

Automated Red Deer Algorithm with Deep Learning Enabled Hyperspectral Image Classification

B. Chellapra^{1,*}, D. Manohari², K. Periyakaruppan³ and M. S. Kavitha⁴

¹Department of Information Technology, Karpagam Institute of Technology, Coimbatore, 641032, Tamilnadu, India

²Department of Computer Science and Engineering, St. Joseph's Institute of Technology, Chennai, 600119, India

³Department of Computer Science & Engineering, SNS College of Engineering, Coimbatore, 641107, India

⁴Department of Computer Science & Engineering, SNS College of Technology, Coimbatore, 641035, India

*Corresponding Author: B. Chellapra. Email: chellapra@gmail.com

Received: 14 March 2022; Accepted: 14 April 2022

Abstract: Hyperspectral (HS) image classification is a hot research area due to challenging issues such as existence of high dimensionality, restricted training data, etc. Precise recognition of features from the HS images is important for effective classification outcomes. Additionally, the recent advancements of deep learning (DL) models make it possible in several application areas. In addition, the performance of the DL models is mainly based on the hyperparameter setting which can be resolved by the design of metaheuristics. In this view, this article develops an automated red deer algorithm with deep learning enabled hyperspectral image (HSI) classification (RDADL-HIC) technique. The proposed RDADL-HIC technique aims to effectively determine the HSI images. In addition, the RDADL-HIC technique comprises a NASNetLarge model with Adagrad optimizer. Moreover, RDA with gated recurrent unit (GRU) approach is used for the identification and classification of HSIs. The design of Adagrad optimizer with RDA helps to optimally tune the hyperparameters of the NASNetLarge and GRU models respectively. The experimental results stated the supremacy of the RDADL-HIC model and the results are inspected in terms of different measures. The comparison study of the RDADL-HIC model demonstrated the enhanced performance over its recent state of art approaches.

Keywords: Hyperspectral images; image classification; deep learning; adagrad optimizer; nasnetlarge model; red deer algorithm

1 Introduction

Recently, hyperspectral (HS) images have gained considerable interest from the researcher. Since the HS images are strong in resolving energy for acceptable spectrum with a wide-ranging function in mining, [1], armed forces, clinical and climatic fields. The application of HS images has functioned on imaging spectrometers [2]. In the past decades, the imaging spectrum has been invented. It is employed for images in transparent, ultraviolet, mid-infrared, and closer-infrared areas of electromagnetic waves. Therefore, the HS images comprised of several bands, densely data, and maximal spectral resolution. [3] The



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

computation approach of HS remote sensing images includes noise limitation, image correction, classification, transformation, and dimension reduction. With the consideration, the advantages of productive spectral data, classifier methods [4].

Deep learning (DL) based methods have attained stimulating efficiency in enormous application. In DL, the Convolution Neural Network (CNN) performs a significant part in computing visual-based problems [5]. CNN is biological-based and multi-layer classification of DL method that employs neural network (NN) trained with dedicated image pixel value for generating the outcomes [6]. With the huge source of training information and creative implementation on graphics processing unit (GPU) [7], CNN has exceeded the conventional methods for human tasks, on object prediction, huge vision-based tasks, face analysis, house value digit classification, scene labeling, and image classification. After that, the visual task in CNN has applied in another region namely speech examination [8]. It can be attained by forecasting the potential groups of techniques to learn visual image content, by presenting the current outcomes on visual image classification and visual-based problems [9,10].

In [11], a new hybrid dilated convolution guided feature filtering and enhance network (HDCFE-Net) technique was presented for classifying HS images (HSI). The dilated convolutional is decrease the spatial feature loss with no decreasing the receptive domain and is attain distant feature. It is also be integrated with standard convolutional with no losing their original data. Cai et al. [12] presented a new Triple attention Guided Residual Dense and bidirectional long short-term memory (Bi-LSTM) network (TARDB-Net) for avoiding repetitive features but improving feature fusion abilities that ultimately enhance the capability for classifying HSI. Mou et al. [13] presented a novel graph based semisupervised network named a non-local graph convolutional network (non-local GCN). Different present CNN and recurrent neural network (RNN) that obtain pixel or patch of an HSI as inputs, this network gets the entire image (combining both labeled as well as unlabeled data). Especially, a non-local graph was initially computed. At last, the semisupervised learning of network was completed by utilizing cross-entropy error entire labeled instances.

Cao et al. [14] projected a new DL technique to HSI classifier that combines both active learning and DL as unified frameworks. Primary, it can be trained a CNN with restricted amount of labeled pixels. Then, it can actively choose one of the informative pixels in the candidate pool for labeling. Afterward, the CNN has been fine tuned with novel trained set created by including the recently labeled pixel. At last, Markov random field (MRF) was employed to enforce class labels to more boost the classifier performance. Hong et al. [15] presented a solution for addressing this problem by locally removing invariant features in HSI from both spatial as well as frequency domains, utilizing this technique called invariant attribute profiles (IAPs).

This paper develops an automated red deer algorithm with deep learning enabled HSI classification (RDADL-HIC) technique. The proposed RDADL-HIC technique comprises a NASNetLarge model with Adagrad optimizer. Moreover, RDA with gated recurrent unit (GRU) classifier is used for the identification and classification of HSIs. The design of Adagrad optimizer with RDA helps to optimally tune the hyperparameters of the NASNetLarge and GRU models respectively. The experimentation outcomes results stated the enhanced outcomes of the RDADL-HIC model and the results are inspected interms of different measures.

2 The Proposed Model

In this article, a new RDADL-HIC technique has been presented to effectively determine the HSI images. The RDADL-HIC technique mainly involves a NASNetLarge model with Adagrad optimizer. Followed by, RDA with GRU classifier is used for the identification and classification of HSIs. The design of Adagrad optimizer with RDA helps to optimally tune the hyperparameters of the NASNetLarge and GRU models respectively.

2.1 Feature Extraction

At the primary level, the features are produced from the HS images by the use of NASNetLarge model [16]. The NASNet-Large segmentation network is a kind of CNN (as shown in Fig. 1) that comprises encoding and decoding that is afterward a classifier layer. There are 2 important variances in our method relating to Segnet that utilizes the pre-training VGG16 network to the encoded. Our Nasnet-Large-decoder net utilizes the primary 414 layers of Nasnet-Large net (that is network highly training for ImageNet classifier) as the encoded for decomposing image. It can be select the initial 414 layers as the size of last layer was nearby the size of original images, thus it is not losing more data. When it can be select the final layer as the feature extracting layer, it is destroyed important structural data of objects as the final layer is further appropriate to classifier, instead of segmentation. It does not utilize the pre-training weighted but retrained the net utilizing novel data for fitting Nasnet-Large was considerably distinct in ImageNet. Also, the decoded was distinct and there are no pooling indices in this method as Nasnet-Large net is produced brief data to decode. The suitable decoded is up-sample their input feature map utilizing max pooling layer.

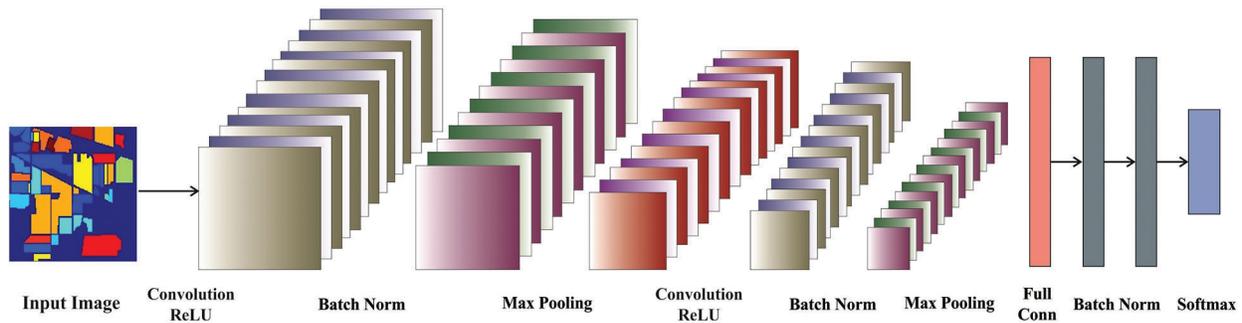


Figure 1: Structure of CNN

There are 4 blocks from the decoded. All the blocks start with up-sampling that is expand the feature map, then convolutional and rectified linear unit (ReLU). The batch normalization layer is then executed for all of the maps. The primary decoder that is nearby the final encoded is generate a multi-channel feature map. It can be related to Segnet that it creates a distinct amount of sizes and channels as its encoded input. The last output of final decoder layer was able to a trainable softmax classification that generates a K channel image of probability in which K is the amount of class labels. The forecasted segmentation equals the class with maximal probability at all the pixels.

For tuning the hyperparameters that exist in the NasNet model, the Adagrad optimizer is applied. Adagrad is a technique to gradient-based optimized that adjusts the rate of learning to the parameters. Adagrad enhanced significantly and it can be utilized to train large scale neural network (NNs) at Google. Besides, it is utilized Adagrad for training Word Embedded. It can be utilized for setting gradient of object function was represented by $g_{i,j}$, in terms of parameter θ_i at timestep t .

$$g_{t,i} = \nabla_{\theta} J(\theta) \tag{1}$$

The SGD upgrade to all the parameters θ_i at all the timestep t then develops:

$$\theta_{t+1,i} = \theta_{i,j} - \alpha \cdot g_{t,i} \tag{2}$$

Their update rule, Adagrad changes the common rate of learning at all the timesteps t to all the parameters θ_i dependent upon the past gradient which is calculated for θ_i :

$$\theta_{t+1,i} = \theta_{i,j} - \frac{\alpha}{\sqrt{G_{t,ii} + \varepsilon}} \cdot g_{t,i} \quad (3)$$

$G_t \in R^{d \times d}$ at this time is a diagonal matrix in which all diagonal elements i , i refers the sum of squares of the gradients in terms of θ_i up to timestep t , but ε refers the smoothing term which avoids division by zero. The default value of the rate of learning for Adagrad is 0.01. The structure of GRU model is depicted in Fig. 2.

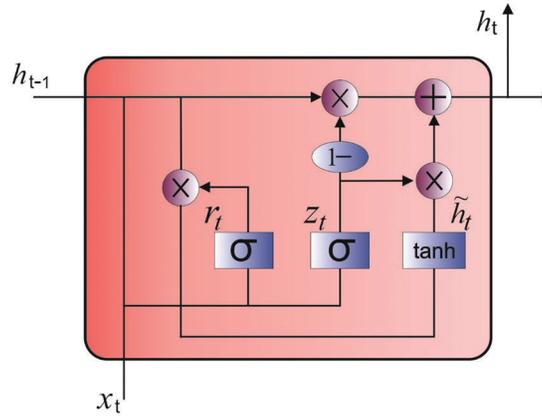


Figure 2: GRU model

2.2 HS Image Classification

For HS image classification, the GRU model is employed which determines the class labels efficiently [17]. In GRU, the two vectors in long short term memory (LSTM) cells are associated with one vector o_t . One gate controller controls the integrated form of forgetting and input gates. When z_t output is one, the forget gate was opened, and input gate was closed, when z_t was zero, the forget gate was closed, and the input gate was opened. Instinctively, the reset gate determines by integrating an input with previous memory, and the upgrade gate elects the prior memory information has preserved to estimate the new state. The variance in the LSTM, however the variations formerly determined by:

$$r_t = \sigma(W_{xr}^T \cdot x_t + W_{or}^T \cdot o_{t-1} + b_r). \quad (4)$$

$$z_t = \sigma(W_{xz}^T \cdot x_t + W_{oz}^T \cdot o_{t-1} + b_z). \quad (5)$$

$$\tilde{o}_t = \tanh(W_{x\tilde{o}}^T \cdot x_t + W_{o\tilde{o}}^T \cdot (r_t \otimes o_{t-1}) + b_{\tilde{o}}). \quad (6)$$

$$o_t = z_t \otimes o_{t-1} + (1 - z_t) \otimes \tilde{o}_t. \quad (7)$$

whereas W_{xr} , W_{xz} , $W_{x\tilde{o}}$ denotes weight matrices to corresponding input vector, W_{or} , W_{oz} , $W_{o\tilde{o}}$ denotes the weight matrices of prior time step, and b_r , b_z , $b_{\tilde{o}}$ represent bias.

2.3 Parameter Optimization

In order to effectually determine the parameters involved in the GRU model, the RDA is exploited [18]. RDA is analogous to other meta-heuristics, initiated by an arbitrary population viz. red deer (RD) counterpart. An optimal RD from the population is carefully chosen and named the “male RD (MRD)”, then the residue was called “hinds” [18]. The iteration presentation of method was coined. To arithmetical process the RDA initiates by Eq. (8) while a key population to the RD is created:

$$RD = X1, X2, X3, \dots, X_{Nvar} \tag{8}$$

Next, the fitness of each member from the population is calculated as follows:

$$Value = f(RD) = f(X1, X2, X3, \dots, X_{Nvar}) \tag{9}$$

MRD is try to boost the grace by roaring in this phase. Therefore, the outcomes, the roaring method could fail or succeed. Once the key functions of neighbor are MRD, it is replaced with the prediction once it is larger than the previous MRD. The succeeding formula is shown below:

$$male_{new} = \begin{cases} male_{old} + a_1(UB - LB) * a_2 + LB, & \text{if } a_3 \geq 0.5 \\ male_{old} - a_1(UB - LB) * a_2 + LB, & \text{if } a_3 < 0.5 \text{ is less than } 0 \end{cases} \tag{10}$$

UB and LB bounds the searching region to create apt male neighborhood solution. The top and lower limits of searching region. Clear the present location of MRD is $male_{old}$, and the following location is $male_{new}$. The sum of male commanders is estimated by the following:

$$N_c = round(\gamma \cdot N_{male}) \tag{11}$$

whereas N_c denotes the sum of commanders that naturally males, γ indicates the arbitrary amount among [0,1] N_{male} characterizes the sum of whole males. It is noted that γ denotes the method initials value. It is value ranges within [0,1]. Eventually, the sum of stags is assessed by the succeeding formula:

$$N_s = N_{male} - N_c \tag{12}$$

The fighting method can be modelled by the following 2 arithmetical formulas:

$$new_1 = \frac{C + S}{2} + b_1((UB - LB) * b_2 + LB) \tag{13}$$

The subsequent equation is employed to calculate commander standardization power.

$$P_n = \left| \frac{V_n}{\sum_{i=1}^{a_i} V_i} \right| \tag{14}$$

The sum total of hinds of a harem is calculated by:

$$N.harem_n = round(P_n \cdot N_{hind}) \tag{15}$$

whereas N_{hind} denotes the complete quantity of hinds. It could be deer mating presentation was executed by commander with proportion of hinds from their harem.

$$N.harem_n^{mate} = round(\alpha \cdot N.harem_n) \tag{16}$$

The sum total of hinds from the n_{th} harem that mate with leader represent $N.harem_n^{mate}$. Regarding the solution space, it selects $N.harem_n^{mate}$ of $N.harem_k$ arbitrarily. In general, the mating technique has described in the following:

$$offs = \frac{C + Hind}{2} + (UB - LB) \times c \tag{17}$$

It chooses a harem arbitrarily (k) and permits the male commander to mate with β percentage of harem hinds. The quantity of hinds from the harem that mate with commander is estimated by the succeeding formula:

$$N.harem_k^{mate} = round(\beta \cdot N.harem_k) \quad (18)$$

whereas $N.harem_k^{mate}$ denotes the quantity of hinds from the k_{th} harem that mate with commander. It is noted that the mating procedure was implemented by Eq. (17). Fig. 3 depicts the process involved in RDA. The distance among a stag and hind from J dimension space is estimated by [19]:

$$d_i = \left\{ \sum_{j \in J} (stag_j - hind_j^i)^2 \right\}^{\frac{1}{2}} \quad (19)$$

Now d_i denotes the distance among the i_{th} hind and stag.

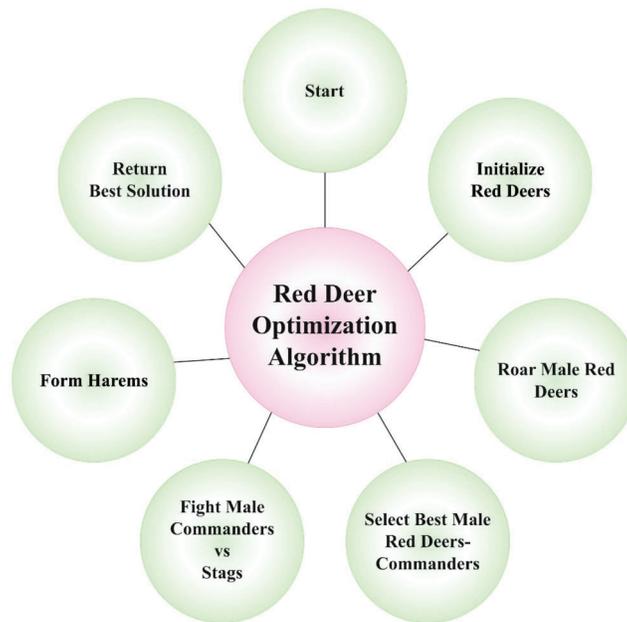


Figure 3: Process involved in RDA

3 Results Analysis

The performance validation of the RDADL-HIC model is examined using Indian Pines dataset (IPD) and Pavia university dataset (PUD), available at http://www.ehu.eus/ccwintco/index.php/HS_Remote_Sensing_Scenes. The first IPD includes 10249 data instances under 16 class labels. The second PUD encompasses 42776 data instances with 9 class labels.

Fig. 4 demonstrates the training accuracy (TA) and validation accuracy (VA) offered by the RDADL-HIC model on IPD. The figure indicated that the RDADL-HIC model has provided closer TA and VA values with an increase in epoch count. It is observable that the VA is certainly higher than TA. Fig. 5 validates the training loss (TL) and validation loss (VL) provided by the RDADL-HIC model on IPD. The figure designated that the RDADL-HIC model has delivered lower TL and VL with an increase in epoch count. It is noticeable that the VL is definitely lower compared to TL.

Fig. 6 illustrates the precision-recall investigation of the RDADL-HIC model on the IPD. The figure indicated that the RDADL-HIC model has accomplished maximum precision-recall values on the distinct class labels.

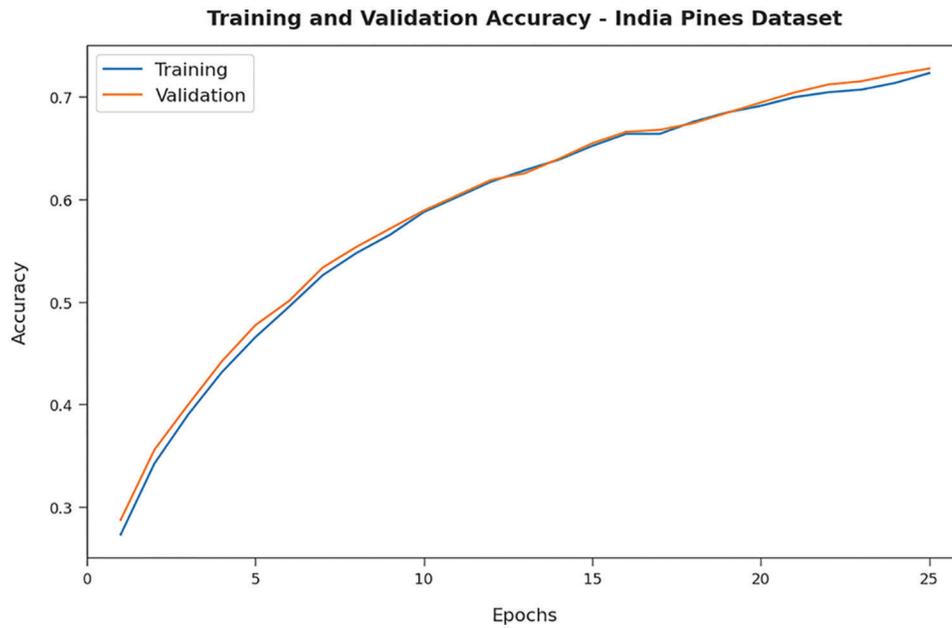


Figure 4: TA and VA of RDADL-HIC model on IPD

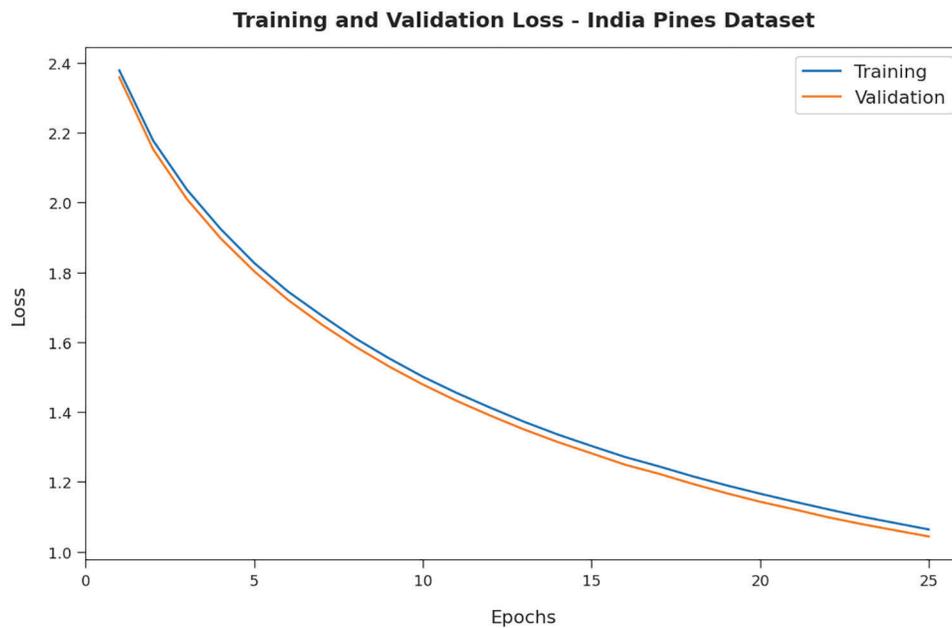


Figure 5: TL and VL of RDADL-HIC model on IPD

Fig. 7 exemplifies the ROC assessment of the RDADL-HIC model on the IPD. The figure indicated that the RDADL-HIC model has gained improved ROC values on the distinct class labels.

Next, the experimental validation of the RDADL-HIC model is tested using PUD. Fig. 8 illustrates the sample input image with its ground truth version. Fig. 9 proves the TA and VA obtainable by the RDADL-HIC model on PUD. The figure indicated that the RDADL-HIC model has provided closer TA and VA values with an increase in epoch count. It is observable that the VA is certainly higher than TA.

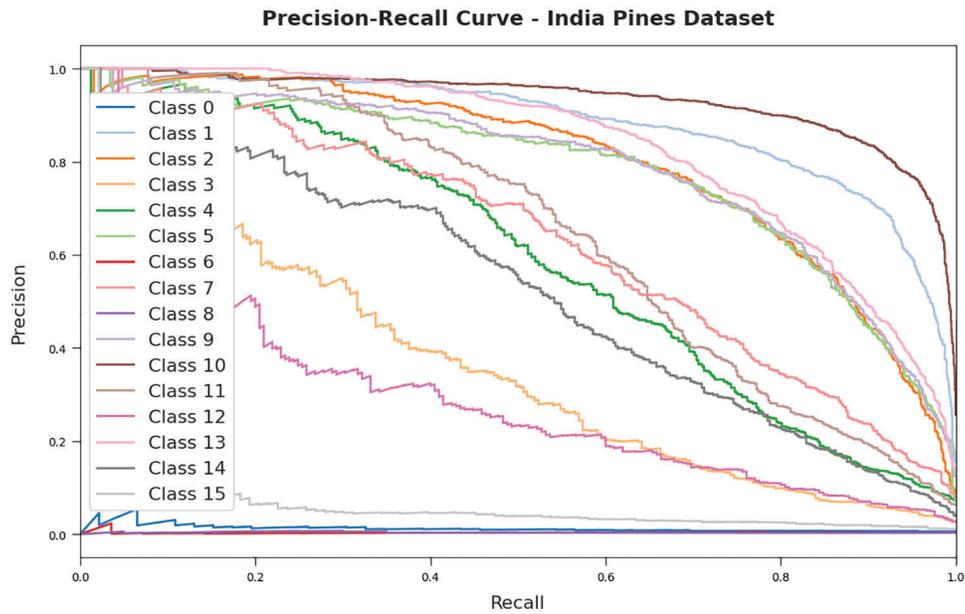


Figure 6: Precision-recall curves of RDADL-HIC model on IPD

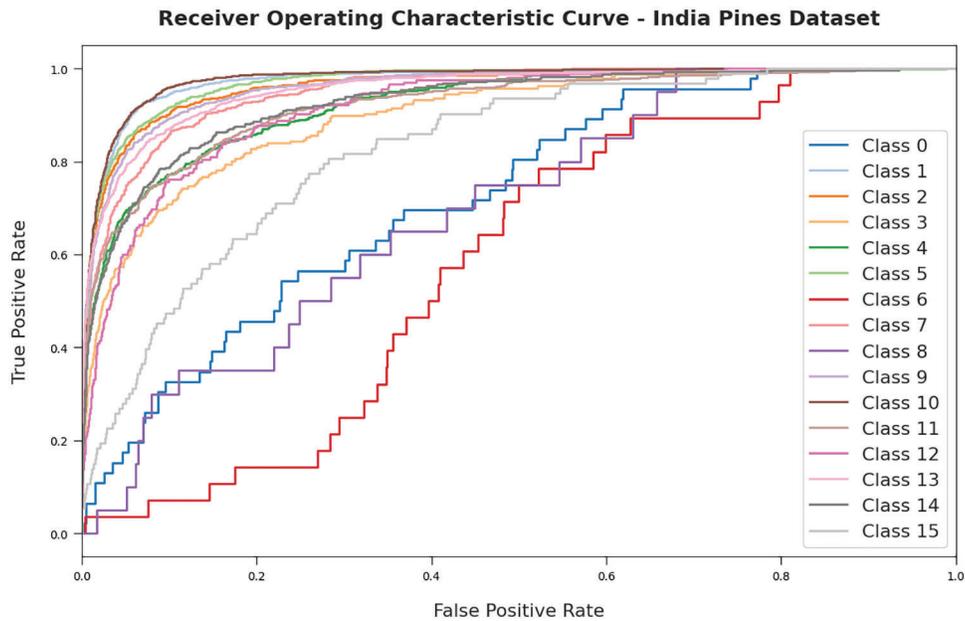


Figure 7: ROC curves of RDADL-HIC model on IPD

Fig. 10 confirms the TL and VL delivered by the RDADL-HIC model on PUD. The figure designated that the RDADL-HIC model has delivered lower TL and VL with an increase in epoch count. It is noticeable that the VL is definitely lower compared to TL.

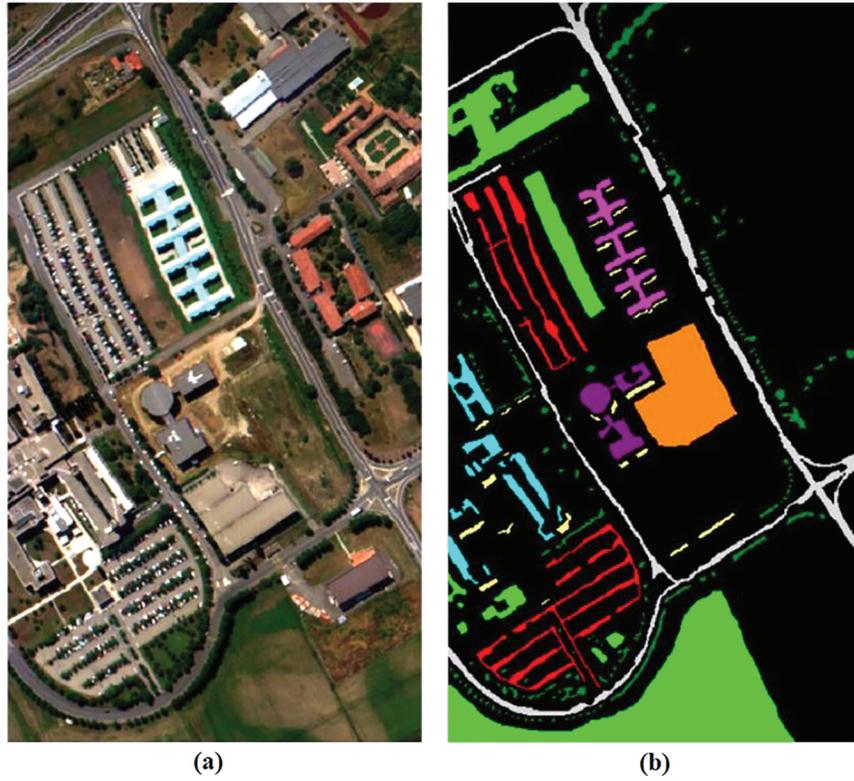


Figure 8: PUD. (a) Sample Image, (b) Ground Truth

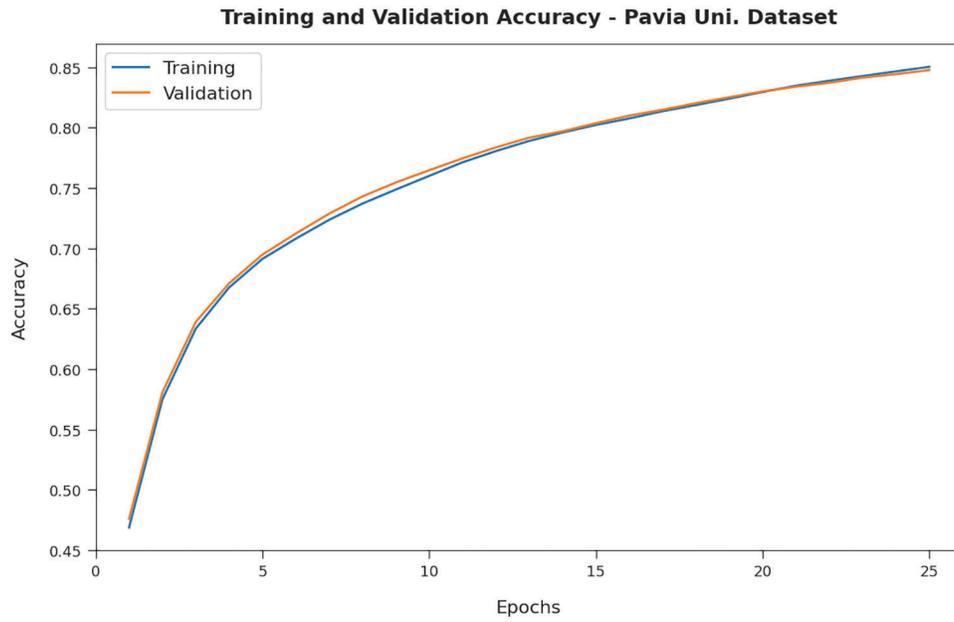


Figure 9: TA and VA of RDADL-HIC model on PUD

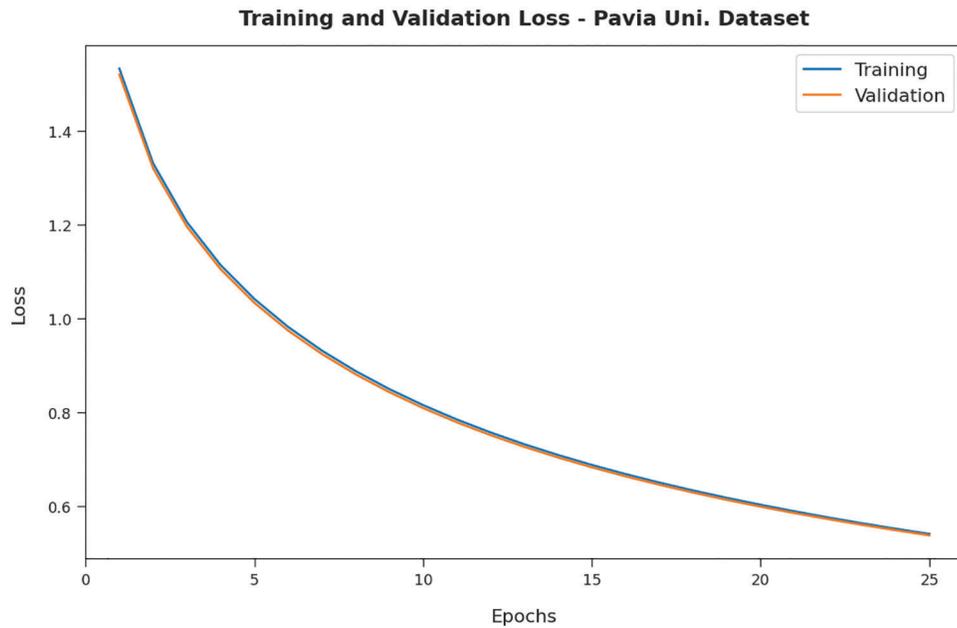


Figure 10: TL and VL of RDADL-HIC model on PUD

Fig. 11 illustrates the precision-recall study of the RDADL-HIC model on the PUD. The figure indicated that the RDADL-HIC model has accomplished maximum precision-recall values on the distinct class labels.

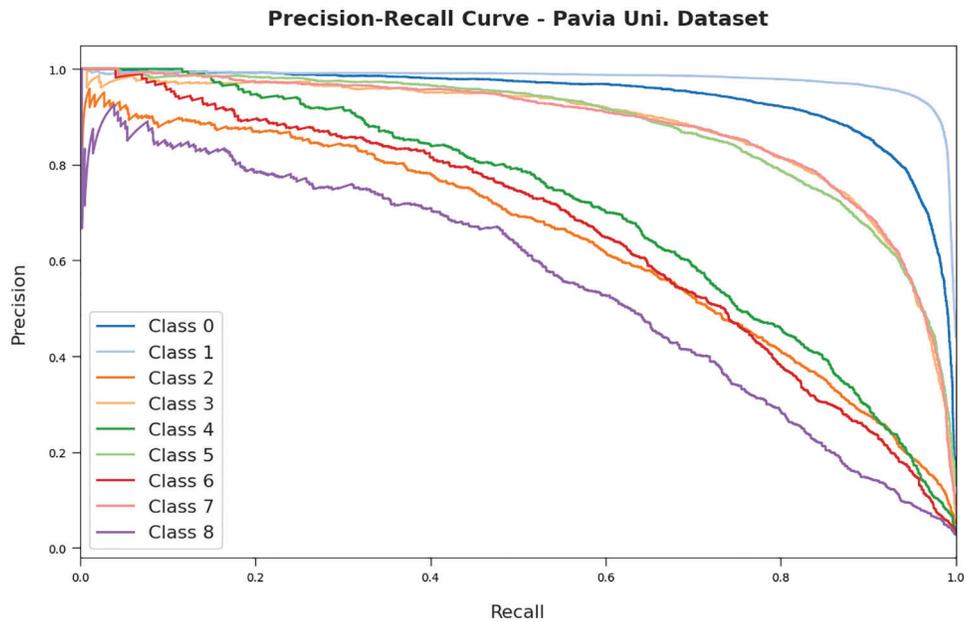


Figure 11: Precision recall curves of RDADL-HIC model on PUD

Fig. 12 exemplifies the ROC assessment of the RDADL-HIC model on the PUD. The figure indicated that the RDADL-HIC model has gained improved ROC values on the distinct class labels. For ensuring the

enhanced outcomes of the RDADL-HIC model, a detailed comparative examination with IPD and PUD datasets is made in Tab. 1 [20,21].

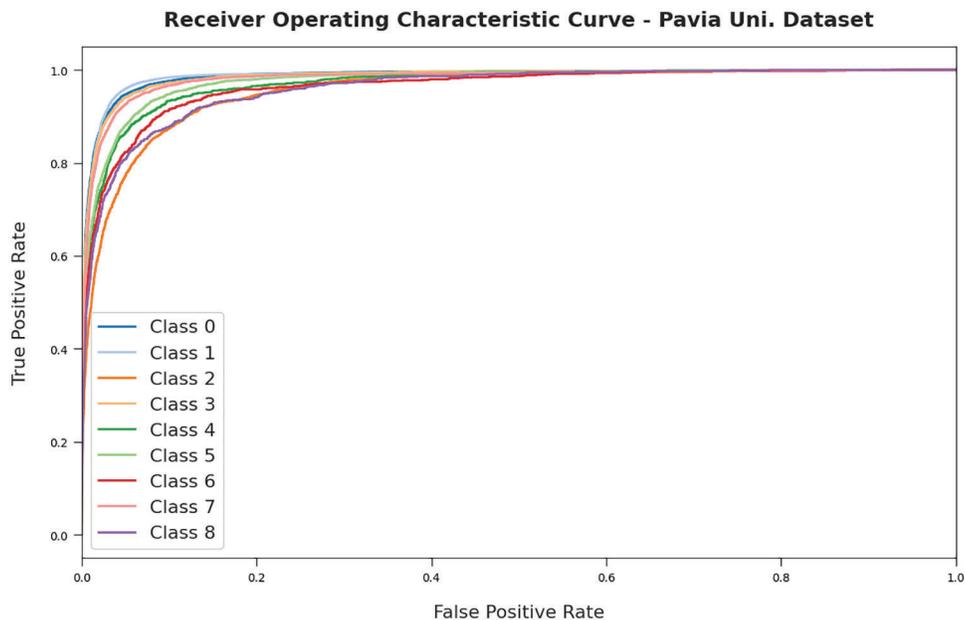


Figure 12: ROC curves of RDADL-HIC model on PUD

Table 1: HSI classification of RDADL-HIC model with existing models

Methods	Accuracy (%)		Testing time (s)	
	India pines dataset	Pavia uni. dataset	India pines dataset	Pavia uni. dataset
RDADL-HIC	96.88	95.90	1.23	1.19
CNN-HSIC	94.65	94.00	2.22	1.91
CNN model	90.63	92.02	2.11	1.77
SVM model	87.70	90.31	2.90	2.62
DNN model	87.51	90.74	3.17	3.09
CNN-LeNet	87.96	89.41	2.87	2.11
SRST algorithm	88.68	82.44	2.52	2.77

Fig. 13 demonstrates a brief accuracy examination of the RDADL-HIC model with existing models on IPD and PUD. The figure indicated that the support vector machine (SVM), deep neural network (DNN), CNN-LeNet models have shown poor outcomes with minimal values of accuracy. Though the CNN-HSIC and CNN models have attained moderate accuracy values, the presented RDADL-HIC model has accomplished maximum accuracy on both datasets. For instance, on IPD, the RDADL-HIC model has attained higher accuracy of 96.88% whereas the CNN-HSIC, CNN, SVM, DNN, CNN-LeNet, and SRST models have obtained lower accuracy of 94.65%, 90.63%, 87.70%, 87.51%, 87.96%, and 88.68% respectively. Similarly, on PUD, the RDADL-HIC model has reached increased accuracy of 95.90% whereas the CNN-HSIC, CNN, SVM, DNN, CNN-LeNet, and SRST models have gotten reduced accuracy of 94%, 92.02%, 90.31%, 90.74%, 89.41%, and 82.44% respectively.

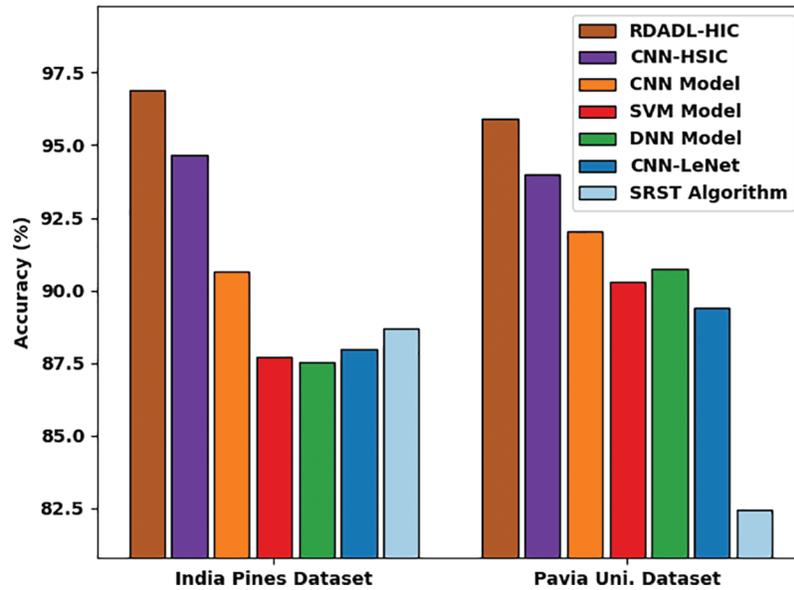


Figure 13: Comparative accuracy analysis of RDADL-HIC model with existing models

Fig. 14 exhibits a detailed testing time (TT) inspection of the RDADL-HIC model with recent approaches on IPD and PUD. The results represented that the SVM, DNN, CNN-LeNet models have exposed worse performance with maximum values of TT. Although the CNN-HSIC and CNN models have attained reasonable TT values, the presented RDADL-HIC model has gained lower values of TT on both datasets. For instance, on IPD, the RDADL-HIC model has accomplished reduced TT of 1.23 s whereas the CNN-HSIC, CNN, SVM, DNN, CNN-LeNet, and SRST models have attained increased TT of 2.22 s, 2.11 s, 2.90 s, 3.17s, 2.87 s, and 2.52 s respectively.

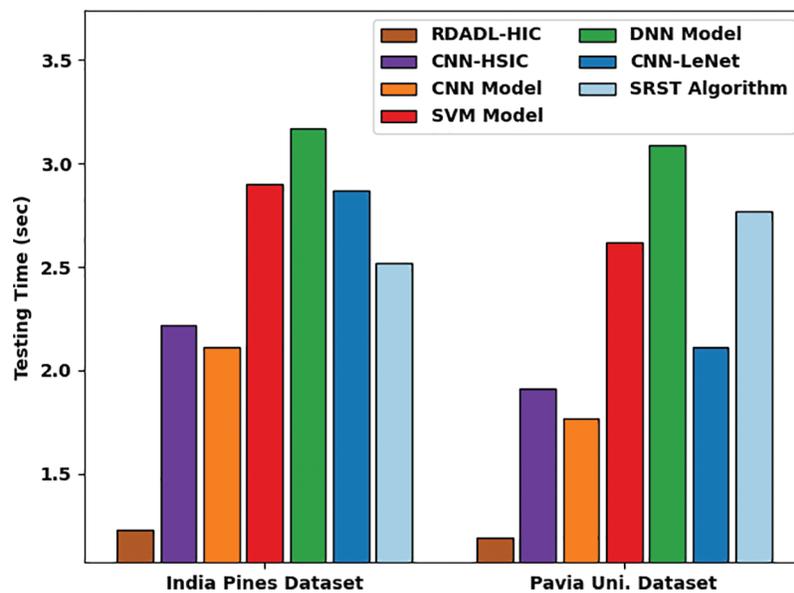


Figure 14: Comparative TT analysis of RDADL-HIC model with existing models

Likewise, on PUD, the RDADL-HIC model has obtained lower TT of 1.19s whereas the CNN-HSIC, CNN, SVM, DNN, CNN-LeNet, and SRST models have gotten reduced TT of 94%, 1.91 s, 1.77 s, 2.62 s, 3.09 s, and 2.11 s respectively. When the results and discussions are observed, it is concluded that the RDADL-HIC model has accomplished maximum performance on HS image classification.

4 Conclusion

In this study, a novel RDADL-HIC technique has been presented to effectively determine the HSI images. The RDADL-HIC technique mainly involves a NASNetLarge model with Adagrad optimizer. Followed by, RDA with GRU classifier is used for the identification and classification of HSIs. The design of Adagrad optimizer with RDA helps to automatically modify the hyperparameters of the NASNetLarge and GRU models respectively. The experimental results stated the superior results of the RDADL-HIC model and the results are inspected interms of different measures. The comparison study of the RDADL-HIC model demonstrated the enhanced performance over its recent approaches. In future, the proposed model can be employed in real time application such as agriculture, mineral exploration, etc.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi *et al.*, “Deep learning for hyperspectral image classification: An overview,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6690–6709, 2019.
- [2] M. Imani and H. Ghassemian, “An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges,” *Information Fusion*, vol. 59, no. 2, pp. 59–83, 2020.
- [3] M. Zhang, H. Luo, W. Song, H. Mei and C. Su, “Spectral-spatial offset graph convolutional networks for hyperspectral image classification,” *Remote Sensing*, vol. 13, no. 21, pp. 4342, 2021.
- [4] R. Hang, Q. Liu, D. Hong and P. Ghamisi, “Cascaded recurrent neural networks for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5384–5394, 2019.
- [5] R. Hang, Z. Li, Q. Liu, P. Ghamisi and S. S. Bhattacharyya, “Hyperspectral image classification with attention-aided CNNs,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2281–2293, 2021.
- [6] L. Mou and X. X. Zhu, “Learning to pay attention on spectral domain: A spectral attention module-based convolutional network for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 1, pp. 110–122, 2020.
- [7] X. R. Zhang, J. Zhou, W. Sun and S. K. Jha, “A lightweight CNN based on transfer learning for COVID-19 diagnosis,” *Computers, Materials & Continua*, vol. 72, no. 1, pp. 1123–1137, 2022.
- [8] W. Sun, G. Z. Dai, X. R. Zhang, X. Z. He and X. Chen, “TBE-Net: A three-branch embedding network with part-aware ability and feature complementary learning for vehicle re-identification,” *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2021.
- [9] F. Luo, L. Zhang, B. Du and L. Zhang, “Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5336–5353, 2020.
- [10] W. Song, S. Li, L. Fang and T. Lu, “Hyperspectral image classification with deep feature fusion network,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3173–3184, 2018.
- [11] R. Liu, W. Cai, G. Li, X. Ning and Y. Jiang, “Hybrid dilated convolution guided feature filtering and enhancement strategy for hyperspectral image classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

- [12] W. Cai, B. Liu, Z. Wei, M. Li and J. Kan, "TARDB-Net: Triple-attention guided residual dense and BiLSTM networks for hyperspectral image classification," *Multimedia Tools and Applications*, vol. 80, no. 7, pp. 11291–11312, 2021.
- [13] L. Mou, X. Lu, X. Li and X. X. Zhu, "Nonlocal graph convolutional networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8246–8257, 2020.
- [14] X. Cao, J. Yao, Z. Xu and D. Meng, "Hyperspectral image classification with convolutional neural network and active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 7, pp. 4604–4616, 2020.
- [15] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya *et al.*, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 3791–3808, 2020.
- [16] C. J. Stolz, F. Y. Genin, T. A. Reitter, N. E. Molau, R. P. Bevis *et al.*, Effect of SiO₂ overcoat thickness on laser damage morphology of HfO₂/SiO₂ Brewster's angle polarizers at 1064 nm. In: *Laser-Induced Damage in Optical Materials: 1996*. SPIE, the international society for optics and photonics, Boulder, CO, United States, pp. 265–272, 1996.
- [17] R. Dey and F. M. Salem, "Gate-variants of Gated Recurrent Unit (GRU) neural networks," in *2017 IEEE 60th Int. Midwest Symp. on Circuits and Systems (MWSCAS)*, Boston, MA, pp. 1597–1600, 2017.
- [18] A. M. F. Fard, M. H. Keshteli and R. T. Moghaddam, "Red deer algorithm (RDA): A new nature-inspired meta-heuristic," *Soft Computing*, vol. 24, no. 19, pp. 14637–14665, 2020.
- [19] J. Banumathi, A. Muthumari, S. Dhanasekaran, S. Rajasekaran, I. V. Pustokhina *et al.*, "An intelligent deep learning based xception model for hyperspectral image analysis and classification," *Computers, Materials & Continua*, vol. 67, no. 2, pp. 2393–2407, 2021.
- [20] W. Hu, Y. Huang, L. Wei, F. Zhang and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *Journal of Sensors*, vol. 2015, no. 2, pp. 1–12, 2015.
- [21] Y. Wu, G. Mu, C. Qin, Q. Miao, W. Ma *et al.*, "Semi-supervised hyperspectral image classification via spatial-regulated self-training," *Remote Sensing*, vol. 12, no. 1, pp. 159, 2020.