ALGORITHM DEVELOPMENT FOR HYPERSPECTRAL IMAGE CLASSIFICATION

M.Tech. Thesis

By

ABHISHEK DARWAI

DEPARTMENT OF ASTRONOMY, ASTROPHYSICS & SPACE ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

May 2025

ALGORITHM DEVELOPMENT FOR HYPERSPECTRAL IMAGE CLASSIFICATION

A THESIS

Submitted in partial fulfillment of the requirements for the award of the degree

of

Master of Technology

by

ABHISHEK DARWAI

DEPARTMENT OF ASTRONOMY, ASTROPHYSICS & SPACE ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

May 2025

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **ALGORITHM DEVELOPMENT FOR HYPERSPECTRAL IMAGE CLASSIFICATION** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF TECHNOLOGY** and submitted in the **DEPARTMENT OF ASTRONOMY, ASTROPHYSICS & SPACE ENGINEERING, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from 23/07/2023 to 16/05/2025 under the supervision of **Dr. Saurabh Das, Associate Professor, IIT Indore**.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

16/05/2025

Signature of the student with date
(NAME OF THE M.TECH. STUDENT)

This is to certify that the above statement made by the candidate is correct to the best of my/our knowledge.

16/05/2025
Signature of the Supervisor of
M.Tech. thesis (with date)
(NAME OF SUPERVISOR)

ABHISHEK DARWAI has successfully given his/her M.Tech. Oral Examination held on **05/05/2025**

16/05/2025
Signature(s) of Supervisor(s) of MTech thesis
Date:

Programme Coordinator, M.Tech.
Date: 16-May-2025

Manoneeta Chakraborty

Convenor, DPGC
Date: 16/05/2025

HoD, DAASE
Date:

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis guide, Dr. Saurabh Das, for his continuous support, insightful guidance, and encouragement throughout the course of this work. His expertise and mentorship have been invaluable.

I extend my sincere thanks to my seniors Lekhraj, Vaibhav, Kavita, and Parvathy for their valuable suggestions, constant motivation, and generous help at every stage of the project.

I am also grateful to my dear friends Sneha, Anand, and Sana for their unwavering support, positivity, and companionship that made this journey memorable.

Abstract

Recent advancements in remote sensing technology have significantly enhanced our ability to monitor and interpret the Earth's surface with high precision and consistency. Among these advancements, the shift from multispectral to hyperspectral imaging marks a pivotal transformation. Hyperspectral imaging, with its hundreds of contiguous and narrow spectral bands, enables more accurate material and terrain classification due to its superior spectral resolution.

In this work, we explore the potential of deep learning-based methods to classify hyperspectral images by leveraging both spatial and spectral information. Our study implements and compares traditional 2D Convolutional Neural Networks (2D CNNs), which focus on spatial feature extraction, and 3D Convolutional Neural Networks (3D CNNs), which simultaneously capture spatial and spectral correlations, offering a more holistic representation of hyperspectral data.

In addition to these baseline models, we propose and three novel architectures tailored specifically for hyperspectral image classification: SpectraEdgeNet, which incorporates edge-aware spectral filtering for sharper class boundaries; SmoothSpectralNet, which enhances spectral continuity and intra-class consistency through spectral smoothing; and HyperViTNet, a Vision Transformer-inspired model that effectively captures long-range dependencies in spectral information. These models were evaluated on two benchmark datasets—Pavia University and Indian Pines—each offering diverse and challenging scenarios for hyperspectral classification. Experimental results demonstrate that the proposed models outperform traditional CNN-based approaches in terms of accuracy and generalization, highlighting their potential for advanced hyperspectral image analysis in remote sensing applications. The HyperViTNet model performs the better than other two proposed model.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYM

HSI – Hyperspectral Imaging.

2D CNN - Two-dimensional convolutional neural network.

3D CNN - Three-dimensional convolutional neural network.

SpectraEdgeNet - A Spatially-Filtered 3D Convolutional Neural Network for Enhanced Hyperspectral Image Classification.

SmoothSpectralNet -A Gaussian-Enhanced 3D CNN Architecture for Robust Hyperspectral Classification.

HyperViTNet- A CNN-Transformer Hybrid Framework for Spectral-Spatial Feature Learning.

HSIC - Hyperspectral Image Classification

SVM - Support Vector Machine

ICA - Independent Component Analysis

t-SNE - t-distributed Stochastic Neighbor Embedding

Chapter 1

Introduction

1.1 Overview of Remote Sensing

Remote sensing is a scientific technique which enables the acquisition, observation, and interpretation of information about objects, events without any physical contact. It involves collecting data from a distance, typically through airborne or spaceborne platforms like satellites or drones. These platforms capture electromagnetic energy either reflected or emitted from Earth's surface, providing valuable insights into natural and man-made features. Remote sensing systems are generally classified on the basis of their energy source into two categories: passive and active sensors.

Passive sensors depend on natural sources of energy, primarily sunlight. They detect energy that is either reflected such as visible light or re emitted like thermal radiation from the Earth. These sensors operate effectively only when sunlight is available and include devices like cameras, optical scanners, and microwave radiometers. In contrast, the active sensors emit their own energy to illuminate the target and measure the backscattered signal. This allows them to collect data regardless of the time of day or weather conditions. Common examples include RADAR and LiDAR systems. Fig.1 shows difference between active and passive instrument.

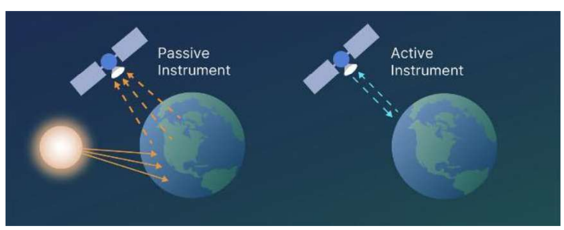

Fig. 1 Difference between active and passive instrument [Credits- NASA]

The use of remote sensing has significantly transformed how we view and analyze the Earth’s surface. High-resolution imagery makes it possible to monitor land use, vegetation, water bodies, and infrastructure in great detail. This technology is particularly critical in sectors such as agriculture, forestry, environmental monitoring, and resource management. By converting energy measurements into images and analytical products, remote sensing supports informed decision-making across a wide range of disciplines.

The development of remote sensing began with multispectral imaging, which captures data across a few broad spectral bands. Although effective for large-area assessments, it had limited ability to distinguish between materials with similar spectral characteristics. Hyperspectral imaging marked a major advancement by capturing hundreds of narrow, continuous spectral bands across electromagnetic spectrum. Each pixel in the hyperspectral image contains spectral signature, allowing for precise material identification and more advanced environmental and industrial applications.

1.2 Hyperspectral remote sensing

Multispectral imaging was one of the earliest breakthroughs in remote sensing, enabling the capture of images across the limited number of broad spectral bands—typically including visible and near-infrared wavelengths. This technique proved useful for observing large-scale features such as vegetation cover, water bodies, and land use. However, its ability to differentiate between materials with similar spectral characteristics was limited. While it provided a solid foundation for remote sensing applications, the level of detail it offered wasn't always sufficient for more precise analysis. Fig. 2 shows hyperspectral remote sensing setup.

This limitation led to a major leap forward with the introduction of **hyperspectral imaging (HSI)**. Unlike multispectral systems, hyperspectral sensors collect data in hundreds of narrow, continuous spectral bands. This provides a much richer and more detailed view of the Earth's surface. With HSI, scientists can now identify specific minerals, detect subtle chemical compositions, and distinguish between different crop types or manmade structures with far greater accuracy. This fine spectral resolution has made HSI an invaluable tool for applications in environmental monitoring, urban planning, precision agriculture, and resource management.

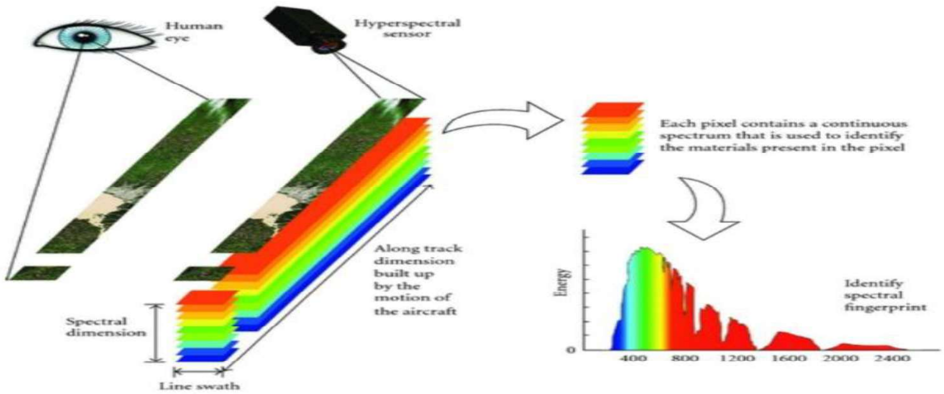

Fig. 2 Hyperspectral remote sensing [1]

That said, the strength of hyperspectral imaging—its high dimensionality—also introduces complexity. Processing and analysing such massive volumes of data is a significant challenge. Traditional methods often fall short, which is why researchers are increasingly turning to advanced deep learning algorithms. These models are particularly used for handling complex and high-dimensional datasets and can uncover patterns and insights that would otherwise be difficult to detect. This integration of artificial intelligence with hyperspectral data analysis is helping to unlock its full potential.

What makes HSI truly unique is how it combines imaging with spectroscopy. While the human eye sees in just three-color bands that are red, green, and blue—HSI divides light into hundreds of narrow spectral bands. Each of these bands generates a grayscale image, and together they form what's known as a "data cube." Each pixel in this cube contains detailed spectral information that serves as a signature for the material it represents [2]. This capability allows scientists to identify and map the chemical and physical properties of different surfaces. As a result, HSI has gained significant attention in fields like biomedical imaging, geoscience, environmental surveillance, object detection, and remote sensing, making it one of the most reliable technologies in the data-driven world today. Fig. 3 shows difference between multispectral and hyperspectral bands whereas Fig. 4 shows spectral signature of hyperspectral data.

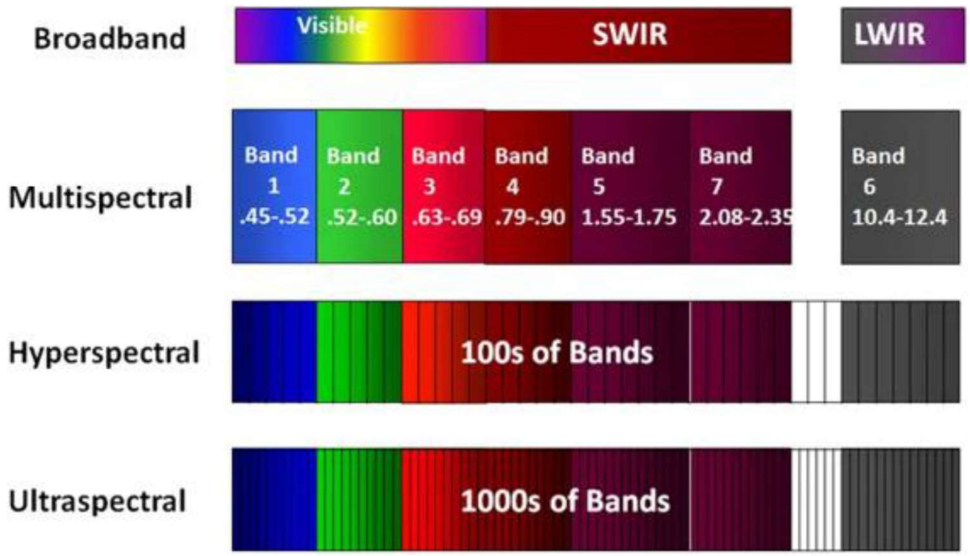

Fig. 3 Difference between Multispectral and Hyperspectral [1].

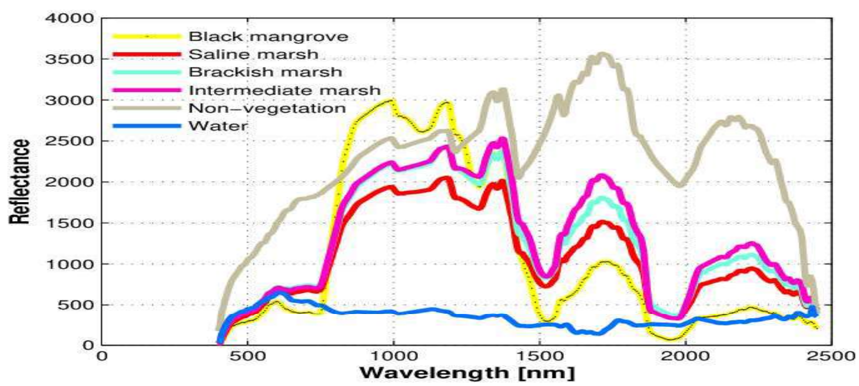

Fig. 4: Mean spectral signatures of the hyperspectral data [2].

1.3 Problem Statement

Hyperspectral image (HSI) classification is a critical task in the field of remote sensing with applications in agriculture, , urban planning , environmental monitoring, and defence. However, existing classification algorithms face several key challenges that limit their effectiveness. One major issue is spectral mixing, where pixels contain mixed signatures from multiple materials, leading to misclassification. Additionally, high dimensionality of hyperspectral data results in increased computational complexity, making many traditional machine learning methods inefficient [3]. Current approaches, such as Random Forests, Support Vector Machines (SVM), and deep learning models, often struggle with long execution times, sensitivity to noise, and difficulty in distinguishing subtle spectral variations between similar classes.

To address these challenges, this research aims to develop a classification algorithm that improves both accuracy and computational efficiency for hyperspectral image classification. The proposed method will focus on processing time through optimized feature selection and dimensionality reduction while enhancing classification performance by effectively handling spectral variability and mixed pixels. Key objectives include improving discrimination between spectrally similar classes, minimizing overfitting in high-dimensional data, and ensuring robustness with limited training samples. By overcoming these limitations, the new algorithm will provide a more reliable and scalable solution for hyperspectral image analysis, benefiting a wide range of real-world applications.

1.4 Organization of thesis

This research work is organized into six chapters to present a comprehensive study on hyperspectral image classification: Chapter 1 introduces the research background, objectives, and overall structure of the thesis. Chapter 2 provides fundamental concepts of hyperspectral remote sensing, explaining key principles and characteristics of hyperspectral data. Chapter 3 presents a critical review of existing literature, analyzing previous approaches and identifying research gaps in hyperspectral image classification. Chapter 4 discusses established classification algorithms and introduces the novel methodology proposed in this research, detailing its architecture and implementation. Chapter 5 describes the experimental datasets (Pavia and Indian Pines), including their specifications, preprocessing steps, and preparation for classification tasks. Chapter 6 demonstrates experimental results, provides comparative analysis with existing methods, and discusses findings through quantitative metrics and visual interpretations. Chapter 7 discusses conclusion and future work.

Chapter 2

Introduction to Hyperspectral Remote Sensing

2.1 Hyperspectral Imaging

Hyperspectral imaging (HSI) was first introduced in the 1980s by A. F. H. Goetz at NASA's Jet Propulsion Laboratory, with the development of the Airborne Imaging Spectrometer (AIS) [4]. Today, NASA continues to collect HSI data using advanced instruments like the Airborne Visible Infrared Imaging Spectrometer (AVIRIS), which captures over 200 spectral bands across visible to mid-infrared wavelengths [5]. Each pixel in the resulting data represents a unique spectral signature of the object it covers. Sensor systems are defined by different resolutions: spatial (the smallest object distinguishable), spectral (the smallest wavelength difference detected), and radiometric (the smallest energy level recorded). These depend on factors like altitude, viewing angle, and signal-to-noise ratio. However, image quality can be affected by noise, bad pixels, and atmospheric interference, making preprocessing essential. Technological advancements have improved spatial resolution in newer HSI sensors. For example, Hyperion (onboard EO-1) offers 30m resolution, while airborne sensors like ROSIS-3 and CASI-1500 can achieve resolutions as fine as 1.7m and provide rich spectral data [6]. These improvements help in accurately identifying and differentiating ground objects using contextual spectral and spatial information.

HSI primarily includes detecting known and unknown materials, classifying and segmenting images based on dominant objects or materials, and estimating the composition of mixed materials within a single pixel—known as hyperspectral unmixing. HSI data is structured as a cube of spectral measurements where each pixel provides highly localized information. However, this pixel-level representation often lacks structural context and may not align well with the broader interpretation needed for real-world decision-making [7]. Since HSI captures data across many narrow, continuous spectral bands, each image layer reflects the scene at a specific wavelength. Together, these layers form a 3D data cube in which each pixel contains a detailed spectrum, revealing how light interacts with the materials present in that region.

2.2 Application of Hyperspectral Remote Sensing

The vast amount of data provided by hyperspectral imaging (HSI) cannot be fully leveraged using existing image analysis methods. Given the broad range of real-world applications, such as agriculture monitoring, land use, ecological and environmental assessment, mineral exploration, target detection, change detection, surveillance, and ground cover classification, significant research is dedicated to the pre- and post-processing of HSI data [8]. The processing tasks in HSI are numerous, but they can generally be categorized into several key functions. These include classification (assigning labels to each pixel), dimensionality reduction (simplifying the data for easier analysis), spectral unmixing (estimating the material composition of each pixel), target and anomaly detection (identifying rare spectral signatures), and change detection (identifying important changes in consecutive HSI scenes). This dissertation primarily focuses on hyperspectral image classification (HSIC), particularly on integrating supervised and the semi-supervised classification techniques to improve performance using various classifiers, such as generative, discriminative, parametric classifiers and ensemble. The goal of HSIC algorithm is to efficiently assign unique label to each of the sample, ensuring it fits well within a specific class.

Chapter 3

Literature Review

Hyperspectral imaging has emerged as one of the powerful analytical tool in remote sensing, offering unprecedented capabilities for material identification and classification. Unlike conventional imaging systems that capture only a few spectral bands, hyperspectral sensors record hundreds of narrow, contiguous bands ranging from the visible to infrared wavelengths [9]. This rich spectral data enables detailed characterization of materials on the basis their unique spectral signatures, which arise from the interaction between the electromagnetic radiation and molecular vibrations. This rich spectral data enables detailed characterization of surface materials based on their unique spectral signatures, which arise from the interaction between electromagnetic radiation and molecular vibrations [10]. The classification of hyperspectral imagery presents several fundamental challenges that have driven extensive research efforts. Spectral variability remains a primary concern, where identical materials can exhibit different spectral responses due to intrinsic factors like compositional variations and surface roughness, as well as extrinsic factors including atmospheric conditions, illumination geometry, and sensor characteristics. Another significant challenge is the curse of the dimensionality, where the high number of bands relative to available training samples leads to statistical instability and overfitting in classification models. Additionally, the presence of mixed pixels containing multiple materials further complicates the classification process [11].

Traditional machine learning approaches have formed the foundation of hyperspectral image classification. Support Vector Machines gained early popularity due to effectiveness in high-dimensional spaces and ability to handle limited training samples through kernel methods [12]. Random Forest classifiers offered robust performance by combining multiple decision trees and providing built-in feature importance analysis. However, these conventional methods often struggled to capture the spectral-spatial relationships in hyperspectral data and showed limitations in handling nonlinear data distributions.

To address the dimensionality challenge, researchers developed various feature reduction techniques. Principal Component Analysis (PCA) emerged as one of the most used linear transformation method for dimensionality reduction, while more sophisticated approaches like manifold learning algorithms attempted to preserve nonlinear relationships in the data [13]. Band selection methods, including information-theoretic approaches and evolutionary algorithms, provided alternative strategies to identify the most discriminative spectral bands. These preprocessing steps significantly improved computational efficiency while attempting to maintain classification accuracy.

The integration of spatial information with spectral data marked a great advancement in classification performance. Recognizing that neighbouring pixels in natural scenes often share similar characteristics, researchers developed spatial-spectral methods that substantially improved classification accuracy. Techniques like extended morphological profiles employed mathematical morphology operations to extract spatial features, while Markov Random Fields provided probabilistic frameworks for modelling spatial dependencies [14]. Superpixel-based approaches and segmentation methods offered alternative ways to incorporate spatial context, demonstrating particular effectiveness in heterogeneous landscapes.

Recent years have shown the growing dominance of deep learning in HSIC. CNN revolutionized the field by automatically learning hierarchical features which capture both spectral and spatial information. 3D CNN architectures specifically designed for hyperspectral data demonstrated remarkable success in various applications. More recently, transformer-based models have shown promise in capturing long-range dependencies within spectral signatures [15]. While these advanced methods achieve state-of-the-art performance, they often require substantial computational resources and labelled datasets, presenting practical challenges for real-world implementation.

Current research directions focus on addressing the limitations of existing approaches. Semi-supervised and the self-supervised learning techniques aim to reduce the dependency on large labelled datasets. Lightweight neural network architectures seek to maintain high accuracy while reducing computational demands [16]. There is growing interest in physics-informed machine learning that incorporates domain knowledge about spectral characteristics and mixing phenomena. Explainable AI methods are being developed to improve the interpretability of complex models, which is particularly important for scientific applications.

The field continues to evolve with emerging technologies and novel applications. Hyperspectral sensors are becoming more compact and affordable, enabling new deployment scenarios. The integration of hyperspectral data with other sensors, such as LiDAR (Light Detection and Ranging) and thermal imaging, opens new possibilities for multimodal analysis [17]. Real-time processing capabilities are being enhanced through edge computing and optimized algorithms. These developments are expanding the practical applications of hyperspectral classification while presenting new research challenges in data fusion, computational efficiency, and system integration.

Chapter 4

Hyperspectral Image Classification Algorithms

4.1 Use of Classification Models

Hyperspectral imaging (HSI) offers vast potential for material identification and scene analysis due to rich spectral information. Each pixel in an HSI image captures hundreds of narrow and the contiguous spectral bands, making it possible to analyse subtle differences in materials and their compositions. However, this richness in data also presents significant challenges—mainly the curse of dimensionality, high data redundancy, and complex spectral variability. To effectively classify such data, machine learning and deep learning algorithms have been explored, aiming to extract quality features from both the spectral and spatial domains.

Traditional machine learning methods, such as k-Nearest Neighbours (k-NN), Support Vector Machines (SVM), Random Forests (RF), and Gaussian Mixture Models (GMM), have been widely used due to their simplicity and interpretability [18]. However, these methods typically rely heavily on hand-crafted features and often treat spectral bands independently, which can limit their ability to capture spatial context and complex feature hierarchies. To overcome these limitations, sparse representation and manifold learning approaches such as PCA, Independent Component Analysis (ICA), and t-distributed Stochastic Neighbour Embedding (t-SNE) have also been employed for feature extraction and dimensionality reduction with varying success [19].

In the recent years, deep learning has emerged as a game-changing solution for hyperspectral image classification. Among the most widely used models are 2D CNN, which are highly effective at capturing spatial information. However, since 2D CNN primarily operate on spatial patches, they may overlook the crucial spectral dependencies between adjacent bands. On the other hand, 3D CNN are designed to process both spectral and spatial dimensions jointly by performing convolutions across all three dimensions of the HSI cube. While they achieve better spectral–spatial feature integration, their computational cost and memory requirements can be significant.

To strike a balance between performance and efficiency, researchers have proposed Hybrid 2D-3D CNNs, which use a combination of 2D convolutions (to capture spatial structure) and 3D convolutions (to integrate spectral–spatial features). This hybrid approach has proven particularly powerful in scenarios where the trade-off between computational cost and classification accuracy must be carefully managed.

4.2 Convolutional neural networks

CNN is the class of deep learning models. Unlike traditional neural networks, CNNs use the spatial structure in images using convolutional layers that apply filters (kernels) to local regions of the input data. Hyperspectral image classification, CNNs have become a cornerstone due to its capabilities to extract both spatial and spectral features with minimal preprocessing [20]. Fig. 5 shows CNN architecture.

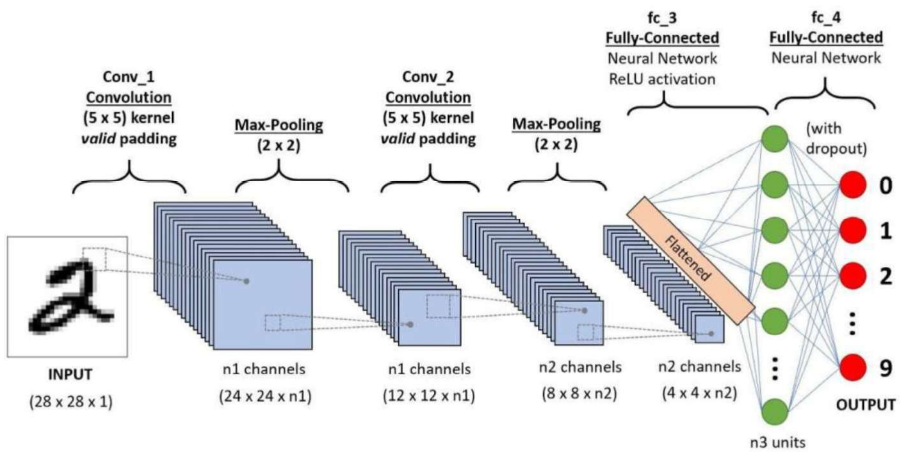

Fig. 5 Convolutional Neural Network Architecture

A CNN architecture comprises several key layers:

- **Convolutional Layers:** These layers apply a set of filters to the input data to extract features like edges, textures, or more complex patterns. For HSI, 2D CNNs apply these filters to spatial patches, while 3D CNNs extend this concept to capture spectral–spatial relationships by convolving across spectral bands and spatial dimensions simultaneously.
- **Activation Layers:** Non-linear activation functions such as ReLU (Rectified Linear Unit) introduce non-linearity into the model, enabling it to learn complex patterns. ReLU is commonly used in HSIC due to its simplicity and effectiveness.

- **Pooling Layers:** These layers reduce the spatial dimensions of the feature maps, typically by applying a maximum or average operation over local patches. This helps decrease computational load and provides a degree of spatial invariance.

- **Fully Connected (Dense) Layers:**Fully connected layers combine the extracted features to perform final classification, mapping them to the appropriate class labels.

- **Dropout and Batch Normalization:** Dropout layers randomly deactivate a portion of neurons during training to prevent overfitting, while batch normalization helps stabilize and accelerate the training by normalizing layer inputs.

For hyperspectral data, CNNs are particularly advantageous because they can be designed to exploit both spectral continuity and spatial correlations—essential for distinguishing between subtle material differences that may appear similar in RGB but vary in infrared or shortwave bands.

4.2.1 Workflow of CNN

Hyperspectral image classification is one of the challenging task due to the high dimensionality, redundant spectral information, and limited labelled data. CNNs have emerged as powerful tools in this domain because they can learn complex spatial and spectral patterns directly from the raw data. However, their performance can be greatly enhanced through thoughtful data preprocessing and training strategies.

- **EDA (Exploratory Data Analysis)** helps in understanding the structure and distribution of hyperspectral data, identifying class imbalance, noise, or missing values, and choosing appropriate preprocessing methods.
- **PCA (Principal Component Analysis)** is commonly used to reduce the number of spectral bands by retaining the most informative components, thus addressing the curse of dimensionality and reducing computational load without significantly compromising classification accuracy.
- **One-Hot Encoding** is essential when dealing with categorical labels (like land cover classes). It transforms these labels into a machine-readable format that enables the CNN to effectively learn multi-class classification.

- **Early Stopping** is a regularization technique used during training. It monitors the model's validation performance and halts training when performance stops improving, thus preventing overfitting—especially important when working with limited annotated hyperspectral data.
- **Pooling -** Pooling is a key operation in CNNs that reduces spatial dimensions while retaining only the important features. The most common types are max pooling (selecting maximum values) and average pooling (calculating mean values) from local regions. This down sampling helps decrease computational complexity and provides translation invariance to small shifts in input data. Pooling layers effectively summarize feature maps while maintaining the most salient information for subsequent network layers**.**

4.2.2 2D Convolutional Neural Networks

2D Convolutional Neural Networks (2D CNNs) have been widely used in image classification tasks due to their ability to effectively learn and extract spatial features from images. In the domain of HSIC, 2D CNNs have shown strong performance when spatial context plays a critical role in distinguishing between different land-cover or material classes.

A 2D CNN processes the data by applying 2D convolutional filters over spatial patches of the image. These filters scan across the height and width of the input data—capturing patterns like edges, textures, and spatial arrangements—without directly considering the spectral information across bands. In HSI applications, each pixel has hundreds of spectral values. When using 2D CNNs, it is common to treat each pixel's spectral vector as a stack of channels (similar to RGB channels in standard images), or to reduce spectral dimensions using techniques like the Principal Component Analysis (PCA) beforehand.

The basic building blocks of a 2D CNN include:

- **Convolutional Layers**: It apply a set of learnable filters to extract spatial features from the input.
- **Activation Functions**: Often ReLU (Rectified Linear Unit), introduce non-linearity after each convolution.

- **Pooling Layers**: Downsample feature maps, reducing computation and emphasizing the most important features (e.g., using max-pooling).
- **Fully Connected Layers**: Act as classifiers based on the learned features from earlier layers.
- **Dropout and Batch Normalization**: Often added to improve generalization and speed up training.

4.2.3 3D Convolutional Neural Networks

Hyperspectral images (HSI) consist of hundreds of narrow and the contiguous spectral bands, providing a rich source of information for analysing and identifying materials on the Earth's surface. Unlike conventional RGB images, HSIs contain both spatial and spectral information, and effective classification requires methods that can jointly analyze these dimensions. This is where 3D Convolutional Neural Networks (3D CNNs) become especially powerful.

Unlike 2D CNNs that only process spatial features by convolving over height and width, 3D CNNs extend the convolution operation into the spectral dimension, meaning the filters operate across height, width, and spectral depth simultaneously. This allows 3D CNNs to capture both spatial patterns and spectral signatures in a single architecture, making them particularly well-suited for hyperspectral data where spectral continuity holds key discriminative information.

The basic structure of a 3D CNN includes layers such as:

- **3D Convolutional Layers**: These use 3D kernels to scan the hyperspectral cube, extracting joint spatial–spectral features from small volumetric patches.
- **3D Pooling Layers**: These layers reduce dimensionality while preserving important features across all three dimensions, making the model more efficient.
- **Activation Functions (e.g. ReLU)**: Introduce non-linearity, enabling the network to learn complex feature relationships.

- **Fully Connected Layers**: These consolidate features for the final classification step.

4.3 Proposed algorithms

Furthermore, to push classification accuracy and generalization even further, several novel algorithmic enhancements have been developed. The following sections in this chapter will introduce three innovative frameworks tailored for HSIC.

4.3.1 SpectraEdgeNet *(Spatial filter + 3D CNN)*

SpectraEdgeNet: A Spatially-Filtered 3D Convolutional Neural Network for Enhanced Hyperspectral Image Classification.

SpectraEdgeNet is a custom-designed framework that integrates a spatial filtering stage before 3D convolution to improve the accuracy of hyperspectral image classification. The process begins by applying a 9×9 spatial filter to each spectral band of the input hyperspectral cube. This filter acts as an edge enhancer and noise suppressor, allowing the model to better highlight important spatial boundaries and structures within the scene.

Once the spatial filter is applied, a new enhanced hyperspectral cube is generated that retains the original spectral content but with improved spatial feature clarity. This pre-processed cube is then passed through a 3D CNN, which simultaneously learns spectral and spatial features from the refined input.

By emphasizing spatial edges and reducing local noise, the initial filtering stage helps the 3D CNN focus on more meaningful patterns, leading to better feature extraction. This layered design not only boosts classification accuracy, especially in complex and heterogeneous regions, but also improves the generalization capability of the model across varied datasets. Fig. 6 shows architecture of SpectraEdgeNet.

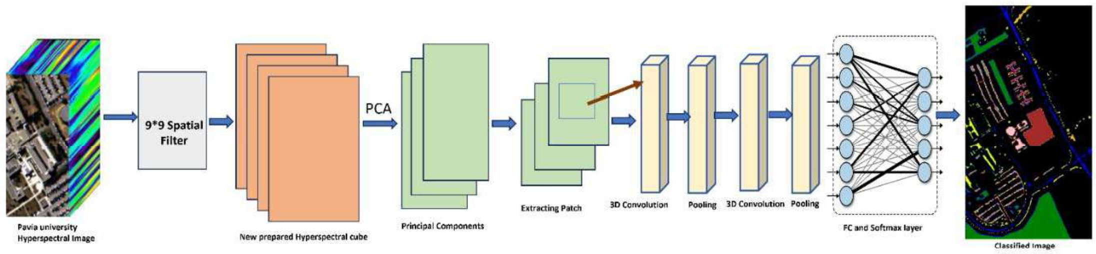

Fig. 6 Architecture of SpectraEdgeNet

4.3.2 SmoothSpectralNet *(Gaussian filter + 3D CNN)*

SmoothSpectralNet: A Gaussian-Enhanced 3D CNN Architecture for Robust Hyperspectral Classification

SmoothSpectralNet is a tailored deep learning model for hyperspectral image classification that introduces a Gaussian smoothing filter before applying 3D convolution. In this approach, the original hyperspectral cube is first passed through a Gaussian filter, which performs localized averaging across the spatial dimensions of each spectral band. This reduces fine-grain noise and suppresses minor pixel-level variations while preserving important global structures.

The main purpose of applying the Gaussian filter is to smooth the spatial content without significantly distorting the spectral information. Hyperspectral data often suffers from sensor noise and abrupt local fluctuations, especially in real-world field conditions. By applying Gaussian smoothing, the model ensures that the input to the 3D CNN is cleaner and more consistent.

Once the smoothed hyperspectral cube is generated, it is fed into a 3D CNN, which captures both spectral and spatial dependencies. The cleaner input helps the 3D CNN extract more stable and representative features, leading to improved classification performance and better robustness across different classes and noise conditions. Fig. 7 shows architecture of SmoothSpectralNet.

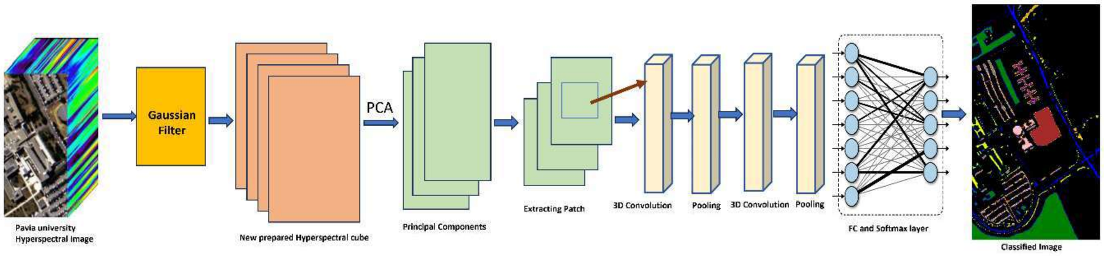

Fig. 7 Architecture of SmoothSpectralNet

4.3.3 HyperViTNet *(CNN + Transformer fusion)*

HyperViTNet: A CNN-Transformer Hybrid Framework for Spectral-Spatial Feature Learning in Hyperspectral Images is a deep learning architecture designed to effectively capture both local and global contextual information in HSIC.

The process begins by applying a Gaussian filter to the input hyperspectral cube to reduce noise and enhance local consistency. This preprocessed cube is then simultaneously fed into two parallel branches. The first branch employs a 3D CNN to jointly extract spectral and spatial features, leveraging its strength in modeling local neighborhood structures. The second branch consists of a Transformer block, which excels at capturing long-range dependencies and global contextual relationships within the data. The outputs from both the CNN and Transformer branches are then concatenated to form a unified feature representation. Finally, these fused features are passed through a classifier that generates the final classification map, offering improved accuracy and robustness in HSI. Fig. 8 shows block diagram of HyperViTNet.

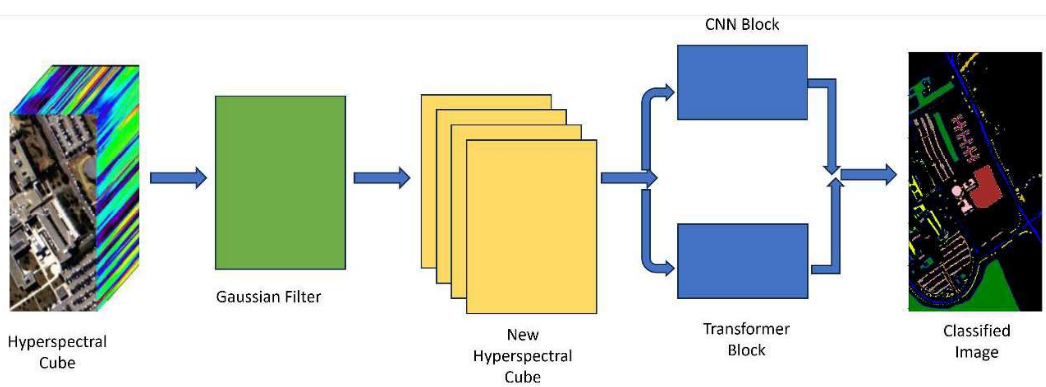

Fig. 8 Block diagram of HyperViTNet

Chapter 5

Dataset

5.1 Description of Pavia university dataset

The Pavia University (PaviaU) dataset is one of the most commonly used benchmark datasets in hyperspectral remote sensing. It was captured by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor during a flight over the University of Pavia in northern Italy. This urban area features a mix of man-made structures and natural elements, making it suitable for testing and evaluating classification algorithms that require both spatial and spectral variability.

The dataset consists of a hyperspectral image of size 610 × 340 pixels, with each pixel containing a spectrum recorded across 103 spectral bands in the visible and near-infrared range (after removing noisy and water absorption bands from the original 115). Each band corresponds to a specific wavelength, capturing detailed information that allows fine discrimination between materials. Table 1 shows distribution of Pavia dataset into 9 classes.

Class Number	**Class Name**	**Samples**
1	Asphalt	6631
2	Meadows	18,649
3	Gravel	2099
4	Trees	3064
5	Painted Metal Sheets	1345
6	Bare Soil	5029
7	Bitumen	1330
8	Self-Blocking Bricks	3682
9	Shadows	947

Table 1: Details of Pavia university dataset.

5.2 Description of Indian Pines datatset

The Indian Pines dataset is a classic benchmark in hyperspectral remote sensing, originally collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over agricultural land in Northwestern Indiana, USA. It is one of the earliest and most widely studied datasets in the field and is especially valuable for testing hyperspectral image classification and unmixing algorithms.

The dataset contains a hyperspectral image of size 145 × 145 pixels, with each pixel comprising a spectrum recorded across 220 spectral bands spanning the 400 to 2500 nm wavelength range (covering the visible to short-wave infrared spectrum). However, after removing bands corrupted by water absorption and noise, 200 spectral bands are typically retained for processing.

Indian Pines primarily represents agricultural and forested areas, featuring a high degree of spectral similarity between different crop types. Table 2 shows the distribution of Indian Pines dataset into 16 classes.

Class Number	**Class Name**	**Samples**
1	Alfalfa	46
2	Corn notill	1428
3	Corn mintill	830
4	Corn	237
5	Grass pasture	483
6	Grass trees	730
7	Grass pasture mowed	28
8	Hay windrowed	478

9	Oats	20
10	Soybean notill	972
11	Soybean mintill	2455
12	Soybean clean	593
13	Wheat	205
14	Woods	1265
15	Buildings Grass Trees Drives	386
16	Stone Steel Towers	93

Table 2 Details of India Pines dataset.

Chapter 6

Results and Discussion

6.1 Hyperspectral Image Classification of Pavia university dataset

As discussed in the previous chapter the dataset is divided into nine different classes and unlabeled class is considered as class zero. The distribution of classes of Pavia University dataset is as follows. Fig. 9 shows Pavia university dataset classes distribution.

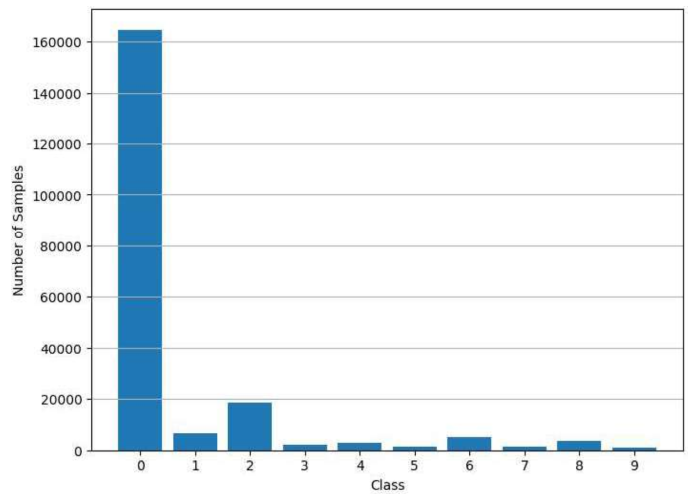

Fig. 9: Pavia University dataset classes distribution.

The Pavia hyperspectral dataset has served as a benchmark for evaluating the performance of various deep learning-based classification algorithms, particularly in the context of incorporating spatial-spectral information. Initially, a standard 2D CNN was employed, which processes only the spatial dimensions of the hyperspectral image while ignoring its rich spectral content. This baseline method achieved an average classification accuracy of **0.8972 ± 0.0492** and a corresponding kappa coefficient of **0.91**, indicating moderately strong classification performance. However, when a 3D CNN architecture was applied—capable of capturing both spectral and spatial dependencies simultaneously—the performance improved significantly. The 3D CNN model achieved a higher accuracy of **0.9459 ± 0.0388**, with a kappa coefficient of **0.937**, demonstrating the importance of spectral information in enhancing classification accuracy.

Building upon these foundations, three novel algorithms were proposed, each introducing unique enhancements to further exploit the spatial-spectral structure of hyperspectral images. The first among them, SpectraEdgeNet, integrates a spatial edge-aware filtering mechanism with a 3D CNN using a 9×9spatial window. This method effectively emphasizes edge information while preserving the spectral characteristics, resulting in a notable improvement in performance with an accuracy of **0.9535 ± 0.0412** and a kappa coefficient of **0.9411**. The second proposed method, SmoothSpectralNet, incorporates a Gaussian smoothing filter before applying the 3D CNN. This preprocessing step helps reduce spectral noise and enhances the continuity of spectral signatures, leading to an even better accuracy of **0.97615 ± 0.0319** and a kappa coefficient of **0.953**, reflecting superior consistency in classification.

The most advanced of the proposed algorithms, HyperViTNet, introduces a hybrid architecture that combines CNN-based feature extraction with Transformer-based attention mechanisms. This fusion leverages the local feature learning ability of CNNs with the global contextual modeling capacity of Transformers, allowing the network to capture both fine-grained and long-range dependencies in hyperspectral data. As a result, HyperViTNet achieved the highest performance among all tested models, with an accuracy of **0.9897 ± 0.0349** and a kappa coefficient of **0.983**. These results underline the potential of Transformer-based models and hybrid approaches in pushing the boundaries of hyperspectral image classification. Fig. 10 shows ground truth and predicted labels for Pavia dataset.

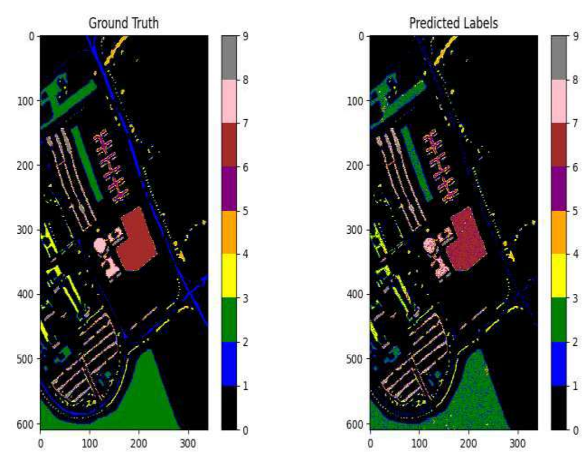

Fig. 10 Ground truth and one of the predicted labels of Pavia dataset.

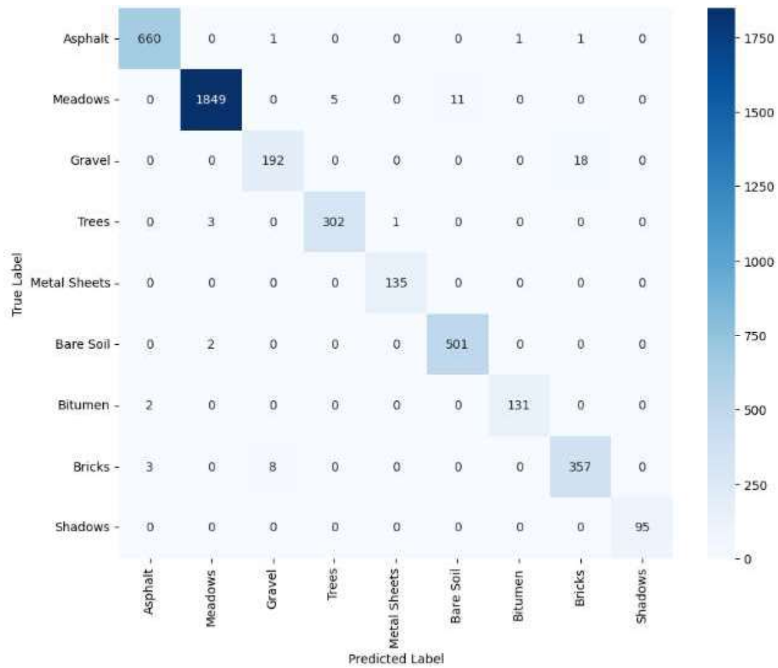

Fig. 11 Confusion Matrix for 2D CNN

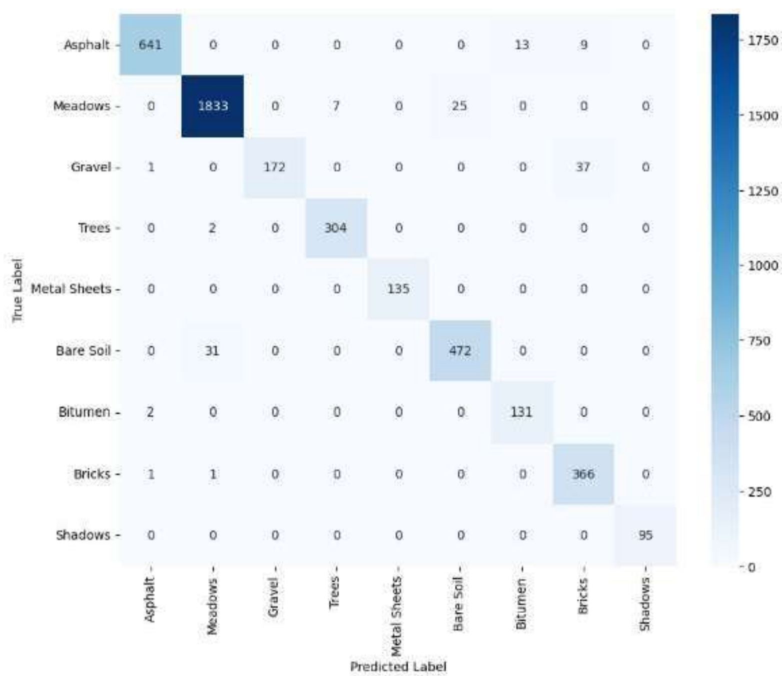

Fig. 12 Confusion Matrix for 3D CNN

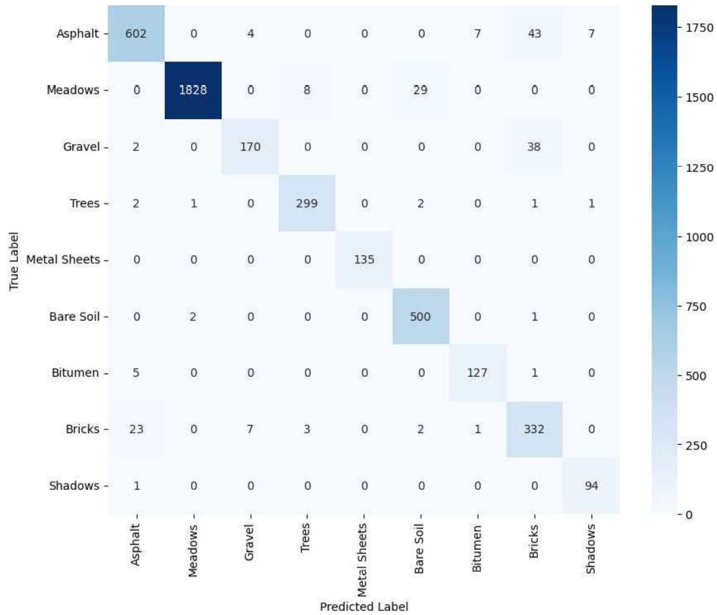

Fig. 13 Confusion Matrix for SpectraEdgeNet

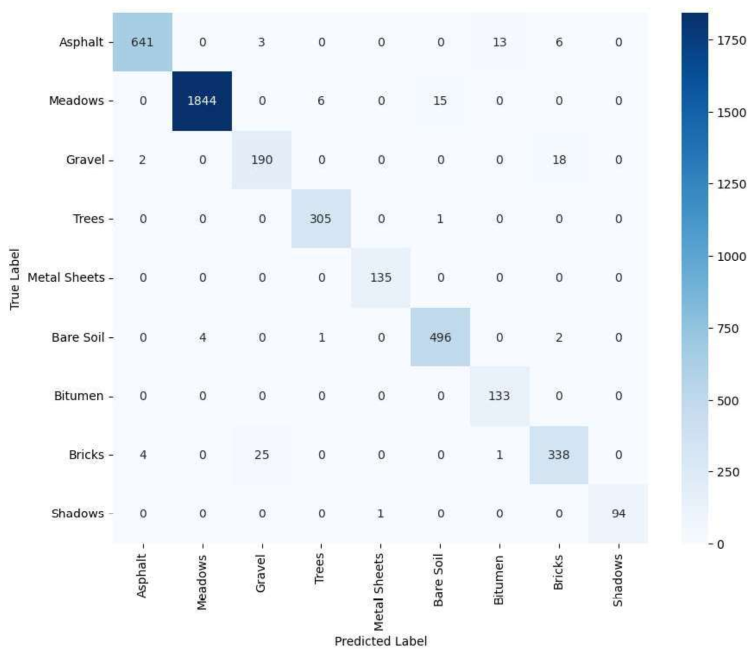

Fig. 14 Confusion Matrix for SmoothSpectralNet

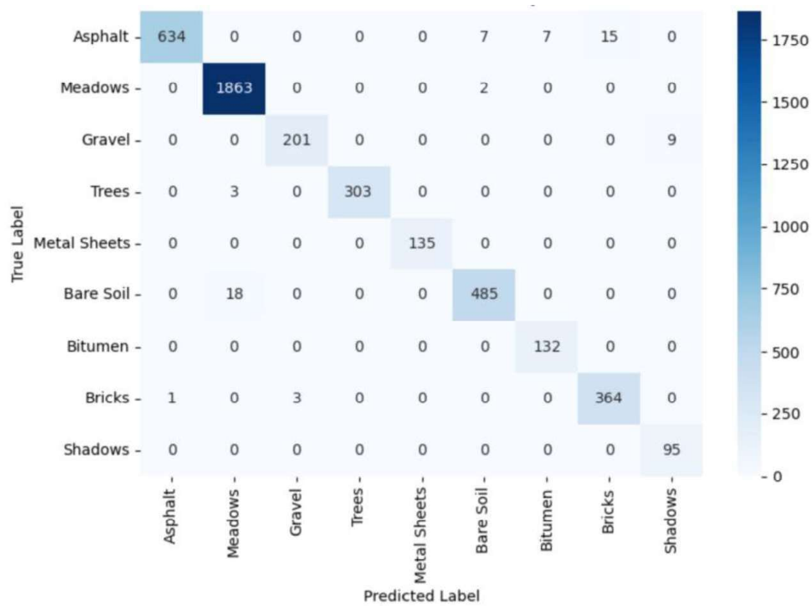

Fig. 15 Confusion Matrix for HyperViTNet

Model	Accuracy and standard deviation	Kappa value
2D CNN	0.8972±0.0492	0.9121
3D CNN	0.9459±0.0388	0.9376
SpectraEdgeNet	0.9535 ± 0.0412	0.9411
SmoothSpectralNet	0.97615 ± 0.0319	0.9536
HyperViTNet	**0.9897 ± 0.0349**	**0.983**

Table 3: Accuracy and Kappa coefficient comparison of different models for Pavia dataset.

6.2 Hyperspectral Image Classification of Indian Pines dataset.

As discussed in the previous chapter the dataset is divided into 17 different classes and unlabeled class is considered as class zero. The distribution of classes of Indian Pines dataset is as follows. Fig. 16 shows classes distribution of Indian Pines dataset.

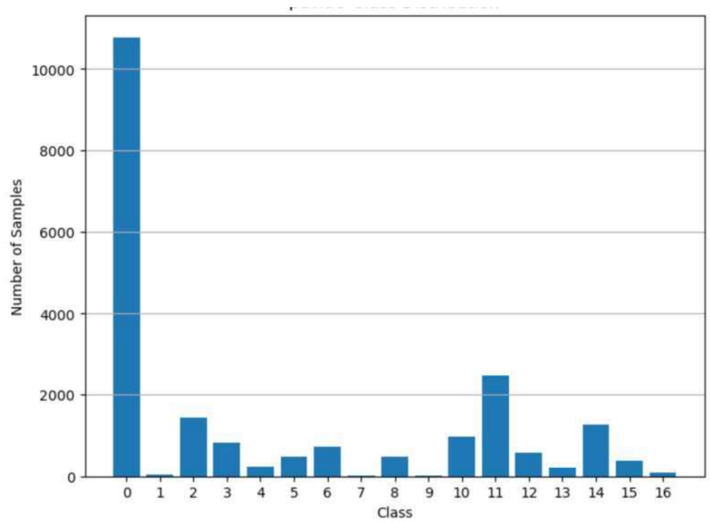

Fig. 16 Indian Pines dataset Classes distribution.

The Indian Pines hyperspectral dataset has long been recognized as a challenging and informative benchmark for testing hyperspectral image classification models due to its high spectral dimensionality and the presence of spectrally similar classes. To establish a baseline, a 2D Convolutional Neural Network (CNN) was first applied. This model, which leverages only the spatial information while disregarding the spectral richness of the data, achieved an average classification accuracy of **0.8767 ± 0.0356** and a kappa coefficient of **0.9215**. While this result reflects a reasonably good performance, it leaves significant room for improvement through spectral domain exploitation. Accordingly, a 3D CNN architecture was utilized to simultaneously capture both spatial and spectral correlations inherent in hyperspectral data. This model delivered improved results with an accuracy of **0.9184 ± 0.0491** and a kappa coefficient of **0.9319**, demonstrating the advantage of incorporating the spectral dimension in the learning process.

Building on these baseline approaches three advanced models were proposed to further enhance classification performance by optimizing the extraction and integration of spatial-spectral features. The first, SpectraEdgeNet, combines a spatial edge-aware filtering technique with a 3D CNN using a 9×99 \times 99×9 spatial kernel. This method enhances edge preservation and spatial feature clarity, resulting in a classification accuracy of **0.9401 ± 0.0493** and a kappa value of **0.9396**, indicating improved boundary delineation and inter-class discrimination. The second proposed model, SmoothSpectralNet, introduces a Gaussian smoothing filter as a preprocessing step prior to the 3D CNN. This filter reduces spectral variability and noise while maintaining the underlying spectral structure, leading to a further boost in performance with an accuracy of **0.9606 ± 0.0387** and a kappa coefficient of **0.9517**, showcasing improved spectral continuity and model robustness.

The final and most advanced of the proposed models, HyperViTNet, fuses CNN with Transformer-based attention mechanisms to capture both local and global features. By combining the spatial feature extraction strengths of CNNs with the contextual reasoning power of Transformers, HyperViTNet achieves an enhanced understanding of spatial-spectral dependencies across the entire image. This hybrid architecture yielded the highest performance among all methods tested on the Indian Pines dataset, with a remarkable classification accuracy of **0.9792 ± 0.0359** and a kappa coefficient of **0.9791**. These results confirm the efficacy of integrating attention-based mechanisms with traditional deep learning techniques in achieving state-of-the-art performance in hyperspectral image classification. Fig. 17 shows ground truth and predicted labels for Indian Pines dataset.

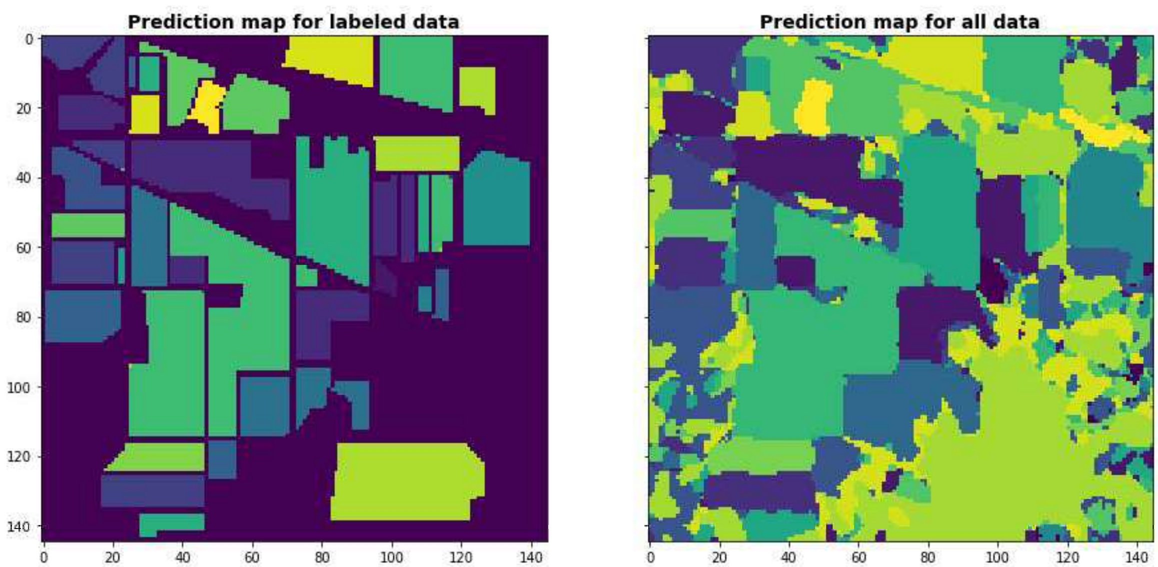

Fig. 17 Ground truth and one of the predicted labels of Indian Pines dataset.

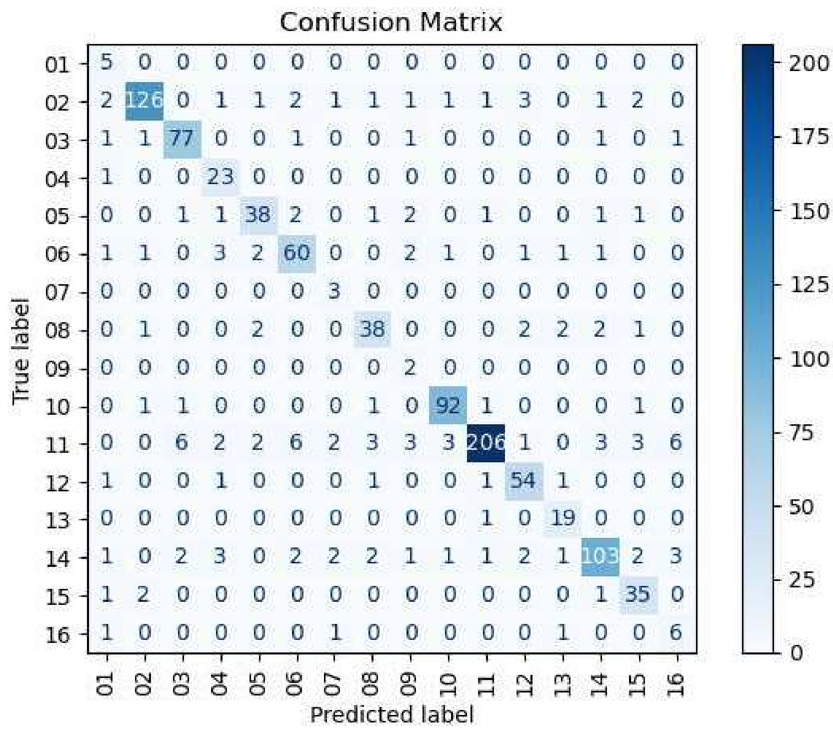

Fig. 18 Confusion Matrix for 2D CNN

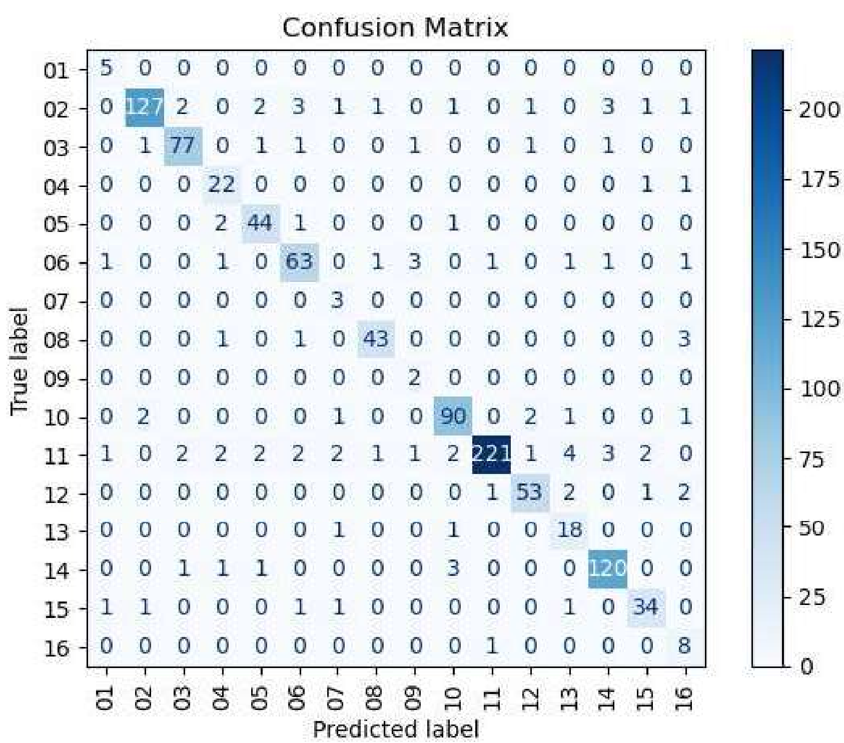

Fig. 19 Confusion Matrix for 3D CNN

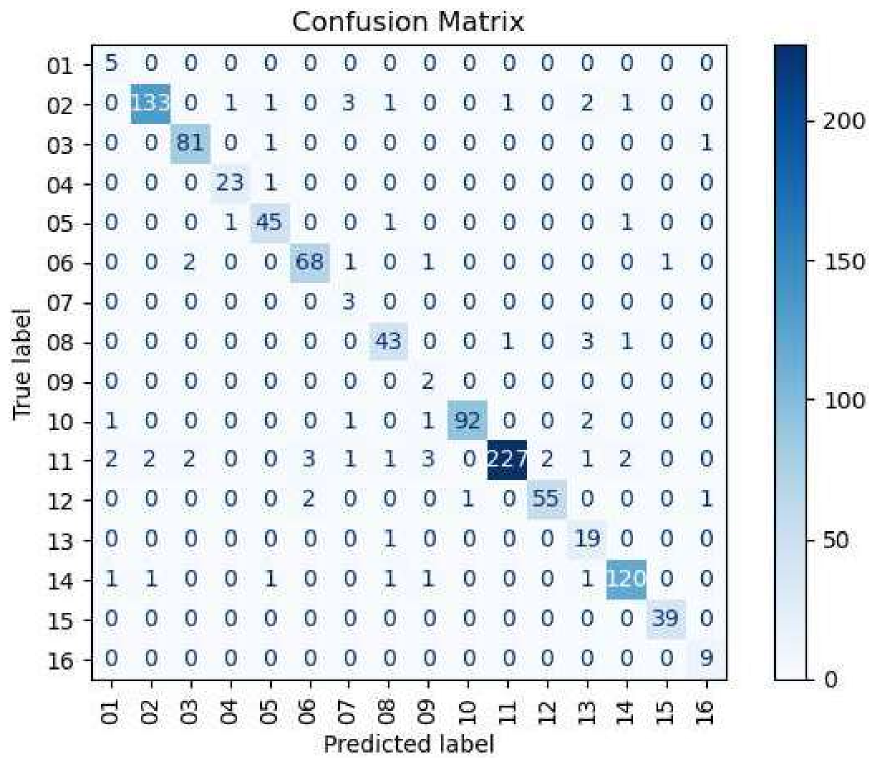

Fig. 20 Confusion Matrix for SpectraEdgeNet

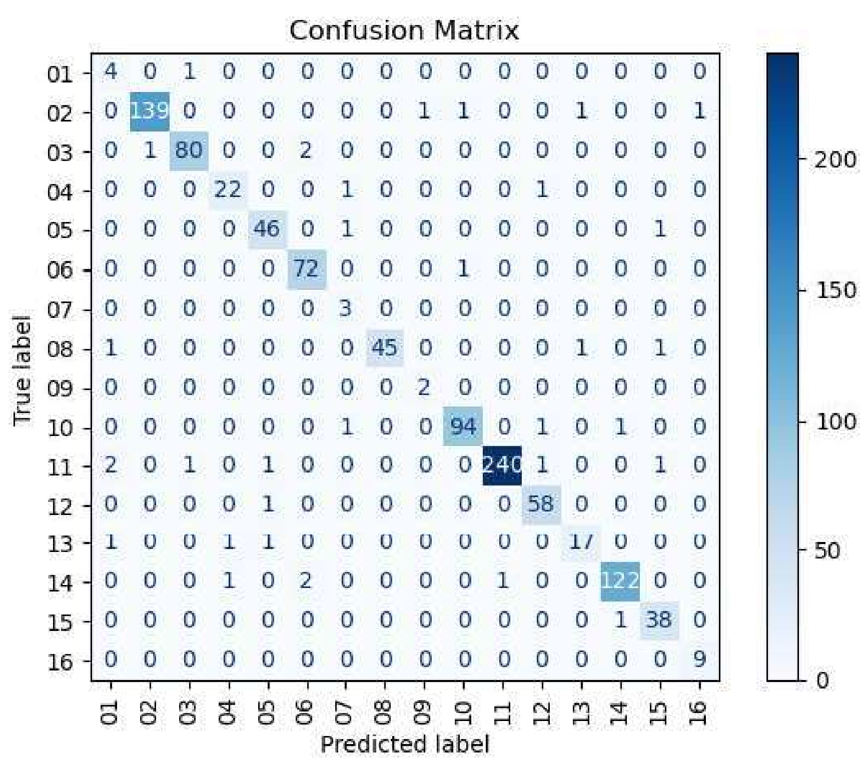

Fig. 21 Confusion Matrix for SmoothSpectralNet

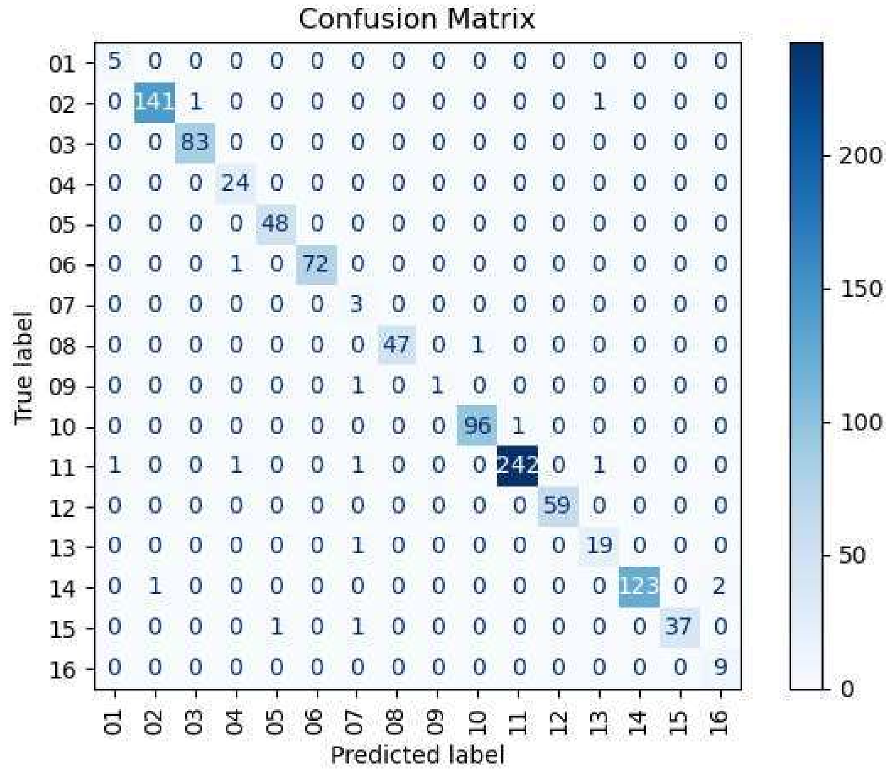

Fig. 22 Confusion Matrix for HyperViTNet

Model	Accuracy and standard deviation	Kappa value
2D CNN	0.8767 ± 0.0356	0.9215
3D CNN	0.9184 ± 0.0491	0.9319
SpectraEdgeNet	0.9401 ± 0.0493	0.9396
SmoothSpectralNet	0.9606 ± 0.0387	0.9517
HyperViTNet	**0.9792 ± 0.0359**	**0.9791**

Table 4: Accuracy and Kappa comparison of different models for Indian Pines dataset.

Chapter 7

Conclusion and Future Work

The experimental results clearly show that the proposed models—SpectraEdgeNet, SmoothSpectralNet, and HyperViTNet—consistently outperform conventional 2D and 3D CNNs in terms of both classification accuracy and kappa scores across standard hyperspectral datasets like Pavia and Indian Pines. These improvements underline the benefits of combining spatial filtering, spectral smoothing, and attention-based feature fusion for better spatial-spectral representation learning. While the current models achieve state-of-the-art performance, there is still scope for further enhancement. Future work will focus on improving model accuracy even further while also reducing computational complexity to make the models faster and more efficient. This is particularly important for practical deployment in real-time or resource-constrained environments. Moreover, we intend to extend the evaluation to additional and more diverse hyperspectral datasets, which will help assess the generalization capability of the proposed methods and fine-tune them for broader remote sensing applications.

REFERENCES

[1] Penna, B., Tillo, T., Magli, E., & Olmo, G. (2007). Transform coding techniques for lossy hyperspectral data compression. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(5), 1408–1421. https://doi.org/10.1109/tgrs.2007.894565

[2] Iordache, M., Bioucas-Dias, J. M., & Plaza, A. (2012). Total variation spatial regularization for sparse hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(11), 4484–4502. https://doi.org/10.1109/tgrs.2012.2191590

[3] Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Martinez-Gonzalez, P., & Garcia-Rodriguez, J. (2018). A survey on deep learning techniques for image and video semantic segmentation. *Applied Soft Computing*, *70*, 41–65. https://doi.org/10.1016/j.asoc.2018.05.018

[4] Qian, S. (2021). Hyperspectral satellites, evolution, and development history. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 7032–7056. https://doi.org/10.1109/jstars.2021.3090256

[5] Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., & Steininger, M. (2003). Remote sensing for biodiversity science and conservation. *Trends in Ecology & Evolution*, *18*(6), 306–314. https://doi.org/10.1016/s0169-5347(03)00070-3

[6] Barth, R. (2018). *Vision principles for harvest robotics : sowing artificial intelligence in agriculture*. https://doi.org/10.18174/456019

[7] Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., & Chanussot, J. (2013). Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine*, *1*(2), 6–36. https://doi.org/10.1109/mgrs.2013.2244672

[8] Manikandan, G. E. D. M. (2022). A Special issue on Multidisciplinary research to promote the applications in Biological sciences. *International Journal of Life Science and Pharma Research*, 1–154. https://doi.org/10.22376/ijpbs/ijlpr/sp22/2022.1-154

[9] Ghamisi, P., Yokoya, N., Li, J., Liao, W., Liu, S., Plaza, J., Rasti, B., & Plaza, A. (2017). Advances in Hyperspectral Image and Signal Processing: A Comprehensive overview of the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), 37–78. https://doi.org/10.1109/mgrs.2017.2762087

[10] Ong, C., Carrère, V., Chabrillat, S., Clark, R., Hoefen, T., Kokaly, R., Marion, R., Filho, C. R. S., Swayze, G., & Thompson, D. R. (2019). Imaging spectroscopy for the detection, assessment and monitoring of natural and anthropogenic hazards. *Surveys in Geophysics*, *40*(3), 431–470. https://doi.org/10.1007/s10712-019-09523-1

[11] Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, *28*(5), 823–870. https://doi.org/10.1080/01431160600746456

[12] Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Ali, A. B. M. S., & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artificial Intelligence Review*, *56*(11), 13521–13617. https://doi.org/10.1007/s10462-023-10466-8

[13] Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, *15*(2), 265–286. https://doi.org/10.1198/106186006x113430

[14] Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., & Glocker, B. (2016). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis*, *36*, 61–78. https://doi.org/10.1016/j.media.2016.10.004

[15] Elahi, M., Afolaranmi, S. O., Lastra, J. L. M., & Garcia, J. a. P. (2023). A comprehensive literature review of the applications of AI techniques through the lifecycle of industrial equipment. *Discover Artificial Intelligence*, *3*(1). https://doi.org/10.1007/s44163-023-00089-x

[16] Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K. A., Elkhatib, Y., Hussain, A., & Al-Fuqaha, A. (2019). Unsupervised machine learning for networking: techniques, applications and research challenges. *IEEE Access*, *7*, 65579–65615. https://doi.org/10.1109/access.2019.2916648

[17] Xu, Z., Jiang, W., & Geng, J. (2023). Dual-Branch dynamic modulation network for hyperspectral and LiDAR data classification. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–13. https://doi.org/10.1109/tgrs.2023.3285097

[18] Mohammadi, M. (2022). *Novel techniques of computational intelligence for analysis of astronomical structures*. https://doi.org/10.33612/diss.196895383

[19] Anowar, F., Sadaoui, S., & Selim, B. (2021). Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). *Computer Science Review*, *40*, 100378. https://doi.org/10.1016/j.cosrev.2021.100378

[20] Shin, H., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, *35*(5), 1285–1298. https://doi.org/10.1109/tmi.2016.2528162