Deep Learning Network Optimization for Analysis and Classification of High Band Images

Manju Sundararajan[1], S.J Grace Shoba[2], Y. Rajesh Babu[3], P N S Lakshmi[4]
Assistant Professor, Department of ECE, Velammal Engineering College, Chennai, India[1]
Associate Professor, Department of ECE, Velammal Engineering College, Chennai, India[2]
Assistant Professor, Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India[3]
Assistant Professor, Dept of CSE, Aditya College of Engineering, Surampalem, Andhra Pradesh, India[4]

***Abstract*—Examination and categorization of high-band pictures are used to describe the process of analysing and classifying photos that have been taken in many bands. Deep learning networks are known for their capacity to extract intricate information from images with a high bandwidth. The novelty lies in the integration of adaptive motion optimization, spectral-spatial transformer for categorization, and CNN-based feature extraction, enhancing high-band picture search efficiency and accuracy. The three primary parts of the technique are adaptive motion for optimization, spectral-spatial transformer for categorization, and CNN-based feature extraction. Initially, hierarchical characteristics from high-band pictures using a CNN. The CNN method enables precise feature representation and does a good job of matching the image's high and low features. This transformer module modifies the spectral and spatial properties of pictures intended for usage, enabling more careful categorization. This method performs better when processing complicated and variable picture data by integrating spectral and spatial information. Additionally, it is preferable to incorporate adaptive motion algorithms into offering the deep learning network training set. During training, this optimization technique dynamically modifies the motion parameter for quicker convergence and better generalization performance. The usefulness of the suggested strategy is demonstrated by researchers through numerous implementations on real-world high-band picture datasets. The challenges of hyperspectral imaging (HSI) classification, driven by high dimensionality and complex spectral-spatial relationships, demand innovative solutions. Current methodologies, including CNNs and transformer-based networks, suffer from resource demands and interpretability issues, necessitating exploration of combined approaches for enhanced accuracy. In high-band image evaluation and classification applications, the approach delivers state-of-the-art performance and python-implemented model has a 97.8% accuracy rate exceeding previous methods.**

Keywords*—Deep learning networks; Convolutional Neural Network (CNN); spectral-spatial transformer; adaptive motion optimization; high-band image analysis

I. Introduction

Hundreds of consecutive spectrum bands that have elevated resolutions make up the remotely sensed hyperspectral data that is gathered by hyperspectral sensors [1]. Hyperspectral images (HSIs) are being effectively utilized in precision farming, development, monitoring of the environment, and many other disciplines, according to recent research in the field of remote identifying. Among the common features of an HSI is its ability to capture high-resolution one-dimensional spectrum data that describes the physical property in addition to scenario data within the target image's two-dimensional space [2]. Numerous dimensionality reduction techniques are now being used with hyperspectral imagery. Existing approaches can be categorized in two classes based on how much the physical importance of the initial information has been preserved, extraction of features and choosing features (also known as band selection). Through combining several initial characteristics into a single feature, feature extraction finishes the transformation of the initial information from high-dimensional space to low-dimensional space. As everyone is aware, extraction of features works effectively for reducing dimensions, but because spectral architecture is destroyed, it is unable to preserve the physical meaning of each band [3]. The goal of band selection is to maximize given performance indices by selecting a band subset after the unique band set. Band selection, as opposed to feature extraction, can produce a band subset that more accurately captures the original data of the different types of land cover [4].The two main kinds of current band selection techniques are supervised band selection and unsupervised band selection, which involves the application of priori-label data. A great deal of label details is typically needed for supervised band selection, yet in most instances, getting labels for hyperspectral data can be exceedingly challenging [5]. Thus, uncontrolled remote sensing encompasses the majority of current effort. To be more precise, there are three types of current unsupervised band selection techniques: clustering-based, heuristics search-based, and ranking-based [6].

Choosing the type of data has an important effect on the outcomes in the feature classification area. Not every band in the hyperspectral photo has data; certain bands, such as ones impacted by different atmospheric conditions, are meaningless and reduce the efficacy of the classification. Additionally, redundant bands are present, which might impede the process of learning and lead to inaccurate predictions [7]. If the size of space pictures becomes so great it takes numerous instances to identify the connections among each band to the scene, even those bands, which carry sufficient information regarding the situation, might not be able to accurately forecast the categories .The findings in the field of features categorization are greatly affected by the kind of data selected. Each of the groups in the hyperspectral image have data; certain bands are useless and lessen the effectiveness of the classification, which include those affected by various conditions in the atmosphere [8]. Furthermore, duplicate bands exist, which could hinder training and result in wrong projections. When space photos are so big that it requires multiple views to discern the relationships between every band as well as the scene, especially bands that contain sufficient details about the circumstances aren't always able to predict the category with any degree of accuracy (Hughes phenomenon). With the goal to reflect the initial hyperspectral data contained in the raw data, the choice of features algorithms seek to choose an appropriate amount of spectral band information [9]. The goal of feature extraction techniques, as opposed to feature selection techniques, is to use the unprocessed information to create an entirely novel low-dimension descriptors. Like previously said, selecting features techniques outperform extraction of features techniques within the subsequent area. Hyperspectral analysis of images needs entire unprocessed information from the spectrum bands whenever extraction of features techniques is used. When compared with feature extraction techniques, methods for selecting features only use a portion of the raw data's spectrum bands, and by employing those chosen bands rather than every one of the originally selected bands, the expense of acquiring hyperspectral information is further decreased [10].

Even while such methods can yield positive outcomes, the modification of HSIs data can occasionally destroy crucial details, resulting in data degradation. As such, these techniques are not necessarily the best options for reducing dimensionality as compared to conventional band selection procedures. Regarding the band selection approach, given the initial band space of the HSIs, the one that is most useful and unique band subgroup is selected by itself [11]. In a nutshell, the searching method and the assessment of the criteria function are both of the main components of band selection technologies. Although the latter assesses an evaluation for every band subset chosen using the initial selection of bands using a suitable criteria function, the earlier method uses an effective search technique to find an especially prejudiced band subset out of all possible subsets [12].Finding the most effective spectrum combination among all band possibilities might occur at a very high computational expense. Using algorithms to haphazardly hunt for the minimal reduction was a different approach. The two types of band selection techniques that are now in use include supervised band selection and unsupervised band selection [13]. The training information for the previous method must be labelled. It appears that the majority of the time there's a lack of labelled information accessible, which raises the significance of unsupervised band selection for implementations. A new and efficient technique for hyperspectral band selection is presented by the clustering-based techniques. Nonetheless, the primary challenge in the hyperspectral band selecting process is determining whether to determine the separation among the bands and how to choose relevant bands [14].

High-band visuals are essential for environmental monitoring and precision farming, but they can be difficult to interpret because of their redundant bands and complexity. Conventional methods such as band selection and feature extraction are not as good at successfully lowering dimensionality while maintaining spectral information. This research presents a novel method to improve high-band picture processing using spectral-spatial transformers, deep learning networks, and adaptive motion optimization. This methodology seeks to achieve much better classification accuracy than current approaches by merging spectral and geographical data. Real-world applications confirm its effectiveness, employing a Python-based model to achieve accuracy rate. This work offers useful solutions for remote sensing applications and related sectors, making a significant contribution to the advancement of high-band image processing.

Key contributions are as follows,

- Novel integration of convolutional neural networks for hierarchical feature extraction from high-band pictures, enhancing categorization precision and knowledge.
- Introduction of a spectral-spatial converter, augmenting categorization accuracy by leveraging both spectral and spatial information.
- Implementation of adaptive motion algorithms to improve generalization performance and convergence speed, advancing efficiency in image processing.
- Demonstrates state-of-the-art capabilities in image processing and categorization, surpassing existing techniques in precision and computational effectiveness.
- Addresses limitations of conventional approaches, propelling forward the field of remote sensing image analysis, with potential applications in agriculture, urban planning, and environmental monitoring.

The remaining section of this work is structured as follows: Section II covers similar work and a full evaluation of it. Section III offers details on the problem statement. Section IV provides a detailed discussion of the suggested method. Section V presents and examines the results of the tests, as well as a comprehensive comparison of the proposed technique to current standard procedures. Section VI, the last section, represents where the paper is finished.

II. Related Works

Bera and Shrivastava [15] suggest that owing to the special qualities of HSI data, HIS organization is unique of the most difficult tasks in the hyperspectral remote sensing sector. This is made up of a huge amount of bands that exhibit robust interactions in both the spatially and spectrum realms. In addition, it becomes challenging with fewer training examples. To try to tackle these issues, researchers developed a deep convolutional neural network (CNN) based spatially extraction of features method for HSI classification below. They demonstrated the impact of seven distinct optimizers on the deep neural network model through the field of classification using HSI since optimization techniques are crucial for the deep CNN model's development. The study employed seven distinct optimization techniques, the better performance of the deep CNN algorithm using the the Adam optimizer for HSI classification was demonstrated by comprehensive research results on four hyperspectral remote sensing data sets. This endeavour will eventually compare the effectiveness of several optimization techniques classification include increased computational complexity due to processing volumetric data, which may require substantial computational resources and time for training and inference. Features in 3D CNNs may be challenging compared to 2D CNNs, making it harder to understand how the model. The limitations of employing 3D CNN models for hyperspectral image (HSI) classification include increased computational complexity due to processing volumetric data, which may require substantial computational resources and time for training and inference. Additionally, the interpretability of the learned features in 3D CNNs may be challenging compared to 2D CNNs, making it harder to understand how the model is making decisions based on both spectral and spatial information.

Hong et al. [16] explains that the ability to capture tiny spectrum variations, hyperspectral (HS) images allow for accurate recognition of substances. They have been defined by roughly contiguous spectral data. Convolutional neural networks (CNNs) have demonstrated their outstanding capacity for analysing locally contextual data, making them a potent feature extractor in high-spatiality picture categorization. Yet, because of the constraints of their built-in network backbone, CNNs are unable to efficiently extract and record the ordered properties of spectrum signatures. They suggest a unique backbone network named Spectral Former or approach HS image classification via a sequence viewpoint using transformers to address this problem. Spectral Former can acquire spectral localized sequence data from adjacent bands of high-spatiality pictures, going above band-wise depictions seen in traditional transformers, or producing group-wise spectral embedded data. Additionally, researchers develop a cross-layer skip link by continually acquiring the ability to fuse "soft" residues between layers, therefore minimizing the risk of missing crucial data during the layer-wise propagating procedure. This allows researchers to transfer memory-like elements through shallow to deep layers. It is important to note that what was suggested Spectral Former is a very adaptable backbone networks which may be used with input which are patch- or pixel-wise. Through comprehensive trials, researchers assess the suggested Spectral Former's ability to classify on three HS datasets, demonstrating its advantages beyond traditional transformer and attaining significant improvements above state-of-the-art backbone networking. To continue to make the transformers-based design more useful for the HS image classification task, it will look into ways to further improve the system in the future. For instance, they may use attention or autonomous learning. It may also try to create a lightweight transformers-based networks in order to lower the network's complexity without sacrificing effectiveness. To create deeper models that are easier to understand, researchers are also interested to incorporate additional spectral band physical properties and previous understanding about HS pictures into the suggested structure. Further study ought to concentrate on increasing the amount of ignored and linked encoders in the CAF module since it is a significant factor that might potentially improve the suggested Spectral Former's classification accuracy.

Hong et al. [17] states that because convolutional neural networks (CNNs) can record spatial-spectral feature representations, they have gained significant interest in the field of hyperspectral (HS) picture categorization. However, they are still not very good at modelling connections among samples and evaluation, going above the constraints of grid sampling. Throughout this research, researchers conduct a detailed both qualitative and quantitative investigation on CNNs and GCNs with respect to HS image classification. Traditional GCNs are typically quite computationally expensive because they need building an adjacency matrix for each of information, especially for large-scale remote sensing (RS) situations. To accomplish that, researchers create a novel mini-batch GCN (henceforth referred to as miniGCN) that enables mini-batch training of large-scale GCNs. Additionally, this miniGCN can improve the ability to classify and infer information from data that is not sampled with no re-training networks. Moreover, fusing CNNs and GCNs is a natural way to overcome a single model's effectiveness barrier because they may retrieve distinct kind's unique HS features. because miniGCNs may execute batch-wise training of networks (allowing the integration of CNNs and GCNs). Comprehensive tests, carried out on three HS datasets, show that miniGCNs are superior to GCNs and that the tried fusion procedures outperform the single CNN or GCN models. In the years to come, to fully utilize the rich spectrum information found in high-resolution photos, researchers will explore the potential combinations of various deep neural networks with the miniGCNs as well as create more sophisticated fusion modules, such as balanced fusion.

Roy et al. [18] suggest that because of numerous contiguous narrowband built on top of one another, hyperspectral images (HSIs) offer extensive spectral-spatial data. The decision-making process of useful spectral-spatial kernel features can be challenging because of band correlations and disturbance. CNN using fixed-size receptive fields (RFs) are frequently used to overcome issue. Forward and reverse propagations are employed to maximize the network's performance, these techniques cannot allow neurons to efficiently modify RF sizes and cross-channel connections. In order to collect discriminatory spectrum-spatial characteristics for the classification of HSI using an entire training way, researchers describe within this paper an attention-based adaptable spectrum-spatial kernel enhanced residual networks (A2S2K-ResNet) using spectrum focus. Specifically, the suggested network employs a successful feature recalibration (EFR) strategy to enhance the accuracy of classification and trains specific 3-D convolutional kernels in order to simultaneously retrieve spectral–spatial characteristics utilizing enhanced 3-D ResBlocks. In comparison to the current approaches under investigation, , the suggested A2S2K-ResNet tackles the problem of choosing useful spectral-spatial kernel features in hyperspectral image classification. However, significant sensitivity to hyper parameters adjustment and computing burden throughout training could be constraints.

Fu et al. [19] discusses that, since hyperspectral imagery (HSI) contains a wealth of spectrum and spatial information, an entirely novel principal component evaluation (PCA) and segmented-PCA (SPCA)-based multiscale 2-D-singular spectrum analysis (2-D-SSA) combining look at is offered for paired spectrum–spatial HSI extraction of features as well as grouping. At first, the overall spectrum of the items and the relationships between nearby bands are all taken into consideration when applying the PCA and SPCA methods for dimensional spectra reductions. After extracting a wealth of spatial characteristics at various scales from the SPCA dimension-reduced visuals, multiscale 2-D-SSA is used, and PCA is utilized once more to decrease the number of dimensions. Multiscale spectrum–spatial features (MSF-PCs) are created by fusing the acquired multiscale spatial characteristics into the world spectral characteristics obtained by PCA. The support vector machine (SVM) classification is used to assess the obtained MSF-PCs' performance. Tests conducted on four standard HSI data sets showed that, in situations whenever a limited quantity of sample training specimens are accessible, the suggested technique works better than other cutting-edge feature extraction techniques, such as multiple deep learning techniques. Super pixel-alike segmentation may be used in subsequent research to increase the effectiveness of the suggested strategy.

Uddin, Mamun, and Hossain [20] explains along narrow spectral wavelength ranges are employed to capture the hyperspectral remote sensing pictures (HSIs), which are intended to record the most important details of terrestrial objects. Regarding real-world uses, classification accuracy is frequently not cost-effectively acceptable when utilizing the complete original HSI. Band reduction strategies—which may be further subdivided into extraction of features and feature selection methods—are used to improve the classification outcome of high-strength images. The linear unsupervised statistical transformation known as Principal Component Analysis, or PCA, is often used for obtaining characteristics from HSIs. In this article, nonlinear variations of PCA such as Kernel Entropy Component Analysis (KECA) and Kernel-PCA (KPCA) are being studied alongside linear variations. For this aim, KECA uses Renyi quadratic entropy measurement The research study shows that methods for extracting features may be better in classification than utilizing the complete original dataset, even though they are more expensive. While FPCA provides a decent categorization outcome using the least amount of time and space complexity, MNF delivers the best precision in classifying. Future evaluations of the aforementioned PCA-based feature extraction techniques may lead to the proposal of certain likely hybrids techniques, including fusing MNF with SPCA and SSPCA in a manner that executes traditional MNF in place of PCA in SPCA and SSPCA. Additionally, PCA-based methods for extracting features may be used in conjunction with additional information mining techniques like deep extraction of features and spectral-spatial feature extraction to analyse

Researchers have proposed various approaches to address the challenges of hyperspectral image (HSI) classification. One approach involves the use of deep convolutional neural networks (CNNs) with different optimization techniques to extract spatial features efficiently. Another method utilizes transformers, such as the Spectral Former, to capture spectral features sequentially, achieving significant improvements in classification accuracy. Additionally, graph convolutional networks (GCNs) and their fusion with CNNs have been explored to model connections among samples, with miniGCNs demonstrating superior performance. Attention-based adaptable spectrum-spatial kernel enhanced residual networks (A2S2K-ResNet) have been introduced to capture discriminatory spectrum-spatial characteristics effectively. Furthermore, a PCA-based multiscale 2-D-singular spectrum analysis fusion approach has been proposed for feature extraction and classification, showing promising results, particularly in scenarios with limited training samples. Overall, these approaches aim to enhance classification accuracy by leveraging spectral and spatial information effectively while addressing computational complexities and interpretability challenges.

III. Problem Statement

Due to the unique properties of hyperspectral imaging (HSI) data, such as the high dimensionality of spectral bands and the intricate relationships between spectral and spatial information, the area of HSI classification faces several difficulties [19] [15]. To overcome these obstacles and increase classification accuracy, researchers have looked at a number of methodologies, such as deep convolutional neural networks (CNNs), transformer-based networks, attention processes, and fusion procedures [17]. All approaches, however, have their drawbacks, including the requirement for significant processing resources, interpretability problems, sensitivity to hyperparameters, and computational complexity. Furthermore, a key factor in classification effectiveness is the selection of feature extraction methods, such as Principal Component Analysis (PCA) including its nonlinear variants. To create better HSI classification solutions, further study is required to examine combination methods and incorporate them alongside other information mining techniques, even if these techniques show promise for improving classification accuracy.

IV. Proposed Spectral-Spatial CNN with Adaptive Momentum for High Band Image Classification

The proposed Spectral-Spatial CNN with Adaptive Momentum integrates spectral and spatial features for high-band image classification, leveraging adaptive momentum optimization for efficient training and improved performance. This method aims to enhance high-bandwidth image analysis and classification using deep learning techniques. It uses Convolutional Neural Networks (CNNs) for feature extraction, incorporates spectral spatial transformer to capture both spectral and spatial information Furthermore, it uses Adaptive Momentum Optimization Algorithm to enhance the training algorithm to improve performance and efficiency. Overall, this technique optimizes the type of excessive-bandwidth pictures through advanced feature extraction and green optimization strategies. Proposed method framework is shown in Fig. 1.

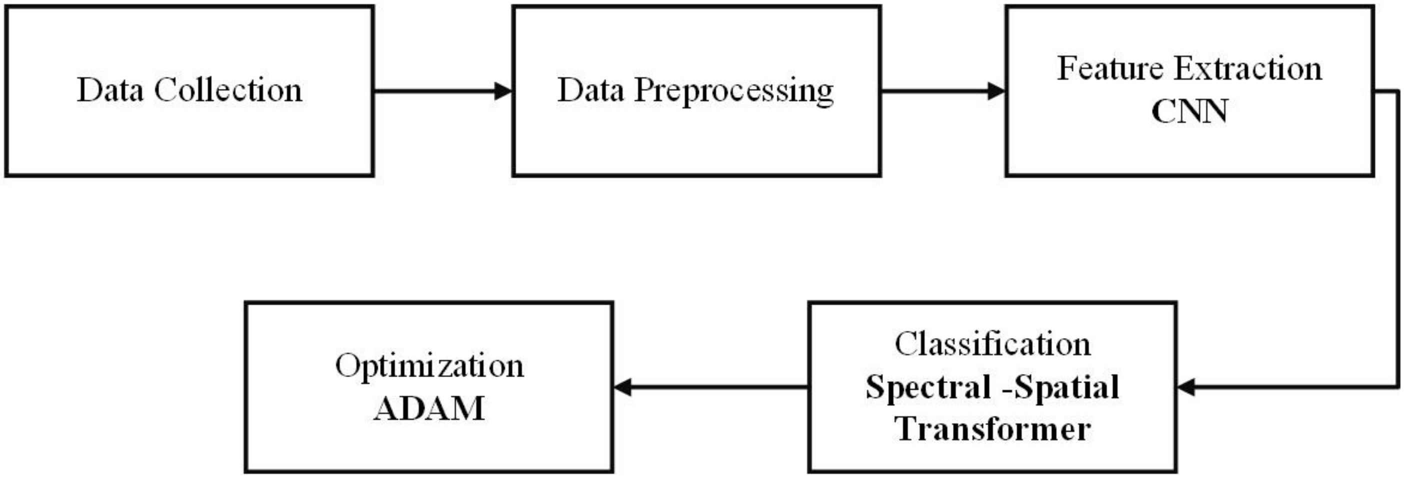

Fig. 1. Proposed spectral-spatial CNN with adaptive momentum for high band image classification.

A. Data Collection

The initial high-resolution images were taken in 1992 over North-Western Indiana, USA, utilizing the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensors. The HS images possesses 145 145 pixels during an average ground sampled distances (GSD) of 20 m and 220 spectral bands having a 10 m spectroscopic resolution, covering its wavelength range of 400 nm to 2500 nm. There are 200 spectral bands remaining, or 1–103, 109–149, and 1642–219, after 20 noise and water absorption bands are uninvolved.16 important groupings in the region beneath investigation were recently looked at. Table II provides a listing of this class names & overall quantity of samples utilized throughout this classifying activity for assessment and training. The geographic distribution of both training and testing sets is shown as well, which helps to duplicate the grouping results [17].

B. Data Pre-processing

Gathering information is a significant step in ensuring that the information is reliable and ready for machine learning analysis. It involves several stages, including Noise Reduction, Data Normalization

1) Noise reduction: Applying filters or denoising techniques to mitigate noise introduced during image acquisition or transmission.

2) Data normalization: Normalization is used to convert data to a similar scale and avoid characteristics with higher values from dominating the study. The following normalization approaches were used:

a) Min-max scaling and z-score standardisation: Min-Max Scaling converts attribute values to a constant scale from 0 to 1, making them comparable. By subtracting the lowest value and dividing by space, it ensures consistency regardless of the original size of the objects. This approach is important for algorithms that are sensitive to input quantities, promoting fair representation in variables. It is especially convenient when dealing with different units or heterogeneous ranges in a data set. To ease convergence, many models' data was normalized to 0 as a mean and 1 as a standard deviation [21].

C. CNN-based Feature Extraction

Comparing to the preceding AlexNet model, VGGNet is a straightforward but efficient model that takes into account the amount of depth of suitable layers without adding additional parameters overall. Consequently, they employed structure akin to VGG. Thirteen layers of convolution and three fully linked layers make up the original VGG's sixteen layers. CNN-Based HSI Spatial Feature Extraction for Every 2.1 The BN layer and Relook operations come between the fourth of the 22 convolutional layers that are used, while the maximal pooling layer is introduced between the third, fourth, seventh, tenth, and thirteenth convolutional layers. CNN is frequently used for applications involving image processing including segmentation, identification, and classifications because of its powerful ability to extract spatial properties from images. For HSI, it has a plethora of spatial information. With CNN, the current study effectively isolates the spatial components of HSI.CNN has many different types of architecture.

Comparing to the preceding Alex Net model, VGGNet is a straightforward but efficient approach that takes into account the size of suitable layers without adding additional variables overall. Consequently, they employed design similar to VGG. Thirteen convolutional layers and three fully linked layers make up the original VGG's sixteen layers. CNN-Based HSI Spatial Feature Extraction for Each 2.1 the BN layer and ReLU operations come between the fourth of the 22 convolutional layers, while the maximum pooling layer is introduced between the subsequent, fourthly, seventh place, a tenth and thirteenth layers of convolution.

It's possible that using all 16 layers isn't the best option for extracting HSI spatial features. The way due to its strong capacity for obtaining spatial characteristics from images, CNN is frequently employed for image processing tasks like segmentation, identification, and classifying. It provides a wealth of geographic data for HSI. This study successfully extracts the spatial components of HSI using CNN. Creating an appropriate CNN framework is essential to creating an effective HSI classification. In the experimentation section, researchers created a deep CNN that is similar to VGG for extracting HSI spatial characteristics. The extracted features from the CNN are then utilized for subsequent processing steps, such as classification. After CNN processes HSI data and extracts features, these features are used in subsequent processing steps such as classification. These extracted features represent the spatial patterns learned by CNN during the training process. Information on the spatial distribution of various features of the HSI data was recorded. Using these features, the subsequent classification algorithm can make a more informed decision about the class labels in the different regions of the HSI, using the spatial information encoded in the features extracted by the CNN has been used [22].

D. Spectral-Spatial Transformer for Classification

The CNN output is handled and sent into a Transformer encoder for band categorization in hyperspectral imaging (HSI). CNN uses a local connection in order to extract adjacent properties from the inputs. HSI often has multiple bands. As a result, CNN finds it challenging to acquire spectrum associations across great distances. The association between each pair of band is obtained through the self-attention process. For instance, there are 224 bands in the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). During the process of learning, a matrix in a form of 224 224 may be generated via attention to oneself. The connection among both bands is represented by every component within the matrix.

The Transformer encoder, the central component of this concept, is the additional component. There are d encoding units in the Transformers encoder, and CNN employs local connections for every encoding unit to obtain nearby characteristics from inputs. Since HSI typically has numerous bands, this is challenging for CNN to determine spectral correlations over great distances. The connection between each pair of band can be obtained by the attention to oneself process. For instance, there are 224 bands in the Airborne Visible/Infrared Image Spectrometer (AVIRIS). During the method of learning, a 224 × 224 matrix may be created via self-attention. The connection among the two bands is represented by every component in the matrix. Comprises of an MLP layer, multi-head attention, layer normalization, and residue connections. Prior to every multi-head focused MLP layer in every encoding block, a normalization level is included and additional connections are planned following each of these layers.

Let n denote model, wherein HS1 ($b'_{1,} b'_{2,} \dots \dots\ b'_{n,}$)dimensionality of the CNN-extracted features, represent the number of n bands of The Transformer's encoder encodes every band as a function of the overall contextual data with the goal of capturing the interactions between all n bands of HSI. In particular, three accessible weighting matrices have been identified: queries (Q), values (V) of dimension dv, and keys (K) of dimension dk. The search query containing all keys is computed using the dot products, and the weights that are placed for each value are subsequently calculated using the function known as softmax. The following is the definition of attention's output is given in Eq. (1)

$$Attention(q,k,v) = softmax(\frac{qk^t}{\sqrt{d'_k}})\ \mathrm{V}, \qquad (1)$$

where, d'_k is the dimension of K.

Projecting the questions, keys, and values multiple times (h times) using distinct and learned projection is advantageous. The outcomes were then combined. This call this method attention in multiple heads. A head is the name given to every outcome of those concurrent attentional calculations is given in Eq. (2)

$$multihead(q,k,v) = concat(head_1, \dots\dots, head_h)w^{o'} \quad (2)$$

The values of the weights the fact that are obtained through the multi-head attention system are then sent to the MLP layer, resulting in 512-dimensional production features. In this case, MLP is made up of two layers that are completely interconnected that have a nonlinearity called the Gaussian error linear unit (GELU) stimulation among it. The ReLU variant known as GELU is described as Eq.. (3)

$$GELU = x'\Phi(x') = x'.\frac{1}{2}[1+\mathrm{erf}(x'/\sqrt{2})] \qquad (3)$$

where, $\Phi(x')$designates the normal Gaussian cumulative distribution function, erf(x') $=\int_0^{x'} e^{'-t'^2}$dt

There is usually a normalization layer preceding the MLP layer that additionally normalizes neurons to shorten the duration of training but additionally solves the disappearing or expanding gradients issue. The normalization of the layer, represented by Eq. (4)

$$\mathrm{a}:a'^{-l}_{I} = \frac{g'^{I}}{\sigma^{I}}.(a_i^{-l} - \mu^{I}) + \mathrm{b'} \qquad (4)$$

where in is the normalized total of the input, and l and l stand for the corresponding expectations and variances at the lth layer. The learnt shifting parameter is denoted by b, and the newly acquired scaling factor by gl. [23].Spectral-Spatial CNN Architecture is shown in Fig. 2.

E. ADAM Optimization

The adaptive learning rate for each parameter used in the gradient-based training process is estimated by the Adaptive Momentum (Adam) method. This is an extremely basic and highly computationally effective method for stochastic optimization which requires limited storage and incorporates first-order gradients. The suggested method is applied to high-dimensional parameter space machine learning problems using large data sets that compute the rate of learning for different parameters separately of assumptions such as initial and second-order aspects. The Adam form of mathematics is as the following Eq. (5) to Eq. (8)

$$y_t = \delta'_1 \times y_{t-1} - (1-\delta'_1) \times h_t \qquad (5)$$

$$x_t = \delta'_2 \times x_{t-1} - (1-\delta'_1) \times {h_t}^2 \qquad (6)$$

$$\Delta_{w'_t} = -\eta' \frac{y_t}{\sqrt{x_{t+\epsilon}}} \times h_t \qquad (7)$$

$$w_{t+1} = w_1 + \Delta w_t \qquad (8)$$

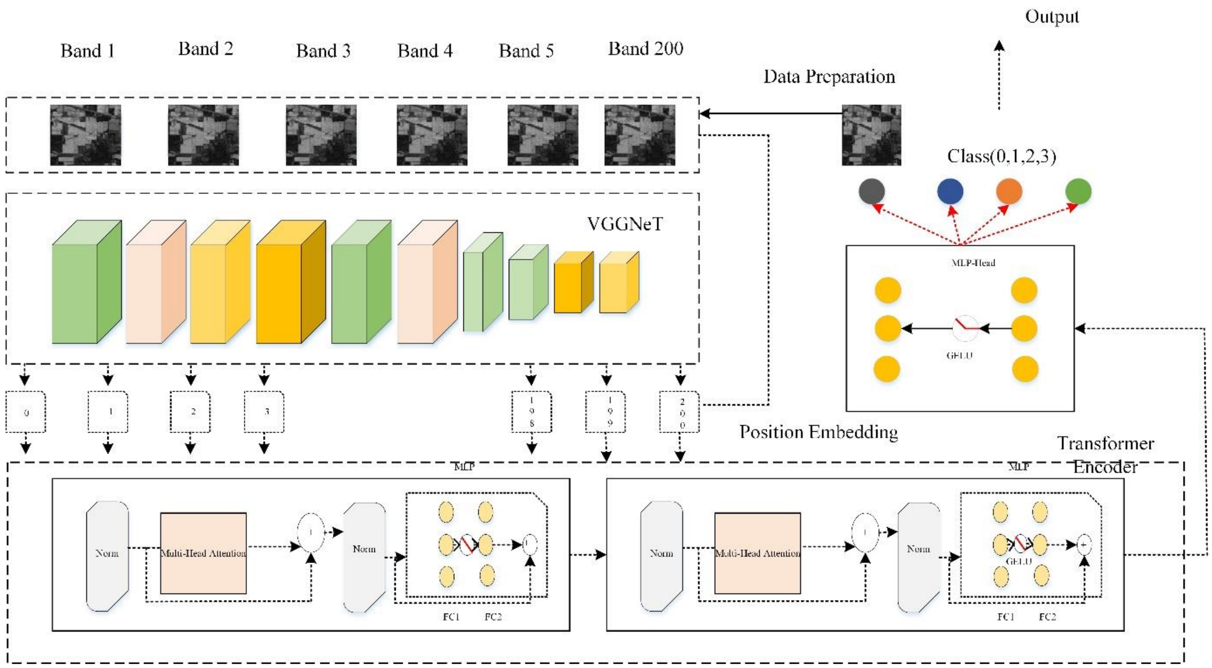

Fig. 2. Spectral-spatial CNN architecture.

where,

η: Preliminary learning rate

h_t: Gradient at time t along h

y_t: Exponential average of gradient along y_t

x_t:Exponential average of squares of gradient along x_t

$\delta'_{1,}\delta'_2$Hyperparameters.

Adam optimizer lowers the overall computing cost, uses fewer resources for execution, and maintains its invariance when gradients are rescaled diagonally. Large data sets, hyper parameters, noisy data, insufficient gradients, and irregular problems requiring minor tweaking are just some of the challenges this resolves. Alpha is the setup parameter for Adam. This is a learning rate or step size; an elevated number (e.g., 0.3) is probably employed because it enables rapid acquisition rather than a lower value and produces flawless results while training.

V. Result and Discussions

In the study, a conducted extensive experiment on real-world high-band picture samples to assess the suggested method's efficacy. The results demonstrate that the method outperforms existing methods in terms of accuracy and performance in high-band image analysis and classification tasks. Specifically, the method achieved state-of-the-art results by leveraging three main components: CNN-based feature extraction, spectral-spatial transformer for classification, and adaptive motion for optimization. Through the use of a convolutional neural network (CNN), it successfully extracted hierarchical features from high-band images, capturing both low and high-level features for detailed representation. Additionally, the incorporation of a spectral-spatial transformer module facilitated more judicious classification by considering both spectral and spatial characteristics of the images. Furthermore, the introduction of adaptive motion algorithms improved the training process of the deep learning network, leading to faster convergence and enhanced generalization performance. Overall, the research contributes to the advancement of remote sensing image analysis by providing a robust framework that enables deep learning networks to optimize classification accuracy and performance for high-band images. Table I represents Classes from the Indian Pines Dataset for every class.

TABLE I. Classes from the Indian Pines Dataset for every class

Class No	Class Name	Training	Testing
1	Grass Pasture	40	1453
2	Oats	40	434
3	Wheat	40	456
4	Corn	20	76
5	Soybean Clean	20	543
5	Grass Trees	10	34
7	Soybean Notill	10	45

Table I presents the classes from the Indian Pines dataset along with their corresponding training and testing sample sizes. The dataset comprises six classes: Grass Pasture, Oats, Wheat, Corn, Soybean Clean, and GrassTrees. Each class is assigned a unique class number, and the table details the number of samples allocated for training and testing within each class. For instance, the GrassPasture class has 40 samples for training and 1453 samples for testing, while the GrassTrees class has 10 samples for training and 34 samples for testing. This information provides a breakdown of the dataset's composition, aiding researchers in understanding the distribution of classes and sample sizes for model training and evaluation.

Taking into account the amount of band used, Fig. 3 displays the Indian Pines dataset's overall precision. The total amount of bands is shown by the X-axis, whereas the overall accuracy ratio is shown by the Y-axis. The diagram shows that accuracy increases as the number of bands grows, but after reaching around five bands, further additions yield diminishing returns while adding more bands initially leads to significant improvements in accuracy, there comes a point where the marginal benefit of additional bands becomes negligible, and further additions may not significantly enhance the complete performance of the classification archetypal.

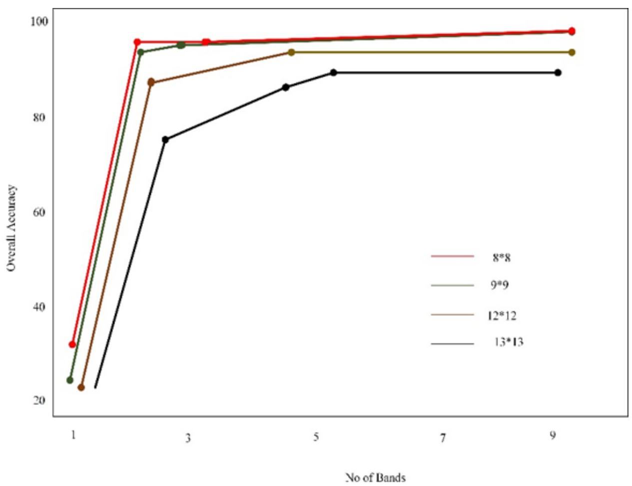

Fig. 3. Overall accuracy of the Indian pines.

Fig. 4 demonstrates the percentage of samples utilized across three distinct datasets: Pavia, Indian Pines, and Houston 2013. It highlights that as the percentage of samples used upsurges, the complete accuracy tends to rise for all datasets. However, each dataset exhibits unique characteristics, resulting in varying rates of accuracy improvement. Pavia dataset shows a steady increase in accuracy with sample percentage, while Indian Pines initially demonstrates rapid gains followed by a plateau, and Houston 2013 exhibits a more gradual increase. These nuances emphasize the importance of dataset-specific considerations when determining the optimal sample size for achieving maximum accuracy in classification tasks.

Fig. 5 illustrates the convergence behaviour of the Adam optimization algorithm over iterations during the training of a machine learning model. Initially, the loss decreases rapidly as the algorithm adjusts the model parameters using adaptive learning rates and momentum. As training progresses, the rate of improvement slows down, indicating convergence towards a minimum point. The graph may exhibit fluctuations due to the adaptive nature of the algorithm, but overall, it demonstrates a consistent decrease in loss over time. The stability and efficiency of Adam optimization make it a popular choice for training deep neural networks. The X-axis represents the steps or iterations in the optimization process, ranging from -1 to 3, while the Y-axis indicates the value of the loss function, ranging from 3.0 to 8.0. This graph depicts how Adam Optimization efficiently navigates through the loss landscape, with blue dots indicating specific points on its path, demonstrating its effectiveness in finding optimal solutions in machine learning model.

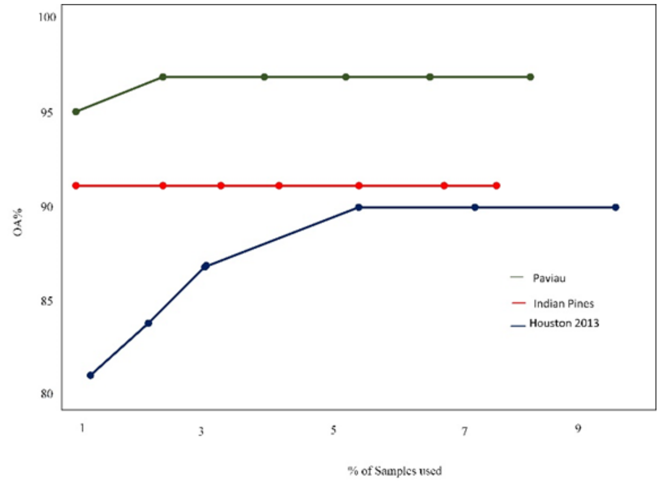

Fig. 4. Samples utilized across three distinct datasets.

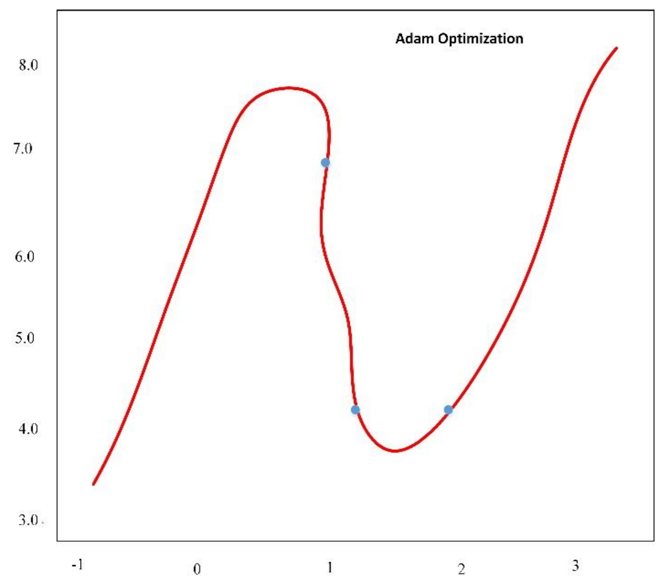

Fig. 5. ADAM optimizer.

Fig. 6 visually signifies the performance of a model during training and evaluation. The x-axis typically indicates epochs or iterations, while the y-axis represents the loss metric. As training progresses, the training loss typically decreases, indicating improved model fit to the training data. Meanwhile, the testing loss, often evaluated on a separate validation set, can help assess generalization performance; ideally, it should decrease initially but stabilize or increase if overfitting occurs, forming distinct patterns aiding in model diagnosis and optimization.

The Receiver Operating Characteristic (ROC) curve is shown in Fig. 7. It serves as an illustration depiction of a binary classification algorithm's effectiveness. Plotting the False Positive Rate (1 - Specificity) versus the True Positive Rate (Sensitivity) at various thresholds is what it does. A ROC curve which approaches the top-left region of the chart, signifying a high degree of sensitivity along with a small positive error rate, suggests an ideal classifier. Greater values for AUC indicate improved class discrimination. The region underneath the ROC curve (AUC) measures the classifier's overall efficacy. When it comes to the compromise among the two qualities in tasks involving classification, ROC modelling offers insightful information.

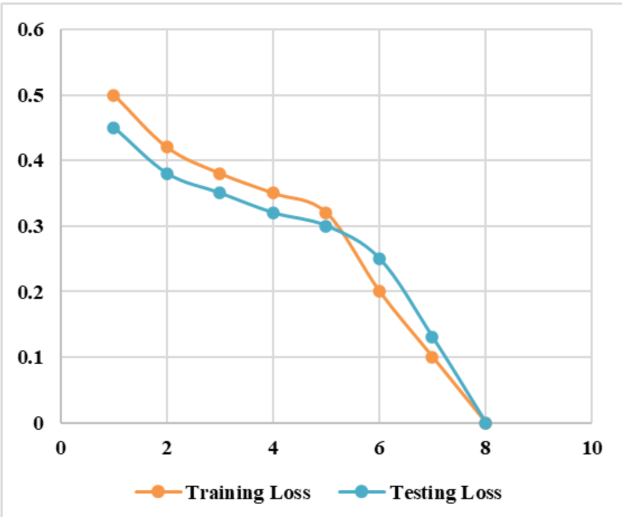

Fig. 6. Training loss and testing loss.

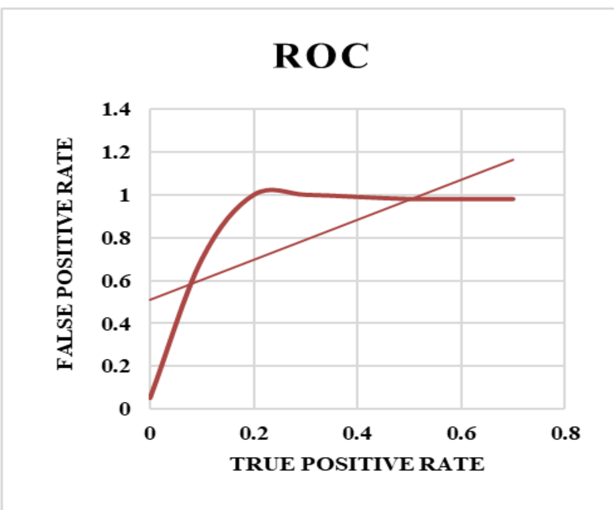

Fig. 7. ROC curve.

Table II compares the performance of various methods in high-band image classification, focusing on accuracy, precision, recall, and F1 score metrics. The Convolutional Neural Network (CNN) achieves 85.77% accuracy, with precision, recall, and F1 score values of 75.87%, 86.09%, and 82%, respectively. The Generative Adversarial Network (GAN) demonstrates superior performance, achieving 97.75% accuracy, with precision, recall, and F1 score values of 90%, 98%, and 95.98%, respectively. The Spectral-Spatial CNN method combines spectral and spatial information, resulting in 95% accuracy, with precision, recall, and F1 score values of 88.56%, 97%, and 95%, respectively.

TABLE II. PERFORMANCE PARAMETERS OF DIFFERENT CLASSIFICATION METHODS

Method	Accuracy	Precision	Recall	F1score
CNN	85.77	75.87	86.09	82
GAN	97.75	90	98	95.98
Spectral-Spatial CNN	95	88.56	97	95
Proposed Spectral-Spatial CNN WITH ADAM	97.8	97	96.8	98

The proposed Spectral-Spatial CNN with ADAM optimization achieves the highest performance, with an accuracy of 97.8%, precision of 97%, recall of 96.8%, and an outstanding F1 score of 98%. These results underscore the efficacy of the proposed approach in high-band image classification, suggesting its suitability for practical applications as shown in Fig. 8.

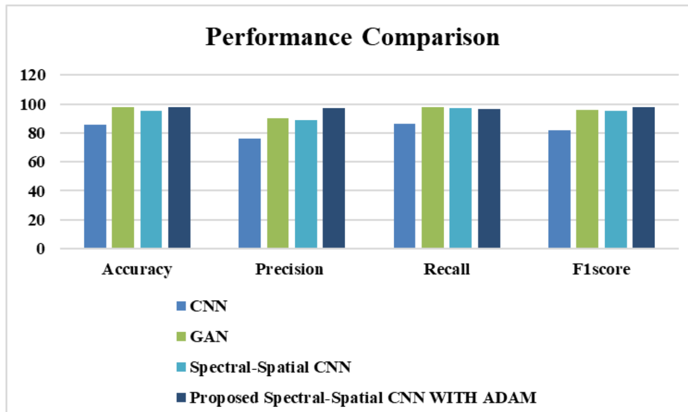

Fig. 8. The performance comparison of different classification methods.

The study explores the effectiveness of a proposed method for high-band image analysis in remote sensing applications. The method significantly outperforms existing methods, achieving state-of-the-art results in accuracy and performance. The approach utilizes CNNs for feature extraction, capturing hierarchical features and improving classification accuracy. The introduction of a spectral-spatial transformer module enhances classification performance by considering both spectral and spatial features. This module enhances robustness and adaptability in processing complex and varied image data sets. Additionally, adaptive motion algorithms optimize the training process of the deep learning network, achieving faster convergence and improved generalization performance. This strategy is highly effective in scenarios where training data may be diverse or noisy. The study contributes to the advancement of remote sensing image analysis by providing a comprehensive framework that integrates state-of-the-art techniques in deep learning and optimization. Future research may explore additional refinements and extensions to the proposed method, such as integrating multi-modal data sources or incorporating additional optimization techniques. This will further enhance the capabilities of image analysis systems for remote sensing and related domains, enabling more accurate and efficient analysis of Earth observation data [15].

VI. Conclusion and Future work

The method for high-band image analysis and classification exhibits promising efficiency and accuracy, leveraging a combination of spectral-spatial transformer, CNN-based feature extraction, and adaptive momentum optimization. By integrating deep learning techniques and optimization algorithms, this framework effectively captures both spatial and spectral data, leading to improved classification accuracy. The flexible momentum optimization approach further enhances the training process by dynamically modifying the momentum parameter, resulting in faster convergence and better generalization. These findings underscore the efficacy of the proposed technique across various high-band image analysis applications. Moving forward, future research directions could focus on exploring attention mechanisms to enhance feature extraction and classification precision, investigating new optimization methods tailored for high-band picture analysis, extending the framework to accommodate multi-modal or multi-temporal datasets, and conducting comprehensive tests on larger and more diverse datasets to confirm scalability and resilience. Ultimately, these efforts aim to enhance the methodology's effectiveness and adaptability for high-band picture assessment and classification in real-world scenarios.

References

[1] S. Mei, Y. Geng, J. Hou, and Q. Du, “Learning hyperspectral images from RGB images via a coarse-to-fine CNN,” Sci. China Inf. Sci., vol. 65, pp. 1–14, 2022.

[2] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, “Hyperspectral images super-resolution via learning high-order coupled tensor ring representation,” IEEE Trans. Neural Netw. Learn. Syst., vol. 31, no. 11, pp. 4747–4760, 2020.

[3] C.-H. Lin, S.-H. Huang, T.-H. Lin, and P. C. Wu, “Metasurface-empowered snapshot hyperspectral imaging with convex/deep (CODE) small-data learning theory,” Nat. Commun., vol. 14, no. 1, p. 6979, 2023.

[4] H. Fu et al., “A novel band selection and spatial noise reduction method for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–13, 2022.

[5] D. AL-Alimi, M. A. Al-qaness, Z. Cai, A. Dahou, Y. Shao, and S. Issaka, “Meta-learner hybrid models to classify hyperspectral images,” Remote Sens., vol. 14, no. 4, p. 1038, 2022.

[6] B. Xu, X. Li, W. Hou, Y. Wang, and Y. Wei, “A similarity-based ranking method for hyperspectral band selection,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 11, pp. 9585–9599, 2021.

[7] C.-I. Chang, Y.-M. Kuo, S. Chen, C.-C. Liang, K. Y. Ma, and P. F. Hu, “Self-mutual information-based band selection for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 7, pp. 5979–5997, 2020.

[8] A. Hennessy, K. Clarke, and M. Lewis, “Hyperspectral classification of plants: A review of waveband selection generalisability,” Remote Sens., vol. 12, no. 1, p. 113, 2020.

[9] L. Agilandeeswari, M. Prabukumar, V. Radhesyam, K. L. B. Phaneendra, and A. Farhan, “Crop classification for agricultural applications in hyperspectral remote sensing images,” Appl. Sci., vol. 12, no. 3, p. 1670, 2022.

[10] L. Mou, S. Saha, Y. Hua, F. Bovolo, L. Bruzzone, and X. X. Zhu, “Deep reinforcement learning for band selection in hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–14, 2021.

[11] B. Barman and S. Patra, “Variable precision rough set based unsupervised band selection technique for hyperspectral image classification,” Knowl.-Based Syst., vol. 193, p. 105414, 2020.

[12] R. N. Patro, S. Subudhi, P. K. Biswal, and F. Dell’acqua, “A review of unsupervised band selection techniques: Land cover classification for hyperspectral earth observation data,” IEEE Geosci. Remote Sens. Mag., vol. 9, no. 3, pp. 72–111, 2021.

[13] M. Ramamurthy, Y. H. Robinson, S. Vimal, and A. Suresh, “Auto encoder based dimensionality reduction and classification using convolutional neural networks for hyperspectral images,” Microprocess. Microsyst., vol. 79, p. 103280, 2020.

[14] F. He, F. Nie, R. Wang, W. Jia, F. Zhang, and X. Li, “Semisupervised band selection with graph optimization for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 12, pp. 10298–10311, 2020.

[15] S. Bera and V. K. Shrivastava, “Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification,” Int. J. Remote Sens., vol. 41, no. 7, pp. 2664–2683, 2020.

[16] D. Hong et al., “SpectralFormer: Rethinking hyperspectral image classification with transformers,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–15, 2021.

[17] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, “Graph convolutional networks for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 7, pp. 5966–5978, 2020.

[18] S. K. Roy, S. Manna, T. Song, and L. Bruzzone, “Attention-based adaptive spectral–spatial kernel ResNet for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 9, pp. 7831–7843, 2020.

[19] H. Fu, G. Sun, J. Ren, A. Zhang, and X. Jia, “Fusion of PCA and segmented-PCA domain multiscale 2-D-SSA for effective spectral-spatial feature extraction and data classification in hyperspectral imagery,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–14, 2020.

[20] M. P. Uddin, M. A. Mamun, and M. A. Hossain, “PCA-based feature reduction for hyperspectral remote sensing image classification,” IETE Tech. Rev., vol. 38, no. 4, pp. 377–396, 2021.

[21] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera, “Big data preprocessing: methods and prospects,” Big Data Anal., vol. 1, no. 1, pp. 1–22, 2016.

[22] C. Yu, R. Han, M. Song, C. Liu, and C.-I. Chang, “Feedback attention-based dense CNN for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–16, 2021.

[23] X. He, Y. Chen, and Z. Lin, “Spatial-spectral transformer for hyperspectral image classification,” Remote Sens., vol. 13, no. 3, p. 498, 2021.