Received: July 21, 2023. Revised: October 3, 2023.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Spatial-Spectral Sparse Optimized CNN for Hyper Spectral Image Classification

Sneha G Venkateshalu[1*] **Santhosh Laxman Deshpande[1]**

[1]*Visvesvaraya Technological University, India*
* Email: snehavenkateshalu@gmail.com

Abstract: The proposed spatial-spectrum sparse aware CNN (SSCNN) approach with a CNN architecture employing sparse aware fusion/compression technique to reconstruct high-quality HSI images, departing from prior approaches. The developed model comprises three phases: creating the dictionary, including sparse coding into the dictionary, and evaluating the quality of the parametric reconstruction for the fusion of spatial data. From fused HSI colors, multiscale features are extracted in the second phase. For classification, spatial, and spectral optimization used, a soft margin decision boundary method to reduce the misclassification error. The regularisation factor C was added to manage the commutation among margin and misclassification sensitivity. SSCNN network enables the training of high-level semantic features of fused bands with the best multi-categorization accuracy of 99.46% and kappa of 99.37% in addition to localization-preserved information within the 143.97 secs elapsed time. The findings of the proposed are helpful in spatial data analysis development in the study area.

Keywords: Fusion with localization-preservation, Optimization, Dictionary, Multi-classification, Soft-margin boundary.

1. Introduction

Hyperspectral images (HSI), which offer a wealth of spectra and spatial data, are being generated due to the explosive growth of satellite imagery. The primary goal of HSI labeling for the majority of these sorts of applications is to anticipate the label for the class for each pixel in the image. The "curse of dimension" causes the HSI categorization to be distinct from and be more difficult than conventional issues with categorization. The term "curse of dimension" describes the process of organizing and analyzing data in highly dimensional environments that are not possible in low-dimensional environments, such as the multidimensional reality of daily life. In areas like statistical sampling, combinations, and algorithms for learning, data mining, and databases, in addition, physically cursed occurrences can occur. These issues all share the common aspect that as a size rises, spatial capacity also rises. The underlying thread across these issues is that when dimension rises, the size of the space expands so quickly that the amount of information gets sparse. The total quantity of information required to produce a credible result frequently increases exponentially with dimension.

The process of learning image data is challenging and time-consuming. To learn data from visuals at an impressively high rate and with little processing duration, an efficient approach is needed. The technique of creating HSI input patches shortens the overall duration for learning. To obtain fused HSI, an effective sparse aware fusion/compression and restoration method utilising the CNN framework is presented. A dictionary-based learning approach is suggested as a means of assisting the fusion process and resulting in merged HSI. Overlapping patches are optimised via the fusion process. However, for optimal efficiency, this better architecture is required. Our method can provide a great compressed HSI after being trained over a variety of HSI with distortion characteristics included. Utilising a sparse linear restoration technique, HSI are merged and this is possible to lessen the additive white Gaussian noise. Without using any method data, the sparse aware lexicon approach to learning allows for flexible training of sparse aware lexicons. Evaluation

effectiveness outperforms learning effectiveness by combining effective fusion performance and patch-wise categorization. An advantageous and crucial feature of employing dictionaries is their ability to minimise sparsity mistakes due to its computational efficiency. The optimised distribution of probability using exponential is used for eliminating the duplicate characteristics from the associated HSI and fusing. The dimensionality-reduced hyperspectral image's pixels can still be directly examined in terms of reality thanks to the sparse CNN-based method used in this case. In another manner, the reduced image of dimensions pixels will be associated to the context's reflecting characteristics. Additionally, an approach built around sparse-CNN may effectively minimise distortion from hyperspectral imaging.

Principal component analysis (PCA), a dimensionality reduction technique, is adapted to group the initial bands of HSI information within subsets of closely tested neighboring bands but is at the finest optimal for familiar pattern classification problems [2]. After dimensionality contraction, analysis is adopted. A common parametric analysis strategy, usually used after dimensionality reduction, is based on maximum likelihood assessment (MLE) of posterior probabilities [3]. A limitation of methods such as PCA and MLE and their modification is that the class-dependent distribution is Gaussian [4]. Real observations are not Gaussian data but are very diverse in extreme cases. Under such circumstances, PCA may fail as a dimensionality reduction technique.

Decreasing dimensionality methods are often created to decrease the dimensionality of the feature space without sacrificing valuable data. Numerous wavelength bands, often thousands, can be found in HSI data. These bands are generally profoundly related. Dimensionality contraction aims to reduce computational complications [1]. As the statistic of images shared on standard networks such as Facebook increases, multi-label image annotation becomes increasingly essential. The objective is to evolve a method to interpret new images and a few consistent keywords from a reference set. Multi-label image elucidation targets to enroll the relationship among observed features and predefined concepts (labels) [5]. Over the past decades, existing efforts have focused on developing handcrafted visual features to improve the accuracy of multi-label annotation.

In order to maximize the prediction power of nearest neighbor type algorithms, combined learning of the scale which describes the most nearby neighbors is not feasible. Tag propagation concludes tags using weighted combinations of the existence or absence of adjoining tags [6].

The discriminatory analysis process examines one tag at a time for analysis and entirely neglects the rest of the specified tags. As an additional feature, multiple tags can be used concurrently, and the data composed in the multi-label region can be used to advance classification performance. A distance metric is used to compute the gap among precedents in the multi-label space. The technique [8] solves the familiar multi-label classification problem and reduces the effect of noise on classification by integrating information in the multi-label space into a single- or two-view-learned discriminative classifier. A powerful tagging approach must appropriate two strict requirements: Limited labeled training examples and noisy termed training examples. A discriminative model [7] called SpSVM-MC uses both termed and unlabeled data over semi-parametric regularization and incorporates multi-label constraints in the optimization.

It is challenging for classifiers to attain high precision because of the considerable bands in HSI, particularly where there are few training samples. Due to the method's usage of limitless examples of training, low-dimensional representations for hyperspectral imaging are sought after through the process of dimension reduction procedure. The feature extraction and feature selection procedure is part of the dimension reduction approach suggested to deal with the categorization of hyperspectral information. The initial multidimensional spatial information is converted into low-dimensional data by extracting features because data from distinct categories are typically easier to distinguish in the latter space. Channels that contribute the least to the categorization are discarded, while the most informative and productive channels are reserved.

The adopted CNN architecture uses sparsity-aware fusion/compression techniques to fix up high-quality HSI images, and flow of proposed as shown in Fig. 1. Our model abides by a dictionary formation stage, a combination of sparse coding in the dictionary phase, and a parametric restoration quality estimate phase. First, the dictionary-building phase divides the input image into several thin slices. Each image is split into specific contrasting patches. These patches are fed into the model SSCNN. The output is used to create a dictionary during the training phase. In the next stage, sparsity-aware integration and reconstruction techniques are employed in the dictionary learning method, and the produced dictionary is used for examining to improve compression quality. This is how his high-quality HSI is created. In the final stage, the built HSI is inspected for quality.

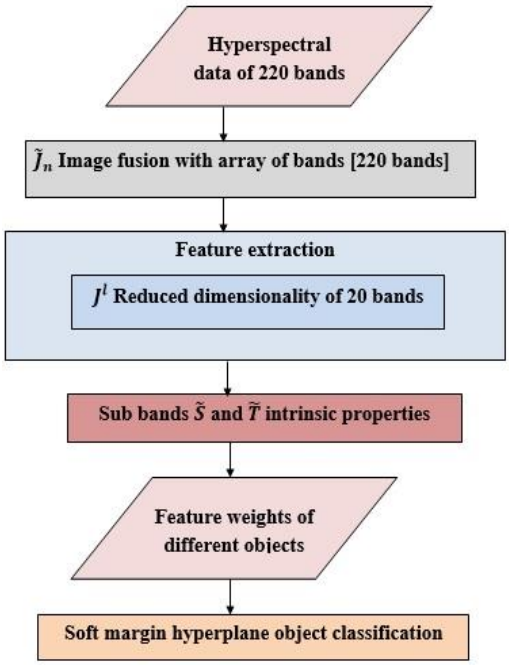

Figure. 1 Flow of proposed multi-label classification

A large extent of real-time HSI info is feasible that needs to be sent to other stations and reconstructed. However, various existing state-of-the-art technologies have a few obstacles, such as immense time dimension, broad power consumption, large HSI datasets, high slew rate, etc., and the essential existing technique are existing technologies are very slow in terms of processing speed. Therefore, in our model, two techniques were adopted to evaluate high reconstruction quality and to maintain high processing speed, CNN and sparse aware fusion/compression and reconstruction technique in integration with dictionary learning methods, to attain enhanced feature fusion.

The present proposed methodology attempted to reduce HSI data misclassification utilizing the targets of an approach as followed:

- Target.1: Spatial hyperspectral data fusion to evaluate high reconstruction quality and to maintain high processing speed.
- Target.2: Feature extraction of fused hyperspectral data sparse aware fusion/compression and reconstruction technique in integration with dictionary learning methods to attain enhanced feature fusion.
- Target.3**:** Multi-label classification of space data with minimized misclassification error introducing soft-margin utilizing slack variable to encounter in making decision boundary.

2. Survey

The morphological patterns (MPs) suggested classifying data from remote sensing more accurately. MPs are created utilizing attributes that hold most of the data's information, including the PCA-derived elements. A new unsupervised approach called unpredictable PCA (NLPCA) auto-associative neural network was developed to condense the semantic depth of hyperspectral data with a small number of elements. Enhanced MPs constructed using NPCA characteristics to achieve precise classification. Linear correlations between spectral bands can be found using linear approaches. However, NLPCA finds combined linear as well as nonlinear relationships. The information content is spread proportionally throughout each element of NLPCA [8]. Principal components are poorly interpreted. Principal components are linear combinations of the raw data's attributes, although these components are more difficult to understand. Since most engineering issues are nonlinear and principal component analysis is a linear method, nonlinearities PCA was implemented. The compromise between minimising dimensionality and information loss. The SSCNN intake HSI data, fuse the object class bands and classify into patch wise without losing its intrinsic properties utilizing optimization feature extraction. Overlapping patches are optimised via the fusion process. Furthermore sparse-CNN can effectively mitigate distortion caused by hyperspectral images.

A statistical approach for discriminating between signals called Independent part discriminant analysis (ICDA) used to satellite imagery categorization is centered on the adaptation of a Bayesian categorization rule to a signal made up of independent components (ICs). The converted elements are independently attributed to the Independent component analysis (ICA) transform array when information is transferred onto a separate space, making computation of the multimodal volume value as the sum of the unitary values simpler. After computing the weighting values for every independent element using a nonparametric kernel density estimator, the categorization allocation was made using the Bayes rule. Multiple hyperspectral pictures were subjected to the ICDA approach evaluation using support vector machines to examine multiple data set parameters, including urban/rural area, amount of the learning set, and kind [9]. It is not explained in Bayesian analysis how to choose a prior. It may result in posterior distributions with strong previous influence. It frequently has a significant computational cost, particularly in models with numerous parameter choices. SVM algorithm-based

approach evaluation is not appropriate for large data sets. When the target classes are overlapping and the data set includes more noise, SVM does not perform very well. The SVM will perform poorly when there are more training data samples than features for each data point. The SSCNN contribute a combo of effective class bands and sub bands fusion and patch-wise categorization reduces overlapping of classes, evaluation effectiveness ranks higher than learning effectiveness. Optimized distribution of probabilities adopted to remove the duplicate characteristics. This lessen the patch variance to prevent over-fitting.

The categorization of hyperspectral satellite images using both the number of input samples and the length of spectral elements is examined using the method of linear discriminant analysis (LDA). The approach presented problems that needed regularisation to be resolved. The regularisation factor is adjusted using the regularised LDA (RLDA). In RLDA, the eigenvalues of the scattered array are greater than zero; values of eigenvalues are regularised as opposed to the scattered array itself. When the regularisation factor goes to 0, standard RLDA involves computing the inverse of the aggregate scattered array (regularised). Deeper neural network models, notably repetitive ones, have triumphed in much computational intelligence and recognize pattern challenges [10, 21, 22, 23]. When the distributions' means are same, linear discriminant analysis (LDA) is unable to identify a new axis that renders the classes linearly separable. The uncertainty included in the prior probability function and the background knowledge cannot be generally represented or handled by the Bayesian technique. The computation of the posterior is a challenge in the use of Bayesian approaches. The SSCNN novelty provide equivalent dictionaries developed at the learning stage and implying soft-margin hyperplane adding slack variable facilitate to identify and classify varied object classes.

A significant obstacle in employing neural networks to classify hyperspectral images along fewer training samples represents difficulty during training. To integrate spatial and spectral information, pixel patching is usually used to train models to aggregate this problem. An artificial neural network (ANN) model is managed by significant loss (ANNC). When trained along spectral information, ANNC provides discriminatory spectral features applicable for consequent analysis functions. A CNN-based dimensional attribute integration (CSFF) algorithm integrates geographical data into the spectral features extracted by ANNC. A CNN-based discriminant model for estimating even if two pixels exist in the same class. During the examination stage, the local structure is predicted and characterized as a fitted convolution kernel that applies a discriminant standard to the pixel combination created by the test pixel and individual of its neighbors. Spectral, spatial features are created by a convolution process among the predicted kernels and the interrelated spectral features in the local region [11, 18, 19, 20]. The position and orientation of an object are not encoded by CNN. Lack of spatial invariance with respect to the input data and large amount of training data is needed for the existing methods. SSCNN address HSI patches conception and built dictionary for bands fused with sub bands overcome spatial invariance for given amount of HSI dataset.

Support vector machine (SVM) classifiers designed to handle the theoretically crucial difficulty of using binary SVMs for several class issues in data with hyperspectral characteristics are used in hyper dimensional characteristic regions. More accurate classification is provided through SVMs, such as the maximum probability and multilayer perceptron neural network classifiers. Without the need to use a feature reduction technique, SVMs will be able to analyse Hyperspectral data through hyper-dynamic feature space. SVMs formulated directly as a multiclass optimization problem, as the number of classes that are to be discriminated simultaneously and the number of parameters to be estimated increases considerably in a multiclass optimization formulation. The SVM technique involves easy modification of its hyperplane variables and the discovery of an ideal hyperplane that permits the greatest separation between the imminent learning collection and an independent hyperplane. The solution of the following convex quadratic programming problem can be determined as an optimal hyperplane. Support vectors refer to the training samples corresponding to no zero weights. The regularisation factor C value is used to determine the penalty ascribed to mistakes; slack variables are added to allow the non-separability of input; the bigger the penalty related to misclassified instances [12, 15, 16, 17]. SVMs can be quite slow and can use up a lot of memory when the dataset has numerous features, making them unsuitable for large datasets with multiple features. Not appropriate for missing value datasets: SVMs cannot accept missing values; they require entire datasets without any gaps. Adopting soft-margin in SVM allows some misclassification with only two classes not more than that. But spatial data contains more than two variety of classes. Considering this case our SSCNN approach deployed to classify different object classes with convolution filters.

With actual information in the structure of labeled data is often scarce, statistical categorization hyperspectral data is large in dimensionality, and reflect multiple categories of data that are occasionally extremely mixed together. The classification algorithms that are produced have a low generalization. Whenever analyzing hyperspectral data, random forest types of filters inside a binary hierarchical multiclassifier framework enhance classifier generalization, particularly if the amount of learning data is constrained. The number of attributes that are chosen at every node of the tree depends on the amount of related data used for learning in the binary hierarchical classifier (BHC), which combines collecting with learning samples. Enormous outcome issues with space are often broken down employing binary trees, which are created employing a number of fragmenting methods utilizing a number of characteristics and output categories. Using spectral patterns that are spatially unstable to reflect random feature subset selection techniques effectiveness concerning generalisation might be useful [13]. Complications with current binary classification is limited to handling two-class problems, whereas a multi-class classifier can theoretically handle any number of classes. This is the major distinction among a binary classifier and a multi-class classifier. Proposed SSCNN implying multiple filters dealt as a multi-class classifier.

The kernels of the methods described for pattern analysis are computationally effective, robust, and reliable. For the classification of remotely sensed data, kernel-based classifiers have been developed and applied. Nonparametric ranked feature extraction (NWFE), which is based on a linear transformation, is employed to derive hyperspectral imagery attributes. The kernel approach is utilized to broaden the nonlinear and linear transformation of the NWFE to the kernel-based NWFE in the case of linear and nonlinear transformation. KNWFE will carry out Generalized Discriminant Analysis, Independent Factor Assessment, Core Based Primary Factor Analysis, and Resolution limits Feature Retrieval. This strategy lessens the impact of the kernel matrix's singularity on the resolution of eigenvalues [14]. The complexity associated with credit allocation path chains of potentially understandable causal relationships among choices and effects—distinguishes deep from advanced learners. Principle component analysis (PCA), deep learning architecture, and logistic regression are all included in the system. Specifically, stacked auto encoders try to obtain relevant high-level features as a deep learning architecture. Confirming the layered auto encoders' suitability for spectrum related to information segmentation [24, 25, 26]. Nonparametric ranked feature extraction approach utilized generalized discriminant analysis fail to discriminant the information with varied in variance of data consuming more training process time. Noise reduction is not possible at every instance point. Due to complexity in the variance dataset, PCA couldn't properly classify. PCA lose the content while dimensionality reduction. SSCNN framework introduced with HSI fusion conception and HSI rebuilt which can't fail losing data during bands fusion with dimensionality reduction.

Initially, only a small amount of features are extracted using factorization in order to do spectrum characteristic preliminary processing. The spatial and spectrum features classes have been combined using a straightforward synthesis. Low-rank second-order pooling is used to attain categorization and improve the precision of the convolution features. The length of LAMFN's runtime is average [27]. For HSI classification reasons, EMFFN sufficiently retrieves multiscale information through simultaneous multipath of three steps. In the EMFFN, there are two subnetworks that are for multiscale spectral and spatial data. Spectrum the cascaded dilated convolutional network (CDCN) was created to derive multiscale characteristics and acquire a wider suitable field with long-range input. It is suggested to use a parallel multipath network (PMN) for collecting large-, middle-, and smaller-scale spatial variables throughout every phase. In consecutive feature fusion, layers of features are combined, and shallower feature maps function faster in training [28]. The limits of current HSI characteristics (spectral and spatial) obtaining and using approaches that utilize CNN. The far-off inter-band associations are frequently overlooked, considering spectral properties. In order to allow the use of convolutional kernels using constrained responsive fields to extract the non-local spectral inter-band associations, a unique spectral bands non-localization (SBNL) technique is presented. The derived spatially multiscale characteristics for spatial features are typically segregated in various channels. To make use of the cross-relationships that exist between multiscale characteristics, there is a unique multiscale-share inception block (MSIB). Adaptive feature fusion (AFF), simple to use component for HSI categorization, has been added [29]. The second order pooling takes larger memory to find single desired solution. Dilated convolutional network reduces the spatial dimensions resolution at the output feature map. Adaptive feature fusion causes noise for the high dimensions. SSCNN employs max pooling provide

solution for the multi-label output inclusive of eliminating noise for the high dimensional spatial data.

3. Proposed approach spatial-spectral optimization approach: sparse aware convolution neural network for hyperspectral data fusion

The understood vision parameters are the basis of neural network operation. A combination of its superior fusion and restoration capabilities, neural networks made up of convolutions have attracted the attention of several academics in the last few years. Relatively the assistance of its genuine feed-forward approaches and effective learning, it handles enormous computing and completes a variety of jobs with ease. CNN focuses mostly on the processing of image data. The CNN design is arranged in a hierarchal manner, and deep training CNN algorithms, in particular, use a variety of layers that are concealed; with the use of dictionary strategies for learning, a unique Convolution Neural Network was employed, utilizing sparse operations to produce outstanding integrated HSI. Independent of the component kind, the CNN framework may combine HSI depending on past information and effective instruction.

A. Image reconstruction

The data that makes up the HSI is separated into a number of regions and HSI classification followed as shown in Fig. 2. The phase of testing to fuse the information from the HSI attribute; this is, and finally, HSI is rebuilt. According to the image restoration block structure, it comprises two phases: the training phase as well as the testing phase.

B. Training phase

The most significant phase of the image data analysis is learning. Learning determines the procedure for controlling the image data analysis process. An effective learning program can continually guarantee any technique's outstanding effectiveness. However, learning image data is a lengthy and sophisticated procedure requiring much time. As a result, an effective method is required to learn data from images at a phenomenally high pace and requires minimal processing time. In learning, samples are continually taken, and variables are chosen randomly. However, an improved design is necessary for optimal performance. A powerful sparse aware fusion/compression and restoration approach utilizing the CNN framework is introduced to get fused HSI.

Here, we employed parallel GPU analyzing with the CUDA programming architecture to learn huge HSI significantly quicker. The HSI kind and regions are used to estimate these parameters. Longer training times are required for data that is highly dimensional.

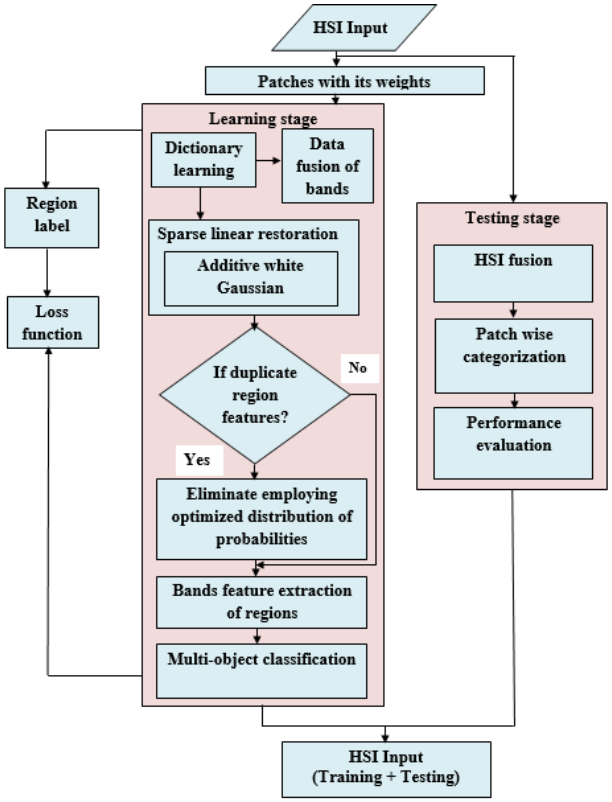

Figure. 2 HSI classification using SSCNN

Initial training may be used to accomplish regularisation, and tiny normative vectors are simpler to alter. Hyperspectral images were used for learning our algorithm. Our algorithm can be trained on various HSI with distortion characteristics present and deliver an excellent compressed HSI. The variation among each layer was utilized to derive strengths for the learning stage of the SSCNN model, and the resulting weights were subsequently fed on through the learning stage, where a dictionary was built to aid in attaining high-quality fusion of the input parameters HSI with regions. We can calculate an efficient learning set utilizing massive data sets via this fashion, and our findings show that our learning method was effective because it outperformed all current conventional HSI fusion techniques when performance measures, including PSNR, MSE, and SSIM, were taken into account.

C. Dictionary based learning

In order to help the fusion process, produce merged HSI, a dictionary method of learning is proposed. Suppose attribute k and incoming picture patch data i are combined.

$$\mathbb{Y} = i + \mathrm{k} \tag{1}$$

Patch-wise fusion of features is carried out in this specific instance. Every patch's size is calculated as m

× m, with k typically falling among 8 and 12. A rebuilt HSI is built by merging all of the patches into within the structure after they have all been separately fused. The fusion procedure is used to optimise overlapping patches. Our dictionary is dependent on the dictionary $\mathbb{D}$, with dimension is estimated to be $m^2 \times n$, when $n > m^2$.

D. Sparse aware dictionary learning methodology

A component that is feature set is offered as a start-up procedure in a dictionary approach to learning. All regions taken from the HSI are fused using a sparse linear restoration approach. It is possible to lessen the AWG (additive white Gaussian noise) noisy mistake by,

$$\min\|\alpha\|\ l_0 \text{ s.t } \|\mathbb{D}_\alpha - \mathbb{Y}\|_2 \leq \in \tag{2}$$

When a preset parameter is given $\in$ and l_0 pseudo normalisation may be calculated applying$\|\cdot\|$ and $\mathbb{Y}$, is specified as a distorted picture patch. With the aid of the l_0 pseudo-normalizing variable, sparsity features are enforced. However, the l_0 norm is unable to offer a comprehensive solution to the convex normalisation issue. The parameter $\in$ is selected with the knowledge knowing the deviation of the approximate norm deviation $\|\mathbb{D}_\alpha - \mathbb{Y}\|_2$ may be used to describe it. The sliding window approach may be used within to accomplish quality fusion, and the cumulative approach can be used to manage conflicting regions.

The highest restoration quality picture is evaluated using the dictionary understanding method. Consequently, the likely options are to: Use a vocabulary that has been specifically constructed; or, learn an overall, adjustable lexicon for the absence of noise HSI collection while taking diverse space phenomena into account. We have shown the lexicon system for learning produces superior outcomes based on findings of the trial. The finest results are seen where images with noise learn to adjust. In contrast to interclass and discord dictionaries, which employ visual regions, multi-class dictionaries are capable of incorporating a variety of discrete sparse characteristics.

The sparse optimization challenge might be minimized by lexicon systems of learning. considering both approaches, the efficiency of acquiring dictionaries may be improved by employing picture regions. Dictionary acquisition and patch categorization methods are lightning-fast and only need 15% of the learning time compared to fuse. The sparse aware lexicon approach to learning may flexibly train sparse aware lexicons without needing any methods data. Here, the first lexicon is created during the learning process. Following that, enabling high-quality reconstruction of HIS, relevant learnt characteristics (i.e., dictionaries) are put into effect.

E. Testing phase

The following part provides an overview of our evaluation capabilities. A combination of effective fusion performance and patch-wise categorization, evaluation effectiveness ranks higher than learning effectiveness. It is possible to lessen the patch variance to prevent over-fitting. Equivalent dictionaries developed throughout the learning stage are made available during evaluation to minimise an attribute with fewer distortion.

$$\hat{l} = \bar{l} + \frac{(\mathbb{F}\{(\mathbb{Y}-\bar{\mathbb{Y}})^2\} - \sigma^2)\ (\mathbb{Y}-\bar{l})}{\mathbb{F}\{(\mathbb{Y}-\bar{\mathbb{Y}})^2\}} \tag{3}$$

where, σ depicts noise level and equation (3) defines fused image mathematical representation. It involves computation of $\bar{l}$and $\mathbb{F}\{(\mathbb{Y}-\bar{\mathbb{Y}})^2\}$. By using dictionary training methods to filter the mixture distorted images while taking into account patch length of m k, it is possible to compute i efficiently. The advantage and key aspect of using dictionaries is their capacity to reduce sparsity errors with their computational efficacy. The duplicate characteristics within the corresponding HSI and fusing are removed using the optimized distribution of probabilities performs employing exponential,

$$\mathrm{i}^{(t+1)} = \mathrm{i}^{(t)} + \eta\ [\sum_{r=1}^{K} \quad N_r^- * \psi_r\ (N_r * i^{(t)}) + \frac{\lambda}{\sigma^2}(\mathbb{Y} - \mathrm{i}^{(t)})] \tag{4}$$

Where N_r^- denotes the image's centre pixel, the convergence is determined by the $*$ sign, and the step size is determined by the η sign. Consequently assumptions are used to create an acceptable probability λ, the trade-offs noise level σ affects how things turn towards the end λ.

4. Stages of spatial statistical analysis

A. Target 1: Spatial hyperspectral data fusion

O is an array of the pixels and C is the original image's depth, and they are separated with N categories of identical spectrum lengths. The $J = (J_1, \ldots, J_O) \in \mathcal{S}^{C \times O}$ is known as the initial hyperspectral image. Every category array of bands is indicated by $C_1, C_2, \ldots, C_N$. Each group of images is subjected to the sparse-CNN based image fusion, with the fused bands obtained are then processed deeper:

$$\tilde{J}_n = \frac{\sum_{o=1}^{C_n} J_n^o}{C_n} \tag{5}$$

Where C_n is the overall amount of bands in the m_{th} category, $\tilde{J}_n$ is the *n*th fused band, n is the n_{th} category, and J_n^o is the oth band in the *n*th category of the raw hyperspectral data. The sparse CNN-based approach in this case is capable of making ensure that the pixels of the dimensionality-reduced hyperspectral image can still be directly examine in terms of reality as indicated by the Einstein's berg colour notion. In another way, the reflecting properties of the context will be connected to the reduced image of dimensional pixels. Additionally, an approach built around sparse-CNN may effectively minimise distortion from hyperspectral imaging.

B. Target 2: Fused hyperspectral data feature extraction

Following are the divisions of the reduced dimensions of HSI($\tilde{J}$) among multiple subgroups of neighboring bands:

$$\hat{J}^l = \begin{cases} \left(\tilde{J}_{(l-1)A+1,\ldots},\tilde{J}_{(l-1)A+A}\right), & l = 1,2,\ldots,\left\lfloor\frac{N}{A}\right\rfloor \\ \left(\tilde{J}_{N-A+1,\ldots},\tilde{J}_N\right), & l = \left\lceil\frac{N}{A}\right\rceil \neq \left\lfloor\frac{N}{A}\right\rfloor \end{cases} \quad (6)$$

Where $\hat{J}^l$ denotes the *l*th subgroup, $\left\lfloor\frac{N}{A}\right\rfloor$ denotes either the highest or lowest integer that is not bigger and smaller than $\frac{N}{A}$. The number of bands in every subgroup is indicated by the A.

The coefficient of reflection and shading elements are obtained for every subcategory $\hat{J}^l$ using optimization intrinsic feature extraction established with sparse CNN, and the solution is as outlined below:

$$\left(S^{l^*}, T^{l^*}\right) = \arg\min_{S^l,T^l} E\left(\hat{J}^l, \hat{S}^l, \hat{T}^l\right) \quad (7)$$

Where S^l and T^l symbolize for the *l*th subcategory coefficient of reflection and shading elements, correspondingly. To derive the space object inherent attributes, consisting are an *n*-dimensional attribute matrix $\tilde{S}$ the following: may be employed in further analysis, the reflectance elements of the various subcategory are merged.

$$\tilde{S} = \begin{pmatrix} S^1 \\ \cdots \\ S^l \\ \cdots \\ S^{\left\lceil\frac{N}{A}\right\rceil} \end{pmatrix} \in \mathcal{S}^{N\times O} \quad (8)$$

and shading module is computed as follows

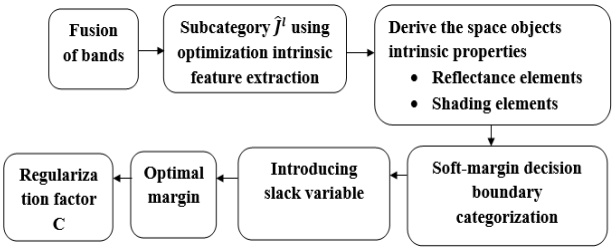

Figure. 3 SSCNN approach statistical method

$$\tilde{T} = \frac{1}{J^l}\tilde{S} \quad (9)$$

C. Target 3: Multi-label classification of space data

Using Eq. (8) and (9) each objects intrinsic properties i.e., shading and reflectance properties is extracted and it is trained with sparse CNN. As the space object classification is multi-label classification problem.

Further, some sample have higher sample and some have very less sample leading to class imbalance problem. In order to accomplish categorization, introducing a soft-margin decision boundary made for separability of data. The CNN hyperplane linear approach is outlined as

$$y = w^T\phi(X) + b \quad (10)$$

where modified feature space $\phi(X)$ is used. The minimum distance between the chosen hyperplane to the nearest point within the collected data constitutes the boundary. To build a choice margin hyperplane within the CNN issue to maximise the data collection range. The distance between the objective tag for every data point is t_i where $t \in \{1,-1\}$. Since the challenge in the present instance wasn't one that can be solved in a linear fashion so the soft-margin CNN is chosen and add the slack variable $\varphi_i \geq 0$. φ enables the categorisation of pattern incorrectly. The data unit is incorrectly categorised when $\varphi_i > 1$. The following unfairness restriction is necessary at the same time to precise as accordance:

$$t_i\left(W_i^T\phi_i(X) + b\right) \geq \quad 1 - \varphi_i \quad (11)$$

The expected objective of our algorithm has to produce $y(x_n) > 0$ for regions with $t_n = 1$ and $y(x_n) < 0$ for points with $t_n = 1$. Formulation gives a measurement in hard margin between the choice border and data point x_n utilizing the equation:

$$\frac{t_n y(x_n)}{\|W\|} = \frac{t_i\left(W^T\phi(X)+b\right)}{\|W\|} \quad (12)$$

Table 1. Total number of training and testing samples in Indian pines dataset

Number	Classes	Total Samples
1	Alfalfa	46
2	Buildings Grass Trees Drives	386
3	Corn notill	1428
4	Corn mintill	830
5	Corn	237
6	Grass pasture	483
7	Grass trees	730
8	Grass pasture moved	28
9	Hay windrowed	478
10	Oats	20
11	Soybean notill	972
12	Soybean mintill	2455
13	Soybean clean	593
14	Stone Steel Towers	93
15	wheat	205
16	Woods	1267

The optimal margin is determined resolving a following issue

$$\underset{W,b}{\operatorname{argmax}} \frac{1}{\|W\|} \min_{n}(t_n(W^T\phi(X)+b)) \qquad (13)$$

The CNN issue ultimately changes with reduction of an issue and adding soft-margin slack parameter

$$\underset{W,b,\varphi}{\operatorname{argmax}} \frac{1}{2}\|W\|^2 + C\sum_{n=1}^{N}\varphi_n \, s.tt_n(W^T\phi(X)+ b) \geq 1-\varphi_{n;}\varphi_n \geq 0;\ n=1,2,\dots,N \qquad (14)$$

The regularisation factor C regulates the trade-off among margins with misclassification tolerance. Additionally, because the information being collected may be separated nonlinearly, utilizing the kernel optimization method to turn it into an additional straightforward dimensionality. These are able to derive an overall optimal from the model considering the CNN issue is a convex optimization issue. The optimal parameters considered utilize the optimal decision boundary to divide the dataset among various labels. When compared to the current hyperspectral spatial object characterization model, which is demonstrated through experiments that follow, the novel approach achieves better precision.

5. Experimental outcomes and analysis

A. Dataset description

An AVIRIS camera that was placed above Indiana's northern-western border obtained data needed for the indian pines dataset. Assuming a wavelength of $0.4-2.5\times10^{-6}$ meters with 224 bands and 145×145 pixels provides for the collection of hyperspectral data. The adoption of IP is justified by the fact that approximately two thirds of the region surveyed is agricultural in nature, while the other one-third of the observed regions are forests and other types of existing vegetation. Additionally, IP data includes two-lane highways, residences, low-lying structures, including tiny roadways. Some are other crops that are still in the initial phases of development, accounting for less than 5% of the total data gathered in IP. According to Table 1, the ground truth data is made up of a total of 16 crops (also known as labels).

B. Experimental setup, design and results

The size of the convolution kernel is 128×1×40 ×40 for the first and last layers. The size of other Conv kernel is 128×64×40×40. The learning rate is set to 0.1, epoch size is set to 10, and band fusion is set to 20. While using spectral-spatial data, SSCNN substantially outperforms LMAFN, EMFFN, and LSGSAN. The suggested SSCNN's OA, Kappa coefficient, and effective multi-label categorization system significantly show enhancement in the Indian Pines data set than EMFFN. Most of these small-scale categories are found in the indian pines data set, causing it difficult for EMFFN to discriminate these in particular when there are not many learning pixels available. Therefore, regardless of whether there are few learning pixels, the suggested SSCNN can nevertheless generate superior results than existing cutting-edge techniques.

Time cost is another crucial consideration for the categorization task operation, particularly in real-world purposes, along with exactness and categorization patterns. Table 2 show that the suggested SSCNN's learning time for learning data sets and testing time for testing data sets are substantially shorter compared to those of the LMAFN, EMFFN, and LSGSAN. However, it is difficult to ensure the precision and outcomes associated with these approaches categorization patterns.

The suggested SSCNN, in comparison, strikes a superior balance between effectiveness, categorization patterns, and accuracy. The proposed SSCNN minimized the misclassification error with better accurate results as shown in Fig. 3 (c) with the ground truth classification maps of Indian pines data for LMAFN (d), EMFFN (e), and LSGSAN (f). Introducing of soft margin with slack variable maximizing the decision boundary, so that the classification errors are minimized with accurate classification as in Fig. 3 (c).

Table 2. Comparative analysis w.r.to fusion and categorization for Indian pines dataset category of LMAFN, EMFFN, LSGSAN with novelty SSCNN

Features	Model			
	Low-Rank Constrained Attention-Enhanced Multiple Feature Fusion Network [LAMFN [27]]	Enhanced Multiscale Feature Fusion Network, (CDCN) and (PMN)[28]	Spectral band non-localization (SBNL), multiscale-share inception block (MSIB) and plug-and-play adaptive feature fusion (AFF)[29]	Proposed SSCNN Spatial-Spectral Optimization
Dataset pixel-size	145*145	145*145	145*145	145*145
Kernel-size	3*3*3	7*7	7*7	128*1*40*40
Number of bands	200	200	200	224
Bands rank	16	20	10	20
Method/Algorithm	Attention-enhanced spatial–spectral feature learning	Adaptive feature fusion with dilated convolution	Spectral band non-localization	Spectral gradient descent optimization
Spectral feature pre-processing	Factor analysis	Feature fusion	Adaptive feature fusion	Patches band fusio n
Pooling	Second-order	Max	Max	Max
Activation function	Relu	Relu	Relu	Leaky-relu
Loss function	Sigmoid	Sigmoid	Sigmoid	Soft-max log
Number of classes	16	16	16	16
Learning rate	0.001	0.001	0.001	0.1
Training epochs	100	200	120	10
Enhancement feature	Cascaded feature fusion with second order pooling	Sparse constraints with discriminative spectral–spatial features	Adaptive feature fusion with residual connection	Dictionary learning
Process learning time(s)	29.1	245.92	106.743	45.209
Elapsed time(s)	103.9	33.73	105.56	143.97
Overall accuracy (%)	78.15	98.85	98.11	99.46
Kappa*100	75.41	98.36	97.84	99.37

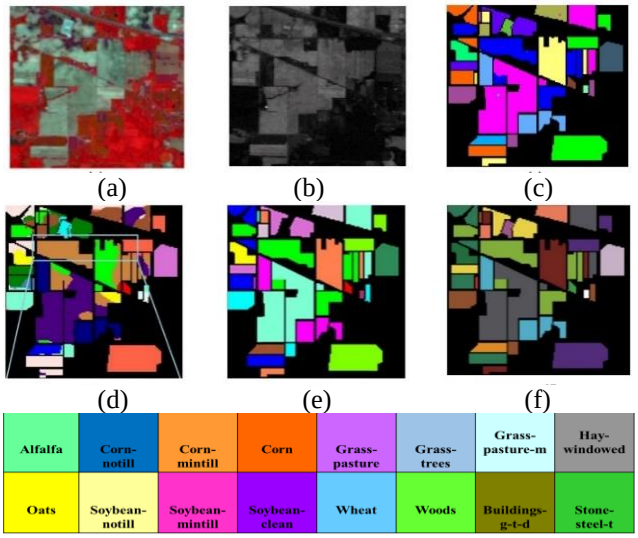

Figure. 3 Classification maps of indian pines data with: (a) pseudo image, (b) SSCNN fusion, (c) SSCNN ground truth (GT), (d) LMAFN GT, (e) EMFFN GT, and (f) LSGSAN GT

C. Framework of the proposed SSCNN algorithm

Input:

1) An HSI with ground truth

2) Initializing an incoming picture patch data i with attribute k as shown in Table 3 notation list

Step 1: Patch's size calculated as m*m with k, Y=i+k and attribute k falling between 8 and 12.

Step 2: Initializing leakyRelu with '0', sigmoid Y = 1/(1+exp(-i)) and weights are initialized with Gaussian distribution mean of 0.

Step 3: Fusion of individual respective patches is denoted by $\tilde{J}_n$ and proposed dictionary is dependent on the $\mathbb{D}$ with dimensions, is estimated to be $m^2 * n$, $n > m^2$.

Table.3 Notation list

Sl.no	Notations	Interpretation	Sl.no	Notations	Interpretation
1.	K	Incoming data attribute	16.	C	Original image's depth
2.	i	Incoming picture patch data	17.	N	Spectrum lengths
3.	$m \times m$	Patch's size	18.	$J \in S^{C \times O}$	Initial hyperspectral image
4.	$\mathbb{D}$	Dictionary	19.	$C_1, C_2, \ldots, C_N$	Array of bands
5.	$\in$	Preset parameter	20.	$\hat{J}_n$	N_{th} fused band
6.	L_0	Pseudo normalization	21.	C_n	Overall amount of bands
7.	$\mathbb{Y}$	Patch-wise fusion of features	22.	$\hat{J}^i$	I_{th} subgroup of bands
8.	$\|\mathbb{D}_\alpha - \mathbb{Y}\|_2$	Approximate norm	23.	A	Number of bands in every subgroup
9.	$\hat{I}$	Fused image	24.	S^i	Coefficient of reflection
10.	σ	Noise level	25.	T^i	Coefficient of shading
11.	N_r^-	Image's centre pixel	26.	$y = w^T\phi(X) + b$	CNN linear hyperplane
12.	$*$	Convergence	27.	$\phi(X)$	Feature space of input data
13.	η	Step size	28.	t_i	Objective tag for every data point
14.	λ	Acceptable probability	29.	φ_i	Slack variable
15.	O	Array of the pixels	30.	w^T, b	Weights and bias of input data point

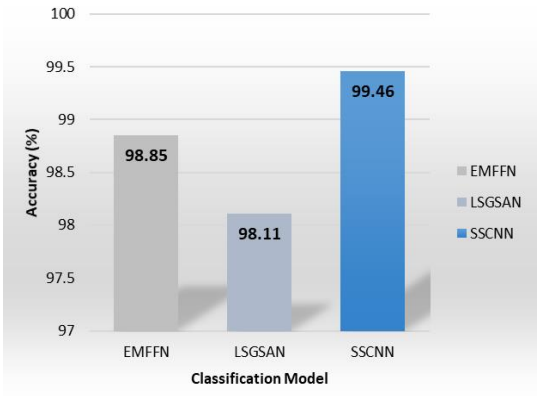

Figure. 4 Overall accuracies of classification models

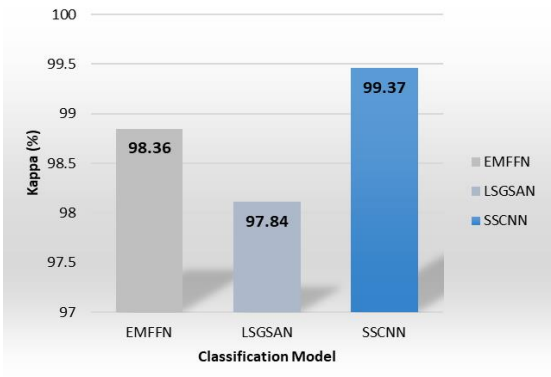

Figure. 5 Kappa value of classification models

Step 4: The duplicate features within the corresponding HSI and fusing are removed employing optimized distribution of probabilities using exponential $i^{(t+1)}$.

Step 5: Apply softmax function along gradient descent algorithm and hyperparameters batch size of 128, dropout rate '0.5', learning rate '0.1', batch normalization '1e-5', softmax loss '1e-6'.

Step 6: Adopting SVM in CNN for the classification utilizing adaptive thresholding with minimum peak threshold '100', minimum noise threshold '10', noise threshold margin '0.5', gnr threshold margin '0.2'.

Step 7: Introducing a soft margin CNN hyperplane linear approach is the optimal decision boundary to divide the HSI dataset among various labels.

Output: A matrix with two dimensions that stores the HSI labels.

D. Estimation of PSNRSSIM for fusion and classification accurate analysis mse=mean(e^2), psnr_cur=10*log10(255^2/mse), mse_cur = mse/100. Where e =A(:)-B(:) ; A is the ground truth image and B is the original image. peaksnr = psnr(A, ref); mssim= mean2(ssim_map). SSIM evaluate the variance between the properties (luminance, contrast and structure) of the pixels and similarity index of the images with each pixel, while PSNR just measures the absolute error between the pixels.

The overall accuracies of Classification models as shown in below Fig. 4. The Kappa value of Classification models are given in Fig. 5.

6. Conclusion

The current approached statistical models couldn't provide the correctness of the objects region classification from HSI data, with longer learning process time. The proposed SSCNN classify spectral data employing neural network, patches learning with dictionary storage of fusion band process, and reduces classifying time. The methodology utilised soft-margin decision boundary classification to obtain the accurate outcome of specific object regions including intrinsic parameters. Overall model improved to optimise the processing time with reduced model space, where the space and time are changeless state. The SSCNN categorization paradigm has been created, combining the reduction of dimensions, spatial-spectral extraction of characteristics, and nonlinear feature subsequent processing to improve the accuracy of categorization for small-scale HSI problems. The central concept of the suggested SSCNN is the efficient fusing of multi-type features with adequate variance, minimal redundancy, and minimal misclassification error. Firstly, the incoming image data is divided into patches which are given for further learning and stored in the dictionary is used to reduce the dimensionality of hyperspectral data, and the spectral features are pre-learned from the learned dictionary of data covariance. The SSCNN-enhanced feature learning unit is employed to improve the

hyperspectral features even further. Due to effective fusion performance and patch-wise categorization, testing efficiency ranks higher than training efficiency. The SSCNN can efficiently emphasize the important spatial-spectral feature regions for classifying and corresponding the spatially compressed hyper-spectrum information to the four-dimensional space of features.

A composite softmax is applied at last to convert into a normalized probability distribution, used to aggregate the features learned by different stages. Finally, the classification is realized by using max pooling with low-rank constraints to learn compact class vectors for each class for faster learning deployed leaky activation functions for further deep classification convolution. We can always get a global optimum from the SSCNN model since the CNN issue is a convex optimization problem. When compared to the current hyperspectral space object classification model, the suggested SSCNN is able to achieve improved fusion correctness and classification by using soft-margin slack variables, utilizing the optimum decision boundary to divide our dataset into several labels.

7. Future work

The same methodology will be implemented for video space dataset and classifying objects with correlating the image frames in consideration of space and time metrics.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization, santhosh laxman deshpande and sneha venkateshalu; methodology, software, validation, formal analysis, investigation, data curation, sneha venkateshalu; writing—original draft preparation, writing—review and editing, visualization, project administration, sneha venkateshalu; supervision, santhosh laxman Deshpande.

Acknowledgment

I always love to be gratitude towards my supreme, parents. Thankful for my guide and friend being as supporters to complete my work.

References

[1] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 50, No. 4, pp. 1185–1198, 2012.

[2] S. Prasad and L. M. Bruce, "Limitations of principal components analysis for hyperspectral target recognition", *IEEE Geoscience and Remote Sensing Letters*, Vol.5, No.4, pp. 625–629, 2008.

[3] S. Di. Zenzo, S. D. Degloria, R. Bernstein, and H. G. Kolsky, "Gaussian Maximum Likelihood and Contextual Classification Algorithms for Multicrop Classification Experiments Using Thematic Mapper and Multispectral Scanner Sensor Data", *IEEE Transactions on Geoscience and Remote Sensing, IEEE Transactions On Geoscience And Remote Sensing*, Vol. Ge-25, No. 6, pp.815–824, 1987.

[4] A. M. Martinez and A. C. Kak, "PCA versus LDA", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2, pp. 228–233, 2001.

[5] F. Wu, Z. Wang, Z. Zhang, Y. Yang, J. Luo, W. Zhu, and Y. Zhuang, "Weakly Semi-Supervised Deep Learning for Multi-Label Image Annotation", *IEEE Transactions on Big Data*, Vol. 1, No. 3, pp. 109–122, 2015.

[6] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "TagProp: Discriminative metric learning in nearest neighbor models for image auto-annotation", In: *Proc. of IEEE International Conf. on Computer Vision, Lear, Inria Grenoble*, pp. 309–316, 2009.

[7] Y. Li, M. Yang, Z. Xu, and Z. Zhang, "Multi-view learning with limited and noisy tagging", In: *Proc. of International conf. on Artificial Intelligence, Zhejiang University*, China, pp. 1718–1724, 2016.

[8] G. Licciardi, P. R. Marpu, J. Chanussot, J. A. Benediktsson, "Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles", *IEEE Geoscience and Remote Sensing Letters*, Vol. 9, No. 3, pp. 447–451, 2012.

[9] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with Independent component discriminant analysis", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 49, No. 12, pp. 4865–4876, 2011.

[10] T. V. Bandos, L. Bruzzone, and G. C. Valls, "Classification of hyperspectral images with regularized linear discriminant analysis", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 47, No. 3, pp. 862–873, 2009.

[11] A. J. X. Guo, and F. Zhu, "A CNN-Based spatial feature fusion algorithm for hyperspectral imagery classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 57, No. 9, pp. 7170–7181, 2019.

[12] F. Melgani and L. Bruzzone, "Classification of Hyperspectral Remote Sensing Images with Support Vector Machines Melgani", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 42, No. 8, pp. 1778–1790, 2004.

[13] J. S. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 43, No. 3, pp. 492–501, 2005.

[14] M. Hasheminejad, "Optimized kernel Nonparametric Weighted Feature Extraction for Hyperspectral Image Classification", *Journal of Information Systems and Telecommunication*, Vol. 10, No. 38, pp. 111–119, 2022.

[15] L. Ran, Y. Zhang, W. Wei, and Q. Zhang, "A hyperspectral image classification framework with spatial pixel pair features", *Sensors Article*, Switzerland, Vol. 17, No. 10, 2017.

[16] Y. Xu and L. Zhang, "Beyond the Patchwise Classification: Spectral-Spatial Fully Convolutional Networks for Hyperspectral Image Classification", *IEEE Transactions on Big Data*, Vol. 6, No. 3, pp. 492–506, 2020.

[17] M. Pesaresi, and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 39, No. 2, pp. 309–320, 2001.

[18] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 43, No. 3, pp. 480–491, 2005.

[19] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of Hyperspectral Images by Exploiting Spectral-Spatial Information of Superpixel via Multiple Kernels", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 53, No. 12, pp. 6663–6674, 2015.

[20] G. Li, L. Li, H. Zhu, X. Liu, and L. Jiao, "Adaptive Multiscale Deep Fusion Residual Network for Remote Sensing Image Classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 57, No. 11, pp. 8506–8521, 2019.

[21] T. Hill, L. Marquez, M. O'Connor, W. Remus, "Artificial Neural Network Models For Forecasting And Decision Making", *International Journal of Forecasting*, Vol. 10, No.1, pp. 5-15, 1994.

[22] P. G. Zhang, "Neural Networks For Classification: A Survey", *IEEE Transactions on Systems Man and Cybernetics Part C (Applications and Reviews)*, Vol. 30, No. 4, pp. 451-462, 2000.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks", *Communications of the ACM*, Vol. 60, No. 6, pp. 84–90, 2017.

[24] J. Schmidhuber, "Deep Learning in neural networks: An overview", *Journal In Neural Networks*, Vol. 61, pp. 85–117, 2015.

[25] X. X. Zhu, D. Tuia, L. Mou, G. S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources", *IEEE Geoscience and Remote Sensing Magazine*, Vol. 5, No. 4, pp. 8–36, 2017.

[26] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 7, No. 6, pp. 2094–2107, 2014.

[27] F. Feng, Y. Zhang, J. Zhang, and B. Liu, "Low-Rank Constrained Attention-Enhanced Multiple Spatial–Spectral Feature Fusion for Small Sample Hyperspectral Image Classification", *International Journal of Remote Sensing*, Vol. 15, No.2, p. 304, 2023.

[28] J. Yang, C. Wu, B. Du, and L. Zhang, "Enhanced Multiscale Feature Fusion Network for HSI Classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 59, No. 12, pp. 10328-10347, 2021.

[29] H. Gao, Z. Chen, and F. Xu, "Adaptive spectral-spatial feature fusion network for hyperspectral image classification using limited training samples", *International Journal of Applied Earth Observation and Geoinformation*, Vol. 107, 2022.