International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I1P121

Volume 72 Issue 1, 209-218, January 2024
© 2024 Seventh Sense Research Group®

Original Article

Performance Evaluation of Hyperspectral Image Classification Methods: A Comparative Study

Tilottama Goswami[1], Yaksha Kasturi[2], Kommareddy Anvitha[3], Kovvur Ram Mohan Rao[4]

[1,2,3,4]*Department of Information Technology, Vasavi College of Engineering, Telangana, India.*

[1]*Corresponding Author : tgoswami@staff.vce.ac.in*

Received: 02 September 2023 Revised: 11 December 2023 Accepted: 23 December 2023 Published: 07 January 2024

Abstract - *Hyperspectral Images (HSIs) offer an extensive wealth of spectral-spatial information through their numerous contiguous narrow bands. However, selecting relevant spectral-spatial kernel features creates a challenge as it involves dealing with noise and band correlation. The classification of hyperspectral images plays a crucial role in remote sensing, and various methods have been proposed to tackle this challenge. This paper presents a compilation and comparative study of the recent state-of-the-art deep learning architectures for the HSI classification tasks: Attention-Based Adaptive Spectral-Spatial Kernel ResNet (A2S2K-ResNet), Residual Network (ResNet), Contextual CNN, Deep Pyramidal Residual Networks (DPyResNet), and SpectralSpatialRN (SSRN). These methods are evaluated on four datasets: Indian Pines, Salinas, Botswana, and Kennedy Space Center, which are commonly used for land cover classification in hyperspectral imaging. The performance evaluation of the classification methods is based on overall accuracy and computational time efficiency. The A2S2K-ResNet architecture demonstrates superior classification capabilities compared to the others followed by Contextual Net.*

Keywords - *Deep Learning, Hyperspectral Image (HSI) Classification, Performance Analysis, Residual Network (ResNet), Spectral-Spatial Information.*

1. Introduction

Remote sensing relies heavily on Hyperspectral Image (HSI) classification, an important task that requires the identification of diverse materials and land covers from high-dimensional hyperspectral data. Hyperspectral images are inherently complex due to the representation of each pixel as a spectrum of reflectance values across multiple wavelengths, resulting in a significant abundance of spectral bands. The challenging nature of hyperspectral image classification stems from the spectral variability observed among different classes, compounded by the presence of noise in the data.

Deep learning-based techniques have significantly improved hyperspectral image classification in recent times. The noteworthy aspect of these approaches is their ability to extract meaningful high-level features from raw hyperspectral data automatically.

This capability has a direct positive impact on enhancing the accuracy of classification outcomes. Various state-of-the-art deep learning frameworks have been implemented to classify hyperspectral images, including A2S2K-ResNet [1], ResNet [2], Contextual CNN [3], DPyResNet [4], SSRN [5] and HybridSN [6].These models have demonstrated exceptional performance and have yielded promising results in HSI classification.

He et al. [2] present a framework to simplify the training of networks with a large number of layers. ResNet is an acclaimed deep residual network renowned for its successful application in diverse computer vision tasks.

Lee et al. [3] proposed a convolutional neural network called Contextual CNN, a deep learning model that leverages the contextual information of neighboring pixels to improve classification accuracy.

Paoletti et al. [4] introduced an architecture exclusively for the HSI data called DPyResNet, which is a customized adaptation of ResNet that integrates spatial and spectral information through the utilization of 2D-3D convolutional layers. This modification enables DPyResNet to achieve improved performance for HSI classification.

Zhong et al. [5] designed SSRN, which utilizes a stacked sparse autoencoder to extract informative spectral-spatial features. This methodology significantly enhances the accuracy of classification tasks. Recently, Roy et al. [1] presented a new architecture, A2S2K-ResNet, that combines the advantages of both spatial and spectral processing. The model utilizes a hybrid 2D-3D approach, combining a 2D CNN for spatial feature extraction and a 3D CNN for spectral feature extraction.

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

This architecture has shown promising results in various HSI classification tasks, including crop classification, mineral detection, and urban land use mapping. Roy et al. proposed HybridSN [6], a hybrid architecture that combines 3D and 2D CNNs in a hierarchical manner, facilitating the incorporation of spatial and spectral information.

Since the implementation of this model is not accessible to the public, it has been excluded from the comparative study despite being a recent approach. Lv and Wang's comprehensive review [7] provides a broad comparison of various techniques in hyperspectral image classification. However, it fails to conduct an in-depth comparative analysis that specifically focuses on the latest deep learning models.

This gap underscores the need to evaluate advanced deep learning architectures thoroughly. To address this issue, this paper presents a thorough comparative study of state-of-the-art models by assessing their performance across multiple hyperspectral datasets. The objective is not only to fill this crucial gap but also to expand the knowledge base within the HSI classification field with insightful findings useful for future research and practical applications. The results will complement Lv and Wang's work while offering valuable insights into HSI classification methods based on modern deep-learning approaches.The structure of the paper is laid out in the following manner: Section 2 discusses the challenges faced in HSI images for land cover classification. Section 3 elaborates on the important concepts and components of deep learning models that are useful for learning and enabling the automatic extraction of discriminative features to classify. Section 4 provides a concise overview of the state of art methods specifically developed for land cover classification. Section 5 provides insights into how A2S2K-Resnet achieves better results and suggests potential applications. Section 6 reports the comparative study of five methods on four benchmark datasets related to HSI classification. The paper concludes by outlining prospective areas for future research in Section 7.

2. Problem Definition

The high dimensionality and complexity of hyperspectral data pose significant challenges in land cover classification using hyperspectral imagery. Hyperspectral images are composed of numerous spectral bands that gather data on the reflectance characteristics across varying wavelengths.

This results in a high-dimensional feature space, which can make classification challenging due to the curse of dimensionality. Moreover, hyperspectral data can exhibit complex spectral variability and spatial variability due to factors such as atmospheric interference, sensor noise, and variability in illumination and viewing geometry. Another challenge is the presence of spectral variability within land cover types, which can lead to misclassification.

For example, vegetation can exhibit different spectral signatures depending on factors such as species, age, and health status. Similarly, soil can vary in its reflectance properties depending on factors such as moisture content and mineral composition. Therefore, effective classification algorithms need to be able to handle spectral variability within land cover types while still maintaining discrimination between different land cover classes.Furthermore, the choice of classification algorithm and its parameterization can greatly affect the accuracy of classification results. Some algorithms may perform better on certain types of land cover or under certain conditions, while others may be more robust to noise and variability. Therefore, it is important to carefully evaluate the performance of different classification algorithms and their parameter settings to select the best approach for a given application. Hyperspectral imaging technology captures information about objects and materials at a very fine spectral resolution, enabling the detection of subtle differences in the reflectance properties of different land cover types. The utilization of machine learning algorithms on hyperspectral data enables automated classification of diverse land cover types, encompassing vegetation, water, soil, and urban areas. This automated classification process serves as a valuable tool for numerous applications, including environmental monitoring, land management, and urban planning. The goal is to achieve high accuracy in land cover classification, which requires the development of robust layers to learn hierarchical representations of data.

3. Related Work

Deep learning learns hierarchical representations of data using artificial neural networks and has gained significant attention and achieved remarkable success in the HSI classification domain. These methods harness the power of neural networks to learn discriminative features automatically from the vast amount of spectral information available in HSIs. Deep learning models for HSI classification typically consist of several key components like convolutional layers, pooling layers, residual connections, normalization layers, attention mechanisms, etc. The methods under consideration for performance evaluation leverage the following components and concepts to learn complex patterns from the HSI image datasets.

3.1. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) [8] are applied to HSI classification. These networks automatically learn spatial and spectral features from input hyperspectral images. The CNN architecture consists of various components such as convolutional, dropout layer, batch normalization, pooling and fully connected layers. Convolutional layers utilize filters to perform localized feature extraction across the input image, enabling the detection of intricate patterns and spatial relationships.

Pooling layers are used for downsampling, reducing the spatial dimensionality, and extracting essential features. Fully connected layers connect these extracted features to the final classification layer.

3.2. Residual Connections

Residual connections [2], also known as skip connections, are an integral part of deep learning models, particularly in architectures like Residual Network (ResNet). Residual connections are designed to tackle the degradation issue that arises with the addition of more layers to a neural network. They are achieved by introducing shortcut connections that directly transmit the input from one layer to a deeper layer. In certain works, like [12, 13], a small number of intermediary layers were directly linked to auxiliary classifiers to mitigate issues related to vanishing or exploding gradients. Within [12], an “inception” layer was introduced, comprising a shortcut branch along with several deeper branches.

These connections facilitate the network’s ability to learn residual mappings, capturing the disparity between the desired output and the input. By doing so, they facilitate the training of deeper networks by allowing the gradient to flow more directly, mitigating the challenge of vanishing gradients. Residual connections have demonstrated the ability to enhance the optimization process, improve network convergence, and enable the construction of deep neural networks with numerous layers while preserving or even enhancing their performance.

3.3. Attention Mechanisms

Attention mechanisms [1], [9] have become a powerful tool in deep learning models. Attention mechanisms enhance the capability of models to emphasize relevant components of the input data while disregarding irrelevant or noisy information. This selective focus enables the model to allocate its attention resources more effectively, improving its ability to process and extract meaningful features. This is achieved by assigning weights to different components of the input, depending on their significance.

These weights are acquired during training and can dynamically change for each input. By prioritizing informative features while disregarding less relevant ones, attention mechanisms enhance the model’s capacity to capture significant patterns and achieve accurate predictions.

This selective attention enhances the model’s interpretability and robustness, enabling it to handle complex and high-dimensional hyperspectral data effectively. Attention mechanisms have been successfully employed in various deep learning architectures, contributing to significant improvements in hyperspectral image classification performance.

3.4. Pyramidal Structures

Pyramidal structures [4] in deep learning refer to architectures that involve hierarchical or multi-scale representations of data. These structures are particularly useful in HSI classification, where capturing both local and global context is essential. Pyramidal structures typically involve multiple levels or layers, with each level capturing information at a different scale or level of abstraction.

This can be achieved through various mechanisms such as pooling, convolutional operations, or feature aggregation. By incorporating pyramidal structures into the network architecture, models can effectively capture fine-grained details at lower levels while simultaneously learning high-level semantic representations at higher levels. This hierarchical approach enables the network to capture both local and global context, improving its ability to discriminate between different land cover classes in hyperspectral images.

3.5. Adaptive Pooling

In the realm of HSI classification, adaptive pooling [10] serves as a prevalent technique within deep learning models. Its purpose is to effectively manage input data that exhibits diverse spatial dimensions, ensuring optimal processing and analysis. Adaptive pooling, unlike traditional pooling layers, dynamically adjusts pooling window sizes based on input data.

This enables flexible feature extraction and captures spatial information regardless of image size. By dynamically adjusting the pooling windows, it enables the model to maintain spatial information at different scales, ensuring that relevant features are preserved while reducing the dimensionality of the input data. Adaptive pooling can be implemented using various methods, such as average pooling or max pooling. This technique plays a crucial role in capturing meaningful spatial features from hyperspectral images and contributes to improved classification accuracy.

4. Deep Learning Models/Architectures

This section presents a compilation of recent state-of-the-art deep learning architectures designed specifically for HSI land cover classification.

4.1. Attention-Based Adaptive Spectral–Spatial Kernel ResNet (A2S2K-ResNet)

Hyperspectral images offer a wealth of spectral-spatial information, presenting hundreds of consecutive narrow bands. However, the challenge lies in selecting informative spectral-spatial kernel features amid noise and band correlations. Convolutional neural networks (CNNs) with fixed-size receptive fields (RFs) [17] are used to overcome this. Roy et al. [1] proposed A2S2K-ResNet to overcome the above limitations. The architecture is organized as follows:

4.1.1. Attention-Based Adaptive Spectral–Spatial Kernel Module

This module enables the network's neurons to effectively learn spectral-spatial features dynamically. This adjustment allows for the incorporation of convolutional kernels from various RF sizes, thereby capturing multi-scale information. Operations within the Module:

- *Kernel Split*: In this operation, the input feature maps are split into multiple branches, each of which focuses on a different RF size. Essentially, this involves applying convolutional kernels of varying sizes to the input data. These branches process the input data with different receptive fields, capturing diverse spatial patterns.
- *Fusion*: The kernel branches, each having processed the input with a specific RF size, are then fused together. This fusion mechanism combines multi-scale information captured by different RF sizes.
- *Selection*: After the fusion of multi-scale information, the module performs an automatic selection process. This selection involves identifying discriminative spectral-spatial kernel feature maps that contribute significantly to the classification task. The selection captures the most relevant information with adaptive RF sizes. Essentially, this step focuses on retaining the most informative channels for further processing.

4.1.2. Efficient Feature Recalibration Module

Capturing long-range nonlinear cross-channel dependencies in feature maps is crucial for improving classification performance. An Efficient Feature Recalibration (EFR) is introduced in [17]. It is used to recalibrate features across channels, allowing the network to adaptively weigh the importance of different features during the classification process.

4.1.3. Modified Spectral–Spatial Residual Network

The core building blocks of ResNet referred to as ResBlocks, are stacked together depth-wise to facilitate bidirectional information flow within the network. Following the sequence of ResBlocks, a Global Average Pooling (GAP) layer is introduced. Finally, the feature maps are fed into a fully connected layer and further by a softmax activation function. This step assigns a probability distribution across different classes to each pixel in the hyperspectral image. The pixel is assigned the class with the highest probability.

4.2. Residual Network (ResNet)

ResNet [2] tackles the issue of training deep neural networks by incorporating residual connections. In traditional deep networks, each layer learns to approximate the underlying mapping of the input to the output. However, as the network becomes deeper, it becomes increasingly difficult for the network to learn this mapping accurately. This issue is referred to as the degradation problem. ResNet tackles this challenge by including residual connections, which allow the network to learn residual mappings effectively.

Rather than directly learning the desired mapping, ResNet specifically aims to capture the discrepancy between the input and output, known as the residual. By adding this residual to the original input, the network can effectively adjust the output to achieve accurate results. ResNet variants [18] - [23] are good to deal with the vanishing gradient problem.

The ResNet framework is composed of multiple residual blocks, each containing a series of convolutional layers, which are succeeded by batch normalization and then activated by ReLU functions. The skip connection allows the residual to bypass the convolutional layers, enabling the network to learn residual mappings efficiently. He et al. [2] experimented with different depths of ResNet, ranging from a few layers to over a hundred layers, and demonstrated that deeper networks achieve better performance. The models underwent training on the ImageNet dataset, surpassing previous approaches. It effectively tackles the degradation problem in deep networks by introducing residual connections, leading to easier training of very deep networks without accuracy degradation.

ResNet architecture allows for scalability, with deeper networks consistently showing improved performance. However, ResNet has some limitations, including increased computational complexity, sensitivity to hyperparameters, limited interpretability due to its deep architecture, and reliance on quality training data.

4.3. Contextual CNN

The Contextual CNN [3] architecture proposed by Lee et al. is a deep learning architecture designed specifically for processing hyperspectral images. The focus of this model is to incorporate contextual information from neighboring pixels to increase the accuracy of categorization results. By collectively combining local spatial and spectral information, the model captures intricate contextual relationships within neighboring pixel vectors. This enables the network to build a joint spatio-spectral feature map that amalgamates spectral and spatial attributes. The subsequent integration into a fully convolutional network allows for accurate pixel-level predictions, effectively addressing tasks like pixel-wise image labeling.

This network distinguishes itself by its increased depth and breadth compared to existing models. Contextual CNN excels in discovering local contextual relationships by effectively harnessing the intricate connections between neighboring individual pixel vectors. Lee et al. [3] introduce an original Fully Convolutional Network (FCN) [15] for hyperspectral image (HSI) classification. The network comprises multiple convolutional layers and is structured as follows: multi-scale filter bank, succeeded by two blocks of convolutional layers and three convolution layers. The first component is a multi-scale filter bank. The model exploits local interactions between spatial and spectral aspects by introducing them at the initial network stage.

The filter block is followed by convolutional layers that involve residual learning. AlexNet[16] comprises five convolutional layers with three fully connected layers. The last three layers are adept at recognizing local features, like AlexNet.

During training, specific convolutional layers integrate dropout techniques and ReLU activation is applied to convolutional layers, as well as residual learning modules. Local Response Normalization is applied to normalize the outputs from the initial two convolutional layers. The dimensions of all data units within the architecture remain consistent in height and width, with changes only in depth.

4.4. Deep Pyramidal Residual Networks (DPyResNet)

The Deep Pyramidal Residual Networks (DPyResNet) [4] model addresses the challenges of effectively managing and classifying extensive hyperspectral data cubes. This novel architecture progressively refines feature extraction through a multi-layered framework. Unlike conventional CNNs, where feature maps maintain a consistent size across layers, DPyResNet progressively widens the feature map dimensions as the network depth increases. The model employs pyramidal bottleneck residual blocks, each consisting of three convolutional layers that facilitate the gradual inclusion of a diverse array of feature map locations.

This structure helps gain an understanding of spectral-spatial patterns across multiple scales. The pyramidal bottleneck residual blocks act as knowledge amplifiers, progressively integrating spatial and spectral information while ensuring a manageable computational complexity. The resulting feature maps contain a complex blend of multi-scale patterns, thus enabling the network to discriminate between diverse classes within hyperspectral data.

As hyperspectral data comprises multiple bands of spectral information and their corresponding spatial arrangement, the model exploits this synergy by processing data through successive layers of convolutional operations. These operations extract increasingly abstract features, Capturing details at different levels of detail. By maintaining this hierarchical perspective, the model develops an acute understanding of the specific spectral characteristics and their spatial context. The model’s ability also extends to its adaptability to different spatial sizes.

The pyramid-like architecture accommodates varying scales, making it suitable for real-world applications where input sizes can differ significantly. Moreover, the model’s resilience to different percentages of training data is noteworthy. Deep learning techniques often necessitate substantial labelled samples for training; however, DPyResNet consistently demonstrates improved performance regardless of the amount of training data available, indicating ts potential for effective deployment in data-limited scenarios.

4.5. Spectral–Spatial Residual Network (SSRN)

The Spectral-Spatial Residual Network (SSRN) [5] addresses the challenge of declining accuracy while demonstrating adaptability to diverse datasets. Through an innovative 3-D deep learning framework, SSRN facilitates the automatic extraction of spectral-spatial representations necessary for accurate image classification.

The design of SSRN involves consecutive spectral and spatial residual blocks, which serve as their fundamental building blocks. This sequential arrangement allows SSRN to effectively capture intricate spectral signatures and spatial contexts in hyperspectral images.

SSRN makes use of residual learning, which mitigates the accuracy drop phenomenon. By incorporating residual connections, SSRN ensures that each layer learns and contributes incremental information, thus promoting deeper, more meaningful representations. This mechanism proves especially effective in hyperspectral image analysis, where subtle spectral variations and spatial patterns are of great importance.

Further contributing to SSRN’s capability is the strategic application of batch normalization (BN). The fundamental idea behind BN is to normalize the activations of each layer within the network based on a batch of input data.

This normalization process serves to stabilize and standardize the distribution of the data flowing through the network during training. By ensuring that the inputs to each layer have similar statistical properties, BN addresses the common issue of internal covariate shift.

This phenomenon occurs as the distributions of intermediate activations change during training, often leading to slower convergence and requiring careful tuning of learning rates. BN reduces this challenge by normalizing the inputs, effectively moderating the adverse impact of covariate shift, and enabling more stable and efficient training.

The BN operation at each convolution layer also reduces the number of iterations for training from hundreds of thousands in [14] to only a few hundred.

SSRN performs consistently well across different datasets, making it an appealing choice for a wide range of remote applications. It offers reliable classification sensing performance even when limited labelled data is available.

5. Comparative Analysis

This section delves into the reasons behind the exceptional performance of A2S2K-ResNet and explores the potential applications for each of the mentioned models from the previous section.

5.1. Integration of Attention Mechanisms

Attention mechanisms enable the network to enhance feature representation adaptively. This results in more accurate classification, particularly in complex scenes which have a mixture of various land cover types. The attention mechanisms enable the network's learning to focus on the most relevant spectral and spatial features of the hyperspectral images as it dynamically allocates more computational resources to important features while ignoring less relevant features.

5.2. Adaptive Spectral-Spatial Kernels

A2S2K-ResNet uses adaptive kernels that can adjust their size and shape according to the spectral and spatial characteristics of the input data. These adaptive kernels allow the network to capture both coarse and fine features effectively. This improves the classification accuracy of heterogeneous data across diverse regions.

5.3. Residual Learning Framework

Building on the standard ResNet framework, A2S2K-ResNet incorporates residual learning, which can train deeper networks efficiently without the problem of vanishing gradients. These residual connections support the development of deeper, more complex networks without the risk of overfitting. With the inclusion of shortcut connections, learning residual functions, and better feature representation, this framework enhances the model's performance and generalization ability.

5.4. Comprehensive Feature Learning

By combining spectral and spatial information along with attention and adaptive kernel mechanisms, the model learns a more extensive set of features compared to traditional methods. This approach ensures that the model does not completely depend on only spectral information, but it also utilizes spatial context effectively, leading to more accurate classifications.

Complex models such as A2S2K-ResNet and DPyResNet are capable of extracting detailed features. However, they come with the risk of increased resource requirements and training difficulties. On the other hand, simpler models like ResNet provide a balance between depth and efficiency, which makes them suitable for diverse applications. A2S2K-ResNet has high complexity due to the integration of attention mechanisms and spectral-spatial kernels in the architecture, which adds more parameters and computational depth, enhancing the model's ability to focus on relevant features but also adding to the training complexity.

Due to these features, A2S2K-ResNet is suited for applications where detailed feature extraction is crucial, like in precision agriculture or mineralogy. The complexity of ResNet varies depending on the number of layers used in the architecture, but it is simpler than most other models.

As the residual connections are fairly straightforward, it is easier to interpret. ResNet is broadly applicable across various tasks due to its balance of depth and efficiency, particularly for general image classification tasks like facial recognition, object detection in autonomous vehicles, and medical imaging. DPyResNet is more complex than standard ResNet because of the pyramidal structure that increases the depth and number of parameters. The depth and scale variation of the network makes it difficult to interpret in certain situations. This model is useful for applications that require high levels of feature accuracy, like military surveillance or geological exploration. Contextual CNN is a moderately complex model that extracts contextual information by making use of more layers but not as many parameters as attention mechanisms.

This model is good for applications where context plays an important role in image classification, such as disaster assessment or urban planning. SSRN incorporates both spectral and spatial features, increasing the number of layers and parameters, but it uses residual connections to train the deep network efficiently. It is ideal for applications that require analysis of both spectral and spatial data, like in vegetation mapping or geological studies.

Each of these models possesses unique strengths that make them suitable for a variety of applications. A suitable model can be chosen based on the requirements of the application, computational resources available and the ideal balance between the interpretability and complexity of the model.

6. Results and Discussion

A comparative evaluation of five state-of-the-art methods with all four benchmark datasets is provided in this section. The related works used for comparison do not report the model performance for all four datasets. The experiments reveal which deep learning architecture, in general, is best suited for HSI classification.

6.1. Datasets

The experiments have been performed on four benchmark datasets for land cover classification, namely Salinas, Botswana, Indian Pines, and Kennedy Space Center. All four datasets are summarized in Table 1.

Table 1. Overview of the dataset's characteristics

Description	SA	BW	IP	KSC
Sensor	AVIRIS	Hyperion	AVIRIS	AVIRIS
Spatial Dimension	512x217	1476x256	145x145	512x614
Spectral Bands	224	145	224	176
Landcover	16	14	16	13
Total size	54129	3248	10249	5202

The Salinas (SA) dataset is obtained over the Salinas Valley in California using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [11] sensor with a fine spatial resolution of 3.7 meters and 224 spectral bands and 20 water absorption bands are discarded. It has a spatial dimension of 512x217 pixels and labels for 16 classes representing vegetables, bare soils, and vineyard fields.

The Botswana (BW) dataset is acquired by NASA's EO-1 satellite of the Okavango Delta in Botswana, Africa. The Hyperion sensor on EO-1 captures data with a pixel resolution of 30 meters, operating in 242 bands with a spectral resolution of 10 nm per window. The dataset has a spatial dimension of 145x145 pixels and labels for 16 classes of land cover types.

The AVIRIS collects the Indian Pines (IP) dataset [11] sensor over the Indian Pines test site in North-western Indiana, USA, consisting of 224 spectral bands having a spatial dimension of 145x145 pixels and has labels for 16 vegetation classes.

The Kennedy Space Center (KSC) dataset is captured by the AVIRIS [11] sensor with a spatial resolution of 18m over the Kennedy Space Center in Florida. It includes 224 spectral bands, which are narrowed down to 176 after removal of low signal-to-noise (SNR) ratio bands. The dataset consists of 5202 samples in total, covering the wavelength range from 400nm to 2500nm and labels for 13 wetland classes.

6.2. Performance Metrics

Overall accuracy, kappa coefficient, and time efficiency were used as metrics for the comparative study.

6.2.1. Overall Accuracy

The performance of the classification task is quantitatively measured using overall accuracy. It is the ratio of correctly classified samples to the total number of input samples. It is calculated using the following formula:

$$P_{\mathrm{OA}} = \frac{\sum_{i=1}^{c} x_{ii}}{\sum_{i=1}^{c} \sum_{j=1}^{c} x_{ij}}$$

6.2.2. Kappa Coefficient

The Kappa coefficient is a statistical metric used to measure the accuracy of a classification model by considering the possibility of random chance agreement. It ranges from -1 to 1, where 1 implies perfect agreement, 0 implies agreement equivalent to chance, and negative values suggest less-than-chance agreement. This metric is particularly valuable in scenarios with imbalanced class distributions, providing a more balanced assessment of classifier performance than mere accuracy.

It has been calculated using a confusion matrix, which outlines the actual versus predicted classifications and is essential for comparing classifiers and assessing rater reliability in various fields. It is calculated using the following formula:

$$P_{\mathrm{KC}} = \frac{n \sum_{i=1}^{c} x_{ii} - n \sum_{i=1}^{c} (x_{i+} * x_{+i})}{n^2 - \sum_{i=1}^{c} (x_{i+} * x_{+i})}$$

Where n is the total number of observations, x_{ii} is diagonal, x_{i+} is the sum of elements of each row, and x_{+i} is the sum of elements of each column.

6.2.3. Time Efficiency

Time efficiency refers to the speed and computational efficiency of deep learning architectures when applied to hyperspectral image (HSI) classification tasks. It measures how quickly the models can process and analyze HSI data, including both training and testing phases.

6.3. Comparative Results

The analysis parameters are already described in section 5.2. The implementation code of all five methods is publicly available [1]. All experiments are performed on Google Colab with the NVIDIA Tesla T4 GPU, which has 2,560 CUDA Cores, 320 Tensor Cores, 16 GB of GDDR6 GPU memory, and a memory bandwidth of 300 GB/s. Python version 3.10.12 has been used. The maximum number of epochs specified for training is 200. Some models converge and achieve satisfactory performance well before the maximum number of epochs.

Table 2. Comparative analysis of methods based on overall accuracy (in %)

Method	SA	BW	IP	KSC
A2S2K-ResNet	99.509 ± 0.003	99.439 ± 0.003	98.599 ± 0.003	99.617 ± 0.002
ResNet	98.611 ± 0.002	96.213 ± 0.009	92.707 ± 0.005	86.909 ± 0.043
Contextual CNN	96.071 ± 0.023	97.348 ± 0.013	91.436 ± 0.028	96.492 ± 0.004
DPyResNet	98.529 ± 0.003	95.052 ± 0.010	94.979 ± 0.016	93.917 ± 0.023
SSRN	97.136 ± 0.005	99.107 ± 0.003	98.251 ± 0.002	99.426 ± 0.002

Table 3. Analysis of methods in terms of Kappa Coefficient

Method	SA	BW	IP	KSC
A2S2K-ResNet	0.9945 ± 0.0036	0.9939 ± 0.0029	0.9840 ± 0.0038	0.9957 ± 0.0026
ResNet	0.9845 ± 0.0026	0.9589 ± 0.0102	0.9166 ± 0.0061	0.8536 ± 0.0483
Contextual CNN	0.9549 ± 0.0198	0.9712 ± 0.0139	0.9428 ± 0.0179	0.9609 ± 0.0041
DPyResNet	0.9836 ± 0.0034	0.9464 ± 0.0111	0.9021 ± 0.0318	0.9322 ± 0.0253
SSRN	0.9878 ± 0.0025	0.9903 ± 0.0028	0.9801 ± 0.0023	0.9936 ± 0.0017

Table 4. Analysis of methods based on time efficiency in terms of average training time (in sec)

Method	SA	BW	IP	KSC
A2S2K-ResNet	1863.801	207.322	619.1	510.882
ResNet	3313.729	153.097	744.755	250.374
Contextual CNN	3041.426	148.521	449.24	283.771
DPyResNet	2622.924	134.112	398.038	304.703
SSRN	3192.853	58.043	192.421	394.093

Table 5. Analysis of methods based on time efficiency in terms of average testing time (in sec)

Method	SA	BW	IP	KSC
A2S2K-ResNet	60.843	5.177	22.119	10.412
ResNet	147.302	9.605	28.675	11.546
Contextual CNN	41.729	3.947	16.705	8.796
DPyResNet	153.921	9.945	29.999	12.849
SSRN	85.077	5.445	17.457	8.002

Table 2 shows the overall accuracy-based analysis results. For the SA dataset, A2S2K-ResNet outperformed the other models with a score of 99.50%. ResNet and DPyResNet demonstrate competitive performance, with overall accuracies of 98.61% and 98.53%, respectively.

A2S2K-ResNet again achieved the highest overall accuracy of 99.439% for the BW dataset. SSRN and Contextual CNN delivered comparable results of 99.107% and 97.348%, respectively. A2S2K-ResNet achieved the highest overall accuracy for the Indian Pines dataset with a score of 98.59%, closely followed by SSRN with 98.25%. A2S2K-ResNet and SSRN also showed competitive performance with an accuracy of 99.61% and 99.42%, respectively, for the KSC dataset.

Table 3 shows the analysis of the methods based on the kappa coefficient. In this assessment, A2S2K-ResNet emerged as the top performer across multiple datasets, showcasing robust performance.

Notably, for the Salinas dataset, A2S2K-ResNet attained a Kappa coefficient of 0.9945 with a minimal deviation of ±0.0036, outperforming other methods such as ResNet, DPyResNet, and SSRN. Similarly, in the Botswana dataset, A2S2K-ResNet maintained its superiority with a Kappa coefficient of 0.9939 ± 0.0029, closely followed by ResNet and DPyResNet.

While Contextual CNN exhibited commendable performance with a Kappa coefficient of 0.9712 ± 0.0139 for the Indian Pines dataset, A2S2K-ResNet and SSRN dominated the KSC dataset with coefficients of 0.9957 ± 0.0026 and 0.9936 ± 0.0017 respectively, demonstrating their robustness in classification accuracy when evaluated using Kappa analysis.

Table 4 shows the time efficiency-based analysis results in terms of average training time, and Table 5 shows the average testing time. It is observed that A2S2K-ResNet performs the best in the SA dataset, with an average training time of 1863.801 seconds and an average testing time of 60.843 seconds. A2S2K-ResNet also achieves the best performance in the BW dataset, with an average training time of 207.322 seconds and an average testing time of 5.177 seconds. However, for the Indian Pines dataset, SSRN achieves the best performance with an average training time of 192.421 seconds and an average testing time of 17.457 seconds. For the KSC dataset, Contextual CNN performs with an average training time of 250.374 seconds and an average testing time of 11.546 seconds.

7. Conclusion

In this study, a comprehensive performance analysis of several state-of-the-art HSI Classification methods: A2S2K-ResNet, ResNet, Contextual CNN, DPyResNet and SSRN on the four land cover benchmark datasets, namely Salinas, Botswana, Indian Pines, and Kennedy Space Center has been conducted. Our findings highlight the varying performance of these methods across different datasets. A2S2K-ResNet consistently performed well in terms of overall accuracy and time efficiency, showcasing its effectiveness in achieving high classification accuracy. Contextual CNN and SSRN also demonstrated competitive performance, but SSRN was comparatively slower. ResNet and DPyResNet offered unique advantages in terms of interpretability and contextual information utilization. These results provide valuable insights into the strengths and limitations of these methods and offer guidance for selecting the most suitable approach for diverse hyperspectral image classification tasks. In the future, the authors will explore ensemble learning strategies further to improve the classification performance of hyperspectral image analysis methods.

Author Contributions

TG & KR - conceptualized the problem; TG, YK, KA, KR - collaborated on methodology; YK and KA- conducted experiments, TG - validation, YK - original draft preparation, TG -review and editing, KA-visualization and result analysis; KR - contributed to tools. TG & KR-approved; Vasavi College of Engineering, Hyderabad - funding acquisition. All authors read and approved the manuscript.

Acknowledgments

The authors wish to extend their gratitude to Vasavi College of Engineering, Hyderabad, for their support in facilitating this research.

References

[1] Swalpa Kumar Roy et al., “Attention-Based Adaptive Spectral–Spatial Kernel ResNet for Hyperspectral Image Classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 9, pp. 7831-7843, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[2] Kaiming He et al., “Deep Residual Learning for Image Recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[3] Hyungtae Lee, and Heesung Kwon, “Going Deeper with Contextual CNN for Hyperspectral Image Classification,” *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4843-4855, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[4] Mercedes E. Paoletti et al., “Deep Pyramidal Residual Networks for Spectral-Spatial Hyperspectral Image Classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 740-754, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[5] Zilong Zhong et al., “Spectral–Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 847-858, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[6] Swalpa Kumar Roy et al., “HybridSN: Exploring 3-D-2-D CNN Feature Hierarchy for Hyperspectral Image Classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 2, pp. 277-281, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Wenjing Lv, and Xiaofei Wang, “Overview of Hyperspectral Image Classification,” *Journal of Sensors*, vol. 2020, pp. 1-13, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi, “Understanding of a Convolutional Neural Network,” *2017 International Conference on Engineering and Technology (ICET)*, Antalya, Turkey, pp. 1-6, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[9] Qilong Wang et al., “ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks,” *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, pp. 11531-11539, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Hsi-Chin Hsin, and Chien-Kun Su, “Adaptive Pooling for Convolutional Neural Networks with Arbitrary Input Sizes,” *IEEE 3rd Eurasia Conference on IOT, Communication and Engineering (ECICE)*, Yunlin, Taiwan, pp. 196-198, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Robert Green et al., “Imaging Spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS),” *Remote Sensing of Environment*, vol. 65, no. 3, pp. 227-248, 1998. [CrossRef] [Google Scholar] [Publisher Link]

[12] Christian Szegedy et al., “Going Deeper with Convolutions,” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 1-9, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[13] Chen-Yu Lee et al., “Deeply-Supervised Nets,” *Arxiv*, pp. 1-10, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[14] Ying Li, Haokui Zhang, and Qiang Shen, “Spectral-Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network,” *Remote Sensing*, vol. 9, no. 1, pp. 1-21, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[15] Jonathan Long, Evan Shelhamer, and Trevor Darrell, “Fully Convolutional Networks for Semantic Segmentation,” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 3431-3440, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[16] Alex Krizhevsky, Ilya Sutskever, and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems*, vol. 1, pp. 1097-1105, 2012. [Google Scholar] [Publisher Link]

[17] Xiang Li et al., “Selective Kernel Networks,” *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, pp. 510-519, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[18] Qilong Wang et al., “ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks,” *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, pp. 11531-11539, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] Kaiming He et al., “Identity Mappings in Deep Residual Networks,” *Proceeding European Conference on Computer Vision (ECCV)*, pp. 630-645, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[20] Sergey Zagoruyko, and Nikos Komodakis, “Wide Residual Networks,” *Arxiv*, pp. 1-15, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[21] Fei Wang et al., “Residual Attention Network for Image Classification,” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 6450-6458, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[22] Dongyoon Han, Jiwhan Kim, and Junmo Kim, “Deep Pyramidal Residual Networks,” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Honolulu, HI, USA, pp. 6307-6315, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[23] Ionut Cosmin Duta et al., “Improved Residual Networks for Image and Video Recognition,” *Arxiv*, pp. 1-22, 2020 [CrossRef] [Google Scholar] [Publisher Link]