

Received 16 March 2024, accepted 7 April 2024, date of publication 15 April 2024, date of current version 2 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3388457

## RESEARCH ARTICLE

# Enhancing Hyperspectral Image Classification for Land Use Land Cover With Dilated Neighborhood Attention Transformer and Crow Search Optimization

GANJI TEJASREE AND AGILANDEESWARI LOGANATHAN 

School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India


Corresponding author: Agilandeewari Loganathan (agila.l@vit.ac.in)

**ABSTRACT** The classification of Land Use Land Cover (LULC) can be accomplished with the help of hyperspectral imaging, which is a cutting-edge technology. Nevertheless, despite its efficacy, the utilization of hyperspectral images for LULC classification continues to present difficulties and demands a significant amount of time. The limited availability of training samples for hyperspectral images poses a challenge in achieving accurate classification of LULC. Nevertheless, through meticulous deliberation and examination, this impediment can be surmounted. To tackle the task of LULC classification, we have developed a Dilated Neighbourhood Attention Transformer (DNAT). Firstly, we employ LeNet-5 to extract features from the provided data. Subsequently, we perform band selection using Crow Search Optimization (CSO). Following the extraction of features and selection of bands, we proceed to classify LULC. In our study, we used the Salinas, Indian Pines (IP), and Washington DC Mall datasets for LULC classification. The performance of our proposed classification approach is evaluated using the commonly used metrics, namely, Average Accuracy (AA), Overall Accuracy (OA), and Kappa Coefficient (KC). We have achieved 99.85% as OA, 99.83% as AA, and 99.73% as KC for the Salinas Dataset. This is the highest accuracy we have achieved using the DNAT classifier. The experimental results proved beyond a reasonable doubt that the proposed method achieved the highest possible performance, surpassing all prior methods.

**INDEX TERMS** Crow search optimization, dilated neighborhood attention transformer, hyperspectral image, land use land cover classification, LeNet-5.

## I. INTRODUCTION

The use of hyperspectral imaging has been a huge boon to remote sensing as a whole. It has become an essential technology in the industry due to the accurate and comprehensive information it provides about the physical and chemical properties of objects that are imaged. The technique involves capturing and processing images across a wide range of electromagnetic frequencies, providing a comprehensive view of the spectral signature of the objects. The history of hyperspectral imaging dates back to the 1970s when remote sensing researchers started exploring the use of remote

The associate editor coordinating the review of this manuscript and approving it for publication was Mauro Gaggero .

sensing technologies to study the Earth's surface. At that time, traditional remote sensing techniques relied on a limited number of spectral bands to capture images, which provided only a limited amount of information about the objects being imaged. Hyperspectral imaging systems can now capture images across hundreds of spectral bands with remarkable accuracy and precision, thanks to recent developments in sensor technology [1]. To obtain an image of a scene, hyperspectral imaging systems use light with a variety of wavelengths, from ultraviolet to infrared. These systems can tell you a lot about the physical and chemical characteristics of the objects in the image just by looking at how much light reflects at different wavelengths. A three-dimensional data cube is produced, with information regarding the reflectance

of light at each wavelength contained in each pixel [2]. Hyperspectral imaging is a sophisticated technology that finds extensive application in diverse industries, including agriculture, environmental monitoring, mineral exploration, healthcare, and military sectors. Hyperspectral imaging is widely employed in agriculture to effectively monitor the health of crops, identify pests and diseases, and optimize the use of irrigation and fertilizers [3]. When it comes to healthcare and the military, hyperspectral imaging is a game-changer. Its non-invasive method has revolutionized medical imaging, leading to more precise and time-saving cancer diagnoses. Hyperspectral imaging is an essential technology for military operations such as target detection, reconnaissance, and situational awareness, which helps our troops to remain ahead of enemy advances [4]. New advancements in sensor technology continue to expand the capabilities of hyperspectral imaging. For example, the development of small and lightweight hyperspectral sensors has made it possible to capture high-resolution hyperspectral images from drones and other unmanned aerial vehicles. These advancements are expected to open up new applications for hyperspectral imaging in areas such as precision agriculture, disaster management, and infrastructure monitoring [5].

An important and very accurate way to get useful information about the surface of the Earth is LULC classification using remote sensing [6], [7], [8]. Land use refers to the specific way in which humans make use of the land. It represents the actions and features of humans that are connected to a specific geographical area. Multiple land uses can coexist on the same type of land cover. For instance, a forest can serve as a site for commercial logging or be officially designated as a national park for recreational purposes. By segregating land use, it becomes possible to examine the effects of human activities and evaluate the effectiveness of management strategies. Land cover pertains to the tangible attributes found on the surface of the Earth. These attributes can be either inherent to nature or created by humans and are frequently examined and quantified from a distance using satellite imagery, aerial photography, or on-site investigations. A single land cover type, such as a forest, can serve multiple purposes, such as being designated as a national park or used for logging activities. Segmenting land cover enables autonomous examination of the Earth's surface attributes. One state-of-the-art technique that can take images of the Earth's surface with astonishing detail is hyperspectral imaging. This technology yields unparalleled precision and accuracy by analyzing and interpreting the spectral signatures of different surface materials. The hyperspectral remote sensing images that are obtained consist of numerous closely arranged and distinct spectral bands, providing a comprehensive understanding of surface characteristics [9]. The LULC classification process using hyperspectral images involves several stages. The first stage is the acquisition of hyperspectral data, which is typically done using specialized sensors mounted on aircraft or satellites. The acquired data

is then preprocessed to correct for atmospheric and sensor-induced effects, as well as to remove unwanted noise and artifacts. After preprocessing the hyperspectral data, the following stage is to extract useful features that can accurately distinguish between various land cover classes. Many things influence how well LULC classifications turn out. These include how well the hyperspectral data is processed, which machine learning algorithms are used, and the quality of the data itself. Agricultural monitoring, land-use planning, urban planning, forest management, and environmental monitoring are just a few of the many practical uses for LULC classification using hyperspectral images [10]. Although it has made significant progress in the classification of LULC, this field still encounters various obstacles that hinder its precision and efficiency. Hyperspectral images, which are used for LULC classification, consist of a substantial quantity of spectral bands. As the number of features (spectral bands) increases, the process of training and applying classification models becomes computationally burdensome. Obtaining substantial quantities of excellent labeled data (accurate and reliable information) for training classification models can be costly, time-consuming, and restricted to specific geographical areas. Accurate classification heavily relies on the careful selection of the most informative features from the high-dimensional data. Conventional feature extraction techniques may not be ideal for hyperspectral images due to their distinct characteristics. Training and implementing intricate, deep learning models, which are frequently successful for LULC classification, can require a significant amount of computational power, particularly when resources are limited.

To maximize the accuracy of the classification process, feature extraction techniques are used to extract the most valuable information from the data. Feature extraction improves the precision of classification algorithms by choosing the most pertinent data, facilitating the differentiation of different objects or materials within an image [11]. Additionally, feature extraction can reduce the computational complexity of the analysis, making it more efficient to process large datasets. Feature extraction greatly improves the interpretability of data by reducing its dimensionality. This results in a more efficient recognition of patterns and connections, empowering data analysts to make well-informed decisions with certainty [12]. Various effective methods exist for extracting features, such as PCA (Principal Component Analysis) [13], LDA (Linear Discriminant Analysis) [14], t-SNE (Stochastic Neighboring Embedding) [15], and Autoencoders (AE) [16]. An extremely useful method for precisely determining the subspace of principal components from input vectors is Principal Component Analysis (PCA). Linear Discriminant Analysis (LDA) is a powerful method for distinguishing between different classes in a low-dimensional hyperspace. Nevertheless, these global linear algorithms may exhibit suboptimal performance when confronted with nonlinear scattered data scenarios. To tackle this issue, researchers have proposed utilizing

sophisticated nonlinear feature extraction algorithms for hyperspectral images. These algorithms, such as Isometric Mapping (ISOMAP), Local Linear Embedding (LLE) [17], and Spherical Stochastic Neighboring Embedding (SSNE) [18], are specifically tailored to process individual features, specifically spectral features, as input. Through the application of these methods, scientists can effectively extract more advanced characteristics from hyperspectral images with certainty. The feature extraction processes can be particularly useful in scientific applications, where the goal is to understand the underlying physical processes that are reflected in the image. However, there are also some potential disadvantages to feature extraction. One risk is that important information may be lost during the process, particularly if inappropriate methods are used [19]. Additionally, feature extraction can introduce bias into the analysis if the methods used are not carefully selected or if the results are not validated. When working with big datasets, feature extraction can be especially difficult because of how time-consuming and computationally intensive [20]. To overcome the issue of time-consuming we have used LeNet-5 for extraction of the features from the given data.

Hyperspectral imaging is a highly potent and sophisticated technology that allows for the detection and identification of materials and objects by analyzing their distinct spectral signatures. Hyperspectral imaging surpasses traditional imaging techniques by capturing a wide range of spectral bands, from ultraviolet to infrared, instead of being limited to only three bands. This state-of-the-art technology has demonstrated significant efficacy in diverse domains, such as agriculture, geology, ecology, and medicine. The precision and accuracy of this technology make it indispensable for any application that necessitates the identification and analysis of materials and objects [21]. This increased spectral resolution allows for the identification of subtle differences in the reflectance and absorption of light by different materials, thereby enabling highly accurate classification and analysis of objects in the image [22]. Choosing the suitable bands is a crucial and determining stage in hyperspectral image analysis. This is because not all spectral bands are equally informative for every application. For instance, in mineral exploration, specific spectral bands may be more indicative of certain minerals than others. In agriculture, certain bands may be more useful for detecting crop stress or disease [23]. A novel band selection technique, distinct from conventional methods, has been proposed in the literature. Instead of choosing a limited number of bands, this method promotes specific endmembers for the process of spectral unmixing. Researchers are currently concentrating on optimization-based band selection methods as a means to enhance classification accuracy [24]. This has been accomplished through the extensive use of global optimization algorithms like Genetic Algorithms (GA) [25], Particle Swarm Optimization (PSO) [26], and Firefly Algorithm (FA) [27]. Having said that, the slow convergence speed of the Gravity Search Algorithm (GSA) makes it seem like it might not be the ideal choice for global search

band selection [28]. Although genetic algorithms are great at solving complicated optimization problems, dealing with all of the parameters makes their execution a real challenge. This can cause genetic algorithms to become stuck in local optima at times, but with the right adjustments and preparation, they can perform admirably in many different contexts. This method is not appropriate for global search band selection. Utilizing firefly algorithms results in a low band discovery rate, low accuracy rate, and very slow convergence speed. To overcome the above-mentioned issues, we have used Crow Search Optimization (CSO) to select the appropriate bands from the given data [29]. With the right selection of spectral bands, hyperspectral imaging can be useful in many different contexts. The practicality and usefulness of artificial intelligence have been proven in many different settings. It can detect and track pollution, locate and map mineral deposits, aid in the discovery of previously unknown archaeological sites, and monitor the health of crops and forests, among other things. In each of these applications, hyperspectral imaging provides a wealth of detailed information that is not available through traditional imaging techniques, making it an invaluable tool for scientific research and analysis [30].

Hyperspectral image classification is an advanced procedure that employs state-of-the-art hyperspectral sensors to capture intricate images of the Earth's surface. Hyperspectral images differ from traditional RGB images in that they contain numerous spectral bands, with each band representing a specific narrow range of wavelengths. The technique of image classification is highly dependable and accurate, and it is extensively employed in diverse applications. This means that hyperspectral images can reveal a wealth of information about the surface features of the Earth, including the chemical composition of soils, the health of vegetation, and the presence of minerals. To extract this information, analysts use a range of techniques to classify pixels within the hyperspectral image into different categories. Employing algorithms for data analysis and statistical models for material or object differentiation are effective approaches to achieving desired outcomes [31]. To automate the process of accurately and quickly analyzing massive amounts of data, machine learning and deep learning algorithms are the preferred choices. Multiple studies have demonstrated the efficacy of algorithms like neural networks, support vector machines (SVM), and random forests. We can trust that our data analysis will be done quickly and accurately when you use them. These algorithms use statistical models to identify patterns in the hyperspectral data and classify them into different land cover classes such as forests, water bodies, agricultural lands, and urban areas [32]. The Convolutional Neural Network (CNN) and Deep Learning Algorithm are highly efficient techniques for LULC classification. The concealed layers of CNN adeptly extract the essential characteristics for this task, rendering it the preferred option for precise and dependable outcomes [33]. A Deep-CNN is the optimal method for categorizing Hyperspectral Images (HSIs) according to their

spectral attributes. The 1D-CNN is employed to extract spectral characteristics from hyperspectral images, but it necessitates input in the form of a one-dimensional vector. The 2D-CNN is a dependable approach for classifying hyperspectral images. The design of this system is tailored to consider both the spectral and spatial features of the images, leading to enhanced classification results. The 2D-CNN is a highly suitable choice for hyperspectral image analysis applications because it combines both spectral and spatial characteristics. However, the 2D-CNN might overlook specific relevant information derived from spectral features [34]. To overcome the issues mentioned above, we have used DNAT. By using this we can classify the data.

### A. MOTIVATION AND CONTRIBUTIONS

The in-depth research and observations made by several researchers worldwide have led to an enhanced understanding of the significance of LULC classification analysis. As part of this process, remote sensing scientists collaborated with satellite data providers worldwide to gather real-time satellite images, which were used for the LULC classification. To ensure the accuracy of the classification, publicly available benchmark datasets were pre-processed, and features were subsequently extracted. The meticulous curation of bands is a pivotal element in attaining accurate categorization of hyperspectral images. This study has provided us with a thorough comprehension of the importance of LULC classification analysis, thereby bolstering our confidence in this research field. The method's capacity to enhance the precision of land cover maps is genuinely remarkable. This approach has demonstrated its efficacy in enhancing the precision of these cartographic representations, which serve various critical functions such as land-use allocation, environmental analysis, and natural resource management. The study aims to devise sustainable solutions that benefit both humanity and the environment in the long run. The contribution of our work is to classify LULC using hyperspectral images. We have attempted DNAT for LULC classification using hyperspectral images for the first time. The feature extraction is done using the LeNet-5 architecture to extract the significant features from the LULC data. After extracting the features from the dataset, the extracted features are sent to the band selection. After selecting the informative bands using Crow Search Optimization (CSO), the DNAT classifier will classify the hyperspectral images.

The paper is meticulously structured into discrete sections, each fulfilling a precise objective. Section II contains a comprehensive literature review. In section III, the proposed methodology is elucidated with clarity and confidence. Section IV contains the experimental results, which are presented in a comprehensive and influential manner. The findings are comprehensively examined and analyzed in section V. Ultimately, readers will encounter a cogent and logical conclusion to the paper.

## II. RELATED WORKS

### A. FEATURE EXTRACTION

Extracting meaningful features from hyperspectral images is crucial for correct ground cover analysis. The curse of dimensionality, however, means that classifications utilizing the original data can end up less accurate. This obstacle has been successfully handled by employing robust approaches like feature extraction and selection [35]. This research study presents a novel method for categorizing bands in hyperspectral imaging by utilizing Normalized Mutual Information (NMI) and Minimum Noise Fraction (MNF) to identify distinctive characteristics. The NMI method is employed for feature selection, while the performance analysis is conducted using a kernel Support Vector Machine (SVM). The suggested strategy has exhibited substantial enhancements in classification accuracy and computing efficiency compared to other established methods, as evidenced by the analysis of two actual hyperspectral imaging datasets.

In terms of sampling, an Adaptive Multi-perceptual field implicit Guided Sampling Generative Adversarial Network (AMGSGAN) algorithm greatly improves long-range perception [36]. The three powerful modules that make up the generator are adaptive guided implicit sampling, precoding, and feature extraction from multi-perceptual fields. These modules collaborate harmoniously to provide exceptional outcomes. In addition, the discriminator module employs a CNN architecture exclusively, using robust approaches like batch normalization and adaptive average pooling layers. AMGSGAN has demonstrated superior performance compared to other algorithms at resolutions of 4x, 8x, and 16x, establishing itself as the leading algorithm in the field of AI-powered algorithms.

Superpixelwise Adaptive Singular Spectral Analysis (SpaSSA) is a remarkable method for extracting features from hyperspectral images by trustingly using local spatial information [37]. Each superpixel acquired from an over-segmented HSI undergoes a combination of conventional SSA and 2-D SSA, which are applied with precision. For feature extraction, we employ either 2-D SSA or SSA, with the former being dynamically adaptable to the size of the superpixel and the latter using SSA as a fallback. Regarding computational complexity and classification accuracy, experimental findings on three datasets show that SpaSSA firmly beats both SSA and 2-D SSA. With principal component analysis (PCA), SpaSSA-PCA can surpass many state-of-the-art methods and significantly enhance the precision of land cover analysis.

### B. BAND (FEATURE) SELECTION

The authors of this article [38] presented a state-of-the-art technique called Dimensionality Reduction Method based on Rough Set Theory (DRM-RST) for feature selection using rough set theory. The hyperspectral images was regarded as a decision system, and the process of selecting effective features was conducted based on information entropy. The



performance of DRM-RST on dimensionality reduction was assessed using the Washington D.C. Mall and New York datasets. The findings indicated that DRM-RST successfully deleted around 184-185 superfluous bands, attained an overall accuracy of 94% by the utilization of an SVM classifier, and decreased computation time by almost 85%. Additionally, the study demonstrated that DRM-RST had superior performance in terms of accuracy and stability when compared to other approaches in the context of dimensionality reduction. Ultimately, their research introduces an innovative and advanced technique for reducing the complexity of computations and successfully tackling the problem of high dimensionality.

Band selection is a technique used to reduce the amount of data in HSI for machine learning. It involves selecting important bands and removing redundant and noisy information. In [39], an unsupervised method for band selection and feature extraction has been proposed. The method trains a sub-neural network to identify important bands and project the data to a reduced and informative space. Results show that the proposed method enhances classification accuracy and reduces computational time and data storage requirements compared to other approaches.

A study analyzed the performance of algorithms Random Forest (RF), Rotation Forest (RoF), and Canonical Correlation Forest (CCF) for hyperspectral imagery classification using three datasets [40]. The CCF algorithm achieved the highest classification accuracies with statistically significant improvements compared to RF and RoF. The individual learners in the CCF ensemble were found to be more diverse and accurate.

The authors have proposed a solution to extract invariant features from hyperspectral imagery using the Invariant Attribute Profiles (IAPs) method in [41]. IAPs extract spatial invariant features with filter banks, convolutional kernels, and superpixel segmentation. Invariant behaviors are modeled using a continuous histogram of oriented gradients in a Fourier polar coordinate. The result comprehensively represents spatial-frequency invariant features suitable for HSI classification.

### C. CLASSIFICATION

Innovative multiclass spatial-spectral GAN (MSGAN) is proposed in [42]. Two generators generate spatial and spectral samples in MSGAN. Discriminators also extract spatial-spectral features and output multiclass probabilities. New adversarial objectives are created for multiclass scenarios. The discriminator accurately classifies training samples into their true classes and generates samples into any class with equal probability. Generators deliberately deceive the discriminator. Training the discriminator with the generators improves its performance through adversarial learning. This is done with discriminative samples.

The spectral-spatial random patches network (SSRPNet) is proposed to classify hyperspectral images in [43]. It uses

image random patches as convolution kernels without training. The SSRPNet spectral-spatial feature is combined with shallow, deep, spectral, and spatial features to form a high-dimensional vector. The high-dimensional vector is then classified by graph-based learning algorithms. By randomly selecting a subset of features from a small sample point to build a graph, these algorithms can achieve high classification accuracy.

The spectral-spatial feature tokenization transformer (SSFTT) method is designed to capture spectral-spatial features, as well as high-level semantic features in [44]. Firstly, a module is created to extract low-level features using spectral-spatial feature extraction. This module comprises a 3-D convolutional layer and a 2-D convolutional layer. These layers are used to extract superficial spectral and spatial characteristics. Furthermore, a feature tokenizer is introduced that employs Gaussian weighting to assist in the conversion of features. Subsequently, the transformed attributes are inputted into the transformer encoder module to acquire feature representation and to enhance the process of learning. Ultimately, a linear layer is used to determine the initial token that can be learned in order to obtain the label for the sample.

In [45], an MS-RPNet is a multiscale superpixelwise classification network. Superpixel-based S3-PCA and Random Patches Network-based 2D-SSA are used in this network. The proposed framework uses data-driven and S3-PCA to incorporate global and local spectral knowledge at the super-pixel level. 2D-SSA simultaneously removes noise and extracts spatial features. After random patch convolution and RPNet cascade procedures, the final characteristics are obtained. Layered extraction enhances land cover features by combining spatial information into multi-scale spatial features. The final fusion features are classified using Support Vector Machines (SVM) to produce the final classification results.

A novel classification framework called spatial-spectral Transformer (SST) is introduced for hyperspectral image (HSI) classification in [46]. The proposed SST utilizes a well-structured CNN to extract spatial features. Additionally, a modified version of the Transformer, known as Dense-Transformer, is introduced to capture sequential spectra relationships. Finally, a multilayer perceptron is employed to carry out the ultimate classification task.

The ghost-module architecture, combined with a CNN-based HSI classifier, aims to minimize computational expenses while achieving an efficient and high-performing classification method in [47]. They assess the effectiveness of our new approach by comparing it to nine established HSI classifiers and five enhanced deep-CNN architectures.

In [48], a novel visual attention-based method for classifying HSI is proposed. More precisely, they enhance a ResNet by integrating attention mechanisms to more effectively capture and analyze the spectral-spatial information present in the data. Their novel approach computes a mask that is

utilized to select the most favorable features obtained from the network for classification.

### III. PROPOSED METHOD

Deep learning techniques have shown significant efficacy in the categorization of LULC through the utilization of hyperspectral imagery. However, the presence of limited labels still presents certain challenges. To address these challenges, we have employed Natural Language Processing models to classify the LULC using hyperspectral images. We employed a DNAT to perform LULC classification. Before conducting direct classification, we performed feature extraction using LeNet-5 and band selection using the Crow Search Optimization (CSO) technique. Following the extraction of features, we have chosen specific bands from the extracted features. Subsequently, we have identified the bands that contain relevant information. Once the bands have been chosen, they are forwarded to the classifier for processing.

We have used LeNet-5 for feature extraction in our work because it showcases the efficiency of CNNs in the image classification tasks. It introduced key elements like convolutional layers, pooling layers, and fully connected layers in CNNs, establishing them as essential architectural elements. To select the informative bands, we have used CSO. It is a metaheuristic optimization algorithm that draws inspiration from the intelligent behavior of crows, which is based on nature. The objective is to identify the most effective hyperparameter configurations for a machine-learning model by emulating the collaborative foraging behavior of crows. It attains exceptional performance in training LULC classification models, and surpassing traditional optimization algorithms. And it will enhance the precision, effectiveness, and resilience of the models. To classify the LULC we have used the DNAT classifier, it will identify and address the constraints of conventional LULC classification techniques when applied to hyperspectral images. The selected bands are the input to the classifier. In the DNAT model, the Transformer architecture captures long-range dependencies in spectral information to achieve a more comprehensive understanding. Dilated convolutions incorporated into the attention mechanism enable the model to effectively capture spatial relationships at various scales within local neighborhoods, thus overcoming the constraints of conventional convolutions.

LeNet-5, a pioneering CNN created in the 1990s, marked a notable breakthrough in the empire of image processing. LeNet-5 obviated the necessity of manual feature engineering, a laborious and error-prone process, by directly extracting features from images. The convolutional layers of the network acquired the ability to identify edges, lines, and other fundamental patterns in images, resulting in concise feature representations that efficiently captured pertinent information. Furthermore, LeNet-5 employed subsampling layers to decrease image dimensions while preserving essential characteristics, leading to a notable enhancement

in computational effectiveness. Crows have an impressive capacity to find their way across search locations, according to the CSO awareness system. Consequently, it is anticipated that the subset of bands discovered by this process would outperform those discovered using greedy algorithms. The equilibrium between exploration and exploitation in CSO facilitates the algorithm in evading local optima and converging towards more advantageous solutions. In addition, CSO is relatively robust to initial parameter settings and can effectively handle various types of hyperspectral data.

DNAT is a modern technique used for classifying hyperspectral images. It combines transformer architecture and dilated convolutions. This method enables the detection of intricate textures and spectral patterns in the images by simultaneously capturing both long-range dependencies and spectral patterns. Dilated convolutions play a crucial role in capturing pixel relationships that are far apart in space, which is essential for hyperspectral image classification. Moreover, the transformer architecture empowers DNAT to take into account the complete image while performing classification, resulting in more thorough comprehension of the scene and enhanced differentiation between spectral signatures. Furthermore, DNAT often utilizes stacked encoder-decoder structures that have different dilation rates. This enables DNAT to extract features at various scales and capture both intricate and broad details of the spectral data. Figure 1 depicts the suggested structure of the DNAT-based LULC categorization.

#### A. DATASET DESCRIPTION

##### 1) DATASET CHARACTERISTICS

Hyperspectral imaging is a powerful technology that has found applications in various fields, ranging from remote sensing to medical imaging. When it comes to building models for hyperspectral data cubes, it is essential to take into account several dataset characteristics that can affect the model's accuracy and generalization. These characteristics include the size of the data cubes, which can impact the computational resources required to process them; the spatial and spectral resolutions, which determine the level of detail captured in the images; the noise levels, which can reduce the quality of the data, and any class imbalance that may be present in the data, which can affect the model's ability to distinguish between different classes. By considering these factors and referring to the dataset characteristics presented in Table 1, which can be used to build models that are better suited for real-world scenarios, ensuring accurate and reliable performance.

##### 2) WASHINGTON DC MALL

Providing a precise and comprehensive depiction of the Washington DC Mall area, the Washington DC Mall dataset is a massive compilation of hyperspectral images. The dataset was rigorously generated in 1995 using the most sophisticated HYDICE sensor, guaranteeing the utmost accuracy

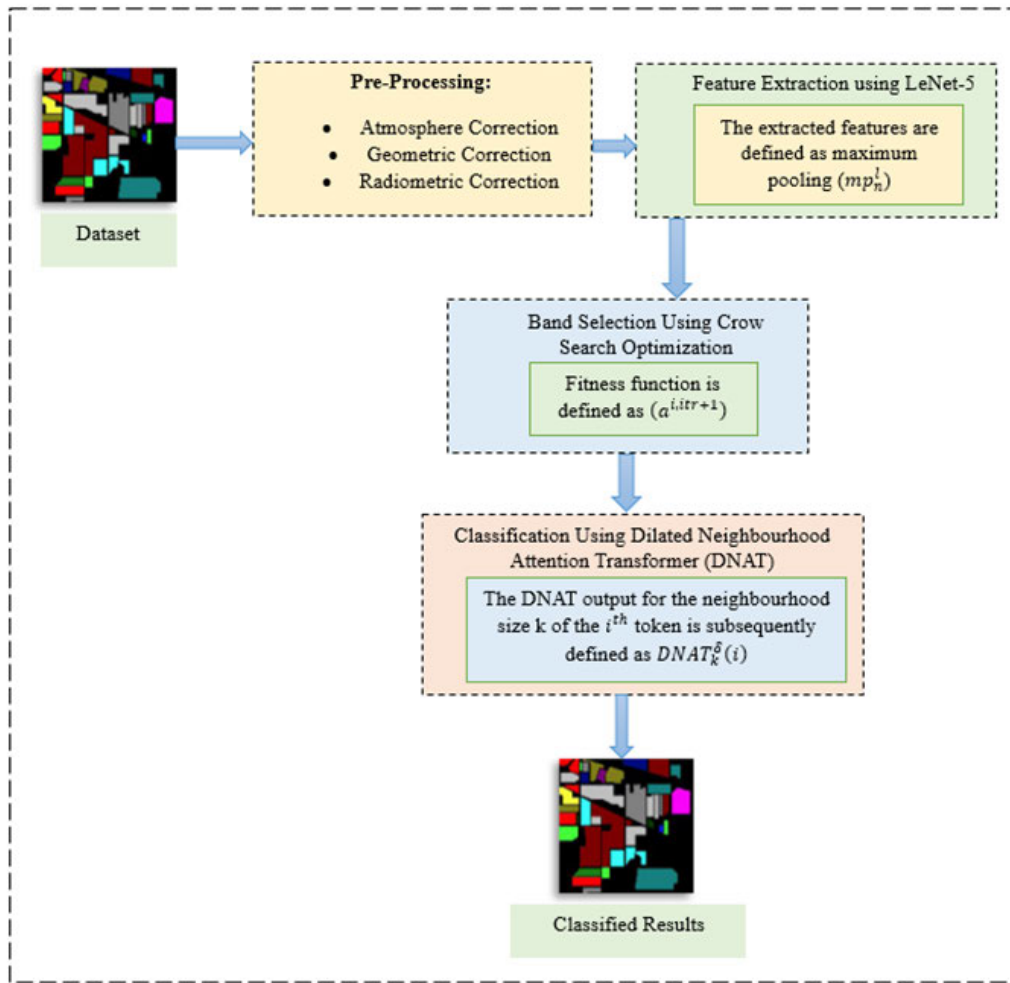


FIGURE 1. Proposed architecture of the DNAT Classifier.

TABLE 1. Dataset characteristics.

	The place of Collection	Equipment used for Acquiring	Categorization Levels	Number of Bands	Band count after denoising	Data Size in Pixels	Spatial Resolution (m)
IP	Indiana, USA	AVIRIS	16	224	220	145*145	20
Salinas	Salinas Valley, California	AVIRIS	16	224	220	86*63	3.7
Washington DC Mall	Washington DC Mall	HYDICE	6	210	191	1280 × 307	-

and precision. The dataset encompasses pixel response in 210 spectral bands, which covers the visible and infrared spectrum range of 0.4 to 2.4  $\mu\text{m}$ , making it an excellent resource for researchers. However, to ensure the accuracy of the data, bands in the 0.9 and 1.4  $\mu\text{m}$  range, which are not transparent due to the atmosphere, were not included, resulting in a total of 191 bands. This dataset comprises 1208 scan lines, each containing 307 pixels, and has a size of approximately 150 Megabytes. The high spatial resolution of 1280 × 307 pixels makes it possible to examine the scene’s detailed features. In addition, the dataset includes a ground truth map, which makes it easier for researchers to identify and classify seven different categories of objects present within the scene. Figure 2 displays the RGB image, Ground Truth, and the number of classes in the dataset about the Washington DC Mall. In Table 2, we can see the dataset details, such as the total number of training and test samples.

TABLE 2. Description of Washington DC Mall dataset.

Name of Classes	Sample size for training	Sample size for testing
Water	335	398
Tree	299	375
Grass	510	1006
Road	203	380
Roof	117	108
Shadow	20	16
Total	1484	2283

### 3) INDIAN PINES (IP)

The dataset was meticulously collected in a northwest Indiana region with a variety of vegetation using a high-tech AVIRIS sensor. Pixels in the high-resolution image are 145 pixels on a side. The image contains 224 distinct spectral bands, with sizes ranging from 0.4 to 2.5  $\mu\text{m}$ , for every pixel. The dataset is organized into sixteen distinct classes; for example, the ‘Oats’ class has twenty-five labeled samples, but the

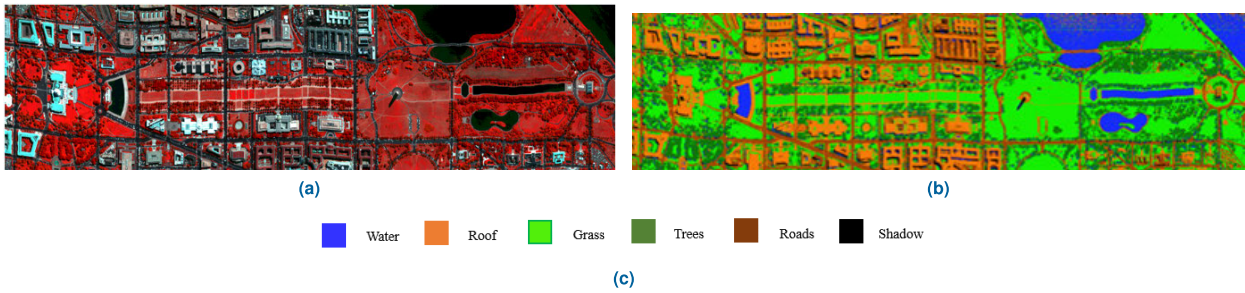


FIGURE 2. Washington DC Mall dataset (a) RGB composite image, (b) Ground truth, and (c) Classes present in the dataset.

TABLE 3. Description of Indian Pines dataset.

S.no	Class	Training	Testing
Land Cover Classes			
1	Alfalfa	5	41
2	Grass-trees	71	659
3	Hay-windrowed	46	432
4	woods	123	1142
5	Grass-pastures	47	436
Land Use Classes			
6	Corn-notill	139	1289
7	Corn	23	214
8	Corn-min	81	749
9	Oats	2	18
10	Soyabean-notill	95	877
11	Soyabean-clean	58	535
12	Soyabean-mintill	240	2215
13	Grass-pasture-mowed	3	25
14	Wheat	20	185
15	Stone-Steel-Towers	9	84
16	Buildings-Grass-Trees	38	348
Total		1000	9249

‘Soyabean-min-till’ class has twenty-four hundred and fifty-five labeled pixels. Figure 3 displays a ground truth reference map and an RGB composite image to provide a thorough visual depiction. Additionally, Table 3 describes the classes in detail, including their training and testing samples, thereby providing a clear and concise framework for analysis.

4) SALINAS

The Salinas dataset is a valuable resource for image analysis that was obtained using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). This dataset contains a total of 16 classes that can be used to identify various features in the images. The images are uniform in size, measuring 512x217 pixels, and have a spatial resolution of 3.7 meters per pixel, providing high-quality details. Out of the 204 spectral bands, only the high-quality bands have been selected to ensure reliable and accurate data. To better understand the dataset, you can refer to Figure 4 for the RGB composite image and ground truth reference map. Additionally, Table 4 provides detailed information about each class, including their training and testing samples.

B. METHODOLOGY

1) LENET-5

LeNet-5 is a specialized version of the CNN that is specifically designed to analyze pre-processed images and

TABLE 4. Description of the Salinas dataset.

S.no	Class Name	Training	Testing
Land Cover Classes			
1	Fallow	11	1965
2	Stubble	22	3978
3	Broccoli_green_weeds_1	11	1998
4	Broccoli_green_weeds_2	20	306
5	Corn_senesced_green_weeds	18	3260
Land Use Classes			
6	Celery	19	3560
7	Lettuce_romainc_4wk	9	1059
8	Lettuce_romainc_5wk	11	1916
9	Lettuce_romainc_6wk	5	911
10	Lettuce_romainc_7wk	6	1064
11	Fallow_rough_plow	8	1386
12	Fallow_smooth	14	2664
13	Vineyard_untrained	40	7228
14	Vineyard_vertical_trellis	10	1797
15	Grapes_untrained	62	11209
16	Soil_vineyard_develop	34	6169
Total		300	53829

extract feature vectors. The LeNet-5 architecture consists of two convolutional layers, two max-pooling layers, and a fully connected layer, making it one of the most extensive CNN models currently accessible. The LeNet-5 algorithm employs a robust method of extracting features by utilizing multiple convolutional kernels in each layer. These kernels are then convolved with the input matrix. LeNet-5 is widely regarded as a highly dependable and effective technique for analyzing images and extracting features. In Figure 5, the architecture of the LeNet-5 is given.

Considering the input data as given in Equation (1)

$$x_2 = x_{2ij} | i = 1, 2, \dots, I, j = 1, 2, \dots, J. \quad (1)$$

In this case, I stand for the input image, and J for the data quantity in the input image. The representation of LeNet-5’s convolution kernel is given in Equation (2)

$$CK = ck_{p,q} | p = 0, 1, \dots, SCK - 1, q = 0, 1, \dots, SCK - 1. \quad (2)$$

A key component of the output from the convolutional layer is the convolution kernel size, which is dictated by the variable SCK. The output of the convolutional layer,



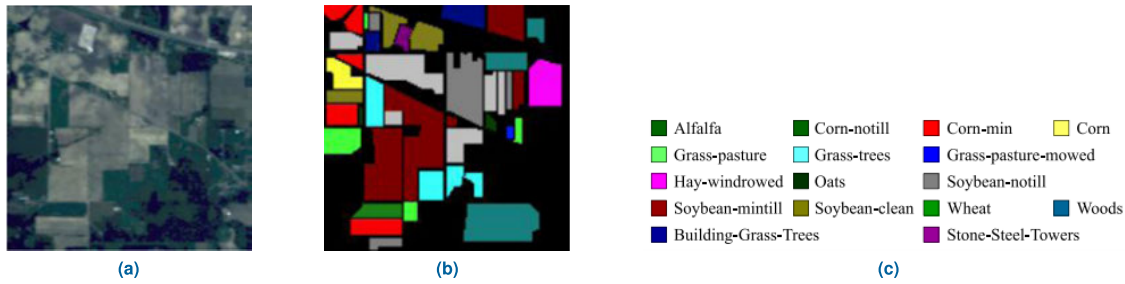


FIGURE 3. Indian Pines dataset (a) RGB composite image, (b) Ground truth, and (c) Classes present in the dataset.

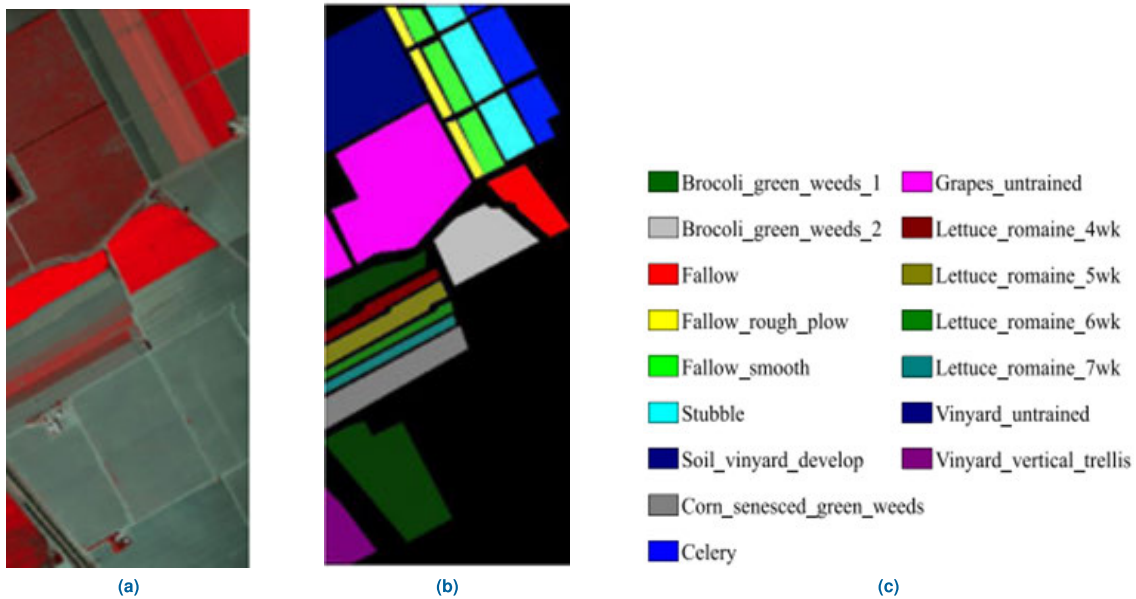


FIGURE 4. Salins dataset (a) RGB composite image, (b) Ground truth, and (c) Classes present in the dataset.

an integral part of the process, is represented by Equation (3)

$$OC_{i,j} = \left\{ f \left( \sum_{p=0}^{SCK-1} \sum_{q=0}^{SCK-1} ck_{p,q} x_{2i+m, j+q} + ot \right) \right\} \quad (3)$$

$i=1,2,\dots,I, j=1,2,\dots,J$

The activation function is denoted as  $f(\cdot)$ , the offset term as ‘ $ot$ ’, and the convolution result as  $OC_{i,j}$ . LeNet-5 employs the Rectified Linear Unit (RELU) activation function, similar to AlexNet.

Using the pooling layer to lower the data’s dimensionality allows feature extraction to be accomplished. In order to get the most valuable points, this pooling layer employs the maximum pooling method. The definition of maximum pooling as  $pool(mp_n^{l-1})$  is given in Equation (4).

$$mp_n^l = pool(mp_n^{l-1}). \quad (4)$$

The preceding layer,  $mp_n^l$  and  $mp_n^{l-1}$ , are represented by  $n$  and  $n^{th}$ , respectively.

Usually, the fully connected layer functions as the ultimate layer of the CNN. Every neuron in the network employs

the RELU activation function, establishing connections with neurons in the previous levels. The purpose of this completely linked layer is to combine and integrate the localized information that helps to differentiate between various classes. Equation (5) represents the outcome of the completely linked layer (l).

$$fc_n^l = f(ck^l \cdot fc_n^{l-1} + ot^l). \quad (5)$$

Combining convolution, pooling, RELU activation, and fully connected layers allows for feature extraction from an input image.

### C. CROW SEARCH OPTIMIZATION (CSO)

CSO is a metaheuristic optimization method that operates on a population-based approach. The underlying principle of CSO is rooted in the collective behavior of crow flocks. Crows reside in communal clusters. They investigate culinary establishments and commit the most exceptional ones to memory. Crows often engage in food site reconnaissance by closely observing and tracking other crows to identify potential sources of food, which they then exploit by pilfering. Furthermore, when a crow perceives the presence of

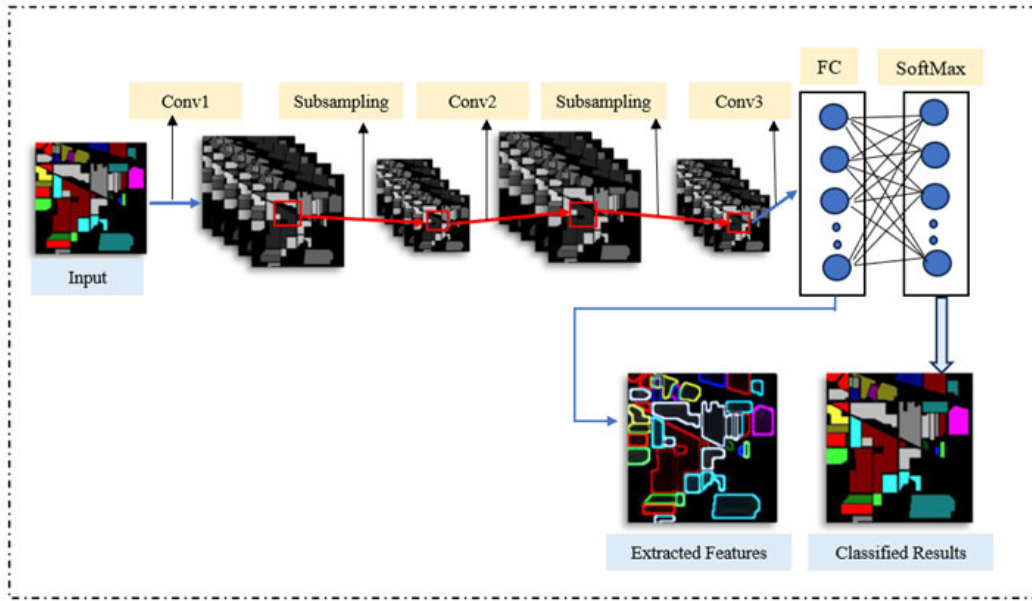


FIGURE 5. Architecture of the LeNet-5.

another crow tailing it, it will promptly escape to an arbitrary destination instead of its usual feeding grounds. In Figure 6, the selected band subset from the IP dataset has been given.

Let's examine a search space with  $d$  dimensions and a population of  $N$  crows. At iteration 'itr', the position of each crow  $i$  is represented by a vector  $a^{i, itr} = [a_1^{i, itr}, a_2^{i, itr}, \dots, a_d^{i, itr}]$ . Each crow symbolizes a prospective resolution for the issue. The variable  $A$  represents the spatial coordinates of all crows, as indicated by Equation (6).

$$A = \begin{bmatrix} a_1^1 & a_2^1 & \dots & a_d^1 \\ a_1^2 & a_2^2 & \dots & a_d^2 \\ \dots & \dots & \dots & \dots \\ a_1^N & a_2^N & \dots & a_d^N \end{bmatrix}. \quad (6)$$

The fitness of each crow's position is evaluated using an objective function. Every crow possesses a memory that retains the location that is presently the most suitable. During iteration 'itr', the memory of each crow  $i$  is represented by a vector  $B^{i, itr} = [B_1^{i, itr}, B_2^{i, itr}, \dots, B_d^{i, itr}]$  valve regurgitation, also known as mitral insufficiency, is a condition where the mitral valve of the heart does not close properly, causing blood to flow backward into the left atrium. This can lead to various symptoms and complications. A matrix, denoted as MEM, stored the memory of all crows and was represented in Equation (7)

$$MEM = \begin{bmatrix} B_1^1 & B_2^1 & \dots & B_d^1 \\ B_1^2 & B_2^2 & \dots & B_d^2 \\ \dots & \dots & \dots & \dots \\ B_1^N & B_2^N & \dots & B_d^N \end{bmatrix}. \quad (7)$$

Essentially, the MEM variable stores the whereabouts of the food that each crow has concealed. At each iteration, the

$i^{th}$  crow generates a new location by randomly following the  $j^{th}$  crow. There are two likely states in this situation:

**Stage 1:** The  $j^{th}$  Crow is clueless to the fact that the  $i^{th}$  Crow is chasing it.

$$a^{i, itr+1} = a^{i, itr} + r_i \times FL^{i, itr} \times (B_j^{i, itr} - a^{i, itr}). \quad (8)$$

The flight length, denoted as FL, is a measure of the distance covered by a crow in a single journey. The variable  $r_i$  represents a random value ranging from 0 to 1.

**Stage 2:** The  $j^{th}$  Crow is apprised of the fact that the  $i^{th}$  Crow is looking for it.

$$a^{i, itr+1} = \text{Any Random Position}. \quad (9)$$

CSO employs a variable known as Awareness Probability (AP) to quantify the probability that a crow is cognizant of being chased. The merging of AP Equation (8), Equation (9) is done as follow Equation (10):

$$a^{i, itr+1} = a^{i, itr+1} + 1 = a_i, itr + r_i FL^{i, itr} B_j, itr - a_i, itr \quad \text{whenever } j > AP_{j, itr} \\ \text{AnyRandomPositionOtherwise} \quad (10)$$

Equation (10) shows that Within the context of the CSO algorithm, when a crow detects the presence of other crows trailing behind it, it relocates itself to a randomly selected location within the search space. Consequently, CSO is more efficient in investigating unexplored domains.

The crow at the  $i^{th}$  position updates its stored information using Equation (11).

$$B^{i, itr+1} = a^{i, itr+1} \text{ if } Fitness^{i, itr+1} \text{ is better than } Fitness(B_i, itr) \\ \text{NoChangeotherwise} \quad (11)$$

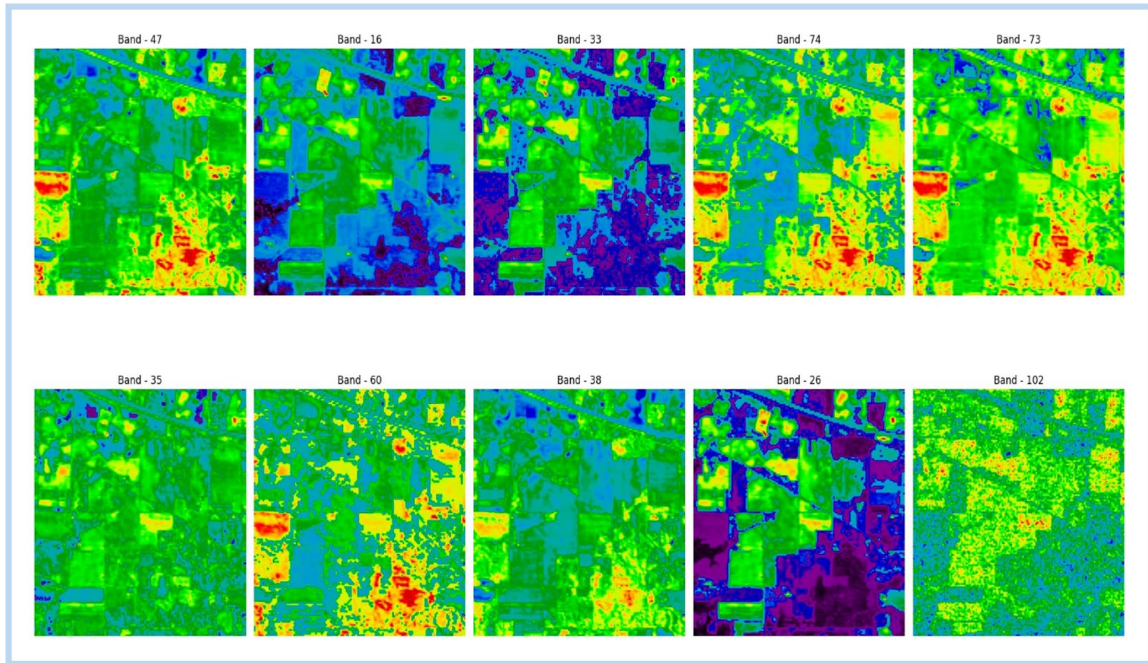


FIGURE 6. Selected bands from Indian Pines dataset.

The fitness function is calculated by using Equation (12)

$$B^{i,itr} = a^{i,itr} + B_{ij} + MEM \quad (12)$$

Here, the fitness function represents the fitness value. It is calculated by using the position of each crow represented by a vector  $a^{i,itr}$ , and distance calculation is done by using Bhattacharya distance calculation is defined by using  $B_{ij}$ , which is given in Equation (13), and the Maximum information between bands, i.e., MEM is also calculated. it is given in Equation (7).

Here,  $B_{ij}$  is the Bhattacharyya distance. It is defined as

$$B_{ij} = \frac{1}{8}(m_i - m_j)^T \left( \frac{\sum_i + \sum_j}{2} \right) (m_i - m_j) + \frac{1}{2} \ln \left[ \frac{|\sum_i + \sum_j|/2}{|\sum_i|^{1/2} |\sum_j|^{1/2}} \right] \quad (13)$$

Here,  $m_i, m_j$  are representing the mean vector of the classes and  $\sum_i, \sum_j$  are representing the class covariance

#### D. NEIGHBOURHOOD ATTENTION(NA)

Neighbourhood Attention (NA) refers to a unique attention mechanism that enables individual pixels to concentrate on their adjacent pixels. This mechanism is commonly referred to as sliding-window attention. Its purpose is to facilitate the localization of self-attention, a potent mechanism that has significantly advanced deep learning. In contrast to Swin's WSA and SWSA, NA uses overlapping sliding windows to function without sacrificing translational equivariance. This implies that the model's output remains unchanged even in the presence of image displacement in the input.

The NA's sliding window pattern bears a resemblance to SASA, but it confines the self-attention mechanism solely to its adjacent neighbors, thereby imposing a direct constraint on self-attention. As the window size of NA increases, it exhibits self-attention. The significance of this lies in the fact that self-attention has demonstrated its efficacy as a potent mechanism in deep learning, enabling models to capture extensive interdependencies present in the input data effectively. Nevertheless, a significant obstacle associated with sliding-window attention is its potential for high computational costs, as there are currently no deep learning or CUDA libraries that offer direct support for these operations. In order to address a particular issue, the implementation of NA was used in conjunction with N ATTEN, an extension including very efficient CPU and GPU kernels. By combining NA with other attention modules like Window Self Attention (WSA) / a pixel-shifted Window Self Attention (SWSA), NA is possible to outperform them in terms of both performance and memory consumption. The Neighborhood Attention Transformer (NAT) is a sophisticated model that effectively utilizes the capabilities of NA to capture local interconnections and the hierarchical architecture of the Swin Transformer to capture global interdependencies. Neural Architecture Transformation (NAT) is a robust and adaptable technique in the field of deep learning that may be efficiently used for a wide range of tasks, including image identification, segmentation, and creation. The NAT and Swin models differ primarily in their approach to downsampling layers. The NAT model employs convolutions with overlapping regions, whereas the Swin model utilizes convolutions with non-overlapping patches. It is safe to say that NAT models have a deeper architecture and smaller inverted bottlenecks



---

**Algorithm:** LULC classification using DNAT

---

**Input:** Pre-processed HSI image  $H_i$

**Output:** Classified HSI image  $C_o$

---

**Step 1: LeNet-5 based Feature Extraction**

$$fc_n^l = \text{LeNet-5}(H_i, mp_n^l, \text{ReLU})$$

Here,  $H_i$  is the input for LeNet-5,  $mp_n^l$  is the pooling layer, ReLU is the activation function.

**Step 2: Band Selection**

$$S_b = \text{CSO}_{B^{i,\text{itr}}}(fc_n^l)$$

$$B^{i,\text{itr}} = a^{i,\text{itr}} + B_{ij} + \text{MEM}$$

The objective function is defined by using  $B^{i,\text{itr}}$ . The position of each crow of  $i$  is represented by a vector  $a^{i,\text{itr}}$ , distance calculation between the bands is calculated by using Bhattacharya distance i.e.,  $B_{ij}$ , maximum information which crow is having i.e., MEM.

Based on the maximum information of the bands & less distance between the bands were selected.

**Step 3: Classification**

$$C_o = \text{DNAT}_{A_v}(S_b)$$

$$A_v = A_i^k, V_i^k, NA_k(i), NA_k(i), P_j^\delta(i), A_j^{(\delta,k)}, V_i^{(\delta,k)}, \text{DNAT}_k^\delta(i)$$

Here,  $A_i^k$  is represented by NA weights, Neighbouring values are given as  $V_i^k$ ,  $NA_k(i)$  is the NA output, Dilated values are defined by  $P_j^\delta(i)$ , DNAT weights are defined by using  $A_j^{(\delta,k)}$ ,  $V_i^{(\delta,k)}$  is the DNAT neighbouring values, and the output of the DNAT is defined  $\text{DNAT}_k^\delta(i)$ .

---

**FIGURE 7.** Algorithm of LULC classification using DNAT.

to get parameter counts and FLOPs that are on par with Swin models. Nevertheless, NAT demonstrates exceptional outcomes in image classification, highlighting the efficacy of overlapping convolutions. Furthermore, NAT demonstrates strong performance on tasks that follow, highlighting its versatility for various applications.

Local attention-based models are highly efficient and effective in a wide range of vision tasks because they possess the ability to preserve locality. Nevertheless, they fail to encompass the global context, which is an essential component of vision. Self-attention is a mechanism that can capture the global context by attending to all positions in a feature map. Receptive fields of localized attention processes are tiny and progressively grow, much like convolutions. In contrast, self-attention utilizes a receptive field that spans the entire size of the input. This enables it to acquire a comprehensive global context, which proves beneficial in diverse vision tasks such as object detection and segmentation. Various studies have investigated the concept of global receptive fields in vision, such as Non-local Neural Networks. Non-local Neural Networks employ self-attention mechanisms to capture dependencies that span across long distances. Nevertheless, operations that have unbounded global receptive fields typically encounter significant computational complexities

in comparison to operations that have restricted receptive fields, which can be either local or sparse. Nevertheless, the investigation of global receptive fields in vision continues to be a vibrant field of study, and further progress in this domain could potentially enhance performance in diverse vision-related tasks.

### E. DILATED NEIGHBOURHOOD ATTENTION TRANSFORMER (DNAT)

This section will provide a detailed explanation of DNAT, which we propose as an expansion of NA. DNAT's flexible architecture can handle a variety of hyperspectral datasets with different resolutions and characteristics. DNAT efficiently extracts local and global contextual information from hyperspectral images using dilated convolutions and attention mechanisms. This combination lets the model extract features at different scales to find complex patterns and subtle details in the data. It uses neighborhood attention to focus on relevant hyperspectral spatial connections. The model can accurately classify LULC by considering neighboring pixels with different dilation rates to capture spatial dependencies at various scales. Dilated convolutions improve model receptivity while reducing computational complexity. DNAT handles hyperspectral images well, making it ideal for large-scale



applications with limited computational resources. We will analyze the consequences of DNAT on the receptive field and then discuss our model, DNAT. In addition, we will offer precise information on how to implement DNAT and seamlessly integrate it with the existing N ATTEN package. The Algorithm of the proposed DNAT classifier for LULC classification is given in Figure 3-C. In Figure 8 (a) Architecture of the Dilated Neighbourhood Attention Transformer(DNAT), (b) DNAT Block

To maintain simplicity, we restrict our notations to single-dimensional NA and DNAT. Let X be an input matrix in  $R^{n \times d}$ , where each row represents a d-dimensional token vector. There are two more linear projections of X that we have: Q for query and K for key. On top of that, there are relative positional biases (B (i, j)) between any two tokens i and j. We multiply matrices in order to determine the neighborhood attention weights, represented as  $A_i^k$ , for the  $i^{th}$  token with a neighborhood size of k. Specifically, we multiply the query projection of the  $i^{th}$  token with the key projections of its k nearest neighboring tokens defined in Equation (14).

$$A_i^k = \begin{bmatrix} Q_i K_{p_1(i)}^T + B_{(i, p_1(i))} \\ Q_i K_{p_2(i)}^T + B_{(i, p_2(i))} \\ \vdots \\ Q_i K_{p_k(i)}^T + B_{(i, p_k(i))} \end{bmatrix}. \quad (14)$$

Here,  $p_j(i)$  represents the  $j^{th}$  nearest neighbor of i. We define neighboring values,  $V_i^k$ , as a matrix where each row represents the k nearest neighboring value projections of the  $i^{th}$  token given in Equation (15).

$$V_i^k = [V_{p_1(i)} T V_{p_2(i)} T \dots V_{p_k(i)} T] T. \quad (15)$$

A linear projection of X, denoted by the symbol V, may be used to represent the Neighborhood Attention output for the  $i^{th}$  token, given a neighborhood size of k. This formulation delineates the procedure through which the attention mechanism analyzes the input data, proficiently capturing the contextual information within the vicinity. The definition of Neighborhood Attention output offers a thorough comprehension of the underlying mechanisms involved, facilitating enhanced analysis and interpretation of the outcomes defined in Equation (16).

$$NA_k(i) = softmax \left( \frac{A_i^k}{\sqrt{d}} \right) V_i^k. \quad (16)$$

The sign  $\sqrt{\quad}$  denotes the mathematical operation of finding the square root. In this particular context, the symbol “d” is used to denote the scaling parameter, whereas it also signifies the embedding dimension. In the context of DNAT, when a dilation value  $\delta$  is provided, we define  $p_j^\delta(i)$  as the  $j^{th}$  closest neighbor of token i that meets the condition:  $j \bmod \delta = i \bmod \delta$ . The  $\delta$ -dilated neighborhood attention weights for the  $i^{th}$  token, with a neighborhood size of k, may be defined as

$A_j^{(k, \delta)}$  in Equation (17).

$$A_i^{(k, \delta)} = \begin{bmatrix} Q_i K_{p_1^\delta(i)}^T + B_{(i, p_1^\delta(i))} \\ Q_i K_{p_2^\delta(i)}^T + B_{(i, p_2^\delta(i))} \\ \vdots \\ Q_i K_{p_k^\delta(i)}^T + B_{(i, p_k^\delta(i))} \end{bmatrix}. \quad (17)$$

We define  $\delta$ -dilated neighboring values for the  $i^{th}$  token with a neighborhood size of k as  $V_i^{(k, \delta)}$  is in Equation (18).

$$V_i^{(k, \delta)} = [V_{p_1^\delta(i)}^T \ V_{p_2^\delta(i)}^T \ \dots \ V_{p_k^\delta(i)}^T] T. \quad (18)$$

The DNAT output for the neighborhood size k of the  $i^{th}$  token is subsequently defined in Equation (19).

$$DNAT_k^\delta(i) = softmax \left( \frac{A_i^{(k, \delta)}}{\sqrt{d_k}} \right) V_i^{(k, \delta)}. \quad (19)$$

DNAT offers a noteworthy architectural hyperparameter: per-layer dilation settings. The maximum value for the dilation parameter is defined as the quotient of  $\lceil \frac{n}{k} \rceil$ , where n represents the number of tokens and k denotes the size of the kernel or neighborhood. The purpose of this is to guarantee the presence of precisely k expanded neighbors for every token. The minimum value is consistently set at 1, which is the same as vanilla NA. Consequently, the dilation value in each layer of the model will be a hyperparameter that depends on the input and can take any integer  $\delta$  within the range of  $[1, \lceil \frac{n}{k} \rceil]$ . The flexibility of receptive fields is achieved through the changeability of dilation values.

## IV. EXPERIMENTAL ANALYSIS

### A. EVALUATION METRICS

Employing a dependable set of assessment criteria is essential when evaluating a LULC classification algorithm. Metrics such as the Kappa coefficient (KC), Overall Accuracy (OA), and Average Accuracy (AA) can yield dependable and precise results. When it comes to testing pixels and using confusion metrics to assess the classifier’s performance confidently, these metrics are widely accepted and considered the gold standard [8].

OA: The overall accuracy metric is considered to be of utmost importance when it comes to measuring the quality of image classification algorithms. This metric represents the percentage of pixels that are accurately classified in an image compared to all the pixels in that image. Its significance lies in the fact that it allows us to evaluate the performance of an algorithm and serves as a benchmark for comparing different methods. Therefore, it’s crucial to ensure that the overall accuracy is as high as possible while developing image classification algorithms. The following Equation (20) depicts how overall accuracy is calculated:

$$OA = \frac{1}{T} \sum_{c=1}^C T_{cc}. \quad (20)$$

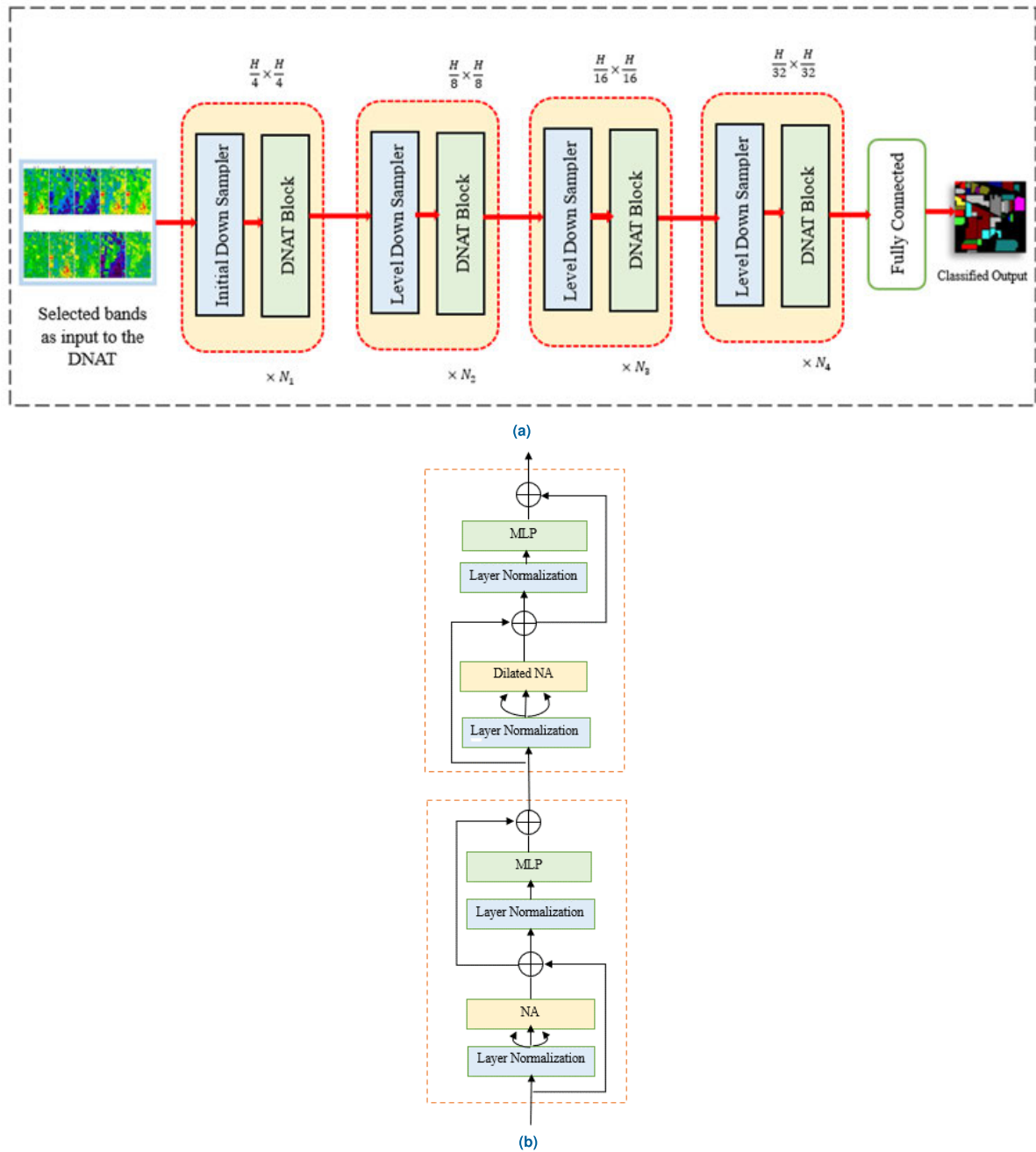


FIGURE 8. (a) Architecture of the Dilated Neighbourhood Attention Transformer (DNAT), (b) DNAT Block.

In this case, T stands for the selected classifier’s confusion matrix and  $T_{cc}$  for the number of testing pixels.

AA: It’s important to note that in image classification, both average accuracy and per-class accuracy are crucial metrics. It is crucial to acknowledge that per-class accuracy is a more precise metric compared to average accuracy, as it quantifies the percentage of accurately classified pixels in each class. Below is Equation (21), we can see the AA equation.

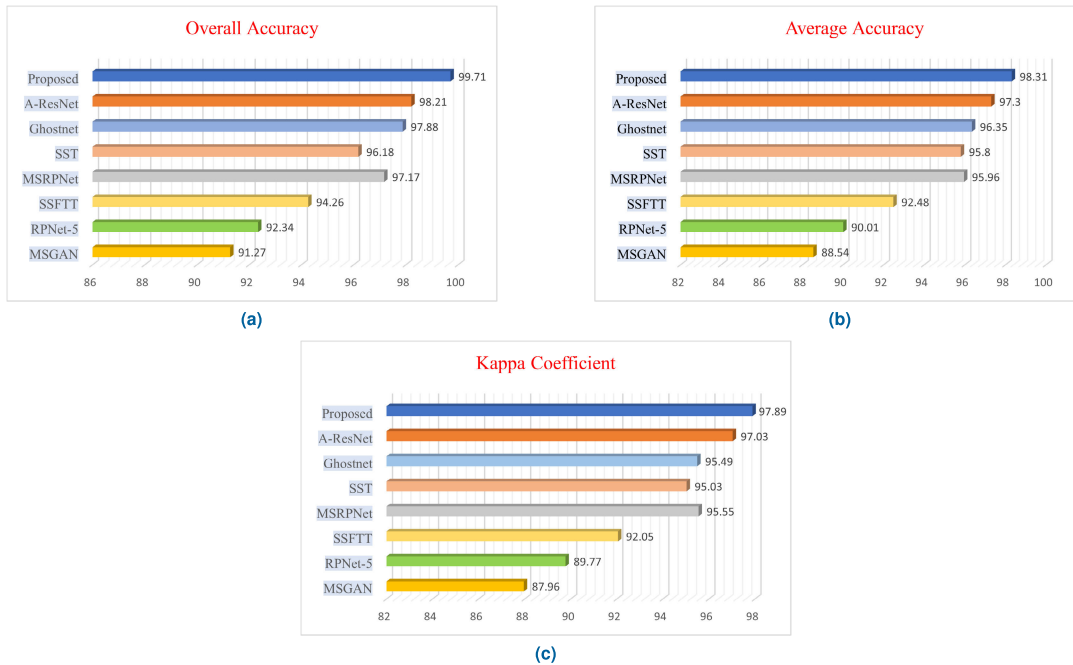
$$AA = \frac{1}{C} \sum_{c=1}^C \frac{T_{cc}}{\sum_{c'}^C T_{cc'}}. \quad (21)$$

Here, T represents the number of difficult pixels, and  $T_{cc}$  signifies the confusion matrix of a given classifier.

KC: The Kappa coefficient is a statistical measure that serves to account for the possibility of chance agreement when evaluating the degree of concordance between two raters or measurements. In essence, this coefficient adjusts the observed agreement (OA) by reducing its value to account for the likelihood that the agreement could have arisen by chance. By taking chance agreement into account, the Kappa coefficient produces a more accurate and dependable estimate of the degree of agreement between the two measures. This,

**TABLE 5.** Class-wise results of the proposed method along with existing systems using the Washington DC Mall dataset.

S.No	Class Name	MSGAN	RPNet-5	SSFTT	MSRPNNet	SST	Ghostnet	A-ResNet	Proposed
1	Water	99.02	94.75	95.27	95.15	95.75	97.97	97.35	98.17
2	Tree	77.95	91.42	93.29	97.73	94.84	93.28	98.73	99.74
3	Grass	91.16	77.59	85.95	96.19	95.62	96.36	95.51	97.28
4	Road	79.05	99.48	98.88	95.57	97.94	99.39	98.05	99.73
5	Roof	86.49	81.96	86.64	96.12	93.51	92.25	98.12	98.08
6	Shadow	97.59	94.88	94.88	95.05	97.17	98.87	96.05	97.16
	OA	91.27	92.34	94.26	97.17	96.18	97.88	98.21	99.71
	AA	88.54	90.01	92.48	95.96	95.80	96.35	97.30	98.31
	KC	87.96	89.77	92.05	95.55	95.03	95.49	97.03	97.89

**FIGURE 9.** Classification accuracy of Proposed method Vs State-of-the-art systems for Washington DC Mall dataset (a) Overall accuracy, (b) Average accuracy, and (c) Kappa coefficient.

in turn, enhances the validity and interpretability of the results obtained from these measures. And it is given in Equation (22).

$$KC = \frac{\frac{1}{T} \sum_c T_{cc} - \frac{1}{T^2} (\sum_{c'} T_{cc'}) (\sum_{c'} T_{c'c})}{1 - \frac{1}{T^2} (\sum_{c'} T_{cc'}) (\sum_{c'} T_{c'c})}. \quad (22)$$

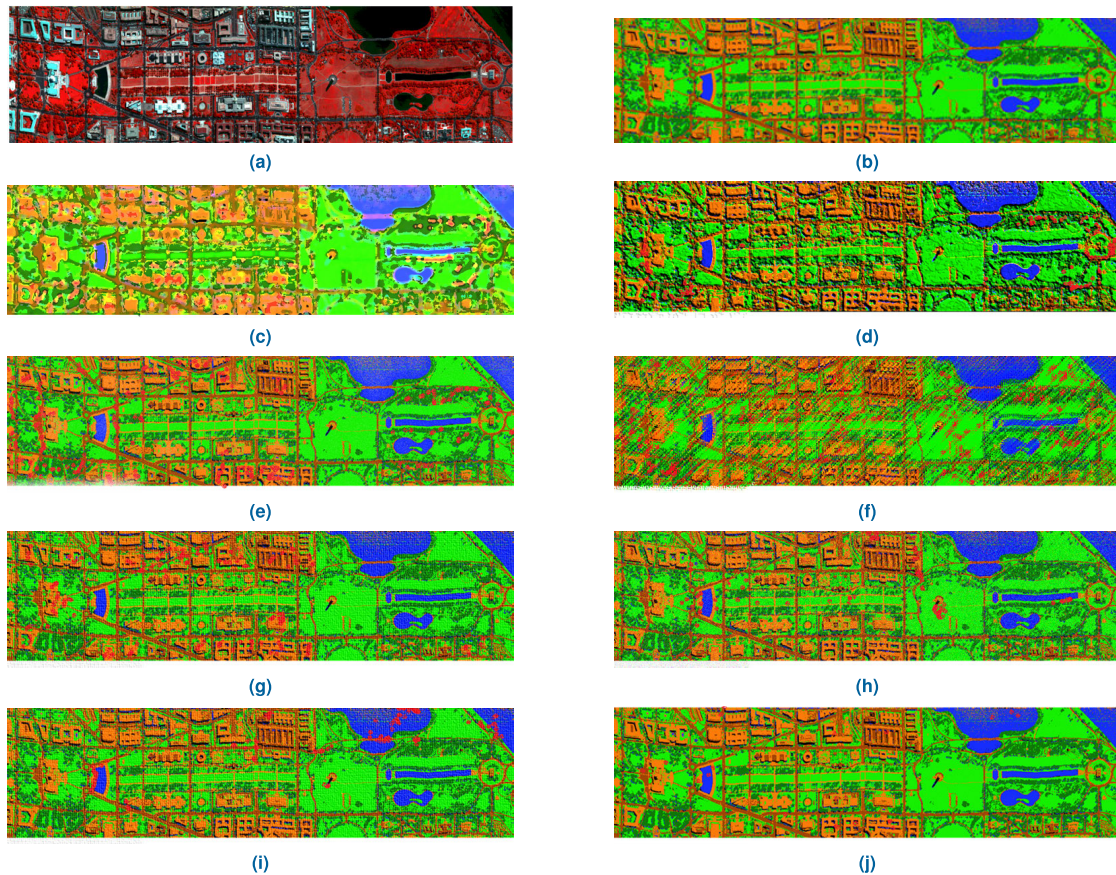
### B. COMPARISON WITH EXISTING TECHNIQUES

To evaluate the precision of the classification achieved using our suggested approach (DNAT), we performed a comprehensive analysis by comparing it with several cutting-edge algorithms. We meticulously assessed the outcomes derived from each algorithm, taking into account multiple metrics and parameters. By undertaking this rigorous procedure, we successfully assessed the efficacy of our suggested approach in attaining the highest level of classification performance. We have used the following algorithms to compare our model effectiveness, and they are MSGAN (Multiclass Spatial-Spectral Generative Adversarial Networks) [42], RPNet (Random Patches Network) [43], SSFTT (spectral-spatial feature tokenization transformer)

[44], MSRPNNet [45], SST (spatial-spectral Transformer) [46], Ghostnet [47], A-ResNet [48]. Table 8 gives the AA, OA, and KC of all the datasets with existing systems.

- 1) MSGAN: The MSGAN model employs a dual-generator approach to produce samples that incorporate both spatial and spectral information. The discriminator has been meticulously designed to extract accurate spectral and spatial characteristics. Consequently, it offers highly reliable probabilities for a significant number of categories. The model establishes novel adversarial goals for the task of multiclass classification. The discriminator is responsible for determining whether training samples are part of the genuine classes or if generated samples are part of any of the classes with equal likelihood. The adversarial learning process involves deliberately designing the generators to challenge the discriminator in a cutthroat battle. By producing high-quality samples that are better able to differentiate between classes, this process greatly enhances the discriminator's classification abilities.





**FIGURE 10.** Classification maps of the Washington DC Mall dataset (a) RGB Image (b) Ground truth (c) SST, (d) Ghostnet, (e) MSGAN, (f) RPNNet, (g) SSFTT, (h) A-ResNet, (i) MSRNet, and (j) Proposed approach.

- 2) RPNNet: The Random Patches Network (RPNNet) method utilizes random hyperspectral image (HSI) patches as convolution kernels without the need for pre-training. RPNNet seamlessly integrates shallow and deep convolutional features, leading to multi-scale analysis that adeptly adjusts to HSI classification, even in the presence of objects with varying scales. We conducted comprehensive experiments on three benchmark HSI datasets to evaluate RPNNet, as well as two variations, RandomNet and RPNNet-single.
- 3) SSFTT: An extremely effective method for obtaining spectral-spatial data and complicated semantic features is the SSFTT. It is based on a cutting-edge spectral-spatial feature extraction module that uses a 2-D and 3-D convolution layer. These layers are highly skilled at extracting low-level features. These layers are highly effective at capturing surface-level spectral and spatial characteristics. Furthermore, a feature tokenizer that utilizes Gaussian weighting is utilized to convert the features. These converted features are then fed into the transformer encoder module to be processed and learned. Ultimately, a linear layer is utilized to ascertain the initial trainable token, which is crucial for obtaining the sample label.
- 4) MSRPNet: This study introduces a new classification network for hyperspectral imaging (HSI) called MS-RPNNet. The MS-RPNNet framework utilizes the Random Patches Network (RPNNet) to integrate multiscale superpixel-wise RPNNet with 2D singular spectrum analysis (2D-SSA). This innovative framework is designed to effectively incorporate data-driven techniques that merge local and global spectral information at the super-pixel level. Projects that necessitate sophisticated spectral analysis greatly profit from its utilization. The implementation of 2D-SSA improves the network's ability to remove noise and extract spatial features. The final characteristics are obtained by employing random patch convolution and other procedures that rely on RPNNet's cascade structure. The process of extracting layered spatial data and creating multi-scale spatial features significantly enhances the characteristics of different land covers. To achieve the most precise classification results, the final step involves utilizing SVM to cluster the amalgamated attributes. The proposed framework presents a viable solution to the HSI classification problem, showcasing its efficacy in achieving enhanced classification performance.
- 5) SST: The spatial-spectral Transformer (SST) is a new classification framework designed specifically for

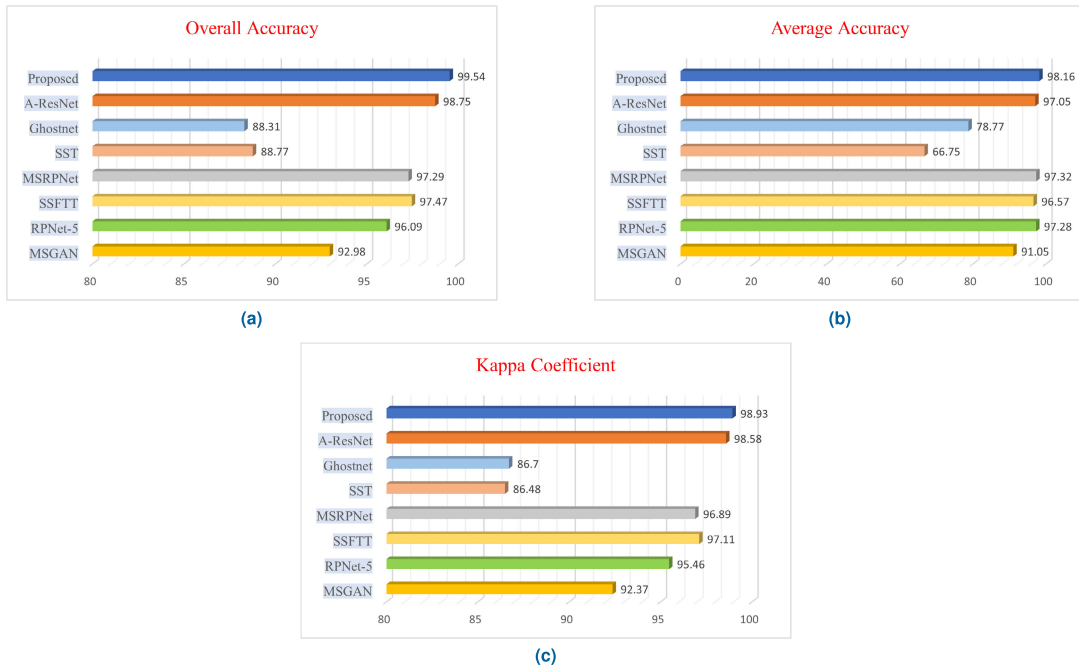


**TABLE 6. Class-wise results of the Proposed method compared with State-of-the-art systems for the Indian Pines dataset**

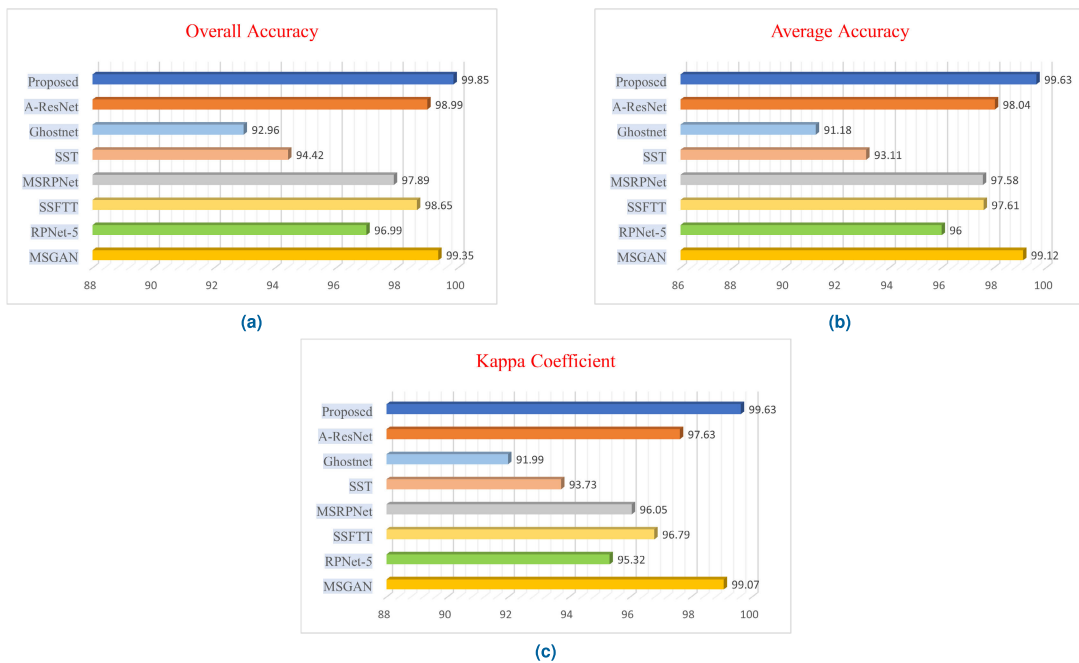
S.No	Class Name	MSGAN	RPNet-5	SSFTT	MSRPNNet	SST	Ghostnet	A-ResNet	Proposed
Land Cover Classes									
1	Alfalfa	82.12	96.25	95.12	86.25	35.40	82.4	89.23	94.26
2	Grass-trees	94.90	94.02	97.67	96.19	93.50	89.19	97.69	96.99
3	Hay-windrowed woods	92.3	96.59	98.87	98.53	82.15	73.42	99.29	98.91
4	Grass-pastures	92.6	99.20	91.55	97.38	97.16	54.95	92.24	94.89
5		89.0	99.00	96.32	97.12	42.96	89.64	99.02	99.54
Land Use Classes									
6	Corn-notill	97.4	99.69	99.54	98.83	90.52	94.46	99.77	99.84
7	Corn	81.1	91.67	100.00	99.01	22.13	0.0	93.04	96.35
8	Corn-min	99.8	99.91	100.00	97.19	50.81	99.52	100.00	100
9	Oats	73.1	100	88.89	96.26	5.88	92.0	90.59	92.28
10	Soyabean-notill	94.1	96.36	97.71	99.21	91.54	93.26	98.57	98.66
11	Soyabean-clean	98.8	93.08	98.69	98.05	70.19	89.88	99.37	99.37
12	Soyabean-mintill	91.8	97.72	98.13	99.32	97.29	80.85	97.14	99.94
13	Grass-pasture-mowed	94.6	99.70	97.28	99.99	42.96	94.25	100.00	100
14	Wheat	98.0	99.21	99.91	99.91	81.24	98.13	99.57	99.84
15	Stone-Steel-Towers	95.8	95.76	98.84	97.32	61.08	49.7	99.58	99.73
16	Buildings-Grass-Trees	81.4	98.37	95.54	97.67	80.15	78.64	99.72	99.99
	OA	92.98	96.09	97.47	97.29	88.77	88.31	98.75	99.54
	AA	91.05	97.28	96.57	97.32	66.75	78.77	97.05	98.16
	KC	92.37	95.46	97.11	96.89	86.48	86.70	98.58	98.93

**TABLE 7. Class-wise results of the Proposed method compared with State-of-the-art systems for the Salinas dataset**

S.No	Class Name	MSGAN	RPNet-5	SSFTT	MSRPNNet	SST	Ghostnet	A-ResNet	Proposed
Land Cover Classes									
1	Fallow	99.92	94.15	97.84	97.73	91.98	82.69	97.51	99.95
2	Stubble	99.49	95.82	96.34	98.79	99.53	87.23	97.94	99.99
3	Broccoli_green_weeds_1	100.0	95.16	97.52	95.14	92.58	98.76	98.01	99.01
4	Broccoli_green_weeds_2	99.12	98.28	96.39	95.98	97.50	95.27	97.27	99.92
5	Corn_senesced_green_weeds	99.57	97.36	98.89	94.67	91.47	99.07	97.68	99.86
Land Use Classes									
6	Celery	99.34	98.89	97.99	94.52	98.65	96.42	98.17	100.00
7	Lettuce_romainc_4wk	99.89	88.73	95.71	97.08	99.87	84.63	94.83	99.88
8	Lettuce_romainc_5wk	98.23	95.19	96.86	98.13	89.80	86.91	96.77	99.77
9	Lettuce_romainc_6wk	99.68	97.99	99.29	99.85	79.54	98.99	99.99	99.99
10	Lettuce_romainc_7wk	99.79	97.69	96.37	98.90	96.37	75.22	99.92	99.92
11	Fallow_rough_plow	99.58	94.28	97.91	99.47	91.98	95.86	99.06	99.60
12	Fallow_smooth	99.29	95.29	99.55	98.99	95.59	96.13	98.27	100.00
13	Vineyard_untrained	99.18	98.74	96.81	98.76	81.42	87.35	97.63	99.76
14	Vineyard_vertical_trellis	98.27	97.19	98.97	98.54	65.07	98.49	99.73	99.94
15	Grapes_untrained	98.41	94.66	99.16	97.98	98.58	99.34	99.84	99.84
16	Soil_vineyard_develop	96.19	96.77	96.19	96.79	99.46	95.59	98.89	99.84
	OA	99.35	96.99	98.65	97.89	94.42	92.96	98.99	99.85
	AA	99.12	96.00	97.61	97.58	93.11	91.18	98.04	99.83
	KC	99.07	95.32	96.79	96.05	93.73	91.99	97.63	99.73



**FIGURE 11.** Classification accuracy of proposed method Vs State-of-the-art systems for Indian Pines dataset (a) Overall accuracy, (b) Average accuracy, and (c) Kappa coefficient.



**FIGURE 12.** Classification accuracy of proposed method Vs State-of-the-art systems for Salinas dataset (a) Overall accuracy, (b) Average accuracy, and (c) Kappa coefficient.

classifying hyperspectral imagery (HSI). SST makes use of a carefully crafted CNN for the extraction of spatial properties and a DenseTransformer, a variant of the Transformer model, for the capture of relationships between sequential spectra. A multilayer perceptron is used to complete the classification task. To further improve the model’s generalizability and combat overfitting, we have integrated a method known as

dynamic feature augmentation into the SST (SST-FA). Furthermore, to tackle the issue of a restricted number of training samples in hyperspectral image (HSI) classification, the technique of transfer learning has been integrated with spectral-spatial-temporal (SST) analysis, resulting in the creation of an alternative classification framework known as transferring-SST (T-SST). In order to mitigate the problem of overfitting

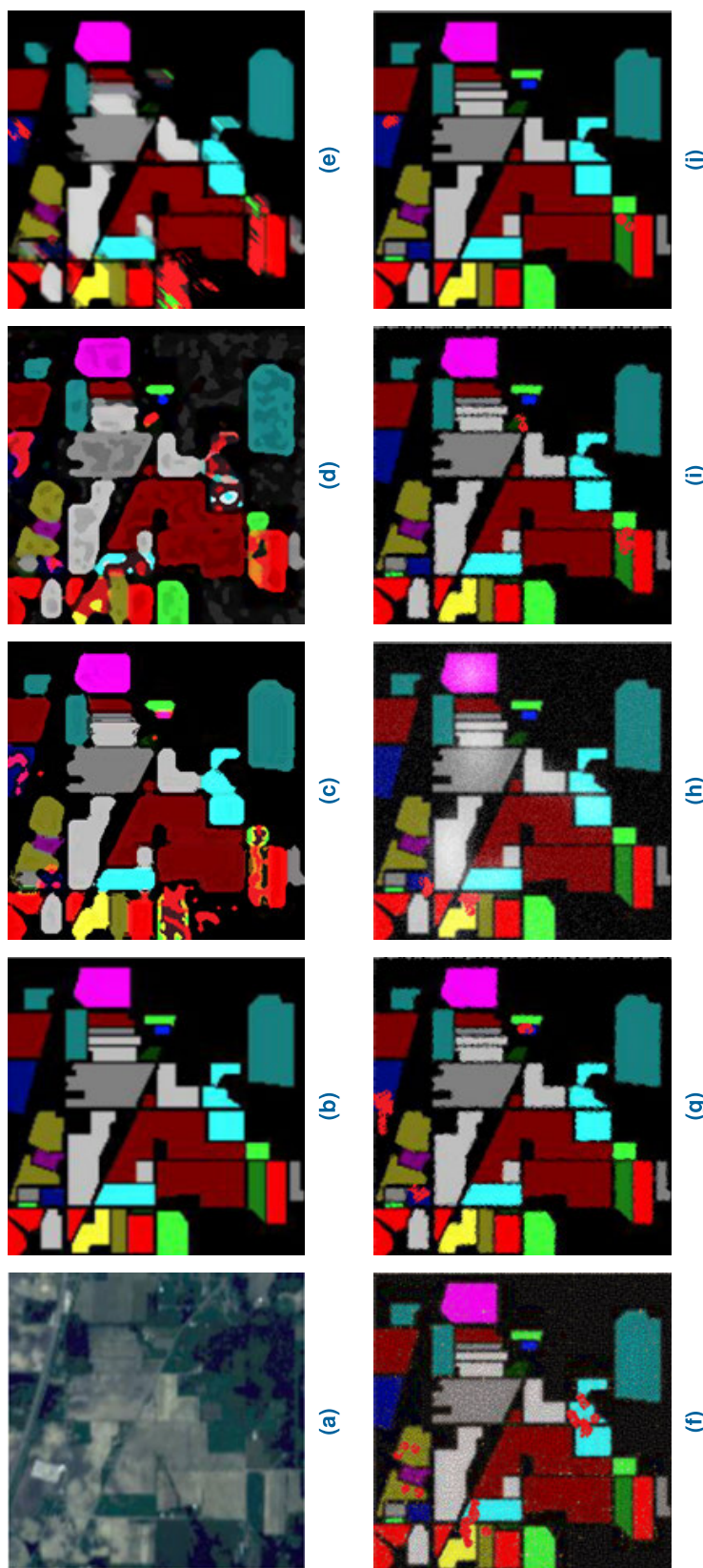
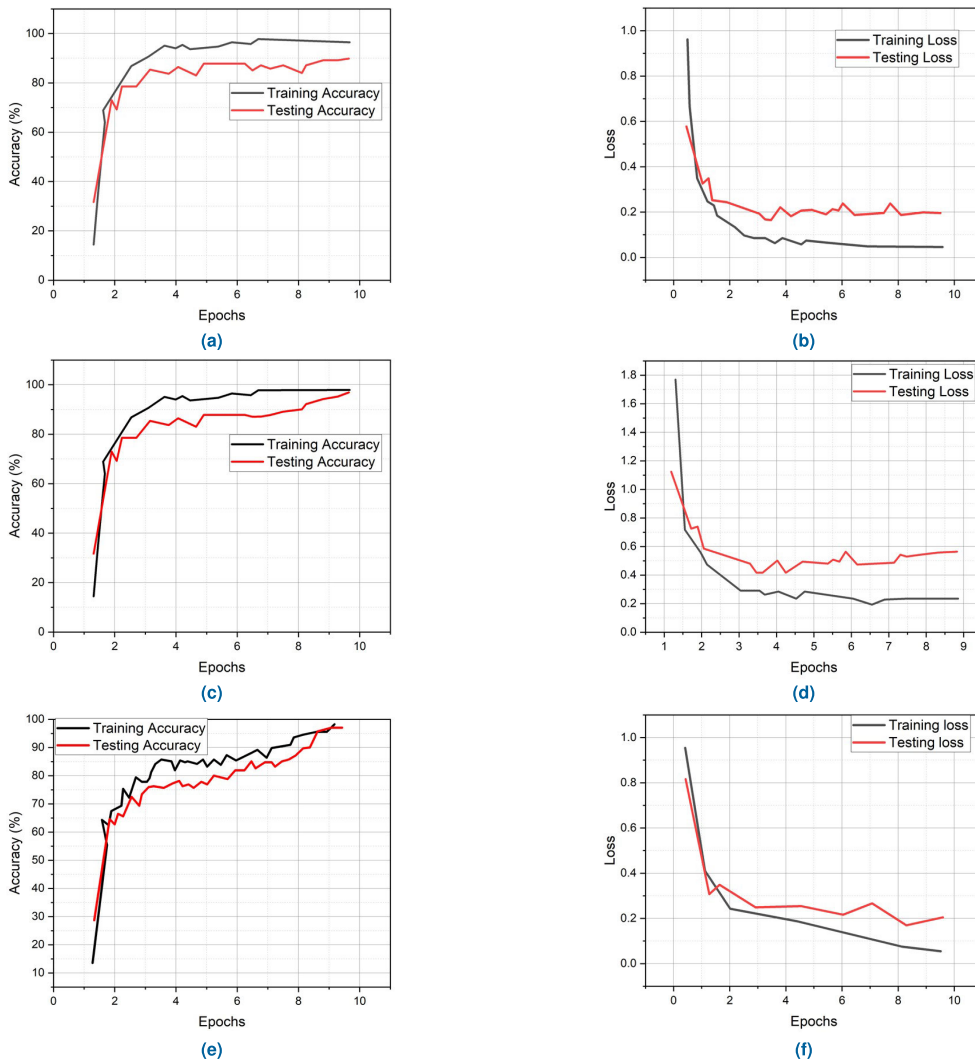


FIGURE 13. Classification maps of Proposed method Vs State-of-the-art systems for Indian Pines dataset (a) RGB composite image (b) Ground truth (c) SST, (d) Ghostnet, (e) MSGAN, (f) RPNNet, (g) SSFT, (h) A-ResNet, (i) MSRNet, and (j) Proposed approach.



**FIGURE 14.** (a) Accuracy (b) Loss of Washington DC Mall dataset, (c) Accuracy (d) Loss of Salinas dataset, and (e) Accuracy (f) Loss of Indian Pines dataset for the proposed method.

and enhance the precision of classification, the T-SST-based classification framework (T-SST-L) proposes the utilization of a method known as label smoothing. The SST, SST-FA, T-SST, and T-SST-L methods are evaluated using three widely used hyperspectral datasets.

- 6) Ghostnet: This study presents a groundbreaking method to greatly decrease the computational complexity of a CNN-based HSI classifier while maintaining a high level of classification accuracy. By combining the ghost-module architecture with the CNN-based HSI classifier, we have effectively improved efficiency and reduced computational costs, thereby establishing the proposed method as a highly effective solution for HSI classification tasks. The proposed method is assessed against nine established HSI classifiers and five enhanced deep CNN architectures, utilizing five widely recognized HSI datasets for the purpose of algorithm benchmarking.

- 7) A-ResNet: This research paper presents a state-of-the-art method for classifying HSI that utilizes visual attention to improve the precision of the outcomes. By integrating attention mechanisms into a ResNet model, their groundbreaking approach significantly improves the data's representation of spectral-spatial information. The suggested technique efficiently generates a mask that, given a set of network features, can identify and choose the most important ones for classification. This method represents a major advance in the field because of how well it produces accurate results.

**C. DISCUSSION**

Using the Washington DC Mall dataset, a comparative analysis was conducted to evaluate the effectiveness of the proposed DNAT method compared to the current techniques. The study found that existing methods frequently failed to accurately depict the different categories. The current methods, such as SST, exhibit lower accuracy in classifying



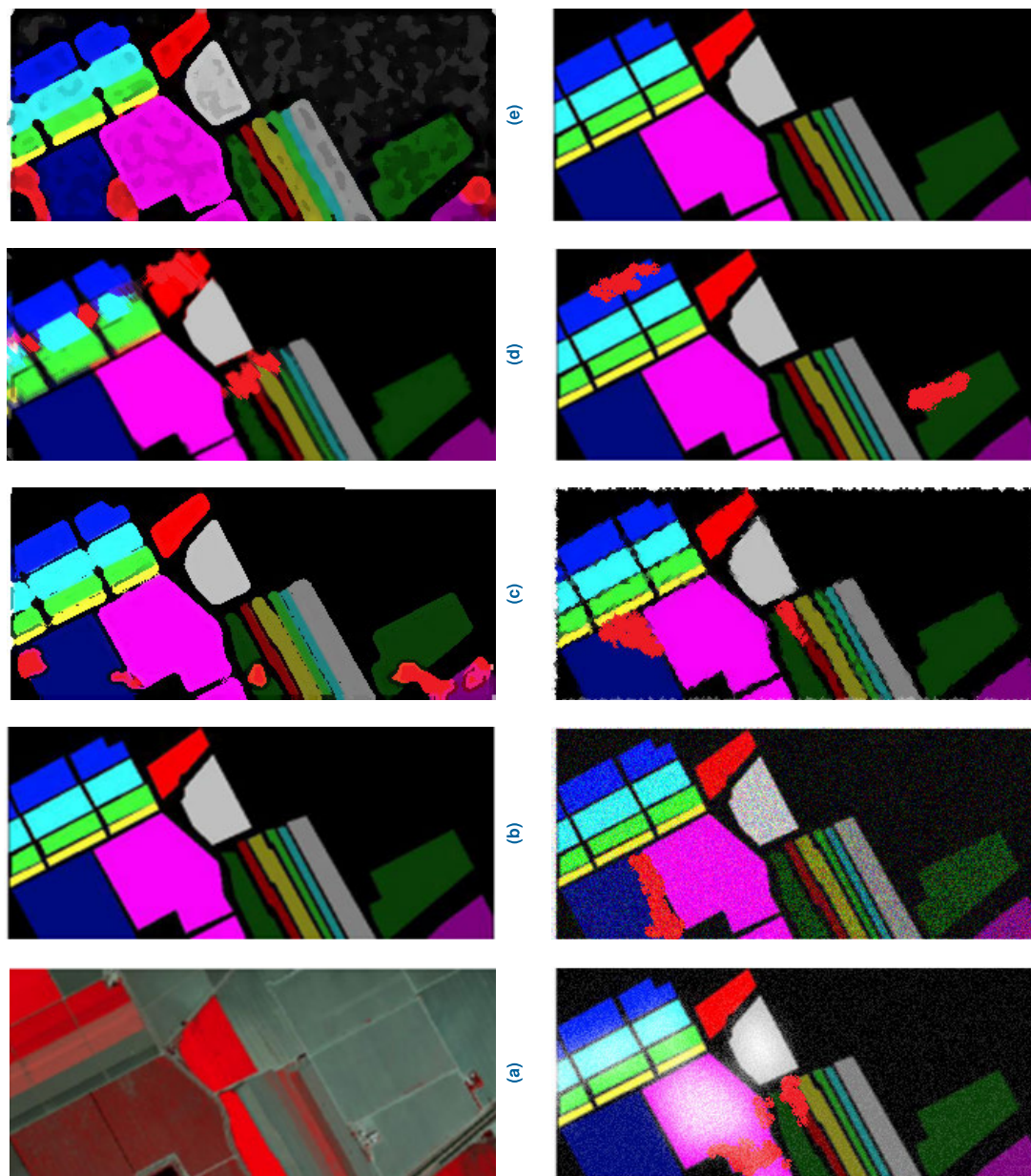


FIGURE 15. Classification maps of Salinas dataset (a) RGB image, (b) Ground truth, (c) SST, (d) Ghostnet, (e) MSGAN, (f) RpNet, (g) SSFIT, (h) A-ResNet, (i) MSRNet, and (j) Proposed approach.

**TABLE 8.** Comparative analysis of the classification results of proposed Vs. State-of-the-art systems.

Algorithm	Classification Index	Washington DC MALL	Indian Pines	Salinas
MSGAN	OA	91.27	92.98	99.1
	AA	88.54	91.05	99.3
	KC	87.96	92.37	99.0
RPNet	OA	92.34	96.09	96.99
	AA	90.01	97.28	96.00
	KC	89.77	95.46	95.32
SSFTT	OA	94.26	97.47	98.65
	AA	92.48	96.57	97.61
	KC	92.05	97.11	96.79
MSRPNet	OA	97.17	97.29	97.89
	AA	95.96	97.32	97.58
	KC	95.55	96.89	96.05
SST	OA	96.18	88.77	94.42
	AA	95.80	66.75	93.11
	KC	95.03	86.48	93.73
Ghostnet	OA	97.88	88.31	92.96
	AA	96.35	78.77	91.18
	KC	95.49	86.70	91.99
A-ResNet	OA	98.21	98.75	98.99
	AA	97.30	97.05	98.04
	KC	97.03	98.58	97.63
Proposed Method	OA	99.71	99.54	99.85
	AA	98.31	98.16	99.83
	KC	97.89	98.93	99.83

roofs and trees. The accuracy of the classes “tree” in Ghostnet, “roof” and “road” in MSGAN, “grass” and “road” in RPNNet, “grass” and “roof” in SSFIT, “grass” and “shadow” in A-ResNet, and “water”, “road”, and “shadow” in MSRNet has been found to be lower compared to the accuracy achieved by the proposed DNAT method. Table 5 displays the classification accuracy of the Washington DC Mall dataset using both existing methods and a proposed method. It indicates that the existing methods have lower accuracy compared to the proposed method. Furthermore, Figure 9 depicts the accuracy graphs of the proposed method in contrast to the existing methods. The classification maps of the Washington DC Mall dataset, both the current systems and the suggested approach, are presented in Figure 10.

Figure 9 is the visual representation of the classification accuracies of the proposed method and the existing methods of the Washington DC Mall dataset. Figure 9 (a) shows that the OA, Defines the percentage of total pixels that are correctly classified. By using our proposed method we have achieved better individual class accuracy, for the class tree we got 99.74% OA. Figure 9 (b) shows that the Average AA is calculated by using the average of the percentage of pixels that are correctly classified per class; we have achieved 98.31% as an AA. Figure 9 (c) represents the KC; it is a quantitative measure of the level of correctly classified pixels by the relationship between the chance agreement and the actual agreement. By using our proposed method, we got 97.89% of KC. Based on these constraints, we have drawn Figure 9 with our proposed system along with existing systems.

Figure 11 displays the classification accuracies of the proposed method and existing methods for the Indian Pines dataset. Figure 11 (a) illustrates that the OA represents the proportion of total pixels that are accurately classified. Through the implementation of our proposed methodology, we have successfully attained improved accuracy for each class. Specifically, we achieved an impressive 100% overall accuracy for the class corn-min, and we have got low-class accuracy for Alfalfa, i.e., 94.26%. In Figure 11 (b), the AA is determined by calculating the average of the percentage of pixels that are accurately classified for each class. In this case, we have achieved an AA of 98.16%. Figure 11 (c) depicts the KC, which is a quantitative metric that assesses the accuracy of pixel classification by comparing the level of chance agreement with the actual agreement. Our proposed method achieved a KC score of 98.93%. By considering all these values, we have drawn a diagram of the proposed system, as well as the existing systems, in the form of Figure 11.

Figure 12 illustrates the classification accuracies of both the proposed method and existing methods for the Salinas dataset. Figure 12 (a) demonstrates that the OA represents the ratio of correctly classified pixels to the total number of pixels. By implementing our proposed methodology, we have achieved enhanced accuracy for every class. Our results show that we achieved a remarkable 100% overall accuracy for the celery class and Fallow\_smooth class. Figure 12

(b) calculates the AA by determining the average of the percentage of pixels that are accurately classified for each class. Our achievement, in this case, is an AA rating of 99.83% Figure 12 (c) illustrates the KC, a quantitative measure that evaluates the precision of pixel classification by comparing the degree of random agreement with the actual agreement. The method we proposed obtained a KC score of 99.73%. After taking into account all of these values, we have drawn a diagram illustrating the proposed system, as well as the state-of-the-art systems, in the form of Figure 12.

Comparing the proposed DNAT approach to established methods using the IP dataset assessed its efficacy. The study showed that current methods often misrepresented diverse categories. Existing methods like SST, Ghostnet, MSGAN, RPNNet, SSFIT, and A-ResNet had lower accuracy in classifying Alfalfa, corn, woods, oats, grass, grass pastures, and Building grass pastures. DNAT outperformed these methods in accuracy and efficiency. Table 6 showed the IP dataset’s high classification accuracy, demonstrating the DNAT method’s superiority over other methods. The accuracy graphs of the proposed method compared to existing methods are shown in Figure 9. In Figure 12, the classification maps of current and proposed systems show the DNAT approach’s clear success. The study concluded that the DNAT method is promising for accurately representing dataset classes. It’s more accurate and efficient than current methods. In Figure 14 (a) Accuracy, (b) Loss of Washington DC Mall dataset has given represented. Figure 14 represents (c) Accuracy, (d) Loss of Salinas dataset, and in Figure 14 (e) Accuracy, (f) Loss of Indian Pines dataset is given. To know the accuracy and loss of our proposed method, we have trained our model with 100 epochs; in the graph, it is given as 1, 2,.. 10 units. Here, 1 unit represents 10 epochs.

The effectiveness of the proposed DNAT approach was evaluated by performing a comparative analysis with established techniques using the Salinas dataset. The study revealed that the current methods often lacked precision in representing various categories. The accuracy of existing methods, including SST, Ghostnet, MSGAN, RPNNet, SSFIT, and A-ResNet, was found to be lower when classifying different categories such as Fallow, Celery, Vineyard\_vertical\_trellis, Lettuce\_romainc\_4wk, Lettuce\_romainc\_5wk, and Vineyard\_untrained. Both the accuracy and efficiency of the DNAT approach were higher than those of these other methods. The proposed DNAT method outperformed the existing methods, as shown in Table 7 that presented the classification accuracy of the Salinas dataset. Figure 13 depicted the classification maps of the current systems and the proposed method, effectively demonstrating the superior performance of the DNAT approach. An additional Figure 10 presented the accuracy graphs of both the suggested approach and current methods, facilitating a thorough evaluation of their respective performance. To summarize, the study concluded that the DNAT approach is a highly promising method for precisely

representing various classes in datasets. It outperforms current methods in terms of both precision and effectiveness.

## V. CONCLUSION

Hyperspectral imaging is an advanced technology that has demonstrated great utility in the classification of LULC. Even though it is effective, the utilization of hyperspectral images for the classification of LULC continues to be fraught with challenges and requires a considerable amount of time. Accurate classification of LULC is difficult to achieve due to the limited number of training samples that are available for hyperspectral images. This is the primary challenge that arises. In order to accomplish the task of LULC classification, we attempted the classification using a Dilated Neighbourhood Attention Transformer, also known as a DNAT. To begin, we make use of LeNet-5 in order to extract features from the data that has been provided. After that, we use CSO to select the informative bands for better performance. Following feature extraction and meticulous band selection, we move on to the next stage of our investigation, namely LULC classification using Salinas, IP, and Washington DC Mall datasets. AA, OA, and KC were the metrics we used to assess the efficacy of our classification strategy. The proposed method achieved the highest level of performance, surpassing the existing methods, according to our experimental results. The proposed model could fail to recognize important contextual details from hyperspectral images if it excessively emphasizes local characteristics. This has the potential to adversely affect tasks such as LULC classification, which requires a thorough understanding of the broader landscape context. By creating computationally optimized versions of DNAT or investigating alternative attention mechanisms that produce comparable outcomes with reduced complexity. To enable the DNAT model to take into account both local and global information, resulting in a greater understanding of the scene.

## REFERENCES

- [1] D. Lorente, N. Aleixos, J. Gómez-Sanchis, S. Cubero, O. L. García-Navarrete, and J. Blasco, "Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment," *Food Bioprocess Technol.*, vol. 5, no. 4, pp. 1121–1142, Nov. 2011, doi: [10.1007/s11947-011-0725-1](https://doi.org/10.1007/s11947-011-0725-1).
- [2] H. Huang, L. Liu, and M. Ngadi, "Recent developments in hyperspectral imaging for assessment of food quality and safety," *Sensors*, vol. 14, no. 4, pp. 7248–7276, Apr. 2014, doi: [10.3390/s140407248](https://doi.org/10.3390/s140407248).
- [3] D. Caballero, R. Calvini, and J. M. Amigo, "Hyperspectral imaging in crop fields: Precision agriculture," in *Data Handling in Science and Technology*, vol. 32. Amsterdam, The Netherlands: Elsevier, 2020, pp. 453–473, doi: [10.1016/b978-0-444-63977-6.00018-3](https://doi.org/10.1016/b978-0-444-63977-6.00018-3).
- [4] G. Tejasree, "Gradient boosting ensemble method for in-vivo brain tumour classification using hyperspectral images," *Indian J. Comput. Sci. Eng.*, vol. 13, no. 5, pp. 1660–1672, Oct. 2022, doi: [10.21817/indjcs/2022/v13i5/221305179](https://doi.org/10.21817/indjcs/2022/v13i5/221305179).
- [5] T. Adão, J. Hruška, L. Pádua, J. Bessa, E. Peres, R. Morais, and J. Sousa, "Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry," *Remote Sens.*, vol. 9, no. 11, p. 1110, Oct. 2017, doi: [10.3390/rs9111110](https://doi.org/10.3390/rs9111110).
- [6] C. Chen, X. Yang, S. Jiang, and Z. Liu, "Mapping and spatiotemporal dynamics of land-use and land-cover change based on the Google Earth engine cloud platform from Landsat imagery: A case study of Zhoushan island, China," *Heliyon*, vol. 9, no. 9, Sep. 2023, Art. no. e19654, doi: [10.1016/j.heliyon.2023.e19654](https://doi.org/10.1016/j.heliyon.2023.e19654).
- [7] C. Chen, J. Liang, G. Yang, and W. Sun, "Spatio-temporal distribution of harmful algal blooms and their correlations with marine hydrological elements in offshore areas, China," *Ocean Coastal Manage.*, vol. 238, May 2023, Art. no. 106554, doi: [10.1016/j.ocecoaman.2023.106554](https://doi.org/10.1016/j.ocecoaman.2023.106554).
- [8] G. Tejasree and A. L., "A novel multi-class land use/land cover classification using deep kernel attention transformer for hyperspectral images," *Earth Sci. Informat.*, vol. 17, no. 1, pp. 593–616, Dec. 2023, doi: [10.1007/s12145-023-01109-1](https://doi.org/10.1007/s12145-023-01109-1).
- [9] M. B. Stuart, A. J. S. McGonigle, and J. R. Willmott, "Hyperspectral imaging in environmental monitoring: A review of recent developments and technological advances in compact field deployable systems," *Sensors*, vol. 19, no. 14, p. 3071, Jul. 2019, doi: [10.3390/s19143071](https://doi.org/10.3390/s19143071).
- [10] A. Vali, S. Comai, and M. Matteucci, "Deep learning for land use and land cover classification based on hyperspectral and multispectral Earth observation data: A review," *Remote Sens.*, vol. 12, no. 15, p. 2495, Aug. 2020, doi: [10.3390/rs12152495](https://doi.org/10.3390/rs12152495).
- [11] Q. Shenming, L. Xiang, and G. Zhihua, "A new hyperspectral image classification method based on spatial-spectral features," *Sci. Rep.*, vol. 12, no. 1, p. 1541, Jan. 2022, doi: [10.1038/s41598-022-05422-5](https://doi.org/10.1038/s41598-022-05422-5).
- [12] M. Prabukumar and S. Shrutika, "Band clustering using expectation-maximization algorithm and weighted average fusion-based feature extraction for hyperspectral image classification," *J. Appl. Remote Sens.*, vol. 12, no. 4, p. 1, Nov. 2018, doi: [10.1117/1.jrs.12.046015](https://doi.org/10.1117/1.jrs.12.046015).
- [13] B. Kumar, O. Dikshit, A. Gupta, and M. K. Singh, "Feature extraction for hyperspectral image classification: A review," *Int. J. Remote Sens.*, vol. 41, no. 16, pp. 6248–6287, 2020.
- [14] B. Liu, X. Yu, P. Zhang, A. Yu, Q. Fu, and X. Wei, "Supervised deep feature extraction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 1909–1921, Apr. 2018.
- [15] B. M. Devassy and S. George, "Dimensionality reduction and visualisation of hyperspectral ink data using t-SNE," *Forensic Sci. Int.*, vol. 311, Jun. 2020, Art. no. 110194.
- [16] W. Yu, M. Zhang, and Y. Shen, "Spatial revising variational autoencoder-based feature extraction method for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1410–1423, Feb. 2021.
- [17] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 441–454, Mar. 2005.
- [18] D. Lunga and O. Ersoy, "Spherical stochastic neighbor embedding of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 857–871, Feb. 2013.
- [19] E. Escobar-Linero, F. Luna-Perejón, L. Muñoz-Saavedra, J. L. Sevillano, and M. Domínguez-Morales, "On the feature extraction process in machine learning. An experimental study about guided versus non-guided process in falling detection systems," *Eng. Appl. Artif. Intell.*, vol. 114, Sep. 2022, Art. no. 105170.
- [20] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, pp. 1–74, Mar. 2021.
- [21] G. Tejasree and L. Agilandeeswari, "An extensive review of hyperspectral image classification and prediction: Techniques and challenges," *Multimedia Tools Appl.*, pp. 1–98, 2024.
- [22] S. S. Sawant, P. Manoharan, and A. Loganathan, "Band selection strategies for hyperspectral image classification based on machine learning and artificial intelligent techniques—Survey," *Arabian J. Geosci.*, vol. 14, no. 7, pp. 1–10, Mar. 2021.
- [23] G. A. Blackburn, "Hyperspectral remote sensing of plant pigments," *J. Experim. Botany*, vol. 58, no. 4, pp. 855–867, Nov. 2006.
- [24] H. Nhaila, A. Elmaizi, E. Sarhrouni, and A. Hammouch, "A novel filter approach for band selection and classification of hyperspectral remotely sensed images using normalized mutual information and support vector machines," in *Information Systems and Technologies to Support Learning*. Cham, Switzerland: Springer, Oct. 2018, pp. 521–530.
- [25] A. Zhang, "Hyperspectral band selection using crossoverbased gravitational search algorithm," *IET Image Process.*, vol. 13, no. 2, pp. 280–286, Feb. 2019.



- [26] D. Saqui, J. H. Saito, A. D. C. Lucio, E. J. Ferreira, D. C. Lima, and J. P. Herrera, "Methodology for band selection of hyperspectral images using genetic algorithms and Gaussian maximum likelihood classifier," in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, vol. 30, Dec. 2016, pp. 733–738.
- [27] M. Zhang, J. Ma, and M. Gong, "Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 773–777, May 2017.
- [28] H. Su, B. Yong, and Q. Du, "Hyperspectral band selection using improved firefly algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 68–72, Jan. 2016.
- [29] S. Valdivia, R. Soto, B. Crawford, N. Caselli, F. Paredes, C. Castro, and R. Olivares, "Clustering-based binarization methods applied to the crow search algorithm for 0/1 combinatorial problems," *Mathematics*, vol. 8, no. 7, p. 1070, Jul. 2020.
- [30] Z. Ting-ting and L. Fei, "Application of hyperspectral remote sensing in mineral identification and mapping," in *Proc. 2nd Int. Conf. Comput. Sci. Netw. Technol.*, Dec. 2012, pp. 103–106.
- [31] W. Lv and X. Wang, "Overview of hyperspectral image classification," *J. Sensors*, vol. 2020, pp. 1–13, Jul. 2020.
- [32] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral-spatial hyperspectral image classification: An overview and new guidelines," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2018.
- [33] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.
- [34] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 277–281, Feb. 2020.
- [35] W. Zhao and S. Du, "SpectralSpatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.
- [36] C. Zhu, R. Dai, L. Gong, L. Gao, N. Ta, and Q. Wu, "An adaptive multi-perceptual implicit sampling for hyperspectral and multispectral remote sensing image fusion," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 125, Dec. 2023, Art. no. 103560.
- [37] G. Sun, H. Fu, J. Ren, A. Zhang, J. Zabalza, X. Jia, and H. Zhao, "SpaSSA: Superpixelwise adaptive SSA for unsupervised spatial-spectral feature extraction in hyperspectral image," *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 6158–6169, Jul. 2022.
- [38] Z. Wang, S. Liang, L. Xu, W. Song, D. Wang, and D. Huang, "Dimensionality reduction method for hyperspectral image analysis based on rough set theory," *Eur. J. Remote Sens.*, vol. 53, no. 1, pp. 192–200, Jan. 2020.
- [39] Q. Frederick, T. Burks, A. Watson, P. K. Yadav, J. Qin, M. Kim, and M. A. Ritenour, "Selecting hyperspectral bands and extracting features with a custom shallow convolutional neural network to classify citrus peel defects," *Smart Agricult. Technol.*, vol. 6, Dec. 2023, Art. no. 100365.
- [40] I. Colkesen and O. H. Ertekin, "Performance analysis of advanced decision forest algorithms in hyperspectral image classification," *Photogramm. Eng. Remote Sens.*, vol. 86, no. 9, pp. 571–580, Sep. 2020.
- [41] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. X. Zhu, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3791–3808, Jun. 2020.
- [42] J. Feng, H. Yu, L. Wang, X. Cao, X. Zhang, and L. Jiao, "Classification of hyperspectral images based on multiclass spatial-spectral generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5329–5343, Aug. 2019.
- [43] Y. Xu, B. Du, F. Zhang, and L. Zhang, "Hyperspectral image classification via a random patches network," *ISPRS J. Photogramm. Remote Sens.*, vol. 142, pp. 344–357, Aug. 2018.
- [44] L. Sun, G. Zhao, Y. Zheng, and Z. Wu, "Spectral-spatial feature tokenization transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.
- [45] H. Chen, T. Wang, T. Chen, and W. Deng, "Hyperspectral image classification based on fusing S3-PCA, 2D-SSA and random patch network," *Remote Sens.*, vol. 15, no. 13, p. 3402, Jul. 2023.
- [46] X. He, Y. Chen, and Z. Lin, "Spatial-spectral transformer for hyperspectral image classification," *Remote Sens.*, vol. 13, no. 3, p. 498, Jan. 2021.
- [47] M. E. Paoletti, J. M. Haut, N. S. Pereira, J. Plaza, and A. Plaza, "GhostNet for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10378–10393, Dec. 2021.
- [48] J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and J. Li, "Visual attention-driven hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8065–8080, Oct. 2019.



**GANJI TEJASREE** received the Bachelor of Engineering degree in computer science and engineering from the Chaitanya Bharathi Institute of Technology (Autonomous, affiliated to Osmania University), Hyderabad, in 2018, and the Master of Technology degree from the Vardhaman College of Engineering (affiliated to Jawaharlal Nehru Technological University Hyderabad), Hyderabad, in 2020. She is currently pursuing the Ph.D. degree with the School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore. Her research interests include image processing, machine learning, and deep learning.



**AGILANDEESWARI LOGANATHAN** received the bachelor's degree in information technology and the master's degree in computer science and engineering from Anna University, in 2005 and 2009, respectively, and the Ph.D. degree from the Department of Software Systems and Engineering, School of Computer Science Engineering and Information Systems (SCORE), VIT Vellore. She was a Professor with the Department of Software Systems and Engineering, SCORE, VIT Vellore.

She has around 19 years of teaching experience and published more than 60 articles in peer-reviewed reputed journals. Her reputed publications include research articles in peer-reviewed journals, namely *Expert Systems With Applications*, *Multimedia Tools and Applications*, *International Journal of Remote Sensing*, *IEEE Access*, *Journal of Ambient Intelligence and Humanized Computing*, and *Journal of Applied Remote Sensing* indexing at Thomson Reuters with an average impact factor of 5. She also published more than 20 engineering books as per Anna University Syllabus. She produced two Ph.D. candidates. Her areas of research interests include image and video watermarking, image and video processing, hyperspectral and multispectral image classification, neural networks, cryptography, fuzzy logic, machine learning, the IoT, information-centric networks, and remote sensing. She is currently a Lifetime Member of the Computer Society of India. She is a Peer Reviewer of journals, including *Neural Computing and Applications*, *IEEE Access*, *Pattern Recognition*, *International Journal of Remote Sensing*, *Array*, *Artificial Intelligence Review*, *Informatics in Medicine Unlocked*, *Neurocomputing*, *Computers*, *Electrical Engineering*, *Journal of King Saud University-Computer and Information Sciences*, *IET Review*, and *Journal of Engineering Science and Technology (JESTEC)*.

...