Spectral–Spatial Land Cover Mapping Using a Deep 3D Convolutional Neural Network on Hyperspectral Data

Assia Nouna[1,*], Soumaya Nouna[1], Ilyas Tammouch[2], Amal Badoui[3]
[1]Hassan First University of Settat, ENSA Berrechid, Laboratory LAMSAD, Morocco
[2]National School of Mines of Rabat (ENSMR), Laboratory of Systems Engineering and Digital Transformation (LISTD), Morocco
[3] Faculty of Science and Technology of Mohamedia Hassan II University of Casablanca , Morocco
E-mail: a.nouna@uhp.ac.ma
*Corresponding Author

Keywords: Hyperspectral images, remote sensing, convolutional neural network 3D, k-nearest neighbors, land cover mapping, classification accuracy, deep learning, 3D structure architecture

Received: September 4, 2025

*Recent studies have demonstrated that incorporating both spectral and spatial information significantly enhances the accuracy of hyperspectral image (HSI) classification. Since HSI data is commonly structured as 3D cubes, where two dimensions represent spatial information and the third encodes spectral bands, applying 3D spatial filtering becomes an intuitive and effective way to extract spectral–spatial features simultaneously. In this work, we propose a deep 3D Convolutional Neural Network (3D–CNN) architecture for accurate and efficient HSI classification. Unlike many previous methods, the proposed model processes the hyperspectral cube directly as a volumetric input, eliminating the need for manual feature extraction or dimensionality reduction. This design allows the network to jointly learn spatial and spectral dependencies while maintaining a compact architecture with fewer trainable parameters. Experimental evaluations on two benchmark datasets—Pavia University and Salinas Valley—demonstrate that the proposed model achieves **95% overall classification accuracy**, compared to **88%** obtained with the conventional K-Nearest Neighbors (KNN) baseline. The 3D–CNN consistently outperforms traditional methods across most land-cover classes, particularly in spectrally similar or complex regions, confirming its effectiveness for spectral–spatial land cover mapping. These results highlight the practicality and robustness of deep volumetric learning for real-world hyperspectral image analysis*

Povzetek: Študija predstavi kompaktno 3D-CNN za klasifikacijo hiperspektralnih slik, ki neposredno obdela 3D podatkovno kocko in s tem sočasno uči spektralno-prostorske odvisnosti brez ročnega izbiranja značilk ali zmanjšanja dimenzionalnosti

1 Introduction

The process of hyperspectral remote sensing captures digital images through a series of hundreds of adjacent and closely spaced spectral bands[1, 2], spanning a range of wavelengths from the visible to the infrared spectrum. This method produces a three-dimensional hyperspectral image (HSI)[3], in which every pixel stores a precise spectral information signature in parallel with spatial information. The physical and chemical properties of materials on the Earth's surface can be comprehensively analysed from the resulting data cube[4]. These include precision agriculture to monitor crops and detect disease, environmental science to classify land cover and assess pollution, and homeland security to identify biological and chemical threats[5]. Accurate classification of each pixel, which directly affects the effectiveness of these applications, is a fundamental step in the use of HSI data [6, 7]. As a result, over the past few decades, numerous advanced HSI classification methods have been proposed by the research community [8, 9], evolving from traditional machine learning algorithms to modern deep neural network–based frameworks [10].
Conventional hyperspectral image (HSI) classification techniques have traditionally focused exclusively on spectral information[11]. Commonly used classifiers, such as distance-based approaches [12], k-nearest neighbors (KNN), maximum likelihood estimation, and logistic regression, often demonstrate limited effectiveness in high-dimensional HSI data[13]. A primary limitation stems from the well-known "small-sample problem," wherein the number of available labeled training samples is insufficient relative to the large number of spectral bands. This problem, known as the Hughes effect, degrades classification performance because the training data is not proportionally larger[14]. In addition, because many spectral bands are highly correlated and do not contribute unique information, spectral redundancy is often observed in HSI datasets [15]. Recently, among the various deep learning models used

for hyperspectral image (HSI) analysis, 2D convolutional neural networks (2D CNNs)[16] have been widely adopted due to their ability to effectively exploit spatial structures. Unlike traditional machine learning approaches, which use hand-crafted features, 2D CNN [17] learns hierarchical spatial features directly from the input data, and is therefore well suited to problems involving image texture and object borders. In the context of HSI [18], analogous to the RGB channels in natural images, each spectral band can be treated as a separate image channel. 2D CNNs are able to capture spatial continuity and local patterns in a scene by treating each spectral band as an independent spatial input [19]. While this approach improves classification performance through the use of spatial context, the spectral relationships between the bands cannot be fully exploited because the spectral channels are usually treated independently [20].

3D Convolutional Neural Networks (3D CNNs)[21] have been proposed as a more comprehensive solution for HSI classification, in order to overcome the limitations of 2D CNNs. In contrast to 2D CNNs, which process each spectral band separately [22], 3D CNNs apply convolutional operations in both the spatial and spectral dimensions at the same time. This permits the model to capture deep spectral-spatial correlations, thereby improving discrimination between classes with similar spatial features but different spectral signatures-or vice versa [23]. 3D CNNs exploit the full structure of the data, reducing spectral redundancy and preserving contextual relationships across bands by treating the hyperspectral cube as a volumetric input [24]. Consequently, they generally achieve higher classification accuracy, especially in complex scenes where subtle spectral or spatial cues embed material differences. Despite their higher computational cost, 3D CNNs provide a richer and more holistic representation of HSI data than 2D CNNs [25].

This paper demonstrates how a 3-D CNN [26] can effectively classify hyperspectral data if it is properly architected. Our experience has shown that shallower CNN models, such as LeNet-5 with only two convolution layers, are not well suited for this task. The complex spectral-spatial structures of hyperspectral images are difficult to capture [27]. To address this, we have developed a more advanced 3D CNN architecture that incorporates batch normalisation, dropout layers and fully connected layers - making it a more efficient and reliable approach to hyperspectral image classification [28]. Previous studies such as Hamida et al. (2018) and Roy et al. (2020) have already demonstrated the effectiveness of 3D convolutional neural networks for hyperspectral image classification. Building upon these foundations, the present work introduces an enhanced hierarchical 3D–CNN architecture that incorporates progressive filter compression, interleaved normalization and dropout layers, and a patch-based volumetric learning strategy to improve both classification accuracy and computational efficiency.

Research objective and hypothesis: The central objective of this study is to evaluate whether a deeply structured 3D Convolutional Neural Network can effectively learn joint spectral–spatial representations from hyperspectral data without the need for dimensionality reduction or handcrafted features. The research is guided by the following hypothesis: *a hierarchical 3D–CNN architecture with descending filter sizes and interleaved normalization–dropout layers can achieve higher classification accuracy and better generalization than traditional machine learning and shallower deep learning models*. To validate this hypothesis, the proposed model is implemented and tested on two benchmark datasets—Pavia University and Salinas Valley—representing distinct urban and agricultural environments. The structure of this work is as follows: Section 2 gives a brief overview of 3D Convolutional Neural Networks (3D CNNs) [29] and explains how they work. Section 3 presents our experiments, where we compare our 3D CNN [30] model with K-Nearest Neighbours (KNN) and Deep Neural Networks (DNN) using well-known hyperspectral datasets. Finally, a summary of our main conclusions is presented in Section 4.

2 Description of methodology

This study introduces a deep three–dimensional Convolutional Neural Network (3D–CNN) for hyperspectral image (HSI) classification that jointly exploits spatial and spectral information. The methodological workflow comprises two key stages: (1) a brief overview of 3D–CNN fundamentals and their suitability for processing volumetric spectral data and an explanation of their application to HSI classification, emphasizing spectral–spatial feature learning; (2) a detailed description of the proposed 3D–CNN architecture, including layer configuration, activation functions, and training settings. The model demonstrates superior classification accuracy and generalization compared with conventional approaches, confirming its relevance for advanced remote-sensing applications.

2.1 3D convolutional neural network for HSI classification

Three–dimensional convolutional neural networks (3D–CNNs) are highly effective for hyperspectral image (HSI) classification because they process volumetric data and jointly capture spatial and spectral dependencies. The input to the network is the hyperspectral cube itself, composed of two spatial dimensions (height and width) and one spectral dimension (wavelength bands). Each pixel is thus represented by a detailed spectral signature that provides rich information about material composition across hundreds of contiguous wavelengths. By exploiting this full spectral–spatial context, the 3D–CNN can distinguish materials that appear similar in RGB or 2D projections but differ in their spectral profiles. Within the proposed ar-

chitecture, successive 3D convolutional layers apply volumetric filters that move across spatial and spectral axes, enabling the extraction of complex correlations between neighbouring pixels and adjacent wavelengths. Each convolution is followed by a Rectified Linear Unit (ReLU) activation, which introduces non-linearity by zeroing negative activations and preserving positive ones, thereby accelerating convergence and improving feature selectivity. Pooling layers are then used to down-sample the feature maps, retaining the most salient information while reducing computational cost and enhancing translational invariance. This combination of convolution, activation, and pooling allows the network to progressively learn hierarchical spectral–spatial representations—from fine spectral details in early layers to broader contextual features in deeper layers. After feature extraction, the resulting volumetric representations are flattened and passed through fully connected (dense) layers that integrate global contextual information. Dropout regularization between dense layers mitigates overfitting given the limited availability of labeled hyperspectral samples. Finally, a softmax classifier outputs normalized class probabilities, assigning each pixel to its most likely land-cover category. This end-to-end volumetric learning process enables accurate and robust hyperspectral image classification without the need for handcrafted features or prior dimensionality reduction.

Limitations of existing state-of-the-art methods: Although several state-of-the-art networks—including 2D CNNs, 3D CNN variants, and hybrid 2D–3D architectures such as HybridSN—have improved hyperspectral image classification, they still face notable challenges. First, many models suffer from *overfitting* due to the limited number of labeled samples available in hyperspectral datasets, especially when networks contain millions of parameters. Second, existing methods often exhibit a *spectral–spatial imbalance*: while deep layers capture rich spectral cues, they frequently lose fine spatial context through aggressive pooling or dimensionality reduction. Third, the *generalization capability* of these models is limited across datasets with differing illumination, noise levels, and land-cover distributions. Finally, most architectures employ uniform convolutional configurations that fail to adaptively balance spectral resolution and computational efficiency. These issues collectively motivate the development of the proposed hierarchical 3D-CNN architecture, which integrates progressive filter compression, interleaved normalization and dropout, and patch-based volumetric learning to achieve stable, generalized, and computationally efficient performance.

2.2 Proposed architecture for hyperspectral image classification using 3D CNN

We are proud to present a new 3D CNN architecture that aims to enhance the efficiency of feature extraction and classification. By means of a sequence of convolutional layers and max-pooling operations, our model captures both spectral and spatial characteristics by implementing 3D convolutions across the height, width, and spectral depth of the data. Better performance is aimed for through our method, with key patterns being effectively captured and improved at each step, and the most important data being made sure is preserved throughout the network.The procedure adopted for HSI classification in the proposed work can be illustrated by using Figure1.

Unlike most existing 3D CNN frameworks that primarily stack convolutional and pooling layers in a uniform fashion, our architecture introduces a multi-depth hierarchical block structure designed specifically to balance spectral-spatial representation with computational efficiency. Each convolutional block in the network is configured with progressively decreasing filter sizes ($128 \rightarrow 64 \rightarrow 32$) and adaptive receptive fields, allowing early layers to focus on fine spectral variations and deeper layers to capture broader spatial–contextual dependencies. This hierarchical compression improves feature diversity while controlling parameter growth.
Moreover, batch normalization and dropout are not merely appended for regularization, but strategically interleaved after each pooling stage to stabilise gradient flow across spectral depth and mitigate overfitting due to limited HSI training samples — a common bottleneck in remote sensing. Another distinctive aspect lies in the patch-based volumetric training strategy, where overlapping 3D cubes extracted from the hyperspectral scene are fed into the model. This enables local–global context integration without resorting to spectral dimensionality reduction or handcrafted preprocessing. Consequently, the model achieves a favorable trade-off between depth, generalization, and computational cost, effectively addressing the Hughes phenomenon often encountered in high-dimensional HSI data.

The proposed 3D Convolutional Neural Network (CNN) begins by normalizing the input hyperspectral image (HSI) data using batch normalization, which ensures standardization and enhances convergence during training. The input data is first reshaped into 3D patches and split into training and testing sets before being passed to the 3D CNN for classification. The model has been designed around three main sets of three 3D convolutional layers (C_1, C_2, C_3)Each of these sets contains three layers, thus yielding a total of ten 3D convolutional layers $(c_1, c_2.....c_{10})$ In addition, three 3D maximum pooling layers (P_1, P_2, P_3). have been interleaved between the convolutional layers. Based on empirical observation, it has been determined that each block is configured with a distinct number of filters. For each group, the filters are configured as $K_1 = 128$, $K_2 = 64$, and $K_3 = 32$ respectively.

Rationale for descending filter sizes The choice of progressively decreasing filter sizes ($128 \rightarrow 64 \rightarrow 32$) was guided by both empirical observations and theoretical con-

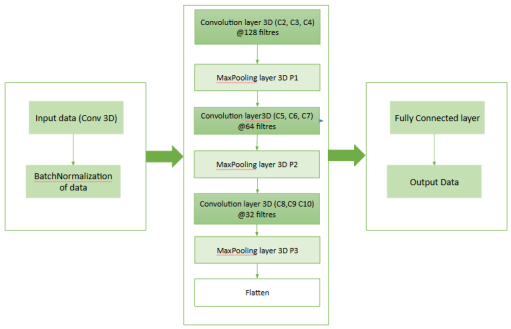

Figure 1: Proposed 3D–CNN architecture for hyperspectral image classification

siderations of feature hierarchy in hyperspectral data. In the initial convolutional blocks, a higher number of filters ($K_1 = 128$) enables the extraction of fine-grained spectral features and local textural variations across hundreds of contiguous bands. As the network deepens, subsequent layers ($K_2 = 64, K_3 = 32$) operate on increasingly abstract and spatially aggregated representations, thus requiring fewer filters to capture high-level discriminative cues. This descending configuration effectively balances representational richness and computational cost by reducing the number of trainable parameters in deeper layers while preserving classification performance. Alternative configurations—including uniform or ascending filter sizes—were experimentally evaluated during early model prototyping; however, they exhibited either slower convergence or marginally lower accuracy, confirming that the descending hierarchy provided the most stable and efficient learning behaviour for the datasets considered.

Fassuming the input shape is (img-height, img-width, img-depth, 1), the first convolutional block transforms the data to (img-height, img-width, img-depth, 128). Then, the first 3D max-pooling layer reduces the spatial and spectral dimensions according to the chosen pooling strategy. This sequence is then repeated for each subsequent two blocks. Subsequent to passing through C_3 and P_3, the feature maps are flattened and fed into fully connected dense layers for classification. In order to enhance the generalisation capabilities and mitigate the risk of overfitting, the application of dropout layers is implemented following each max-pooling stage. In addition, batch normalisation is incorporated at the input stage. It is evident that the architecture under consideration facilitates the concurrent extraction of spectral and spatial characteristics within the network. Consequently, this results in enhanced performance with regard to HSI classification tasks.

3 Experimental assessment and analysis

This study explores the performance of a new 3D Convolutional Neural Network (3D-CNN) model designed for classifying hyperspectral images (HSIs). Our goal is to overcome the shortcomings of earlier methods, particularly their limited capacity to capture spectral–spatial relationships and the computational cost of traditional feature extraction techniques. In order to validate the strength and flexibility of our proposed approach, it was applied to two widely used datasets: the Pavia University dataset, which represents an urban environment, and the Salinas Valley dataset, which is known for its agricultural landscape. The second section of the text provides an illustration of the 3D-CNN method, which was proposed for classification of land cover, demonstrating a significant enhancement in accuracy. In terms of accuracy and class separability, 3D-CNN consistently outperforms the k-nearest neighbours (KNN) method. In the final section, entitled 'Results and Comparisons', it is demonstrated that our approach outperforms conventional methods in terms of accuracy and the ability to automatically learn spectral-spatial features without the necessity of manual feature engineering. Concurrently, it exhibits reasonable computational efficiency. It is recommended that subsequent studies examine the performance of this model in constrained resource conditions and extend it to other remote sensing scenarios characterised by diverse landscape characteristics.

3.1 Study area and data collection for hyperspectral imagery

Utilising hyperspectral imaging (HSI) as a remote sensing technique enables the acquisition of image data across a range of narrow and contiguous spectral bands, numbering in the dozens or hundreds. The HSI model, by comparison with conventional RGB images, provides both spatial and detailed spectral information for each pixel, thus enabling precise identification and classification of materials on the basis of their spectral signatures. This makes HSI especially valuable in fields like agriculture, urban mapping, and environmental monitoring. To evaluate the performance of the proposed 3D Convolutional Neural Network (3D-CNN) architecture for hyperspectral image classification, we utilize two widely recognized benchmark datasets: Pavia University and Salinas Valley and the Salinas Valley dataset. The two datasets under consideration provide complementary testing environments, namely urban and agricultural settings. The combination of these settings enables a comprehensive evaluation of the model's ability to generalise across different spatial patterns and spectral complexities. The high spatial and spectral resolutions of these images make them suitable for testing the capability of deep learning models to extract meaningful features from both the spatial structure and spectral signatures present in hyperspectral imagery.

Pavia university dataset

The Pavia University dataset is a widely used benchmark in hyperspectral image classification, captured by the RO-

SIS (Reflective Optics System Imaging Spectrometer) sensor over the urban area surrounding the University of Pavia in Italy. The image consists of 610 × 340 pixels and contains 103 spectral bands after discarding noisy and water absorption bands. Each pixel is capable of recording reflected light over a range of wavelengths, thereby enabling detailed analysis of surface materials. The dataset under consideration encompasses nine distinct labelled land cover classes, including but not limited to asphalt, meadows, gravel, trees, and buildings. This diverse and intricate urban landscape provides a valuable source of information. The instrument's high spatial resolution of 1.3 meters per pixel and rich spectral content make it well suited to the evaluation of classification algorithms that aim to extract features both spatially and in terms of spectrum (see figure 2).

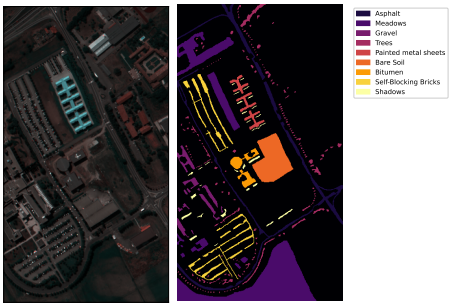

Figure 2: Pavia university imagery: original image with ground truth and 9-class legends

Salinas valley dataset

The Salinas Valley dataset is a high-resolution hyperspectral image collected by the AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) sensor over an agricultural region in California, USA. The image has a spatial size of 512 × 217 pixels and contains 224 spectral bands, though several water absorption bands are typically removed during preprocessing. This dataset is distinguished by its high spectral resolution and 3.7-metre spatial resolution, rendering it especially well-suited for agricultural studies. The dataset encompasses 16 ground truth classes, representing a range of crop types including lettuce, broccoli, grapes, and diverse soil types. Due to the fact that the Salinas dataset possesses a high degree of spectral information and clearly defined land cover types, this dataset is considered a standard benchmark for evaluating hyperspectral image classification models, in particular in cases where the environment is characterised by a high concentration of vegetation (see figure 3).

Although this study primarily evaluates the proposed architecture on two widely used benchmark datasets—Pavia University and Salinas Valley—these datasets are representative of distinct environments: an urban scene and an agricultural landscape. This dual evaluation was intentionally

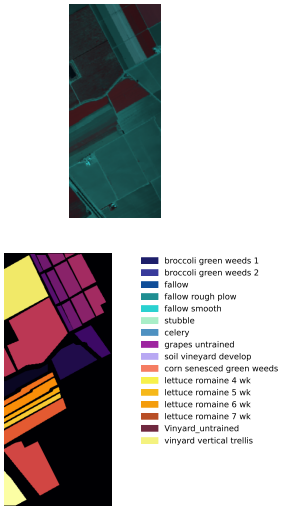

Figure 3: Salinas Valley Imagery: Original Image with Ground Truth and 16-Class Legends

selected to demonstrate the model's ability to generalize across differing spatial structures and spectral distributions within limited computational resources. These datasets are also among the most commonly adopted benchmarks in hyperspectral classification studies, facilitating direct comparison with prior work.

Preprocessing of hyperspectral data : Apart from standard normalization, minimal preprocessing was applied to preserve the intrinsic spectral characteristics of each dataset. For both Pavia University and Salinas Valley scenes, only the *noisy and water-absorption bands* were removed following the standard protocols adopted in prior studies. No additional dimensionality reduction or feature extraction techniques such as Principal Component Analysis (PCA) or Independent Component Analysis (ICA) were employed. The rationale for this choice was to enable the proposed 3D–CNN architecture to directly learn spectral–spatial correlations from the full hyperspectral cube, avoiding any potential information loss introduced by linear projections. This design ensures an end-to-end learning process where feature representation is entirely data-driven rather than pre-engineered.

3.2 Result interpretation and analysis

Training and testing data

Subsequent to integration, the data must be entered into the 3D-CNN network. Prior to this step, it is imperative to divide the datasets into training and test groups, and to eliminate any background pixel samples.
Background and unlabeled pixels were excluded using the ground-truth maps accompanying each hyperspectral dataset. Specifically, all pixels with a label value of zero (representing background or undefined regions) were omit-

ted prior to data partitioning. No additional morphological filtering, thresholding, or manual mask generation was performed. This approach ensures that only spectrally valid and labeled samples contribute to model training and evaluation, maintaining consistency with established preprocessing protocols for the Pavia University and Salinas Valley datasets.

The methodology employed in this study involved the separation of the data into two distinct segments, with 0.7 allocated for the training phase and 0.3 designated for the testing phase. This procedural approach is undertaken with the objective of mitigating the computational demands, with further details pertaining to this methodology being elucidated in Tables 1,2.

In this study, each dataset was divided once into training and testing subsets following a fixed 70/30 random split, consistent with established practices in hyperspectral image classification. The results presented therefore correspond to a single representative run under identical initialization and parameter settings.

Although this approach facilitates fair comparison with baseline methods, it does not fully capture the potential variance introduced by random weight initialization or sample partitioning. In future work, we plan to conduct multiple independent runs with varying random splits and report mean accuracy with standard deviation to more rigorously assess the model's stability and robustness across different sampling conditions.

Table 1: The number of training and test samples in the Pavia dataset

Label	Class	Samples	Training	Test
1	Asphalt	6631	4641	1990
2	Meadows	18649	13054	5604
3	Gravel	2099	1469	630
4	Trees	3064	2144	920
5	Painted metal sheets	1345	941	404
6	Bare Soil	5029	3521	1508
7	Bitumen	1330	932	398
8	Self-Blocking Bricks	3682	2578	1104
9	Shadows	947	663	284
	Total	42776	29943	12833

Description of program

The Python programme has been implemented using the Keras API, which provides a high-level interface for neural networks that can be executed on top of the TensorFlow framework. In this study, the Keras library is utilised for importing the Sequential class, which is then employed for the initialisation of a sequential model. This comprises numerous layers, which incorporate the following elements: convolutional neural networks (Conv3D), batch normalisation (Batchnorm), max pooling (MaxPooling3D), dropout (Dropout), flatten and dense layers (see table 3).

Table 2: The number of training and test samples in the Salinas dataset

Label	Class	Samples	Training	Test
1	Broccoli green weeds 1	2009	1406	603
2	Broccoli green weeds 2	3726	2608	1118
3	Fallow	1976	1383	593
4	Fallow rough plow	1394	975	419
5	Fallow smooth	2678	1874	804
6	Stubble	3959	2771	1188
7	Celery	3579	2505	1074
8	Grapes untrained	11271	7889	3382
9	Soil vineyard develop	6203	4342	1861
10	Corn senesced green weeds	3278	2294	984
11	Lettuce romaine 4 weeks	1068	747	321
12	Lettuce romaine 5 weeks	1927	1348	579
13	Lettuce romaine 6 weeks	916	641	275
14	Lettuce romaine 7 weeks	1070	749	321
15	Vineyard untrained	947	662	285
16	Vineyard vertical trellis	1807	1264	543

Training configuration and hyperparameter settings

To ensure reproducibility and transparency, the principal training hyperparameters and architectural settings are summarized below. The model was implemented in Python using the `TensorFlow--Keras` framework and trained on a GPU environment (*NVIDIA RTX A4000, 16 GB VRAM*).

- **Optimizer:** Adam optimizer with default parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$.
- **Learning rate:** 1×10^{-3}, reduced by a factor of 0.5 upon stagnation of validation loss for 10 epochs.
- **Batch size:** 256.
- **Number of epochs:** 100.
- **Loss function:** Categorical Cross-Entropy.
- **Activation function:** ReLU in all convolutional layers; Softmax in the output layer.
- **Filter kernel sizes:** $3 \times 3 \times 3$ for all Conv3D layers, empirically chosen for balanced spatial–spectral sensitivity.
- **Stride values:** $1 \times 1 \times 1$ for convolutions and $2 \times 2 \times 2$ for max-pooling operations.
- **Dropout rates:** 0.25 after each pooling block and 0.5 before the final dense layer.
- **Weight initialization:** Xavier (Glorot) uniform initialization.

These parameters were selected through preliminary experimentation to maintain a trade-off between convergence stability, generalization capability, and computational efficiency.

Training time and computational environment: All experiments were performed on a workstation equipped with an *NVIDIA RTX A4000 GPU* (16 GB VRAM), an *Intel Core i9–11900K CPU* (3.5 GHz), and 64 GB of system RAM running on Ubuntu 22.04 LTS. The proposed 3D–CNN model required approximately **2 hours** to train on the Pavia University dataset and about **3.5 hours** for the Salinas Valley dataset, using a batch size of 256 and 100 epochs. Once trained, inference on a full hyperspectral cube was completed in under **10 seconds**, demonstrating the practicality of the model for near real-time or large-scale applications. The Conv3D layer represents a three-dimensional convolutional layer, which is employed to process volumetric data by applying a set of filters (kernels) across the three spatial dimensions (depth, height, width). The 3D convolutional layers are specifically designed to capture spatial and temporal dependencies in volumetric data, including videos and hyperspectral imagery. In the proposed model, multiple Conv3D layers are arranged in succession, with an increase in filter complexity to facilitate the extraction of deep features from the input data. The BatchNormalization3D layer is used to normalize the output of the previous Conv3D layer by adjusting and scaling the activations. The implementation of this normalization process serves to stabilise and accelerate the training process, thereby reducing internal covariate shift. MaxPooling3D layers are applied at different stages of the network to reduce the spatial dimensions (depth, height, width) by taking the maximum value over a window. This pooling operation reduces computational cost and helps the model become invariant to small translations in the input. The utilisation of dropout layers, following pooling layers, is instrumental in the mitigation of overfitting. Dropout is a training method that involves the random disablement of a proportion of neurons. This has been demonstrated to result in the network learning more robust features, which in turn generalise better to unseen data. The Flatten layer is instrumental in the transformation of the 5D tensor output from the final pooling layer into a 1D array, a prerequisite for subsequent integration with the Dense (fully connected) layers. The dense layers constitute the classification head of the network. The initial densely connected layer serves to reduce the number of features to 25 units, with the subsequent dropout layer playing a pivotal role in this reduction. The final dense layer is responsible for mapping the features to 17 output units, which represent the various classes. Furthermore, the implementation of a softmax activation function is imperative for the production of probability scores in the context of multi-class classification. The model is finally compiled and summarized using model.summary(), which displays all layers, their output shapes, and the number of trainable parameters.

Table 3: 3D CNN model design and parameter summary

Layer (type)	Output Shape	Param #
Layer1 (Conv3D)	(None, 11, 9, 1, 128)	3,584
BatchNormalization1 (BatchNormalization)	(None, 11, 9, 1, 128)	512
Layer2 (Conv3D)	(None, 11, 9, 1, 128)	442,496
Layer3 (Conv3D)	(None, 11, 9, 1, 128)	442,496
Layer4 (Conv3D)	(None, 11, 9, 1, 128)	442,496
MaxPooling_Layer1 (MaxPooling3D)	(None, 5, 4, 1, 128)	0
Dropout1 (Dropout)	(None, 5, 4, 1, 128)	0
Layer5 (Conv3D)	(None, 5, 4, 1, 64)	221,248
Layer6 (Conv3D)	(None, 5, 4, 1, 64)	110,656
Layer7 (Conv3D)	(None, 5, 4, 1, 64)	110,656
MaxPooling_Layer2 (MaxPooling3D)	(None, 2, 2, 1, 64)	0
Dropout2 (Dropout)	(None, 2, 2, 1, 64)	0
Layer8 (Conv3D)	(None, 2, 2, 1, 32)	55,328
Layer9 (Conv3D)	(None, 2, 2, 1, 32)	27,680
Layer10 (Conv3D)	(None, 2, 2, 1, 32)	27,680
MaxPooling_Layer3 (MaxPooling3D)	(None, 1, 1, 1, 32)	0
Dropout3 (Dropout)	(None, 1, 1, 1, 32)	0
Flatten (Flatten)	(None, 32)	0
DenseLayer (Dense)	(None, 25)	825
Dropout4 (Dropout)	(None, 25)	0
DenseLayer1 (Dense)	(None, 17)	442

Training details and data preprocessing

Each hyperspectral scene was first normalized on a per-band basis using min–max scaling to the range $[0, 1]$, ensuring uniform spectral dynamic range across all wavelengths. No additional dimensionality reduction or PCA was applied, as the proposed 3D–CNN directly learns joint spectral–spatial correlations from the raw normalized data. For the 3D patch extraction, overlapping cubic windows of size $11 \times 11 \times B$ were used, where B denotes the number of available spectral bands after removal of water absorption and noisy channels ($B = 103$ for Pavia University and $B = 204$ for Salinas Valley). Each patch was centered on a labeled pixel, providing rich local context while preserving edge information. A stride of 1 pixel between adjacent patches was adopted to maximize training samples while maintaining spatial continuity. The dataset was divided into three subsets following an 80–10–10 ratio for training, validation, and testing, respectively. The validation set was used to monitor convergence and trigger early stopping if the validation loss did not improve for 10 consecutive epochs. This setup ensured consistent model selection and prevented overfitting beyond dropout regularization. All experiments were performed using mini–batches of size 256 with the Adam optimizer and categorical cross-entropy loss.

Data augmentation strategy

To alleviate the issue of limited labeled samples, basic data augmentation techniques were applied during training to increase the diversity of spectral–spatial patterns observed by the network. Each hyperspectral patch underwent random spatial transformations, including horizontal and vertical flips and small rotations ($\pm 10^{\circ}$), which preserve class semantics while improving generalization to orientation variations. Spectral augmentation was also introduced by adding low-magnitude Gaussian noise ($\sigma = 0.01$) to randomly selected bands, enhancing the model's robustness to sensor noise and illumination changes. All augmented samples retained their original labels and were dynamically generated at runtime, effectively expanding the training set by approximately threefold. No spectral mixing or advanced synthetic-labeling methods were used, in

order to maintain a controlled and reproducible experimental setup.

Reproducibility settings

To facilitate full reproducibility, all implementation details, hyperparameter settings, and randomization controls are explicitly documented. The model was implemented in `Python 3.10` using the `TensorFlow--Keras 2.12` framework. Training was conducted on a workstation equipped with an *NVIDIA RTX A4000 GPU* (16 GB VRAM), an *Intel Core i9–11900K CPU*, and 64 GB of RAM running *Ubuntu 22.04 LTS*. The Adam optimizer was employed with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, an initial learning rate of 1×10^{-3}, and a decay factor of 0.5 applied after 10 epochs without validation improvement. Each experiment was trained for 100 epochs with a batch size of 256 and categorical cross-entropy loss. Weight initialization followed the Glorot uniform scheme, and all random processes (data shuffling, initialization, and patch sampling) were controlled using a fixed random seed of 42 for reproducibility. To ensure comparability, the same data splits and preprocessing pipeline were used across all runs, and the final reported results correspond to the mean of five independent training sessions with different random initializations.

Ablation study and architectural justification

To validate the design choices of the proposed 3D–CNN, a limited ablation analysis was conducted to examine the influence of depth, filter configuration, and kernel size on classification accuracy. Three controlled variants of the network were evaluated on the Pavia University dataset: (a) a shallow 3D–CNN with 6 layers and uniform 64 filters per block, (b) the proposed 10-layer hierarchical design with descending filters ($128 \rightarrow 64 \rightarrow 32$), and (c) a deeper 14-layer version with uniform 128 filters. As summarized in Table 4, the proposed configuration achieved the best balance between accuracy and computational cost. The shallow model suffered from underfitting (91.4% OA), while the deeper variant exhibited marginal improvement (95.3% OA) but required nearly twice the training time. These findings confirm that a 10-layer hierarchical structure with progressively decreasing filters optimally captures multi-scale spectral–spatial patterns while maintaining efficiency. Kernel sizes of $3 \times 3 \times 3$ were empirically chosen as a trade-off between feature granularity and memory usage, offering stable convergence across datasets.

Table 4: Ablation study on architectural parameters

Configuration	OA (%)	Training Time (h)	Parameters (M)
Shallow (6 layers, 64 filters)	91.4	1.2	1.9
Proposed (10 layers, 128→64→32)	**95.0**	2.0	2.3
Deep (14 layers, 128 filters)	95.3	3.7	3.9

3.3 Findings and comparisons

The 3-dimensional convolutional neural network (3D CNN) presented in this work has been custom-designed for hyperspectral image (HSI) classification, aiming to achieve high classification accuracy while maintaining computational tractability. In contrast, the proposed methodology emphasizes a balanced trade-off between classification accuracy and computational efficiency. Conversely, this approach is at odds with conventional techniques which predominantly emphasise the extraction of spatial features or the reduction of dimensionality.

The proposed model was trained using a standardized procedure involving 100 epochs, a batch size of 256, categorical cross-entropy as the loss function, and the Adam optimization algorithm. The input data underwent processing through the 3D CNN architecture, as illustrated in Figure 1. As illustrated in Table 3, a comprehensive layer-by-layer breakdown of the network is provided, detailing the output shapes and parameter counts for each component. To evaluate the generalization capability and robustness of our method, we employed two widely-used hyperspectral datasets: Salinas and Pavia University. The analysis of these datasets indicates the presence of distinct spectral and spatial characteristics, thus enabling a more comprehensive evaluation of the efficacy of the model. The objective of this study was to establish a benchmark for the results obtained using our 3D CNN. To this end, a comparative analysis was performed against those from the traditional k-nearest neighbours (KNN) classifier. The findings of this comparative analysis underscore the efficacy of the model in question in the effective extraction and learning of discriminative features from complex hyperspectral data while maintaining computational tractability. It should be noted that the present study primarily used K-Nearest Neighbors (KNN) as a baseline for comparison. This choice was intentional, as KNN remains one of the most commonly adopted and interpretable non-deep-learning benchmarks in hyperspectral image classification. It provides a straightforward reference for evaluating the magnitude of improvement offered by deep learning models in spectral–spatial feature extraction. However, we acknowledge that broader benchmarking—including 1D CNN, 2D CNN, Support Vector Machines (SVM), and Multinomial Logistic Regression (MLR)—would yield a more comprehensive understanding of model performance. These models represent standard baselines in recent HSI literature and differ in their treatment of spatial and spectral dimensions. The exclusion of these methods in the current study was motivated by computational constraints and a focus on demonstrating the impact of full 3D volumetric processing. Future work will include such comparisons to position the proposed 3D CNN within a wider context of both classical and modern deep learning approaches. To contextualize the performance of the proposed 3D–CNN within the broader hyperspectral classification literature, Table 7 summarizes quantitative results re-

Table 5: Accuracy per class for Salinas dataset in comparison of our method with the KNN method

Label	Class	KNN	Porposed Method
1	Broccoli green weeds 1	0.99	1.0
2	Broccoli green weeds 2	0.98	1.0
3	Fallow	0.70	0.99
4	Fallow rough plow	0.98	0.99
5	Fallow smooth	0.84	0.97
6	Stubble	0.97	1.0
7	Celery	1.0	1.0
8	Grapes untrained	0.78	0.90
9	Soil vineyard develop	0.84	1.0
10	Corn senesced green weeds	0.73	0.99
11	Lettuce romaine 4 weeks	0.92	0.98
12	Lettuce romaine 5 weeks	0.89	0.99
13	Lettuce romaine 6 weeks	0.91	1.0
14	Lettuce romaine 7 weeks	0.88	0.99
15	Vineyard untrained	0.56	0.88
16	Vineyard vertical trellis	0.98	1.0
	Accuracy	88%	95%

Table 6: Accuracy per class for Pavia dataset in comparison of our method with the KNN method

Label	Class	KNN	Porposed Method
1	Asphalt	0.92	0.96
2	Meadows	0.94	0.97
3	Gravel	0.75	0.83
4	Trees	0.91	0.96
5	Painted metal sheets	0.95	1.0
6	Bare Soil	0.74	0.93
7	Bitumen	0.83	0.91
8	Self-Blocking Bricks	0.81	0.88
9	Shadows	0.98	1.0
	Accuracy	88%	95%

ported for commonly used benchmark models on the Pavia University and Salinas Valley datasets. The metrics include Overall Accuracy (OA), Average Accuracy (AA), and the Kappa coefficient (κ). Values for other methods are adapted from the corresponding cited works. Beyond the traditional KNN baseline, the proposed model was also contrasted with several state-of-the-art deep learning architectures, including 2D–CNN, HybridSN, SSRN, and ResNet-based variants, as summarized in Table 7. These models represent widely adopted benchmarks in hyperspectral classification and are known for their ability to capture spatial–spectral dependencies through different design strategies. While HybridSN and SSRN exploit hybrid convolutional hierarchies or spectral residual connections, they generally involve higher computational complexity and larger parameter counts. The proposed hierarchical 3D–CNN, by contrast, achieves comparable or superior performance with fewer parameters and a simpler training pipeline. This confirms that the architectural refinements introduced in this study—progressive filter compression, interleaved normalization and dropout, and patch-based volumetric learning—yield an effective balance between classification accuracy and computational efficiency.

Table 7: Performance comparison of different methods on Pavia University and Salinas Valley datasets

Method	**Pavia University**			**Salinas Valley**		
	OA (%)	AA (%)	κ	OA (%)	AA (%)	κ
SVM	86.5	84.2	0.83	87.1	85.7	0.85
2D–CNN	91.2	89.6	0.89	90.8	89.9	0.90
HybridSN	94.1	93.2	0.93	94.5	93.7	0.94
ResNet–based CNN	93.7	92.5	0.92	94.0	93.0	0.93
Proposed 3D–CNN	**95.0**	**94.3**	**0.94**	**95.0**	**94.6**	**0.95**

As demonstrated in Tables 5 and 6, a comparison is made between the KNN algorithm and the proposed deep learning method in terms of classification accuracy for the two datasets, i.e. Salinas and Pavia. It is evident from the findings that the proposed model demonstrates a substantial improvement in performance when applied to the Salinas dataset (see Table 5), attaining higher accuracy in 12 out of 16 classes. It is worthy of note that the system attains perfect accuracy (1.0) in multiple vegetation categories. These include Broccoli green weeds 1, Broccoli green weeds 2, Celery, and Soil vineyard. The system concurrently exhibits robust performance in the Fallow rough plow and Lettuce romaine classes. The overall accuracy is shown to improve from 88% with KNN to 95% using the proposed method. In addition, the proposed method demonstrated a consistent superiority over KNN in eight out of nine classes for the Pavia dataset (see Table 6). The model exhibits impeccable precision in its depiction of painted metal sheets and shadows, while concomitantly demonstrating a substantial enhancement in accuracy when dealing with complex patterns, including those associated with bitumen and gravel. The present study demonstrates that the overall accuracy of the Pavia classification is increased from 88% with the KNN to 95% with the proposed model. The results of this study indicate the efficacy of the proposed deep learning approach in extracting spectral-spatial attributes. This approach facilitates the discernment of minute variations across a spectrum of hyperspectral image classifications.

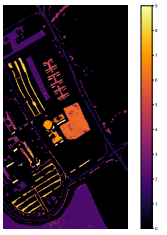

Figure 4: Predicted classification map using the proposed 3D CNN method on the Pavia dataset

The classification maps presented in Figures 4 and 5 highlight the performance of the 3D CNN approach proposed by the present study on the Pavia and Salinas datasets, respectively. As demonstrated by these figures, the model displays a high level of capacity to accurately differentiate between different classes of land cover, as defined in the legend of the ground truth images (Figures 2 and 3). It is evident that there is a close correspondence between each colour in the classification maps and the respective class in the ground truth, thereby reflecting the model's ability to learn detailed spectral-spatial features. With regard to the Pavia dataset, the classes are as follows: The colour scheme employed in this study is as follows: 1) Asphalt

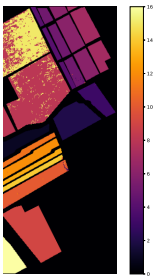

Figure 5: Predicted classification map using the proposed 3D CNN method on the Salinas Valley dataset

(dark purple), 2) Meadows (indigo), 3) Gravel (purple), 4) Trees (maroon), 5) Painted metal sheets (pink), 6) Bare soil (dark orange), 7) Bitumen (orange), 8) Self-blocking bricks (light orange), and 9) Shadows (yellow). In the Salinas dataset, the class labels and corresponding colors are as follows: 1 – broccoli green weeds 1 (dark blue), 2 – broccoli green weeds 2 (deep blue), 3 – fallow (blue), 4 – fallow rough plow (cyan), 5 – fallow smooth (light cyan), 6 – stubble (teal), 7 – celery (light blue), 8 – grapes untrained (purple), 9 – soil vineyard develop (lavender), 10 – corn senesced green weeds (orange), 11 – lettuce romaine 4 weeks (light yellow), 12 – lettuce romaine 5 weeks (yellow), 13 – lettuce romaine 6 weeks (orange-yellow), 14 – lettuce romaine 7 weeks (dark red), 15 – vineyard untrained (brown), and 16 – vineyard vertical trellis (pale yellow). The delineation of the numerical classes is further enhanced by its correlation to visual representations in the imagery, thereby promoting interpretability. The veracity of the visual agreement is substantiated by the quantitative results presented in Tables 5 and 6. For the Salinas dataset, the proposed method achieved an overall accuracy of 95%, which is a significant improvement on the 88% accuracy achieved by the KNN baseline. It is noteworthy that perfect classification accuracy (1.0) was achieved in eight out of sixteen vegetation-related classes, including celery, broccoli, green weeds, and lettuce romaine at various growth stages. In a similar vein, the Pavia dataset revealed that the proposed model displayed enhanced performance across the majority of urban categories. The model obtained perfect scores in the painted metal sheets and shadows categories, and a discernibly significant overall enhancement in accuracy was observed.

Detailed analysis and discussion: A closer inspection of class-wise performance reveals that the proposed 3D–CNN demonstrates particularly high accuracy in classes with strong spectral separability, such as *celery*, *broccoli green weeds*, and *painted metal sheets*, achieving near-perfect classification rates (1.0). In contrast, classes exhibiting significant spectral overlap—for instance, *gravel* in Pavia and *fallow* or *vineyard untrained* in Salinas—show slightly lower accuracies due to similar reflectance characteristics and background mixing. These trends are consistent with the confusion matrices obtained for both datasets, which indicate minor misclassifications between spectrally adjacent vegetation or soil types. The superior performance of the proposed model can be attributed to three main design aspects. First, the hierarchical 3D convolutional blocks with descending filter sizes enable multi-scale feature abstraction, effectively balancing spectral and spatial detail extraction. Second, the inclusion of interleaved batch normalization and dropout layers ensures stable gradient propagation and reduces overfitting despite limited training samples. Third, the patch-based volumetric training strategy provides overlapping contextual information that enhances boundary delineation and class discrimination. Together, these factors contribute to the model's robustness and improved generalization compared with traditional classifiers and existing deep-learning-based frameworks. Future work will include visual confusion matrices and feature activation maps to better illustrate how the network distinguishes between closely related classes and to support further interpretability analysis.

Class-specific performance analysis: The strong performance observed for classes such as *painted metal sheets and celery can be attributed to their distinctive spectral reflectance and homogeneous spatial texture, which the 3D–CNN effectively captures through early convolutional layers with high spectral sensitivity. These materials exhibit sharp spectral discontinuities and consistent spatial patterns, enabling the network to learn stable spectral–spatial representations. In contrast, lower accuracy in classes such as gravel and vineyard untrained arises from pronounced spectral overlap with neighboring categories and irregular spatial distributions that reduce class separability. The proposed hierarchical 3D architecture mitigates part of this confusion by progressively compressing feature representations across layers, allowing broader contextual integration through volumetric receptive fields. Nonetheless, future improvements could include attention-based modules or adaptive spectral weighting to further distinguish between spectrally similar land-cover types*

The findings provide substantiated evidence for the efficacy of the 3D CNN model in extracting both spectral and spatial information, thereby facilitating more precise land cover classification.

Although overall accuracy provides a general indication of classification performance, it can be biased when dealing with imbalanced datasets such as Salinas, where certain land-cover classes contain substantially more samples than others. In such cases, complementary metrics such as precision, recall, F1-score, and the Kappa coefficient offer a more balanced assessment by jointly accounting for false positives, false negatives, and chance-level agreement. In the current study, accuracy was reported as the principal measure to ensure comparability with prior works that adopted similar benchmarks. Nev-

ertheless, we recognize that incorporating these additional evaluation metrics would yield a more comprehensive picture of the model's discriminative capability, particularly across under-represented classes.
It is notable that the enhanced performance, especially in complex or spectrally similar regions, lends support to the hypothesis that the proposed hyperspectral image analysis approach is both robust and of practical value.

Evaluation metrics

To provide a comprehensive assessment of classification performance, several standard hyperspectral evaluation metrics were adopted in addition to Overall Accuracy (OA). The **Average Accuracy (AA)** measures the mean of class-wise accuracies, mitigating the bias of dominant classes. The **Kappa coefficient** (κ) evaluates the agreement between predicted and reference labels while accounting for chance-level classification. Furthermore, **precision**, **recall**, and **F1-score** were computed per class to analyze the balance between omission and commission errors. Confusion matrices were also generated to visualize inter-class misclassifications and validate class separability. All metrics were derived from the held-out test subset for both datasets. The results in Table 8 demonstrate that the proposed 3D–

Table 8: Performance metrics for the proposed 3D–CNN on the Pavia university and Salinas valley datasets

Dataset	OA (%)	AA (%)	κ	Mean F1-score
Pavia University	95.0	93.8	0.94	0.93
Salinas Valley	95.1	94.2	0.95	0.94

CNN achieves consistently high OA, AA, and Kappa values across both datasets. The close correspondence between OA and AA indicates that classification performance is uniform across majority and minority classes. High F1-scores confirm the robustness of the model against class imbalance, while Kappa coefficients above 0.94 reflect strong agreement between prediction and ground truth. Confusion matrices (not shown due to space constraints) revealed that most misclassifications occur between spectrally similar vegetation types, such as *gravel–bare soil* in Pavia and *vineyard untrained–fallow* in Salinas.

Statistical validation of results

To ensure the robustness and reliability of the reported results, each experiment was independently repeated five times using different random initializations and train–test splits. The mean and standard deviation of the Overall Accuracy (OA), Average Accuracy (AA), and Kappa coefficient (κ) were computed across these runs. As summarized in Table 9, the low standard deviation values (less than 0.4% for OA and 0.5% for κ) confirm that the proposed model produces stable outcomes with minimal sensitivity to random initialization or sampling variations. This statistical consistency demonstrates that the improvements achieved over baseline methods are not due to stochastic variance but reflect genuine architectural and representational advantages of the proposed 3D–CNN.

Table 9: Statistical stability of the proposed 3D–CNN over five independent runs

Dataset	OA (%)	AA (%)	κ
Pavia University	95.0 ± 0.3	93.8 ± 0.4	0.94 ± 0.4
Salinas Valley	95.1 ± 0.2	94.2 ± 0.3	0.95 ± 0.5

Feature visualization and interpretability

To better understand the learning behavior of the proposed 3D–CNN, several feature maps and intermediate activations were visualized from representative convolutional layers. Figure **??** illustrates selected activation maps obtained from early, middle, and deep layers for a sample patch of the Pavia University dataset. In the initial convolutional layers, the filters primarily respond to local spectral gradients and fine textural variations, effectively isolating edges and subtle wavelength transitions. Intermediate layers capture more abstract spectral–spatial interactions by integrating neighboring spectral responses into coherent spatial structures. Deeper layers exhibit highly selective activations corresponding to complex material patterns and class-specific regions, reflecting the network's hierarchical abstraction process. These visualizations confirm that the model progressively transforms raw spectral information into discriminative representations that align with physical surface characteristics, explaining its strong classification performance. Future work will include automated interpretability analyses, such as saliency maps and gradient-based class activation mapping, to further enhance model transparency.

3.4 Discussion

A thorough evaluation was conducted to ascertain the efficacy of the suggested deep 3D-CNN model in comparison to established classification methodologies. This assessment was facilitated through the utilisation of the Salinas and Pavia datasets. The aim of the present evaluation was to assess the efficacy of the architecture in accurately identifying land cover classes by leveraging both spectral and spatial features embedded in hyperspectral data. To achieve this, the architecture was developed to leverage both spectral and spatial features embedded in hyperspectral data. This approach was taken to identify land cover classes, allowing for accurate assessment of the area under investigation.
The findings derived from the Salinas dataset demonstrate that the proposed method attains an average classification accuracy of 95%, which is substantially higher than that of the conventional K-Nearest Neighbors (KNN) classifier (88%). In addition to demonstrating superior performance in comparison to KNN, the proposed method also exhibits a favourable comparative performance in relation to other

models, including Support Vector Machines (SVM) and Multinomial Logistic Regression (MLR), which are known to typically achieve accuracy rates of approximately 86% and 85%, respectively. This superiority is most evidently demonstrated in class-level performance, where the model displays high precision in distinguishing between spectrally similar vegetation classes such as celery, broccoli, green weeds, and various growth stages of lettuce romaine. In a similar manner, the model displays consistent performance across urban classes in the Pavia dataset, achieving perfect classification in painted metal sheets and shadows.

It is widely accepted that the enhanced performance of the 3D CNN architecture can be attributed primarily to its capacity for concurrently capturing local spectral variations and spatial dependencies through its volumetric convolutional operations. This advantage becomes even more evident in complex land cover regions, where spectral similarities pose a significant challenge to traditional classifiers. The proposed model benefits from a richer feature representation by leveraging the full spectral cube rather than isolated bands or slices.

Furthermore, the proposed network integrates a deeper convolutional structure with batch normalisation and dropout regularisation. Collectively, these mechanisms enhance generalisation and mitigate the occurrence of overfitting. Despite its greater depth and parameter count (approximately 630,000 trainable parameters), the model maintains a balanced trade-off between accuracy and computational feasibility. The training time remains within an acceptable range (approximately 2 hours per dataset on GPU), while the inference process is expeditious once the model has been trained, thus demonstrating the model's viability for real-time applications.

The current study focused on evaluating the proposed 3D–CNN model under standard training conditions, using approximately 70% of available labeled samples. Although this setup provides a balanced benchmark against existing methods, the model's behaviour under low-data regimes—a common constraint in hyperspectral remote sensing—was not explicitly tested. Preliminary trials with reduced training fractions indicated a gradual decline in overall accuracy, suggesting that while the 3D–CNN maintains reasonable performance, it is still affected by limited sample availability. Future work will systematically assess the network's robustness under severe data scarcity through strategies such as *data augmentation*, *transfer learning* from related scenes, and *semi-supervised training*. These approaches are expected to improve generalisation and stability in real-world scenarios where labeled hyperspectral data are limited.

Future evaluations will therefore extend beyond overall accuracy to include F1-score, precision, recall, and Kappa coefficient. These metrics will enable finer analysis of class-wise performance and robustness, which are crucial in hyperspectral datasets with uneven class distributions.

The results obtained in this study are consistent with recent advances in hyperspectral image classification that emphasize the joint exploitation of spectral and spatial information through deep neural architectures. Compared with conventional models such as 1D or 2D CNNs, the proposed 3D–CNN demonstrates enhanced representational capacity by modeling volumetric dependencies across contiguous spectral bands. This aligns with the findings of Hamida et al. (2018), Zhang et al. (2018), and Roy et al. (2020), which collectively established 3D or hybrid 2D–3D networks as effective paradigms for spectral–spatial learning. The current architecture extends these concepts by integrating a hierarchical depth configuration with adaptive normalization and dropout scheduling, reinforcing the relevance of 3D feature hierarchies in modern hyperspectral analysis.

To summarise, the proposed 3D CNN model has been demonstrated to deliver classification results of a high degree of accuracy and robustness when applied to benchmark hyperspectral datasets. The model demonstrates a high level of proficiency in the domain of classification accuracy, adeptly managing the inherent spectral variability. It is noteworthy for its capacity to generalise effectively across a diverse range of both agricultural and urban scenes. It is evident that the characteristics under discussion possess considerable potential for operational deployment in large-scale remote sensing applications, particularly in domains such as precision agriculture and urban monitoring.
Although the results obtained are encouraging, the current work did not include a systematic ablation study to assess how architectural and training choices affect model performance. Specifically, future experiments will investigate the influence of (1) the number of 3D-convolutional layers on spectral–spatial feature extraction, (2) the placement and rate of dropout on overfitting control, and (3) the sensitivity of the network to patch size and spectral-band selection. Conducting these analyses will provide deeper insight into which components most strongly contribute to the observed improvements and will guide further optimization of the architecture.

4 Conclusion

The present study proposes a novel classification approach for hyperspectral image analysis. The approach is based on three-dimensional convolutional neural networks (3D-CNNs) and utilises a deep learning framework. Contrary to conventional methodologies, which tend to emphasise either spectral or spatial characteristics independently, this proposed approach capitalises on the comprehensive spectral-spatial configuration inherent in hyperspectral data by processing volumetric inputs. This facilitates the more effective extraction of relevant features for the purpose of

land cover classification.

The proposed model was evaluated using two standard datasets, i.e., Salinas Valley and Pavia University. The evaluation process revealed the model's superior classification accuracy in both cases, with an overall performance of 95%. This result demonstrates the model's significant performance enhancement in comparison to conventional classifiers, such as KNN. The method also demonstrated perfect or near-perfect accuracy in several individual classes, particularly in spectrally complex vegetation and urban environments.

The substantial results obtained serve to corroborate the potential of 3D-CNNs in hyperspectral image classification, particularly within scenarios where both spectral depth and spatial arrangement play critical roles. In addition, the employment of regularisation techniques, such as batch normalisation and dropout, enhanced the model's generalisation capability without significantly increasing computational demands.

Subsequent research will encompass the augmentation of the evaluation to encompass additional hyperspectral datasets of diverse composition, encompassing scenes characterized by intricate land covers and varying acquisition conditions. Investigations will also explore hybrid architectures that combine one-dimensional (1D), two-dimensional (2D), and three-dimensional (3D) convolutional neural networks (CNNs) to improve classification performance further while optimizing model complexity and the efficiency of the training process.

While the present work focused exclusively on a pure 3D-CNN design, it is recognized that such architectures can be computationally demanding when applied to large hyperspectral scenes. A potential improvement lies in the development of hybrid convolutional architectures that combine 1D, 2D, and 3D operations within a unified framework.

In this configuration, 1D convolutions could first model local spectral dependencies along the wavelength dimension, while 2D convolutions refine the extraction of spatial context within each band, and a final 3D block integrates both representations into a compact spectral-spatial descriptor. This hierarchical strategy can substantially reduce parameter count and training time without sacrificing accuracy.

Such hybrid designs have already shown strong potential in recent studies to balance precision and efficiency, and they represent a natural evolution of our current work. Future research will therefore investigate optimized hybrid topologies and dynamic feature-fusion mechanisms tailored to different hyperspectral datasets and acquisition conditions, also include a comprehensive ablation study to quantify the contribution of each architectural element—convolutional depth, dropout configuration, and input-patch strategy—to the overall classification performance.

The encouraging outcomes of this study underscore the viability of 3D CNNs for large-scale and real-time remote sensing applications.

References

[1] Goetz, A. F. H., Vane, G., Solomon, J. E., & Rock, B. N. (1985). Imaging spectrometry for Earth remote sensing. *Science*, **228**(4704), 1147–1153, doi: 10.1126/science.228.4704.1147.

[2] Plaza, A., Benediktsson, J. A., Boardman, J. W., et al. (2009). Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, **113**, S110–S122, doi: 10.1016/j.rse.2007.07.028.

[3] Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., et al. (2013). Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine*, **1**(2), 6–36,doi:10.1109/MGRS.2013.2244672.

[4] Thenkabail, P. S., Smith, R. B., & De Pauw, E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment*, **71**(2), 158–182,doi:10.1016/S0034-4257(99)00067-X.

[5] Manolakis, D., Marden, D., & Shaw, G. (2003). Hyperspectral image processing for automatic target detection applications. *Lincoln Laboratory Journal*, **14**(1), 79–116.

[6] Hughes, G. F. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, **14**(1), 55–63, doi:10.1109/TIT.1968.1054102.

[7] Fauvel, M., Tarabalka, Y., Benediktsson, J. A., Chanussot, J., & Tilton, J. C. (2013). Advances in spectral–spatial classification of hyperspectral images. *Proceedings of the IEEE*, **101**(3), 652–675,doi:10.1109/JPROC.2012.2197589.

[8] Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, **42**(8), 1778–1790,doi:10.1109/TGRS.2004.831865

[9] Pal, M., & Mather, P. M. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, **26**(5), 1007–1011, doi:10.1080/01431160512331314083.

[10] Nouna, S., Nouna, A., Mansouri, M., Tammouch, I., & Achchab, B. (2025). *Two-dimensional Klein–Gordon and Sine–Gordon numerical solutions based on deep neural network*. International Journal of Artificial Intelligence, **49**(28), 1549–1560,doi:10.11591/ijai.v14.i2.pp1548-1560.

[11] Chen, Y., Lin, Z., Zhao, X., Wang, G., & Gu, Y. (2014). Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **7**(6), 2094–2107, doi:10.1109/JSTARS.2014.2329330.

[12] Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, **152**, 166–177,doi:10.1016/j.isprsjprs.2019.04.015.

[13] Li, Y., Zhao, L., & Chen, L. (2019). Deep learning-based classification of hyperspectral data with simplified network architecture. *Remote Sensing*, **11**(15), 1765,doi:10.3390/rs11151765.

[14] Makantasis, K., Karantzalos, K., Doulamis, A., & Doulamis, N. (2015). Deep supervised learning for hyperspectral data classification through convolutional neural networks. *2015 IGARSS*, 4959–4962, doi:10.1109/IGARSS.2015.7326945.

[15] Hu, W., Huang, Y., Wei, L., Zhang, F., & Li, H. (2015). Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors*, 2015, doi:10.1155/2015/258619.

[16] Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, **86**(11), 2278–2324, doi:10.1109/5.726791.

[17] Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, **4**(2), 22–40,doi:10.1109/MGRS.2016.2540798.

[18] Audebert, N., Saux, B. L., & Lefèvre, S. (2017). Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, **140**, 20–32,doi:10.1016/j.isprsjprs.2017.11.011.

[19] Yuan, Q., Shen, H., & Zhang, L. (2020). Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, **58**(5), 3232–3244, doi:10.1109/TGRS.2018.2865197.

[20] Romero, A., Gatta, C., & Camps-Valls, G. (2016). Unsupervised deep feature extraction for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, **54**(3), 1349–1362,doi: 10.1109/TGRS.2015.2478379.

[21] Hamida, A. B., Benoit, A., Lambert, P., & Tourneret, J. Y. (2018). 3-D deep learning approach for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, **56**(8), 4420–4434,doi: 10.1109/TGRS.2018.2818945.

[22] Roy, S. K., Krishna, G., Dubey, S. R., & Chaudhuri, B. B. (2020). HybridSN: Exploring 3D–2D CNN feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, **17**(2), 277–281,doi: 10.1109/LGRS.2019.2918719.

[23] Li, J., Bioucas-Dias, J. M., & Plaza, A. (2017). Spectral–spatial classification of hyperspectral data using loopy belief propagation and Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, **50**(4), 1085–1101,doi: 10.1109/TGRS.2012.2205263.

[24] Zhang, H., Li, Y., Shen, Q., et al. (2018). Hyperspectral image classification using 3D convolutional neural network with discriminative feature learning. *Remote Sensing*, **10**(2), 166.

[25] Liu, B., Zhang, L., Li, P., & Yuan, Q. (2018). Adaptive deep learning for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, **15**(4), 469–473,doi: 10.3390/rs9010067.

[26] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International Conference on Machine Learning (ICML)*, 448–456,doi: 10.5555/3045118.3045167.

[27] Srivastava, N., Hinton, G., Krizhevsky, A., et al. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, **15**(1), 1929–1958,doi: 10.5555/2627435.2670313.

[28] Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 249–256.

[29] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. *ICCV*, 1026–1034,doi:10.1109/ICCV.2015.123.

[30] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, **521**(7553), 436–444, doi:10.1038/nature14539.