International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 16, No.1 (Jul-24)

http://dx.doi.org/10.12785/ijcds/160133

Deep Learning Based Hyperspectral Image Classification: A Review For Future Enhancement

Anish Sarkar[1], Utpal Nandi*[1], Nayan Kumar Sarkar[2], Chiranjit Changdar[3] and Bachchu Paul[1]

[1]*Department of Computer Science, Vidyasagar University, Rangamati, Midnapore, West Bengal, INDIA*
[2]*Faculty of Engineering, Assam Down Town University, Guwahati, Assam, INDIA*
[3]*Department of Computer Science, Belda College, Belda, West Bengal, INDIA*

Received 16 Jan. 2024, Revised 2 Apr. 2024, Accepted 19 Apr. 2024, Published 1 Jul. 2024

Abstract: The use of Hyperspectral Image(HSI) has become prevalent in many sectors due to its ability to identify detailed spectral information (i.e., relationships between the collected spectral data and the object in the HSI data) that cannot be obtained through ordinary imaging. Traditional RGB image classification approaches are insufficient for hyperspectral image classification(HSIC) because they struggle to capture the subtle spectral information that exists within hyperspectral data. In the past few years, the Deep Learning(DL) based model has become a very powerful and efficient non-linear feature extractor for a wide range of computer vision tasks. Furthermore, DL-based models are exempt from manual feature extraction. The use of this stimulus prompted the researchers to use a DL-based model for the classification of Hyperspectral Images, which yielded impressive results. This motivation inspired the researchers to develop a DL-based model for the classification of hyperspectral images, which performed well. Deeper networks might encounter vanishing gradient problems, making optimization more difficult. To address this issue, regularisation and architectural improvements are being implemented. One of the key issues is that the DL-based HSIC model requires a large number of training samples for training, which is an important concern with hyperspectral data due to the scarcity of public HSI datasets. This article provides an overview of deep learning for hyperspectral image classification and assesses the most recent methods. Among all studied methods SpectralNET offers significantly better performance, due to the utilization of wavelet transformation.

Keywords: Hyperspectral image, hyperspectral image classification, deep learning based hyperspectral image classification, deep learning

1. INTRODUCTION

The sole purpose of Hyperspectral Image(HSI) processing is to acquire meaningful data from the spectral bands captured by the sensor at specific distances without having to come into contact with the object of interest [1]. HSI processing technology can capture data in hundreds of discrete, successive spectral bands across a wide range of the electromagnetic spectrum, including visible light (0.4-0.7 m) and short wave infrared (0.7-2.4 m). This gives a full study of the spectrum, allowing for the extraction of detailed spectral information. Moreover, HSI is the only method for evaluating the luminosity properties of objects in the mid to far infrared band [2]. Because HSI offers comprehensive data about each pixel, popular methods for multispectral and RGB images have encountered numerous challenges and cannot be used directly.

HSI has a multitude of useful applications, including municipal planning, natural resource prospecting, forestry management, and ecological sustainability [3], [4], [5], [6], [7], [8]. HSI has recently been used in the defence sector for a variety of reasons spanning from finding landmines to charting coastal regions. HSI has also been used on spacecraft, airplanes, and watercraft to collect precise spectral data for a variety of purposes [9], [10], [11], [12].

This paper focuses on Hyperspectral Image Classification (HSIC), which has piqued the interest of scientists leading to better capability for identifying land use and land cover, recognizing environmental threats, and city planning. Initially, machine learning-based algorithms were used for HSIC. The constant evolution of machine-learning-based methods has improved classification accuracy over time. Deep learning's breakthrough has had a significant impact on the precision of HSIC, making it among the most significant advances in this area. This article intends to provide an overview of some DL-based HSIC strategies that have been developed in recent years. The performance of different methods are compared and analyzed to find out the which method is better than others for which dataset. The motivational backgroud behind this survey study have been discussed in Section 2. Section 3 explains HSI and

E-mail address: anishsarkar983@gmail.com, nandi.3utpal@gmail.com, nayankrsarkar@gmail.com, chiranjit_changdar@yahoo.co.in, ableb.paul@gmail.com

its classification. Then, the application of deep learning in HSIC is discussed in Section 4. After that, recent deep learning based HSIC methods are explained in Section 5. The comparative evaluations are explained in Section 6. And in the last we conclude our study at Section7.

2. Motivation behind this survey

Deep learning has emerged as a potent technique for tackling complicated challenges in the world of technological developments. Among its numerous uses, HSIC has received a lot of interest. The comparative research of deep learning-based HSI classification systems is an intriguing and hard endeavour with enormous potential to revolutionise numerous sectors.

The objective of this comparative study of deep learning-based HSIC systems is to uncover the genuine capabilities of these algorithms in this area. We hope to determine the most effective and efficient ways for the classification of HSIs by comparing and contrasting various procedures. This type of research offers the potential to develop and improve existing methodologies, opening the door to a wide range of applications.

Researchers want to find unique insights, identify creative solutions, and inspire future breakthroughs in this interesting subject by undertaking this comparative study of deep learning-based HSIC systems. The outcomes of this study have the potential to impact the future of hyperspectral image analysis and bring in a new era of intelligent information extraction. The applications are diverse and transformational, ranging from boosting precision agriculture practises to assisting disaster response operations. The ambition to leverage the full potential of deep learning and hyperspectral imaging drives this effort, generating a synergy with the potential to revolutionise industries and positively benefit society as a whole.

3. Hyperspectral image and its classification

A hyperspectral image (HSI) can be depicted as a 3-dimensional hypercube $HC \in R^{BD(ROW \times COL)}$ as shown in Figure 1, which contains one-dimensional spectral information and two-dimensional spatial information for each sample. The number of spectral bands BD and the spatial dimension $ROW \times COL$ of the cube are both incorporated in the representation.

A. Spectral representation

Each pixel of the hypercube is distinguished from its neighbours in spectral representation, and each pixel is treated separately based on its own unique one-dimensional spectral signature. Spectral representation can be stated mathematically as $HC \in R^{BD}$, where BD denotes either the total number of spectral bands present or only the pertinent spectral bands selected using different band selection methods. In general, pixels with only the necessary spectral bands are picked for processing rather than all of the spectral bands. This selection procedure greatly reduces the overall dimension of the hypercube, minimizing redundancy and achieving greater class separability without a notable drop in information.

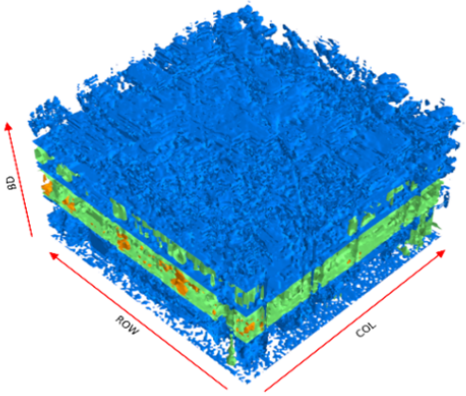


Figure 1. An overview of Hypercube

Bands for a particular dataset can be chosen either supervised or unsupervised. Without any prior knowledge of the class labels associated with the data, unsupervised algorithms pick the relevant spectral bands. This suggests that the decision is made only on the basis of the data's spectral features, with no preconceived class labels. Two well-known unsupervised band selection approaches are principal component analysis(PCA) and locally linear embedding [13]. Supervised band selection methods employ labelled samples to gain an understanding of the data distribution. Their purpose is to guarantee that data points from the same class are placed near to one another. Band selection techniques includes linear discriminant analysis(LDA), local Fisher discriminant analysis(LFDA) [14]. local discriminant embedding [15], and nonparametric weighted FE [16]. [14].

B. Representation of HSI in spatial domain

To escape the restrictions of HSI representation in spectral domain, another strategy is to utilize the spatial data of the pixels, which is structured in a matrix, $hc_i \times R^{(ROW \times COL)}$. This matrix is made up of ROW rows and COL columns of numerical numbers, providing for a comprehensive image depiction. Each band contains its own individual pixels, which are all represented in this matrix. Because of the strong correlation between neighboring pixels, pixels near each other are apt to be identical in character. As a result, when working with spatial depiction, it is extremely crucial to take into account information from adjacent pixels. This can be done by using a kernel- or pixel-centric window [17]. Extracting spatial information from HSI cubes. In addition, Deep Neural Networks (DNNs) have also been employed for this purpose.

The image texture provides significant spatial context for HSI. A texture inspection technique, such as the Gabor filter, can successfully capture textural information at a

variety of scales and orientations. Similarly, Local Binary Pattern(LBP) can express spatial textures in a rotation-invariant manner. Furthermore, DNNs can derive spatial information from HSI by treating them as image patches, as opposed to the conventional method of treating individual pixel as a spectral vector.

C. Spectra-Spatial Representation

Both spectral and spatial information are combined in this representation. Mathematically this representation of HSI can be expressed as $HC \in R^{BD(ROW \times COL)}$. This means that the vector of each pixel is examined using spectral characteristics while also taking into consideration spatial relevant data. In hyperspectral imaging(HSI), Spectra-Spatial representation techniques that use both spatial and spectral models frequently link the spatial data with the spectral vector [18]. In this literature most of the methods are based on this spectra-spatial representation of HSI.

D. HSI classification (HSIC)

The key focus of HSIC is to give a unique label to each individual pixel vector within an image cube based on its spectral and spatial characteristics. An HSI cube is mathematically stated as $HC = [hc_1, hc_2, hc_3, ..., hc_{BD}]^T \in R^{BD(ROW \times BD)}$, where BD is the count of spectral bands, each with $ROW \times COL$ samples that can be allocated to CLS distinct classes. The class identifier cls_i is assigned to the i^{th} sample in the HSI cube, $hc_i = (hc_{1,i}, hc_{2,i}, hc_{3,i}, ..., hc_{BD,i})^T$. One way to look at the classification issue is as an optimization problem. This means that a mapping function, $f_c(.)$, is used to transform the input data HC in a way that results in a matching level CLS. By doing so, the function reduces the discrepancy between the anticipated and the real output.

$$CLS = f_C(HC, \theta) \tag{1}$$

The parameter θ needs to be adjusted in order for transformation of the input data HC to be done in the form of $f_c: HC \rightarrow CLS$.

4. Application of deep learning for HSIC

Traditional RGB, monochrome, and multispectral imaging approaches cannot be directly used to HSI due to its unique engineering and statistical qualities with high-dimensional spectra-spatial data. As a result, techniques for HSI classification based on machine learning have been created. These methods typically necessitate the use of engineering skills and domain knowledge to build a collection of important features created by humans. Hand-crafted features can represent numerous picture attributes; therefore, they can be used with analyzed data, but they may not function with real data. Because the optimal features differ widely amongst data sets, it is challenging to strike a balance between robustness and discriminability. DL-based approaches were created to solve the drawbacks of standard models by learning the behaviour of any data without understanding its statistical distribution [19] and extracting linear and nonlinear features without any predetermined knowledge.

Considering the outlined DL potentials, there are several additional hurdles to consider when applying DL to HSI data. The existence of numerous continuous and narrow spectral bands with higher spectral resolution and lower spatial resolution throughout the electromagnetic spectrum, combined with a dearth of training data, causes the majority of these issues. While pixels that contain a lot of data about their spectral characteristics are advantageous for classification, the complexity of the calculations involved becomes a major issue.

Furthermore, as the number of factors increases, the complexity of processing such high-dimensional data increases. The “curse of dimensionality” refers to the occurrence of trouble with classification as the number of dimensions increases, and it significantly decreases the efficacy of supervised learning [20]. The model may be prone to overfitting due to a lack of adequate data for training and/or dependability problems (for example, the training examples may not contribute any additional features to the model or may have similar structures). When the number of frequency bands in the information is significantly less than the quantity of labelled training data, the Hughes effect, as described in [21], is noticed. Various variables, such as extreme variance within the same class due to uncontrolled reflectance values produced by outside effects and the existence of instrument noise during the recording process, may have a negative impact on HSIC [22]. Spectral mixing is a difficulty caused by the spatial resolution of HSI being too low or average. When the resolution of HSI pixels is decreased, they encompass a greater area of land, causing the problem of mixed spectral signatures. This makes distinguishing between dissimilar materials utilizing their spectral reflectance values challenging, especially in boundary areas where interclass similarity is strong [23]. The following are some of the major issues encountered when applying DL to HSIC.

A. Difficulties in training procedure

The Nondeterministic Polynomial Time Complete (NP-complete) problems in the area of DL for HSIC predict the results of training and optimizing DNNs by altering parameters extremely difficult [24], [25], [26]. This leads to the general assumption that training DNNs can be quite challenging [19], particularly with HSI where a substantial number of parameters must be adjusted. Nonetheless, the recent development of numerous optimization methods for Deep Convolutional Neural Networks (CNNs) has made the convergence process easier. Among the successful CNN optimization techniques frequently employed for any classification task is stochastic gradient descent [27] and its momentum variant [28], RMSProp [29], Adam [30], diffGrad [31], RAdam [32], gradient centralization (GC) [33], and AngularGrad [34].

B. Scarcity of Training data

Supervised DNNs require a large amount of training data in order to prevent substantial overfitting [35]. However, Hyperspectral Images (HSIs) have a high dimensionality that makes them challenging to handle, which is exacerbated by the scarcity of annotated training data. This creates a challenge for DNNs to be effective for HSIC as it requires extensive adjustments and tuning during the training phase [36].

C. Huge computational cost

DNNs encounter numerous challenges, particularly when dealing with significant amounts of data. This necessitates more memory bandwidth a higher processing expense, and the use of more memory storage [37]. Nonetheless, the use of modern data processing methods, such as distributed and parallel systems [38], [39], and high-performance computing (HPC) [23], can aid to overcome this issue. DNNs can now process enormous volumes of data more effectively and efficiently because to these strategies. Additionally, the use of distributed and parallel architectures aids in lowering the processing cost associated with big data collections. Furthermore, when dealing with massive amounts of data, HPC can help increase the performance of DNNs. DNNs can handle big data volumes using these sophisticated processing techniques, resulting in better results.

D. Degradation of training accuracy

DNNs are often believed to result in more complicated characteristics being extracted from data [40]; however, this may not be the case. As the network grows deeper, the gradients may erupt or lessen [41], which can have a significant detrimental effect on the overall performance of the model [40]. This is because adding more layers to a network can result in a gradient that is either too large or too small, which can prevent the model from learning the optimal parameters. Therefore, simply adding more layers to a network does not result in higher accuracy for all systems.

5. Recently developed Deep-Learning models for Hyperspectral Image Classification

In recent years lots of DL–based HSIC techniques were developed with the aim of increased accuracy. In this section, we will discuss some of the very recently developed classification models for HSI.

A. HSIC models based on 2D CNN

1) SP-CNN

Convolutional neural network(CNN) based pixel-wised HSI classification frameworks are already proven methods for spectral feature extraction, but with the network depth increment the spatial information of HSI cube losses gradually which leads to a classification result with lower accuracy. The training period of the network would be prolonged if we deepened it due to the necessity of tuning a large number of internal parameters. To address this issue F. Xie et al. proposed an HSIC model called Super Pixel CNN (SP-CNN) [42], where they used CNN architecture for spectral feature extraction of the HSI cube, and they used Super Pixel maps to extract the spatial information which is later fused the spectral features for further classification. Generally, there are two process modules in their proposed SP-CNN, the first one is down-sampling and the second one is Up-sampling. Convolution and pooling procedures are employed in the down-sampling methodology to derive the spectral characteristics from the HSI cube. In the up-sampling process module, the spatial information is recovered with the spectral information. Instead of pixels, they used Super-pixel which dramatically reduces the number of classification samples. The overall structure of their proposed SP-CNN method is shown in Figure 2.

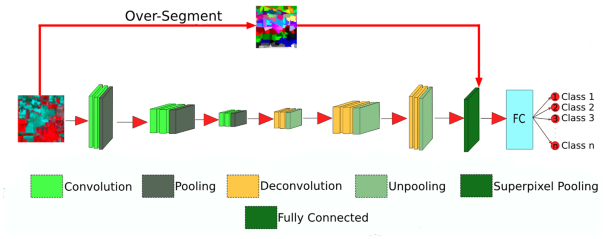


Figure 2. Architecture of Super Pixel Convolution Neural Network (SP-CNN) [42]

2) FCSN

Similar to the common segmentation jobs we have to focus on identifying the labels of each pixel for HSI. Recent researchers have primarily concentrated on the development of various DNNs to enhance the performance of HSI Classification for various public HSI datasets. H. Sun et al. found two major concerning generalization capabilities issues with those techniques: the first one is whether deep convolutional networks that have been designed are resistant to changes in spatial land-cover distributions, and the other question is the classification whether the results of classification at the boundaries of land-cover regions are accurate or not. They posed these two issues because, in actual uses, the spatial land-cover distributions of HSIs are changeable and complex due to shifts in imaging viewpoint or urban renewal. To overcome these issues, they proposed an HSI classification model called “Fully Convolutional Segmentation Network (FCSN)” [43]. The fundamental architecture of the FCSN for HSI classification and the conventional CNN-based network is similar, but the key distinction is that the FCSN does not utilize any Fully Connected (FC) layer, Flattening layer, or Global Pooling layer as core operations. The use of FC or global pooling layers in CNN-based techniques may result in spatial information loss in an HSI cube. This is because these layers are responsible for transforming the feature maps into vectors, and there is a risk of losing the spatial connections between the various spectral bands throughout this conversion. CNN-based methods generally concentrated on classifying the center pixel of the HSI cube, whereas their proposed FCSN concentrated on classifying all pixels of the HSI cube. To

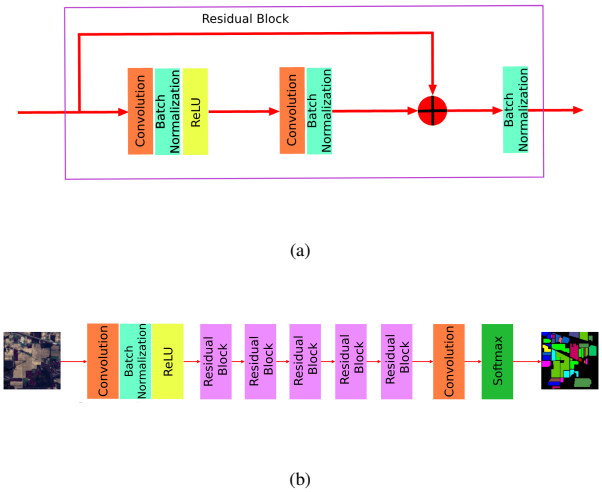


Figure 3. Architecture FCSN [43]: (a) Residual Block, (b) FCSN

implement FCSN they first deploy a residual connection to build a residual block and then they stack of residual block to implement their proposed model. The structure of FCSN and their proposed residual block are shown in Figure 3.

3) LS²CM

CNNs have become among the most common techniques for classifying HSIs. However, one major disadvantage of CNN-based systems is that they are numerically intensive, despite the large number of internal parameters they contain. To increase the accuracy of classification more complex CNN-based models are adopted, which makes the overall system computationally heavier. To solve this issue Z. Meng et al. proposed a computationally lightweight spectra spatial convolutional module LS²CM [44] to replace the conventional convolution layer of the HSI classification model, which dramatically reduces the number of parameters. They used two LS²CM based residual blocks in their proposed model, which were similar to the residual block of ResNET. In their proposed model they used 1×1 convolution for spectral feature extraction and 3×3 convolution for spatial feature extraction. Then they combined the input of the first LS²CM block and the output of the second LS ²CM block by using an additive shortcut operation. After each layer of the first LS²CM block, they implemented the BN and ReLU activation layer. They also did not use the ReLU component in the second LS²CM block. The diagram of LS²CM is shown in Figure 4.

4) SPRN

X. Zhanag et al. introduced an HSIC method called Spectral Partitioning Residual Network (SPRN) [45] by combining the Spectral Partitioning (SP) approach with 2D CNN techniques, which reduced the computational complexity of the overall methods. The input patches are partitioned into several groups with uniform breadth and spectral dimension and passed into some parallel convolution networks with an order of one group in one network

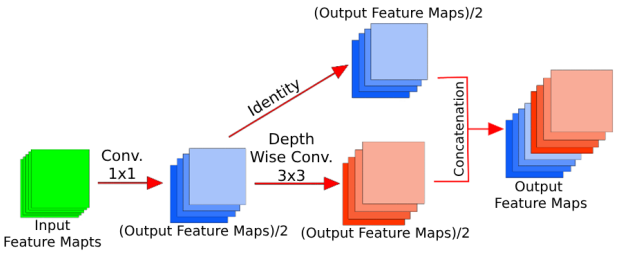
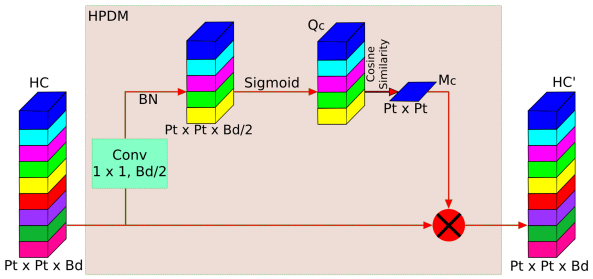


Figure 4. Diagram of their proposed LS²CM module [44]

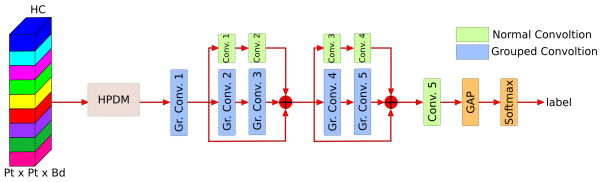
for feature extraction in their proposed methods. In real they used a group of convolutions instead of a parallel convolution network which gives them a similar outcome. They deployed residual blocks to make possible communication between internal convolution layers. Their used residual block typically has two branches: one local and one global. S parallel CNNs, which are made up of two paired convolutional layers, are used to develop the local branch. They extracted local spectral characteristics in a specific channel interval using clustered convolutional layers. The global ramify is made up of two standard convolution layers which are used for global spectral features from the entire input channel. They used the gap layer to combine spatial and spectral characteristics, as well as to fuse local spectral features with global spectral features. A Homogeneous Pixel Detection Module (HPDM) was employed by them to improve the operation of the Deep SPRN they created. This module generates a weight mask that can identify the significance of each pixel in a segment of the HSI cube. Measuring the commonalities between the middle and neighboring pixels yields the weight mask. The pixels having similar spectral signatures to respect the center pixel will get higher scores rather than the pixels having lower similarities. The architecture of the Homogeneous Pixel Detection Module (HPDM) and Spectral Partitioning Residual Network is shown in Figure 5.

5) SpectralNET

In general, two types of CNNs are used in HSIC. The first is the 3D CNN, which is good for spectral feature extraction but computationally heavy. The second one is the 2D CNN, which is computationally light and good for spatial feature extraction but not good for multiresolution processing of images. Some researchers also investigated a composite model composed of 3D CNN and 2D CNN, although their performance appeared to be limited for many databases. To overcome these limitations, T. Chakraborty et al. proposed a two-dimensional CNN-based model called SpectralNET [46], where they used wavelet transformation instead of three-dimensional CNN to bring out the spectral feature maps. On their proposed model they used factor analysis instead of PCA for the purpose of band selection. They deployed wavelet transform for spectral feature extraction and 2D CNN for spatial feature deployment. Then the feature channels wise concatenated and sent to dense layer



a. Architecture of Homogeneous Pixel Detection Module



b. Architecture of Spectral Partitioning Residual Network

Figure 5. Architecture of HPDM module and SPRN [45]

for classification. To prevent the problem of overfitting, a global mean pooling layer was introduced by the authors at the conclusion of all convolution layers prior to being transmitted to the dense layer. They also employed two dropout layers along with a batch normalization layer to prevent the problem of overfitting.

B. LKSSAN

G. Sun et al. proposed Larger Kernel Spectral and Spatial Attention Network(LKSSAN) [47] is a patch-based attention HSI classification network that includes data preparation, spectral-spatial attention module(SSAM), and classification components. The problems discussed include the difficulty of leveraging long-range 3-D features as well as the computing overhead. Spatial patching generates a large number of 3-D patches for local feature utilization. SSAM uses large kernel attention(LKA) and convolutional feed-forward (CFF) to prioritize informative long-range information through adaptive weighting. CFF enables the flexible recovery of spatial information in semantic features. The classification module refines 3-D feature maps using simple multilayer perception(SMLP) to produce class probability maps. In addition, a scale expansion block facilitates the use of spectral-spatial relationships in multilayer feature maps.

C. Residual Block

1) HResNETAM

We know that the Hyperspectral Image can have huge dimensional nonlinear data. Most multiscale feature extractors built on CNN are incapable of extracting both local and global characteristics at the same time. Hierarchical Feature Extractor helps to ease these drawbacks to a certain degree. However, the use of the layer-wise method may be responsible for the disappearance of gradients, as it requires

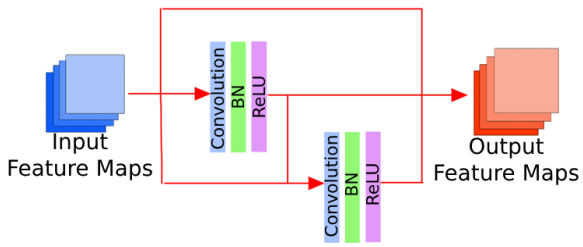


Figure 6. Block Diagram Hierarchical Residual Unit [48]

a large number of labeled samples for training. To solve this issue Z. XUE et al. proposed a network called Hierarchical Residual Network with Attention Mechanism (HResNETAM) [48]. In order to enhance the capacity of features to accurately classify HSI, the proposed approach incorporates independent spectral and spatial attention modules in conjunction with hierarchical spectral and spatial feature extractors. In their proposed feature extractor, they incorporated three tiers of residual blocks to produce several feature map subgroups with various receptive fields. Each residual block consists of three layers: 1. Convolutional (Conv) Layer, 2. Batch Normalization(BN) Layer and 3.Rectified Linear Unit(ReLU) activation layer. They doubled this hierarchical feature extractor to separately derive the spatial and spectral characteristics. To depict the interdependencies between channels, they used a spectral attention module in conjunction with their spectral feature extractor. They first reshaped and transposed the original input in the spectral attention module, then multiplied it with the original input, and the outcome of multiplication was transmitted into a softmax activation layer to yield the attention map. The architecture of their proposed Hierarchical Residual Unit is shown in Figure 6.

2) Ghostnet

Obtaining information from Hyperspectral Data cubes is a very challenging and time-consuming problem. As a solution to this issue, CNNs have been extensively used for HSIC in recent years. Although CNN-based HSI classification algorithms are very efficient, processing them takes a lot of time and consumes a lot of memory. So, it is very important to develop a lightweight CNN-based highly accurate HSI classification model that can be used for today's applications on mobile and embedded systems for different platforms. M. E. Paoletti et al. [49] proposed an HSI classification approach by combining a ghost-module framework with a CNN-based classifier that reduced the computational expenses and achieved a higher performance and accuracy. To reduce the spectral dimension of the input HSI cube, their proposed network Ghostnet implements a simple stem unit, which consists of a convolution, normalization, and activation layer. Then they used a stack of three ghost bottlenecks, which helped their model to avoid the problem of overfitting and also helped to avoid the problem

of degradation and gradient loss during the forward and backward propagation. To address the declining-accuracy phenomenon, they used the shortcut connection and they designed their ghost bottleneck block based on residual bottleneck architecture [50], [51]. In their proposed model every bottleneck is a combination of two stack ghost modules. Each ghost module is a combination of a primary convolution layer to generate essential feature maps of the input features, and a grouped convolution layer (the size of the group equals the number of input channels) which confirms that each linear kernel will be applied on only on channel of the essential feature maps. Then they concatenated the output block of the primary convolution layer with the grouped convolution layer and sent it to the next ghost bottleneck block. To enhance the channel-wise feature response, they deployed a Squeeze-and-Excitation (SE) block between every two ghost module blocks. Their proposed SE block is a combination of one adaptive average pooling layer and two pointwise convolution layers. Just like residual bottleneck [40], except for the first ghost bottleneck, the number of channels is increased in the first ghost module and then decreased in the second ghost module of each ghost bottleneck. We can see that the first module extends thrice the number of channels, then the SE module first squeezes and then expands them again to combine spatial information across the channel, and at the end, the second ghost module compresses the number of channels to the desired one. They have to manage the size of input features by deploying several convolutions in the shortcut connection before performing the final sum because the second ghost bottleneck increased the number of input features at the time of output. Finally, they deployed a convolution pooling layer that collects all the feature maps and converts them into a vector before sending it to a classifier, which is a combination of two layers Fully Connection (FC), and Multi-Layer Perception(MLP).

D. 3D CNN

1) SCNN

Today's DL-based HSIC algorithms face the following key challenges: Mixed pixels undermine the credibility of original spectral data with spatial information. The Sandwich CNN (SCNN) model by H. Gao et al. [52] address this issue by working on preprocessed data features, substituting spatial aspects with spectral data. They introduce a Spectral Feature Enhancement (SFE) module to improve spectral data extraction and preserve spatial information. This module includes two spectral blocks, each consisting of three components: The convolution (Conv) layer, Batch Normalization (BN) layer, and ReLU activation layer. They also use a Spatial block for spatial data extraction using pointwise convolution. A multi-scale feature fusion method is employed to leverage different spatial scales. After spatial feature extraction, another spectral feature extraction block, similar to the first, is used to capture abstract characteristics from spatially fused spectral features. This enables the utilization of spectral characteristics from multiple positions. The overview of SCNN is shown in Figure 7.

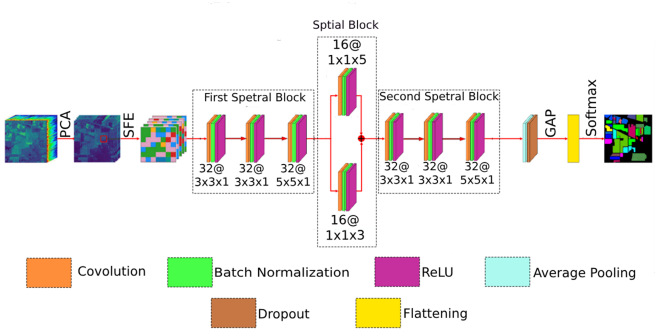


Figure 7. Block Diagram of Sandwich Convolution Neural Network (SCNN) [52]

2) SSAD

We all know that every deep learning-based method has a common problem: it requires a huge number of labeled samples to achieve higher accuracy. Generally, these labeled samples can be collected in two ways: 1. Field Investigation, 2. Visually recognition from high-resolution images. Labeled samples that are collected from field investigations can provide higher accuracy, but it is a very expensive process, and acquiring an adequate number of marked samples for the training process can be challenging at times. To address this issue, they posit a method called the self-supervised learning approach with adaptive distillation (SSAD) [53] for training DNNs using a multitude of unlabeled samples. Their proposed techniques are divided into two modules: 1. Adaptive knowledge distillation with Self Supervised Learning(SSL) 2. SSL with three-dimensional transformation. The knowledge distillation strategy in the adaptive knowledge distillation module is essentially a transfer learning strategy in which knowledge from one highly précised larger network is transferred to a small size network, where the larger network is utilized for soft label generation and the training of the smaller network is guided by those generated soft labels. The soft labels are prepared by comparing the resemblance of the unlabeled samples to the specified object classes. Their proposed soft labels, which are adaptive in nature, include the possibility of unlabeled samples about each preset object class. We all know that horizontal rotation in the spatial realm has no effect on pixel prediction in HSI classification. As a result, in the 3D transformation portion, the Hyperspectral cube is rotated and mirrored in the spatial domain to improve the resilience of the HSIC model. By rotating the ground-truth map horizontally in a similar manner that the input HSI is turned, the cross-entropy loss can be determined. In their implementation, they established four horizontal rotations: $0^{\cdot}$, $90^{\cdot}$, $180^{\cdot}$, and $270^{\cdot}$. In the spectral domain, they also rotate the hyperspectral cube to determine the arrangement of the spectral sequence. They defied two types of spectral sequences: the first one is the frequencies arranged in ascending order, which is labeled as 1, and the second is the frequencies arranged in descending order, which is labeled as 0. The Progressive Convolution Network (PCN)

is a full CNN that can perform geometric transformations in both spatial and spectral domains. The HSI rotates with the outcome of forward propagation in the PCN.

E. Hybrid Methods

1) Consolidated CNN (C-CNN)

We all know that deep learning technique has a common problem of overfitting when we use these techniques with data having higher dimensions. Chang et. al. used ReLU activation, L2 Regularization, and Dropout to overcome this issue of overfitting for deep learning with high dimensional data in their proposed HSI classification model called "Consolidated CNN (C-CNN)" [54]. They used two units to build their proposed model. They used one unit for Spectral feature learning and they used another unit for Spatial feature learning. In their Spectral Feature Learning unit, they used three consecutive 3D Convolution Layer having the convolution and kernel size of (8, (3×3×7)), (16, (3×3×5)) and (32, (3×3×3)). They have deployed a Max-pooling layer to succeed in the operations of consecutive three convolution layers. The Spectral feature learning unit was fed into the Spatial feature learning unit, utilizing three 2D Convolution layers with convolution kernel sizes of (128, (1×1)), (256, (3×3)) and (64, (1×1)) that were followed by a Max-pooling layer. Hyper Spectral Image has a large number of channels, which may be a concern for higher computational complexity. To overcome this problem, they deployed 1×1 2D Convolution to cut down on both the size of the convolution kernel and the complexity while retaining all of the features. This type of convolution also provided them a better learning ability and generalization.

6. Comparitive Evaluation

To demonstrate the benefits and downsides of their recommendations, research investigations frequently give a full experimental evaluation. However, different experimental settings, such as varying numbers or proportions of training, validation, and evaluation samples, may be used in these works. To achieve a fair comparison of works of literature, it is vital to adopt the same experimental circumstances.

The experimental parameters consist of the same samples (geographic locations should not differ between models) and the number of samples to be picked for each training session in the process of cross-validation. Because these samples are typically picked at random, it is very possible that they will vary for various models if implemented at different times.

The majority of current literature has a problem with training and test sample overlap. Although the training and validation samples are chosen at random, the complete dataset is used for testing, resulting in a biassed model with high accuracy. To avoid this, the samples in this work are chosen at random, but the intersection between them is empty.

TABLE I. Overview of three popular HSI datasets

Particulers	**IP** [55]	**PU** [56]	**SA**[58]
Year	1992	2001	2001
Source	AVIRIS	ROSIS-03	AVIRIS
Spatial	145 × 145	610 × 340	512 × 217
Spectral	220	115	224
Wavelength	400-2500	430-860	-
Samples	21025	207400	54129
Classes	16	9	16
Sensor	Aerial	Aerial	Aerial
Resolution	20m	1.3m	3.7m

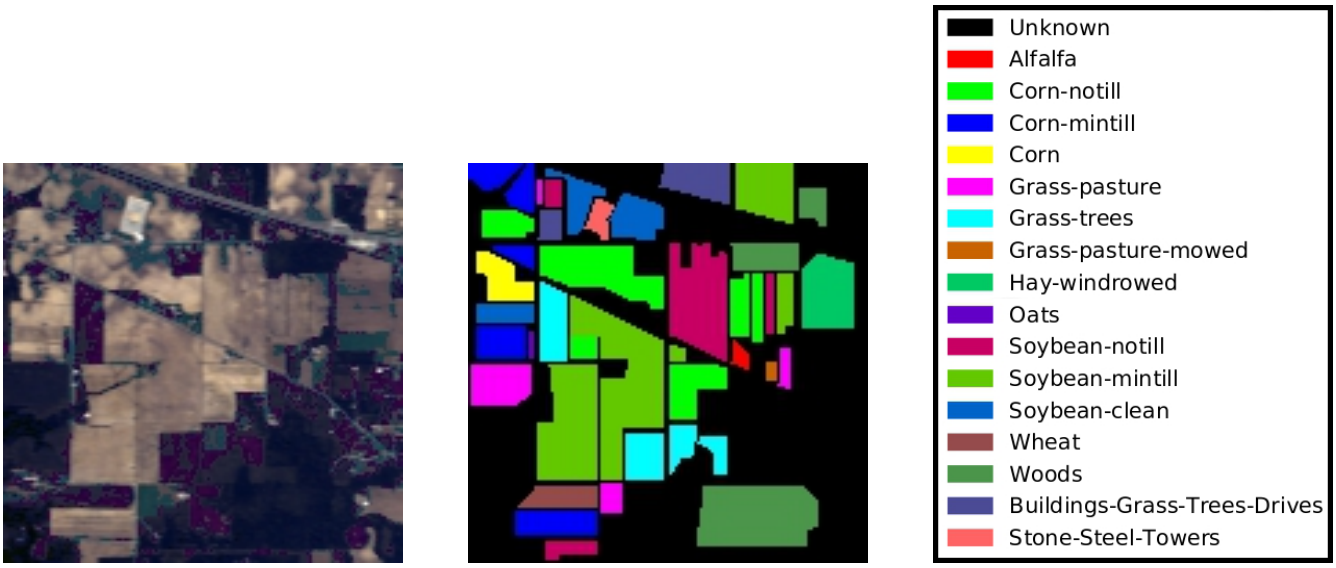
A. Utilized Datasets

In order to conduct a comparative experimental evaluation, we utilized the three datasets that are most widely used: Indian Pines(IP), University of Pavia(PU), and Salinas Scene(SA). These datasets have a wide range of applications in remote sensing, making them ideal for comparison studies to gain deeper insights into the methods discussed in the preceding part. The overall information of the used classes was presented in a tabular format in Table I and Table II.

The Indian Pines (IP) [55] dataset was acquired using the AVIRIS instrument from the Indian Pines test location in northeastern Indiana. The sample had 224 spectral bands in the wavelength range from 400 nm to 2500 nm after removing 24 blank and corrupted bands. The image had a resolution of 20 m per pixel (MPP) and was 145x145 pixels in size. It was made up of 16 different plant classes, each with its own description and ground truth maps. Figure 8 depicted the RGB origin and Ground Truth for the Indian Pines (IP) dataset.

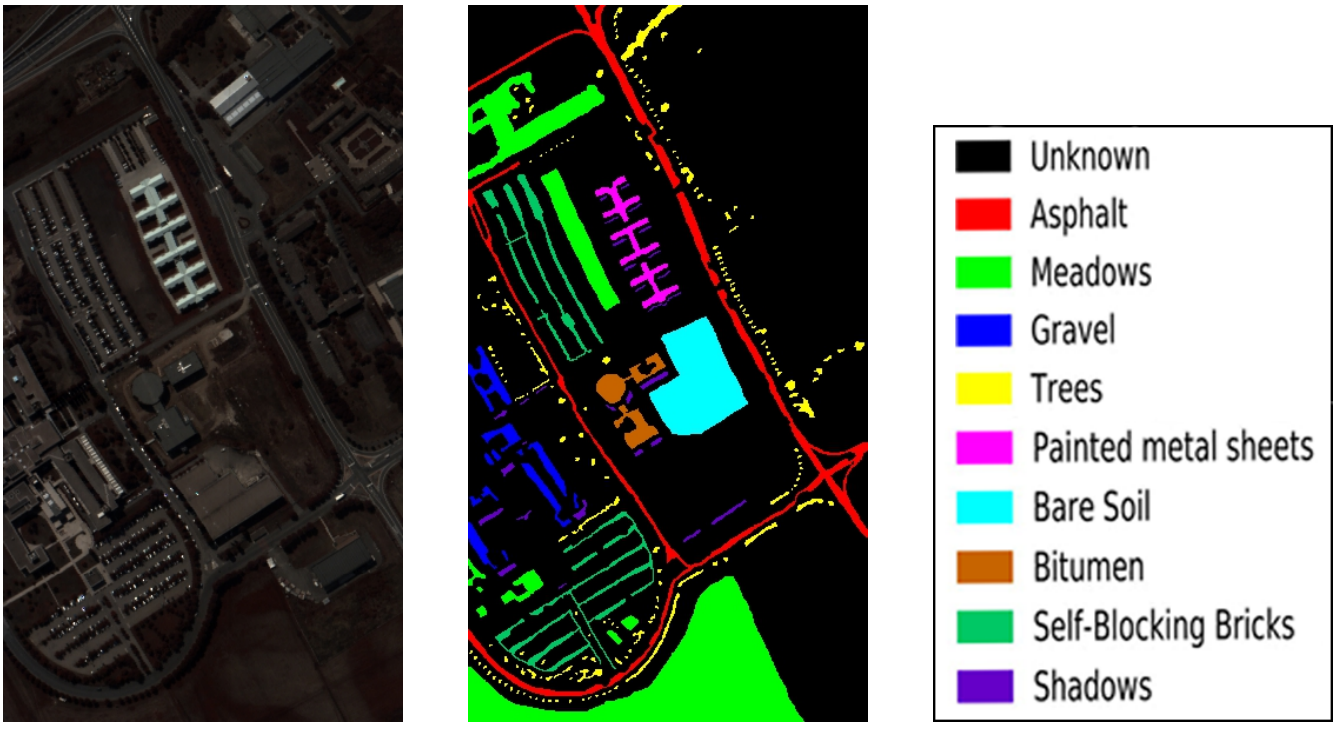
The University of Pavia (PU) [56] dataset is a large collection of hyperspectral images captured by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor [57] in an agricultural region near Pavia, Italy. The dataset is 610x340 pixels in area and has a resolution of 2.5 m per pixel (MPP), with 103 spectral bands varying in wavelength from 430 to 860 nm. Asphalt, fields, gravel, trees, metal sheets, bare soil, bitumen, brickwork, and shadows are among the nine types of ground cover. The RGB origin and Ground Truth for the University of Pavia (PU) dataset are displayed in Figure 9.

The Salinas Scene [58] dataset is an accumulation of hyperspectral imaging data that was gathered over Salinas Valley, California by the AVIRIS sensor. The dataset consists of 145x145 pixels and 224 spectral bands, with each pixel representing an area of approximately 1.3 meters by 1.3 meters on the ground. The area covered was 512 lines by 217 samples. The image, which included flora, barren soils, and wine fields, was only available as at-sensor radiance data. The Salinas ground truth was made up of 16 different classes. Figure 10 depicted the RGB origin and Ground Truth for the Salinas scene (SA) dataset.



(a) (b) (c)

Figure 8. The overview of IP dataset [55]: (a) RGB origin, (b) Ground truths , (c) legend



(a) (b) (c)

Figure 9. The overview of PU dataset [56]: (a) RGB origin, (b) Ground truths , (c) legend

TABLE II. The landcover categories of the three datasets

Class No.	IP [55]	PU [56]	SA [58]
1	Alfalfa	Asphalt	Brocoli_green_weeds_1
2	Corn-notill	Meadows	Brocoli_green_weeds_2
3	Corn-mintill	Gravel	Fallow
4	Corn	Trees	Fallow_rough_plow
5	Grass-pasture	Painted Metal Sheets	Fallow_smooth
6	Grass-trees	Bare Soil	Stubble
7	Grass-pasture-mowed	Bitumen	Celery
8	Hay-windrowed	Self-Blocking Bricks	Grapes_untrained
9	Oats	Shadows	Soil_vinyard_develop
10	Soybean-notill		Corn_senesced_green
11	Soybean-mintill		Lettuce_romaine_4wk
12	Soybean-clean		Lettuce_romaine_5wk
13	Wheat		Lettuce_romaine_6wk
14	Woods		Lettuce_romaine_7wk
15	Building-Grass-Trees-Drives		Vinyard_untrained
16	Stone-Steel-Towers		Vinyard_vertical

Figure 10. The overview of SA dataset [58]: (a) RGB origin, (b) Ground truths , (c) legend

B. Comparative Results and Analysis

The experiments were carried out on a Dell Inspiron 15 Gaming 5577 laptop equipped with an Intel i5 7300HQ CPU, 8GB DD4 2400MHz RAM, and an NVIDIA GTX 1050 4GB GPU. To guarantee a fair comparison, we restricted the analysis to 15 spectral bands across all datasets and employed a consistent patch size of 25×25. The datasets were divided into training and testing sets using a 70:30 ratio, with 70% utilized for training and 30% for testing. Every method underwent 100s epochs during the training period.

Several recent studies indicate that certain image classification methods are effective. SCNN, SpectralNET, SP-CNN, C-CNN, HResNETAM, LS^{2}CM, FCSN, Ghostnet, SPRN, and SSAD are some of these methods. These studies compared the experimental results of these methods, which are discussed in the previous sections, in order to determine which are the most effective. We utilized three metrics to analyze classification performance and selected the most effective HSIC model based on equations 2, 3, and 4: Overall Accuracy (OA), Average Accuracy (AA), and Kappa (K).

$$OA = \frac{\sum_{p=1}^{Nos} Curr(p,p)}{T} \tag{2}$$

In equation 2, OA represents the overall accuracy, T represents the total number of samples for each class, Nos represents the total number of landcover classes in the dataset, and $Curr(p,p)$ represents the algorithm's current classified samples.

$$AA = \frac{1}{T} \frac{\sum_{p=1}^{Nos}(Curr(p,+))}{\sum_{p=1}^{Nos}(Curr(+,p))} \tag{3}$$

Equation 3 uses the following notation: AA stands for average accuracy, while T represents the total number of landcover classifications. $Curr(p,+)$ indicates the total number of samples in which landcover objects are classified into specific classes, and $Curr(+,p)$ specifies the total number of samples in which landcover objects actually belong to specific classes.

$$Kp = \frac{T\sum_{p=1}^{Nos} Curr(p,p) - \sum_{p=1}^{Nos}(Curr(p,+)Curr(+,p)}{T^2 - \sum_{p=1}^{Nos}(Curr(p,+)Curr(+,p)} \quad (4)$$

In equation 4, the term Kp refers to Kappa. $Curr(p,+)$ defines the total number of samples, in which landcover objects are split into certain classes. where $Curr(+,p)$ defines the total number of samples in which landcover objects genuinely belong to a specific class.

On Table III, Table IV and Table V we have shown the class wised accuracy of the different HSIC methods for the which we have studied for the public HSI dataset named PU, IP and SA respectively. Table VI contains a comparison table of OA, AA, and Kp values for various methods. This table allows you to quickly compare the performance of the methods under consideration.

On Table III, we can see that SpectralNET [46], outperforms for all of the classes of University of Pavia datasets. For the first and eighth class of University of Pavia datasets named as 'Asphalt' and 'Self-Blocking Bricks' respectively, HResNETAM [48] also gives the identical performance like SpectralNET [46]. In case of the seventh class of University of Pavia datasets named as 'Bitumen' SCNN [52], SP-CNN [42] gives the similar performance like SpectralNET [52]. For the nineth class of University of Pavia datasets named as 'Shadows' FCSN [43] can also provides highest accuracy that is 100% like SpectralNET [46]. HResNETAM [48] can give more than 98% accuracy for eight classes out of nine classes of University of Pavia dataset. For the fifth class of University of Pavia dataset named as "Painted Metal Sheets" almost all of methods which we have stuidied give more than 99% accuracy except SSAD [53]. From the Table III we can find that SCNN [52], SpectralNET [46] and HResNETAM [48] gives a consistence performance for almost all of the classes of University of Pavia dataset, and in case of the other HSIC methods we can see a performance massive drops for few classes.

We can find from the Table IV, that SpectralNET [46] outperforms for most of the classes of Indian Pines dataset. For the nineth class of Indian Pines dataset named as 'Oats' SCNN [52] and SP-CNN [42] gives the identical accuracy and also perform better than SpectralNET [46]. In case of fifth class of Indian Pines dataset named as 'Grass-pasture' SP-CNN [42] gives better accuracy than SpectralNET [46]. In case of seventh class of Indian Pines dataset named as "Grass-pasture-mowed", SCNN [52], SpectralNET [46], SP-CNN [42] and SPRN [45] shows identical performance. For the eighth, thirteenth and fourteenth class of Indian Pines datasets named as 'Hay-windrowed', 'Wheat' and 'Wood' SCNN [52] and SpectralNET [46] gives similar performance. For the first, fourth and sixth class of Indian Pines dataset named as 'Alfalfa', 'Corn' and 'Grass-pasture-mowed' SpectralNET [46] and SP-CNN [42] provides the similar performance. From the Table IV we can found the that the SCNN [42], HResNETAM [46], LS2CN [47] and SPRN [45] provides the consistence performance, while other HSIC methods showcasing a performance drop for the few classes of University of Pavia datasets.

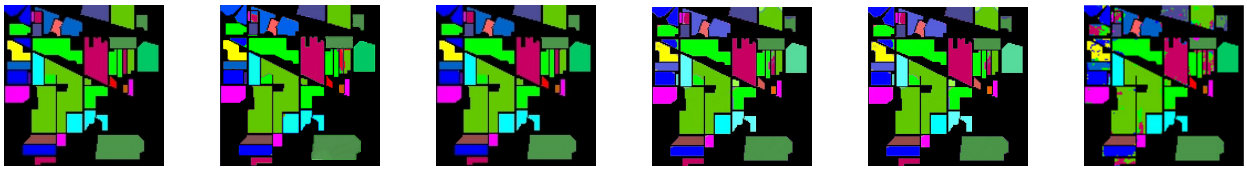
On Table V, we can see that SpectralNET [46] performs better than other HSIC methods for all of the classes of Salinas Scene dataset. For the first class of Salinas Scene dataset named as '*Brocoli_green_weeds_1*' SCNN [52] and SpectralNET [46] performs identical, and for the second class of Salinas Scene datasets named as '*Brocloli_green_weeds_2*' SCNN [52], SpectralNET [46] and SPRN [45] showcasing similar performance. In case of third class of Salinas Scene datasets named as 'Fallow' SCNN [52], SpectralNET [46] and SP-CNN [42] provides similar performance. In case of nineth class of Salinas Scene dataset, C-CNN [54] and and SPRN [45] giving similar performance like SpectralNET [46]. For the twelfth, thirteenth and fourteenth class of of Salinas Scene dataset named as "*Lettuce_romaine_5wk*", "*Lettuce_romaine_6wk*" and "*Lettuce_romaine_7wk*" SP-CNN [42] also provide the similar performance like SpectralNET [46]. Besides SpectralNET [46], C-CNN [54] also provides very good performance that is more than 97% for all of classes present in Salinas Scene datasets. On the Table V, it is clearly visible that SCNN [52], SpectralNET [46], C-CNN [54], HResNETAM [48], LS2CN [44], FCSN [43], Ghostnet [49] and SPRN [45] almost giving a consistence performance for all of the classes in Salians Scene dataset, while SP-CNN [42] shows a drop of accuracy performance for few classes of Salinas Scene dataset. We primarily compare HSI classification methods in this survey using three predefined datasets: 1. Indian Pines (IP), 2. University of Pavia (PU), and 3. Salinas Scene. Some methods, however, have only been tested on one or two of these datasets. The effectiveness of all the convolutional feature extractors previously stated has been assessed. The evaluation's findings are shown in Table VI. We all know that the results of the HSIC algorithms are graphically represented in the form of classification maps. In the resultant classification map, each pixel is labeled with a specific class in the form of unique color. Overall, a classification map is very helpful for detailed understanding and analysis of objects of different class on earth's surface. The classification of different HSIC methods that we have studied for the Indian Pines dataset and Salinas Scene dataset are shown in the Figure 11 and Figure 12 respectively.

TABLE III. Comparison of class wised accuracy for PU dataset

Class No.	SCNN [52]	SpectralNET [46]	SP-CNN [42]	C-CNN [54]	HResNETAM [48]	LS^{2}CM [44]	FCSN [43]	Ghostnet [49]	SPRN [45]	SSAD [53]
1.	97.61	**100**	91.12	97.63	100	96.93	95.31	92.04	98.48	86.63
2.	99.33	**100**	98.73	99.75	99.83	99.62	92.96	97.77	99.41	84.53
3.	91.84	**100**	99.67	83.75	96.61	89.51	78.45	82.04	85.36	82.21
4.	99.55	**100**	78.34	94.63	98.99	94.70	96.32	96.12	95.41	91.62
5.	97.27	**100**	99.35	99.73	99.27	99.17	99.54	99.13	99.89	93.92
6.	95.65	**100**	97.10	98.69	98.62	98.26	95.50	71.83	97.82	89.43
7.	**100**	**100**	**100**	95.31	99.81	91.12	93.74	88.01	98.92	91.83
8.	96.51	**100**	99.46	91.46	100	96.61	95.94	97.23	96.13	89.02
9.	97.54	**100**	88.46	90.21	99.78	96.38	100	98.16	97.51	90.02
OA	97.83	**99.99**	93.18	97.08	99.8	97.58	93.79	92.83	97.92	86.80
AA	97.82	**99.98**	93.78	94.57	99.29	95.81	93.19	91.37	96.75	88.80
Kp	97.64	**99.98**	92.36	96.13	99.72	96.79	91.86	90.20	97.23	83.40

TABLE IV. Comparison of class wised accuracy for IP dataset (The best performances are marked by bold)

Class No.	SCNN [52]	SpectralNET [46]	SP-CNN [42]	C-CNN [54]	HResNETAM [48]	LS^{2}CM [44]	FCSN [43]	Ghostnet [49]	SPRN [45]	SSAD [53]
1.	97.62	**100**	**100**	42	97.05	95.95	87.80	82.44	96.560	99.32
2.	99.30	**100**	89.62	58.15	98.20	99.72	84.44	89.19	98.99	80.91
3.	99.20	**100**	98.12	54.27	98.77	99.54	86.48	73.19	97.24	79.72
4.	99.53	**100**	**100**	18.34	98.32	99.91	93.91	54.95	98.73	83.51
5.	95.86	99.01	**99.78**	80.90	96.52	96.08	94.24	89.64	95.49	82.12
6.	99.24	**100**	**100**	92.99	95.34	98.39	97.26	94.47	99.54	83.02
7.	**100**	**100**	**100**	70.00	98.51	99.63	88.00	0	**100**	89.42
8.	**100**	**100**	99.56	98.47	98.89	99.28	99.78	99.53	99.98	97.82
9.	**100**	78.003	**100**	76.50	99.17	99.45	61.12	92.01	94.45	96.01
10.	96.45	**100**	88.86	67.67	98.90	99.52	80.21	93.36	96.90	80.41
11.	99.32	**100**	89.36	83.06	98.82	97.72	83.75	89.88	98.15	82.91
12.	99.45	**100**	95.37	45.38	98.57	99.09	88.18	80.85	98.82	81.21
13.	**100**	**100**	99.42	76.06	99.11	99.00	98.91	94.25	98.97	91.51
14.	**100**	**100**	99.43	95.70	97.72	99.01	95.51	98.13	99.72	91.82
15.	97.70	**100**	99.43	54.27	95.42	99.09	74.64	49.72	96.08	87.11
16.	98.81	98.01	**100**	59.24	99.18	99.45	98.80	74.64	95.95	97.71
OA	98.93	**99.86**	94.45	73.33	97.08	98.41	87.96	88.31	98.29	84.30
AA	98.90	**99.84**	96.43	67.06	97.05	97.63	88.31	78.77	97.85	87.8
Kp	98.78	**99.98**	93.44	69.43	97.08	97.48	86.30	86.30	98.05	79.90



(a) (b) (c) (d) (e) (f)

Figure 11. Comparison among outputs of different methods for IP: (a) Ground Truth, (b) Using SCNN, (c) Using Spectralnet, (d) Using SPCNN, (e) Using SPRN, (f) Using SSAD

TABLE V. Comparison of class wised accuracy for SA dataset (The best performances are marked by bold)

Class No.	SCNN [52]	SpectralNET [46]	SP-CNN [42]	C-CNN [54]	HResNETAM [48]	LS2CM [44]	FCSN [43]	Ghostnet [49]	SPRN [45]	SSAD [53]
1.	**100**	**100**	**100**	99.94	99.98	93.14	99.16	93.21	99.99	89.02
2.	**100**	**100**	99.89	99.63	99.96	99.09	99.94	94.12	**100**	88.96
3.	**100**	**100**	**100**	99.80	99.92	97.75	97.06	97.32	99.95	99.24
4.	99.89	**100**	99.89	99.49	99.97	96.73	98.09	98.23	98.70	98.78
5.	99.94	**100**	98.45	99.08	99.97	99.42	99.21	96.33	98.91	97.89
6.	99.99	**100**	99.56	99.94	99.94	97.17	98.60	97.05	99.97	98.32
7.	99.99	**100**	99.26	99.96	99.80	98.77	99.24	99.14	99.97	98.90
8.	99.61	**100**	95.63	98.39	99.83	95.39	99.82	99.23	93.71	99.48
9.	99.83	**100**	99.53	**100**	99.94	97.98	98.31	99.34	**100**	99.66
10.	99.42	**100**	99.64	97.92	99.91	94.75	99.20	95.41	97.08	96.60
11.	99.08	**100**	99.22	99.943	99.97	96.65	98.49	99.23	98.39	95.54
12.	99.26	**100**	**100**	99.93	99.97	96.98	99.99	99.32	99.96	94.49
13.	96.72	**100**	**100**	99.15	99.90	97.13	99.18	99.14	99.97	98.21
14.	96.55	**100**	**100**	98.90	99.96	97.48	99.18	99.31	99.05	98.11
15.	99.36	**100**	87.36	98.87	99.97	98.51	99.59	98.24	93.32	98.02
16.	99.43	**100**	99.49	99.29	99.94	95.85	99.34	95.38	98.67	97.52
OA	99.58	**100**	95.99	99.22	99.93	97.10	97.42	95.43	97.46	96.84
AA	99.27	**100**	95.97	99.39	99.91	97.37	96.52	96.21	98.45	95.40
Kp	99.53	**100**	95.46	99.13	99.45	97.09	95.71	91.05	97.17	95.40

TABLE VI. Comparison of OA, AA and Kappa of the described methods (The best performances are marked by bold)

	Indian Pines			University of Pavia			Salinas Scene		
	OA	AA	Kp	OA	AA	Kp	OA	AA	Kp
SCNN [42]	98.93	98.9	98.78	98.93	98.9	99.78	99.58	99.27	99.53
SpectralNET [43]	**99.86**	**99.84**	**99.98**	**99.99**	**99.98**	**99.98**	**100**	**100**	**100**
SP–CNN [44]	94.45	96.43	93.44	93.18	93.78	92.36	95.99	95.97	95.46
C–CNN [45]	73.33	67.06	69.43	97.08	94.57	96.13	99.22	99.39	99.13
HResNETAM [46]	97.08	970.5	97.08	99.8	99.29	99.72	99.93	99.91	99.45
LS^2CM [47]	98.41	97.63	97.48	97.58	95.81	96.79	97.10	97.37	97.09
FCSN [48]	87.96	88.31	86.3	93.79	94.19	91.86	97.42	96.52	95.71
Ghostnet [49]	88.31	78.77	86.3	92.83	91.37	90.2	95.43	97.46	96.84
SPRN [52]	98.29	97.85	98.05	97.92	96.75	97.23	97.46	98.45	97.17
SSAD [53]	84.30	87.80	79.90	86.80	88.8	83.40	96.84	95.40	95.40

On the Figure 11, it is clearly visible that the classification maps generated by SCNN [52], and SpectralNET [46] is almost equivalent to original ground truth of Indian Pines datasets. But in case of the classification maps generated by SP-CNN [52], SPRN [45] and SSAD [53] we can find some overlapping of classes. That means SP-CNN [42], SPRN [45] and SSAD [53] had lower classification accuracy compare to SCNN [42] and SpectralNET [43] for the classification of classes of Indian Pines dataset.

On the Figure 12, it is clearly visible that the classification maps generated by SCNN [52], and SpectralNET [46] is almost equivalent to original ground truth of Salinas Scene datasets. But in case of the classification maps generated by SP-CNN [42], C-CNN [54] and SPRN [45] we can find some overlapping of classes. That means SP-CNN [42], C-CNN [54] and SPRN [45] had lower classification accuracy compare to SCNN [52] and SpectralNET [46] for the classification of classes of University of Pavia dataset.

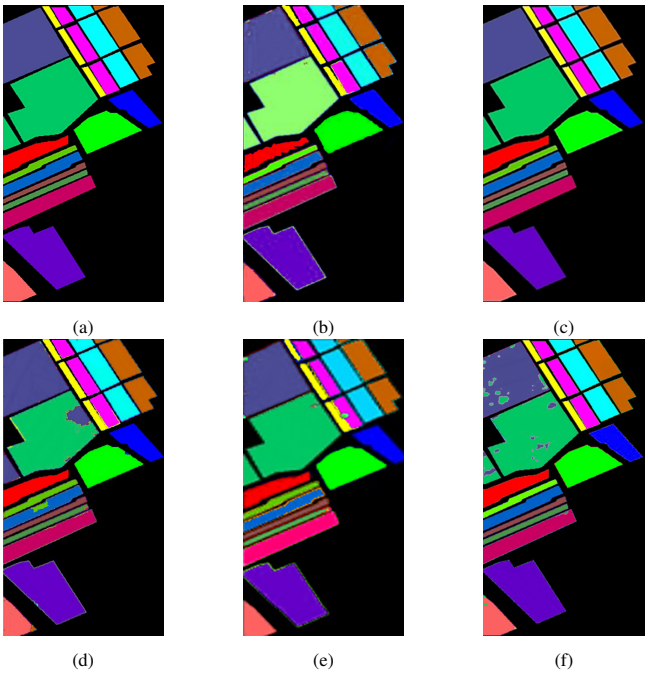


Figure 12. Comparison among outputs of different methods for SA: (a) Ground Truth, (b) Using SCNN, (c) Using Spectralnet, (d) Using SP-CNN, (e) Using C-CNN, (f) Using SPRN

In summary, the comparative results and analysis of the numerous HSIC approaches reveal a range of strengths and shortcomings. SCNN significantly improves spectral data extraction while preserving spatial information, resulting in high classification accuracy. SpectralNET's balanced method, which employs wavelet transformation and 2D CNNs, improves performance in capturing both spectral and spatial features. SP-CNN overcomes the challenge of spatial information loss in deep CNNs by adding superpixel maps, resulting in precise classification with minimal computational overhead. The C-CNN's separate units for spectral and spatial feature learning reduce computational complexity, but it may struggle with complex spatial patterns.HResNETAM's attention mechanisms enhanced feature extraction, increasing classification accuracy, but may necessitate more training data. LS^{2}CM's lightweight convolutional modules provide efficiency without sacrificing accuracy, but FCSN's attention to classifying all pixels improves accuracy in complex spatial distributions. Ghostnet is a lightweight approach with higher accuracy, but SPRN uses spectral partitioning and residual networks to minimize computational complexity. Finally, SSAD's self-supervised learning strategy takes the benefit of unlabeled examples to increase performance. Each method has distinct advantages and disadvantages, emphasizing the necessity of selecting a suitable strategy depending on the specific application needs.

7. Conclusion

A common option for practical applications is HSI data because of the fascinating information it contains. Its deployment potential is also being improved by machine learning techniques. With cutting-edge DNNs including SCNN, SpectralNET, SP-CNN, C-CNN, HResNETAM, LS^{2}CM, FCSN, Ghostnet, SPRN, and SSAD, we examined current advancements in HSIC in this article. We selected a few of these studies to run tests on benchmark HSI datasets. The performance of SpectralNET is comparatively better than its counterparts.

Despite the advancements made in HSIC, further work is required to improve generality. The dearth of labelled data is a significant issue because HSI data is hard to come by and DNNs need a lot of labelled training data. To improve classification accuracy, simultaneous exploitation of spectral and spatial data should be investigated. There should be the development of efficient and parallel HSIC architectures to satisfy the computing requirements of time-critical HSIC applications.

References

[1] M. Ahmad, A. Khan, A. M. Khan, M. Mazzara, S. Distefano, A. Sohaib, and O. Nibouche, "Spatial prior fuzziness pool-based interactive classification of hyperspectral images," *Remote Sensing*, vol. 11, no. 9, p. 1136, 2019.

[2] D. Hong, W. He, N. Yokoya, J. Yao, L. Gao, L. Zhang, J. Chanussot, and X. Zhu, "Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 2, pp. 52–87, 2021.

[3] H. Ayaz, M. Ahmad, A. Sohaib, M. N. Yasir, M. A. Zaidan, M. Ali, M. H. Khan, and Z. Saleem, "Myoglobin-based classification of minced meat using hyperspectral imaging," *Applied sciences*, vol. 10, no. 19, p. 6862, 2020.

[4] M. H. Khan, Z. Saleem, M. Ahmad, A. Sohaib, H. Ayaz, and M. Mazzara, "Hyperspectral imaging for color adulteration detection in red chili," *Applied Sciences*, vol. 10, no. 17, p. 5955, 2020.

[5] Z. Saleem, M. H. Khan, M. Ahmad, A. Sohaib, H. Ayaz, and M. Mazzara, "Prediction of microbial spoilage and shelf-life of bakery products through hyperspectral imaging," *IEEE Access*, vol. 8, pp. 176 986–176 996, 2020.

[6] M. Zulfiqar, M. Ahmad, A. Sohaib, M. Mazzara, and S. Distefano, "Hyperspectral imaging for bloodstain identification," *Sensors*, vol. 21, no. 9, p. 3045, 2021.

[7] H. Ayaz, M. Ahmad, M. Mazzara, and A. Sohaib, "Hyperspectral imaging for minced meat classification using nonlinear deep features," *Applied Sciences*, vol. 10, no. 21, p. 7783, 2020.

[8] M. H. Khan, Z. Saleem, M. Ahmad, A. Sohaib, H. Ayaz, M. Mazzara, and R. A. Raza, "Hyperspectral imaging-based unsupervised adulterated red chili content transformation for classification: Identification of red chili adulterants," *Neural Computing and Applications*, vol. 33, no. 21, pp. 14 507–14 521, 2021.

[9] N. Abdallah, "Food quality monitoring using hyperspectral data," Ph.D. dissertation, Politecnico di Torino, 2020.

[10] F. Xing, H. Yao, Y. Liu, X. Dai, R. L. Brown, and D. Bhatnagar, "Recent developments and applications of hyperspectral imaging for rapid detection of mycotoxins and mycotoxigenic fungi in food products," *Critical reviews in food science and nutrition*, vol. 59, no. 1, pp. 173–180, 2019.

[11] M. Ahmad, "Ground truth labeling and samples selection for hyperspectral image classification," *Optik*, vol. 230, p. 166267, 2021.

[12] W. Jia, S. van Ruth, N. Scollan, and A. Koidis, "Hyperspectral imaging (hsi) for meat quality evaluation across the supply chain: Current and future trends," *Current Research in Food Science*, vol. 5, pp. 1017–1027, 2022.

[13] Y. Fang, H. Li, Y. Ma, K. Liang, Y. Hu, S. Zhang, and H. Wang, "Dimensionality reduction of hyperspectral images based on robust spatial information using locally linear embedding," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 10, pp. 1712–1716, 2014.

[14] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis." *Journal of machine learning research*, vol. 8, no. 5, 2007.

[15] H.-T. Chen, H.-W. Chang, and T.-L. Liu, "Local discriminant embedding and its variants," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 2. IEEE, 2005, pp. 846–853.

[16] B.-C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 5, pp. 1096–1105, 2004.

[17] B. Kumar, O. Dikshit, A. Gupta, and M. K. Singh, "Feature extraction for hyperspectral image classification: A review," *International Journal of Remote Sensing*, vol. 41, no. 16, pp. 6248–6287, 2020.

[18] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected topics in applied earth observations and remote sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.

[19] S. Chen and Y. Wang, "Convolutional neural network and convex optimization," *Dept. of Elect. and Comput. Eng., Univ. of California at San Diego, San Diego, CA, USA, Tech. Rep*, 2014.

[20] R. Bellman, *Adaptive Control Processes: A Guided Tour*, 5th ed., ser. Princeton Legacy Library. New Jersey: Princeton University Press, 1961, vol. 2045.

[21] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE transactions on information theory*, vol. 14, no. 1, pp. 55–63, 1968.

[22] ——, "On the mean accuracy of statistical pattern recognizers," *IEEE transactions on information theory*, vol. 14, no. 1, pp. 55–63, 1968.

[23] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and remote sensing magazine*, vol. 1, no. 2, pp. 6–36, 2013.

[24] Q. Nguyen and M. Hein, "Optimization landscape and expressivity of deep cnns," in *International conference on machine learning*. PMLR, 2018, pp. 3730–3739.

[25] M. Ahmad, S. Shabbir, R. A. Raza, M. Mazzara, S. Distefano, and A. M. Khan, "Artifacts of different dimension reduction methods on hybrid cnn feature hierarchy for hyperspectral image classification," *Optik*, vol. 246, p. 167757, 2021.

[26] M. Ahmad, M. Mazzara, and S. Distefano, "Regularized cnn feature hierarchy for hyperspectral image classification," *Remote Sensing*, vol. 13, no. 12, p. 2275, 2021.

[27] L. Bottou *et al.*, "Stochastic gradient learning in neural networks," *Proceedings of Neuro-Nımes*, vol. 91, no. 8, p. 12, 1991.

[28] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural networks*, vol. 12, no. 1, pp. 145–151, 1999.

[29] G. Hinton, N. Srivastava, and K. Swersky, "Neural networks for machine learning lecture 6a overview of mini-batch gradient descent," *Cited on*, vol. 14, no. 8, p. 2, 2012.

[30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[31] ——, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[32] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han, "On the variance of the adaptive learning rate and beyond," *arXiv preprint arXiv:1908.03265*, 2019.

[33] H. Yong, J. Huang, X. Hua, and L. Zhang, "Gradient centralization: A new optimization technique for deep neural networks," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*. Springer, 2020, pp. 635–652.

[34] S. K. Roy, M. E. Paoletti, J. M. Haut, S. R. Dubey, P. Kar, A. Plaza, and B. B. Chaudhuri, "Angulargrad: A new optimization technique for angular convergence of convolutional neural networks," *arXiv preprint arXiv:2105.10190*, 2021.

[35] D. Erhan, A. Courville, Y. Bengio, and P. Vincent, "Why does unsupervised pre-training help deep learning?" in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2010, pp. 201–208.

[36] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," pp. 279–317, 2019.

[37] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, M. Hasan, B. C. Van Essen, A. A. Awwal, and V. K. Asari, "A state-of-the-art survey on deep learning theory and architectures," *electronics*, vol. 8, no. 3, p. 292, 2019.

[38] A. Plaza, D. Valencia, and J. Plaza, "An experimental comparison of parallel algorithms for hyperspectral analysis using heterogeneous and homogeneous networks of workstations," *Parallel Computing*, vol. 34, no. 2, pp. 92–114, 2008.

[39] A. Plaza, J. Plaza, A. Paz, and S. Sanchez, "Parallel hyperspectral image and signal processing [applications corner]," *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 119–126, 2011.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[41] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994.

[42] F. Xie, Q. Gao, C. Jin, and F. Zhao, "Hyperspectral image classification based on superpixel pooling convolutional neural network with transfer learning," *Remote sensing*, vol. 13, no. 5, p. 930, 2021.

[43] H. Sun, X. Zheng, and X. Lu, "A supervised segmentation network for hyperspectral image classification," *IEEE Transactions on Image Processing*, vol. 30, pp. 2810–2825, 2021.

[44] Z. Meng, L. Jiao, M. Liang, and F. Zhao, "A lightweight spectral-spatial convolution module for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[45] X. Zhang, S. Shang, X. Tang, J. Feng, and L. Jiao, "Spectral partitioning residual network with spatial attention mechanism for hyperspectral image classification," *IEEE transactions on geoscience and remote sensing*, vol. 60, pp. 1–14, 2021.

[46] T. Chakraborty and U. Trehan, "Spectralnet: Exploring spatial-spectral waveletcnn for hyperspectral image classification," *arXiv preprint arXiv:2104.00341*, 2021.

[47] G. Sun, Z. Pan, A. Zhang, X. Jia, J. Ren, H. Fu, and K. Yan, "Large kernel spectral and spatial attention networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.

[48] Z. Xue, X. Yu, B. Liu, X. Tan, and X. Wei, "Hresnetam: Hierarchical residual network with attention mechanism for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3566–3580, 2021.

[49] M. E. Paoletti, J. M. Haut, N. S. Pereira, J. Plaza, and A. Plaza, "Ghostnet for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 12, pp. 10 378–10 393, 2021.

[50] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, "Deep pyramidal residual networks for spectral–spatial hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 740–754, 2018.

[51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[52] H. Gao, Z. Chen, and C. Li, "Sandwich convolutional neural network for hyperspectral image classification using spectral feature enhancement," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3006–3015, 2021.

[53] J. Yue, L. Fang, H. Rahmani, and P. Ghamisi, "Self-supervised learning with adaptive distillation for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2021.

[54] Y.-L. Chang, T.-H. Tan, W.-H. Lee, L. Chang, Y.-N. Chen, K.-C. Fan, and M. Alkhaleefah, "Consolidated convolutional neural network for hyperspectral image classification," *Remote Sensing*, vol. 14, no. 7, p. 1571, 2022.

[55] D. A. L. Marion F. Baumgardner, Larry L. Biehl, "220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3," Sep 2015. [Online]. Available: https://purr.purdue.edu/publications/1947/1

[56] X. Huang and L. Zhang, "A comparative study of spatial approaches for urban mapping using hyperspectral rosis images over pavia city, northern italy," *International Journal of Remote Sensing*, vol. 30, no. 12, pp. 3205–3221, 2009.

[57] ——, "A comparative study of spatial approaches for urban mapping using hyperspectral rosis images over pavia city, northern italy," *International Journal of Remote Sensing*, vol. 30, no. 12, pp. 3205–3221, 2009.

[58] A. Plaza, P. Martinez, J. Plaza, and R. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 466–479, 2005.

Anish Sarkar A Ph.D. Scholar in the Department of Computer Science at Vidyasagar University delves into the realms of Computer Vision, Deep Learning, and Image Analysis. He has published few quality articles.

Utpal Nandi With M.Tech (2009), and Ph.D. (2018) from the University of Kalyani, he is presently an Assistant Professor at Vidyasagar University. His focus encompasses data and image compression, Computer Vision, and Deep Learning. Authoring 42+ papers in international journals, conferences and book chapters, he holds patents and copyrights.

Nayan Kumar Sarkar Nayan Kumar Sarkar currently pursuing Ph.D. in Computer Science and Engineering at NERIST, his research focuses on machine learning, deep learning, and image classification. He has contributed to numerous publications in international journals, book chapters, and conferences.

Chiranjit Changdar Chiranjit Changdar holds Ph.D. from the University of Calcutta. As an assistant professor at Belda College, he specializes in soft computing, fuzzy mathematics, and artificial intelligence. He has published extensively and supervises three Ph.D. scholars.

Bachchu Paul He is currently working as an Assistant Professor, Department of Computer Science, Vidyasagar University. His research interest is in Automatic Speech Recognition and Speech Signal Analysis on the native languages using different deep learning approaches.