

Review of Computer Engineering Research

2024 Vol. 11, No. 2, pp. 58-72

ISSN(e): 2410-9142


ISSN(p): 2412-4281

DOI: 10.18488/76.v11i2.3706

© 2024 Conscientia Beam. All Rights Reserved.



A multi-attention residual integrated network with enhanced fireworks algorithm for remote sensing image classification

 **Josephine Anitha Antony¹⁺**
Gladis Dennis²

¹Department of Computer Science, Presidency College, University of Madras, Chennai, India.

Email: anitharesearchscholar@gmail.com

²Bharathi Women's College, Chennai, India.

Email: gladischristopher@gmail.com



(+ Corresponding author)

ABSTRACT

Article History

Received: 22 December 2023

Revised: 14 February 2024

Accepted: 12 March 2024

Published: 5 April 2024

Keywords

Feature fusion
Image classification
Multi-attention
Remote sensing
Residual network.

This research examines a multi-attention residual integrated network with an enhanced fireworks algorithm for remote sensing image classification. Remote sensing (RS) picture classification is important for land cover mapping, environmental monitoring, and urban planning. Remote sensing image classification is important in earth observation since the military and commercial sectors have focused on it. Due to RS data's high complexity and limited labelled examples, classifying RS pictures is difficult. Deep Learning (DL) techniques have made great strides in RS image categorization, expanding this field's potential. This research introduces Multi-Attention Residual Integrated Network with Enhanced Fireworks Algorithm (MAR-EFA) to improve hyper spectral image identification. MARIN-EFA improves feature fusion and removes unneeded features to overcome technique constraints. The suggested method weights features using different attention models. These characteristics are then carefully extracted and integrated using a residual network. Final contextual semantic integration on deeply fused features is done with a Bi-LSTM network. Our population-based Enhanced Fireworks Algorithm (EFA) is inspired by fireworks' explosive performance and optimises MARIN parameters. Attention techniques and an improved optimisation algorithm improve performance over current systems. Numerous Eurosat dataset studies were assessed using various performance indicators. The simulation results show that MARIN-EFA outperforms current methods. The suggested technique shows promise for improving RS picture classification and allowing more accurate and reliable data categorization.

Contribution/Originality: We introduce the Enhanced Fireworks Algorithm (EFA), a population-based optimization algorithm that draws inspiration from the spectacular display of fireworks. The utilization of attention mechanisms and the enhanced optimization algorithm contribute to achieving superior performance compared to existing methods. Extensive experiments were conducted on Eurosat dataset, and the results were evaluated using various performance metrics. The simulation outcomes clearly prove that the MARIN-EFA method outclasses state-of-the-art techniques currently available.

1. INTRODUCTION

An imaging satellite, which might be operated by businesses or governments, is used to generate satellite photographs of the earth. In general, remote sensing may be understood as the process of assessing and collecting data on an event, region, or entity without really being in contact with it [1]. For several uses, including land use categorization, RS data is being considered a crucial resource, especially when combined with Deep Learning

approaches. We now have access to many other types of RS photos, such as synthetic aperture radar (SAR), multi/hyper spectral, and very high resolution (VHR) pictures. Precision farming, city planning, ecological safeguarding, and conservation of natural resources are just a few of the many fields that greatly benefit from land cover categorization with RS images [2]. Land use and land cover (LULC) have emerged as a primary driver of ecological change as human activity has spread over the planet. As a result, fulfilling these objectives relies heavily on LULC extraction with a high degree of precision. For decades, the standard method of producing LULC using single-temporal RS pictures as objects has been to apply a class label to each pixel using machine learning. Several other classification methods have been developed and used for LULC production [3, 4]. Unfortunately, only a few LULC categories are recognized since a single-temporal RS picture only offers the prompt spectrum of the land surface, and the characteristics that may be employed for categorization are rare [5]. For the categories of various land use and land change applications, it is extremely challenging to achieve good classification accuracy using a single-temporal RS picture [6]. Recent years have seen tremendous advancements in RS image categorization thanks to the proliferation of deep learning (DL) methods [7]. Due to the widespread success convolutional neural networks (CNN) in the computer vision community, many CNN-based approaches have been proposed for RS classification. When compared to a different traditional model, this one provides superior classification performance [8]. However, the CNN feature is only partially utilized in this study, as the rich hierarchical data from the deleted convolution layer is not utilised in the CNN feature extraction. One common workaround is downsizing the original scene image to a standard resolution, which results in a loss of informative detail [9].

Recent work has shown that deep learning (DL) algorithms may learn picture features in a task-specific manner, improving classification accuracy over traditional scene classification methods [10]. However, there are two major obstacles that significantly hinder the application of DL approaches [11]: Training the DL approach's module takes a lot of time and a lot of data. It has been shown in a number of studies that current pre-trained CNN techniques may be transferred to different detection tasks, even those not originally designed for them. For the three RS scenes dataset, Nogueira, et al. [12] showed that using a pre-trained CNN as the feature extractor outperformed entirely training a novel CNN. In this paper, a novel method called Multi-Attention Residual Integrated Network with Enhanced Fireworks Algorithm (MAR-EFA) is planned to improve hyperspectral image recognition in remote sensing (RS) image classification. The research aims to address the challenges associated with RS image classification, such as high dimensionality and limited labeled samples, by utilizing deep learning (DL) approaches, attention mechanisms, and an enhanced optimization algorithm. The suggested method focuses on getting rid of features that aren't needed, improving feature fusion, using a residual network to extract and combine features, and using a Bi-LSTM network for contextual semantic integration. The researchers conducted extensive experiments on a standard dataset and evaluated the results using various performance metrics. The results show that the MARIN-EFA method works better than other methods, showing that it has the potential to improve RS image classification and make RS data categorization more accurate and reliable.

The major contributions from the research are:

1. Image augmentation techniques using spectral and spatial augmentation can be beneficial for enhancing the performance and robustness of remote sensing image classification models.
2. Proposal of a novel approach called Multi-Attention Residual Integrated Network with Enhanced Fireworks Algorithm (MARIN-EFA) to improve hyperspectral image recognition.
3. Incorporation of different attention models to assign varying weights to different features improves the classification performance.
4. Application of a Bi-LSTM network for contextual semantic integration on the deeply fused features enhances the understanding of spatial relationships.
5. Introduction of the Enhanced Fireworks Algorithm (EFA) for optimizing the parameters of MARIN, improving convergence and solution quality.

2. LITERATURE SURVEY

The model complexity made available by deep learning makes it possible to train mechanized end-to-end processes for generating visualizations of features from data. According to Yuan, et al. [13] deep learning is a game-changing innovation in the field of machine learning, with impressive results in areas such as object detection, semantic segmentation, picture recognition, etc. A few examples include Liu and Shi [14]; Wang, et al. [15] and Dou, et al. [16]. Due to its flexibility in expressing features and mechanization via expert-free knowledge acquisition, deep learning has been successfully applied in various areas of RS study. In particular, deep learning has become popular for time series classification due to its ability to deal with temporal aspects. Different methods have been used to make LULC mapping better. These include models using convolutional neural networks (CNNs), such as temporal CNNs (TCNNs), and models based on sequential data, such as deep recurrent neural networks (RNNs) and bidirectional long short-term memories (Bi-LSTMs).

We cite the work of Zhong, et al. [6] due to the lengthy and complicated nature of TSI's temporal relationships, it is still an open challenge to build the deep learning architecture necessary to analyze RS time series. Source: Wang, et al. [17]. Reportedly the most effective models for recognizing patterns in the 2D domain, CNN-based algorithms also achieve excellent classification accuracy in the 1D domain, where they may be used to spectral and time series data. Better performance requires a well-designed architecture of the CNN models, which requires more testing, which in turn makes the models obese and complicated. When employing a deep learning architecture to categorize temporal relationships, the lower layers often capture fine-grained differences in temporal scale, while the higher layers concentrate on macro-level patterns. However, deep construction has a minimal impact on classification enhancements when the time series is short or the temporal density is low.

Pan, et al. [18] suggested and investigated the optimal parameter for a CNN-based multidimensional LiDAR land cover categorization architecture. The approach first converts the multispectral 3D Light Detection and Ranging (LiDAR) data to 2D pictures for further processing. Then, utilizing seven fundamental functional layers, CNN models are built, and their hyperparameters are extensively tuned and explained. While Kwan, et al. [19] investigated the recital of two CNN-based computational models for identifying land cover using five bands, the Extended Multi-Attribute Profiles (EMAP) technique expands this limited number of picture bands. In Zhang, et al. [20] introduced a cutting-edge multi-level context-guided classification method with an object (MLCG-OCNN) algorithm. For accurate object discrimination, we introduce a feature-fusing OCNN, which learns higher-level features derived from geometric features, object-level contextual data, and autonomous spectral patterns all at once. This allows us to take advantage of all three types of data simultaneously. The next step is to use contextual assistance at the pixel level to further enhance the per-object categorization outcome.

To improve the performance of LULC classification, Rajendran, et al. [21] provide a hybrid feature optimization model and DL classifiers that aid in foreseeing animal habitat, random elements, declining environmental excellence, and so on. Eurosat, Sat 4, and Sat 6 are used to quantify LULC classes. After that, the choice of a suitable remote sensing reputation, standardization, and histogram equalization technique improve picture quality. Novel cross-attention and graph-convolution integration methods were presented in the work of Pu, et al. [22]. To begin, a PCA model is used to scale down the size of the hyperspectral picture so that we may extract more meaningful information from it. The models then employ a cross attention strategy to distribute importance. The final step is to use the created deep features and the connections between the deep features to finish the forecasts of hyperspectral data. A unique RS image classification technique based on stacked denoising auto encoders (DAE) is described by Liang, et al. [23] which is driven by DL approach. At first, the DAE stacked layer is used to build the deep network models. After that, characteristics are acquired by supervised modeling using Back Propagation Neural Networks (BPNN), and the complete systems are improved via error BPNN; all of this is done with noisy input and unsupervised greedy layerwise training techniques.

3. METHODS AND MATERIALS

The MARIN-EFA algorithm's model outline is exemplified in Figure 1. The algorithm utilizes a multi-attention mechanism to fully mine the features of the input information. This mechanism assigns varying weights to different features, allowing for more effective feature extraction. The residual network is then employed to calculate the residual of the multi-attention mechanism. By doing so, the network generates residual features that capture additional representative information. To integrate the deep features obtained from the residual network, a two-way long and short-term memory network (Bi-LSTM) is utilized. This network enables contextual semantic integration, further enhancing the understanding of the features and their relationships.

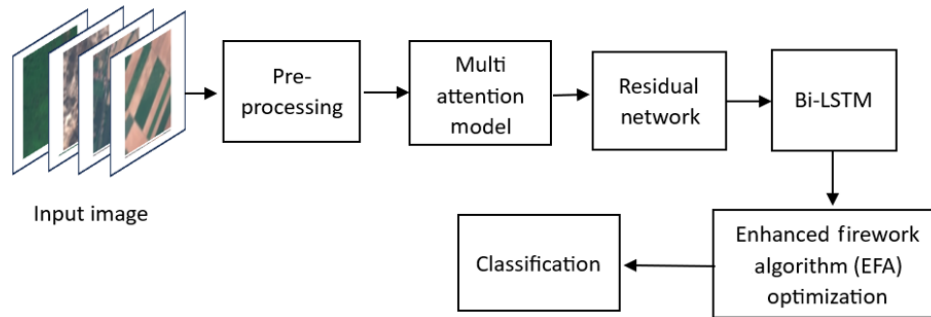


Figure 1. The overview of the process flow of the proposed methodology in classifying the RS images.

To enhance the parameters of the MARIN model, the Enhanced Fireworks Algorithm (EFA) is introduced. EFA is a population-based optimization algorithm that draws inspiration from the explosive behavior of fireworks. It improves the convergence and solution quality of the algorithm, leading to better performance. In the final stage, the Softmax classification function is applied to forecast the class likelihood for each pixel. This function assigns probabilities to different land cover categories, enabling the final recognition of the image.

3.1. Dataset

The dataset provided includes images from the EuroSat dataset, which are organized into two main folders: "EuroSAT" and "EuroSATallBands." The "EuroSAT" folder contains Red, Green, Blue (RGB) images that were collected from the Sentinel Dataset. These images represent different land cover categories and were captured using the Sentinel-2 satellite. On the other hand, the "EuroSATallBands" folder contains tif files that include all the bands of the electromagnetic spectrum as collected from the Sentinel-2 satellite. These files provide a more comprehensive representation of the spectral information for each image. Each image in the dataset has a resolution of 64x64 pixels and a Ground Sampling Distance (GSD) of 10 meters. This GSD refers to the spatial resolution at which the images were acquired. The dataset is organized into class folders, representing different land cover types. These class folders include the images as shown in the Table 1.

Table 1. The classes and the number of images from the classes.

Classes	Image counts
Annual crop	3000
Forest	3000
Herbaceous vegetation	3000
Highway	2500
Industrial	2500
Pasture	2000
Permanent crop	2500
Residential	3000
River	2500
Sea Lake	3000

These class folders allow for easy categorization and analysis of the images based on their respective land cover types. The dataset provides a valuable resource for tasks such as land cover mapping, environmental monitoring, and other applications in the field of remote sensing.

3.2. Augmentation

Let's imagine that the $m \times n$ pixels of the RS picture are laid out in $m \times n$ grids, where m is the number of rows in the image and n is the number of columns. The top left corner of the grid has a coordinate of $(0, 0)$, whereas the bottom right corner has a coordinate of (m, n) . Where (i, j) is a spatial collaborate, $x_{l,k}^t$ is the spectral vector of the labeled test the specimen, and $x_{l,k}^s$ is the spectrum vector of the untrained sample in RS pictures. Sample augmentation with spectral and geographic limitations requires the annotated sample to be assigned the category, thereby increasing the sample size. Prior to utilizing the sample augmentation strategy, all unlabelled samples have their labeled categories set to 0, which indicates that they belong to no specific classification.

At start, we pick a group of samples that includes all classes shown as $SA_{l,k}^c$ that are most distant from the unlabeled sample $x_{l,k}^s$. Comparison between the unlabeled specimen $x_{l,k}^s$ and every component of $SA_{l,k}^c$ is quantified by calculating the spectral angle distance (SD) using Equation 1. If the SD value is low, then the unlabelled sample $x_{l,k}^s$ most likely falls into this group. Therefore, the unlabeled specimen previously categorized as $X_{l,k}^c$ is placed in the SAD group with the lowest SD.

$$SD(\rho, \vartheta) = \arccos\left(\frac{\langle \rho, \vartheta \rangle}{\|\rho\| \|\vartheta\|}\right) \quad (1)$$

ρ, ϑ are the pixel spectral vectors.

For spatial limitations, we take use of local neighbourhood similarities. The unlabelled sample $x_{l,k}^s$'s central pixel is assumed to be extracted as a local window from the hyperspectral picture. The local window has a black central pixel where the unlabelled sample $x_{l,k}^s$ is located, and the neighbourhood-coloured pixels are where the training samples are. The training samples were categorized and labeled using a rainbow of colors. Training samples in the local window are labeled as $SP_{l,k}^d$, where d is the number of distinct classes. Similarity among the unlabeled sample $x_{l,k}^s$ and every group $SP_{l,k}^d$ in the local window is measured using a local neighborhood similarity metric. Assume the training sample with the known label is located in the local window at (a, b) . Following these equations, we may get the local neighborhood resemblance estimate $Y_{l,k}^d$.

$$\gamma = \sqrt{(a - l)^2 + (b - k)^2} \quad (2)$$

$$\theta = SD(x_{l,k}^s, x_{l,k}^t) \quad (3)$$

$$C_{a,b}^d = \frac{1}{\sqrt{\gamma + \theta}} \quad (4)$$

$$Y_{l,k}^d = \sum C_{a,b}^d \quad (5)$$

Where $x_{l,k}^t$ is the local-window training sample. The $C_{a,b}^d$ value indicates the percentage of the unlabeled sample that was contributed by the training sample $x_{l,k}^t$, which also belongs to category $SP_{l,k}^d$. Local neighborhood similarity measurement q is the sum of the inputs from all of the training observations in the local frame that belong to the same group $SP_{l,k}^d$. The unlabeled sample $x_{l,k}^s$ is first allocated to the class with the highest local neighborhood similarity value d . In the absence of any training samples inside the current window, $x_{l,k}^s$, the unlabeled sample, will remain at 0.

3.3. Multi-Attention Mechanism

The multi-attention mechanism, combined with a Bi-LSTM network, provides an effective approach for processing remote sensing images. The multi-attention mechanism assigns varying weights to different features, allowing the model to focus on the most relevant information. The attention mechanism can be represented mathematically as follows:

$$A = \text{softmax}(W_a \cdot f(X) + b_a) \quad (6)$$

Where:

A is the attention weight vector. W_a is the attention weight matrix. $f(X)$ represents the features extracted from the input image. b_a is the bias term. The attention weights vector A represents the importance of each feature, and the softmax function normalizes the weights to ensure their sum equals 1. Next, the attention-weighted features are combined using element-wise multiplication with the original features:

$$F_{att} = A \odot f(X) \quad (7)$$

Where:

F_{att} represents the attention-weighted features. The attention-weighted features are then fed into a Bi-LSTM network to capture temporal dependencies and further integrate contextual information. The Bi-LSTM network processes the sequence of attention-weighted features and can be represented as:

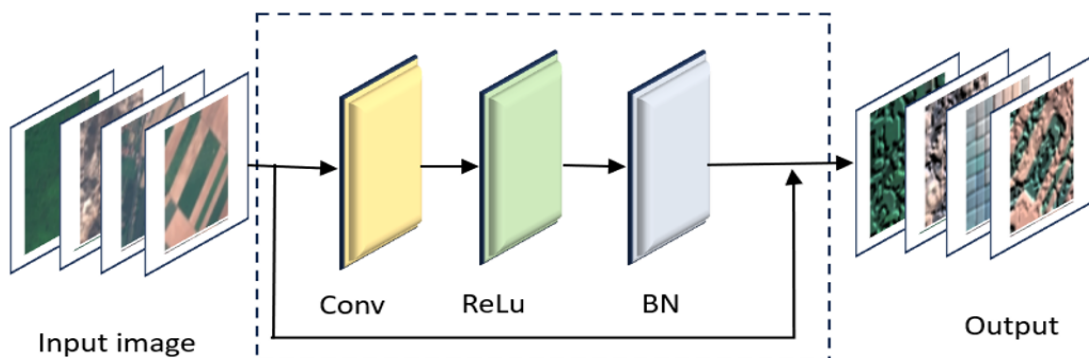
$$H = \text{BiLSTM}(F_{att}) \quad (8)$$

Where:

H represents the hidden states obtained from the Bi-LSTM network. The BiLSTM function applies the forward and backward LSTM operations to the input sequence, enabling the network to capture both past and future dependencies. The hidden states H obtained from the Bi-LSTM network capture the comprehensive contextual information, integrating both local and global features. These hidden states can be used for subsequent tasks such as classification or further analysis of the remote sensing images.

3.4. Residual Learning

Residual learning, also known as residual networks or ResNets, is a deep learning architecture that was introduced to address the issue of vanishing gradients in deep neural networks. It was first proposed by researchers at Microsoft Research in 2015 and has since become a widely adopted technique for various computer vision and natural language processing tasks. Instead of explicitly learning the intended visualization, the system is trained to acquire residual functions via skip connections or shortcut linkages. By introducing these shortcut connections, the network can bypass one or more layers and propagate information directly from earlier layers to later layers. Figure 2 shows the residual learning structure associated with the attention layers. The key benefit of residual learning is that it helps alleviate the vanishing gradient problem, which occurs when training deep networks. The vanishing gradient problem arises because gradients tend to diminish as they propagate backward through many layers, making it difficult for the early layers to learn meaningful representations. By using skip connections, the gradients have a shortcut path to flow directly to earlier layers, allowing the network to better capture and propagate information.



Residual connection

Figure 2. Residual learning network.

Note: Conv: Convolution.
ReLu: Rectified linear unit.
BN: Batch normalization.

Residual units make up the bulk of a residual network's design. Each residual block consists of multiple layers, and the input to a residual block is added to its output using the skip connection. The skip connection helps preserve the information from earlier layers and enables the network to learn residual mappings effectively. Picture classification, object identification, and picture segmentation are only some of the machine vision applications where residual neural systems have reached contemporary performance. They have also been applied to other domains, including natural language processing, where they have been used for tasks like machine translation and text classification. The formula for residual layer is:

$$z = w_k \delta(w_{k-1} x_{k-1}) + x \quad (9)$$

Where z is the layer-k output, w is the layer-input weight matrix, δ is the Relu activation function, and x is the layer-input vector. To combat vanishing gradients and expansion in deep networks, we employ a residual network to not only filter out superfluous information but also mine and fuse features from high-dimensional remote sensing pictures.

3.5. Bi-LSTM Model

An extension of the LSTM neuronal network construction, the Bidirectional Long Short-Term Memory (Bi-LSTM) structure considers data from both previous periods and the foreseeable future. It's made up of a pair of LSTMs, one of which processes the set of inputs ahead of time and the other in reverse. This improves the network's ability to comprehend the input sequence by allowing it to record connections at every step of the way. The Figure 3 shows the Bi-LSTM model architecture. The formulas for a basic Bi-LSTM structure can be described as follows:

$$in_t = \sigma(W_{int} * X_t + W_{inh} ht_{-1} + b_{int}) \quad (10)$$

$$fo_t = \sigma(W_{fot} * X_t + W_{foh} ht_{-1} + b_{fot}) \quad (11)$$

$$op_t = \sigma(W_{opt} * X_t + W_{oph} ht_{-1} + b_{opt}) \quad (12)$$

$$mt_t = ft_t * mt_{t-1} + it_t * \tanh(W_{mt} \cdot [Ht_{t-1}, X_t] + b_{mt}) \quad (13)$$

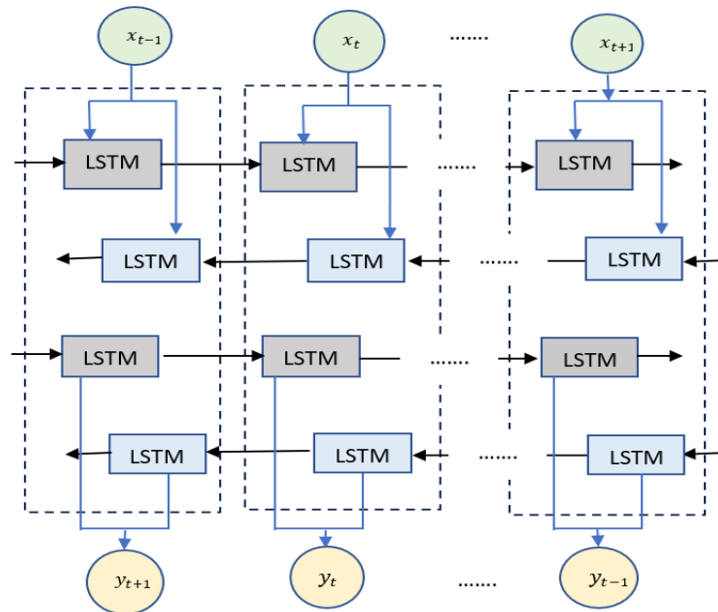


Figure 3. The proposed Bi-LSTM architecture.

These hidden states are concatenated to obtain the final hidden state h_t :

$$h_t = [h_t^f, h_t^b] \quad (14)$$

In the above formulas, the subscripts "t" and "t-1" denote the time step, \odot represents element-wise multiplication, σ denotes the sigmoid activation function, and \tanh signifies the hyperbolic tangent activation function. X_t is the input at time step t, h_t is the hidden state at time step t, mt_t is the cell state at time step t, and the

variables with apostrophes represent the backward LSTM. The outputs of the Bi-LSTM can be found by concatenating the forward and backward hidden states at each time step, resulting in a symbol that encodes information from both past and future backgrounds. This makes Bi-LSTM particularly useful for tasks such as sequence labeling, sentiment analysis, and machine translation, where understanding the context in both directions is important.

3.6. Enhanced Fireworks Algorithm

The Enhanced Fireworks Algorithm (EFA) is a nature-inspired optimization algorithm that is derived from the behavior of fireworks explosions. It is commonly used for solving optimization problems, including hyperparameter tuning in machine learning. The EFA extends the traditional Fireworks Algorithm (FA) by introducing additional mechanisms to enhance its exploration and exploitation capabilities. Here's a general outline of the Enhanced Fireworks Algorithm for hyperparameter tuning:

Algorithm 1: Enhanced fireworks Algorithm for hyperparameter tuning

1. Procedure Initialization ()
 - i. Initialize a population of fireworks with random hyperparameter configurations.
 - b. Evaluate the fitness of each fireworks using a fitness function that measures the performance of the corresponding hyperparameter configuration.
2. Procedure Explosion Operators ()
 - a. Define the explosion amplitude, which controls the spread of fireworks after the explosion. It can be calculated as:

$$i. A = A_{max} * (1 - \exp(-\vartheta * (t - t_i))) + \varepsilon,$$

Where A_{max} is the maximum explosion amplitude, γ is a scaling factor, t is the current iteration, t_i is the iteration when the fireworks were first ignited, and ε is a small constant to avoid division by zero.

3. Determine the number of sparks generated by each fireworks during explosion, denoted as S_i . It can be calculated as:
 - a. $S_i = \text{round}(S_{max} * (\text{fitness}_i / \text{fitness}_{max}))$,
4. Where S_{max} is the maximum number of sparks, fitness_i is the fitness of the fireworks, and fitness_{max} is the fitness of the best fireworks in the population.
5. Generate sparks around each firework with an explosion operator. The new position for each spark is given by:

$$a. x_s = x_i + \text{rand}() * A * (x_i - x_j),$$

Where x_i and x_j are the positions of the current fireworks and another randomly selected fireworks, and $\text{rand}()$ is a random number between 0 and 1.

6. Evaluate the fitness of each spark using the fitness function.
 7. Procedure Selection ()
 - a. Select the best fireworks from the population based on fitness evaluation.
 - b. Replace the worst fireworks in the population with the best fireworks.
 8. Procedure Termination ()
 - a. Repeat steps 2 and 3 until a termination criterion is met.
-

The formulas provided above represent the key equations used in the Enhanced Fireworks Algorithm. They define the explosion amplitude, the number of sparks generated, and the position update for each spark during the explosion process. The algorithm iteratively updates the population of fireworks, exploring different hyperparameter configurations, and selecting the best performers to guide the search towards promising areas of

the hyperparameter space. It's worth noting that specific implementations of the EFA may have variations in the formulas and parameters used, as researchers may introduce modifications or adaptations to suit their specific problem domains. Several improvements can be made to the Enhanced Fireworks Algorithm, to make it work better. These include mechanisms for maintaining diversity, intelligent migration strategies, and adaptive explosion amplitude adjustment. These enhancements aim to strike a balance between exploration (diversity) and exploitation (exploiting promising regions of the search space).

4. EXPERIMENTAL SETUP

The Bi-LSTM model is implemented using the "Keras" package. "Keras" is a Python library that facilitates the deployment of deep learning modules on the "tensor flow" framework. Since the dataset contains nonlinear residual values, "Keras" uses a stacked Bi-LSTM deep learning model to handle them. Table 2 displays the stacked Bi-LSTM model's hyperparameters. Bidirectional LSTM lessens the room for error by taking into account both past data and future predictions. We use the Adam optimizer to maximize the performance of the Bi-LSTM model.

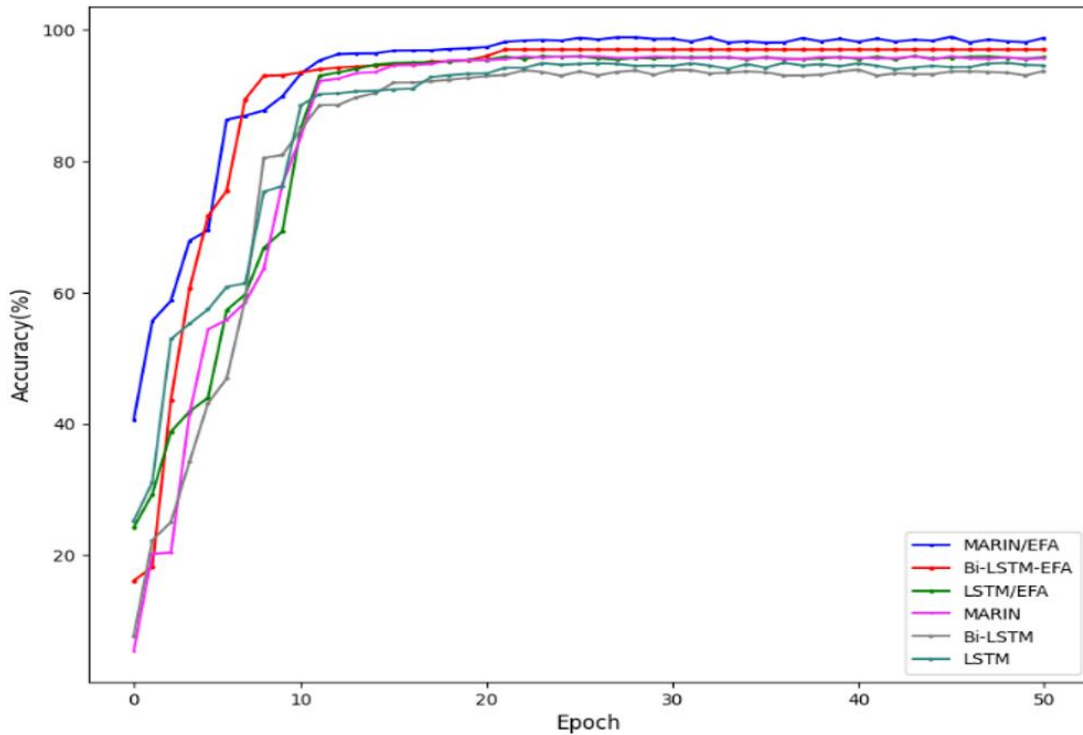
Table 2. Hyper parameter setting for the proposed model.

Hyperparameter	Value
LSTM units	128
Dropout rate	0.5
Recurrent dropout	0.2
Learning rate	0.0001
Batch size	32
Epochs	50
Optimizer	Adam
Loss function	Categorical cross-entropy

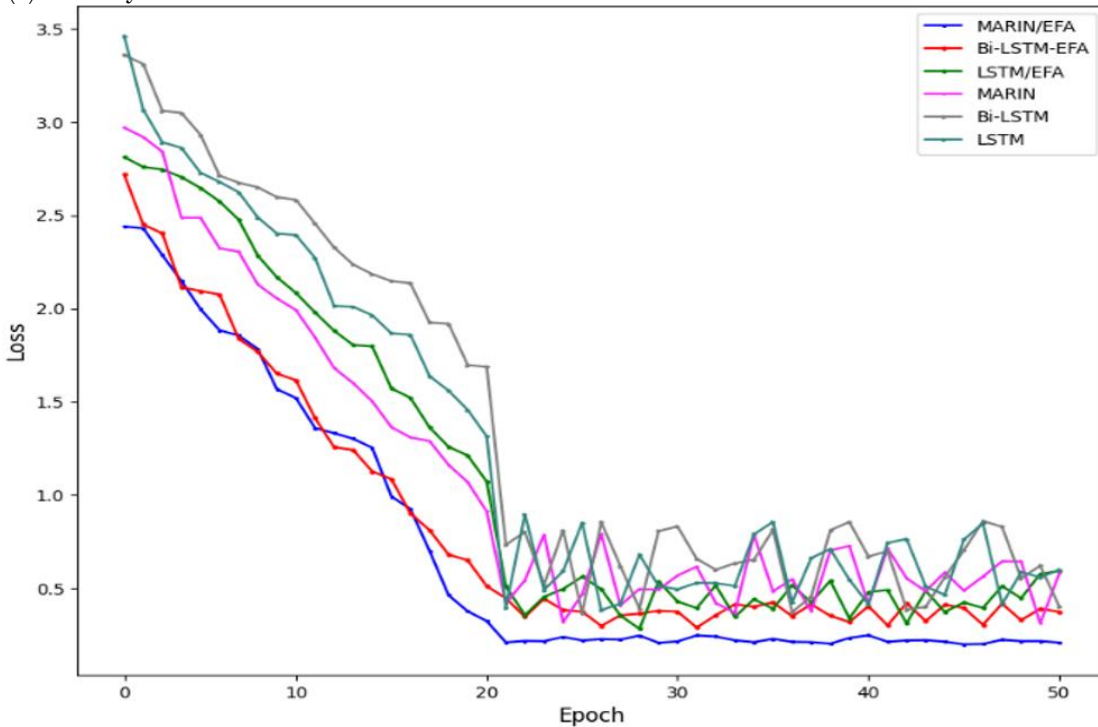
5. RESULTS AND DISCUSSIONS

The models that incorporate fire detection algorithms (MARIN-EFA, Bi-LSTM-EFA, and LSTM/EFA) demonstrate higher accuracy compared to the models without such enhancements (MARIN, Bi-LSTM, and LSTM). The use of algorithms specifically designed for fire detection in remote sensing images appears to significantly improve classification performance. MARIN-Enhanced Fire Algorithm (MARIN-EFA) model achieved the highest accuracy of 99%. The MARIN-EFA algorithm is specifically designed for fire detection in remote sensing images. It seems to be highly effective in accurately identifying fires, resulting in a very high classification accuracy. Bi-LSTM with Enhanced Fireworks Algorithm (Bi-LSTM-EFA) achieved an accuracy of 97%. This approach likely leverages the power of the Bi-LSTM architecture to capture temporal dependencies in the data, along with the fire detection capabilities of the Enhanced Fireworks Algorithm. It performs slightly lower than MARIN-EFA but still delivers a strong performance. LSTM with Enhanced Fireworks Algorithm (LSTM/EFA) achieved an accuracy of 96%. LSTM is another recurrent neural network architecture known for its ability to capture sequential information. By incorporating the Enhanced Fireworks Algorithm, this model demonstrates competitive performance in remote sensing image classification.

The MARIN model attained an accuracy of 96%. This suggests that the original MARIN algorithm, without any enhancements or additional models, is already quite effective for remote sensing image classification. It demonstrates strong performance, but with the addition of further techniques, the accuracy can be improved. The Bi-LSTM model achieved an accuracy of 94%. This model relies solely on the power of the Bi-LSTM architecture to capture temporal dependencies in the data. While it performs slightly lower than the models combined with the fire detection algorithm, it still demonstrates respectable accuracy.



(a) Accuracy



(b) Loss

Figure 4. (a) Accuracy (b) Loss of the models MARIN-EFA, Bi-LSTM-EFA, LSTM/EFA and MARIN, Bi-LSTM, LSTM.

Figure 4 illustrates the (a) Accuracy (b) Loss of the models MARIN-EFA, Bi-LSTM-EFA, LSTM/EFA and MARIN, Bi-LSTM, LSTM.

The LSTM model achieved an accuracy of 95%. Similar to the Bi-LSTM model, LSTM also captures sequential information but without the bidirectional aspect. It performs slightly better than the Bi-LSTM model but falls short compared to the models combined with the fire detection algorithm. In our paper, we evaluated several models for remote sensing image classification as shown in the Figure 4.b, including the MARIN-Enhanced Fire algorithm with a hyperparameter value of 0.2 (which remained constant after epoch 20), Bi-LSTM with a hyperparameter

value of 0.25, LSTM/EFA with a hyperparameter value of 0.28, MARIN with a hyperparameter value of 0.3, Bi-LSTM with a hyperparameter value of 0.35, and LSTM with a hyperparameter value of 0.38.

Table 3. The model's accuracy and loss comparison.

Model	Accuracy	Loss value
MARIN-enhanced fireworks (EFA)	99%	0.2
Bi-LSTM with EFA	97%	0.25
LSTM with EFA	96%	0.28
MARIN	96%	0.3
Bi-LSTM	95%	0.35
LSTM	94%	0.38

When comparing the loss values of these models, we found that the MARIN-Enhanced Fireworks algorithm achieved the lowest loss, followed by the Bi-LSTM-EFA model, as shown in the Table 3. The LSTM/EFA and MARIN models performed slightly worse, while the Bi-LSTM and LSTM models had the highest loss values. These findings indicate that the MARIN-Enhanced Fire algorithm with a hyperparameter value of 0.2 outperformed the other models in terms of loss for remote sensing image classification. The Bi-LSTM-EFA model also showed promising results, suggesting that incorporating the Enhanced Fireworks algorithm can improve the performance of the Bi-LSTM model. However, further analysis and experiments are required to fully understand the strengths and limitations of each model and to determine their suitability for different remote sensing tasks.

The effectiveness of a classification model may be measured with the use of a table called a confusion matrix. In the context of remote sensing image classification, a confusion matrix can be used to measure the accuracy of a classification algorithm in assigning different land cover or land use classes to pixels or regions within an image, as shown in the Figure 5. From the figure it is observed that the annual crop is achieved the minimum accuracy 96% and industrial, pasture, permanent crop, and sea lake achieved the maximum accuracy of 100%.

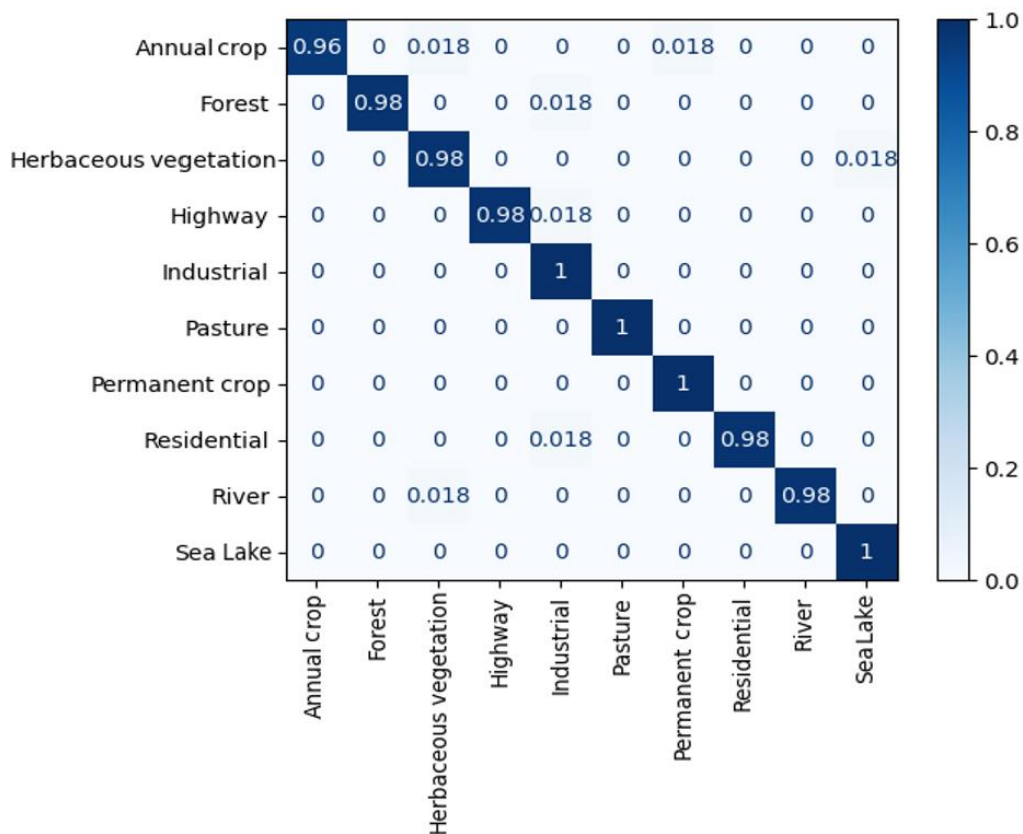


Figure 5. Confusion matrix for the remote sensing image classification of the model MARIN-EFA.

The recital evaluation of the remote sensing image classification algorithm using accuracy, precision, recall, and F1 score metrics provides valuable insights into its effectiveness, as shown in the Table 4. The results demonstrate the algorithm's overall accuracy, as well as its performance for individual land cover classes. The findings of this research contribute to the understanding of the algorithm's strengths and limitations, guiding future improvements in remote sensing image classification methodologies.

Table 4. The performance evaluation of the model MARIN-EFA.

Classes	Accuracy	Precision	Recall	F1-score	AUC
Annual crop	96%	0.97	0.98	0.97	0.98
Forest	98%	0.98	0.97	0.98	0.97
Herbaceous vegetation	98%	0.99	0.98	0.97	0.98
Highway	98%	0.99	0.98	0.99	0.99
Industrial	100%	1.00	0.99	0.99	1.00
Pasture	100%	1.00	0.99	1.00	1.00
Permanent crop	100%	1.00	0.99	1.00	1.00
Residential	98%	0.99	0.98	0.98	0.99
River	98%	0.98	0.97	0.98	0.98
Sea Lake	100%	1.00	1.00	0.99	1.00

In Figure 6, the Receiver Operating Characteristic (ROC) curve is plotted with the True Positive Rate (TPR) on the y-axis and the False Positive Rate (FPR) on the x-axis. Each point on the curve represents the performance of the classification model at a specific threshold value. The ROC curve provides valuable insights into the model's trade-off between sensitivity (TPR) and specificity (1 - FPR). A point on the upper-left corner of the plot indicates high sensitivity (a high proportion of true positive predictions) and low false positive rate (a low proportion of false positive predictions). This configuration represents an excellent classification model with high accuracy.

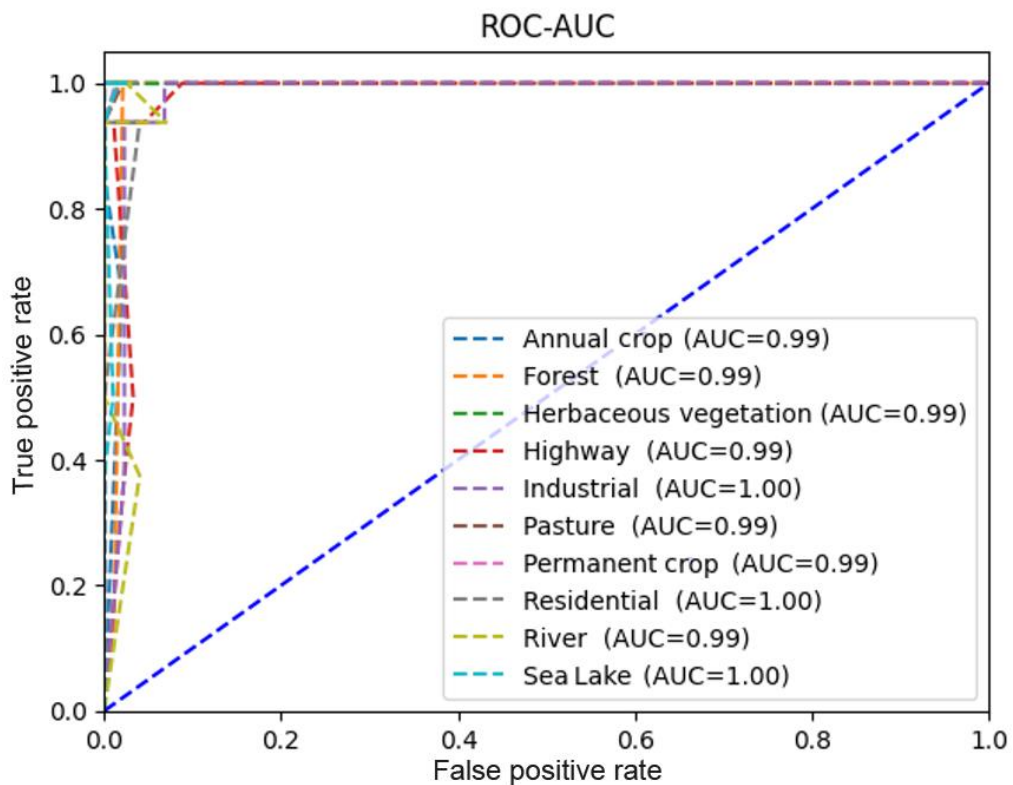


Figure 6. ROC-AUC for the remote sensing image classification on the model MARIN-EFA.

For the ROC curve, we also determine the Area Under the Curve (AUC). The AUC value is a simple way to measure the classification algorithm's efficacy. A higher AUC value indicates better discrimination between different classes. If the AUC value is close to 1, it signifies excellent performance, while a value close to 0.5 suggests poor performance (similar to random guessing). By analyzing the ROC curve and the AUC value, you can assess the classification model's effectiveness in distinguishing between classes. If the ROC curve is close to the upper-left corner and the AUC value is high, it indicates that the model has a strong ability to accurately classify the different classes. On the other hand, if the ROC curve is closer to the diagonal line (indicating a random classifier) and the AUC value is close to 0.5, it suggests that the model's performance is not significantly better than random guessing.

While the proposed approach in the given context offers promising advancements in remote sensing image classification, it is important to acknowledge some limitations that can affect the generalizability and applicability of the method. The experiments and evaluations conducted in the paper are based on a specific dataset (Eurosat). The performance of the proposed method may vary when useful to different datasets with varying characteristics. Further experiments on diverse datasets are needed to validate the generalizability of the approach. The paper does not explicitly describe the computational complications of the proposed method. Deep learning approaches, especially when incorporating attention mechanisms, can be computationally demanding. The practical feasibility and efficiency of the proposed method in real-time or near real-time applications need to be further investigated. Deep learning approaches, including attention mechanisms, are known to be black-box models that lack interpretability. The paper does not explicitly discuss the interpretability or explainability of the proposed method. In many remote sensing applications, interpretability is crucial for decision-making and understanding the reasoning behind classification outcomes.

6. CONCLUSION

In conclusion, remote sensing (RS) image classification is a critical task in multiple domains, ranging from land cover mapping to environmental monitoring and urban planning. However, the classification of RS images poses significant scientific and practical challenges due to the unique characteristics of RS data, such as high dimensionality and limited labeled samples. In recent years, Deep Learning (DL) approaches have shown remarkable advancements in RS image classification, opening new possibilities for improving this field. This paper presents a novel approach called Multi-Attention Residual Integrated Network with Enhanced Fireworks Algorithm (MAR-EFA) to enhance hyperspectral image recognition. MARIN-EFA addresses the limitations of existing methods by eliminating unnecessary features and improving feature fusion. The proposed approach incorporates different attention models to assign varying weights to different features. It employs a residual network to deeply extract and integrate these features and utilizes a Bi-LSTM network for contextual semantic integration of the fused features. The paper introduces the Enhanced Fireworks Algorithm (EFA), a population-based optimization algorithm that draws inspiration from the explosive behavior of fireworks, to optimize the parameters of MARIN. The combination of attention mechanisms and the enhanced optimization algorithm contributes to achieving superior performance compared to existing methods. MARIN-Enhanced Fireworks Algorithm (MARIN-EFA) model achieved the highest accuracy of 99%. Extensive experiments were conducted on a Eurosat dataset, and the results were evaluated using various performance metrics. The simulation outcomes clearly demonstrate that the MARIN-EFA method outperforms state-of-the-art techniques currently available in RS image classification. The proposed approach offers a promising solution for addressing the challenges associated with RS image classification, showcasing its potential to advance this field and enable more accurate and reliable categorization of RS data.

Future work in remote sensing (RS) image classification can focus on several areas to enhance the accuracy and efficiency of classification algorithms. Investigate the fusion of data from multiple sensors (e.g., optical, radar, and LiDAR) and different temporal acquisitions to leverage complementary information for classification. Fusion

techniques can enhance classification accuracy and provide more comprehensive and robust land cover or land use mapping.

Funding: This study received no specific financial support.

Institutional Review Board Statement: Not applicable.

Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.

Competing Interests: The authors declare that they have no competing interests.

Authors' Contributions: Problem formulation, writing, software, quantitative and qualitative analysis, A.J; verification validation, data curation, review, D.G. Both authors have read and agreed to the published version of the manuscript.

REFERENCES

- [1] Z. Zhang, R. Jiang, S. Mei, S. Zhang, and Y. Zhang, "Rotation-invariant feature learning for object detection in VHR optical remote sensing images by double-net," *IEEE Access*, vol. 8, pp. 20818-20827, 2019. <https://doi.org/10.1109/ACCESS.2019.2960931>
- [2] L. Dai, G. Liu, L. Huang, G. Xiao, Z. Xu, and J. Ruan, "Feature transfer method for infrared and visible image fusion via fuzzy lifting scheme," *Infrared Physics & Technology*, vol. 114, p. 103621, 2021. <https://doi.org/10.1016/j.infrared.2020.103621>
- [3] P. Du *et al.*, "Advances of four machine learning methods for spatial data handling: A review," *Journal of Geovisualization and Spatial Analysis*, vol. 4, pp. 1-25, 2020. <https://doi.org/10.1007/s41651-020-00048-5>
- [4] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *International Journal of Remote Sensing*, vol. 39, no. 9, pp. 2784-2817, 2018. <https://doi.org/10.1080/01431161.2018.1433343>.
- [5] C. Hakkenberg, M. Dannenberg, C. Song, and K. Ensor, "Characterizing multi-decadal, annual land cover change dynamics in Houston, TX based on automated classification of Landsat imagery," *International Journal of Remote Sensing*, vol. 40, no. 2, pp. 693-718, 2019. <https://doi.org/10.1080/01431161.2018.1516318>
- [6] L. Zhong, L. Hu, and H. Zhou, "Deep learning based multi-temporal crop classification," *Remote Sensing of Environment*, vol. 221, pp. 430-443, 2019. <https://doi.org/10.1016/j.rse.2018.11.032>
- [7] Z. Lv, T. Liu, and J. A. Benediktsson, "Object-oriented key point vector distance for binary land cover change detection using VHR remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 9, pp. 6524-6533, 2020.
- [8] A. Hagag, M. Amin, and F. E. Abd El-Samie, "Multispectral image compression with band ordering and wavelet transforms," *Signal, Image and Video Processing*, vol. 9, no. 4, pp. 769-778, 2015. <https://doi.org/10.1007/s11760-013-0516-4>
- [9] A. P. Byju, G. Sumbul, B. Demir, and L. Bruzzone, "Remote-sensing image scene classification with deep neural networks in JPEG 2000 compressed domain," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 4, pp. 3458-3472, 2020. <https://doi.org/10.1109/TGRS.2020.3007523>
- [10] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. Paragios, "Building detection in very high resolution multispectral data with deep learning features," presented at the In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, 2015.
- [11] Z. Xu, G. Liu, G. Xiao, L. Tang, and Y. Li, "JCa2Co: A joint cascade convolution coding network based on fuzzy regional characteristics for infrared and visible image fusion," *IET Computer Vision*, vol. 15, no. 7, pp. 487-500, 2021. <https://doi.org/10.1049/cvi2.12046>
- [12] K. Nogueira, O. A. Penatti, and J. A. Dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognition*, vol. 61, pp. 539-556, 2017. <https://doi.org/10.1016/j.patcog.2016.07.001>

- [13] Q. Yuan *et al.*, "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sensing of Environment*, vol. 241, p. 111716, 2020. <https://doi.org/10.1016/j.rse.2020.111716>.
- [14] S. Liu and Q. Shi, "Local climate zone mapping as remote sensing scene classification using deep learning: A case study of metropolitan China," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 164, pp. 229-242, 2020. <https://doi.org/10.1016/j.isprsjprs.2020.04.008>
- [15] Y. Wang, Z. Li, C. Zeng, G.-S. Xia, and H. Shen, "An urban water extraction method combining deep learning and Google Earth engine," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 769-782, 2020.
- [16] P. Dou, H. Shen, Z. Li, X. Guan, and W. Huang, "Remote sensing image classification using deep-shallow learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3070-3083, 2021.
- [17] H. Wang, X. Zhao, X. Zhang, D. Wu, and X. Du, "Long time series land cover classification in China from 1982 to 2015 based on Bi-LSTM deep learning," *Remote Sensing*, vol. 11, no. 14, p. 1639, 2019. <https://doi.org/10.3390/rs11141639>
- [18] S. Pan *et al.*, "Land-cover classification of multispectral LiDAR data using CNN with optimized hyper-parameters," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 241-254, 2020. <https://doi.org/10.1016/j.isprsjprs.2020.05.022>
- [19] C. Kwan *et al.*, "Deep learning for land cover classification using only a few bands," *Remote Sensing*, vol. 12, no. 12, p. 2000, 2020. <https://doi.org/10.3390/rs12122000>
- [20] C. Zhang, P. Yue, D. Tapete, B. Shangguan, M. Wang, and Z. Wu, "A multi-level context-guided classification method with object-based convolutional neural network for land cover classification using very high resolution remote sensing images," *International Journal of Applied Earth Observation and Geoinformation*, vol. 88, p. 102086, 2020. <https://doi.org/10.1016/j.jag.2020.102086>
- [21] G. B. Rajendran, U. M. Kumarasamy, C. Zarro, P. B. Divakarachari, and S. L. Ullo, "Land-use and land-cover classification using a human group-based particle swarm optimization algorithm with an LSTM Classifier on hybrid pre-processing remote-sensing images," *Remote Sensing*, vol. 12, no. 24, p. 4135, 2020. <https://doi.org/10.3390/rs12244135>
- [22] S. Pu, Y. Wu, X. Sun, and X. Sun, "Hyperspectral image classification with localized graph convolutional filtering," *Remote Sensing*, vol. 13, no. 3, p. 526, 2021. <https://doi.org/10.3390/rs13030526>
- [23] P. Liang, W. Shi, and X. Zhang, "Remote sensing image classification based on stacked denoising autoencoder," *Remote Sensing*, vol. 10, no. 1, p. 16, 2017. <https://doi.org/10.3390/rs10010016>

Views and opinions expressed in this article are the views and opinions of the author(s), Review of Computer Engineering Research shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.