Land scene classification using diversity promoting metric learning-convolutional neural network

Kampa Ratna Babu[1], Kampa Kanthi Kumar[2], Akula Suneetha[3]
[1]Department of Computer Engineering, M.B.T.S. Government Polytechnic, Guntur, India
[2]Department of Electronics and Communication Engineering, Tirumala Engineering College, Narasaraopet, India
[3]Department of Computer Science and Engineering, KKR & KSR Institute of Technology and Sciences, Guntur, India

Article Info

Article history:

Received Jun 23, 2024
Revised Sep 3, 2024
Accepted Sep 7, 2024

Keywords:

Convolutional neural network
Diversity promoting metric learning
Land scene classification
Remote sensing scene images
Uncorrelation

ABSTRACT

The land scene classification by remote sensing images predicts semantic class of image blocks by removing visual primitives in remote sensing images. However, there is a problem of within-class diversity and between-class similarity that degrades a performance of scene classification. In this research, the diversity promoting metric learning–convolutional neural network (DPML-CNN) method is proposed for classifying land scene images. The metric learning with convolutional neural network (CNN) maps the same scene image class closer and the different class scenes as far as possible which makes the method much discrimination. The diversity promoting in metric learning is used to reduce the overlapping of the same scene class by uncorrelation of every parameter and provides unique information for those parameters. The UC Merced, AID, and NWPU RESISC45 datasets are utilized in this research for evaluating the proposed DPML-CNN method with evaluation metrics like accuracy and kappa coefficient. The DPML-CNN method reached highest accuracy of 99.27% and 99.84% for 50% and 80% training ratios on the UC Merced dataset when compared to other existing methods like multi-level semantic feature clustering attention (MLFC-Net) and global context spatial attention (GCSA-Net).

Corresponding Author:

Kampa Ratna Babu
Department of Computer Engineering, M.B.T.S. Government Polytechnic
Guntur, India
Email: kratnababu@gmail.com

1. INTRODUCTION

The remote sensing image scene classification maps scene images to particular high-level semantic classes and enables the acquisition of high-level semantic data [1]. Recently, it has been a promising research field in high-resolution image classification of remote sensing [2]. It is majorly utilized in analyzing natural resources, land use and classification of land coverage, detection of disasters, and planning of urban [3]. The classification of scenes supports people in understanding image content, which is highly convenient to the lives of people in various applications like smart cities, detection of remote sensing [4]-[6]. However, intra-class variations are difficult, and inter-class similarity of objects in original sensors maximizes the data combination and logical reason in every scene [7]. For the classification of scene images, the image is initially characterized through a feature encoder and next classified by a classifier [8]. Commonly, there are variations and inconsistencies among data extracted from visual information and comprehension of people in similar data that cause semantic gaps among feature representation and understanding of high-level features [9]. In recent times, convolutional neural networks (CNNs) have developed in area of image scene classification [10].

Variant CNN-dependent techniques have been ruling in the area of remote sensing scene classification [11]. The great achievement of different CNN-based image scene classification techniques is highly attributed to utilization of deep CNN techniques like VGGNet, GoogLeNet, and AlexNet. These techniques extract the meaningful image feature representations that are much discriminative than handcrafted low-level features like texture, spectral, and color features [12]. Though performance of scene image classification is enhanced through meaningful features learned through deep CNN techniques [13], [14], there are two main issues: within-class diversity and between-class similarity. These issues degrade the performance of remote sensing image scene classification which needs to be tackled.

Wang *et al.* [15] suggested a multi-level semantic feature clustering attention (MLFC-Net) method depending on deep convolution neural networks (DCNNs) which extracted much more accurate feature data. The suggested method utilized high spatial data in remote sensing images combining common semantic feature data with clustered data. Then, Rearrange the respective data weight like feature maps and tensor blocks by attention mechanism. The suggested method improved the representation of various difficult aspects with minimum computational cost and better portability. However, the suggested method was not good at discrimination capability among variant scene categories. Thirumaladevi *et al.* [16] presented a method through integrates convolutional neural network (CNN) and transfer learning to land scene classification. The categorization of images was enhanced for scene classification accuracy through transfer learning with networks like AlexNet and visual geometry group (VGG) and compared to conventional feature extraction techniques. Initially, features were recovered from the network's next fully connected layer and assigned support vector machine (SVM) classification. Next, the final layers of the networks were substituted during transfer learning to classify new datasets. The presented method captured the boundary of scene images and recognized it. However, the method was difficult in describing high semantic data in remote sensing images. Chen *et al.* [17] developed a method that depended on global context spatial attention (GCSA) and densely connected convolutional networks for extracting multiple scale scene features named GCSANet. The mixture process was utilized for improving spatial mixture information of remote sensing images and discrete sample space was reduced to enhance smoothness in data space neighborhood. The GCSA was implemented in a densely connected network for encoding contextual data of remote sensing scene images into local features. The developed method has high robustness and stability, by using mix-up operations that enhance the classification accuracy and smoothness of the method. However, the issue of interclass similarity lies in overlapping of similar surfaces between various scenes. Lv *et al.* [18] incorporated the benefits of multiple scale and multiple level features and developed a method that combined global features for identifying global attention features and learning the multi-deep dependencies among variations in spatial scale. Two various feature adaptive fusion methods were implemented for exploring complementary associations of local and global aggregate features that acquire various image scenes. The method explored the nature of global and local features for comprehensively describing the image scenes. However, due to limited data, the method tends to issue over-fitting and less feature generalization capability.

Ma *et al.* [19] presented the homo-heterogenous transformer learning (HHTL) method to classify RS scenes. Initially, a patch generation method was developed for generating patches of homogeneous and heterogeneous data within RS scenes. Next, a double-phase feature learning module (FLM) was introduced for homogeneous and heterogeneous data within RS scenes. In the FLM-dependent vision transformer, both global data and local regions and their context data were captured. At last, developed a classification method that has a fusion submodule and metric learning. However, it has difficult intra-class variations and inter-class similarities in original scenes that maximize the complexity of data integration in every scene. Xu *et al.* [20] introduced a Lie Group deep learning method to classify remote sensing scenes. Initially, extracted shallower and high-level features from images depended on lie group machine learning (LGML) and deep learning for maximizing the capacity of feature representation of the method. Then, the spatial attention mechanism enhanced local semantic features and compressed irrelevance feature data. Finally, feature-level fusion was employed to minimize redundant features and enhance execution performance. Additionally, cross-entropy loss function with label smoothing was utilized to improve classification accuracy. The introduced method minimized the influence of huge similarity classes on scene classification. However, the method can't be concentrated on the spatial relationship of local features. From the overall analysis, the existing methods have limitations like difficulty in describing the high semantic data due to limited representation, issue of interclass similarity, difficulty in intra-class variations and inter-class similarities and can't concentrate on the spatial relationship of local features. The problem of within-class diversity and between-class similarity degrades the performance of remote sensing image scene classification which needs to be tackled. In this research, the proposed DPML-CNN method maximizes interclass variation and minimizes the interclass variation. The metric learning in CNN makes the method much more discriminative and captures both global and local features. The diversity promotion in metric learning minimizes the overlapping of interclass similarity by uncorrelation of every scene parameter. The significant contributions of the research are mentioned as follows:

- The diversity promoting metric learning – CNN (DPML-CNN) method is proposed to classify land scene remote sensing images. The CNN with diversity-promoting metric learning minimized the cross-entropy loss and made the method much more discriminative.
- The metric learning regularization used the diversity promoting term which uncorrelated every parameter and gives unique information for those parameters. This minimizes the overlapping of similar scenes and enhances the method's performance.
- The scale invariant feature transform (SIFT) and local binary patterns (LBP) techniques are used for extracting meaningful features from images which enhanced the classification performance of the DPML-CNN method.

This manuscript is organized as follows: Section 2 explains the process of the proposed framework. Section 3 gives the results and comparison of the proposed method with existing methods. The conclusion of this research is given in Section 4.

2. PROPOSED METHOD

This research, proposed a DPML-CNN method to classify the land scene images from remote sensing scene images. The UC Merced, AID, and NWPU RESISC 45 datasets are utilized in this research and the significant features from the images are extracted by SIFT and LBP methods. Then, the extracted features are classified by using the DPML-CNN method. Figure 1 describes the process of the proposed method framework.

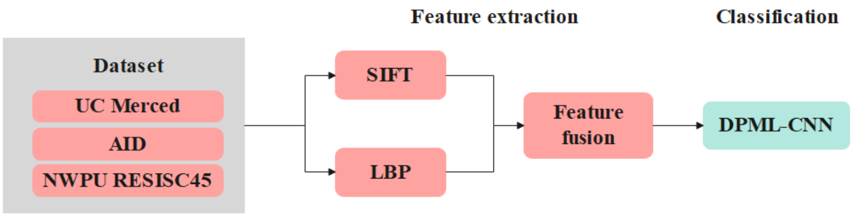

Figure 1. Process of proposed method framework

2.1. Dataset

The datasets used in the research for land scene image classification are UCM [21], AID [22], and NWPURESISC45 [23] datasets. These three datasets contain remote sensing scene images which are utilized in this research to classify land scenes. The detailed explanation of three datasets is described below:

2.1.1. UCM dataset

The UCM dataset is available through University of California Merced and it has 2100 remote sensing scene images. The remote sensing images are separated into 21 scenes and every class has 100 images of size 256 x 256 pixels. The spatial resolution of image is 0.3 m per pixel in color space of red green blue (RGB).

2.1.2. AID dataset

The AID dataset is developed through Wuhan University and it has 10,000 remote sensing scene images. The remote sensing scene images are separated into 30 scene classes with the size of 600 × 600 pixels and the number of images in every class are varied from 200 to 400. The spatial resolution varied from 8 m to 0.5 m per pixel.

2.1.3. NWPURESISC45 dataset

The NWPU dataset is developed through Northwestern Polytechnical University and contains 31,500 remote-sensing scene images in color space of RGB. The remote sensing scene images are separated into 45 scene classes and every class contains 700 scene images with the size of 256 x 256 pixels. The spatial resolution varied from 0.2 m to 30 m. Table 1 represents the description of all three datasets. The images from the three datasets are given as raw input images to the feature extraction stage to extract the local and pixel-based features for classification. The detailed explanation of feature extraction is described as follows.

Table 1. Dataset description

Dataset	Total images	No. of scenes	Size of image
UCM [20]	2100	21	256 x 256
AID [21]	10,000	30	600 x 600
NWPURESISC45 [22]	31,500	45	256 x 256

2.2. Feature extraction

Feature extraction is a process of converting raw image data to certain meaningful representations which minimizes the dimensionality reduction. In General, the high dimensional nature of images can lead to a reduction in classification performance. The feature extraction converts high-dimensional data to low-level data when extracting significant data from scene images. In this research, SIFT and LBP feature extraction techniques are used to extract local and pixel-based features.

2.2.1. Local binary patterns (LBP)

The LBP identified uniform LBP as a critical feature that represents image texture. The uniform LBP is employed for generating occurrences of histograms for the representation of texture features. The LBP characterized the image by spatial data of image texture structure [24], [25]. The LBP is measured by thresholding the neighbor $\{p_i\}_{i=0}^{n-1}$ pixels along the middle pixel p_c for executing an n-bit binary number that is changed to decimal and numerical expression is given as (1). The $d_p = (p_c - p_i)$ represents variance between neighbor and center pixels representing the spatial architecture of center position with local variance vector $[d_0, d_1, \dots, d_p - 1]$. The LBP generated a histogram and which is given in (2).

$$LBP_{n,r}(p_c) = \sum_{i=0}^{n-1} s(p_i - p_c)2^i = \sum_{i=0}^{n-1} s(d_i)2^i, s(x) = \begin{cases} 1, x > 0 \\ 0, x < 0 \end{cases} \quad (1)$$

$$\big(H(m)\big) = \sum_{p=0}^{p} \sum_{j=0}^{J} f\big(LBPs_{n,r}, m \in [0, m]\big), f(x, y) = \begin{cases} 1, \\ 0, \end{cases} \quad (2)$$

The m represents the highest pattern number of LBP. The LBP partitions the image to the fixed size of grid cells for accomplishing the pooling of local texture descriptors. The LBP characterized the image by spatial data of image texture structure.

2.2.2. Scale invariant feature transform (SIFT)

The SIFT method is a majorly utilized shape feature extraction technique. The technique is a key point detector and descriptor technique for extracting meaningful features from images. That is hugely robust to scaling and orientation of the image and is invariant for illumination changes. The SIFT extracted the highest features from low-level resolution images [26], [27]. The SIFT extracted the 128 features from remote sensing scene images by filtering method that processed in four phases. The initial phase detected significant positions from an image by difference of the Gaussian (DoG) method. Next, localization is processed to determine significant features. In final phase, executed key points are changed to feature vectors. The extracted features from the LBP and SIFT methods are fused for the classification process.

2.3. Classification by DPML-CNN method

The classification of land scenes is performed by CNN with diversity-promoting metric learning. The performance of scene classification is enhanced because of meaningful features learned by CNN methods. The proposed CNN with diversity-promoting metric learning minimized the cross-entropy loss and learned the model to be discriminative. The proposed method maps the images from similar scene classes closest and images of various categories as far as possible [28], [29]. A detailed explanation of discriminative CNN and diversity-promoting metric learning is described below. Figure 2 describes the process of the proposed DPML-CNN method.

2.3.1. Diversity promoting metric learning – CNN (DPML-CNN)

The proposed DPML-CNN is used for land scene classification to tackle issues of within-class diversity and between-class diversity. Here extracted the meaningful CNN features for enhancing the performance of classification. The proposed method minimized cross-entropy loss which is an error of softmax classification from the last fully connected (FC) layer utilized in classical CNN methods. The diversity-promoting metric learning on CNN features makes the method much more discriminative and maps the similar scene classes closer together and distant scene classes as far as possible.

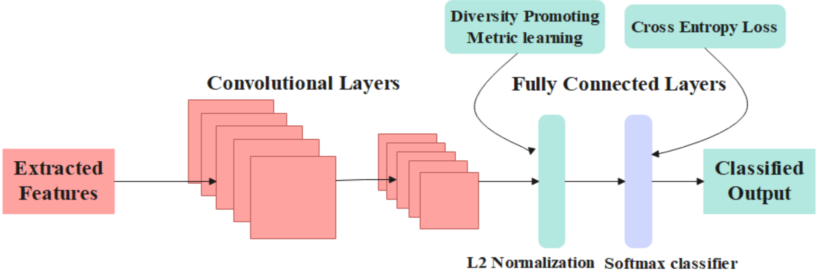

Figure 2. Process of proposed DPML-CNN method

2.3.2. Learning discriminative CNNs

Consider $X = \{x_i | i = 1,2, \dots, N\}$ as a set of training samples and $Y = \{y_i | i = 1,2, \dots, N\}$ as a set of labels in X, where N represents whole count of training samples, $y_i \in R^c$ represents label vector of ground truth in x_i sample along one component and C represents the whole count of image scene classes. The DPML-CNN method contains $L + 1$ layers. The network parameters are $W = \{W_1, \dots, W_L, W_{L+1}\}$ and $B = \{B_1, \dots, B_L, B_{L+1}\}$, here W_l represents filter weights of lth layer and B_l represents respective biases, $l = 1, \dots, L + 1$. The $(L + 1)$ layer represents softmax classification and Lth layer represents result of DPML-CNN method. The input x_i and result of softmax layer $O_{L+1}(x_i)$, $O_L(x_i)$ represents DPML-CNN feature layer and whole intermediate layers $O_l(x_i)$ is given as (3) – (5).

$$O_{L+1}(x_i) = S_{L+1}(W_{L+1} O_L(x_i) + B_{L+1}) \tag{3}$$

$$O_L(x_i) = \frac{S_L(W_L O_{L-1}(x_i) + B_L)}{\|S_L(W_L O_{L-1}(x_i) + B_L)\|_2} \tag{4}$$

$$O_l(x_i) = S_l(W_l O_l(x_i) + B_l) \tag{5}$$

The S_l represents element-wise nonlinear activation functions like softmax. The S_{L+1} represents a softmax function, next for executing the distance in DPML-CNN is L2 normalized for eliminating scale variance. This research proposed a term that consists of cross-entropy loss, diversity promoting metric learning, and weight decay term. Numerical expression is given as (6). The λ_1 and λ_2 represents parameters of tradeoff which controls the relative significance of three terms. The detailed explanation of three terms is described in the below subsections.

$$J = \min\,(J_1(X, W, B) + \frac{\lambda_1}{2} J_2(X, W, B) + \frac{\lambda_2}{2} J_3(W, B) \tag{6}$$

A. Cross-entropy loss

The cross-entropy loss is determined as cross-entropy loss function which minimized classification error in training samples. The $(y_i, \log O_{L+1}(x_i))$ represents the inside product of y_i and $\log O_{L+1}(x_i)$ and N represents the count of training samples in X. and numerical expression is given as (7).

$$J_1(X, W, B) = -\frac{1}{N} \sum_{i=1}^{N} (y_i, \log O_{L+1}(x_i)) \tag{7}$$

B. Diversity promoting metric learning

This learning enforced a CNN method to be much more discriminative, which enhances a feature representation that has little scatter of intraclass and huge separation of interclass. Given every paired training sample (x_i, x_j), its distance of pair-wise feature is calculated through executing Euclidean distance among CNN feature representations. The numerical expression for computing Euclidean distance is given as (8).

$$D(x_i, x_j) = \|O_L(x_i) - O_L(x_j)\|_2 \tag{8}$$

To exploit discriminative feature representation in output layer of the DPML-CNN method, the distances among the same pairs are lesser than dissimilar pairs and there is a huge margin among similar and dissimilar pairs developed by image scene categories Particularly, whether two images shared similar scene label it is considering as similar pair or else consider as dissimilar pair. For protecting the issue of dissimilar pairs being chosen with a similar number of the same pairs. The x_i and x_j represents similar scene class, its feature distance is represented as $D(x_i, x_j)$ is smaller than up-margin τ_1 whether x_i and x_j represents various scene categories, its feature distance is higher than the down-margin τ_2. Numerical expression is given as (9).

$$\begin{cases} D^2(x_i, x_j) < \tau_1, & y_i = y_j \\ D^2(x_i, x_j) > \tau_2, & y_i \neq y_j \end{cases} \tag{9}$$

The τ_1 is utilized for penalizing same-pair distances, τ_2 is utilized for constraining dissimilar distance pairs in training and τ_2 is higher than τ_1. For minimizing the count of parameters in the proposed method, implemented intermediate parameter τ for merging τ_1 and τ_2. Particularly, set $\tau_1 = \tau - 0.05$ and $\tau_2 = \tau + 0.05$ and numerical expression is given as (10). The y_{ij} represents label indicator for paired data (x_i, x_j) and numerical expression is given as (11).

$$0.05 - y_{ij}\left(\tau - D^2(x_i, x_j)\right) < 0 \tag{10}$$

$$y_{ij} = \begin{cases} +1, & y_i = y_j \\ -1, & y_i \neq y_j \end{cases} \tag{11}$$

There is a connection among every similar and dissimilar pair learned in CNN feature space. Through employing the constraint to every similar and dissimilar pair in training phase obtained hinge loss function of diversity promoting metric learning and numerical expression is given as (12).

$$J_2(X, W, B) = \sum_{i,j} h\left(0.05 - y_{ij}\left(\tau - \left\|O_L(x_i) - O_L(x_j)\right\|_2^2\right)\right) \tag{12}$$

C. Diversity promoting and weight deacy

The general remote sensing scenes can't give enough samples for training, this research encourages the metric parameter factors to be diversified for enhancing capability of representation and captures much discriminative data from the restricted count of samples for acquiring good performance. To meet specific requirements, the diversity-promoting priors are implemented as regularization factors in parameters for better representation. It focused on diversification among learned parameters. By implementing diversity-promoting priors in ML, every parameter factor represents unique information from scenes and all factors are combined into a large proportion of training data. Hence, much data is captured through diversified methods, and performance is enhanced. It is developed for minimizing weight magnitudes of W and B and it is used to prevent over-fitting, the numerical expression is given as (13).

$$J_3(W, B) = \sum_{l=1}^{L+1}(\|W_l\|_F^2 + \|B_l\|_2^2) \tag{13}$$

By utilizing diversity-promoting metric learning in CNN, it tackles the issue of within-class diversity and between-class similarity. The overlapping of similarity scenes is tackled by using diversity promoting in metric learning which uncorrelated every factor and represents unique information for every scene. This minimizes the probability of overlapping and increases the performance of the method. By using the DPML-CNN method, it classified the land scenes from remote sensing images with high classification performance.

3. EXPERIMENTAL ANALYSIS

The proposed DPML-CNN method is simulated with the environment of MATLAB 2020a and with system requirements are a Windows 10 operating system, 16 GB of RAM, and an i5 processor. The evaluation metrics used for analyzing the proposed method are Accuracy (%) and Kappa Coefficient (%) with three datasets such as UC Merced, AID, and NWPU RESISC 45. The accuracy is represented as a count of accurately classified images separated through a whole number of testing images; this reflected the overall classification performance of the method. The Kappa coefficient defines the ratio of error minimization among classification and fully random classification.

3.1. Quantitative and qualitative analysis

The proposed DPML-CNN method is evaluated with evaluation metrics of accuracy, kappa coefficient, precision, and f1-score with three datasets. The evaluation is performed based on different training ratios for three different datasets. The separated training ratios for UCM dataset are 50% and 80%, for AID dataset are 20% and 50%, for NWPU RESISC 45 dataset are 10% and 20%. The existing neural networks used for evaluating the proposed DPML-CNN method are deep neural network (DNN), CNN, and region-based CNN (RCNN). Three different tables are described below for three datasets. In Table 2, the classification performance on the UCM dataset is described with evaluation metrics of accuracy, kappa coefficient, precision and f1-score. The proposed DPML-CNN method reached the highest accuracy of 99.27% and 99.84% for 50% and 80% training ratios. The proposed method attained a kappa coefficient of 98.63%, 99.94% for 50% and 80% training ratios on UCM dataset. The proposed method shows effective classification performance on land scenes when compared to other neural networks.

In Table 3, the classification performance on AID dataset is described with evaluation metrics of accuracy, kappa coefficient, precision and f1-score. The proposed DPML-CNN method reached the highest accuracy of 96.23% and 97.91% for training ratios of 20% and 50%. The proposed method attained a kappa coefficient of 94.93%, 98.03% for training ratios of 20% and 50% on AID dataset. The proposed method shows effective classification performance on land scenes when compared to other neural networks.

In Table 4, the classification performance on NWPU RESISC45 dataset is described with evaluation metrics of accuracy, kappa coefficient, precision, and f1-score. The proposed DPML-CNN method reached the highest accuracy of 93.52% and 95.21% for training ratios of 10% and 20%. The proposed method attained a kappa coefficient of 92.78%, and 93.92% for training ratios of 20% and 50% on NWPU RESISC45 dataset. The proposed method shows effective classification performance on land scenes when compared to other neural networks.

Table 2. Classification performance on UCM dataset

Methods	Accuracy (%)		Kappa Coefficient (%)		Precision (%)		F1-score (%)	
	50%	80%	50%	80%	50%	80%	50%	80%
DNN	96.45	97.02	96.17	96.93	93.89	94.56	94.57	95.26
CNN	97.03	97.84	96.89	97.26	94.67	95.78	95.53	96.39
RCNN	97.78	98.63	97.26	98.77	95.71	96.25	96.62	97.74
DPML-CNN	99.27	99.84	98.63	99.94	97.43	98.27	97.71	98.56

Table 3. Classification performance on AID dataset

Methods	Accuracy (%)		Kappa Coefficient (%)		Precision (%)		F1-score (%)	
	20%	50%	20%	50%	20%	50%	20%	50%
DNN	93.82	94.47	92.84	94.73	93.73	95.49	92.71	94.52
CNN	94.78	95.58	93.45	96.04	94.68	96.82	93.59	95.72
RCNN	95.37	96.88	94.02	97.16	95.56	97.61	94.83	96.69
DPML-CNN	96.23	97.91	94.93	98.03	96.62	98.45	95.82	97.28

Table 4. Classification performance on NWPU RESISC45 dataset

Methods	Accuracy (%)		Kappa Coefficient (%)		Precision (%)		F1-score (%)	
	10%	20%	10%	20%	10%	20%	10%	20%
DNN	91.42	92.79	89.28	90.58	88.35	89.92	90.56	91.27
CNN	92.03	93.37	90.46	91.65	89.56	91.49	91.66	92.78
RCNN	92.77	94.68	91.86	92.76	91.45	92.71	92.78	93.63
DPML-CNN	93.52	95.21	92.78	93.92	92.17	93.47	93.12	94.07

3.2. Comparative analysis

The classification performance of the DPML-CNN method is compared with existing land scene methods like MLFC-Net [14], TL [15], GCSA-Net [16], GLFAF-Net [17], HHTL [18], and LGML [19]. Three different comparison tables are described in this section for three datasets with different training ratios. By utilizing diversity-promoting metric learning in CNN, it tackles the issue of within-class diversity and between-class similarity. The overlapping of similarity scenes is tackled by using diversity promoting in metric learning which uncorrelated every factor and represents unique information for every scene. This minimizes the probability of overlapping and increases the performance of the method. By using the DPML-CNN method, it classified the land scenes from remote sensing images with high classification performance. Table 5 describes comparison of DPML-CNN method with other classification methods like MLFC-Net [14], TL [15], GCSA-Net [16], GLFAF-Net [17], HHTL [18] and LGML [19] on the UCM dataset. The proposed method is compared with evaluation metrics of accuracy and kappa coefficient on 50% and 80% training ratios. From Table 5, it is clear that the proposed method performed more effectively than existing algorithms.

Table 5. Comparison of the proposed method on UCM dataset

Methods	Performance Metrics	Training ratio	
		50%	80%
MLFC-Net [14]	Accuracy (%)	98.53	99.66
TL [15]	Accuracy (%)	N/A	95
GCSA-Net [16]	Accuracy (%)	98.32	99.31
GLFAF-Net [17]	Accuracy (%)	97.52	N/A
HHTL [18]	Accuracy (%)	98.87	99.48
LGML [19]	Accuracy (%)	98.67	99.78
	Kappa coefficient (%)	98.31	99.76
Proposed DPML-CNN	Accuracy (%)	99.27	99.84
	Kappa Coefficient (%)	98.63	99.94

Table 6 describes comparison of the DPML-CNN method with other classification methods like GCSA-Net [16], HHTL [18] and LGML [19] on the AID dataset. The proposed method is compared with evaluation metrics of accuracy and kappa coefficient on 20% and 50% training ratios. From Table 6, it is clear that the proposed method performed more effectively than existing algorithms.

Table 6. Comparison of the proposed method on the AID dataset

Methods	Performance Metrics	Training ratio	
		20%	50%
GCSA-Net [16]	Accuracy (%)	95.96	97.53
HHTL [18]	Accuracy (%)	95.62	96.88
LGML [19]	Accuracy (%)	94.79	97.72
	Kappa coefficient (%)	94.57	97.61
Proposed DPML-CNN	Accuracy (%)	96.23	97.91
	Kappa coefficient (%)	94.93	98.05

Table 7 describes comparison of the DPML-CNN method with other classification methods like GCSA-Net [16], HHTL [18] and LGML [19] on the NWPU RESISC45 dataset. The proposed method is compared by evaluation metrics of accuracy and kappa coefficient on training ratios of 10% and 20%. From Table 7, it is clear that the proposed method performed more effectively than existing algorithms.

Table 7. Comparison of the proposed method on NWPU RESISC45 dataset

Methods	Performance Metrics	Training ratio	
		10%	20%
GCSA-Net [16]	Accuracy (%)	93.39	94.95
HHTL [18]	Accuracy (%)	92.07	94.21
LGML [19]	Accuracy (%)	92.62	94.49
	Kappa Coefficient (%)	92.25	94.31
Proposed DPML-CNN	Accuracy (%)	93.52	95.21
	Kappa Coefficient (%)	92.78	93.92

3.3. Discussion

The results of DPML-CNN method are evaluated with existing methods like DNN, CNN and RCNN. The proposed DPML-CNN method achieved the highest accuracies on these datasets, with 99.27% and 99.84% for 50% and 80% training ratios on the UC Merced dataset, 96.23% and 97.91% for 20% and 50% training ratios on the AID dataset, and 93.52% and 95.21% for 10% and 20% training ratios on the NWPU RESISC45 dataset. Additionally, the performance of DPML-CNN method is compared with MLFC-Net [14], TL [15], GCSA-Net [16], GLFAF-Net [17], HHTL [18], and LGML [19]. The DPML-CNN method is proposed for the classification of land scenes from remote sensing scene images. The existing methods have drawbacks of not being good at discrimination capability among variant scene categories [14]. Difficult in describing high semantic data in remote sensing images [15]. The issue of interclass similarity lies in the overlapping of similar surfaces among various scenes [16]. Due to limited data, the method has issues of over-fitting and less feature generalization capability [17]. Difficulty in intra-class variations and inter-class similarities in original scenes maximizes the complexity of data integration in every scene [18], can't be concentrated on the spatial relationship of local features [19]. The problem of within-class diversity and between-class similarity degrades the performance of image scene classification which needs to be tackled.

This research proposed a DPML-CNN method that maximizes interclass variation and minimizes the interclass variation. The metric learning in CNN makes the method much more discriminative and captures both global and local features. The diversity promotion in metric learning minimizes the overlapping of interclass similarity by uncorrelation of every scene parameter. By utilizing diversity-promoting metric learning in CNN, it tackles the issue of within-class diversity and between-class similarity. The overlapping of similarity scenes is tackled by using diversity promoting in metric learning which uncorrelated every factor and represents unique information for every scene. This minimizes the probability of overlapping and increases the performance of the method. By using the DPML-CNN method, it classified the land scenes from remote sensing images with high classification performance.

4. CONCLUSION

The DPML-CNN method is proposed for the classification of land scenes from remote sensing scene images. The metric learning with the CNN method enhanced the discrimination capability of the model, effectively capturing both global and local features for improved classification. The diversity promoted in metric learning reduced the overlapping of interclass similarity through the uncorrelation of every scene parameter. The proposed method addresses the issue of class diversity and between-class similarity by mapping the same scene class closer together and different classes as far apart as possible. The proposed DPML-CNN method achieved the highest accuracies on these datasets, with 99.27% and 99.84% for 50% and 80% training ratios on UC Merced dataset, 96.23% and 97.91% for 20% and 50% training ratios on AID dataset, and 93.52% and 95.21% for 10% and 20% training ratios on NWPU RESISC45 dataset. In the future, various deep learning methods can be used to classify the land scene images for further improving the classification performance.

REFERENCES

[1] T. Bhosale and S. Pushkar, “ATiTHi: deep learning and hybrid optimization for accurate tourist destination classification,” *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 10, 2024.

[2] A. Hamza *et al.*, “An integrated parallel inner deep learning models information fusion with Bayesian optimization for land scene classification in satellite images,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 9888–9903, 2023, doi: 10.1109/JSTARS.2023.3324494.

[3] H. Wang, X. Li, G. Zhou, W. Chen, and L. Wang, “Edge enhanced channel attention-based graph convolution network for scene classification of complex landscapes,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 3831–3849, 2023, doi: 10.1109/JSTARS.2023.3265677.

[4] C. C. Yu, T. Y. Chen, C. W. Hsu, and H. Y. Cheng, “Incremental scene classification using dual knowledge distillation and classifier discrepancy on natural and remote sensing images,” *Electronics (Switzerland)*, vol. 13, no. 3, p. 583, Jan. 2024, doi: 10.3390/electronics13030583.

[5] J. Gawlikowski, P. Ebel, M. Schmitt, and X. X. Zhu, “Explaining the effects of clouds on remote sensing scene classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 9976–9986, 2022, doi: 10.1109/JSTARS.2022.3221788.

[6] J. Ni, K. Shen, Y. Chen, W. Cao, and S. X. Yang, “An improved deep network-based scene classification method for self-driving cars,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–14, 2022, doi: 10.1109/TIM.2022.3146923.

[7] M. N. Razali, E. O. N. Tony, A. A. A. Ibrahim, R. Hanapi, and Z. Iswandono, “Landmark recognition model for smart tourism using lightweight deep learning and linear discriminant analysis,” *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, pp. 198–213, 2023, doi: 10.14569/IJACSA.2023.0140225.

[8] K. Liu, J. Yang, and S. Li, “Remote-sensing cross-domain scene classification: a dataset and benchmark,” *Remote Sensing*, vol. 14, no. 18, p. 4635, Sep. 2022, doi: 10.3390/rs14184635.

[9] C. Xu, J. Shu, and G. Zhu, “Scene classification based on heterogeneous features of multi-source data,” *Remote Sensing*, vol. 15, no. 2, p. 325, Jan. 2023, doi: 10.3390/rs15020325.

[10] S. D. Khan and S. Basalamah, “Multi-branch deep learning framework for land scene classification in satellite imagery,” *Remote Sensing*, vol. 15, no. 13, p. 3408, Jul. 2023, doi: 10.3390/rs15133408.

[11] N. Guo *et al.*, “HFCC-Net: a dual-branch hybrid framework of CNN and CapsNet for land-use scene classification,” *Remote Sensing*, vol. 15, no. 20, p. 5044, Oct. 2023, doi: 10.3390/rs15205044.

[12] S. Wang *et al.*, “HSCNet++: hierarchical scene coordinate classification and regression for visual localization with transformer,” *International Journal of Computer Vision*, vol. 132, no. 7, pp. 2530–2550, Feb. 2024, doi: 10.1007/s11263-023-01982-9.

[13] Y. Yu *et al.*, “C2-CapsViT: cross-context and cross-scale capsule vision transformers for remote sensing image scene classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2022.3185454.

[14] K. C. G. Ganashree, R. Hemavathy, and M. R. Anala, “Land scene classification from remote sensing images using improved artificial bee colony optimization algorithm,” *International Journal of Electrical and Computer Engineering*, vol. 14, no. 1, pp. 347–357, Feb. 2024, doi: 10.11591/ijece.v14i1.pp347-357.

[15] D. Wang, C. Zhang, and M. Han, “MLFC-net: A multi-level feature combination attention model for remote sensing scene classification,” *Computers and Geosciences*, vol. 160, p. 105042, Mar. 2022, doi: 10.1016/j.cageo.2022.105042.

[16] S. Thirumaladevi, K. Veera Swamy, and M. Sailaja, “Remote sensing image scene classification by transfer learning to augment the accuracy,” *Measurement: Sensors*, vol. 25, p. 100645, Feb. 2023, doi: 10.1016/j.measen.2022.100645.

[17] W. Chen, S. Ouyang, W. Tong, X. Li, X. Zheng, and L. Wang, “GCSANet: a global context spatial attention deep learning network for remote sensing scene classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 1150–1162, 2022, doi: 10.1109/JSTARS.2022.3141826.

[18] G. Lv, L. Dong, W. Zhang, and W. Xu, “A global-local feature adaptive fusion network for image scene classification,” *Multimedia Tools and Applications*, vol. 83, no. 3, pp. 6521–6554, Jun. 2024, doi: 10.1007/s11042-023-15519-2.

[19] J. Ma, M. Li, X. Tang, X. Zhang, F. Liu, and L. Jiao, “Homo-heterogenous transformer learning framework for RS scene classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 2223–2239, 2022, doi: 10.1109/JSTARS.2022.3155665.

[20] C. Xu, G. Zhu, and J. Shu, “A combination of lie group machine learning and deep learning for remote sensing scene classification using multi-layer heterogeneous feature extraction and fusion,” *Remote Sensing*, vol. 14, no. 6, p. 1445, Mar. 2022, doi: 10.3390/rs14061445.

[21] “UC merced land use dataset.” http://weegee.vision.ucmerced.edu/datasets/landuse.html.

[22] G.-S. Xia *et al.*, “AID: A Benchmark dataset for performance evaluation of aerial scene classification,” *IEEE Trans. on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017, [Online]. Available: https://captain-whu.github.io/AID/.

[23] “NWPURESISC45.” https://www.tensorflow.org/datasets/catalog/resisc45.

[24] G. Dheepak, A. C. J, and D. Vaishali, “Brain tumor classification: a novel approach integrating GLCM, LBP and composite features,” *Frontiers in Oncology*, vol. 13, Jan. 2023, doi: 10.3389/fonc.2023.1248452.

[25] A. Chater, H. Benradi, and A. Lasfar, “New approach to similarity detection by combining technique three-patch local binary patterns (TP-LBP) with support vector machine,” *IAES International Journal of Artificial Intelligence*, vol. 12, no. 4, pp. 1644–1653, Dec. 2023, doi: 10.11591/ijai.v12.i4.pp1644-1653.

[26] J. S. Sujin and S. Sophia, “High-performance image forgery detection via adaptive SIFT feature extraction for low-contrast or small or smooth copy–move region images,” *Soft Computing*, vol. 28, no. 1, pp. 437–445, Apr. 2024, doi: 10.1007/s00500-023-08209-6.

[27] X. Li, G. Zhu, S. Wang, Y. Zhou, and X. Zhang, “Deep reverse attack on SIFT features with a coarse-to-fine GAN model,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 7, pp. 6391–6402, Jul. 2024, doi: 10.1109/TCSVT.2024.3367808.

[28] A. Sivasubramanian, V. R. Prashanth, T. Hari, V. Sowmya, E. A. Gopalakrishnan, and V. Ravi, “Transformer-based convolutional neural network approach for remote sensing natural scene classification,” *Remote Sensing Applications: Society and Environment*, vol. 33, p. 101126, Jan. 2024, doi: 10.1016/j.rsase.2023.101126.

[29] T. Lu, L. Wan, S. Qi, and M. Gao, “Land cover classification of UAV remote sensing based on transformer–CNN hybrid architecture,” *Sensors*, vol. 23, no. 11, p. 5288, Jun. 2023, doi: 10.3390/s23115288.

BIOGRAPHIES OF AUTHORS

Kampa Ratna Babu received his Ph.D (C.S.E) Degree in the year 2015, M.Tech (Software Engineering) Degree in the year 2003 from Jawaharlal Nehru Technological University Hyderabad, Telangana, India and B.Tech (C.S.E) from Bapatla Engineering College, Nagarjuna University, Andhra Pradesh in the year 2000. He is currently working as Lecturer in Computer Engineering Department, M.B.T.S. Government Polytechnic, Guntur, Andhra Pradesh. His research interests are image processing and pattern Recognition. He has published more than 15 papers in International and National Journals and conferences. He is having 22 years of Teaching Experience. He is a life member of technical association ISTE. He can be contacted at email: kratnababu@gmail.com.

Kampa Kanthi Kumar received his Ph.D. (Electronics & Communication Engineering) from Jawaharlal Nehru Technological University Kakinada, Andhra Pradesh, India in the year 2019, M.Tech. (Computers and Communications) Degree from Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India in the year 2005 and B.Tech (Electronics and Communication Engineering) from Bapatla Engineering College, Nagarjuna University, Andhra Pradesh in the year 2002. He is currently working as a Professor in ECE, Tirumala Engineering College, Narasaraopet, Andhra Pradesh. His research interests are wireless communications and networks, computer networks, signal processing and image processing. He has published more than 32 papers in International and National Journals and conferences. He has 19 years of Teaching Experience. He is a life member of technical association ISTE, IETE, IEAE, IARAIAE. He can be contacted at email: kkanthik@gmail.com.

Akula Suneetha received her Ph.D. (C.S.E) Degree in the year 2022 from Acharya Nagarjuna University, Andhra Pradesh, India, M.Tech (C.S.E) Degree from Jawaharlal Nehru Technological University Hyderabad, Telangana, India in the year 2009 and B.Tech (C.S.E) from Sir C R Reddy College of Engineering, Eluru,Andhra University, A.P, India in the year 2005. She is currently working as Associate professor in the Department of Computer Science and Engineering, KKR and KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh affiliated to JNTUK University, A.P, India. Her research area is Fuzzy Image Processing. She has published various papers in International and National Journals, patent publications and conferences. She is having 18 years of Teaching Experience. She can be contacted at email: akulasuneetha25@gmail.com.