The Bioscan AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES 21(2): 773-799, 2026 www.thebioscan.com

A MULTIMODAL FUSION ARCHITECTURE COMBINING DTICF-3D-CNN–BI-LSTM AND RGB-ENHANCED 2D-CNN FOR EARTH SURFACE HSI CLASSIFICATION

M Preethi [a*], T Idhaya [b], S kanagasankari [c], J Rexy [d]

[a] Department of Computer Science, St. Xavier's College (Autonomous), Palayamkottai, Tirunelveli, Tamil Nadu, India, preethi_cs@stxavierstn.edu.in *** Corresponding author.**

[b] Department of Computer Science, St. Xavier's College (Autonomous), Palayamkottai, Tirunelveli, Tamil Nadu, India, idhaya.cs@stxavierstn.edu.in

[c] Department of Computer Science, St. Xavier's College (Autonomous), Palayamkottai, Tirunelveli, Tamil Nadu, India, kanagasankari_cs@stxavierstn.edu.in

[d] Department of Computer Science, St. Xavier's College (Autonomous), Palayamkottai, Tirunelveli, Tamil Nadu, India, rexy_cs@stxavierstn.edu.in

DOI: https://doi.org/10.63001/tbs.2026.v21.i02.pp773-799

KEYWORDS

Hyperspectral Image, 3-Dimensional Convolutional Neural Network with Bi Directional Long Short Term Memory Network, Markov Random Field, Domain Transform Interpolated Convolution Filter.

Received on:
20-04-2026

Accepted on:
08-05-2026

Published on:
18-05-2026

Abstract

Hyperspectral imaging technology is one of the most efficient and growing technologies derived from Image Sensing technology. HSI classification contains rich spectral information which is used to identify and classify the information of the earth surface. Due to large dimensionality of spectral information, HSI classification still faces several challenges and also creates the problem of detecting the target of the classes accurately. To overcome this problem, we proposed a Spectral-Spatial Classification of HSI using fusion of 3-Dimensional Convolutional Neural Network with Bi Directional Long Short Term Memory Network (3D-CNN-BiLSTM)and 2-Dimensional-CNN (2D-CNN) approach. Finally, Markov Random Field (MRF) is utilized to encourage the neighboring pixel which helps to enhance the classification results. To prove the better performance of the proposed techniques, some of the existing Machine learning and deep learning techniques are compared. From this analysis, it is evaluated that, the proposed techniques provide the best results, when compared to other techniques.

1. Introduction

HSI classification is used for classifying data into different classes according to some constraints. Besides, it is utilized to predict categorical class labels and classifies data based on training set and class labels [1-2]. HSI is gathered by imaging spectrometer and retains uninterrupted spectral information. The large number of spectral information is captured by electromagnetic spectrum from visible to near-infrared regions [3]. As a consequence, high dimensional data cubes are extracted. Due to large dimensionality of spectral information, HSI classification still faces several challenges [4]. There are two types of HSI classification such as spectral classification and spatial-spectral classification. In spectral classification, it only handles the spectral information. In spectral-spatial classification, spectral and spatial features are classified concurrently. It improves the classification performance

1.1 Literature Review

Recent research focuses on deep learning-based spectral–spatial feature extraction, hybrid fusion architectures, and spatial refinement techniques to overcome these limitations. Li et al [5]. proposed a comprehensive review of recent spectral–spatial feature extraction techniques for hyperspectral image classification. The study discussed traditional machine learning, 2D-CNN, 3D-CNN, and hybrid deep learning approaches. The authors explained that 3D-CNN models effectively predict joint spectral–spatial information, while fusion-based architectures given superior performance. This work strongly supports the motivation for combining 3D-CNN and 2D-CNN in spectral–spatial classification. Zhang et al [6]. developed a dual-stream convolutional neural network for hyperspectral image classification. This author discussed on spectral feature extraction, while the other captures spatial contextual information. The extracted spatial-spectral features are combined to improved class discrimination. The results explained that fusion-based CNN models outperform single-stream networks, supporting the need for multi-model fusion as adopted in the proposed work. Luo et al [7]. proposed a hybrid 3D-CNN and Bi-LSTM network for hyperspectral image classification. The 3D-CNN extracts local spectral–spatial features, while the Bi-LSTM models long-range spectral dependencies in both forward and backward directions. These fusion techniques feature representation and classification accuracy. The study directly aligns with the proposed fusion of 3D-CNN and Bi-LSTM for effective spectral–spatial learning. Mou et al [8]. explored recurrent neural networks for hyperspectral image classification by treating spectral bands as sequential data. LSTM units were employed to model spectral dependencies effectively. The results showed improved performance over traditional CNN-based approaches. This work highlights the importance of sequence modeling, which elaborated the use of Bi-LSTM in the proposed framework. Paoletti et al [9]. investigated the effectiveness of 3D convolutional neural networks for hyperspectral image classification. The proposed approach simultaneously processes spatial neighborhoods and spectral bands. This joint learning mechanism significantly improves classification accuracy. The study establishes 3D-CNN as a strong baseline for spectral–spatial feature extraction, forming a core component of the proposed system. Fang et al [10]. proposed a multi-feature fusion framework combining spectral, spatial, and texture information using deep learning. Feature fusion was shown to improve class separability in complex hyperspectral scenes. The work discussed that integrating heterogeneous features yields higher accuracy than single-feature models. This supports the fusion strategy adopted in the proposed work. Zhong et al [11]. developed a deep residual spectral–spatial network for hyperspectral image classification. Residual learning enabled deeper architectures and better feature extraction. The network effectively extracted hierarchical spectral–spatial features. This work highlights the importance of deep architectures for improved classification accuracy. He et al [12]. proposed a graph convolution-based framework to model spatial relationships among neighboring pixels. The graph structure evaluated spatial consistency in classification maps. This approach reduced misclassification noise in homogeneous regions. The study motivated the use of spatial refinement techniques such as Markov Random Fields (MRF). Zhou et al [13]. introduced a fusion-based deep learning model combining 2D-CNN and 3D-CNN architectures. The 2D-CNN captured spatial texture patterns, while the 3D-CNN extracts spectral–spatial features. Feature fusion enhances classification accuracy. This approach closely aligned with the proposed fusion of 3D-CNN-BiLSTM and 2D-CNN. The literature survey concluded that, fusion techniques of 3D-CNN is utilized to provide the better earth surface HSI classification.

1.2 Importance of the Study

Hyperspectral imaging (HSI) provides rich spectral information across numerous bands, making it highly useful for applications such as land cover mapping, agriculture monitoring, and environmental analysis. The proposed fusion-based deep learning model enhances classification accuracy and provides reliable results for practical HSI applications. This work is applied to earth surface hyperspectral image classification, mainly for remote sensing applications such as land cover mapping, agricultural monitoring, vegetation analysis, urban planning, and environmental assessment. In this application, hyperspectral images captured from airborne or satellite sensors contain rich spectral information for each ground pixel. The proposed 3D-CNN–BiLSTM model is used to extract joint spectral–spatial features and to model spectral band dependencies for accurate material identification. The RGB-2D-CNN branch enhances spatial texture information by utilizing chromatic features. Finally, the fusion output refined with MRF improves spatial consistency of neighboring pixels, resulting in more accurate and smooth land cover classification maps of the earth surface.

1.3 Challenges in Hyperspectral Image Classification

HSI classification suffers from high dimensionality, spectral redundancy, and limited labeled samples, leading to the curse of dimensionality. Spectral similarity between different classes and intra-class variability further reduce accuracy. Traditional methods cannot fully capture complex spectral–spatial relationships, while many deep learning models struggle to simultaneously model spectral dependencies and spatial context. These challenges necessitate an effective hybrid fusion approach.

1.4 Research Contribution

This paper proposes a spectral–spatial classification framework that integrates 3D-CNN, BiLSTM, and RGB-based 2D-CNN with MRF refinement. The 3D-CNN extracts joint spectral–spatial features, and BiLSTM captures bidirectional spectral dependencies. The 2D-CNN processes RGB features to enhance spatial texture information. The fused outputs are refined using MRF to improve spatial consistency. Experimental results show that the proposed method outperforms existing machine learning and deep learning techniques.

1.5 Organization of the Paper

The remainder of this paper is organized as follows. Section I presents the introduction and discusses the importance and challenges of hyperspectral image classification. Section II reviews existing machine learning and deep learning approaches related to spectral–spatial HSI classification. Section III describes the proposed multimodal fusion framework, including DTICF-based spatial feature extraction, 3D-CNN for joint spectral–spatial learning, Bi-LSTM for spectral dependency modeling, RGB-based 2D-CNN feature extraction, fusion strategy, and MRF-based spatial refinement. Section IV explains the experimental setup, benchmark datasets. Section V presents the experimental results, performance comparisons with existing methods, and detailed analysis. Finally, Section VI concludes the paper and highlights future research directions.

2. Methodology

The proposed approach of the flow chart is given in the Figure 1. At the initial stage, RGB color features are extracted by using chromaticity computation. The obtained color features are provided to 2D-CNN. Secondly, 3-D patches as $S \times S \times D$ from HSI are extracted and $S \times S$ is denoted by spatial information and spectral information is denoted by D. The spatial features are extracted by DTICF. The

spatial features and extracted patches are fused. The fused information is given to 3D-CNN network. In this 3D-CNN network, 3D convolution layers are used for extracting features. At the starting stage of 3D-CNN, 3D convolution filters with different sizes simultaneously captured the spatial – spectral feature maps. After completing the 3D convolution operation, max pooling operation is used to compress the size of the extracted features. Finally, these Rectified Linear Unit (ReLU) [14] features are provided to Bi-LSTM network. This network is utilized to provide input in both directions. Soft-Max activation function is used in the FC layer which is used to predict the related pixel vector for HSI. In the third processing, the probabilistic classification map of 3D-CNN-Bi-LSTM and 2D-CNN are fused. Finally, MRF is utilized for enhancing the classification results.

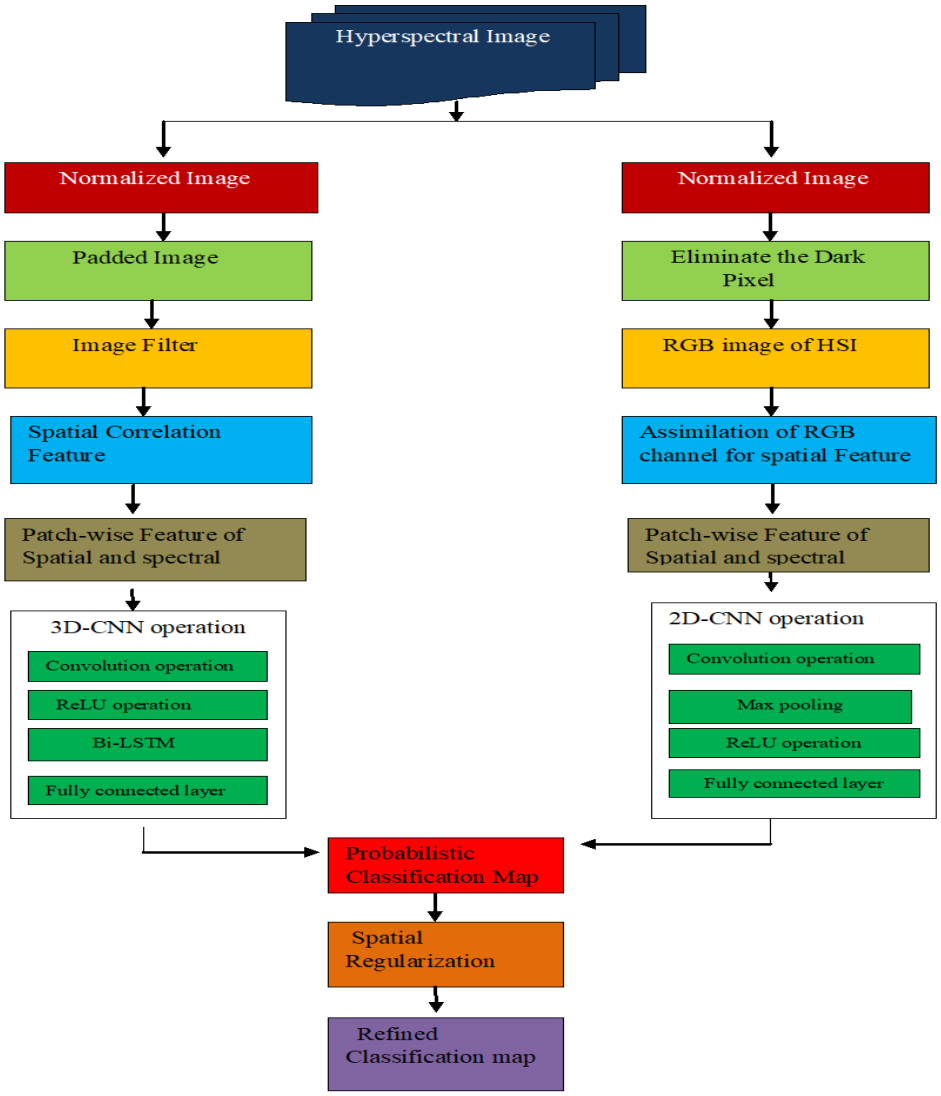


Figure 1: Structure of proposed work

2.1 CNN Operation

CNN is a special kind of neural network and very similar to ordinary neural network [15]. This network is able to learn the weight and bias. CNN is used for major applications such as image recognition, image classification, detection of object and recognizing faces. In CNN, Convolution operation is the most important mathematical operation. Convolution is a specialized kind of linear operation. Convnets are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers. The following Figure 2 shows the mathematical operation of CNN.

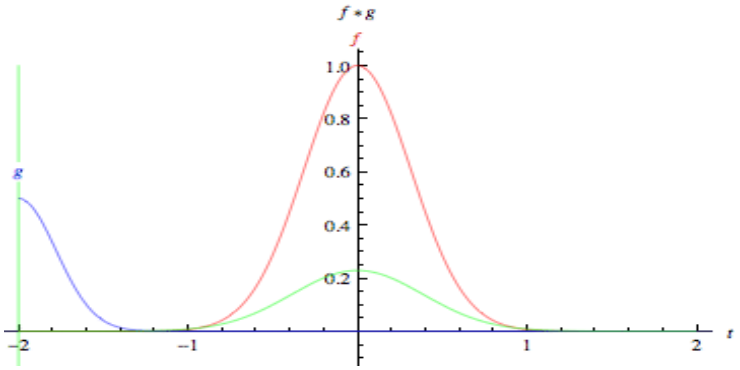


Figure 2: Mathematical operation of CNN

In CNN, input images are performed by some operations such as feature detector, kernel, filter and feature map.

2.2 3D Convolution

3D convolution [16-17] is used to apply a 3-dimensional filter in the data and contains a 3-direction (x, y, and z). It is used to calculate the low level feature representations. It provides the output shape as a 3-dimensional volume space such as cube or cuboid. The below Figure 3 shows the 3D convolution operation. Compared with 2D-CNN, our proposed method of 3D-CNN simultaneously extracts pair of spatial and spectral features.

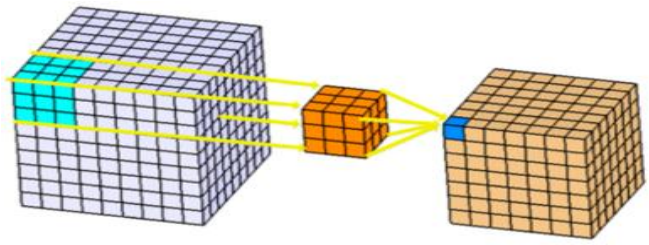


Figure 3: Structure of 3D-Convolution

2.4 The Framework of the Bi-LSTM Network

It is same as LSTM. In LSTM [19], we follow the input in one direction either backwards or forwards. Compared to LSTM, Bi-LSTM follows the input from both directions such as backwards (future to past) or forwards (past to future) [20]. It is different from regular LSTM. It is used to predict the future and past information. The block diagram of LSTM is displayed in Figure 4.

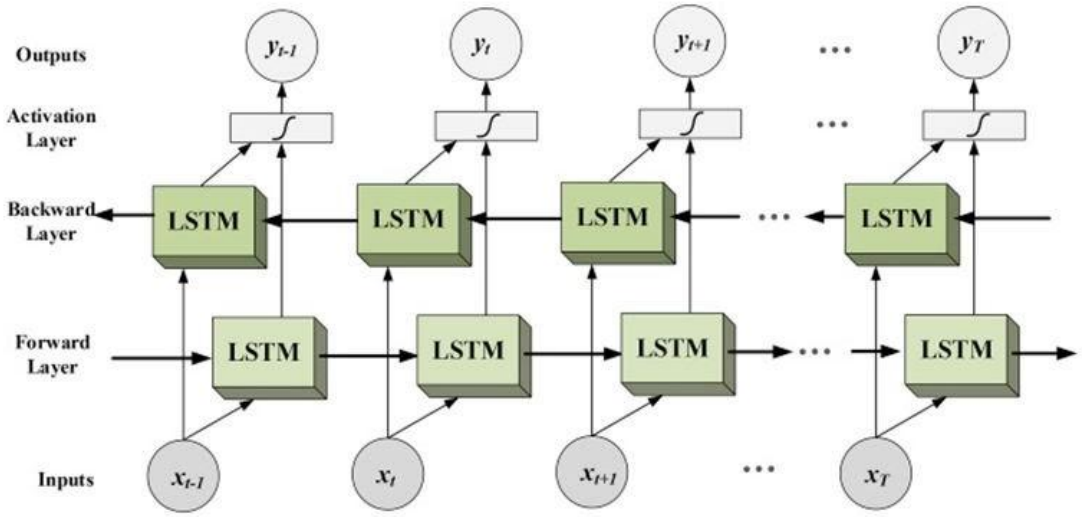


Figure 4: The block diagram of Bi-LSTM

Bi-LSTM network was established by Hochreiter and Schmidedhuber [20]. This network structure overcomes the problems of RNN.

In Bi-LSTM, there are three memory gates such as input, forget and output. The high frequency Intrinsic Mode Function (IMF) is C_{tsh}that is given as an input and h_{i-1} is the output. $CS_{t-1}$is the input of cell state determined by forget gate f_t using a sigmoid function. It is written in Eq. 1.1 [20]:

$$f_t = \sigma(w_f[h_{i-1,}C_{tsh}] + b_f) \quad (1.1)$$

Input gate i_t is used to determine the values that are to be updated to CS_t as in Eq. 1.2 [20]:

$$i_t = \sigma(w_i[h_{i-1,}C_{tsh}] + b_i) \quad (1.2)$$

The output gate o_t is equated in Eq. 1.3 [10]:

$$Cs_t = f_i \otimes CS_{t-1} \oplus i_t \otimes CS_{t-1} \quad (1.3)$$

Consequently, the output of LSTM memory cell is written in Eq. 1.4 & 1.5 [20]:

$$h_N = o_t \Theta Cs_t \quad (1.4)$$

$$y_d = \text{softmax}\ (W_o \cdot h_N + b_o) \quad (1.5)$$

3. Fusion of DTICF-3D-CNN-Bi-LSTM and RGB Channel Assimilation-2D-CNN for HSI Classification

The Fusion of 3D-CNN-Bi-LSTM with 2D-CNN is applied for classification of HSI. The Figure 7 shows the detail structure of proposed work.

2.5 Novel RGB Channel Assimilation

The RGB color space [21-22] is crucial for color object related problem and can also be utilized for color texture extraction. The transformation of RGB image from HSI helps to avoid the redundant information and also reduce the dimensionality problem.This RGB color space is obtained by computing rgb chromaticity value. It helps to improve the classification accuracy than the direct use of R, G, and B. In RGB image, color measures are represented by the science of colorimetric that contains color qualitatively and luminance measures. Color is referred as brightness, hue and saturation. Brightness is denoted as perceived luminance. The hue is referred as amount of red and green of the image. Saturation contains more and more monochromatic light. The RGB image is based on above three components. This component doesn't show for human visual system.

In RGB space, chromaticity coordinates is calculated by chromaticity value. The average value of the RGB color space is obtained by chromaticity coordinates. It has some limitation for derived classes of color representation. It has some crucial problem for selection of color space on object. There are three coordinates such as achromatic value (luminance) and two chromatic coordinates (green-red and yellow-blue). We obtain the chromaticity invariance by using normalization of tristimulus coordinate. By using normalization of tristimulus coordinates, the new tristimulus value r, g, and blue are obtained. In RGB space, the chromaticity value is calculated for chromaticity coordinates. The chromaticity coordinates are the average value of RGB color space.Figure 5 shows the structure of RGB channel Assimillation and summarize the RGB color space in algorithm 1. Figure 6 shows examples of RGB channel.

$$r(R, G, B) = R, \frac{G+R}{2}, \frac{B+R}{2} \quad (1.6)$$

$$g(R,G,B) = \frac{R+G}{2}, G, \frac{B+G}{2} \quad (1.7)$$

$$b(R,G,B) = \frac{G+B}{2}, \frac{R+B}{2}, B \quad (1.8)$$

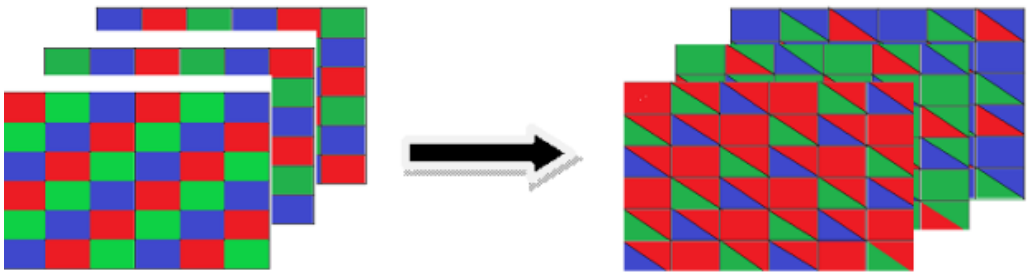


Figure 5: Structure of RGB channel Assimilation

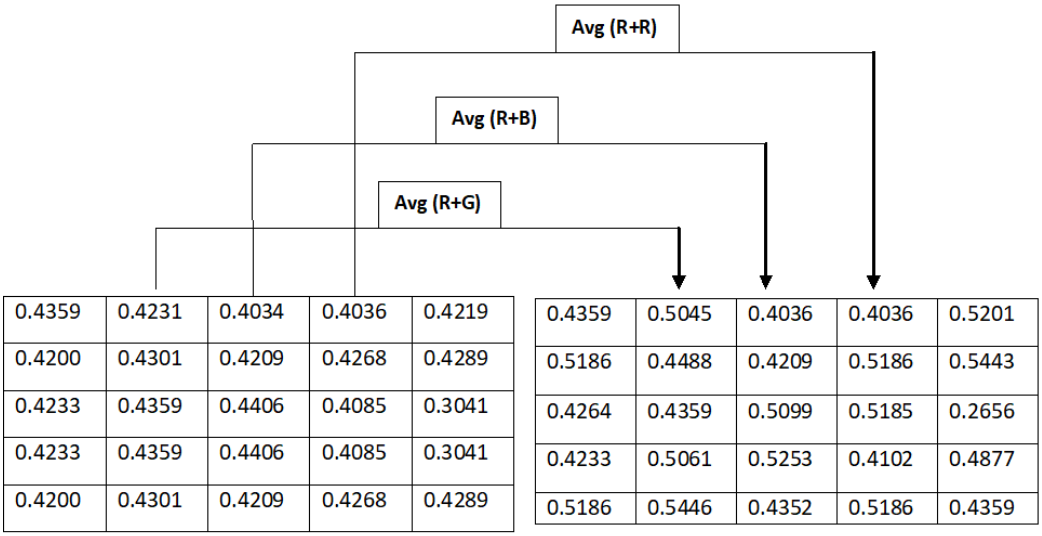


Figure 6: Examples of RGB Channel Assimilation

Input: HSI data $H \in R^{h \times w \times d}$, Number of diagonal value S

Output: RGB channel Patches

1. Set *rvalue* to Red channel value;
2. Set *gvalue* to Green channel value;
3. Set *bvalue* to Blue channel value;
4. For each diagonal value from 1 to 5 % Red channel value
5. Ifdiagonal value is equal to 1
6. Set *a* to rvalue;
7. Set *b* to rvaue;
8. Elseifdiagonal value is equal to 2
9. Set *a* to rvalue;
10. Set *b* to gvalue;
11. Elseifdiagonal value is equal to 3
12. Set *a* to rvalue;
13. Set *b* to bvalue;
14. Elseifdiagonal value is equal to 4
15. Set *a* to rpatch;
16. Set *b* to rpatch;
17. Elseifdiagonal value is equal to 5

18. Set *a* to rpatch;

19. Set *b* to gpatch;

20. End if

21. End for

22. Compute the average of Red Channel Value;

23. Initialize *j* to one;

24. For each diagonal value from -1 to 1

25. Find average of Red channels

26. Increment *j*;

27. End for

28. Repeat step 4 to 26 for Green and Blue Channel vale;

Algorithm 1: RGB Channel Assimilation for HSI Classification

The process of implementation will depict below:

Step1: Normalization

The HSI is normalized as the equation 1.9 [23], where x_{minj} and x_{maxj} are the minimum and maximum value values inx_{ij}, respectively.

$$x_{ij}=\frac{x_{ij}-x_{minj}}{x_{maxj}-x_{minj}} (1.9)$$

Step 2: Remove the dark pixel value

After completed the normalizing process, we remove the dark pixel value for clearing the redundant information. It will help to reduce the burden of high dimensionality of HSI.

Step 3: 3-D Patch Extraction

We extract the 3-D Patches from HSI.3-D patches are $S \times S \times D$ here $S \times S$ are denoted by spatial information and spectral information is denoted by *D [24]*.

Step 4: Color Texture Extraction

For extracting RGB color space features, we obtain the RGB image from HSI patches. RGB color features are excerpted by using chromaticity computation.

Step 5: Classification

After obtaining the RGB color features, these features are classified by 2-Dimensional-Convolutional Neural Network (2D-CNN).

3.1 Domain Transform Interpolated Convolution Filter (DTICF)

Spatial features play a critical role in HSI. Edge preserve filter [25-26] smoothes away noise while retaining sharp edges. The primary advantages of filter are that edges become clearer after filtering rather than becoming blurred. It provides a different weight calculation method for the neighborhood pixels, making the current pixels which are in the same distance have different weights [27-28]. Neighborhood

pixels on the same side ofthe current pixels will have lager weights while neighborhood pixels on the different side of the current pixel will have smaller weights. In general, the HSI retains strong spatial correlation between pixels. The reflection intensity between pixels is represented by spatial correlation. Last few years, many supervised and unsupervised methods were used for spatial and spectral information. However spatial correlation was ignored.

Gastal and Oliveria [29] are proposed a DTICF for image filtering that is edge preserving filtering. It is used to transform 2-dimensional image filtering into one dimensional image filtering. The HSI of DTICF calculates the energy functions.

$$Z_i(u) = \int_{\Omega w} P_w(x) Q(h(u), x) dx$$

$$i = 1,2,\ldots..,n;\ u \in \Omega_w \qquad (1.10)$$

The equations are displayed $P_w$as a reconstructed signal captured by the linear interpolation (inΩ_w). The uninterrupted convolution is conducted by P_w filtering. The normalized box kernel Q is represented by boot function and r is the filter radius.

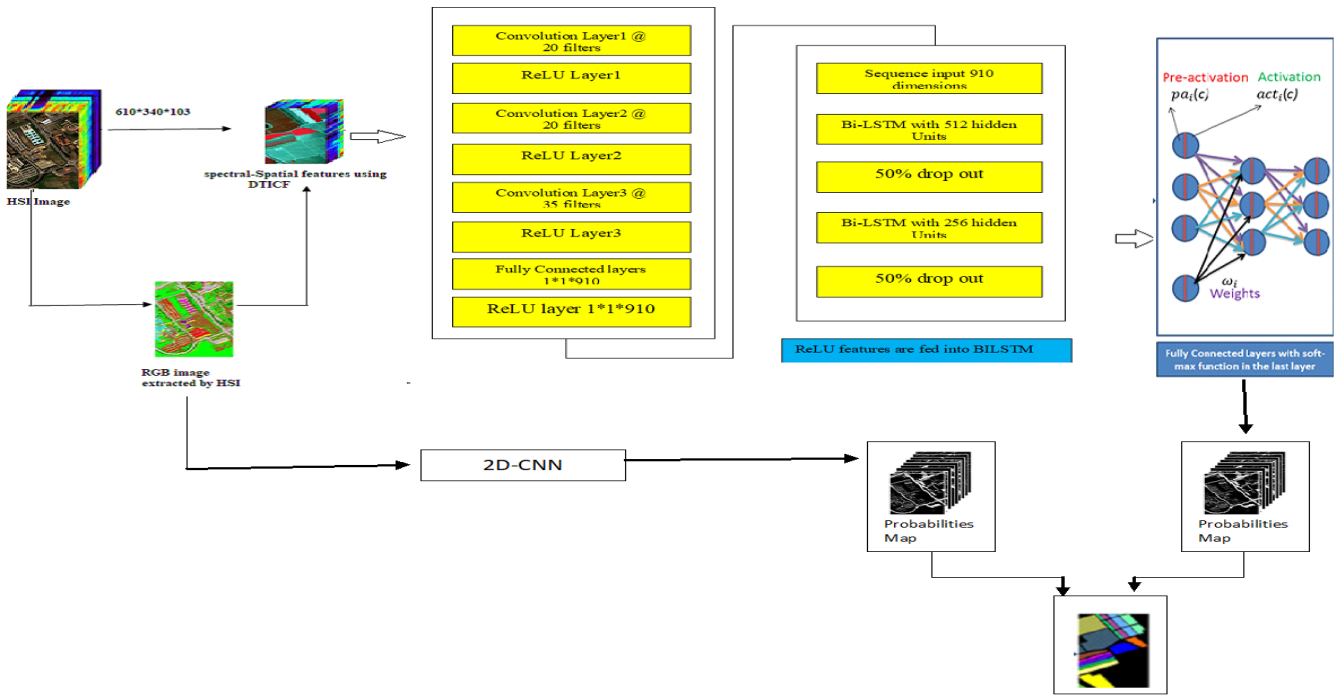


Figure 7: Detail Structure of Proposed Method

$$Q(h(u), x) = \frac{1}{2r} \delta\{|\, g(u) - x \,| \leq r\} \qquad (1.11)$$

$$(E) = \begin{cases} 1 & E \text{ is true} \\ 0 & \text{other} \end{cases} \qquad (1.12)$$

Substituting Equations (1.11) and (1.12) into (1.10), [29]

$$F_i(u) = \frac{1}{2r} \int_{g(u)-r}^{g(u)+r} P_w(x)dx \quad (1.13)$$

$$G(u) = \int_0^u 1 + \frac{\sigma_s}{\sigma_r} \sum_{i=1}^{c} | I_k^{'}(x)dx \quad (1.14)$$

$$\sigma_r = \sqrt{3\sigma_j} \quad (1.15)$$

$$\sigma_{jn} = \sigma_s \sqrt{3} \frac{2^{M-N}}{\sqrt{4^M - 1}} \quad (1.16)$$

By using this filter, the spatial correlation features are extracted. For extracting spatial features, we obtain the RGB image from HSI data. This RGB image and spatial information are applied to DTICF. After that, these spatial features and patch-wise spectral information are fused.

3.2 The network configuration of the 3D-CNN-Bi-LSTM

This fused information is provided to 3D-CNN network. In this network, 3D convolution layer, Max pooling operation and ReLU are performed. The 3D-CNN network contains convolution(c_1, c_2, c_3)and ReLU (R_1, R_2, R_3) layers. There are three types filters are used such ask_1 = 20, k_2= 20 and k_3 = 35 respectively.

The obtained features of the HSI (x_1, y_1,1) are provided as input to 3D-CNN. In 3D-CNN, c_1 with $k_1$filters are utilized in first convolutional layer which data becomes (x_1, y_1,k_1)and (x_2, y_2,k_1). The $c_2$with k_2 filters are utilized in second convolutional layer which data becomes (x_2, y_2,k_2). and (x_3, y_3,k_2). Finally, the third convolution layer btain the data (x_3, y_3,k_3) and ReLU layers. After completing the process of 3D-CNN, ReLU features are provided to Bi-LSTM network. Soft-max function is utilized to obtain the final data. Figure 8 shows the network structure of DTICF-3D-CNN-Bi-LSTM

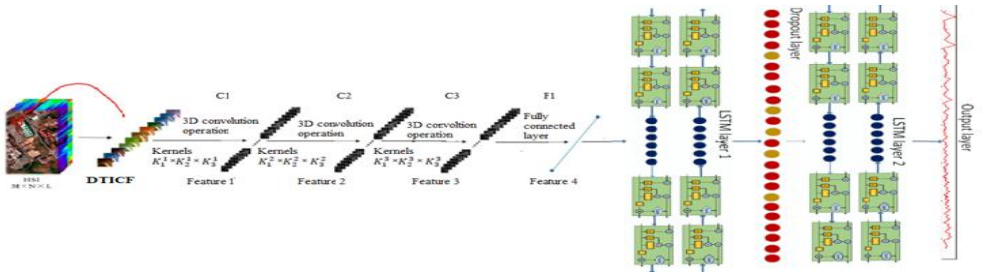


Figure 8: Network Structure of DTICF-3D-CNN-Bi-LSTM

3.3 Fused DTICF-3D-CNN-BILSTM and RGB Channel Assimilation 2D-CNN

The probability map of 3D-CNN-BiLSTM and RGB channel assimilation 2D-CNN are fused. Finally, MRF is utilized for enhancing the fused probabilistic classification map.

In the last stage, MRF is employed to enhance the neighboring pixels. MRF is implemented in probabilistic classification map of the 3D-CNN-Bi-LSTM. These algorithms 2 mainly focused for encouraging the neighboring pixels.

$$\hat{y}=\underset{y\in K^n}{argmax}\left\{\sum_{i=1}^{n}\sum_{k=1}^{K}1\{y_i=k\}log\tilde{y}_{ik}+\mu\sum_{i=1}^{n}\sum_{j\in N(i)}\delta(y_i-y_j)\right\} \quad (1.17)$$

Input: HSI data $H\in R^{h\times w\times d}$, D for Training Data, patches $y = \{y_1, y_2, \ldots, y_n\}$, K for number of labels, T target Labels.
Output: Labels $\hat{y}$.

1. Normalize the input data H;
2. Eliminate the zero values of input data;
3. Set convolution size;
4. For each height of the H
5. For each width of the H
6. If ground truth is greater than 0
7. Extract the patches;
8. End if
9. End for
10. End for
11. Set training size P;
12. Set training data Label Index;
13. Set testing data ;
14. For each class label from 1 to maximum value
15. Find the index of class label;
16. Find the Training range of class;
17. Find the Patches of Training data;
18. Find the Patches of Testing data;
19. End for
20. Compute the padding size;
21. Extract the spatial features by DTICF.
22. // Sigma_s Filter spatial standard deviation;
23. // Sigma_r Filter range standard deviation;
24. // num_iteration Number of iteration to perform;
25. //Joint_image optional image for joint filtering;
26. While Training data i:1 $\rightarrow T$ do
27. Fused the spatial features patches and raw spectral information for training data;
28. Compute another patch for testing;
29. $D_{l^{(k)}}^{(k)}= \{(x_1, y_1), \ldots., (x_{l^{(k)}}, y_{l^{(k)}})\}$;
30. Generate feature maps operation using convolution operation ;
31. f(x) = max (0,x);% ReLU operation;
32. Sequence input by f(x) given Bi-LSTM network using Eq. (1.5);
33. Perform forward and backward sequence to generate the feature;
34. Compute a = $\frac{\exp(o)}{\sum_k \exp(o_k)}$;% Soft-max activation function;
35. end while
36. compute RGB channel assimilation patches (1.6), (1.7), (1.8);
37. Obtain spectral-spatial feature through 2D-CNN;
38. compute probabilistic classification map for 3D-CNN-BiLSTM and

2D-CNN;

39. Compute the classification label $\hat{y}$ using Eq. (1.17).

Algorithm 2: Fusion of 3D-CNN-Bi-LSTM with 2D-CNN of Hyperspectral Image Classification

4 Experiments and Results

All our experiments are implemented by MATLAB R2023. In this portion, the experimental setup is shown and then describes the process of the experiment results. We had done our experiments on two benchmark datasets including Indian pines and pavia University. We measure the classification accuracy using following evaluations metrics such as Overall Accuracy (OA), Average Accuracy (AA) and Kappa Coefficient (KA). OA means the number of pixels classified correctly. AA means the average value of the classification accuracy. KA means statistical measurement of the different class based on confusion matrix [20-21]. In these experiments, we select the sampling proportion for Indian pines data sets and Pavia University data sets. We selected 10% samples for training respectively and we tested basic methods including Support Vector Machine (SVM), subspace multinomial logistic regression with multilevel logistic (MLRsubMLL) and structured sparse logistic regression (SSLR), CNN, CNN-MRF, 3D-CNN and 3D-CNN-MRF respectively. In Table 5.3, 3D-CNN-Bi-LSTM network architecture is explained for our proposed work. The patch-wise input is given to 3D-CNN. During the process of 3D-CNN, 3D-convolution is used to extract the spectral-spatial features concurrently. Padding is also used for predicting the original patch size. In this network, ReLU activation function is utilized for predicting the features. These sequence features are given as input to Bi-LSTM network. In this network, features are captured by forward and backward direction. Dropout layer is also applied to reduce the features size. In Table 1, validation accuracy and loss are also calculated for proposed network.

Table 1: Network structure of the proposed method in Indian pines dataset.

Input Shape	Function	Output shape
3D-CNN	image3dInputLayer([5 5 200 1],"Name","image3dinput")	
(5×5×3)	convolution3dLayer([3 3 3],20, "Name","conv1","Stride",[1 1 1],"Padding",[0 0 1;0 0 1])	(3×3×20)
(3×3×20)	Activation=ReLU	(3×3×20)
(3×3×20)	convolution3dLayer([1 1 3],20, "Name","conv2","Stride",[1 1 2],"Padding",[0 0 1;0 0 1])	(3×3×20)
(3×3×20)	Activation=ReLU	(3×3×20)
(3×3×20)	convolution3dLayer([3 3 3],35, "Name","conv3","Stride",[1 1 1],"Padding",[0 0 1;0 0 1])	(3×3×35)
(3×3×35)	convolution3dLayer([1 1 3],35, "Name","conv4","Stride",[1 1 2],"Padding",[0 0 1;0 0 1])	(1×1×35)
(1×1×35)	Activation=ReLU	(1×1×35)
(1×1×35)	Fully Connected Layer	1×1×910Weights 540×35 Bias 540 ×1
1×1×540	Activation=ReLU 1×1×540	1×1×540
BiLSTM		
540	Dense 1 540 (input layer) Activation=ReLU Sequence input with 540 dimension	540
540	Bi-LSTM 1 with 512 hidden units	1024 Input Weights (4096)
1024	Dropout 50%	1024
1024	Bi-LSTM 2 with 256 hidden units	512 Input Weights (2048)
512	Dropout 50%	512
512	Fully Connected Layer 9	Weight 9×512 Bias 9×1
9	Activation=Softmax	9

Table 2: Validation Accuracy and Validation Loss of 3D-CNN-Bi-LSTM

Epoch	Iteration	Mini-batch Accuracy	Validation Accuracy	Mini-batch Loss	Validation |Loss	Base Learning Rate
1	1	4.63%	36.25%	2.7718	1.0000e-04	1.0000e-04
5	50	36.13%.	----	1.9134	---	1.0000e-04
10	100	52.00%	52.36%	1.2267	1.3463	1.0000e-04
14	150	65.50%	----	0.9721	----	1.0000e-04
19	200	71.75%	64.73%	0.7697	1.2663	1.0000e-04
23	250	81.88%	---	0.5128	---	1.0000e-04
28	300	81.75%	66.01%	0.5128	1.2030	1.0000e-04
32	350	87.75%	---	0.4112	---	1.0000e-04
37	400	89.00%	64.54%	0.3213	1.1812	1.0000e-04
41	450	91.00%	--	0.3213	--	1.0000e-04
46	500	94.46%	66.99%	0.2962	1.1330	1.0000e-04
50	550	98.81%	71.12%	0.2669	1.1106	1.0000e-04

To examine the usefulness of the proposed methods, we compared with SVM, MLRsubMLL, SSLR, CNN, CNN-MRF, 3D-CNN and 3D-CNN-MRF respectively. The first experiments were performed on our proposed method using Indian pine data sets which compares different classification methods [21-22]. The spectral-spatial features are extracted as patch-wise. The classified results are reported in Table 3, where OA, AA and K are displayed. Concentrating on OA value, we conclude that our proposed method able to reach the higher classification accuracy. Another interesting aspect of is that AA, it also shows the higher value in our proposed method.

Table 3: Overall, Average and Individual Class Accuracies (%) and Kappa Statistics of all Competing Methods on the Indian Pines Image Test Set.

Class	SSLR	SVM	MLRsubMLL	CNN	CNN_MRF	3DCNN	IC_3D-CNN_BiLSTM	IC_3DCNN_BiLSTM_MRF	Fusion of 3D-CNN-Bi-LSTM-2D-CNN
1	27.78	46.34	47.12	34.15	31.71	69.69	93.23	91.05	99.98
2	45.36	69.03	86.23	89.57	89.57	89.11	91.20	96.81	100
3	18.07	53.41	55.82	88.62	90.36	91.98	92.95	98.76	100
4	25.93	15.96	13.15	95.12	98.12	97.92	98.68	98.85	100
5	71.76	89.63	93.09	94.01	94.93	96.27	96.20	99.12	100
6	95.55	97.72	99.70	95.59	95.74	94.09	96.18	98.18	100
7	18.18	36.00	36.00	76.00	76.00	79.37	88.52	93.75	100
8	98.95	98.60	100.00	98.84	98.84	98.59	99.18	98.63	100
9	0.00	0.00	0.00	100.00	100.00	99.64	98.38	99.92	100
10	30.63	65.33	81.12	91.99	94.15	93.84	92.09	98.64	100
11	85.03	84.16	95.02	95.07	96.42	89.97	98.39	97.72	100
12	23.84	68.48	90.99	87.43	89.49	91.87	92.20	99.61	100
13	93.90	94.57	98.37	98.37	99.46	98.65	98.84	99.05	100
14	99.01	98.51	99.47	97.98	98.07	98.06	98.31	98.74	100
15	7.47	44.38	55.91	89.91	92.22	95.02	97.04	99.50	100
16	68.92	93.98	97.59	98.80	98.80	99.25	99.12	99.86	100
OA	63.03	77.29	85.93	93.50	94.62	95.24	97.36	98.82	99.84
AA	50.65	66.01	66.70	89.65	90.26	92.71	96.16	98.05	100
KA	56.46	73.79	83.72	92.28	93.40	94.67	96.97	98.61	100

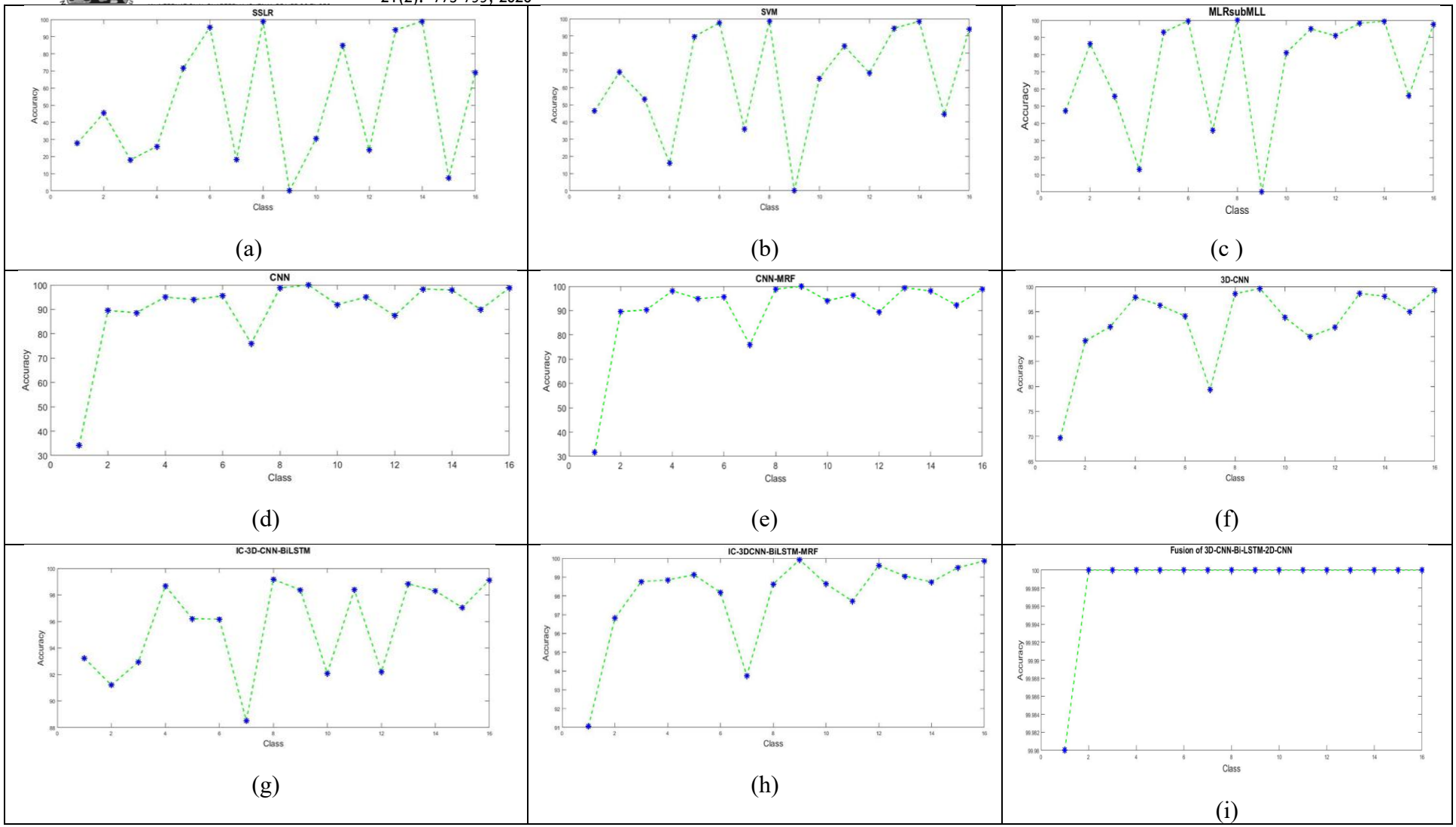

SSLR
Accuracy
Class
(a)
SVM
Accuracy
Class
(b)
MLRsubMLL
Accuracy
Class
(c)
CNN
Accuracy
Class
(d)
CNN-MRF
Accuracy
Class
(e)
3D-CNN
Accuracy
Class
(f)
IC-3D-CNN-BiLSTM
Accuracy
Class
(g)
IC-3DCNN-BiLSTM-MRF
Accuracy
Class
(h)
Fusion of 3D-CNN-Bi-LSTM-2D-CNN
Accuracy
Class
(i)

Figure 9: Classification accuracy for Indian Pines Dataset using (a) SSLR, (b) SVM, (c) MLRsubMLL, (d) CNN, (e) CNN-MRF. (f) 3D-CNN, (g) IC-3D-CNN-Bi-LSTM, (h) IC-3D-CNN-Bi-LSTM-MRF and (i) Fusion of 3D-CNN-Bi-LSTM-2D-CNN.

In Table 3, it is clearly shown that Indian pines data sets are utilized for our proposed method. SSLR and SVM traditional methods got 63.03% and 77.29% respectively. SVM and SSLR obtain the worst classification results. In other existing methods, MLRsubMLL obtains 83.93% classification accuracy. In deep learning methods, CNN and CNN-MRF achieve 93.50% and 94.62% respectively. Compared to CNN, MRF based method provides good classification accuracy. In spectral-spatial based classification methods, 3D-CNN achieves 95.24 % and 96.75 % of the classification accuracy. By using spatial correlation techniques, our proposed method achieves the good classification accuracy. In our proposed method for fusion of 3D-CNN-Bi-LSTM-2D-CNN, the classification accuracy got best performance compared to other machine learning and deep learning methods.

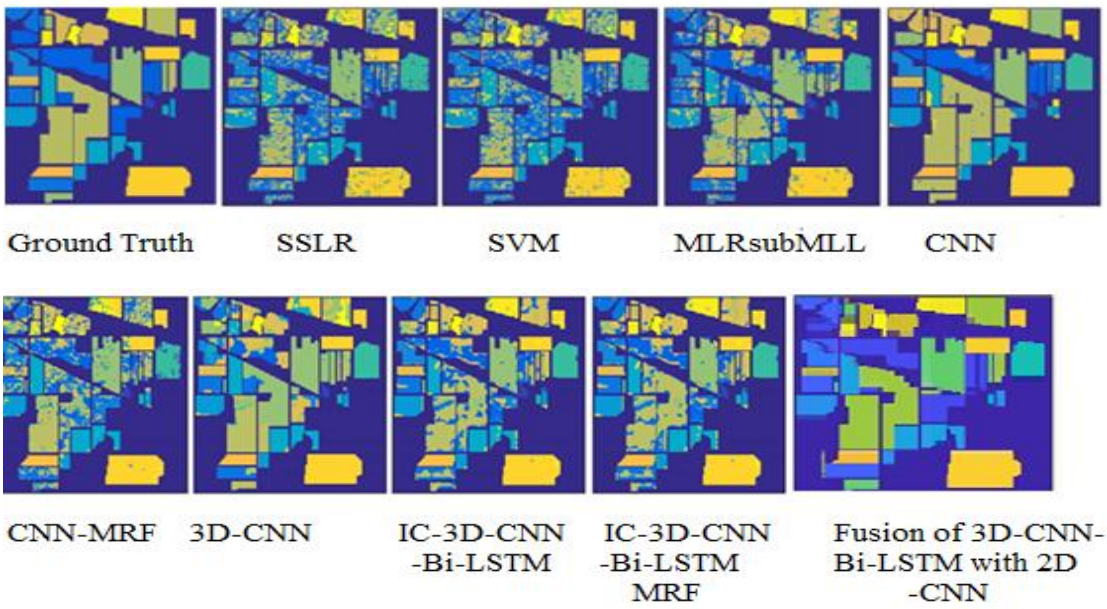


Figure 10: Classification Result by proposed work on the Indian Pines dataset

Figure 9 shows the classification accuracy of Indian Pines dataset using different classification methods for HSI classification. In Figure 10, we conclude that our proposed method provides good classification accuracy compared to other machine learning and deep learning method. SSLR, SVM, and MLRsubMLL obtain worst classification accuracy. Compared to CNN, 3D-CNN classification methods, our proposed Fusion of 3D-CNN-Bi-

LSTM-2D-CNN provides the higher performance for HSI classification. In Figure 11, evaluations metrics are displayed. The classification accuracy of OA and AA is showed. It is clearly shown that, Fusion of 3D-CNN-Bi-LSTM-2D-CNN obtains the good HSI classification.

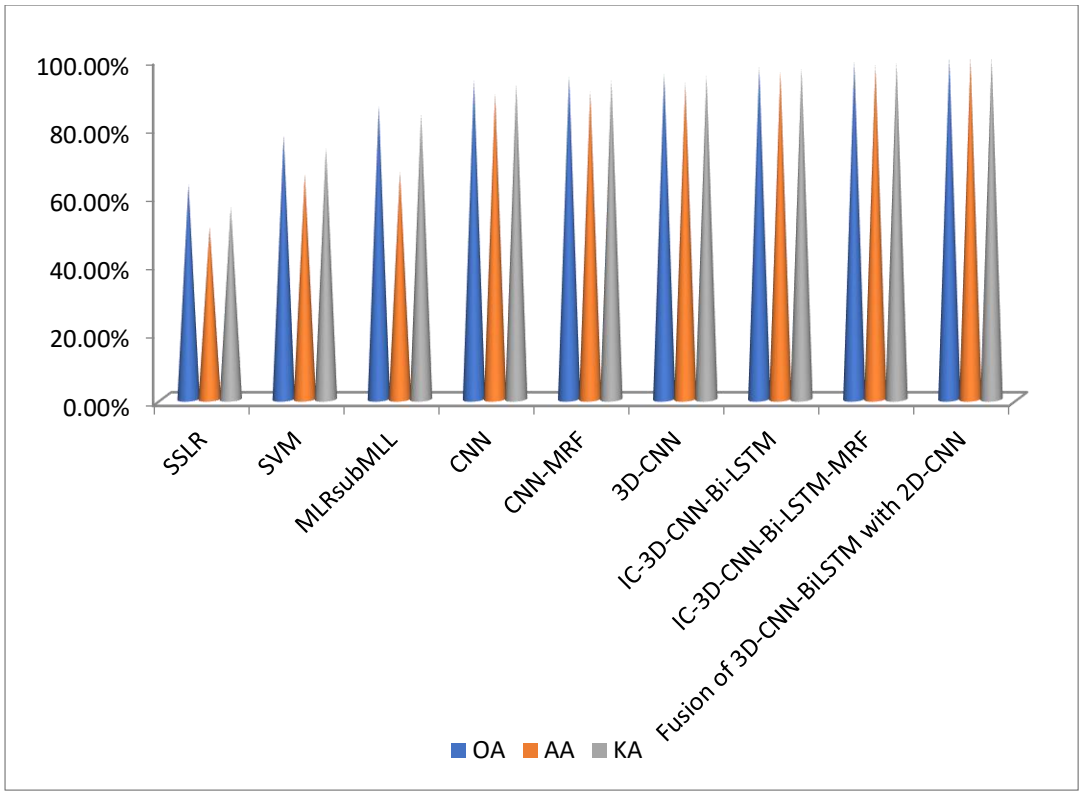
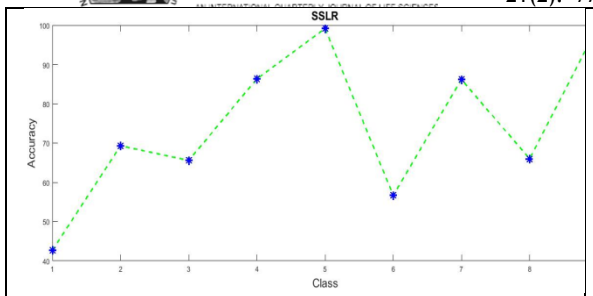


Figure 11: Classification accuracy of proposed method and other classification methods on Indian pines dataset

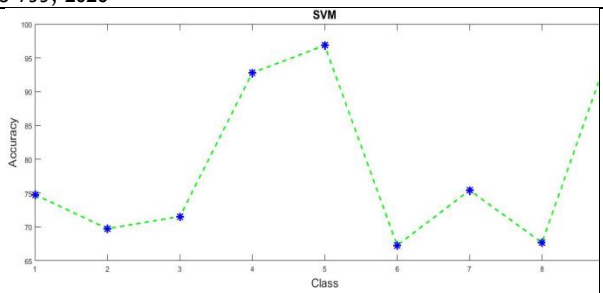
Table 4: Overall, average and individual class accuracies (%) and kappa statistics of all competing and proposed methods on the Pavia university image test set

Class	SSLR	SVM	MLRsubMLL	CNN	CNN_MRF	3DCNN	IC_3DCNN_BiLSTM	IC_3DCNN_BiLSTM_MRF	Fusion of 3D-CNN-Bi-LSTM-2D-CNN
1	42.72	74.78	96.89	96.78	97.92	97.8	98.54	99.29	100
2	69.31	69.75	84.82	96.92	97.38	95.83	97.27	98.30	99.30
3	65.57	71.54	86.69	85.96	87.66	91.92	94.22	95.92	99.73
4	86.34	92.79	94.18	98.78	98.97	97.31	98.46	98.71	100
5	99.16	96.86	97.47	99.92	99.92	98.38	99.04	99.27	100
6	56.66	67.27	96.31	90.10	92.00	95.7	97.17	98.13	99.92
7	86.20	75.43	91.40	84.42	85.27	89.14	91.36	94.81	99.23
8	65.98	67.68	91.93	89.84	91.54	96.31	97.74	99.07	99.89
9	99.67	98.13	99.34	96.80	97.13	99.38	99.49	99.68	100

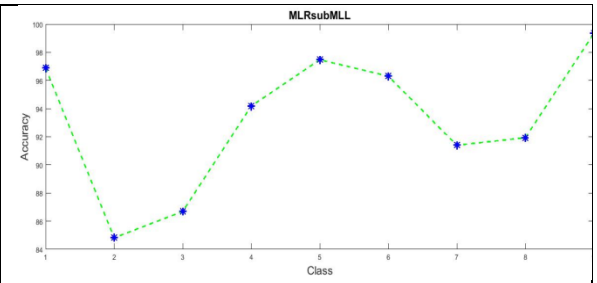
OA	66.52	73.41	90.36	90.50	92.68	94.37	97.93	98.67	99.01
AA	74.62	79.36	93.24	90.44	92.20	95.76	97.03	98.13	99.20
KA	57.75	66.23	87.52	87.46	94.26	93.26	97.33	98.01	99.34


SSLR
Accuracy
Class

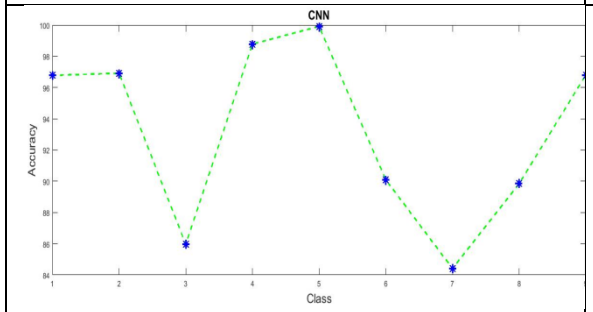
(a)


SVM
Accuracy
Class

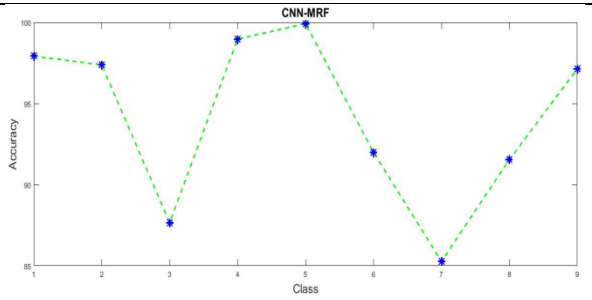
(b)


MLRsubMLL
Accuracy
Class

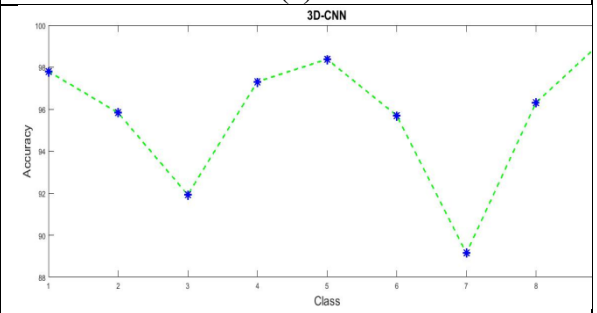
(c)


CNN
Accuracy
Class

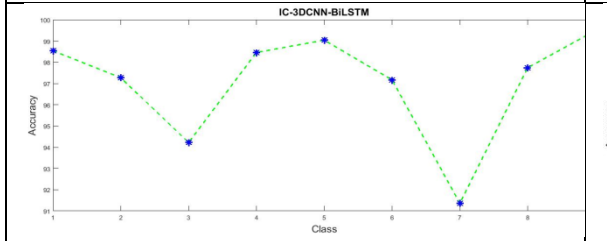
(d)


CNN-MRF
Accuracy
Class

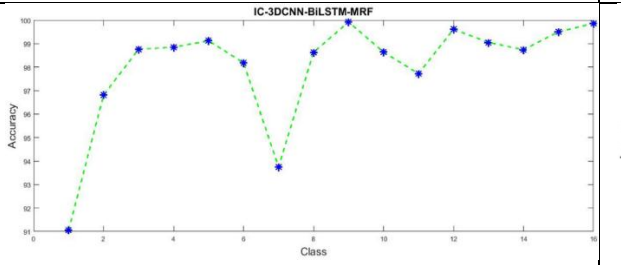
(e)


3D-CNN
Accuracy
Class

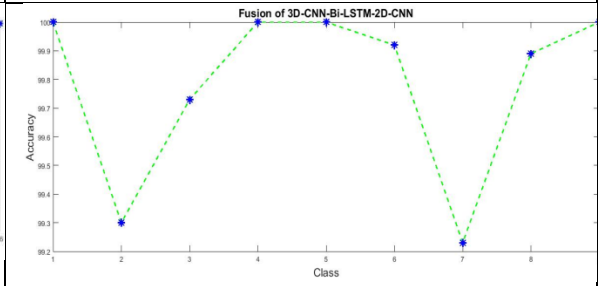
(f)


IC-3DCNN-BiLSTM
Accuracy
Class

(g)


IC-3DCNN-BiLSTM-MRF
Accuracy
Class

(h)


Fusion of 3D-CNN-Bi-LSTM-2D-CNN
Accuracy
Class

(i)

Figure 12: Classification accuracy for Pavia University data set using (a) SSLR, (b) SVM, (c) MLRsubMLL, (d) CNN, (e) CNN-MRF. (f) 3D-CNN, (g) IC-3D-CNN-Bi-LSTM, (h) IC-3D-CNN-Bi-LSTM-MRF and (i) Fusion of 3D-CNN-Bi-LSTM-2D-CNNFigure 5.43: Classification accuracy for the using IC-3D-CNN-Bi-LSTM

In Table 4, it is clearly shown that Pavia university data sets are utilized for our proposed method. SSLR and SVM traditional methods got 66.52% and 73.41% respectively. SVM and SSLR obtain the worst classification results. In other existing methods, MLRsubMLL obtains 90.36% classification accuracy. In deep learning methods, CNN and CNN-MRF achieve 93.50% and 95.68% respectively. Compared to CNN, MRF based method provides good classification accuracy. In spectral-spatial based classification methods, 3D-CNN and 3D-CNN-MRF achieve 94.37% and 97.33% of the classification accuracy. By using spatial correlation techniques, our proposed method achieves the good classification accuracy. In our proposed method for Fusion of 3D-CNN-Bi-LSTM-2D-CNN, the classification accuracy got best performance compared to other machine learning and deep learning methods.

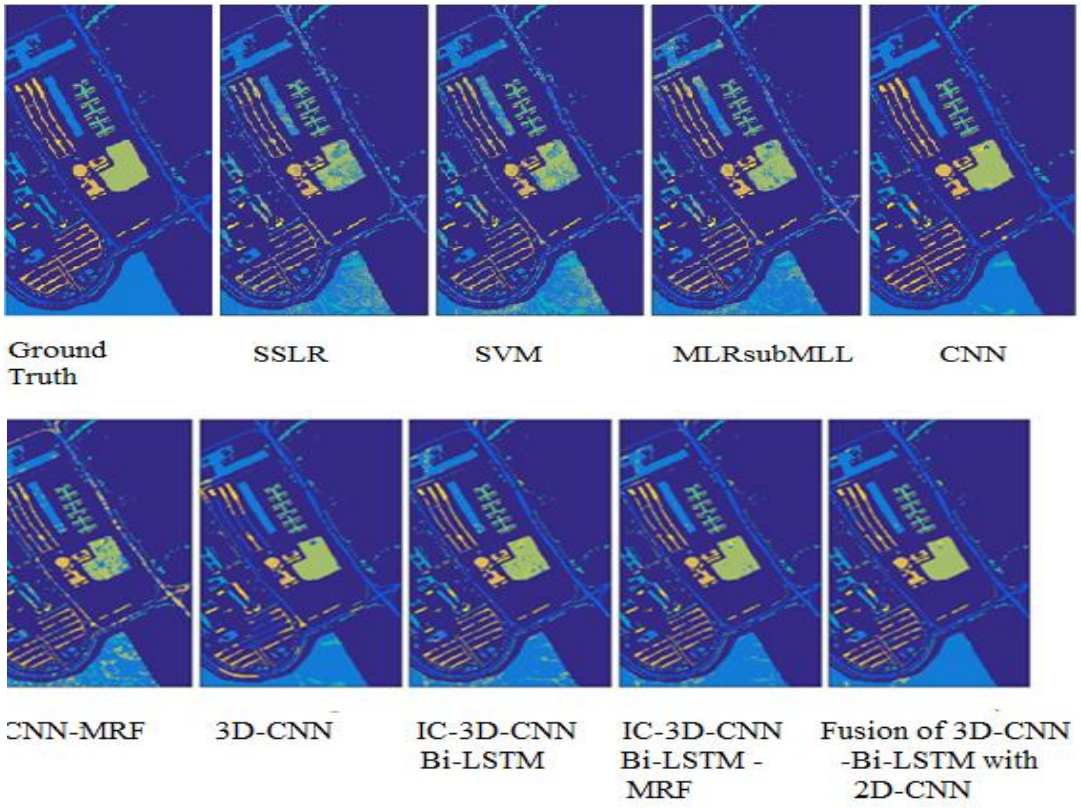


Figure 13: Classification Result by proposed work on the Pavia University dataset

Figure 12 shows the classification accuracy of Pavia University dataset using different classification methods for HSI classification. In Figure 13, we conclude that our proposed method provides good classification accuracy compared to other machine learning and deep learning method. SSLR, SVM, and MLRsubMLL obtain worst classification

accuracy. Compared to CNN, 3D-CNN classification methods, our proposed method provides the higher performance for HSI classification. In Figure 14, evaluations metrics are displayed. The classification accuracy of OA and AA is showed. It is clearly shown that, Fusion of 3D-CNN-Bi-LSTM-2D-CNN obtains the good HSI classification.

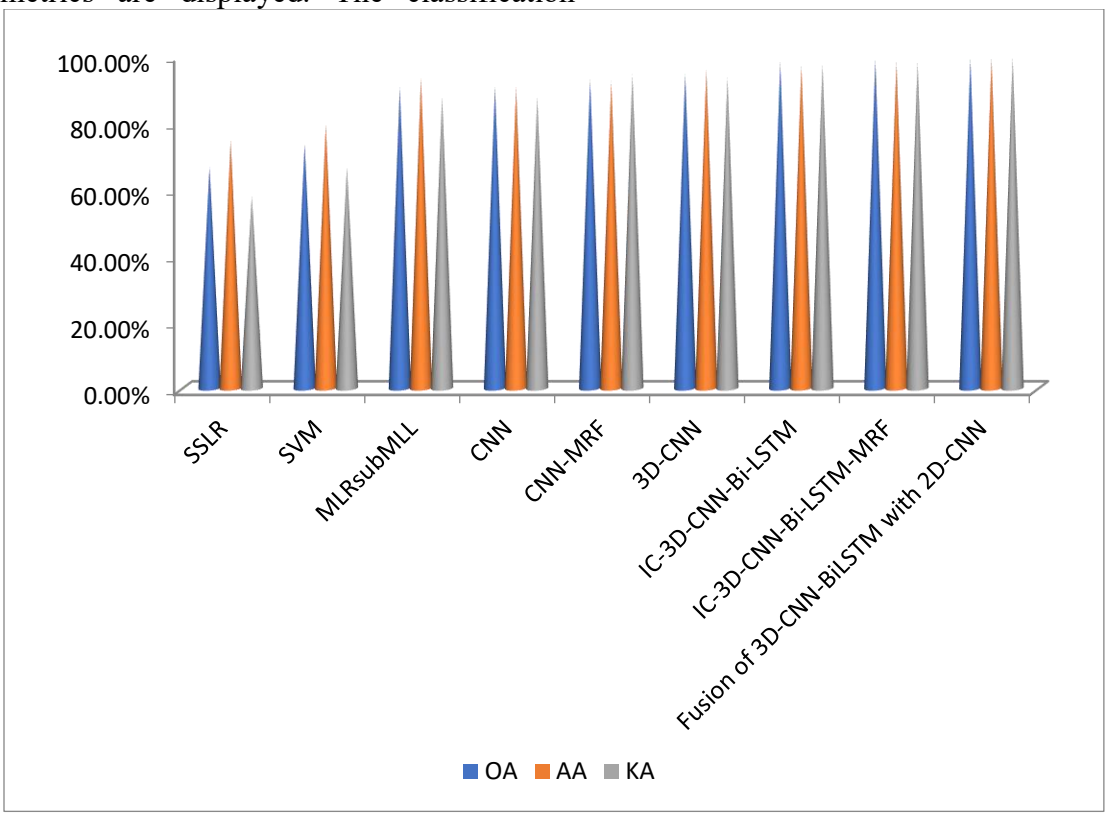


Figure 14: Classification accuracy of proposed method and other classification methods on Indian pines dataset

5. Conclusion

Fusion of 3D-CNN-BiLSTM-2D-CNN framework has been proposed that is utilized to obtain the features of spectral-spatial information. DTICF is applied to HSI to extract the spatial features. These features are provided to the 3D-CNN-BiLSTM.The RGB Assimilation of color features are given to 2D-CNN. The 3D-CNN-BiLSTM and 2D-CNN is fused. Finally, MRF is utilized to improve the features map for smoothing the classification result. Fusion of 3D-CNN-BiLSTM-2D-CNN framework is compared with other HSI classification methods. The experimental result clearly proved that Fusion of 3D-CNN-BiLSTM-2D-CNN based HSI classification obtained the welfare classification accuracy. In future work, we will concentrate on how to reduce computational time across a variety of HSI datasets. For Future enhancement, hybrid classification techniques will be developed for reduce the execution of time. Besides, classification will be carried out by mapping resultant data with ground truth data. Furthermore, the work will enhance this process with most recent optimization techniques to reduce the number of bands for improve the accuracy.

Compliance with Ethical Standards:

Human Ethics and Consent: Not applicable. We don't involve humans and animals for our research.

Funding: Not applicable

Conflict of Interest: We declare that there is no conflict of interest

Data Availability: Data will be made available based on the request

(In case animals were involved) Ethical approval: Not applicable. We don't involve humans and animals for our research.

(And/or in case humans were involved) Ethical approval: Not applicable. We don't involve humans and animals for our research.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

(In case humans are involved) Informed consent: N/A.

Acknowledgement

The authors express their sincere gratitude to the Department of Computer Science, St. Xavier's College (Autonomous), Palayamkottai, Tirunelveli. for the continuous encouragement, research facilities, and technical support.

References

[1] Kaiqiang Zhu, Yushi Chen, and PedramGhamisi,"DeepConvolutional CapsuleNetwork for Hyperspectral Image Spectral and Spectral-Spatial Classification," *RemoteSensing.* 2019.

[2] Wenchao Qi, Xia Zhang, Nan Wang, Mao Zhang and Yi Cen, "A Spectral-Spatial Cascaded 3D Convolutional Neural Network with a Convolutional Long Short-Term Memory Network for Hyperspectral Image Classification," *Remote Sensing*, 2019.

[3] Alberto Signoroni. and MattiaSavardi., "Deep Learning Meets Hyperspectral Image Analysis: A Multidisciplinary Review," *Journal of imaging,* Vol. 52, No. 5, 2019.

[4] Lefei Zhang. and Qian Zhang,. "Simultaneous Spectral Spatial Feature selection and extraction for hyperspectral image, " *IEEE Transactions on Geoscience Remote Sensing,* Vol. 27, No. 6, pp. 2168-2267, 2016.

[5] Li, N., Wang, Z., & Cheikh, F. A. (2024). Discriminating spectral–spatial feature extraction for hyperspectral image classification: A review. *Sensors*, 24(10), 2987. https://doi.org/10.3390/s24102987

[6] Zhang, Y., Li, J., & Plaza, A. (2024). Spectral–spatial hyperspectral image classification using dual-stream convolutional neural networks. *Information Fusion*, 96, 102–115. https://doi.org/10.1016/j.inffus.2024.102115

[7] Luo, F., Du, B., Zhang, L., Zhang, L., & Tao, D. (2022). Feature learning using spatial–spectral hypergraph discriminant analysis for hyperspectral image. *Remote Sensing*, 14(12), 2856. https://doi.org/10.3390/rs14122856

[8] Mou, L., Zhu, X. X., & Benediktsson, J. A. (2021). Hyperspectral image classification using recurrent neural networks. *IEEE Transactions on Geoscience and Remote*

Sensing, 59(1), 36–49. https://doi.org/10.1109/TGRS.2020.2989314

[9] Paoletti, M. E., Haut, J. M., Plaza, J., & Plaza, A. (2022). Deep learning classifiers for hyperspectral imaging: A review. *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5. https://doi.org/10.1109/LGRS.2021.3060862

[10] Fang, L., Li, S., Kang, X., & Benediktsson, J. A. (2023). Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 195, 1–15. https://doi.org/10.1016/j.isprsjprs.2022.11.010

[11] Zhong, Z., Li, J., Luo, Z., & Chapman, M. (2021). Spectral–spatial residual network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 59(1), 1–13. https://doi.org/10.1109/TGRS.2020.2991097

[12]He, M., Li, B., & Chen, H. (2023). Graph-based spectral–spatial classification of hyperspectral images. *Pattern Recognition*, 135, 109108. https://doi.org/10.1016/j.patcog.2022.109108

[13] Zhou, Q., Liu, Y., & Zhang, D. (2024). Fusion of 2D and 3D convolutional neural networks for hyperspectral image classification. *IEEE Access*, 12, 45890–45905. https://doi.org/10.1109/ACCESS.2024.3374218

[14] Tatyana., and Lorenzo., "Classification of Hyperspectral Images With Regularized Linear Discriminant Analysis," *IEEE Transactions on Geoscience Remote Sensing*, Vol. 47, No. 3, pp. 862-873, 2009.

[15] Chen, Y., Lin, Z., and Gu, Y., "Deep learning based classification of hyperspectral data,"*IEEE Journal of Selected Topics inApplied Earth Observations and RemoteSensing*, Vol. 7, No. 6,pp. 2094 - 2107, 2014.

[16] Ji, S., Xu, W., Yang, M., and Yu, K., "3D convolutional neural networks for human action recognition," *IEEE Transaction Pattern Analysis Machine Intelligence,* Vol. 35, No.1, pp. 221-231, 2013.

[17] Chen, Y., Lin, Z., and Gu, Y., "Deep learning based classification of hyperspectral data,"*IEEE Journal of Selected Topics inApplied Earth Observations and RemoteSensing*, Vol. 7, No. 6,pp. 2094 - 2107, 2014.

[18] Hochreiter, S. and Schmidhuber, J., "Long short-term memory," *Neural computation*,Vol. 9, No. 8, pp. 1735-1780, 1997.

[19] Hochreiter, S. and Schmidhuber, J., "Long short-term memory," *Neural computation*,Vol. 9, No. 8, pp. 1735-1780, 1997.

[20] Sugata Banerji., Atreyee Sinha. and Chengjun Liu., "New image descriptors based on color, texture, shape, and wavelets for object and scene image classification,"*Neurocomputing,* Vol. 117, pp. 173-185, 2013.

[21] Constantin Vertan, NozhaBoujemaa, "Color Texture Classification by

Normalized Color Space Representation," *Proceedings 15th International Conference on Pattern Recognition*, Barcelona, Spain,2002.

[22] Mohamed S. Yasein. and Pan Agathokli.,"An Image Normalization Technique based on Geometric Properties of Image Feature Points", *IEEE International Symposium on Signal Processing and Information Technology*, Giza, Egypt, 2007.

[23] Kang, Li, and Benediktsson., "Spectral-spatial hyperspectralimage classification with edge-preserving Filtering," *IEEE Transaction of Geoscience and. Remote Sensing*, Vol. 52, pp. 2666-2677, 2014.

[24] Preethi. M., Velayutham. C. and Arumugaperumal. S., " A novel RGB Channel Assimilation for Hyperspectral Image Classification using 3D-Convolutional Neural Network with Bi-Long Short-Term Memory," Seventh Vol. 70,No. 3, pp. 201-211, 2022.

[25] Xudong Kang. and Shutao., "Spectral-Spatial Hyperspectral Image Classification with Edge-Preserving Filtering," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 552, No. 5, pp. 2666-2677, 2014.

[26] Xudongkang. and puhongduan., " Hyperspectral image visualization with edge- preserving filtering and principal component analysis," *Information fusion*, Vol. 57, pp. 130-143,2020.

[27] Binge Cui. and Xiudan Ma., "Classification of visible and infrared hyperspectral image based on image segmentation and edge preserving filtering," *Infrared Physics and Technology*, Vol. 81, pp. 79-88, 2017.

[28] Oliveira, M.M. and Gastal, E.S.,"Domain transform for edge-aware image and video processing." *ACM Transactions on Graphics (ToG)*, Vol. 30, No. 4, pp. 69. 2011.

[29] Sokolova, M. and Lapalme, G., *"*A systematic analysis of performance measures for classification tasks," *IEEE Transactions on Geoscience Remote Sensing,* Vol. 45, No. 4, pp. 427-437, 2009.

[30] Lee, M. A."Sensitivity of hyperspectral classification algorithms to training sample size," *In Proceeding Workshop Hyperspectral Image Signal Process,* pp. 1-4, 2009.

[31] Bing Tu. "Hyperspectral image classification with multi scale feature extraction," *Remote Sensing*, Vol. 11, pp. 534-550, 2019.

[32] Xu Tang., Qionglin Zhou. and FanboMeng.,"Hyperspectral Image Classification Basedon Spectral Graph and Bidirectional LSTM Network," IEEE International Geoscienceand Remote Sensing Symposium, Brussels, Belgium, 2021.