INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research

International Journal of Computational Science and Engineering

ISSN online: 1742-7193 - ISSN print: 1742-7185
https://www.inderscience.com/ijcse

Black widow optimisation with deep learning-based feature fusion model for remote sensing image analysis

Vaishnavee Vijay Rathod, Dipti P. Rana, Rupa G. Mehta

DOI: 10.1504/IJCSE.2024.10063761

Article History:

Received:	14 June 2023
Last revised:	18 September 2023
Accepted:	22 January 2024
Published online:	21 December 2024

Black widow optimisation with deep learning-based feature fusion model for remote sensing image analysis

Vaishnavee Vijay Rathod*, Dipti P. Rana and Rupa G. Mehta

Department of Computer Science Engineering,
Sardar Vallabhbhai National Institute of Technology,
Surat, Gujarat, India
Email: d20co004@coed.svnit.ac.in
Email: dpr@coed.svnit.ac.in
Email: rgm@coed.svnit.ac.in
*Corresponding author

Abstract: Recently, achieving accurate remote sensing images (RSI) classification has been a primary goal in deep learning, given its extensive applications, including urban planning and disaster management. The performance of existing convolutional neural networks (CNN)-based strategies is primarily influenced by their parameter settings, necessitating automated hyperparameter tuning through metaheuristic methods. The proposed BWODLF-RSI technique integrates black widow optimisation with a deep learning feature fusion model for enhanced RSI analysis. The preliminary processing step is to enhance RSI quality using noise reduction through a Gaussian filter (GF), enhancing contrast with the help of contrast limited adaptive histogram equalisation (CLAHE), and data augmentation to prevent overfitting. It is followed by employing Inception v3 and DenseNet201 to extract and fuse potent features. A critical aspect of this strategy is using black widow optimisation to fine-tune the kernel extreme learning machine (KELM) model, attaining a notable RSI classification accuracy of 94.05%. When tested on UCM and AID datasets, the BWODLF-RSI approach demonstrated superior feature selection and RSI analysis performance.

Keywords: remote sensing; image classification; deep learning; pre-processing; feature fusion.

Reference to this paper should be made as follows: Rathod, V.V., Rana, D.P. and Mehta, R.G. (2025) 'Black widow optimisation with deep learning-based feature fusion model for remote sensing image analysis', *Int. J. Computational Science and Engineering*, Vol. 28, No. 1, pp.56–70.

Biographical notes: Vaishnavee Vijay Rathod received her BE in Electronics and Telecommunication from Thakur College of Engineering and Technology, Mumbai, in 2018 and MTech Electronics and Telecommunication from GHRIET, Nagpur, in 2020. She is pursuing her PhD at the Department of Computer Science and Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat, India. Her areas of interest are image processing, artificial intelligence, machine learning, and computer vision.

Dipti P. Rana works in the Department of Computer Science and Engineering as an Associate Professor at Sardar Vallabhbhai National Institute of Technology (SVNIT) Surat, Gujarat, India. She completed her PhD in Computer Science from the Department of Computer Science and Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat, India. Her areas of interest are data mining, database systems, image processing, artificial intelligence and machine learning.

Rupa G. Mehta works at Sardar Vallabhbhai National Institute of Technology (SVNIT) Surat as a Professor in the Department of Computer Science and Engineering. She completed her BTech, MTech and PhD in Computer Science from the Department of Computer Science and Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat, India. Her areas of interest are big data analytics, data science, artificial intelligence, and machine learning.

1 Introduction

In recent times, the field of remote sensing has evolved significantly, offering more detailed and sophisticated monitoring of the Earth through land-cover and land-use classifications. Advances in spatial resolution have not only improved the data quality and area covered by remote sensing images (RSIs) but has also introduced challenges in managing the enhanced volume of data. Emery and Camps

(2017) report that observing the Earth from geostationary satellites and low Earth orbit has continuously improved. This increase needs a considerable change in how RSIs are managed and used. Zhou et al. (2018) declared that enhanced spatial resolution makes it feasible to develop better methods and provides novel opportunities to advance RSI understanding and analysis, thereby permitting users to investigate the ground surface. However, data enhancement accessibility gave rise to significant problems regarding appropriately handling RSI imagery.

Cheng et al. (2017) determined scene classifiers to classify RSIs into separate sets of useful land-use and land-cover categories. Scene classifier is a primary remote-sensing task and essential for real-time remote-sensing applications, namely land management (Zhou et al., 2013), urban planning, and to describe wildfires, amongst other applications. This extensive usage of RSI classification directed several authors to examine processes for accelerating image retrieval and classifying remote-sensing information. Traditional scene classification technique uses lower-level visual features to characterise the image of interest. This lower-level feature is both global and local. The global feature is extracted in the whole RSI, namely the shape feature (Chen et al., 2016), colour (spectral) feature, and texture feature (Daldegan et al., 2019; Sebai et al., 2015). Recent developments in DL algorithms, especially deep convolutional neural networks (DCNN) architecture, have enhanced advanced speech recognition, visual object detection, and recognition, etc. (Sun et al., 2022; Yao et al., 2020).

This article introduces a new deep learning-based feature fusion method for RSI classification, named BWODLF-RSI system. The projected BWODLF-RSI technique involves pre-processing for improving the RSI quality in three diverse ways such as Gaussian filter (GF)-based de-noising, contrast limited adaptive histogram equalisation (CLAHE) for improving contrast and augmenting the data. Moreover, a set of two DL-based approaches Inceptionv3 and densely connected networks (DenseNet201) are employed. Furthermore, two feature vectors are fused together to boost the general performance of the proposed technique. At last, black widow optimisation (BWO) with kernel extreme learning machine (KELM) approach is applied for RSI classification. To scrutinise the enhanced result of the BWODLF-RSI approach, a detailed simulation examination was implemented on benchmark databases. It is useful in the analysis and interpretation of the RSIs. Deep learning (DL) plays a significant role in RSI classification. The various land use applications like agriculture, urban planning, traffic surveillance, etc. are associated with RSI scene classifications. The monitoring and management strategies can be efficiently built in consideration with present risk factors and this holds one of the major roles of apt RSI classification. Applying this system facilitates more accurate and efficient image analysis, a tool integral to various industries including urban planning, agriculture, and disaster management. Through detailed RSI analysis, stakeholders can undertake more informed planning and management, whether in developing sustainable cities, optimising agricultural yield or effectively responding to natural disasters. This progressive approach to RSI classification stands to play a pivotal role in utilising remote sensing for contemporary needs, underpinned by deep learning technologies.

2 Related work

Boualleg et al. (2019) presented an HR-RSSC technique for deep forest (DF) and feature extraction (FE) for classifiers. The extraction of deep features is carried out in the last convolution layer so that FC layer usage can be avoided, which requires various parameters to tune. Furthermore, the authors trained a DF method, i.e., depending on ensemble learning, which could obtain good efficiency when compared to a single classifier and easier to train some parameters. Zhang et al. (2019) developed a powerful RSI scene classifier method called CNN-Caps Net to take advantage of the advantages of these two methods: Caps Net and CNN. However, this concerns scalability and applicability as the evaluation is done on specific and limited datasets. Han et al. (2018) designed a semi-supervised generative framework (SSGF) that joins the discriminative evaluation approach, DL feature, and self-label approach for concluding scene classification tasks and annotating datasets. According to this, an extended approach (SSGA-E) is developed and estimated by exclusive experiment. Yang et al. (2018) proposed a classification technique DCNN_MSFF based on multi-scale features fusion (MSFF) and DCNN. Next, it can be inputted to the DCNN for extraction features, and distinct scale features of fully connected and convolutional layers are averagely pooled or encoded. At last, the processed feature is combined, and the multi-kernel SVM (MKSVM) is utilised for classifying the scene.

Li et al. (2020) proposed a key region or place capture methodology named key filter bank (KFB); also, KFB could retain global data simultaneously. This approach could integrate with distinct CNN approaches for better efficiency of the high-resolution image scene classifier. Furthermore, to facilitate the real-time task, an end-to-end method named KFBNet, in which KFB is integrated with DenseNet-121, is presented to compare the efficiency with current methodologies. Yang et al. (2022) propose a deep semantic feature extraction method for improving the segmentation accuracy using the feature fusion method to map better features. Feng et al. (2023) combine SVM and ensemble learning. Several models for classification are chosen for the first layer, and in the final layer, the SVM is then applied. Thus, a good ensemble learning model and the generalisation capability were attained. Niu et al. (2020) suggest an ideal urban land classification technique for remote sensing images. A genetic algorithm with K-means mutation operator for RSI improves the classification accuracy.

Venkatesan and Prabu (2022) present a technique for DL-based feature extraction of hyperspectral data. Noise is reduced at the pre-processing stage utilising median filtering and an anisotropic diffusion method. To decrease the vast amount of data, a dimensionality reduction technique is frequently employed. The hyperspectral cube picture dimension is reduced using a discriminative local metric learning method, keeping important parts for future processing. A recurrent neural network algorithm is the last step. Choudhary and Sharma (2023) present a CNN-based categorisation created to learn historical context and categorise the landscape. The study also aimed to improve camouflage efficacy by developing the texture to match a particular terrain's main features. The work proposes a technique for classifying battlefields and creating camouflage textures using a CNN-based model. Guo et al. (2021) introduce MSANet, a new attention-based network for proficient image data segmentation, aiming to preserve contextual and spatial details often lost in current deep learning models. Yang et al. (2021) introduce the diffusion convolutional network (DCNet), a hybrid approach merging CNN and GCNN to enhance semantic image segmentation, a pivotal aspect of scene understanding. Using diffusion convolution as a graph convolutional layer, the DCNet retains spatial attributes while assimilating structural and contextual details.

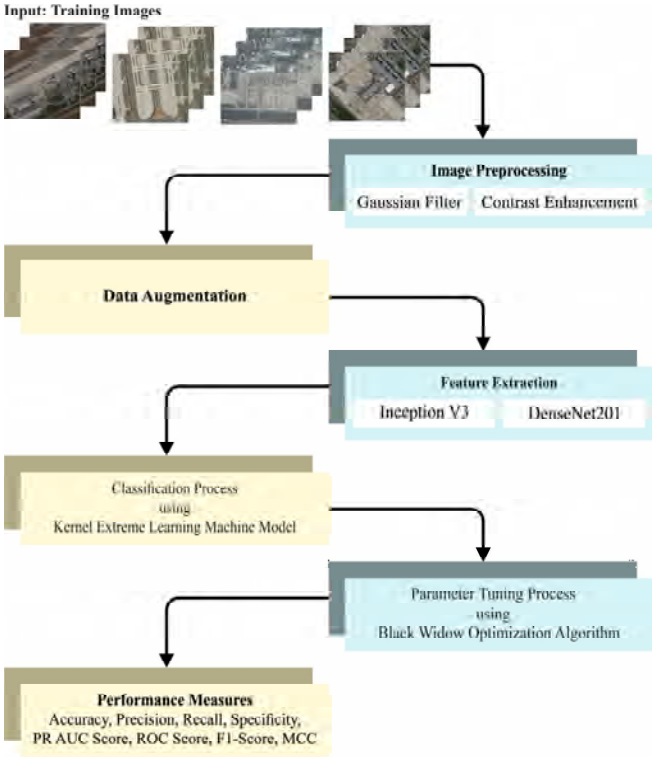
3 The proposed model

Existing methods struggled to effectively manage the complexities and unique characteristics of RSIs, including irregular patterns and varying spatial resolutions. This has resulted in an inability to efficiently capture the intricate details in RSIs and led to underwhelming performance. This article introduces the BWODLF-RSI technique, a novel approach for RSI classification that involves a systematic process of pre-processing, feature extraction, and feature fusion. The pre-processing step employs noise reduction, contrast enhancement, and data augmentation to mitigate challenges posed by noise and artefacts in RSIs. Leveraging the strengths of both Inceptionv3 and DenseNet201 models, the feature extraction process aims to extract relevant features from RSIs more effectively and efficiently, enhancing the analysis and interpretation of these images in various applications. Integrating the black widow optimisation (BWO) algorithm with the kernel extreme learning machine (KELM) model further elevates classification performance, as evidenced through comparative analyses showcased in the study. This innovative approach, outlined in Figure 1, promises to significantly improve the RSI analysis process, paving the way for advanced utilisation in several applications.

3.1 Image pre-processing

In this study, RSI pre-processing is carried out through three techniques to enhance image quality: GF-based noise elimination, CLAHE-based contrast enhancement, and data augmentation. The first employs the cv.GaussianBlur() function with a Gaussian kernel instead of box filters to efficiently remove Gaussian noise, using specified standard deviations to control the blurring process. The second technique involves adaptive histogram equalisation, where the image is divided into 'tiles' or small blocks. The process adjusts the histogram in each tile to improve the image's contrast, confining it to a specific range to mitigate noise amplification. It leverages contrast limiting to prevent noise amplification, with excess pixels distributed uniformly before equalisation. Post equalisation, artefacts at tile borders are reduced using bilinear interpolation. This approach ensures a well-balanced image with pixels representing all regions efficiently, enhancing the image's overall quality.

Figure 1 Workflow of proposed model (see online version for colours)



CLAHE is an adaptive histogram equalisation with contrast limitation by utilising the clip limits and no. of tiles parameter (Kuran and Kuran, 2021). The CLAHE divides the image into $M \times N$ local tiles. In order to all tiles, histogram was calculated separately. To compute the histogram, it can be the primary requirement for calculating average amount of pixel per area utilising equation (1).

$$N_A = (N_X \times N_Y)/N_G \tag{1}$$

In equation (1), N_A refers to the average no. of pixels, N_X signifies the no. of pixels from X dimensional and N_Y stands for the amount of pixel from the Y dimensional and N_G denotes the amount of grey levels. Afterward, it describes clipper limits as in equation (2) for equalising histograms by clipping:

$$N_{CL} = N_A \times N_{NCL} \tag{2}$$

In equation (2), N_{CL} signifies the clipper limits and N_{NCL} indicates the normalising clipper limits amongst 0 and 1. Then, to all the tiles, the clip limits are executed to the histogram utilised in equation (3).

$$H_i = \begin{cases} N_{CL} & if\ N_i \geq N_{CL} \\ N_i & else \end{cases} \quad i = 1, 2, \ldots, 1-1 \tag{3}$$

In equation (3), H_i demonstrated the histogram i^{th} height tiles, N_i defines the i^{th} tile histogram and L implies the amount of grey levels. The entire clipped pixel is calculated utilising equation (4).

$$N_c = (N_X \times N_Y) - \sum_{i=0}^{L-1} H_i \tag{4}$$

In equation (4), N_C stands for the clipped pixels. Afterward the computing N_C, it is also redistributing the pixel which is clipped both evenly/unevenly. For computing the amount of pixel that redistribute, it is utilised in equation (5).

$$N_R = N_C / L \tag{5}$$

In equation (5), N_R implies the amount of pixel that redistribute. Then, normalisation of clipped histogram was carried out utilising equation (6).

$$H_i = \begin{cases} N_{CL} & if\ N_i + N_R \geq N_{CL} \\ N_i + N_R & else \end{cases} \quad i = 1, 2, \ldots, l-1 \tag{6}$$

In equations (4) and (5), undistributed pixels were counted. Until each pixel is redistributed in equation (6) was repeated. Eventually, the cumulative histogram of contextual region is written in equation (7).

$$C_i = \frac{1}{(N_X \times N_Y)} \sum_{j=0}^{i} H_j \tag{7}$$

Afterward, each calculation is done, the histogram of contextual area was equivalent with uniform, Rayleigh, or exponential probability distribution that offers a prefix brightness and quality of visual mentioning that it contains pixel $P(x, y)$ with values of s and four centre points affect the neighbour tile that is named as T_1, T_2, T_3 and T_4. The weight sum has been calculated on these 4 contextual areas. To the resultant image, tiles were combined and artefact deletion amongst the independent tile is complete utilising the bilinear interruption, a new s value that is referred to as s, was attained utilised in equation (8).

$$\begin{aligned} s' = {} & (1-y)\big((1-x) \times T_1(s) + x \times T_2(s)\big) \\ & + y\big((1-x) \times T_3(s) + x \times T_4(s)\big) \end{aligned} \tag{8}$$

After these pre-processing steps like Gaussian filtering and CLAHE technique, de-noised and better contrast image was achieved which was further used as training dataset. The augmentation methods like rotation and flipping to expand sample size for better model training. This reduces the over-fitting error and makes the model learn different situations that may occur during the real-world applications.

3.2 Fusion-based feature extraction

Following image processing, the Inceptionv3 and DenseNet201 models are employed to obtain a valuable set of feature vectors. The Inceptionv3 model is favoured for its efficient use of computing resources, providing high performance without substantially increasing the computational load, compared to its predecessors, Inception V1 and V2. It is also cost-effective and uses auxiliary classifiers for regularisation. Meanwhile, DenseNet201 is known for mitigating gradient loss problem, encouraging robust features, and reducing parameters significantly, which facilitates a more effective image representation. It combines these advantages to enhance the depiction of images while maintaining computational efficiency.

3.2.1 Inception v3 model

Training a CNN from scratch is generally challenging due to the risk of overfitting. Utilising pre-trained CNNs through transfer learning helps to mitigate this issue, allowing for fine-tuning on specific datasets. In this study, several well-known CNN frameworks including DenseNet and ResNet were tested to identify the most efficient models. The Inceptionv3 architecture (Wang et al., 2019) emerged as optimal, benefiting from initialisation by ImageNet weight and fine-tuning on the training set for effective feature map extraction. This version of inception CNN introduces improvements such as factorised convolutional to decrease the number of parameters without compromising network effectiveness, and employs label smoothing for regularisation, enhancing overall performance.

3.2.2 DenseNet model

A CNN utilises classification and feature extraction, often leveraging pretrained networks like DenseNet available in the Keras API. Unlike traditional CNN models where each convolution layer is connected only to the preceding one, DenseNet layers take the feature maps of all preceding layers as inputs, enhancing feed-forward functioning. In this study, the DenseNet201 architecture is employed for its advantages such as preventing gradient vanishing issues, reducing parameter quantities, and facilitating better feature propagation. This architecture utilises various layers including convolutional, transition, and classification layers, with the dense block comprising different sets of convolution layers, promoting feature reuse and offering a more parametrically efficient and trainable network. A convolution layer is a basic component of a NN. The fixed size was utilised for extracting the complicated features of the provided information (Wang and Zhang, 2020). Other than that, there is one convolutional layer, three transition layers, and one classification layer. The dense block consists of 6, 12, 32, and 32 convolution layers. Because of

the capacity to feature reused by consecutive layers, the DenseNet201 employs a condensed network, enabling parametrically efficient and easy-to-train methods.

3.2.3 Feature fusion process

At this stage, the fusion of two feature vectors (namely Inceptionv3 and DenseNet201) can be employed for RSI classification. Fusing the features plays a vital role to classify images that intend to combine many feature vectors. In this study, the fusion of features takes place using the entropy function and it is defined as follows:

$$f_{I_{1\times m}} = \{I_{1\times 1}, I_{1\times 2}, \ldots, I_{1\times n}\} \tag{9}$$

$$f_{D_{1\times p}} = \{D_{1\times 1}, D_{1\times 2}, D_{1\times 3}, \ldots, D_{1\times n}\} \tag{10}$$

Moreover, extracted features are fused in a single vector.

$$Fused(\ features\ vector)_{1\times q} = \sum_{i=1}^{2} \{\ fI_{1\times m},\ fD_{1\times p}\} \tag{11}$$

where f is a fused vector.

3.3 Image classification using KELM model

At this stage, the fused feature vectors are then given to the KELM model to classify the RSIs. KELM is a revised version of ELM, and ELM is an FFNN that comprises hidden, input, and output layers (Chen et al., 2021).

Where $\chi_k = [x_{k1}, x_{k2}, \ldots, x_{km}]^T \in R^m$ stands for input vector, $c_k = [c_{k1}, c_{k2}, \ldots, c_{kn}]^T \in R^m$ shows the expected output vector, $y_k = [y_{k1}, y_{k2}, \ldots, y_{kn}]^T \in R^n$ is the actual resultant vector. The connection weighted are represented as $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_q]^T$ and $\beta = [\beta_1, \beta_2, \ldots, \beta_q]^T$ in which $\alpha_i = [\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{in}]^T$ and $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{in}]^T$. The threshold value was signified as $[b_1, b_2, \ldots, b_q]^T$. The kernel function parameter X and regularisation coefficient C are enriched by SSA. Certain steps are given in the following:

Step 1 Population initialisation. The size of population N and the maximal iterations $iter_{\max}$ are intended. In all the iterations, the P_N sparrow with better activation function for the hidden layer neuron can be represented as g:

$$H\beta = C \tag{12}$$

where $C = [c_1, c_2, \ldots, c_N]^T$,

$$H = \left[h^T(x_1), h^T(x_2), \ldots, h^T(x_N)\right]^T = \begin{bmatrix} g\left(a_1^T x_1 + b_1\right) & g\left(a_2^T x_1 + b_1\right) & \cdots & g\left(a_q^T x_1 + b_q\right) \\ g\left(a_1^T x_2 + b_1\right) & g\left(a_2^T x_2 + b_1\right) & \cdots & g\left(a_q^T x_2 + b_q\right) \\ \vdots & \vdots & \ddots & \vdots \\ g\left(a_1^T x_N + b_1\right) & g\left(a_2^T x_N + b_1\right) & \cdots & g\left(a_q^T x_N + b_q\right) \end{bmatrix} \tag{13}$$

$$H^+\beta = C \tag{14}$$

$$\beta = Pf^+C_t \tag{15}$$

whereas $H^+ = H^T\left(HH^T + \frac{/}{c}\right)^{-1}$ shows the inverse matrix of H, and C indicates the regularised coefficient. KELM can be attained by presenting the kernel function as to ELM, and the kernel matrix was formulated by

$$\Omega_{ELM} = HH^T = h(x_i)h(x_j) = K(x_i, x_j) \tag{16}$$

Thus, the outcome y of KELM can be formulated by equation (17):

$$y = Kelm(x) = \begin{bmatrix} K(x, x_1) \\ K(x, x_2) \\ \vdots \\ K(x, x_N) \end{bmatrix}\left(\Omega_{ELM} + \frac{I}{C}\right)^{-1} C \tag{17}$$

As a kernel function, Gaussian kernel is utilised, viz. $(x_i, x_j) = e^{\frac{-\|x-x\|^2}{2S^2}}$, where S denotes the kernel parameter.

3.4 Parameter tuning using BWO algorithm

In this study, the BWO technique was exploited to optimise parameter adjustment of the KELM model. Hayyolalam and Kazem (2020) proposed the BWO approach based on mathematical modelling as the existence of black widow. The model simulates both microscopic and macroscopic principles inherent in spider population dynamics to find optimised solutions to specific problems. The BWO technique operates through four stages: initialisation, procreation, cannibalism, and mutation, with crucial parameters adjusted to optimise performance. This approach offers faster solutions and superior convergence rates for RSI analysis compared to other algorithms.

The pseudocode of BWO has been demonstrated below:

Algorithm 1 The BWO algorithm

Input: count of population (N), procreating rate (Pp), maximal iteration (T), mutation rate (Pm), cannibalism rate (CR)

Output: best solution of the objective function

Begin

Initialisation of the spider population randomly based on equation (18);

While ($t < T$)

By using Pp, evaluate the count of reproduction '*nr*';

Choose the *nr* parents from the population;

For i =1 to *nr* **do**

Randomly choose two solution solutions as parent from *nr* parent;

Produce children based on equation (19);

Destroy fathers;

By using CR, some children are destroyed;

End for

Retain the remaining children and mother in the novel matrix as novel generation;

Evaluate the count of mutations of children *nm* based Pm;

For i = 1 to nm

Choose the solution for the continued children;

Randomly mutate one component of solutions and produce a novel solution;

End for

Return the better solution;

End while

Outcome the better solution;

End

4 Results and discussion

The BWODLF-RSI approach's test results on a UCM database RSI are detailed in Figure 4, highlighting the substantial enhancement in image quality facilitated by pre-processing. Figure 4(a) exhibits the original test RSI, serving as a benchmark for illustrating the initial image constraints. Meanwhile, Figure 4(b) displays the image post-pre-processing, showcasing the notable enhancements attained through this initial step.

4.1 Dataset used

The performance validation of the BWODLF-RSI approach occurs utilising the UCM and AID dataset (https://captain-whu.github.io/AID/). The proposed BWODLF-RSI technique is simulated using Python 3.8.5 tool with additional packages (tensorflow-gpu==2.6.0, pandas, scikit-learn, seaborn, matplotlib, OpenCV-python, OpenCV-contrib-python, pillow, prettytable, tqdm, landscapes, pyqt5, cmapy, and numpy). The BWODLF-RSI technique's performance was validated using the UCM and AID datasets. The UCM dataset comprises 21 class labels, each with 100 instances of 256 × 256 pixel images, while the AID dataset image size is 600 × 600 pixels, it has 30 class labels with a total of 10,000 images, showcasing various aerial imagery scenes, including both urban and rural settings. Utilising both datasets, which offer a rich diversity of image sizes and categories, not only fosters enhanced performance and generalisation ability for the BWODLF-RSI model but also aids in preventing overfitting. Moreover, the combined datasets ensure a robust and precise evaluation of the model by facilitating a detailed analysis of spatial and spectral information present in the remote sensing datasets.

4.2 Qualitative result analysis

The sample visualisation result analysis of the BWODLF-RSI approach on the test RSI from UCM and AID database is shown in Figure 4. The results demonstrated that the qualitative analysis of images is considerably enhanced by the pre-processing step. Figure 4(a) presents the original test RSI, showcasing the initial state of the image before any processing or enhancement techniques are applied. This raw representation serves as a baseline for comparison and highlights the potential imperfections or limitations in the original RSI. To address these limitations and improve the image quality, a pre-processing step is implemented. The outcome of this pre-processing procedure is displayed in Figure 4(b). The sub-figure reveals the transformed or enhanced version of the test RSI, illustrating the considerable improvement achieved through the pre-processing step of BWODLF-RSI.

Figure 2 UCM sample images (see online version for colours)

Figure 3 AID sample images (see online version for colours)

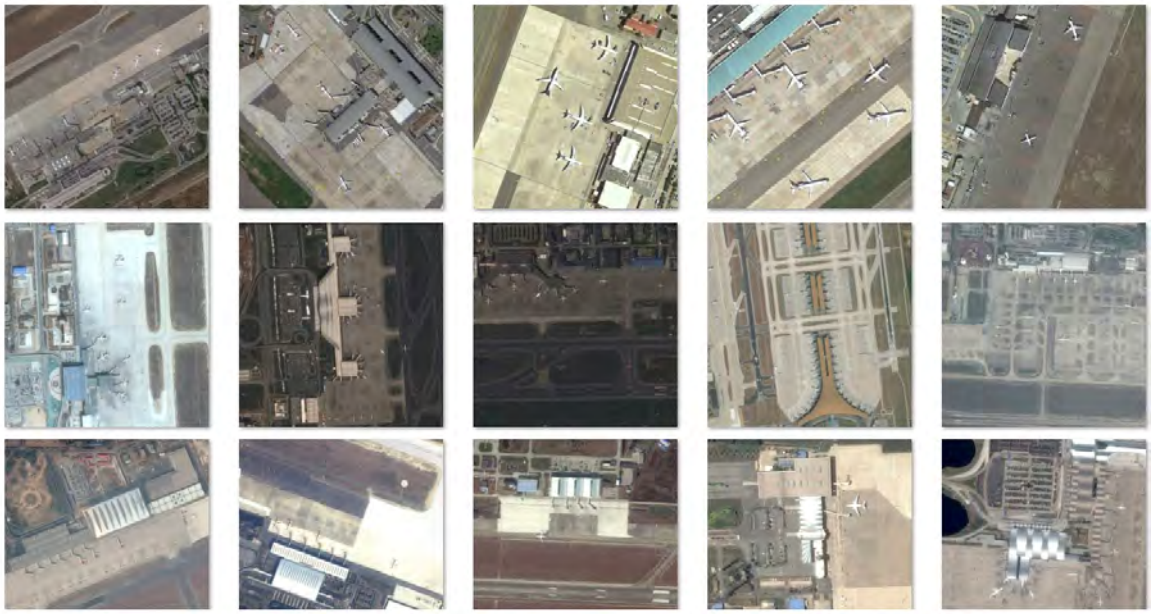
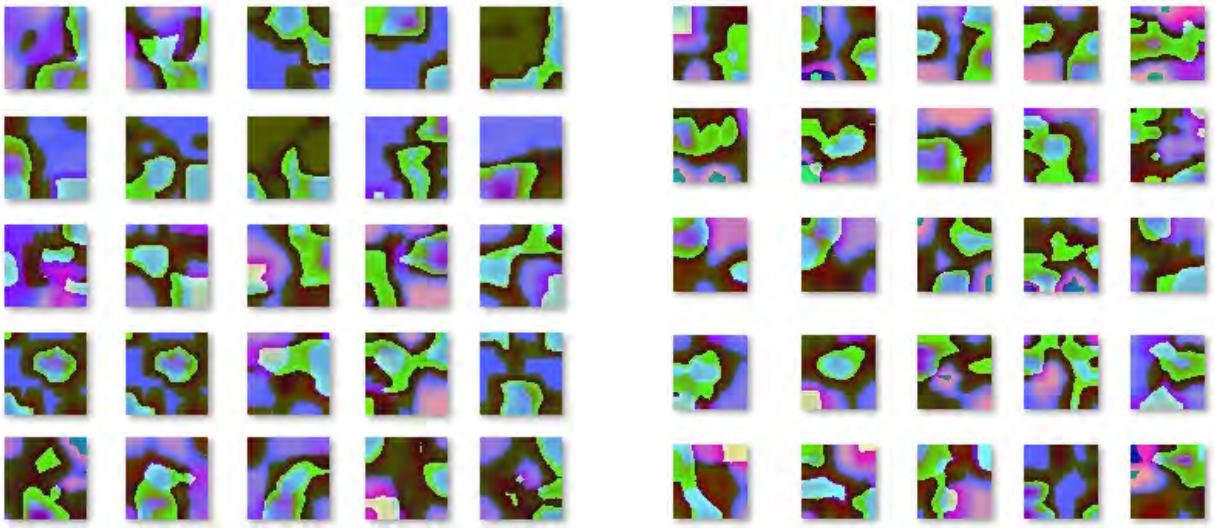


Figure 4 Sample visualisation results on UCM and AID dataset, (a) original image (b) pre-processed image (see online version for colours)

(a) (b)

Figure 5 Sample fused features (a) UCM and (b) AID dataset (see online version for colours)



(a) (b)

Figure 6 Sample visualisation results, (a) original image (b) pre-processed image (c) fused features (d) parameters with and without BWO (see online version for colours)

(a) (b) (c)

Metrics values

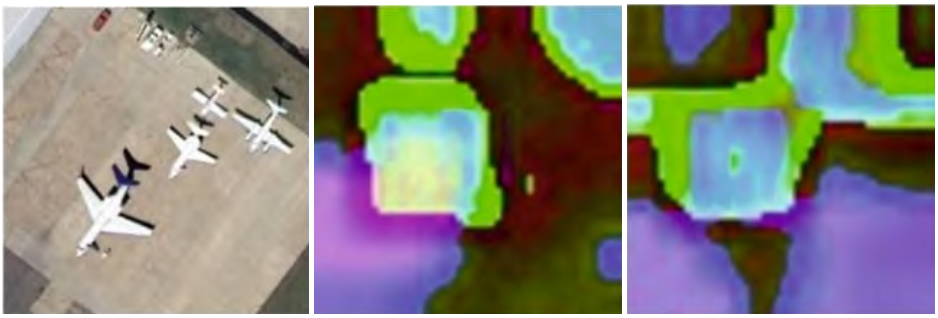
Metrics	Values
Accuracy	0.8855
Precision	0.8871
Recall	0.8855
Specificity	0.9943
PR AUC Score	0.9022
ROC AUC Score	0.9877
F1-Score	0.8846
Mathews Correlation Coefficient	0.8798

Metrics values with BWO

Metrics	Values
Accuracy	0.9464
Precision	0.9469
Recall	0.9464
Specificity	0.9973
PR AUC Score	0.9593
ROC AUC Score	0.9956
F1-Score	0.9461
Mathews Correlation Coefficient	0.9437

(d)

Figure 7 Comparison of feature maps, (a) original image (b) AlexNet (c) proposed model (see online version for colours)



(a) (b) (c)

Figure 5 showcases the fused feature maps produced by the BWODLF-RSI method on the test images from the UCM and AID datasets, as seen in Figures 5(a) and 5(b), respectively. This visualisation proves the method's effectiveness in synthesising and integrating pertinent features, offering a holistic, less redundant representation of the scene, thereby enhancing the precision of analysis efforts. It underlines the technique's ability to enhance discriminative features in the RSIs, illustrating its potential in facilitating more accurate analyses by reducing noise and redundancy.

Figure 5 displays various outputs from the BWODLF-RSI methodology applied to the UCM and AID datasets. Figures 6(a) and 6(b) showcase the original and pre-processed test RSI, respectively, while Figure 6(c) presents the successfully fused feature maps achieved post pre-processing. Figure 6(d) outlines various performance metrics, including accuracy and F1 score, underlining the improved results garnered using the BWODLF-RSI technique both with and without BWO optimisation.

Figure 7 displays a comparative analysis between the proposed BWODLF-RSI model and the traditional AlexNet model. The input image is shown in Figure 7(a). The feature map from AlexNet is in Figure 7(b), while the fused features from BWODLF-RSI are in Figure 7(c). A visual comparison suggests that BWODLF-RSI outperforms AlexNet in RSI analysis, presenting a more comprehensive and discriminative representation of relevant information.

4.3 Result analysis of proposed model on UCM dataset

This section examines BWODLF-RSI model on UCM dataset in three aspects of training/testing (TR/TS) data:

1 entire UCM dataset

2 70:30 of UCM dataset

3 80:20 of UCM dataset.

4.3.1 Result analysis of proposed model on entire UCM dataset

Figure 8 showcases the results of the BWO-KELM and KELM classifier models applied on the entire UCM database using the BWODLF-RSI method. The BWO-KELM model surpassed the KELM in classification performance, exhibiting exceptional precision-recall metrics and achieving a peak ROC of 0.9994.

Table 1 summarises classification results for the KELM model and BWODLF-RSI on the entire UCM database. BWODLF-RSI outperformed KELM with higher accuracy (94.38%), precision (94.40%), recall (94.38%), specificity (99.72%), PR_AUC score (95.61%), ROC (99.48%), F1_score (94.35%), and MCC (94.09%).

Table 1 Result analysis of KELM and BWODLF-RSI approaches under entire UCM dataset

Metrics	*KELM*	*BWODLF-RSI*
Accuracy	88.21	94.38
Precision	88.30	94.40
Recall	88.21	94.38
Specificity	99.41	99.72
PR AUC score	89.92	95.61
ROC score	98.71	99.48
F1-score	88.10	94.35
MCC	87.60	94.09

Figure 8 Results analysis on entire UCM dataset for KELM for BWODLF-RSI method, (a) confusion matrix (b) PR-Curve (c) ROC, and for BWO-KELM (d) confusion matrix (e) PR-curve (f) ROC (see online version for colours)

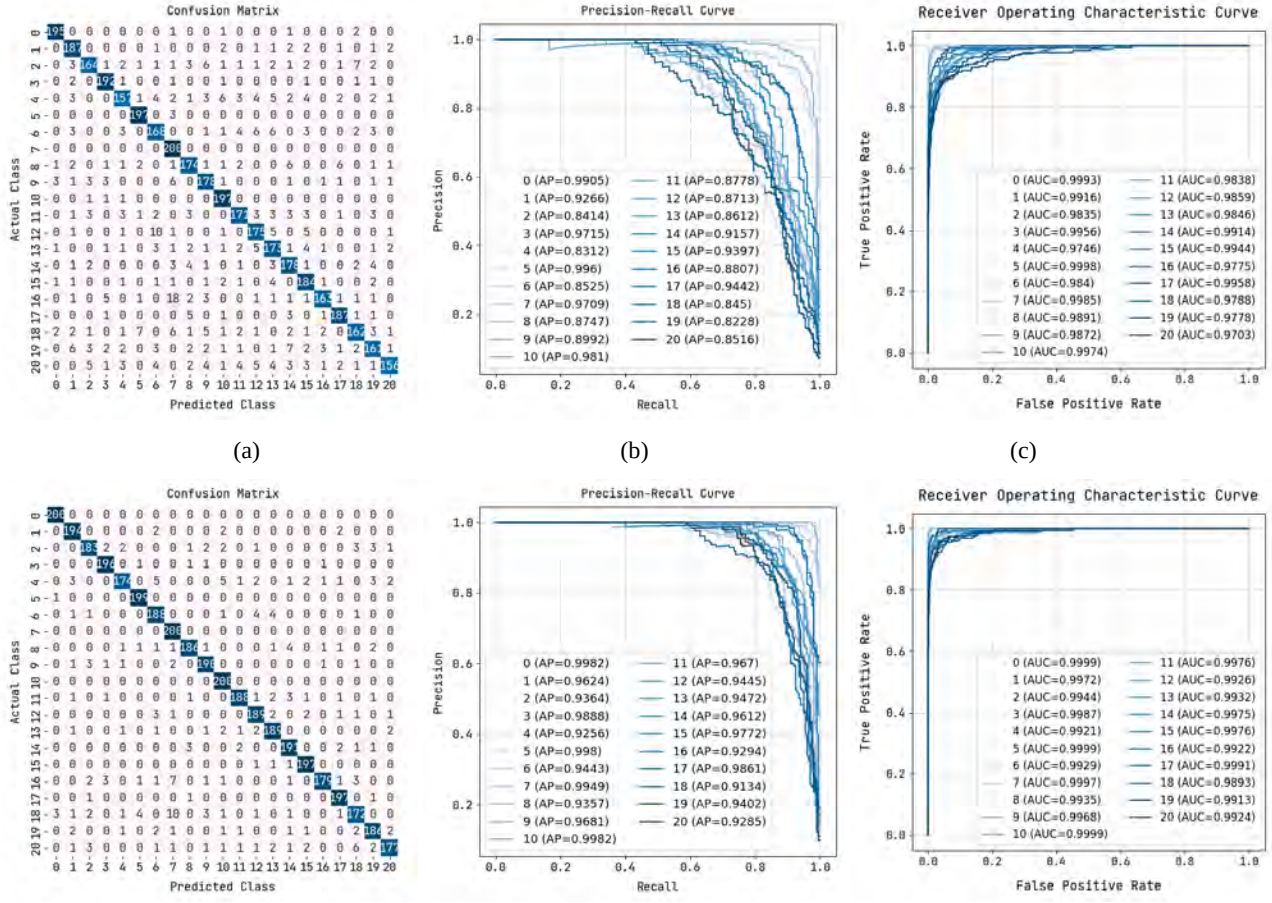


Figure 9 Results analysis on 70:30 of UCM dataset for KELM for BWODLF-RSI method, (a) confusion matrix (b) PR-curve (c) ROC and for BWO-KELM (d) confusion matrix (e) PR-curve (f) ROC (see online version for colours)

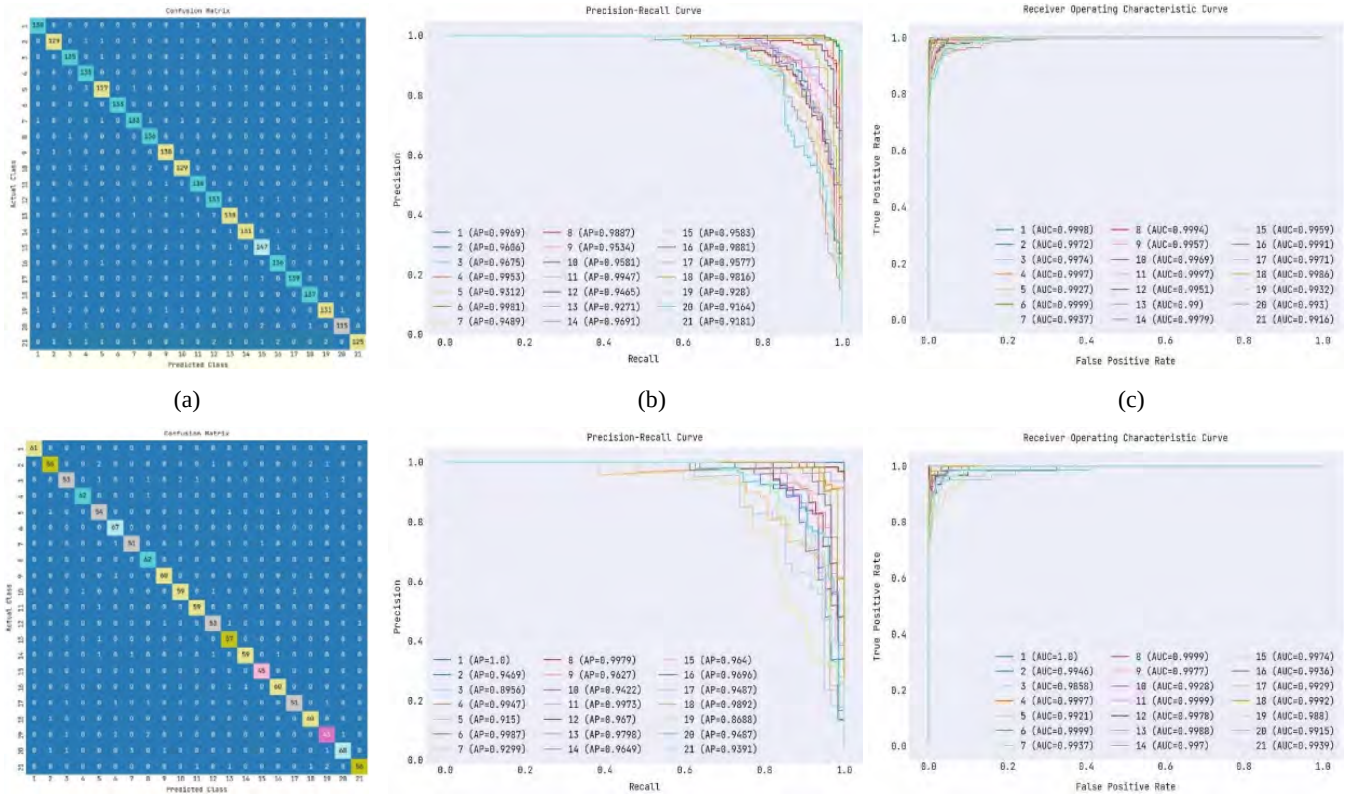
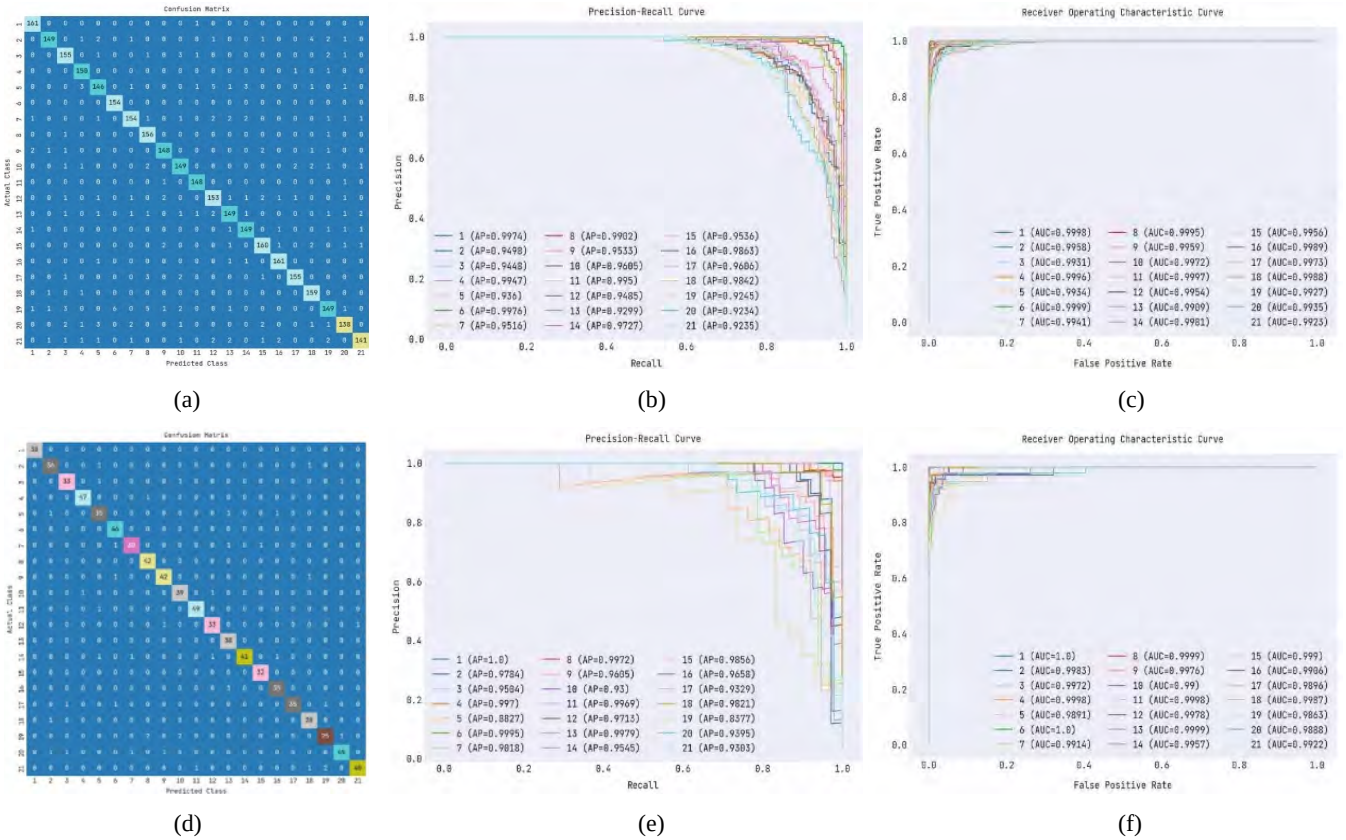


Figure 10 Results analysis on 80:20 of UCM dataset for KELM for BWODLF-RSI method, (a) confusion matrix (b) PR-curve (c) ROC and for BWO-KELM (d) confusion matrix (e) PR-curve (f) ROC (see online version for colours)



4.3.2 Result analysis of proposed model on 70:30 of UCM dataset

In Figure 9, BWO-KELM and KELM classifier models were tested on a 70:30 UCM database split. BWO-KELM outperformed KELM with a peak ROC of 0.9999, showcasing high precision and recall rates and fewer false positives. This data supports the improved performance of the BWODLF-RSI approach.

4.3.3 Result analysis of proposed model on 80:20 of UCM dataset

Figure 10 showcases the results of the BWO-KELM and KELM classifiers applied to the entire UCM database under the BWODLF-RSI method. The BWO-KELM method outperformed the KELM approach, achieving the highest ROC of 0.9998 and illustrating effective positive instance classification with reduced errors as evident from the precision-recall curve.

Table 2 reports a brief classification outcome of the BWODLF-RSI method on the 80:20 and 70:30 of UCM databases. The outcomes stated that the BWODLF-RSI method has obtained better performance under all aspects. It is noticed that the BWODLF-RSI method has demonstrated maximum performance under all TR and TS data.

Table 2 Result analysis of BWODLF-RSI approaches under 70:30 and 80:20 of UCM dataset

	UCM dataset			
Measures	*TR (70%)*	*TS (30%)*	*TR (80%)*	*TS (20%)*
Accuracy	94.84	94.91	94.81	94.96
Precision	94.81	94.92	94.79	95.02
Sensitivity	94.84	94.91	94.81	94.96
Specificity	99.74	99.75	99.74	99.76
F-score	94.79	94.86	94.77	94.93
NPV	99.74	99.75	99.74	99.76

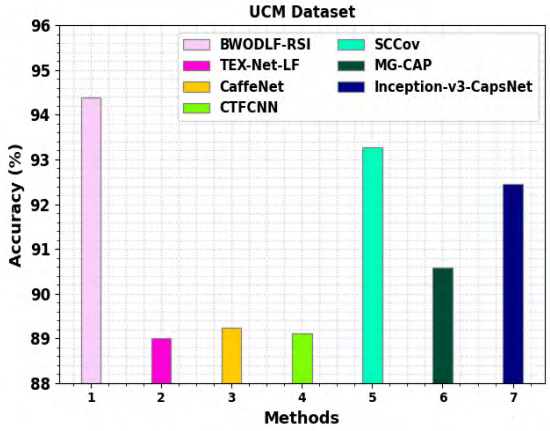
4.3.4 Comparative result analysis on UCM dataset

Table 3 and Figure 11 present a comparative analysis of the BWODLF-RSI model against other DL models using the UCM dataset for image classification. The data reveals that while other models exhibited a performance with accuracy ranging from 89% to 93.28%, the BWODLF-RSI model surpassed them with a peak accuracy of 94.38%. This superior performance indicates the model's substantial potential in applications necessitating precise RSI classification and accurate identification of positive instances.

Table 3 Accuracy analysis of BWODLF-RSI system with other techniques under UCM database

Methods	*Accuracy (%)*
BWODLF-RSI	94.38
TEX-Net-LF	89.00
CaffeNet	89.23
CTFCNN	89.11
SCCov	93.28
MG-CAP	90.59
Inception-v3-CapsNet	92.46

Figure 11 Accuracy analysis of BWODLF-RSI approach under UCM dataset (see online version for colours)



4.4 Result analysis of proposed model on AID dataset

This section examines BWODLF-RSI model on AID dataset in three aspects of training/testing (TR/TS) data:

1 entire UCM dataset

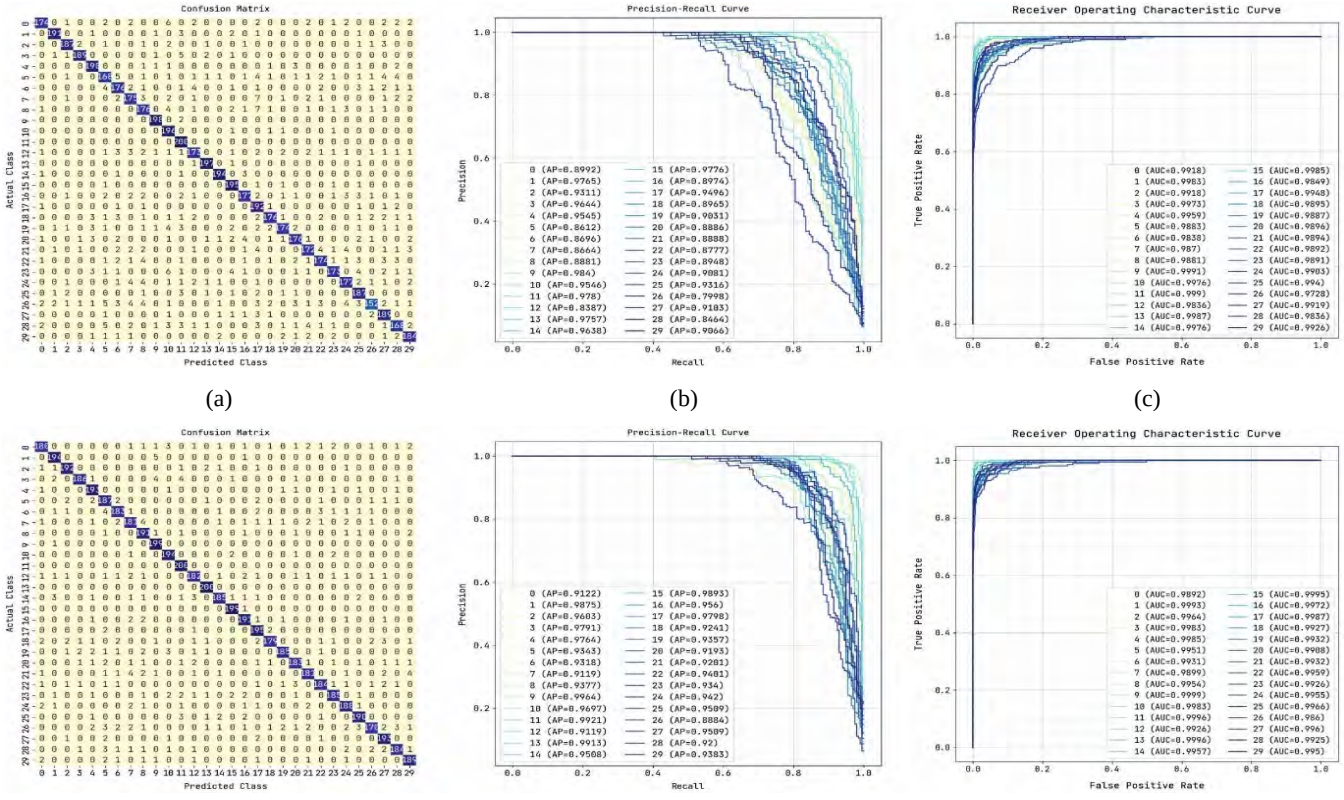
2 70:30 of UCM dataset

3 80:20 of UCM dataset.

4.4.1 Result analysis of proposed model on entire AID dataset

Figure 12 showcases the superior classification outcomes of the BWO-KELM method over the KELM when applied to the entire AID database within the BWODLF-RSI framework, achieving a notable ROC of 0.9994 as evidenced by the PR and ROC curves.

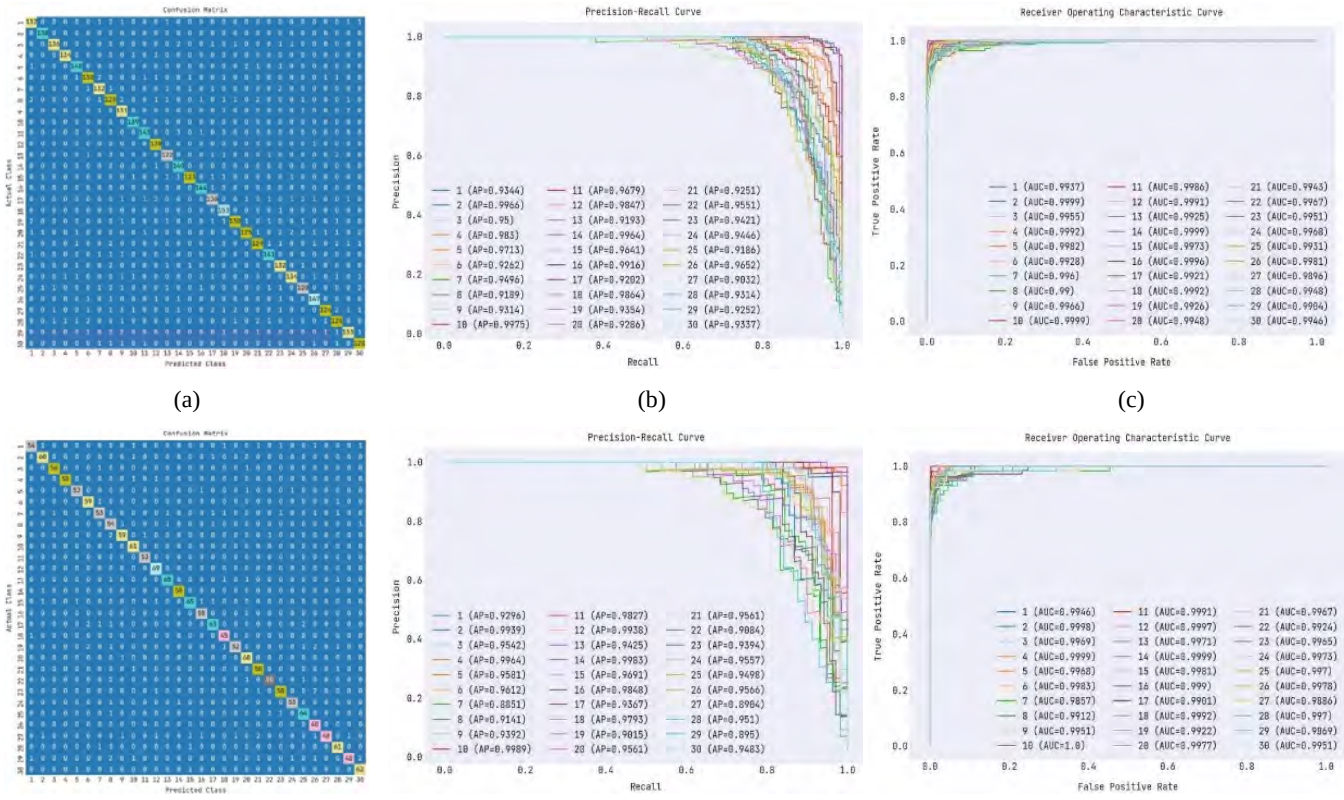
Figure 12 Results analysis on entire AID dataset for KELM for BWODLF-RSI method, (a) confusion matrix (b) PR-curve (c) ROC, and for BWO-KELM (d) confusion matrix (e) PR-curve (f) ROC (see online version for colours)



(a) (b) (c)

(d) (e) (f)

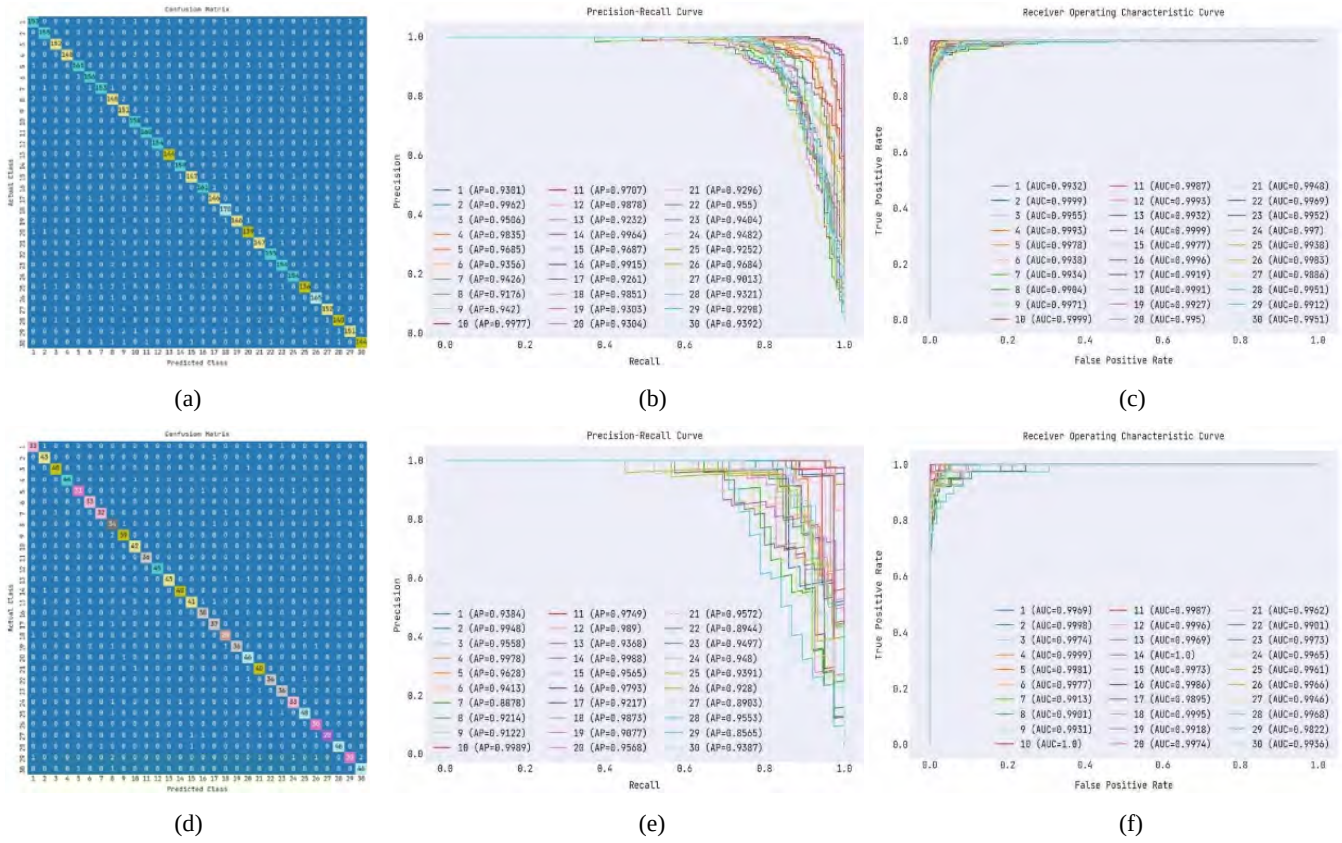
Figure 13 Results analysis on 70:30 of AID dataset for KELM for BWODLF-RSI method, (a) confusion matrix (b) PR-curve (c) ROC and for BWO-KELM (d) confusion matrix (e) PR-curve (f) ROC (see online version for colours)



(a) (b) (c)

(d) (e) (f)

Figure 14 Results analysis on 80:20 of AID dataset for KELM for BWODLF-RSI method, (a) confusion matrix (b) PR-curve (c) ROC, and for BWO-KELM (d) confusion matrix (e) PR-curve (f) ROC (see online version for colours)



(a) (b) (c)

(d) (e) (f)

Table 4 presents the classification outcomes of the BWODLF-RSI system and the KELM method on the entire AID database, highlighting the superior performance of the BWODLF-RSI algorithm. The algorithm showcases improved metrics, including a 94.05% accuracy and a 99.52% ROC score, among others.

Table 4 Result analysis of KELM and BWODLF-RSI approaches under entire AID dataset

Metrics	*KELM*	*BWODLF-RSI*
Accuracy	90.83	94.05
Precision	91.01	94.09
Recall	90.83	94.05
Specificity	99.68	99.79
PR AUC score	91.28	94.77
ROC score	99.12	99.52
F1-score	90.80	94.03
MCC	90.55	93.84

4.4.2 Result analysis of proposed model on 70:30 of AID dataset

Figure 13 highlights the superior performance of BWO-KELM over KELM in classifying a 70:30 AID database split under the BWODLF-RSI framework, showcasing improved precision-recall and ROC curves with a peak ROC score of 0.9999.

4.4.3 Result analysis of proposed model on 80:20 of AID dataset

Figure 14 illustrates the superior performance of BWO-KELM over KELM on the 80:20 AID database for classification utilising the BWODLF-RSI method, demonstrating improved precision-recall and ROC curves with a top ROC score of 0.9999.

Table 5 presents the results of the BWODLF-RSI on 80:20 and 70:30 splits of the AID databases. The method outperformed others in all evaluated aspects, showcasing improved performance across all TR and TS data.

Table 5 Result analysis of BWODLF-RSI approaches under 70:30 and 80:20 of AID dataset

	AID dataset			
Measures	*TR (70%)*	*TS (30%)*	*TR (80%)*	*TS (20%)*
Accuracy	94.80	94.56	94.96	93.62
Precision	94.86	94.63	94.99	93.99
Sensitivity	94.80	94.56	94.96	93.62
Specificity	99.82	99.81	99.83	99.79
F-Score	94.79	94.48	94.94	93.63
NPV	99.82	99.81	99.83	99.79

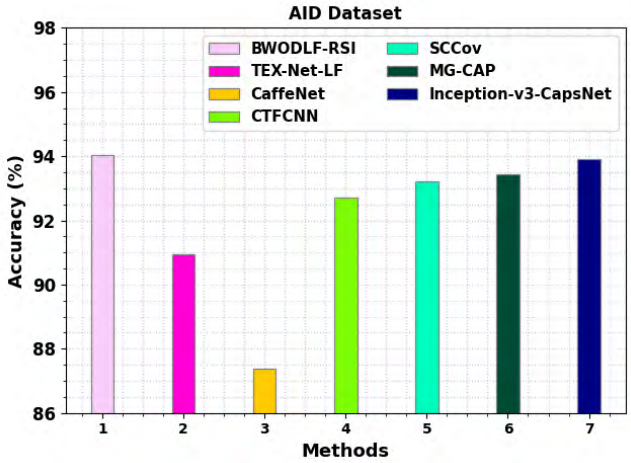
4.3.4 Comparative result analysis on AID dataset

Table 6 and Figure 15 highlight the BWODLF-RSI algorithm's superior performance over other deep learning approaches on the AID dataset, showcasing its potential for precise RSI classification. The TEX-Net-LF and CaffeNet systems recorded lower accuracies of 90.93% and 87.37%, respectively.

Table 6 Accuracy analysis of BWODLF-RSI approach with other techniques under AID dataset

Methods	*Accuracy (%)*
BWODLF-RSI	94.05
TEX-Net-LF	90.93
CaffeNet	87.37
CTFCNN	92.71
SCCov	93.22
MG-CAP	93.43
Inception-v3-CapsNet	93.92

Figure 15 Accuracy analysis of BWODLF-RSI approach under AID dataset (see online version for colours)



The BWODLF-RSI algorithm outperformed other approaches, achieving the highest classification accuracy of 94.05%. Other methods like CTFCNN and Inception-v3-CapsNet also had reasonable accuracies, ranging from 92.71% to 93.92%. The results underline the proposed model's potential to enhance RSI applications, including image recognition and object detection, through fusion techniques and advanced methodologies.

5 Conclusions

A new BWODLF-RSI system has been introduced to better identify various categories within RSIs. Firstly, there's a three-part pre-processing stage; it includes GF-based noise reduction to enhance the focal area, CLAHE-enabled contrast boosting to enhance the features and data augmentation to avoid overfitting during RSI analysis. Following this, a combination of Inceptionv3 and DenseNet201 models was utilised for a strong feature extraction process, laying a firm groundwork for the classification stage. The final step is BWO technique working in tandem with the KELM model, significantly improving efficiency and predictive precision. Through extensive testing on benchmark datasets, this BWODLF-RSI system demonstrated encouraging results, outperforming existing methods with a classification accuracy of 94.05%, compared to 90.93% and 87.37% achieved by TEX-Net-LF and CaffeNet, respectively. Looking ahead, we see opportunities to further refine deep learning architectures for even more accurate RSI classification. Exploring transfer learning for tailored RSI analysis is a promising avenue for enhancing classification outcomes. Our main goal with the BWODLF-RSI system is to facilitate improved analysis and interpretation of RSIs, a tool that finds critical applications for urban planning, disaster management, and weather analysis, among others.

References

Boualleg, Y., Farah, M. and Farah, I.R. (2019) 'Remote sensing scene classification using convolutional features and deep forest classifier', *IEEE Geosci. Rem. Sens. Lett.*, December, Vol. 16, No. 12, pp.1944–1948.

Chen, C., Zhang, B., Su, H., Li, W. and Wang, L. (2016) 'Land-use scene classification using multi-scale completed local binary patterns', *SIViP*, April, Vol. 10, No. 4, pp.745–752, https://doi: 10.1007/s11760-015-0804-2.

Cheng, G., Han, J. and Lu, X. (2017) 'Remote sensing image scene classification: benchmark and state of the art', *Proc. IEEE*, October, Vol. 105, No. 10, pp.1865–1883.

Choudhary, S. and Sharma, R. (2023) 'CNN-based battlefield classification and camouflage texture generation for real environment', *Int. J. of Comp. Sci. and Engg.*, Vol. 26, No. 3, pp.231–242.

Daldegan, G., Roberts, D. and Ribeiro, F. (2019) 'Spectral mixture analysis in Google Earth engine to model and delineate fire scars over a large extent and a long time-series in a rainforest-savanna transition zone', *Rem. Sens. of Env.*, Vol. 232, No. 1, p.111340.

Emery, W.J. and Camps, A. (2017) *Introduction to Satellite Remote Sensing: Atmosphere, Ocean, Land and Cryosphere Applications*, Elsevier, Amsterdam, Netherlands; Cambridge, MA.

Feng, W., Gou, J., Fan, Z. and Chen, X. (2023) 'An ensemble machine learning approach for classification tasks using feature generation', *Conn. Science*, 31 December, Vol. 35, No. 1, p.2231168.

Guo, J., Jiang, Z. and Jiang, D. (2021) 'Multi-level spatial attention network for image data segmentation', *Intl J. of Embed. Sys.*, Vol. 14, No. 3, pp.289–299.

Han, W., Feng, R., Wang, L. and Cheng, Y. (2018) 'A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification', *ISPRS J. Photogram. and Rem. Sens.*, November, Vol. 145, No. 1, pp.23–43.

Hayyolalam, V. and Kazem, A.P.A. (2020) 'Black widow optimization algorithm: a novel meta-heuristic approach for solving engineering optimization problems', *Engg. App. of Arti. Intel.*, Vol. 87, No. 1, p.103249.

https://captain-whu.github.io/AID/ (accessed 18 December 2022).

Kuran, U. and Kuran, E.C. (2021) 'Parameter selection for CLAHE using multi-objective cuckoo search algorithm for image contrast enhancement', *Intel. Sys. with App.*, November, Vol. 12, No. 1, p.200051.

Li, F., Feng, R., Han, W. and Wang, L. (2020) 'High-resolution remote sensing image scene classification via key filter bank based on convolutional neural network', *IEEE Trans. Geosci. Rem. Sens.*, November,. Vol. 58, No. 11, pp.8077–8092.

Chen, H., Miao, F., Chen, Y., Xiong, Y. and Chen, T. (2021) 'A hyperspectral image classification method using multi-feature vectors and optimized KELM', *IEEE J. of Selected Topics in Applied Earth Observ. and Rem. Sen.*, Vol. 14, No. 1, pp.2781–2795.

Niu, L., Pan, M., Zhou, Y. and Xiong, L. (2020) 'Study on optimal urban land classification method based on remote sensing images', *Int. J. Info. and Comm. Tech.*, Vol. 16, No. 4, pp.365–377.

Sebai, H., Kourgli, A. and Serir, A. (2015) 'Dual-tree complex wavelet transform applied on color descriptors for remote-sensed images retrieval', *J. Appl. Rem. Sens.*, October, Vol. 9, No. 1, p.095994.

Sun, K., Luo, X., Ma, L. and Zhu, S. (2022) 'Prior distributions-based data augmentation for object detection', *Intl. J. of Comp. Sci. and Eng.*, Vol. 25, No. 1, pp.34–43.

Venkatesan, R. and Prabu, S. (2022) 'Profound feature extraction and classification of hyperspectral images using deep learning', *Int. J. Intel. Enterprise.*, Vol. 9, No. 3, pp.318–331.

Wang, C. et al. (2019) 'Pulmonary image classification based on Inception-v3 transfer learning model', *IEEE Acc.*, Vol. 7, pp.146533–146541.

Wang, S-H. and Zhang, Y-D. (2020) 'DenseNet-201-based deep neural network with composite learning factor and precomputation for multiple sclerosis classification', *ACM Trans. Multimedia Comp. Comm. Appl.*, July, Vol. 16, No. 2s, pp.1–19, https://doi: 10.1145/3341095.

Yang, D., Du, Y., Yao, H. and Bao, L. (2022) 'Image semantic segmentation with hierarchical feature fusion based on deep neural network', *Conn. Science*, 31 December, Vol. 34, No. 1, pp.1772–1784.

Yang, L., Jiang, Z., Zhou, H. and Guo, J. (2021) 'DCNet: diffusion convolutional networks for semantic image segmentation', *Intl. J. of Embed. Sys.*, Vol. 14, No. 3, pp.300–311.

Yang, Z., Mu, X. and Zhao, F. (2018) 'Scene classification of remote sensing image based on deep network and multi-scale features fusion', *Optik*, October, Vol. 171, No. 1, pp.287–293.

Yao, J., Dai, Y., Ni, Y., Wang, J. and Zhao, J. (2020) 'Deep characteristics analysis on travel time of emergency traffic', *Intl. J. of Comp. Sci. and Eng.*, Vol. 22, No. 1, pp.162–169.

Zhang, W., Tang, P. and Zhao, L. (2019) 'Remote sensing image scene classification using CNN-CapsNet', *Rem. Sens.*, February, Vol. 11, No. 5, p.494.

Zhou, L., Zhou, Z. and Hu, D. (2013) 'Scene classification using a multi-resolution bag-of-features model', *Pattern Recog.*, January, Vol. 46, No. 1, pp.424–433.

Zhou, W., Newsam, S., Li, C. and Shao, Z. (2018) 'PatternNet: a benchmark dataset for performance evaluation of remote sensing image retrieval', *ISPRS J. of Photogram. and Rem. Sens.*, November, Vol. 145, pp.197–209.