Mathematics

AIMS Mathematics, 11(3): 7945–7979.
DOI: 10.3934/math.2026328
Received: 14 December 2025
Revised: 05 March 2026
Accepted: 11 March 2026
Published: 25 March 2026

https://www.aimspress.com/journal/Math

Research article

S^2A-DL-EGBNet: Land Cover classification using an Attention-enabled Distributed Learning-based Encoder Generative Bidirectional Network

Afnan M. Alhassan* and Nouf I. Altmami

Department of Computer Science, College of Computing and Information Technology, Shaqra University, Shaqra 11961, Saudi Arabia

* **Correspondence:** Email: aalhassan@su.edu.sa.

Abstract: Land Cover (LC) classification is an effective technique that categorizes the Earth's surface into urban, forest, and agricultural land classes by utilizing remote sensing data. Nevertheless, the hyperspectral remote sensing images are afflicted with a lack of labeled data, spectral variability, and the curse of dimensionality, which limit their competence in remote sensing applications. Hence, this research presents a Spatial Split Attention-enabled Distributed Learning-based Encoder Generative Bidirectional Network (S^2A-DL-EGBNet) model for accurate LC classification using hyperspectral images (HSIs). The model integrates a distributed learning module to process large datasets, and training is done in parallel by reducing complexity while improving scalability. Also, a Generative Adversarial Network (GAN)-based data balancing is employed to address the class imbalance problem and enables the model's effectiveness. Thereafter, the parallel Bidirectional Long Short-Term Memory (BiLSTM) is integrated to speed up the training process and minimize computation time. The research applies multiple feature extraction techniques, capturing complex features, scaling variations in photographic distortions, and illumination changes from the aspects of input data. Notably, the S^2A-DL-EGBNet model is validated using the Hyperspectral Remote Sensing Scenes dataset, and the S^2A-DL-EGBNet model shows remarkable performance by achieving 98.44% sensitivity, 98.84% accuracy, and 99.24% sensitivity on the Indian Pines dataset for training data of 90%.

Keywords: remote sensing; hyperspectral images; deep learning; Land Cover classification; attention mechanism
Mathematics Subject Classification: 68T07, 68U10, 68T05, 94A08

1. Introduction

The method of fine-grained Land Cover (LC) classification determines how the surface of the Earth is covered and utilized. However, classifying LC is important in disaster management, urban planning, and climate change research [1]. Evaluating LC is mandatory for upholding, monitoring, and planning the utilization of natural resources. Likewise, classifying LC has a direct influence on soil erosion, water, and atmosphere, which is indirectly associated with the issues of the global landscape [2,3]. Various LC classifications are done accurately by categorizing and defining factors such as vegetation types, soil types, crops, and urban areas. Importantly, remote sensing (RS) technologies are utilized to provide meaningful information about land usage patterns, ecosystem dynamics, and environmental changes [4,5]. Modern precision agriculture technology is widely emerging, and RS plays a major role in transforming such strategies [6,7]. In LC analysis, RS plays an essential part, because it allows a detailed ecosystem mapping, natural resources, agricultural, and urban development [8–10]. The volume of information and satellite imagery about the surface of the Earth is issued by modern RS. In LC classification, the accuracy inspires the decision-makers and scholars to make informed decisions on preservation initiatives, decline in biodiversity, and afforestation schemes in forest regions [11]. High-Resolution Remote Sensing (HRRS) determines the images with a spatial resolution of more than 10 m, and they are obtained by aerial platforms or satellites. The classification of LC depending on HRRS images provides support for variation detection, environmental monitoring, and LC mapping [12].

The classification and extraction of LC data are important in land surveys because they offer greater significance [13]. However, integrating RS to perform LC classification provides important benefits across sectors such as urban planning, environmental protection, land resource management, and agriculture [14]. With the development of RS strategies, multimodal images with high resolution, including multispectral imagery (MSI), synthetic aperture radar (SAR), and digital surface models (DSM), capture the detailed information about the surface of the Earth [1]. In the agriculture field, HSI has been widely used to monitor the growth of crops, detect pests, and evaluate soil nutrients. The main principle of HSI is to exploit features of electromagnetic wave radiation with different bands to acquire the spectral information [10,15]. In RS-based agricultural analysis, Vegetation Indices (VIs) act as a basic component, because they provide interpretable, scalable, and rapid insights into biomass, canopy dynamics, and vegetation health [7,16]. To monitor the LC deviations, the parameters such as the normalized difference built-up index (NDBI), land surface temperature (LST), and normalized difference vegetation index (NDVI) are utilized [17]. However, in RS images, the complex features, illumination changes, scale variations, and variability in photographic distortions lead to a complex process in applying traditional models to classify images effectively [14,18].

Mostly, the traditional models rely on surface attributes such as hue, form, texture, and pixel dimension for image classification. However, machine learning (ML) approaches, such as K-nearest neighbors (KNNs), K-means clustering, random forests (RFs), and support vector machines (SVMs), are utilized for RS imagery classification. Typically, these algorithms are associated with some challenges. Hence, deep learning (DL) techniques are used in addressing the spectral, intricate, and spatial characteristics present in the hyperspectral data [5,19]. The issues associated with RS image analysis are overcome by introducing tools like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) [11]. Nonetheless, issues like a lack of an extensive dataset and inadequate transferability arise when applying deep network models for the task of LC classification [14]. The segmentation anything model (SAM) provides generalization capabilities and robust transfer by enabling semantic segmentation and effective instance across several downstream applications [12].

The SAM focuses on segmentation with prompts for users, and its output has limitations on object category information, which makes it rigorous during the generation of classification results [1]. However, the CNN faces challenges like overfitting and high computational complexity with the presence of an imbalanced dataset, which decreases the model performance [2,20].

To address the earlier drawbacks, this research introduces an S^2A-DL-EGBNet model for accurate LC classification. Initially, the methodology processes hyperspectral remote sensing images using standard datasets to improve the adaptability. The obtained input is further applied to the feature extraction model, where the informative regions of the image are extracted. Thereafter, all the features are concatenated and passed into the proposed framework, where the model is trained effectively for accurate LC classification.

Spatial Split Attention enabled Distributed Learning-based Encoder Generative Bidirectional Network: Parallel training of large-scale data can be done by utilizing a distributed learning framework in the S^2A-DL-EGBNet model, which reduces the training complexity and enhances generalizability. However, the stability of the network and the strength of the model are improved by exploiting the hybrid attention mechanisms in the model. Moreover, the model incorporates an encoder architecture, which minimizes the dimensionality of input data by preserving important information that helps the model to learn efficient and robust representations.

The organization of this research is described as follows. The overview of the related works with their limitations is presented in Section 2, whereas Section 3 depicts the system model with a neat diagram, Section 4 provides the detailed model architecture with methodology, Section 5 explains the proposed specifications and experimental setup with results, and Section 6 provides the conclusion by summarizing the future research of the work.

2. Literature review

The methodologies utilized, advantages, and disadvantages of the related articles are illustrated below. Reddy et al. [5] developed a 3D CNN with a self-attention technique for hyperspectral image classification. Typically, the model improves its adaptability and accuracy by incorporating a self-attention mechanism. However, the attention technique in the model leads to improved classification accuracy and provides insights into an informed decision-making process. The primary drawback encountered by the model is that it lacks data fusion collected from different sensors. Li et al. [10] explained an algorithm for LC classification that incorporates a random forest and spectral transformer network structure (RS-Net). The method combines different approaches to map the multidimensional time series for image features, enabling recognition and extraction to improve the accuracy of classification. The relationship between the features was captured with various scales, and the recognition ability of the LC features was improved significantly. The model handles high-dimensional data effectively to achieve high efficacy and performance.

Zhang et al. [1] suggested the FlexiSAM method based on semantic segmentation of multimodal data for LC classification. Here, the model handles the variations that occur in RS imagery, and it treats RGB as the primary modality. However, the model significantly enhances the robustness and classification accuracy by outperforming conventional methods. The model consumes high resources, which is considered the main challenge. Yang et al. [4] developed a multi-level feature adaptation-segment anything model (MLFA-SAM) that exploits a fine-tuning strategy for remote sensing classification. The model utilizes different levels to separate the domains of spectral and spatial data from the input images, and the training parameters of the model were reduced by the SAM. However, the model was validated on multiple datasets, but it has issues with training data characteristics and model architecture.

Li et al. [10] explained the DASR-Net algorithm that integrates the Residual Network (ResNet) with a dual attention network (DAN) in parallel to enhance the classification of LC. The model utilizes a dual attention technique to capture multiscale semantic information, and it enhances classification capabilities with feature extraction for hyperspectral data. The model exhibits a parallel architecture to prevent loss of information and recognize the types of features with high similarity during the classification process. Wang et al. [13] developed a triplet attention mechanism named border-enhanced triple attention mechanism (BETAM) for the classification of LC. The model is exploited with attention techniques that encode contextual features to enrich the representation of local features. Typically, the model is applied in agricultural fields for disease and crop monitoring to identify the crop disease areas effectively. The model offers benefits like better segmentation accuracy and robustness, but it limits its applicability in datasets.

Fayaz et al. [14] suggested Inception-v3 and DenseNet networks for reliable and efficient LC classification. The model was fine-tuned with datasets to enhance the adaptability and reliability strategies. Here, optimized networks play an important role in capturing high-level features in LC classification. The model fails to utilize data fusion technology, and the classification accuracy of the model was low. Aljebreen et al. [3] introduced a technique named the river formation dynamics algorithm with deep learning (LULCC-RFDADL) for LC classification. However, feature extraction is done by incorporating the Dense EfficientNet algorithm, which takes complex information and patterns from high-resolution images. Different applications are utilized by the model in several areas, such as urban planning, environmental monitoring, and natural resource management. Typically, the model exhibits high performance, but it has poor generalization capabilities across different weather conditions.

2.1. Challenges

The limitations and challenges faced by the related works are listed below:

- ➢ The 3D-CNN method with an attention mechanism [1] issued valuable insights on informed decision-making. The limitation of this approach is that it lacks data fusion from different sensors.

- ➢ The RS-Net model [2] managed high-dimensional data when utilizing different algorithms, but the model suffered from high computational cost and time.

- ➢ The MLFA-SAM approach [4] attained better generalizability than related approaches, thereby supporting real-time implication. Nevertheless, the model displayed less sensitivity in extracting small targets.

- ➢ The DASR-Net model [5] can differentiate between complex structures in the land, like understory vegetation and tree species. However, the major drawbacks present here are dataset complexities, which lead to poor robustness.

- ➢ The RFDADL-SOA method [8] achieved prominent performance based on simulation results. Nevertheless, their complexities with the dataset and computation overhead posed challenges.

2.2. Problem definition

In remote sensing, LC classification is the task of classifying the objects or pixels that provide important information for environmental monitoring, land management, urban planning, and disaster response. However, it mainly categorizes the attributes of Earth's surface, such as water and vegetation, and classifies them into distinct classes like agriculture, forest, or urban areas. Typically, the existing methods in LC classification consist of several limitations, such as a lack of computational resources, limited empirical validation, noise sensitivity, low computational efficiency, and complexity of model architecture. To address these issues, this research introduced a method named the S^{2}A-DL-EGBNet model for accurate LC classification. The proposed model incorporates the Pavia Centre Scene, Salinas, and Indian Pines datasets, which are gathered through the Hyperspectral Remote Sensing Scenes dataset [21]. Initially, the hyperspectral remote sensing images are collected from the dataset F, which is mathematically expressed as

$$F = \{R_s, W_s\} \tag{1}$$

where R_s denotes the hyperspectral remote sensing image, and here $R_s \in \Re^{J \times O \times E}$ denotes height, width, and band of the image, respectively, and W_s indicates the categorical pixel-wise mask image for the corresponding hyperspectral sensing image. Here, the image R_s has N a number of pixels, and it is obtained by multiplying the width and height of the image, that is, $N = J \times O$ pixels. Further, each pixel present in the hyperspectral image R_s is mathematically expressed as

$$R_s = \{U_i\}_{i=1}^{N} \tag{2}$$

where N indicates the number of pixels in an image, U_i indicates the i^{th} pixel present in the hyperspectral image, and here $U_i = \{R_s(u,v)\}$ in which (u,v) implies the position of the pixel in an image. Likewise, the mask image is represented as $W_s = \{d_i\}_{i=1}^{N}$, where d_i denotes the label value corresponding to pixel U_i in image R_s. For the Indian Pines dataset, the label values are represented as $W_s(u,v) \in (0,1,...,15)$, and for the Pavia Centre Scene dataset, label values are expressed as $W_s(u,v) \in (0,1,...,8)$. Similarly, the label values for the Salinas dataset are denoted as $W_s(u,v) \in (0,1,...,15)$. After pre-processing the image and extracting the features, the collected pixels along with the feature vector can be fed to the proposed model that classifies the LC. Thereafter, the model's error rate is specified through the Categorical Cross-Entropy (CCE) loss function, which is mathematically denoted as

$$loss_{CCE} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{g} d_{ij} \log\left(d'_{ij}\right) \tag{3}$$

where g indicates the maximum number of classes, d_{ij} represents actual labels present in the j^{th} class of the i^{th} sample, and d'_{ij} indicates the predicted probability for the j^{th} class of the i^{th} sample.

3. System model for Land Cover classification

LC classification is a geospatial technique that classifies physical attributes of Earth's surface, such as water, vegetation, and soil, into distinct classes. However, this technique often uses remote sensing data captured from satellites to create maps for applications like urban planning, disaster management, and environmental monitoring. In this research, the hyperspectral image is collected as input data, and further feature extraction is done to extract important features, where the extracted features are applied to the proposed model. Here, the model is trained well to perform LC classification, and its performance is analyzed by utilizing various evaluation metrics. Figure 1 depicts the system model of the proposed framework.

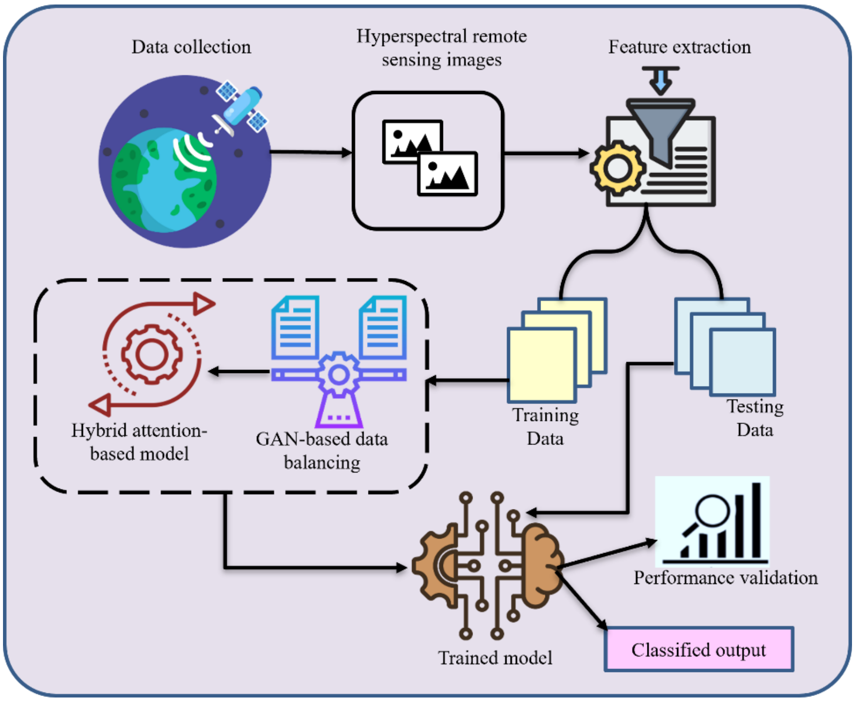


Figure 1. System model for LC classification.

4. Land Cover classification using Spatial Split Attention-enabled Distributed Learning-based Encoder Generative Bidirectional Network.

Remote sensing technology plays an important role because it is widely utilized for LC classification to classify the attributes of Earth's surface by providing certain information based on applications in urban planning, environmental protection, and resource management. Here, the main objective of this research is to introduce a method named the S^2A-DL-EGBNet model to perform LC classification. At first, the Hyperspectral Remote Sensing Scenes dataset [21] is utilized in the proposed model to capture the hyperspectral image. Thereafter, feature extraction takes place for the obtained image data, where the features include vegetative index features, deep statistical vegetative features, resnet151, and hybrid structural pattern features. Specifically, the distributed learning utilizes a data parallelism that generates a copy of the model on multiple devices, which improves the learning of complex features. The given data is parallelly trained over multiple devices for accurate LC classification, which leads to better scalability while reducing complexity. The proposed S^2A-DL-EGBNet model incorporates a Bidirectional Long Short-Term Memory (BiLSTM) network for capturing the patterns present in hyperspectral images and improving the classification outcomes. The utilization of multiple attention techniques enhances computational efficiency and performance of the proposed model. Finally, model training takes place, and it is applied with test data features to perform LC classification. The block diagram of the S^2A-DL-EGBNet model is shown in Figure 2.

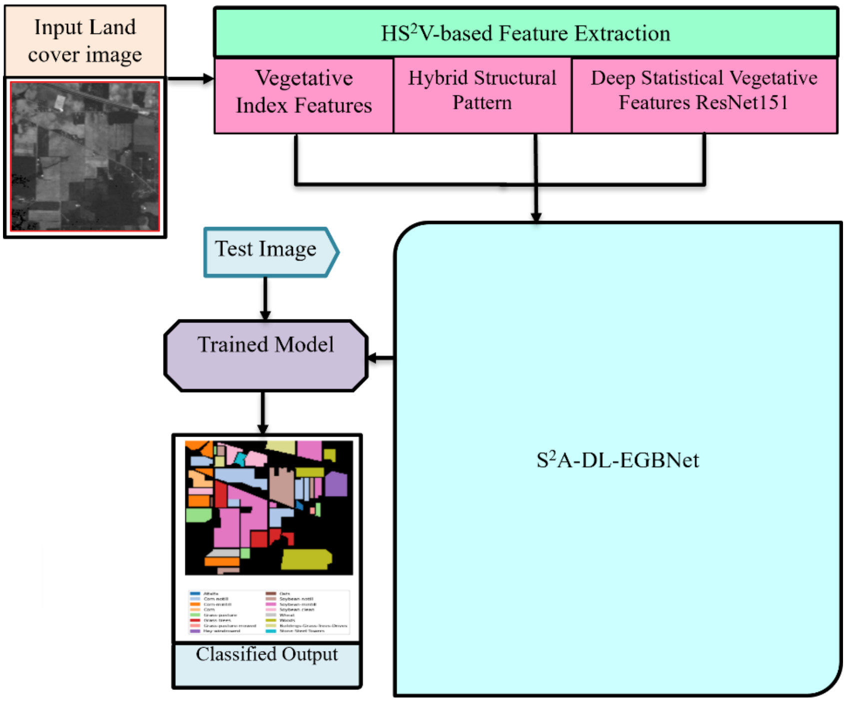


Figure 2. Structure of the S^2A-DL-EGBNet model for LC classification.

4.1. Input data collection

At first, the input hyperspectral images are obtained from the Pavia Centre Scene, Salinas, and Indian Pines datasets, which are collectively taken from the Hyperspectral Remote Sensing Scenes dataset [21]. The dataset utilized to gather the hyperspectral images is mathematically represented as F. Here, the input image shape acquired from the Indian Pines dataset is denoted as R_s, which has 200 spectral bands and 145×145 pixels, and is represented as $[145 \times 145 \times 200]$. Likewise, the input shape of the hyperspectral image collected from the Pavia Centre Scene dataset is denoted as $\hat{R}$, which has the dimensions $[1096 \times 715 \times 102]$ of 102 spectral bands and $[1096 \times 715]$ pixels. Similarly, the Salinas dataset contains 204 spectral bands with $[512 \times 217]$ pixels, and it is denoted as R' a matrix with dimensions $[512 \times 217 \times 204]$. To classify the LC, the R_s input captured from the dataset is applied to the feature extraction process and further fed into the model to get the classified outcome.

4.2. Feature extraction using hybrid structural statistical vegetative features

Feature extraction plays an essential role in identifying and extracting the key characteristics, such as textures, patterns, and spatial information, from the hyperspectral images. Here, the HS^{2}V features are obtained by integrating vegetative index features, hybrid structural pattern features, and deep statistical vegetative features of ResNet 151. However, higher model accuracy is achieved by employing these features, which provides better performance. Moreover, drawbacks like computational complexity are reduced, and the model enhances the interpretability and generalization by transforming the data into meaningful features.

4.2.1. Vegetative index features

Vegetative index features are the combinations of spectral bands obtained from the satellite images that emphasize the specific characteristics of vegetation. These features are crucial for tasks such as yield prediction, crop monitoring, and environmental change detection. The vegetation indices are single values that are calculated from different spectral bands, and they are mainly designed to enhance the contrast of vegetation. Typically, to extract the information of vegetation from two or more spectral bands, these vegetation indices are utilized [22]. At first, a hyperspectral input image R_s is passed as input to all the vegetation features separately. Here, the vegetative features such as enhanced vegetation index (EVI), normalized difference vegetation index (NDVI), global environmental monitoring index (GEMI), soil adjusted vegetation index (SAVI), green normalized difference vegetation index (GNDVI), difference vegetation index (DVI), leaf area index (LAI), visible atmospherically resistant index (VARI), near-infrared (NIR), atmospherically resistant vegetation index (ARVI), and green atmospherically resistant index (GARI) are extracted. All these features are considered powerful tools in remote sensing for improving agricultural practices and managing natural resources. Table 1 summarizes the vegetative index features.

Table 1. Vegetative index features.

Vegetative Index Features	Description	Formula	Dimension
Normalized Difference Vegetation Index	The NDVI is a remote sensing measurement that analyzes the health of vegetation by taking the difference between the visible and near-infrared light it reflects.	$A_1 = \frac{S-\hat{S}}{S+\hat{S}}$ where S denotes the near-infrared reflectance, $\hat{S}$ indicates the red band, and A_1 denotes the NDVI.	$[145 \times 145 \times 1]$
Soil Adjusted Vegetation Index	The SAVI is developed to minimize the brightness of soil influences in spectral vegetation indices.	$A_2 = \frac{S-\hat{S}}{S+\hat{S}+y}(1+y)$ where L indicates the soil adjustment factor, and A_2 denotes the SAVI.	$[145 \times 145 \times 1]$
Green Normalized Difference Vegetation Index	This measure is used to calculate the green and near-infrared bands of an image, which estimates the health of a plant.	$A_3 = \frac{S-K}{S+K}$ where K the green band is, and A_3 represents the GNDVI.	$[145 \times 145 \times 1]$

Continued on next page

Vegetative Index Features	Description	Formula	Dimension
Difference Vegetation Index	This computes the difference between the red band from near-infrared reflectance.	$A_4 = S - \hat{S}$ where A_4 indicates the DVI.	$\left[145\times145\times1\right]$
Enhanced Vegetation Index	This measure aims to optimize the signal of vegetation by improving the sensitivity in biomass regions, and vegetation monitoring is improved through the decoupling of the canopy background signal.	$A_5 = l \times \dfrac{S - \hat{S}}{S + I_1\hat{S} - I_2 S_i^* + L}$ where l denotes a gain factor, I_1, I_2 represents the aerosol resistance coefficients, S_i^* denotes the blue band, and A_5 is the EVI.	$\left[145\times145\times1\right]$
Leaf Area Index	It describes the total amount of area of the leaf per unit of ground surface area.	$A_6 = \dfrac{\log\left(\dfrac{1}{1-S}\right)}{\mathbb{Q}}$ where $\mathbb{Q}$ denotes the coefficient of light extinction, and A_6 denotes the LAI.	$\left[145\times145\times1\right]$
Visible Atmospherically Resistant Index	This measures the health of vegetation by mitigating the atmospheric effects and illumination differences.	$A_7 = \dfrac{\left(K - \hat{S}\right)}{K + \hat{S} - S_i^*}$ Where A_7 is the VARI.	$\left[145\times145\times1\right]$
Near-Infrared	This denotes the electromagnetic spectrum's region that ranges from 780 nm to 2500 nm.	$A_8 = S$ where A_8 denotes the NIR.	$\left[145\times145\times1\right]$
Atmospherically Resistant Vegetation Index	This measure is utilized to correct the influence of the atmosphere, and it is useful in regions with atmospheric aerosol, such as tropical areas.	$A_9 = \dfrac{\left(S - \hat{S}S_i^*\right)}{\left(S + \hat{S}S_i^*\right)}$ where $\hat{S}S_i^* = \hat{S} - \rho\left(S_i^* - \hat{S}\right)$, $\rho = 1$, unless the aerosol is a priori, and A_9 is the ARVI.	$\left[145\times145\times1\right]$
Global Environmental Monitoring Index	This is a nonlinear vegetation index, and it is less sensitive to atmospheric effects.	$A_{10} = \eta\left(1 - 0.25\eta\right) - \dfrac{\hat{S} - 0.125}{1 - \hat{S}}$ where η denotes an intermediate variable, and A_{10} is the GEMI.	$\left[145\times145\times1\right]$
Green Atmospherically Resistant Index	Generally, this index is less sensitive to atmospheric effects and more sensitive to chlorophyll concentrations.	$A_{11} = \dfrac{S - \left(K - \left(S_i^* - \hat{S}\right)\right)}{S - \left(K + \left(S_i^* - \hat{S}\right)\right)}$ where A_{11} indicates the GARI.	$\left[145\times145\times1\right]$

Finally, the concatenated features of the vegetative index are represented as A with the magnitude of $[145\times145\times11]$,

$$A=\left\{A_1\left\|A_2\right\|A_3\left\|A_4\right\|A_5\left\|A_6\right\|A_7\left\|A_8\right\|A_9\left\|A_{10}\right\|A_{11}\right\}. \tag{4}$$

4.2.2. Hybrid structural pattern features

The hybrid structural pattern features are utilized to extract complex features, whereas they improve computational efficiency, enhance model performance, and encompass faster execution of the model by capturing multi-scale information. Initially, the hyperspectral image R_s with dimensions $[145\times145\times200]$ is split into three bands, such as red, green, and blue. Here, the dimensions obtained for three bands are represented as $[145\times145\times3]$. Further, the three bands are resized into the magnitude of $[224\times224\times3]$. Thereafter, the resized information is forwarded into the ResNet-101 model as input, where the output of the second layer in the ResNet-101 model has 64 channels, and the obtained magnitude is denoted as $[112\times112\times64]$. From these 64 bands, the best band's output is considered, that is, the band with the highest visual representation is taken, and it obtains the magnitude $[112\times112\times1]$. Further, the selected band is resized into the dimension of $[145\times145\times1]$. Thereafter, the resized output is applied to grid-based structural pattern features. Here, the powerful texture descriptor Local Binary Pattern (LBP) is computed by comparing the gray level values of neighboring and central pixels. Generally, the LBP operator takes the central pixel of the window as a threshold. Here, the difference between the neighboring and central pixels is calculated, and some values are assigned to the pixels. The descriptor LBP is formulated as follows

$$LBP_{D,M}=\sum_{m}^{M-1}L\left(\mathbb{N}_m-\mathbb{N}_o\right)2^m, \tag{5}$$

$$L(m)=\begin{cases}1, & if\ m\geq 0\\ 0, & if\ m<0\end{cases} \tag{6}$$

where $\mathbb{N}_o$ indicates the central pixel of the image, $\mathbb{N}_m$ denotes the neighboring pixel, D implies the distance between $\mathbb{N}_m$, and $\mathbb{N}_0$ and M represents the radius. Here, rather than considering the traditional LBP, the grid-based structural pattern is exploited to divide images into a grid of overlapping regions. Typically, the hybrid structural features obtain the output H, which has the dimension $[145\times145\times1]$.

4.2.3. Deep statistical vegetative features ResNet 151

The deep statistical vegetative features Resnet 151 are mainly employed in feature extraction to achieve robust feature representation that leads to high accuracy. Moreover, these features enhance the model performance by capturing informative features, and the ResNet 151 model extracts textural, spectral, and spatial features by improving the ability of generalization of the model. Initially, vegetative index features such as the transformed soil adjusted vegetation index (TSAVI), renormalized difference vegetation index (RDVI), and modified soil adjusted vegetation index (MSAVI) are taken as input with dimension $[145\times145\times3]$. The TSAVI index is used to assess the health of a plant by minimizing the interference caused by the color and brightness of the soil. It is expressed as

$$TSAVI = \frac{\left(w\left(S - w.\hat{S} - r\right)\right)}{\hat{S} + w.S - w.r} \tag{7}$$

where w, r represent the intercept and slope of the soil line, respectively, S denotes near-infrared reflectance, and $\hat{S}$ indicates the red band. Moreover, the RDVI index helps to monitor the growth and health of the vegetation by analyzing the reflectance of light from plants. It is mathematically denoted as

$$RDVI = \frac{S - \hat{S}}{\left(S + \hat{S}\right)^{0.5}}. \tag{8}$$

The index MSAVI assesses the health of vegetation in certain areas with exposed soil by reducing the effect of soil brightness on the vegetation signal, and it is expressed as

$$MSAVI = \frac{S - \hat{S}}{S + \hat{S} + L'}\left(1 + L'\right) \tag{9}$$

where L' denotes the primary soil line parameter. Thereafter, the obtained input dimension is resized into the magnitude $[224\times224\times3]$ and forwarded to the ResNet 151 model. Here, the ResNet 151 model tackles the vanishing gradient problem, which leads to faster training and stable performance. The acquired output from the ResNet 151 model is denoted as s, which has the dimension of $[112\times112\times64]$. Next, the output is forwarded to extract statistical information like mean, median, and standard deviation. Table 2 illustrates the statistical features.

Table 2. Statistical features.

Features	Description	Formula	Dimension
Mean	The metric mean defines the color value in the image.	$Y_1 = \frac{1}{\mathbb{Z}}\sum_{f=1}^{\mathbb{Z}} s_f$ where $\mathbb{Z}$ indicates the number of data samples, s_f is the input, and Y_1 denotes the mean.	$[112\times112\times1]$
Median	The median represents the central tendency, and it is calculated for spectral bands in images.	$Y_2 = \frac{\mathbb{Z}+1}{2}$ where Y_2 denotes the median.	$[112\times112\times1]$
Standard deviation	This defines the square root of the distribution variation.	$Y_3 = \sqrt{\frac{1}{\mathbb{Z}}\sum_{f=1}^{\mathbb{Z}}\left(s_f - Y_1\right)^2}$ where Y_3 is the standard deviation.	$[112\times112\times1]$

The concatenated statistical features are denoted as Y with the dimension $[112\times112\times3]$, which is expressed as

$$Y=\{Y_1\|Y_2\|Y_3\}. \tag{10}$$

Thereafter, the output Y is resized to obtain the dimension as $[145\times145\times3]$, which is represented as G. Overall, HS^{2}V features are obtained by performing concatenation, and they are mathematically formulated as

$$X=\{A\|H\|G\} \tag{11}$$

where X denotes the HS^{2}V features extracted for the Indian Pines dataset with the dimension $[145\times145\times15]$, which is then applied to the proposed model. Likewise, features extracted from the Pavia Centre Scene dataset are represented with the magnitude of $[1096\times715\times15]$. Further, the dimension obtained for the Salinas dataset after extracting features is denoted as $[512\times217\times15]$.

4.3. Spatial Split Attention-enabled Distributed Learning-based Encoder Generative Bidirectional Network.

In this research, the S^{2}A-DL-EGBNet model is exploited accurately for LC classification, and it integrates an effective distributed learning technique with a hybrid attention mechanism. Nevertheless, the attention mechanisms are employed to capture the interdependence and long-range dependencies between channels from the perspective of the channel and space. Typically, the exploitation of distributed learning can minimize the overall training time of the model. Moreover, the model's accuracy and stability are enhanced by utilizing the learning mechanism, which is useful for large datasets and complex models. Moreover, the model incorporates an encoder architecture that automatically learns to capture important patterns from the input features. Further, a data balancing technique based on a GAN is utilized to balance the datasets, and it enhances the reliability of the model. The detailed architecture of the S^{2}A-DL-EGBNet model is illustrated in Figure 3.

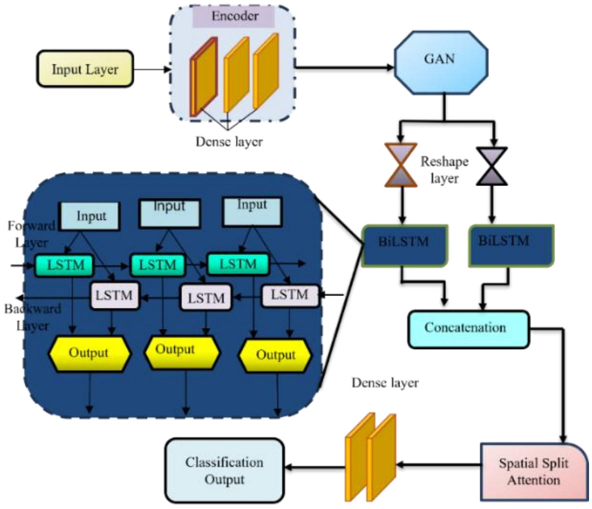


Figure 3. S^{2}A-DL-EGBNet model's detailed architecture.

4.3.1. Encoder

The proposed S^{2}A-DL-EGBNet model contains an encoder network, which takes the feature extracted output X as input data with a dimension $[N\times 15]$. The encoder network contains three dense layers, and the output is further passed into GAN-based data balancing, which is discussed below. Here, the encoder architecture is mainly utilized to understand the spatial location of features, and it explores meaningful features by capturing both frequency and spatial domains. At first, the encoder takes the X input and passes it into the dense layer. Thereafter, the dimension of the dense layer is $[N\times 64]$, which is forwarded again to the dense layer. Further, the data with dimension $[N\times 32]$ is obtained from the dense layer as output. Next, the outcome is again passed into the dense layer, which is denoted as q with the magnitude $[N\times 32]$, and it is further fed into data balancing.

4.3.2. GAN-based data balancing

The data balancing technique helps to tackle the problem of class imbalance, where one class consists of fewer samples than another. Generally, balancing the data for training increases the model's accuracy by improving the diversity and size of data, which helps to avoid the overfitting problems. Here, the GAN-based data balancing technique is used to balance the data q obtained from the encoder network. The GAN generates samples for the minority class, which efficiently makes the data samples more balanced.

At first, B is represented as a generator and Z is denoted as a discriminator, where both are network architectures of the GAN model. Generally, the discriminator is fooled by the generator, whereas the discriminator differentiates between fake and real samples that are generated by the generator [23]. Typically, the function of the generator B is generally expressed as $B(Q)=h$, where h indicates the generated samples of the generator, and Q denotes random noise. Mostly, the generator takes the random noise Q as input that consists of data points c, where Q is determined as a prior distribution. The architecture of the GAN in this research comprises various dense layers. The generator takes the random noise Q as input with dimension $[N\times 32]$ and applies it to a dense layer. Further, the dense layer's outcome with magnitude $[N\times 32]$ is again passed to the dense layer. Finally, the dense layer obtains the dimension as $[N\times 32]$, and the generator generates the samples, which are denoted as h. Likewise, the working of the discriminator is done by taking the fake samples h and the output from the encoder q as input data. Here, the input with $[N\times 32]$ dimension is forwarded into the dense layer, which obtains the output magnitude as $[N\times 32]$. Further, the output is again fed to a dense layer, and it obtains the dimension $[N\times 16]$. Thereafter, the output is again passed to a dense layer that classifies the data as fake or real with a dimension $[N\times 1]$, which is denoted as V.

The discriminator and generator function is mathematically formulated as

$$\min_{B}\max_{Z} V(B,Z)=\mathrm{E}_{q\sim P_{data(q)}}\left[\log Z(q)\right]+E_{Q\sim P_{c(Q)}}[\log(1-Z(B(Q)))] \tag{12}$$

where $P_{c(Q)}$ represents the generator distribution, $B(Q)$ denotes the generator function, $Z(Q)$ and

$Z(B(Q))$ indicates predictions for fake and real data samples, $P_{data}(q)$ describes samples of a real data distribution, and $P(Q)$ denotes the probability appearing from data of the original distribution. The above Eq (12) is known as the optimization function. Figure 4 shows the detailed working principle of GAN-based data balancing.

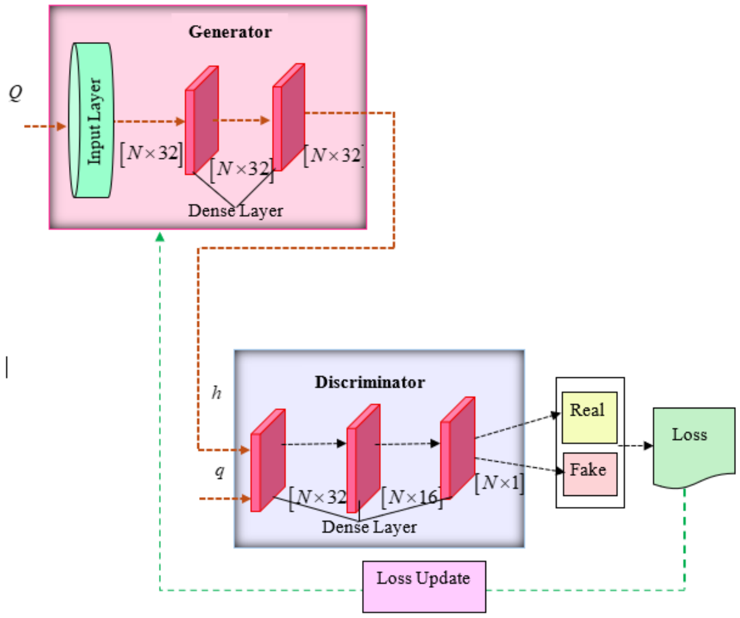


Figure 4. Structure of GAN-based data balancing.

Further, data balancing output obtained with the dimension $[N\times 32]$ is given as input to the BiLSTM blocks. First, the output is resized into the magnitude of $[N\times 32\times 1]$, and it is applied separately to the parallel BiLSTM layers. The BiLSTM layers are employed to represent the contextual information in a better way, and they are comprised of two LSTM networks [24]. Moreover, these networks have the ability to process the inputs in both forward and backward directions. The information is processed from left to right in the forward LSTM, where the hidden state is expressed as $\vec{z}_x = LSTM(n_x, \vec{z}_{x-1})$, and the backward LSTM processes the information from right to left, and its hidden state is represented as $\overleftarrow{z}_x = LSTM(n_x, \overleftarrow{z}_{x+1})$. Lastly, the BiLSTM output is obtained by concatenating the backward and forward states $z_x = [\vec{z}_x, \overleftarrow{z}_x]$. The parallel BiLSTM blocks obtain the output as k and n with dimension $[N\times 32\times 64]$, and they acquired the magnitude $[N\times 64\times 128]$ by performing concatenation, which is formulated as

$$T = \{k \| n\}. \tag{13}$$

Further, the concatenated output T obtains the magnitude $[N\times 64\times 128]$, which is applied to the hybrid attention mechanism.

4.3.3. Spatial Split Attention mechanism

The S^{2}A mechanism is obtained by the combination of spatial [25] and split [26,27] attention mechanisms. Integration of split and spatial attention mechanisms enhances the modeling capabilities and expressive power of the network. Moreover, the S^{2}A mechanism increases the robustness, interpretability, and accuracy of the model. The structure of the S^{2}A mechanism is visualized in Figure 5.

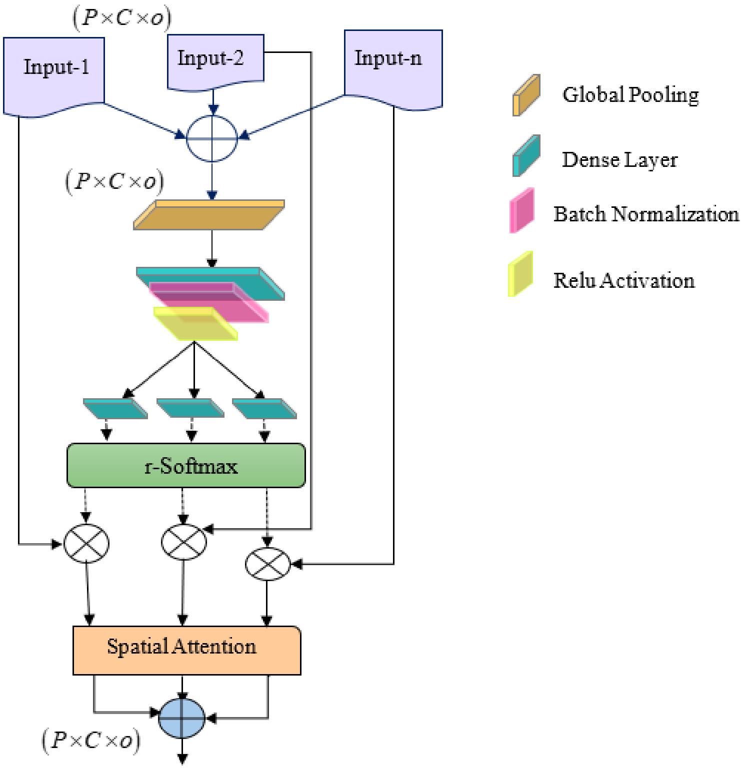


Figure 5. Working of the S^{2}A mechanism.

The Split attention block comprises a feature map, and it can be divided into various splits given by the radix hyperparameter $\Im$. The representation of the split is x_r, where $x_r \in \mathbb{C}^{P\times C\times o}$, and P, C, and o are the outputs from the block with the feature map. However, the contextual global information with embedded channel-wise statistics is collected with global average pooling between the spatial dimensions $\chi \in \mathbb{C}^{o}$. The formulation of the c th component is denoted as:

$$\chi^c = \frac{1}{P\times C}\sum_{a=1}^{P}\sum_{b=1}^{C}\hat{x}^c(a,b), \tag{14}$$

$$\hat{x}^c(a,b) = \sum_{t=1}^{\Im} x_t^c(a,b). \tag{15}$$

Here, the intermediate representation with weighted fusion $\beta \in \mathbb{C}^{P\times C\times o}$ is aggregated by utilizing channel-wise soft attention. Generally, the channel that contains feature maps is produced by employing a combination of weighted splits. Further, the c th channel is computed as:

$$\beta_c(a,b) = \sum_{t=1}^{\Im} \phi_t^c x_t^c \left(a,b\right) \tag{16}$$

where ϕ_t^c represents a soft weight factor given by:

$$\phi_t^c = \begin{cases} \dfrac{\exp\left(\xi_t^c\left(\varepsilon^c\right)\right)}{\sum_{b=1}^{\Im} \exp\left(\xi_b^c\left(\varepsilon^c\right)\right)}, & \Im > 1 \\ \dfrac{1}{1+\exp\left(-\xi_t^c\left(\varepsilon^c\right)\right)}, & \Im = 1 \end{cases}. \tag{17}$$

Here, ξ_t^c indicates each split weight for the cth channel based on ε^c global information. The output from the split attention module is further applied to the spatial attention mechanism.

Similarly, the spatial attention map is produced by employing features with spatial relationships. Mostly, the spatial attention mainly concentrates on the informative part, which is more complementary to the channel attention [28]. Moreover, the max-pooling and average pooling operations are concatenated for calculating the spatial attention along the channel axis, and they generate an efficient feature descriptor. Here, the informative regions are highlighted by applying some operations of pooling with the channel axis, and a map is generated by applying the convolutional layer on the feature descriptor $Z_s\left(p\right) \in \mathbb{R}^{P \times C}$, which encodes where to suppress. At first, pooling operations are used for aggregating the feature map's channel elements that include generating 2D maps $p_{avg}^s \in \mathbb{Q}^{1 \times C \times o}$ $p_{\max}^s \in \mathbb{Q}^{1 \times C \times o}$. The feature maps are concatenated using a convolutional layer by generating a 2D spatial attention map. It is formulated as

$$Z_s(p) = \sigma\left(f^{7\times 7}\left(\left[Avg(p); Max(p)\right]\right)\right) \tag{18}$$

where $f^{7\times 7}$ denotes a convolutional operation with 7×7 filter size, and σ denotes the sigmoid function. The magnitude obtained for the S^{2}A mechanism is $\left[N \times 64 \times 128\right]$, which is denoted as $\aleph$, and it is applied to the dense layer. Here, the dense layer output with dimension $\left[N \times 64\right]$ is forwarded again into the dense layer. At last, the dense layer classifies the output into 16 classes for the Indian Pines dataset, which is represented as d'_{ij}, with the dimension $\left[N \times 16\right]$, respectively.

5. Results and discussion

The section below explains the experimental setup, performance analysis, and dataset description of the S^{2}A-DL-EGBNet model with conventional methods in LC classification.

5.1. Experimental setup

The implementation of the S^{2}A-DL-EGBNet model is done through a Python 3.7 tool using the operating system of Windows 11 operating system with an Intel Core i7 processor, 16 GB of RAM, and 12 GB of GPU memory. The results are carried out by utilizing various libraries like Keras, TensorFlow, and PyTorch. Further, three datasets are used to evaluate the S^{2}A-DL-EGBNet performance. The hyperparameters of S^{2}A-DL-EGBNet are shown in Table 3.

Table 3. Hyperparameters of the S^{2}A-DL-EGBNet model.

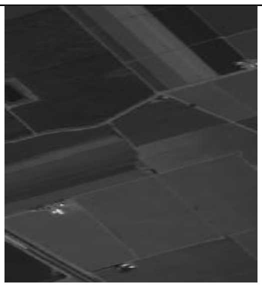
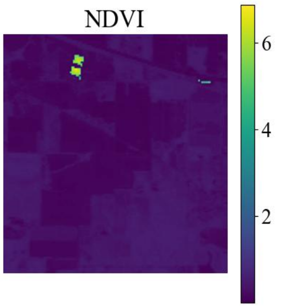
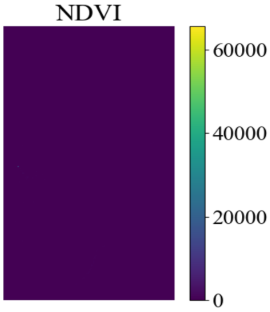
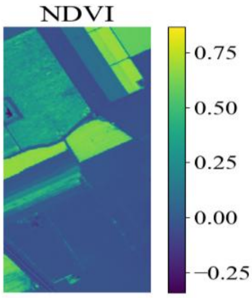
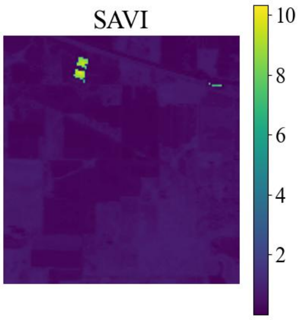
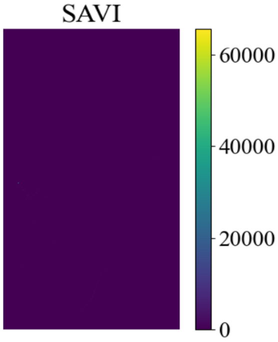
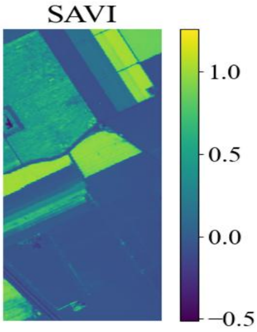
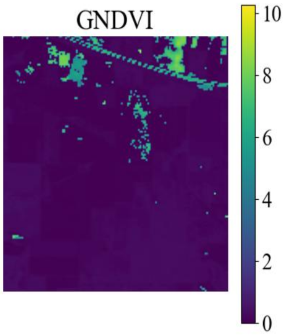
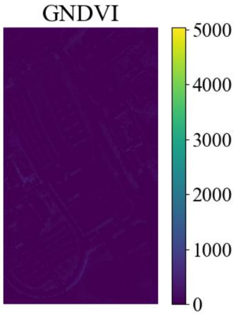
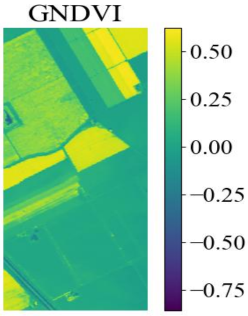
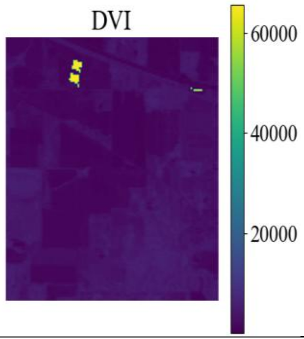
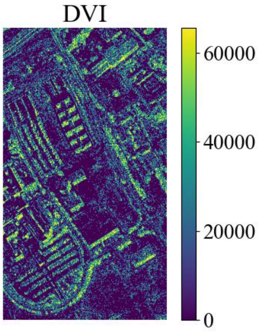
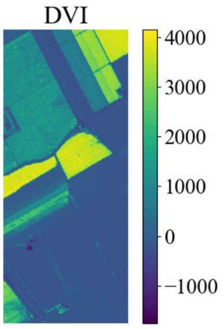
Hyperparameters	Values
Dropout rate	0.3
Learning rate of the GAN	0.0002
Number of epochs	200
Latent space size	100
Discriminator layers	3
Generator layers	3
Discriminator hidden units	256
Generator hidden units	256
Batch size	64
Weight decay	0.0001
BCE loss	True
Gradient penalty	True
Gradient penalty C	10
Learning rate of BiLSTM	0.001
LSTM hidden units	128
Default optimizer	Adam

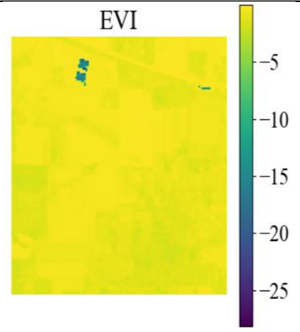
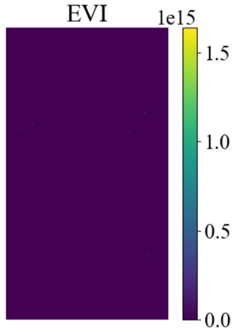
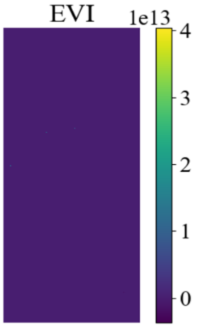
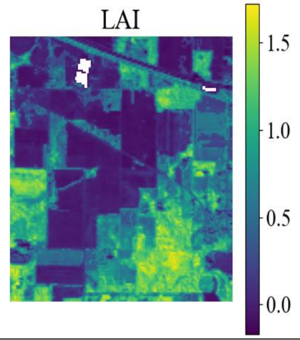
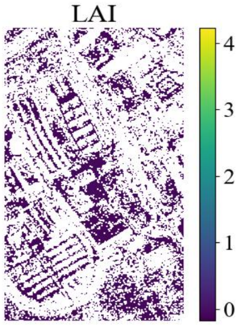
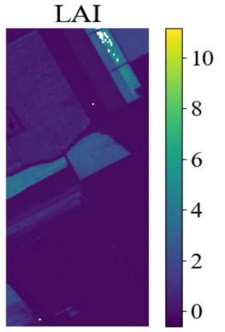
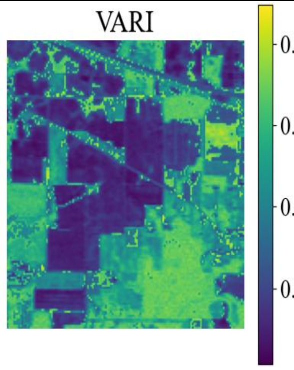
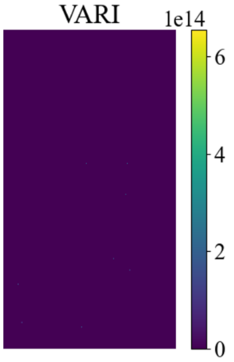
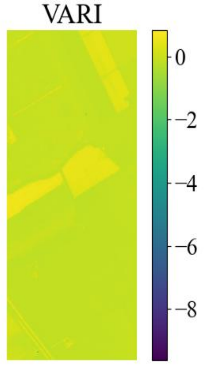
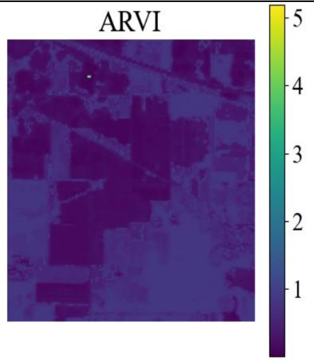
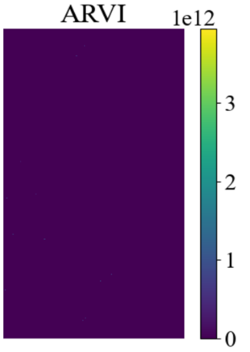
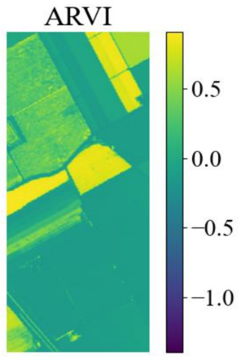
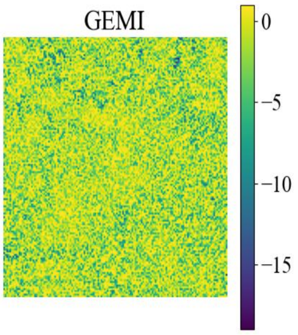
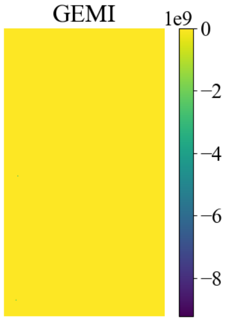
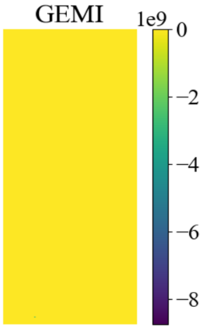
5.2. Dataset description

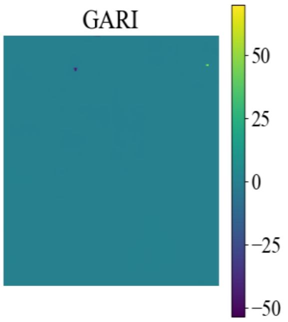
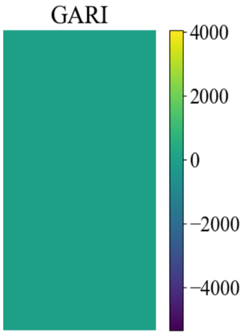
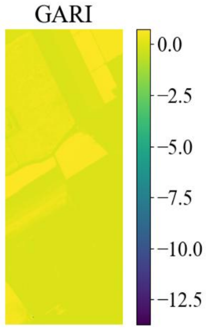
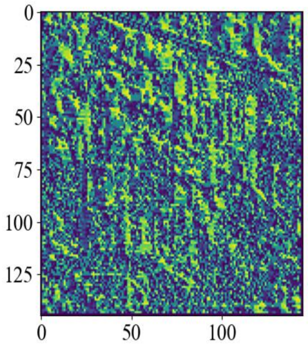
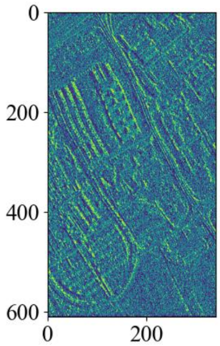
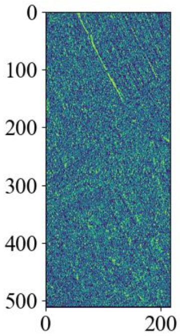
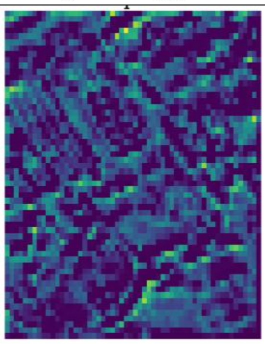
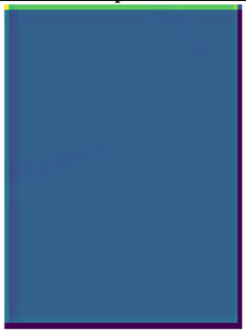
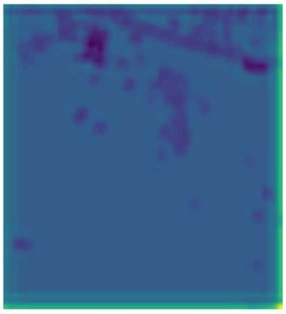
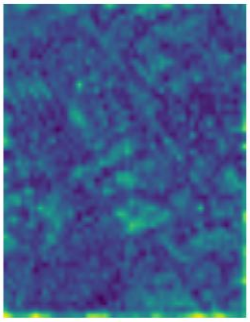
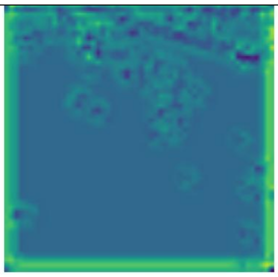
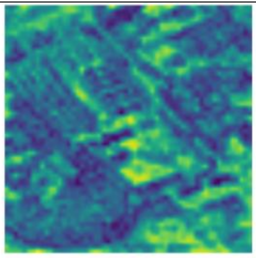
The Hyperspectral Remote Sensing Scenes Dataset [21] contains the images that are collected from satellites or airborne sensors. This dataset is classified into Salinas, Indian Pines, and Pavia Centre Scene sections. The scenes in Indian Pines are collected from different sensors over northwestern Indiana, and they contain 224 spectral reflectance bands. Moreover, the scenes contain one-third forest, two-thirds agriculture, and natural vegetation. Likewise, the Salinas scene was gathered by 224 bands over California, the Salinas Valley, and they obtained it with high resolution. Together, the ground truth of Salinas contains 16 classes. Similarly, the Pavia Scenes consist of 102 spectral bands, and some samples in the images have no information. They contain 9 classes with 1.3 meters of geometric resolution.

5.3. Experimental results

The proposed S^{2}A-DL-EGBNet model's image outcomes using three datasets are briefly shown in Figure 6.

<table>
<tr><th colspan="3"></th><th>Indian Pines Dataset</th><th>Pavia Centre scene Dataset</th><th>Salinas Dataset</th></tr>
<tr><td colspan="3">Input</td>
<td></td>
<td></td>
<td></td></tr>
<tr><td rowspan="4">Feature Extraction</td><td rowspan="4">Vegetative Index Features</td><td>NDVI</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
<tr><td>SAVI</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
<tr><td>GNDVI</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
<tr><td>DVI</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
</table>

		EVI			
		LAI			
		VARI			
		ARVI			
		GEMI			

		GARI			
	Hybrid Structural Pattern				
	Deep Statistical Vegetative Feature ResNet	ResNet 151			
		Mean			
		Median			

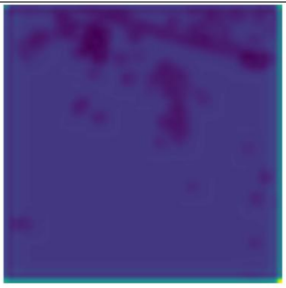
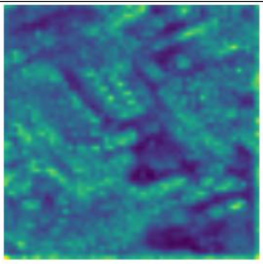
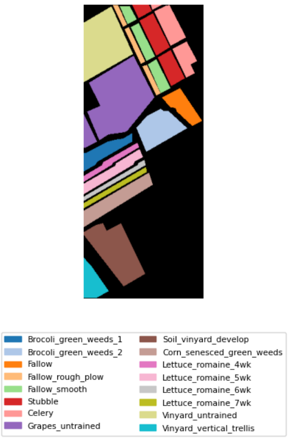
		Standard Deviation			
Output					

Figure 6. S^{2}A-DL-EGBNet model's image outcomes.

5.4. Performance metrics

Accuracy: The metric accuracy denotes the ratio of correctly predicted observations among all the observations, which is expressed as follows:

$$Ac_y = \frac{\upsilon + \delta}{\upsilon + \delta + \alpha + \varphi} \tag{19}$$

where, υ implies the true positive, δ represents the true negative, α denotes false positive, φ is the false negative, and Ac_y indicates the accuracy.

Sensitivity: This indicates the true positive pixel ratio based on the entire positive pixels, and it is expressed as

$$s_n = \frac{\upsilon}{\upsilon + \varphi} \tag{20}$$

where s_n denotes sensitivity.

Specificity: The true negative rate is another name for specificity because it determines the ratio of negatives with respect to negative pixels, which is expressed as

$$s_y = \frac{\delta}{\delta + \varphi} \tag{21}$$

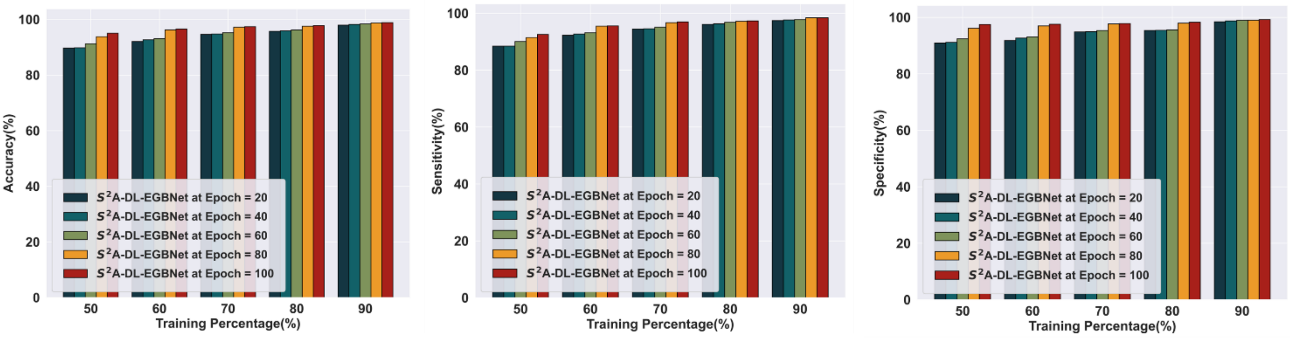
where s_y represents the specificity.

5.5. Performance analysis

This part explores the performance of the S^2A-DL-EGBNet model using the Pavia Centre Scene, the Salinas datasets, and the Indian Pines datasets.

5.5.1. Performance analysis with the Indian Pines dataset

Figure 7 depicts the performance analysis of the proposed S^2A-DL-EGBNet model using Indian Pines. The analysis shows that the accuracy obtained by the proposed model is 97.9%, 98.1%, 98.3%, 98.7%, and 98.8% for the epochs 20, 40, 60, 80, and 100 at 90% of the data. The sensitivity gained by the model S^2A-DL-EGBNet model is 97.3% for epoch 100 in the training data of 80%, and by changing the epoch values, the sensitivity values increased. The specificity attained for S^2A-DL-EGBNet is 99.2% in 90% of the data at epoch 100. Hence, by improving the training data, the model gets better performance with respect to the epoch values.

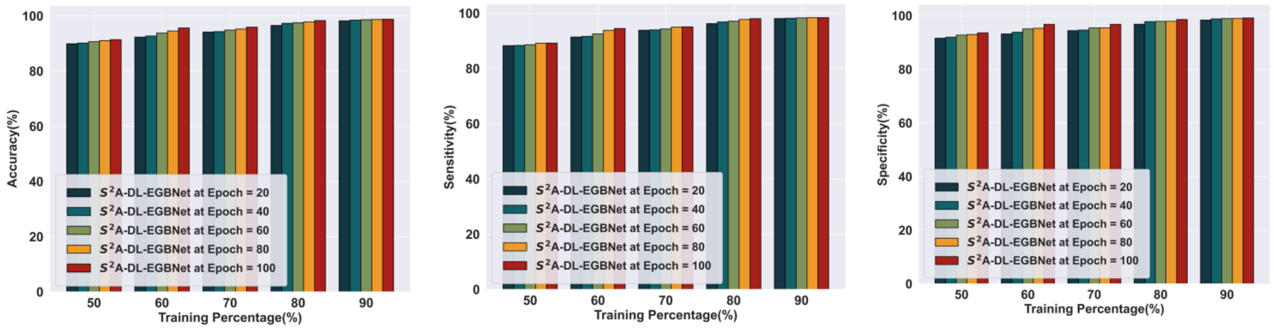


Accuracy Sensitivity Specificity

Figure 7. Performance analysis with the Indian Pines dataset.

5.5.2. Performance analysis with the Pavia Centre Scene dataset

The Pavia Centre Scene dataset is utilized to analyze the S^2A-DL-EGBNet model's performance, which is depicted in Figure 8. From the analysis, the accuracy attained by the proposed model is 98.2% for epoch 20, further it increases to 98.6% at epoch 60, and to 98.8% at epoch 100 with train data of 90%. Moreover, the sensitivity for the proposed model is 98.06% for epoch 100 at 80% of training. At 90% of training data, the specificity for the S^2A- DL-EGBNet model is 99.1% at epoch 100. By maximizing epoch values and training data, the performance of the proposed model gets better results.

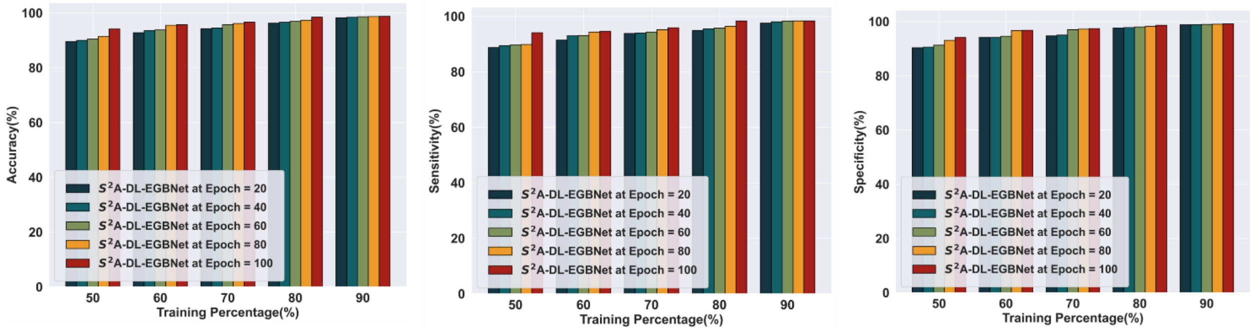


Accuracy Sensitivity Specificity

Figure 8. Performance analysis with the Pavia Centre Scene dataset.

5.5.3. Performance analysis with the Salinas dataset

Figure 9 illustrates the performance analysis of the S^{2}A- DL-EGBNet model using the Salinas dataset at different epochs. With training data 90%, the proposed model obtained the accuracy of 98.2%, 98.4%, 98.6%, 98.70%, and 98.77% for the epochs 20, 40, 60, 80, and 100, respectively, where the accuracy is maximized with increased training data and epochs. The sensitivity and specificity for the S^{2}A- DL-EGBNet model are 98.3% and 98.6% for epoch 100 at 80% training data. The value of all the metrics is improved when the training data and epochs are augmented.



Accuracy Sensitivity Specificity

Figure 9. Performance analysis with Salinas dataset.

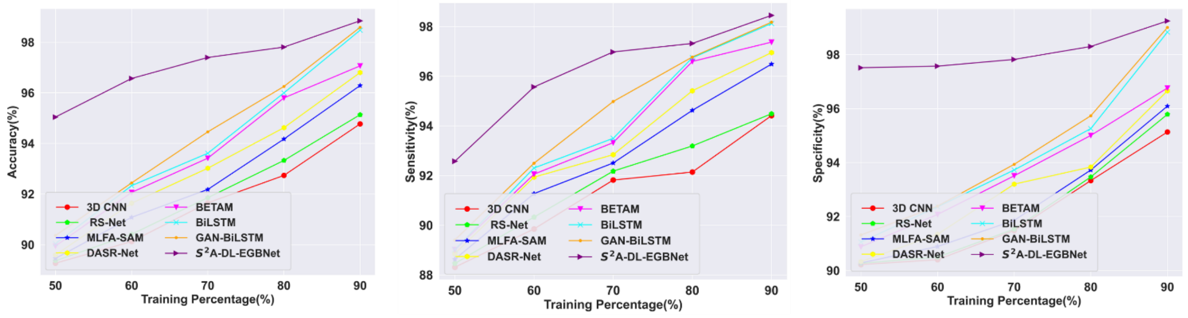
5.6. Comparative methods

The existing 3D CNN [1], RS-Net [2], MLFA-SAM [4], DASR-Net [5], BETAM [6], BiLSTM [24], and GAN-BiLSTM [25] methods are utilized to compare with the S^{2}A- DL-EGBNet model.

5.6.1. Comparative analysis using the Indian Pines dataset

Figure 10 represents the comparative results of the proposed S^{2}A- DL-EGBNet model and existing 3D CNN, RS-Net, MLFA-SAM, DASR-Net, BETAM, BiLSTM, and GAN-BiLSTM approaches using the Indian Pines dataset. With 90% of training data, the accuracy achieved by the

S^{2}A- DL-EGBNet model is 98.8% and it is improved over 4.1%, 3.7%, 2.5%, 2.06%, 1.8%, 0.3%, and 0.2% with the 3D CNN, RS-Net, MLFA-SAM, DASR-Net, BETAM, BiLSTM, and GAN-BiLSTM approaches. The sensitivity of S^{2}A- DL-EGBNet with 90% of training data is 98.4%, which is greater than the 3D CNN by 4.09%, RS-Net by 4.02%, MLFA-SAM by 1.9%, DASR-Net by 1.5%, BETAM by 1.09%, BiLSTM by 0.3%, and GAN-BiLSTM by 0.2%, respectively. The specificity of the proposed S^{2}A- DL-EGBNet model is 99.2%, which is higher than the existing methods 3D CNN, RS-Net, MLFA-SAM, DASR-Net, BETAM, BiLSTM, and GAN-BiLSTM model by 4.1%, 3.4%, 3.1%, 2.6%, 2.5%, 0.4%, and 0.2% with 90% of training data.

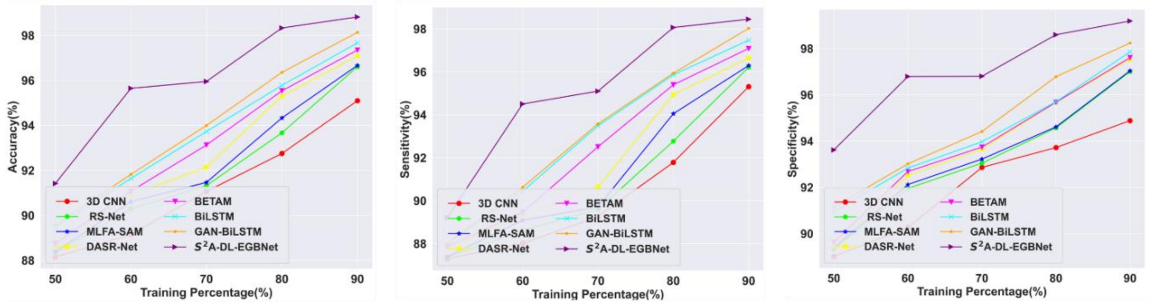


Accuracy Sensitivity Specificity

Figure 10. Comparative analysis with the Indian Pines dataset.

5.6.2. Comparative analysis using the Pavia Centre Scene dataset

The comparative results of the proposed S^{2}A- DL-EGBNet model and existing 3D CNN, RS-Net, MLFA-SAM, DASR-Net, BETAM, BiLSTM, and GAN-BiLSTM approaches using the Pavia Centre Scene dataset are depicted in Figure 11. The accuracy of the S2A- DL-EGBNet model is 98.3%, which is higher than comparative approaches. 5.6% with 3D CNN, 4.7% with RS-Net, 4.06% with MLFA-SAM, 3.07% with DASR-Net, 2.8% with BETAM, 2.5% with BiLSTM, and 2% with GAN-BiLSTM with 80% training data. The sensitivity of the S^{2}A- DL-EGBNet model with 60% and 90% of training data is 94.5%, and 98.4%, respectively. Comparatively, the sensitivity of the S^{2}A- DL-EGBNet model is better than other compared techniques. The specificity value of S^{2}A- DL-EGBNet with 90% of training data is 99.1%, and it is higher than the 3D CNN by 4.3%, RS-Net by 2.2%, MLFA-SAM by 2.1%, DASR-Net by 1.6%, BETAM by 1.5%, BiLSTM by 1.3%, and GAN-BiLSTM by 0.9%.

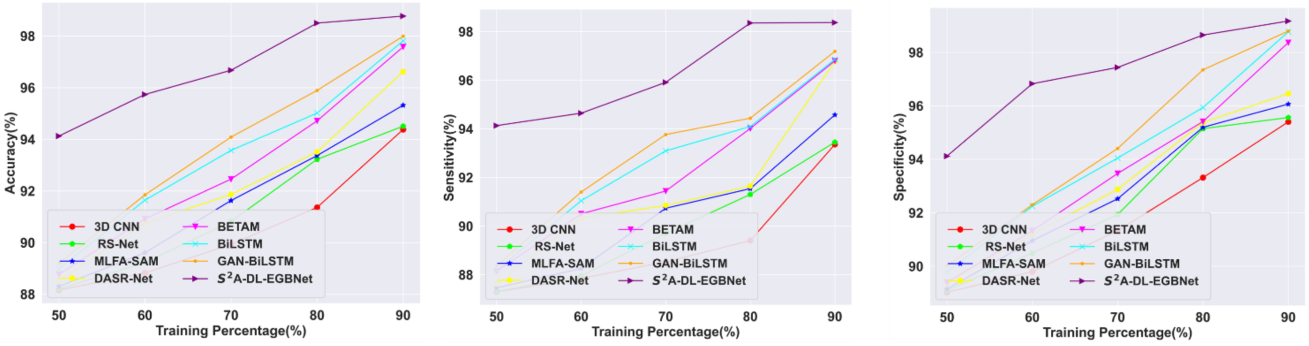


Accuracy Sensitivity Specificity

Figure 11. Comparative analysis with the Pavia Centre scene dataset.

5.6.3. Comparative analysis using the Salinas dataset

Figure 12 reveals the comparative analysis of the proposed S^{2}A- DL-EGBNet model and existing 3D CNN, RS-Net, MLFA-SAM, DASR-Net, BETAM, BiLSTM, and GAN-BiLSTM techniques by utilizing the Salinas dataset. The accuracy of the S^{2}A- DL-EGBNet model is 98.7% for 90% training data, and it shows that the accuracy of the S^{2}A- DL-EGBNet model exceeds the existing approaches. Here, the sensitivity attained by the S^{2}A- DL-EGBNet model is 98.37% for 90% of the train data, and it is with the improvement of 3D CNN, 5% with RS-Net, 3.8% with MLFA-SAM, 1.61% with DASR-Net, 1.60% with BETAM, 1.5% with BiLSTM, and 1.2% with GAN-BiLSTM approaches. The specificity of the S^{2}A- DL-EGBNet approach is 99.1% at 90% of training data, and it is higher than the existing 3D CNN, RS-Net, MLFA-SAM, DASR-Net, BETAM, BiLSTM, and GAN-BiLSTM approaches.



Accuracy Sensitivity Specificity

Figure 12. Comparative analysis with the Salinas dataset.

5.7. Comparative discussion

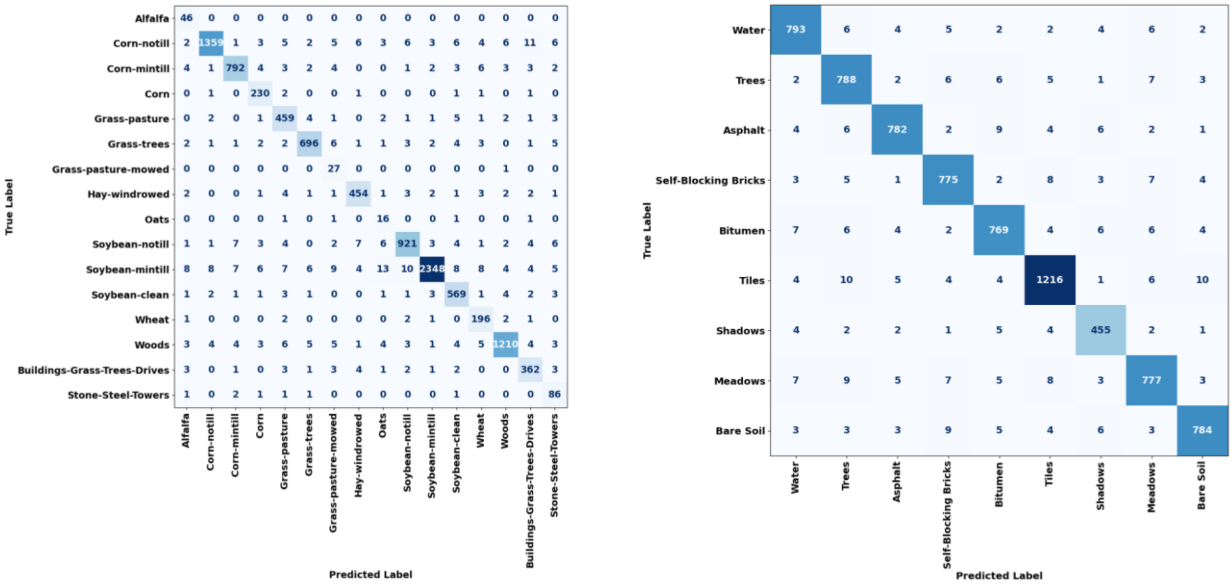
The conventional 3D CNN, RS-Net, MLFA-SAM, DASR-Net, BETAM, BiLSTM, and GAN-BiLSTM approaches are compared with the proposed S^{2}A- DL-EGBNet model, which is described in Table 4. The suggested 3D CNN [1] model suffers from a data scarcity problem because a vast amount of labeled data is needed for training, and the model does not integrate transfer learning and data augmentation to enhance model generalization. The developed RS-Net [2] model faced complexities while handling large datasets, and it has limitations while extracting complex data features. The suggested MLFA-SAM [4] model exhibits low performance because it fails to adopt advanced training methods like fine-tuning strategies. The DASR-Net [5] model had poor generalization capability across seasonal conditions, and it was evaluated with limited datasets. Moreover, the model cannot be applied to large-scale remote sensing data due to computational issues. The BETAM method suggested in [6] exhibits low segmentation accuracy and has dataset limitations that affect the performance of the model. These challenges are faced by the conventional approaches in LC classification. Hence, the proposed S^{2}A- DL-EGBNet model is developed to address the above challenges that reduce the issue of computational complexity, and improve feature representation by enhancing the quality of the dataset by performing accurate LC classification. The incorporation of various feature extraction techniques extracts complex and intricate features that influence the classification, and also eliminates the background noise while focusing on significant features. The application of the GAN rectifies the class imbalance problem, which reduces overfitting issues. Most importantly, the DL improves the S2A-DL-EGBNet model's learning process and minimizes computation time, which enhances generalizability. Moreover, the strength of the hybrid attention mechanisms, the encoder and BiLSTM, preserves the minute features from the input and provides accurate LC classification consistently.

Table 4. Comparative outcome of the S^2A- DL-EGBNet model.

Methods/ Metrics		3D CNN	RS-Net	MLFA-SAM	DASR-Net	BETAM	BiLSTM	GAN-BiLSTM	S^2A-DL-EGBNet
Indian Pines Dataset	**Accuracy**	94.7	95.1	96.2	96.8	97	98.4	98.5	98.84
	Sensitivity	94.4	94.4	96.4	96.9	97.3	98.1	98.1	98.44
	Specificity	95.1	95.7	96	96.6	96.7	98.8	99	99.24
Pavia Centre scene Dataset	**Accuracy**	95	96.6	96.6	97	97.3	97.6	98.1	98.82
	Sensitivity	95.3	96.2	96.2	96.6	97	97.4	98	98.45
	Specificity	94.8	97	97	97.5	97.6	97.8	98.2	99.19
Salinas Dataset	**Accuracy**	94.3	94.5	95.3	96.6	97.5	97.8	97.9	98.77
	Sensitivity	93.3	93.4	94.5	96.7	96.7	96.8	97.1	98.37
	Specificity	95.4	95.5	96	96.4	98.3	98.7	98.7	99.17

5.8. Confusion matrix with a

The confusion matrix measures the performance of the S^2A- DL-EGBNet model by associating the positive prediction labels with predicted labels, and then it analyzes the number of positive reactions that take place papredictions takesplace in the dataset. In the Indian Pines dataset, the model correctly predicted 2348 classes as soybean-mintill, 1359 as corn-notill, 792 as corn-mintill, 230 as corn, 459 as grass pasture, 696 as grass-trees, and so on. Similarly, the remaining dataset analyses the true labels with correctly predicted labels. The classification ability of the S^2A- DL-EGBNet model with the Pavia Centre scene, Indian Pines, and Salinas datasets is visualized in Figure 13.

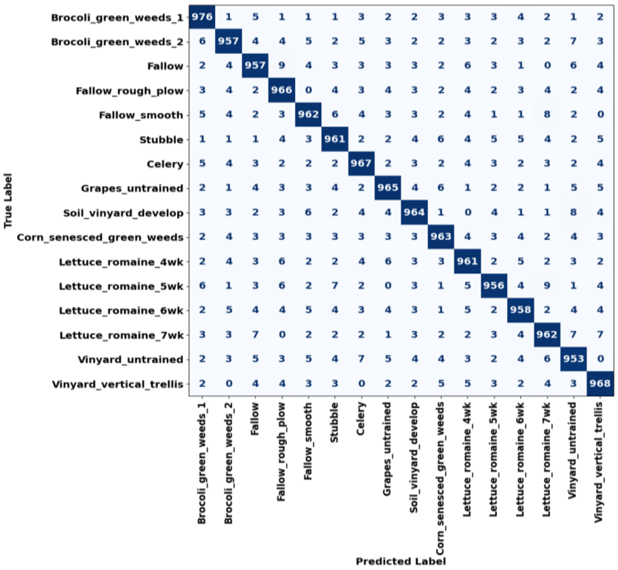


	Alfalfa	Corn-notill	Corn-mintill	Corn	Grass-pasture	Grass-trees	Grass-pasture-mowed	Hay-windrowed	Oats	Soybean-notill	Soybean-mintill	Soybean-clean	Wheat	Woods	Buildings-Grass-Trees-Drives	Stone-Steel-Towers
Alfalfa	46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Corn-notill	2	1359	1	3	5	2	5	6	3	6	3	6	4	6	11	6
Corn-mintill	4	1	792	4	3	2	4	0	0	1	2	3	6	3	3	2
Corn	0	1	0	230	2	0	0	1	0	0	0	1	1	0	1	0
Grass-pasture	0	2	0	1	459	4	1	0	2	1	1	5	1	2	1	3
Grass-trees	2	1	1	2	2	696	6	1	1	3	2	4	3	0	1	5
Grass-pasture-mowed	0	0	0	0	0	0	27	0	0	0	0	0	0	1	0	0
Hay-windrowed	2	0	0	1	4	1	1	454	1	3	2	1	3	2	2	1
Oats	0	0	0	0	1	0	1	0	16	0	0	1	0	0	1	0
Soybean-notill	1	1	7	3	4	0	2	7	6	921	3	4	1	2	4	6
Soybean-mintill	8	8	7	6	7	6	9	4	13	10	2348	8	8	4	4	5
Soybean-clean	1	2	1	1	3	1	0	0	1	1	3	569	1	4	2	3
Wheat	1	0	0	0	2	0	0	0	0	2	1	0	196	2	1	0
Woods	3	4	4	3	6	5	5	1	4	3	1	4	5	1210	4	3
Buildings-Grass-Trees-Drives	3	0	1	0	3	1	3	4	1	2	1	2	0	0	362	3
Stone-Steel-Towers	1	0	2	1	1	1	0	0	0	0	0	1	0	0	0	86

Indian Pines dataset

	Water	Trees	Asphalt	Self-Blocking Bricks	Bitumen	Tiles	Shadows	Meadows	Bare Soil
Water	793	6	4	5	2	2	4	6	2
Trees	2	788	2	6	6	5	1	7	3
Asphalt	4	6	782	2	9	4	6	2	1
Self-Blocking Bricks	3	5	1	775	2	8	3	7	4
Bitumen	7	6	4	2	769	4	6	6	4
Tiles	4	10	5	4	4	1216	1	6	10
Shadows	4	2	2	1	5	4	455	2	1
Meadows	7	9	5	7	5	8	3	777	3
Bare Soil	3	3	3	9	5	4	6	3	784

Pavia Centre scenes dataset



	Brocoli_green_weeds_1	Brocoli_green_weeds_2	Fallow	Fallow_rough_plow	Fallow_smooth	Stubble	Celery	Grapes_untrained	Soil_vinyard_develop	Corn_senesced_green_weeds	Lettuce_romaine_4wk	Lettuce_romaine_5wk	Lettuce_romaine_6wk	Lettuce_romaine_7wk	Vinyard_untrained	Vinyard_vertical_trellis
Brocoli_green_weeds_1	976	1	5	1	1	1	3	2	2	3	3	3	4	2	1	2
Brocoli_green_weeds_2	6	957	4	4	5	2	5	3	2	2	3	2	3	2	7	3
Fallow	2	4	957	9	4	3	3	3	3	2	6	3	1	0	6	4
Fallow_rough_plow	3	4	2	966	0	4	3	4	3	2	4	2	3	4	2	4
Fallow_smooth	5	4	2	3	962	6	4	3	3	2	4	1	1	8	2	0
Stubble	1	1	1	4	3	961	2	2	4	6	4	5	5	4	2	5
Celery	5	4	3	2	2	2	967	2	3	2	4	3	2	3	2	4
Grapes_untrained	2	1	4	3	3	4	2	965	4	6	1	2	2	1	5	5
Soil_vinyard_develop	3	3	2	3	6	2	4	4	964	1	0	4	1	1	8	4
Corn_senesced_green_weeds	2	4	3	3	3	3	3	3	3	963	4	3	4	2	4	3
Lettuce_romaine_4wk	2	4	3	6	2	2	4	6	3	3	961	2	5	2	3	2
Lettuce_romaine_5wk	6	1	3	6	2	7	2	0	3	1	5	956	4	9	1	4
Lettuce_romaine_6wk	2	5	4	4	5	4	3	4	3	1	5	2	958	2	4	4
Lettuce_romaine_7wk	3	3	7	0	2	2	2	1	3	2	2	3	4	962	7	7
Vinyard_untrained	2	3	5	3	5	4	7	5	4	4	3	2	4	6	953	0
Vinyard_vertical_trellis	2	0	4	4	3	3	0	2	2	5	5	3	2	4	3	968

Salinas dataset

Figure 13. Confusion matrix.

5.9. Statistical analysis

The statistical results of the S^{2}A- DL-EGBNet model for the Pavia Centre scene dataset, Salinas dataset, and Indian Pines dataset are tabulated in Tables 5–7, respectively. Various tests, such as the Wilcoxon signed-rank test, Nemenyi post-hoc test, Friedman test, and statistical test, are performed to analyze the statistical significance of the evaluation results.

Table 5. Statistical analysis with the Indian Pines dataset.

Methods/ Metrics		**Indian Pines dataset**							
		Accuracy							
		3D CNN	RS-Net	MLFA-SAM	DASR-Net	BETAM	BiLSTM	GAN-BiLSTM	S^2A- DL-EGBNet
T- test	Chi2-Statistic	1.8010	1.5050	1.7872	0.1905	1.1044	0.3542	1.3653	1.0573
	P-value	0.3679	0.3679	0.3679	0.3679	0.3679	0.3679	0.3679	0.3679
Statistical test	Best	94.773	95.137	96.288	96.803	97.066	98.476	98.588	98.847
	Mean	91.711	92.020	92.637	93.136	93.659	94.091	94.420	97.133
	Variance	3.783	4.219	5.688	6.092	6.515	8.521	8.226	1.630
	STD	1.945	2.054	2.385	2.468	2.552	2.919	2.868	1.277
	Median	91.652	91.856	92.182	93.021	93.417	93.612	94.458	97.398
	T-Statistic	3.1483	3.0352	3.0625	2.9713	2.6696	3.0040	2.9067	2.6846
	P-value	0.0035	0.0039	0.0038	0.0041	0.0056	0.0040	0.0044	0.0055
Wilcoxon signed-rank test	W-Statistic	0.9849	0.1965	0.8318	0.1705	0.0389	0.1880	0.4840	0.9579
	P-value	0.0011	0.0043	0.0045	0.0013	0.0015	0.0045	0.0026	0.0037

Table 6. Statistical analysis with the Pavia Centre scene dataset.

Methods/ Metrics		**Pavia Centre scene dataset**							
		Accuracy							
		3D CNN	RS-Net	MLFA-SAM	DASR-Net	BETAM	BiLSTM	GAN-BiLSTM	S^2A-DL-EGBNet
Friedman test	Chi2-Statistic	1.4667	0.8556	1.7162	0.9925	0.3566	0.7021	1.4814	1.4135
	P-value	0.3679	0.3679	0.3679	0.3679	0.3679	0.3679	0.3679	0.3679
Statistical test	Best	96.105	97.616	97.675	98.097	98.357	98.675	99.141	98.822
	Mean	92.250	93.055	93.300	93.816	94.175	94.666	94.987	96.035
	Variance	6.222	8.154	8.411	9.270	9.399	8.428	9.381	6.903
	STD	2.494	2.856	2.900	3.045	3.066	2.903	3.063	2.627
	Median	92.072	92.330	92.474	93.151	94.134	94.735	95.006	95.953
	T-Statistic	3.091	3.195	3.017	2.812	2.728	2.762	2.713	2.122
	P-value	0.0036	0.0033	0.0039	0.0048	0.0052	0.0050	0.0053	0.0101
Wilcoxon signed-rank test	W-Statistic	0.9251	0.4530	0.3218	0.2640	0.7672	0.9044	0.9625	0.8615
	P-value	0.0016	0.0007	0.0014	0.0005	0.0019	0.0030	0.0009	0.0045

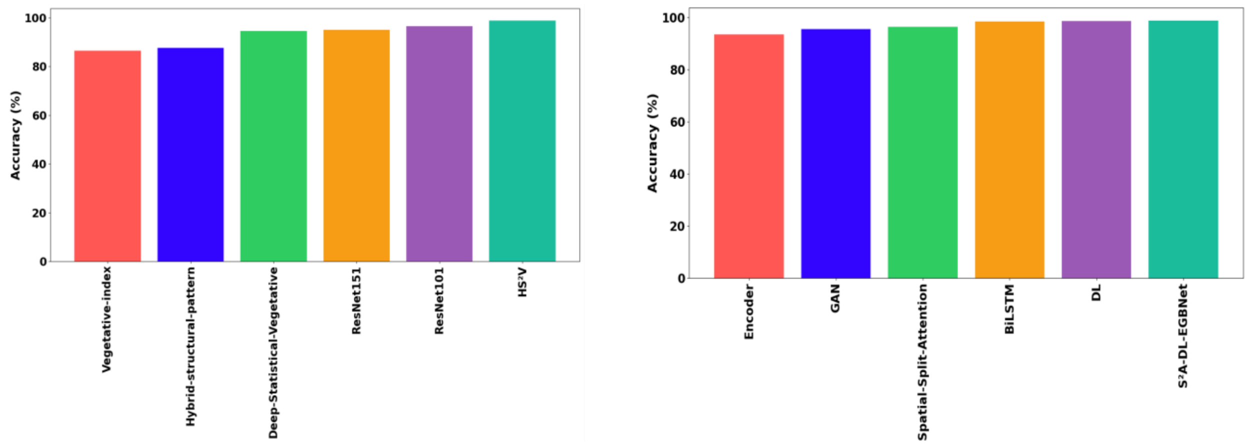
Table 7. Statistical analysis with the Salinas dataset.

Methods/ Metrics		**Salinas dataset dataset**							
		Accuracy							
		3D CNN	RS-Net	MLFA-SAM	DASR-Net	BETAM	BiLSTM	GAN-BiLSTM	S^2A- DL-EGBNet
Friedman test	Chi2-Statistic	0.3883	1.4318	0.3401	1.5460	0.6618	0.8731	1.2578	0.0865
	P-value	0.3679	0.3679	0.3679	0.3679	0.3679	0.3679	0.3679	0.3679
Statistical test	Best	95.389	95.516	96.329	97.630	98.590	98.822	99.002	98.774
	Mean	91.541	92.205	92.654	93.261	93.896	94.414	94.811	96.764
	Variance	4.877	5.612	6.361	7.419	9.281	8.947	9.417	3.014
	STD	2.208	2.369	2.522	2.724	3.047	2.991	3.069	1.736
	Median	90.939	91.825	92.637	92.873	93.465	94.579	95.095	96.677
	T-Statistic	3.091	3.195	3.017	2.812	2.728	2.762	2.713	2.122
	P-value	0.0036	0.0033	0.0039	0.0048	0.0052	0.0050	0.0053	0.0101
Wilcoxon signed-rank test	W-Statistic	3.4853	2.7958	2.9143	3.2084	3.0813	2.9480	2.7309	2.3154
	P-value	0.0025	0.0049	0.0043	0.0033	0.0037	0.0042	0.0052	0.0082

5.10. Ablation study of feature extraction methods

Figure 14 explicates the ablation study of the multiple feature extraction techniques and components used in the proposed model for LC classification. The feature extraction methods, such as vegetative-index, hybrid-structural-pattern, deep-statistical-vegetative, resnet-151, resnet-101, and HS^2V, are compared based on accuracy, which is shown in Figure 14a). The methods obtained accuracy of 86.49%, 87.65%, 94.57%, 95.08%, 96.55%, and 98.85%. Here, after hybridizing all of the feature extraction methods, HS^2V obtained a high accuracy of 98.85% and outperformed other individual methods. Moreover, the application of these techniques extracts complex and important features that boost the accurate LC classification while reducing overfitting issues. Specifically, the depth feature extraction techniques of ResNet-151 and ResNet-101 are used to differentiate the variations in photographic distortions and illumination changes from the remote sensing images. Therefore, these extracted features received from each technique are concatenated to generate significant features.

Figure 14b) depicts the ablation study of multiple components used in the proposed S2A-DL-EGBNet model based on accuracy. Based on the analysis, the S2A-DL-EGBNet model attained a high accuracy of 98.85% and outperformed every component. The model components obtained less accuracy with Encoder at 93.54%, GAN at 95.56%, spatial-split-attention at 96.49%, BiLSTM at 98.48%, and DL of 98.59%. Hence, the benefits of each component enhance the model's performance, which leads to better robustness and scalability in LC classification.



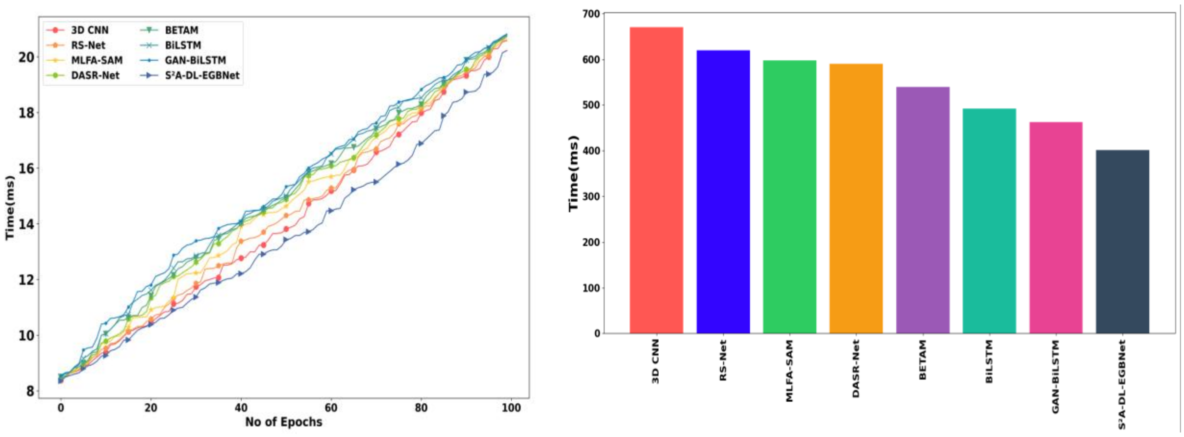
a) Feature extraction methods b) Model components

Figure 14. Ablation study.

5.11. *Computational complexity analysis*

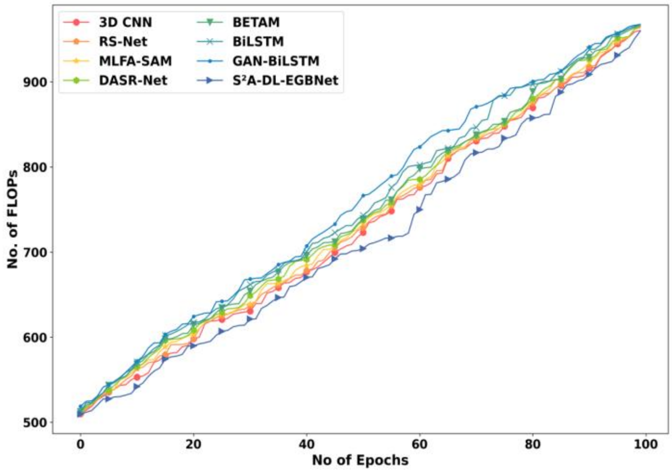
Figure 15a) explicates the computational time required for the proposed model for LC classification, compared with other models. The S2A-DL-EGBNet model utilizes less computation time of 20.23 ms and outperforms other models. The other models utilized high computation time: 3D CNN of 20.59 ms, RS-Net of 20.60 ms, MLFA-SAM of 20.69 ms, DASR-Net of 20.74 ms, BETAM of 20.75 ms, BiLSTM of 20.79 ms, and GAN-BiLSTM of 20.82 ms. Similarly, Figure 15b) illustrates the inference time analysis of the models. The proposed model achieved a faster inference time of 401.95 ms and surpasses the 3D CNN of 670.57 ms, RS-Net of 620.32 ms, MLFA-SAM of 598.77 ms, DASR-Net of 590.58 ms, BETAM of 540.35 ms, BiLSTM of 492.85 ms, and GAN-BiLSTM of 462.37ms. The utilization of DL, GAN data balancing, and a spatial-split-attention mechanism with the BiLSTM model allows it to learn effective features that reduce computational complexity.

Figure 15c) depicts the computational cost analysis in terms of Floating-point Operations Per second (FLOPs), which is performed to evaluate the model's performance in handling the computational workload and resource utilization compared to other existing algorithms. As per the analysis, the proposed S2A-DL-EGBNet model reduced the computation cost to 959.52 FLOPs in 100 epochs. Simultaneously, the 3D CNN had 960.30 FLOPs, RS-Net had 963.69 FLOPs, MLFA-SAM had 965.14 FLOPs, DASR-Net had 965.74 FLOPs, BETAM had 966.01 FLOPs, BiLSTM had 967.37 FLOPs, and GAN-BiLSTM had 967.43 FLOPs. The computational complexity analysis of the S2A-DL-EGBNet model based on computation time, inference time, and FLOPs analysis is depicted in Figure 15.



a) Time complexity

b) Inference time



c) FLOPs analysis

Figure 15. Computational complexity analysis.

5.12. Per-class metrics analysis using the GAN

Figure 16 shows the evaluation of the S2A-DL-EGBNet model for the individual class with and without the GAN based on performance metrics. The application of GAN-based data balancing attained a high precision of 98.87%, recall of 98.05%, and F1-score of 98.46%, which leads to effective performance of LC classification. However, the absence of the GAN creates overfitting issues that affect the model's classification process, with a precision of 92.90%, a recall of 93.59%, and an F1-score of 93.24%. Therefore, the application of GAN-based data balancing is essential in this research to attain a robust LC classification process.

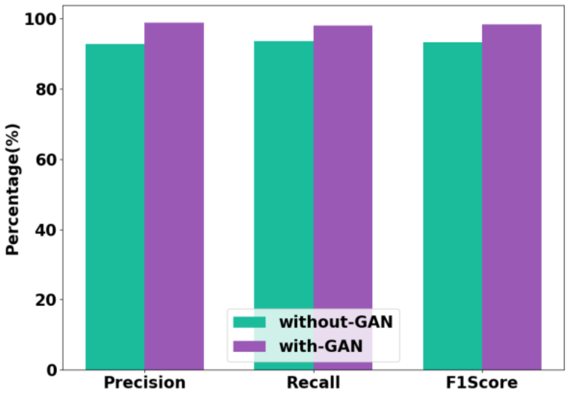


Figure 16. Per-class metrics analysis using the GAN.

6. Conclusions

In this study, the S2A-DL-EGBNet model performed accurate LC classification, which utilizes remote sensing techniques to classify particular categories of ground cover by understanding and analyzing the attributes of the Earth's surface. Notably, the S^2A- DL-EGBNet model employs HS^2V feature extraction to capture the complex attributes from images such as patterns, structures, and textures, allowing for the extraction of informative regions. GAN-based data balancing is exploited to ensure diversity in image samples and enhance generalization. The model is trained effectively in a parallel manner with large datasets to enhance the classification accuracy. The distributed learning module is integrated to offer flexibility to the model, and it enhances scalability by reducing the computational cost. However, the hybrid attention techniques are utilized to preserve spatial saliency and channel-wise diversity. Furthermore, the model's generalization capability is improved by adopting vast heterogeneous hyperspectral data to ensure its utility in a larger context with large-scale scenarios. Based on the experimental results, the model showed impressive outcomes with an accuracy of 98.84%, sensitivity of 98.44%, and specificity of 99.24% on the Indian Pines dataset for 90% training. In the future, the model will be integrated with meta-heuristic algorithms to optimize the hyperparameters to further enhance the classification accuracy with high convergence. Additionally, an explainability technique will be incorporated to enhance the decision-making ability.

Author contributions

Afnan M. Alhassan: Conceptualization, methodology, data curation, formal analysis, investigation, writing – original draft, visualization; Nouf I. Altmami: Validation, supervision, writing – review and editing, project administration, resources. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Data availability and access

This study utilizes publicly accessible datasets.

Acknowledgments

The authors would like to thank the Deanship of Scientific Research at Shaqra University for supporting this work.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Z. Zhang, D. Shu, C. Liao, C. Liu, Y. Zhao, R. Wang, et al., FlexiSAM: A flexible SAM-based semantic segmentation model for land cover classification using high-resolution multimodal remote sensing imagery, *ISPRS J. Photogramm. Remote Sens.*, **227** (2025), 594–612. https://doi.org/10.1016/j.isprsjprs.2025.05.028
2. A. Vali, S. Comai, M. Matteucci, Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: a review, *Remote Sens.*, **12** (2020), 2495. https://doi.org/10.3390/rs12152495
3. M. Aljebreen, H. A. Mengash, M. Alamgeer, S. S. Alotaibi, A. S. Salama, M. A. Hamza, Land use and land cover classification using river formation dynamics algorithm with deep learning on remote sensing images, *IEEE Access*, **12** (2024), 11147–11156. https://doi.org/10.1109/ACCESS.2023.3349285
4. R. P. Sishodia, R. L. Ray, S. K. Singh, Applications of remote sensing in precision agriculture: a review, *Remote Sens.*, **12** (2020), 3136. https://doi.org/10.3390/rs12193136
5. C. K. K. Reddy, A. Daduvy, R. M. Mohana, B. Assiri, M. Shuaib, S. Alam, et al., Enhancing precision agriculture and land cover classification: a self-attention 3D convolutional neural network approach for hyperspectral image analysis, *IEEE Access*, **12** (2024), 125592–125608. https://doi.org/10.1109/access.2024.3420089
6. M. Weiss, F. Jacob, G. Duveiller, Remote sensing for agricultural applications: a meta-review, *Remote Sens. Environ.*, **236** (2020), 111402. https://doi.org/10.1016/j.rse.2019.111402
7. A. Nițu, C. Florea, M. Ivanovici, A. Racoviteanu, NDVI and Beyond: vegetation indices as features for crop recognition and segmentation in hyperspectral data, *Sensors (Basel)*, **25** (2025), 3817. https://doi.org/10.3390/s25123817
8. A. D. Campbell, T. Fatoyinbo, S. P. Charles, L. L. Bourgeau-Chavez, J. Goes, H. Gomes, et al., A review of carbon monitoring in wet carbon systems using remote sensing, *Environ. Res. Lett.*, **17** (2022), 025009. https://doi.org/10.1088/1748-9326/ac4d4d
9. G. A. Sánchez-Azofeifa, K. L. Castro-Esau, W. A. Kurz, A. Joyce, Monitoring carbon stocks in the tropics and the remote sensing operational limitations: from local to regional projects, *Ecol. Appl.*, **19** (2009), 480–494. https://doi.org/10.1890/08-1149.1
10. X. Li, X. Fan, Q. Li, X. Zhao, Rs-net: Hyperspectral image land cover classification based on spectral imager combined with random forest algorithm, *Electronics*, **13** (2024), 4046. https://doi.org/10.3390/electronics13204046

11. X. Li, X. Fan, J. Fan, Q. Li, Y. Gao, X. Zhao, DASR-Net: Land cover classification methods for hybrid multiattention multispectral high spectral resolution remote sensing imagery, *Forests*, **15** (2024), 1826. https://doi.org/10.3390/f15101826
12. H. Yang, Z. Jiang, Y. Zhang, Y. Wu, H. Luo, P. Zhang, et al., A high-resolution remote sensing land use/land cover classification method based on multi-level features adaptation of segment anything model, *Int. J. Appl. Earth Obs. Geoinf.*, **141** (2025), 104659. https://doi.org/10.1016/j.jag.2025.104659
13. G. Wang, J. Chen, L. Mo, P. Wu, X. Yi, Border-enhanced triple attention mechanism for high-resolution remote sensing images and application to land cover classification, *Remote Sens.*, **16** (2024), 2814. https://doi.org/10.3390/rs16152814
14. M. Fayaz, J. Nam, L. M. Dang, H. K. Song, H. Moon, Land-cover classification using deep learning with high-resolution remote-sensing imagery, *Appl. Sci.*, **14** (2024), 1844. https://doi.org/10.3390/app14051844
15. B. Chen, L. Liu, Z. Zou, Z. Shi, Target detection in hyperspectral remote sensing image: current status and challenges, *Remote Sens.*, **15** (2023), 3223. https://doi.org/10.3390/rs15133223
16. R. Vidican, A. Mălinaș, O. Ranta, C. Moldovan, O. Marian, A. Ghețe, et al., Using remote sensing vegetation indices for the discrimination and monitoring of agricultural crops: a critical review, *Agronomy*, **13** (2023), 3040. https://doi.org/10.3390/agronomy13123040
17. A. S. Alademomi, C. J. Okolie, O. E. Daramola, S. A. Akinnusi, E. Adediran, H. O. Olanrewaju, et al., The interrelationship between LST, NDVI, NDBI, and land cover change in a section of Lagos metropolis, Nigeria, *Appl. Geomatics*, **14** (2022), 299–314. https://doi.org/10.1007/s12518-022-00434-2
18. A. Hussain, M. Imad, A. Khan, B. Ullah, Multi-class classification for the identification of COVID-19 in X-ray images using customized efficient neural network, In: *AI and IoT for sustainable development in emerging countries*, Springer, Cham, 2022, 473–486. https://doi.org/10.1007/978-3-030-90618-4_23
19. S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, J. A. Benediktsson, Deep learning for hyperspectral image classification: an overview, *IEEE Trans. Geosci. Remote Sens.*, **57** (2019), 6690–6709. https://doi.org/10.1109/TGRS.2019.2907932
20. G. S. Chadha, A. Panambilly, A. Schwung, S. X. Ding, Bidirectional deep recurrent neural networks for process fault classification, *ISA Trans.*, **106** (2020), 330–342. https://doi.org/10.1016/j.isatra.2020.07.011
21. M. Graña, M. A. Veganzons, B. Ayerdi, Hyperspectral remote sensing scenes dataset, 2025. Available from: https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.
22. V. Vani, V. R. Mandla, Comparative study of NDVI and SAVI vegetation indices in Anantapur district semi-arid areas, *Int. J. Civ. Eng. Technol.*, **8** (2017), 559–566.
23. I. H. El-Shal, O. M. Fahmy, M. A. Elattar, License plate image analysis empowered by generative adversarial neural networks (GANs), *IEEE Access*, **10** (2022), 30846–30857. https://doi.org/10.1109/access.2022.3157714
24. Z. Hameed, B. Garcia-Zapirain, Sentiment classification using a single-layered BiLSTM model, *IEEE Access*, **8** (2020), 73992–74001. https://doi.org/10.1109/access.2020.2988550
25. H. A. Al-Najjar, B. Pradhan, R. Sarkar, G. Beydoun, A. Alamri, A new integrated approach for landslide data balancing and spatial prediction based on generative adversarial networks (GAN), *Remote Sens.*, **13** (2021), 4011. https://doi.org/10.3390/rs13194011

26. Y. Tan, X. Ding, Split-attention CNN and self-attention with RoPE and GCN for voice activity detection, *IEEE Access*, **12** (2024), 156673–156682. https://doi.org/10.1109/access.2024.3486003
27. H. Zhang, C. Wu, Z. Zhang, Y. Zhu, H. Lin, Z. Zhang, et al., Resnest: Split-attention networks, *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2022, 2735–2745. https://doi.org/10.1109/cvprw56347.2022.00309
28. S. Woo, J. Park, J. Y. Lee, I. S. Kweon, Cbam: Convolutional block attention module, *Computer Vision – ECCV*, Springer, Cham, 2018, 3–19. https://doi.org/10.1007/978-3-030-01234-2_1