, vol. 1: IEEE, pp. 278-282.

[27] G. Collell, D. Prelec and K. R. Patil, "A simple plug-in bagging ensemble based on threshold-moving for classifying binary and multiclass imbalanced data," *Neurocomputing,* vol. 275, pp. 330-340, 2018.

[28] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies,* vol. 2, no. 1, pp. 37-63, 2011.

[29] D. G. Altman and J. M. Bland, "Diagnostic tests. 1: Sensitivity and specificity," *BMJ: British Medical Journal,* vol. 308, no. 6943, pp. 1552-1552, 1994.