International Journal of Chemical Studies

P-ISSN: 2349–8528
E-ISSN: 2321–4902
IJCS 2019; 7(6): 1610-1620

Received: 06-09-2019
Accepted: 10-10-2019

MK Pradhan
Department of Statistics and Computer Science, IGKV, Raipur, Chhattisgarh, India

BL Sinha
College of Agriculture, Bhatapara, Chhattisgarh, India

S Ramole
Department of Statistics and Computer Science, IGKV, Raipur, Chhattisgarh, India

Corresponding Author:
MK Pradhan
Department of Statistics and Computer Science, IGKV, Raipur, Chhattisgarh, India

Extended differential pattern-based large scale live active learning model for classification of remote sensing data

MK Pradhan, BL Sinha and S Ramole

Abstract
Recently, the rapid growth of the space-oriented imaging techniques with the tremendous remote sensors facilitates the earth observation in the spatial and spectral domain. The Hyperspectral Images (HSI) have the ability to deliver the detailed information of earth in such domains. An accurate identification of objects from the acquisition system depends on the clear segmentation and classification. Traditionally, clustering and rules-based methods are adopted for classification according to the thresholding effect of image pixel intensity. The clustering process depends on the mean feature of the image pixels with the gray limit. With the spectral limitations, the multi-label segmentation problem affects the clustering and rules adversely. The variations in spectrum cause the changes in the number of rules each and every time that leads to computational complexity and misclassification. This paper proposes the novel classification method based on the textural information obtained from the Extended Differential Pattern (EDP). Initially, the Distributed Intensity Filtering (DIF) removes the noise present in the image and the application of Histogram Equalization (HE) enhances the image quality. The merging and classification of different labels for each image sample are performed through the Extended Differential Pattern (EDP) provides the textural information clearly. With these pattern set, the traditional active learning methods such as Relevance Vector Machine (RVM) and the multi-class SVM classify the HSI patterns that play the major role in remote sensing applications. The comparative analysis between the proposed EDP-AL with the existing algorithms regarding the various parameters overall accuracy, average accuracy and kappa statistics conveys the effectiveness of EDP-AL in remote sensing applications.

Keywords: Active learning, classification, hyperspectral images, multi-class support vector machine

Introduction
Computer vision applications depend on the good set of labeled image for deep analysis. The classification of images is the long history research area in such applications to facilitate the accurate object identification from the acquisition system. The prediction of categories and the location of objects is the fundamental research problem in the remote sensing applications. The creation of good quality training sets with the large size labeled samples requires the substantial human effort. In another aspect, the construction of robust classifier depends on the large size training samples. But, this is a time and cost consuming process. Hence, the trade-off between the best classification and the minimum training samples is the major requirement in the HSI classification.

The introduction of active learning methods [1] in the research supports the built up of the robust classifier with minimum labeling. But, the good classifier training with the minimal labeling cost is the critical task with machine learning algorithms.

The selection of either non-informative or redundant instances induces the difficulties in random selection of unlabeled instances to label it. Hence, the evaluation methods of informativeness of the labeled instances are applicable to reduce the human effort. The selection of informative regions in the HSI requires most uncertainty measures with the capture of relationship between the candidate instances and the classification model in real-time scenarios [2]. This causes the unusual instances for labeling. Hence, the development in active learning model should consider the representativeness of the candidate according to uncertainty features. The numerous developments PF the acquisition models increase the spatial-spectral and temporal resolution imaging systems drastically to improve the classification performance.

The data analysis and the accurate classification requires the implementation of several processing techniques. The developed classification techniques are grouped into two categories [3] such as unsupervised and supervised. First one deals with the division of image into clusters of pixels with the similar characteristics. Decompose of image into clusters is independent of the labeled information by the user. But, the lack of correspondence between the clusters makes the difficulties in the class separation. Alternatively, the establishment of link between the samples and classes explicitly provides the guarantee of accurate classification. But, the manual training sample collections is an obvious problem and thus leads to a semi-automatic training sample selection in which the most useful information is selected on the basis of ranking.

The dependency of samples with the class is based on their own attributes. The attributes of HSI classification are affected by two factors [4] as follows:

- Determination of reflectance spectra through the object and surrounding materials
- Spectral deviation due to the noise from the remote sensing devices.
- Environmental factors with viewing angle

Besides the spectral information, the spatial information such as shape, texture and profile features are necessary to measure the dependence of samples with the class. Morphological Profiles (MP), Support Vector Machine (SVM) and Gray Level Co-occurrence Matrix (GLCM) –based models convey that the spatial information plays the major role in the classification of high –resolution images in remote sensing data. With increase in data dimension, the number of spatial features corresponding to those images also more such as morphological features, GLCM features, wavelet-based features, object-based features and the structural features [5]. Hence, the prediction of optimal features or relevant features for better classification is the important task. The addition of more number of features and the hyperdimensional feature space require the multi-feature analysis models to improve the classification performance. The development of AL methods considered general purpose irrespective of the real annotation cost-oriented procedures that depend on the application. The better annotation process depends on the most informative samples selection from the large size information pools. The summary of traditional methods conveys that the novel learning models are required to provide the trade-off between the cost effectiveness and the accurate classification. This paper employs the suitable texture pattern extraction method to select the most useful information that governs the deep analysis. The technical contributions of proposed EDP-AL are listed as follows:

- The employment of Distributed Intensity Filtering (DIF) and the Histogram Equalization (HE) removes the noise and enhances the quality of images for clear depth information
- An Extended Differential Pattern (EDP)-based texture pattern extraction supports the merging and classification of labels under the fusion of spectrum bands
- The utilization of novel classifiers such as RVM and Multi-class SVM validates the performance of EDP-based texture pattern extraction in terms of classified samples.

The paper organized as follows: The detailed description of the related works on learning and classification models under spectral-spatial domains arediscussed in section II. The implementation process of Extended Differential Pattern (EDP)-based Active Learning (AL) is described in section III. The comparative analysis of EDP-AL with existing methods provided in section IV. Finally, the conclusions about the application of EDP-AL on the remote sensed data presented in section V.

Related work

This section discusses the review of traditional classification and learning models for HSI under the spectral-spatial variations. The uncertainty query selection-based procedures failed to extract the useful information from the large size unlabeled instances. *Li et al.* [2] presented the adaptive learning that combined the density and uncertainty measures together to select the critical instances for labeling. The validation of adaptive learning was performed through object and scene recognition. The formulation of AL framework dependent on both spectral and spatial domain. *Pasolli et al.* [3] utilized the iterative sample selection methodologies to integrate the spatial and spectral features. The explicit computation of Euclidean distance and the Parzen window-based spatial domain analysis included the spatial entropy to handle the high-resolution images. The widely used HSI classification method with the limited training samples called semi-supervised learning. *Ma et al.* [4] presented the novel semi-supervised classification on the basis of multi-decision labeling and deep feature learning. The exploitation of more information was the basic need of better classification. The simultaneous provision of plentiful spatial-spectral features increases the resolution of remotely sensed imagery. Huang *et al.* [5] constructed the SVM ensemble model that combined the multiple spatial-spectral features under both pixel and object levels. They proposed multi-feature SVMs (certainty voting, urban complexity index and the object-based semantic approach providing the accurate results.

An efficient object detection depends on the robustness to the appearance variations for the object. The trade-off between the detection with fewer models and the high operational speed was the major requirement in remote sensing applications. *Ohn-Bar et al.* [6] studied the efficient means for dealing with the intra-diversity in object detection. The employment of AdaBoost detection scheme with pixel look-up features improved the operational speed. The extraction of noticeable objects with limited computational resources considers the saliency detection as the major task in the field of vision community. *Wang et al.* [7] utilized the Manifold Ranking (MR) to reduce the limitations in the salient band selection methods. An access of hyperspetcral data structure rather than similarity rating through MR provided the solution to the ranking problem. The challenging task in the Active Learning (AL) was the selection of most informative samples from the data. *Babae et al.* [8] proposed the novel AL algorithm based on low-rank classifier as the training model and the visualization support data point selection called First Certain Wrong Labeled (FCWL). The composition of logistic regression loss function and the trace-norm learning parameters improved the classification performance. An automated selection of positive and negative samples of multiple parts requires the modification in AL approaches. *Stazoda et al.* [9] presented the Vehicle Detection using Active learning Symmetry (VeDAS) model that utilized the Haar-like features and the AdaBoost classifiers to detect fully visible parts. The integration of AL models with the semi-supervised methods reduces the effort of constructing training

sets for classification. *Polewski et al.* [10] enabled the efficient optimization model through the generalization of Deterministic Annealing Expectation Maximization (DAEM) algorithm for the application of Renyi-regularized model to reduce the expected error.

The replacement of human oracle with the tags representing image makes the AL method as an automated to handle the social context model. *Chatzillari et al.* [11] presented the Social Active Learning for Image Classification (SALIC) to select most appropriate tagged images that expand the training set necessary for complexity reduction. The joint maximization of informativeness and associated confidence level were achieved through the probabilistic framework. The susceptibility to noise or irrelevant features and the lack of statistical irregularities affected the prediction of the weight value of each kernel. *Li Li et al.* [12] addressed the limitations in Distance Metric Learning (DML) to learn the matrix without considering the weight of each class. They proposed the dual layer supervised Mahalanobis kernel for the classification of images. They utilized the SVM-classifier to classify the less-dimensionality data with better performance. Limited labeled samples and the spatial variability of spectral signatures were the major limitations in addition to the dimensionality. *Soomro et al.* [13] proposed the novel bilayer Elastic Net (ELN^2) regression model for classification of hyperspectral images with the spatial-spectral information exploitation. The novel bilayer ELN^2 contained two components namely spectral-only ELN and the spatial contextual driven ELN in two layers. The features extracted from the hyperspectral images represent the diverse characteristics and their combination has the positive impact on the classification performance. *Zhang et al.* [14] formulated the multi-feature HSI classification model as the joint sparse constraint model. They preserved spatial information and utilized the complementary information additionally supported the real-time applications. With the increase of features, suitable kernel model called Multiple Kernel Learning (MKL) was introduced to improve the classification performance. Pre-determination of parameters plays the major role in MKL and they require the prior knowledge about the results that leads to complexity. *Liu et al.* [15] concentrated on the embedding of Extended Multi-Attribute Profile (EMAP) in MKL specific model. They proposed the class specific algorithm that automatically learns the efficient feature set which is necessary for classification.

Real-annotation procedures and the costs were considered to minimize the number of samples to be labeled and added to the training set. The major assumption to implement annotation procedure was all the samples require the equal effort for labeling. *Persello et al.* [16] addressed the problem in active sample selection based on Markov Decision Process (MDP). Besides, they addressed the optimizing the collection of labeled samples and showed the effectiveness of the MDP on the forest inventory controlling. Collaborative Representation (CR) generated the non-sparse codes by using all the atoms leads to interferences. Sparse Representation (SR) selected few samples which cannot reflect within the class variations. Hence CR and SR were combined in the research studies to alleviate the problems. *Li et al.* [17] proposed the fusion of CR and SR on the basis of two following methods: Fusion Representation based Classification (FRC) and the Elastic Net Representation based Classification (ENRC). The convex combinations of penalties (l_1, l_2) used by the ENRC to achieve the balance between CR and SR models. *Arabi et al.* [18] employed definition of typical sets from the Asymptotic Equipartition Property (AEP) for HSI classification. Decompose of end members through Discrete Wavelet Transform (DWT) and the Hidden Markov Model (HMM) facilitated the better classification performance. The accuracy of HSI classification required further improvement which depends on the multi-hypothesis prediction model. *Chen et al.* [19] proposed the spectral band partitioning strategy on the basis of inter-band correlation coefficient to improve the representation power. The integration of spatial and spectral information modeled as the hypothesis predictions by Tikhonov regularization framework. Under small sample size constraints and noise corruption environment, the classification accuracy and the maximum likelihood were increased. The segmentation of Infrared images requires the prior information regarding the shape and appearance of dead trees. The lack of 3D height information leads to confusing dead trees with patches of ground areas. *Polewski et al.* [20] described the AL-based approach to detect the standing dead trees from the infrared imagery. The segmentation of individual trees within the 3D point cloud and the prediction of approximate bounding polygon for each tree within the image through greedy approximation improved the classification performance.

With the increase in the dimensionality of features, the prediction of optimal feature which is adaptable to all the images was the complicated task. *Chunsen et al.* [21] proposed the probabilistic weighed fusion method of multiple features for hyperspectral image classification. They conducted the dimensionality reduction and the feature extraction through minimum noise fraction. The non-linear mapping of input data to the high-dimensional feature space in kernel methods improved the classification performance. *Zhang et al.* [22] presented the spectral-spatial learning methods to integrate those information into the Group Sparse Coding (GSC) through clusters. The incorporation of kernel trick into GSC supported the capture of non-linear relationships. The multiple classifiers, lack of creating single models from the complex data structure and the ingenius design effort were the major limitations from the literature studies. *Zhang et al.* [23] presented the novel sparse ensemble learning algorithm with the integrated spatial and spectral features to develop the joint sparse representation models. Among various AL approaches, the Multi-View (MV)-based AL plays the major role in object detection in remote sensing applications. *Zhou et al.* [24] highlighted the features that affect the efficiency of MV-Al such as a number of samples to be reduced, the quick convergence of learning process, disagreement among the multiple views and the simple model without any assumptions. They employed singularity-based criterion to identify the most informative pixels from the remotely sensed data. The incorporation of disparate features from the several sources provide the diverse information for remote sensing data analysis. *Zhang et al.* [25] proposed the ensemble MKL model that included the multiple features from the multi-sensors-based HSI classification. *Wan et al.* [26] proposed Collaborative Active and Semi supervised Learning (CASSL) that combines the AL and Semi-Supervised Learning (SSL) to improve the learning performance compared to multiclass level uncertainty-enhanced cluster-based diversity (MCLU-ECBD) [27], locally linear embedding (LLE) with manifold Co-Regularization (LLE-mCR) [28] and CASSL-No Pseudo Label Verification (NoPLV) [26]. The adequate learning with less time consumption is an investigating issue in the HSI classification in recent researches. *Sun et al.* [29] discussed the Gaussian Process (GP)-Al approach with various versions

such as GP-full, GP-Init, GP-Random Selection (RS), GP-AL1, GP-AL2 and GP-AL3 heuristics. With the increase in feature size, the prediction of relevant features from the large size is the difficult task. This paper proposes the novel learning models based on the Extended Differential Pattern (EDP)-based texture pattern extraction to improve the classification performance.

Extended-differential pattern-active learning model

This section discusses the implementation details of proposed Extended Differential Pattern (EDP)-based Active Learning (AL) for remote sensing applications. The simultaneous achievement of large size training samples and better classification depends on following models in proposed work as shown in Fig. 1.

1. Distributed Intensity Filtering
2. Extended Differential Pattern
3. RVM classification

The low-quality images with the noise existence affect the depth information which plays the major role in hyperspectral image classification. Initially, the preprocessing through the Distributed Intensity Filtering (DIF) removes the noise present in the images and the integration of Histogram Equalization (HE) enhances the quality of image for clear depth information analysis. Then, the texture pattern information extraction is the important stage in proposed work. The Extended Differential Pattern (EDP) method extracts the necessary texture patterns which highly contribute to relevant information for analysis. With the extracted pattern set, the RVM classifier is used to classify the samples which are necessary for vision data analysis. The comparison between the proposed EDP-based AL models with the multi-class SVM models regarding the various parameters of accuracy, sensitivity, specificity, positive likelihood, negative likelihood and classification rate conveys the effectiveness of proposed EDP-AL models.

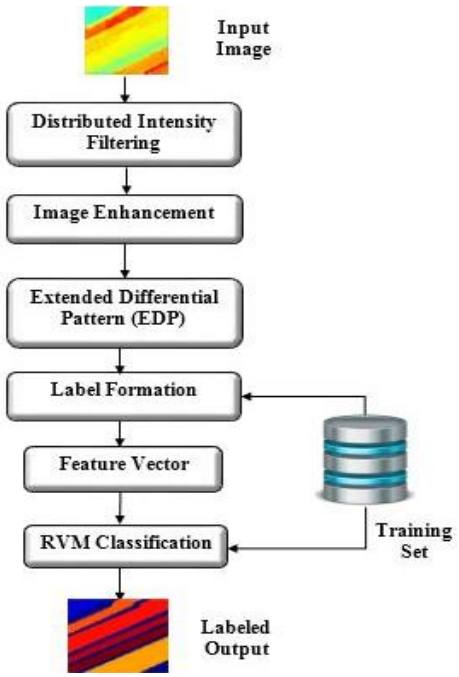


Fig 1: Workflow of proposed EDP-AL

Distributed intensity filtering

The noise present in the input image as shown in Fig. 2 affects the quality of edge information that leads to misclassification and limits the useful information prediction. To remove the noise present in the image, the window with the size of 3×3is formed to project the input image. Fig. 3 shows the image projection window.

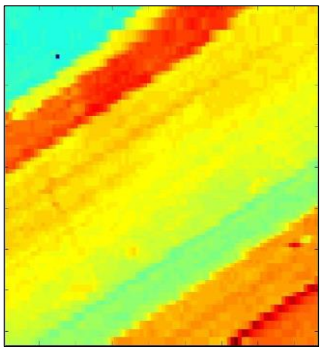


Fig 2: Input Image

W(1)	W(2)	W(3)
W(4)	W(5)	W(6)
W(7)	W(8)	W(9)

Fig 3: Projection Window

The Distributed Intensity Filtering (DIF) is proposed to remove the noise present in the image. The major processes in this filtering are listed as follows:

- Placing the neighborhood around the point to be analyzed
- Analyze the pixel intensities of neighborhood with the center value
- Replace the original pixel value with the analyzed result from the previous step.

Initially, the window is constructed for the image with the row values from $i-1\ to\ i+1$ and the column values from$j-1\ to\ j+1$. Then, the neighborhood moves over the each pixel in the image successively to predict the replacement value. The difference between the center pixel with the boundary is initially estimated and check whether such difference value is greater than the center pixel or not. If the condition is satisfied, then replace the pixel value by using the average value of window elements as follows:

$$I_p(i,j)=\frac{\sum W_{temp}}{n} \quad (1)$$

Where, $I_p(i,j)$= Preprocessed image

W_{temp} = $(W(x),'x' \neq center)$

n = Total number of neighborhoods

Distributed intensity filtering

Input: Load Hyper spectral Image 'I'

Output: Preprocessed Image, 'I_p'

Step 1: Initialize window size (3×3).

Step 2: for (i = 2 to$Row_Size\ (I)-1$) //'i' Row value

Step 3: for (j = 2 to$Column_Size\ (I)-1$) //'j' Column value

Step 4: W=$I_{i-1\ to\ i+1,j-1\ to\ j+1}$

//Project window over image matrix as, W

Step 5: if (($W\ (5) \sim W\ (Boundary)$) >$W\ (5)$)

//Check neighboring Pixel variation.

Step 6: Compute the preprocessed output

I_p (i, j) =$\frac{\sum W_{temp}}{n}$

Step 7: end if
Step 8: end loop 'j'
Step 9: end loop 'i'

Once the neighborhood values are replaced with the estimated value reduces the noise present in the image. Hence, the output from the DIF process contains less noise as shown in Fig. 4. The interpretation of information present in the image not only depends on the noise-free regions of image. The quality of the image is further enhanced for clear analysis of images. The Gaussian modelling is applied to enhance the quality of the input image. The standard deviation in the traditional Gaussian model [30] is reformulated with the Root Mean Square (RMS) value of the difference between each pixel with the overall pixel value as follows:

$$\sigma = \sqrt{\frac{1}{a*b}\sum_{i=1}^{a*b}\left(I_p(i) - \frac{\sum I_p}{a}\right)^2} \quad (2)$$

Where, a, b −Row and column size of the images

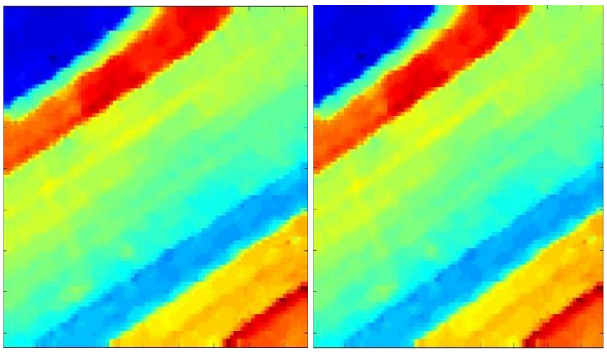


Fig 5: (a) Filtered Image (b) Enhanced Image

The quality of the image is enhanced with the modified standard deviations from equation (2) as follows

$$I_e = \frac{I_p}{\max\left(I_p\left(\frac{\sum I_p}{a}*\sigma\right)\right)} \quad (3)$$

The pixel values of original input image are normalized with the maximum mean values from the enhanced image to provide the good quality image necessary for future processing.

Extended differential pattern

In this stage, the window size is extended to 5×5 for projection of enhanced image. Then, the median value of projected image is computed. Initially, the window over the enhanced image is formed with the size of 5×5. Within this window, the cells with 3×3 is extracted separately. By applying the angle-based difference estimation, the rules required for the vector prediction is formed. The algorithm to compute the patterns in multi-angular form is listed in this section.

The magnitude value corresponding to the difference between the window formation (temp, temp1) are mathematically expressed as follows:

$$mag = \sqrt{\left(double\left(\begin{matrix}\left((temp(3,4) \sim temp(3,3)\right)^2\right) + \\ \left((temp(2,3) \sim temp(3,3)\right)^2\right)\end{matrix}\right)\right)} \quad (4)$$

The comparison between the center pixel with the neighboring pixels is performed and then the decimal coding is performed to extract the patterns. The multiplication is performed with two types of patterns $(Pt2, Pt1)$ that extract the relevant patterns. These relevant patterns are considered as the major role in classification.

Extended differential pattern

Input: Enhanced Image 'I_e'

Output: Texture pattern 'out'

S-1: Initialize 5×5 window matrix
S-2: Project window over the enhanced image (I_e)
For $(i = 3\ to\ (Row\_size) - 2)$
For$(j = 3\ to\ (Column_size) - 2)$
$temp = I_e(i^2 + j^2)$
S-3: Compute the median value for the window
$med = temp(3)$
S-4: Check the difference of center of pixel with the neighborhood
if$temp$ (2, 3)>=med&&$temp$ (2,4)>= med
$I_{gc}(1)$=1;
elseif$temp$ (2, 3)<med&&$temp$(2,4)>=med
$I_{gc}(2)$=2;
elseif $temp$ (2,3)<med&&$temp$ (2, 4)<med
$I_{gc}(3)$=3;
elseif$temp$(2,3)>=med&&$temp$ (2,4)<med
$I_{gc}(4)$=4;
Endif
S-5: Compute the magnitude value from newly formed window by using equation (4)
S-6: Compute the patterns $Pt1 = mag \times I_{gc}$
S-7: For$(i = 2$ to $(Row\_size) - 1)$
For$(j = 2$ to $(Column\_size) - 1)$
Assign the original image to the temporary variable
$temp1 = I_e\ (i, j)$;
S-8: Check the condition
$temp2(i-1, j-1)\ = I_e\ (i-1, j-1) > temp1$;
S-9: Compute the patterns
$Pt2 =\ temp2$
End Loop j
End Loop i
S-10: Perform the bitwise OR operation between two patterns
$out = Pt1.* Pt2$

Active learning

The process of learning with the special program to take the control over the various inputs required for training refers active learning. The major objective of such approaches is at selective input query against the numerous classifiers. The sample selection by using the AL is more discriminative compared to random sampling. The presence of region of uncertainty among the set of training samples is the core idea of the proposed work. The samples available more likely to be classified incorrectly in repetitive training stage. The major steps in the active learning framework are listed as follows:

- The annotation of positive and negative samples used for training is manually and such process refers passive learning
- The queries from the outputs generated by using the passive trained classifiers to get the true and false positive manually

With the above processing steps, the AL reduces the false detection rates while maintaining high detection rates.

Consider the list of labelled samples be $X = (x_i, y_i)_{i=1}^{l}$that mapped into the input space χ of dimension (d). Besides, the unlabeled samples are regarded as $U = (x_i)_{i=l+1}^{l+u}$ pool of candidates. The repeated feeding in classification model with the new labeled pixels improves the performance. The algorithmic steps for traditional AL are listed as follows:

1. Initialize the training sets and pool of candidates, number of pixels added to the classification model for each iteration
2. Train the model with the current training set
3. Compute the user defined heuristic for each candidate in candidate pool
4. According to the score of heuristic, rank is assigned to each candidate
5. Select the most interesting pixels corresponding to the rank values
6. Assign the label to the selected pixels
7. Add the batch to the training set
8. Remove that batch from the pool of candidates

The major requirement for the active learning is the interaction between the user and model. The provision of labelled information with the class-knowledge and the interpretation results of the distributed classes is the basic need for AL training framework. To complete the execution, the relevant pixels are needed which is the crucial task in the traditional framework. With the above processing steps, the AL reduces the false detection rates while maintaining high detection rates.

The patterns extracted from the EDP are assigned as the input values to the AL. The parameters used in AL are illustrated in Table I.

Table I

AL parameters

S. No	Variable	Parameter
1	A	Training Vectors
2	B	Texture patterns
3	X	Total number of classes
4	Y	Total number of attributes
5	θ	Angle between the input layer
6	ϑ	Hidden network layer
7	μ	Input network layer
8	λ	Angle between the hidden layer
9	R	RMS value
10	N	Number of RMS value
11	δ	Classifier coefficient

The exponential function for classified output is initialized with the following metrics

$$H = \frac{(e^{\mu} - e^{-\mu})}{(e^{\mu} + e^{-\mu})} \quad (5)$$

Where, $\mu = \beta * X - \theta$

The angle between the input layer and the hidden layer are related with the following formulation

$$\vartheta = H * Y - \lambda \quad (6)$$

The RMS value necessary to update the classes and attributes are estimated by using following formulation

$$R = \frac{\sqrt{(\alpha - \mathrm{q}*\wedge 2)}}{\mathrm{q}} \quad (7)$$

Where, q= Exponential form of angle update=$\frac{(e^{\vartheta} - e^{-\vartheta})}{(e^{\vartheta} + e^{-\vartheta})}$

The update of classes, attributes and the angles between the layers are sequentially update through the following formulations for all the extracted patterns as follows

$$X = X + \delta * (\beta(n))' * (1 + H) * (1 - H1)*(\Delta \mathrm{q} + Y') + X\rho \quad (8)$$

$$\lambda = \lambda + (-\delta * \Delta \mathrm{q}) + \lambda\rho \quad (9)$$

$$\theta = \theta + -\delta * (1 + H) * (1 - H1)*(\Delta \mathrm{q} + Y') + \theta\rho \quad (10)$$

$$\mathrm{Y} = \mathrm{Y} + \delta * (H' * \Delta \mathrm{q}) + \mathrm{Y}\rho$$

The update values $(X\rho, \lambda\rho, \mathrm{Y}\rho \; and \; \theta\rho)$ are formulated as follows

$$\mathrm{X}\rho = R * \delta * (\beta(n))' * (1 + H) * (1 - H1)*(\Delta \mathrm{q} + Y') \quad (11)$$

$$Y\rho = R * \delta * (H' * \Delta \mathrm{q}) \quad (12)$$

$$\theta\rho = R * -\delta * (1 + H) * (1 - H1)*(\Delta \mathrm{q} + Y') \quad (13)$$

$$\lambda\rho = R * (-\delta * \Delta \mathrm{q}) \quad (14)$$

The distance between the RMS and the count value is constructed for previous and the new update values. The region of classification for the two feature set corresponding to different classes as shown in Fig. 6. From the distance values, the minimum value of distance is computed and the corresponding label is assigned to the input image.

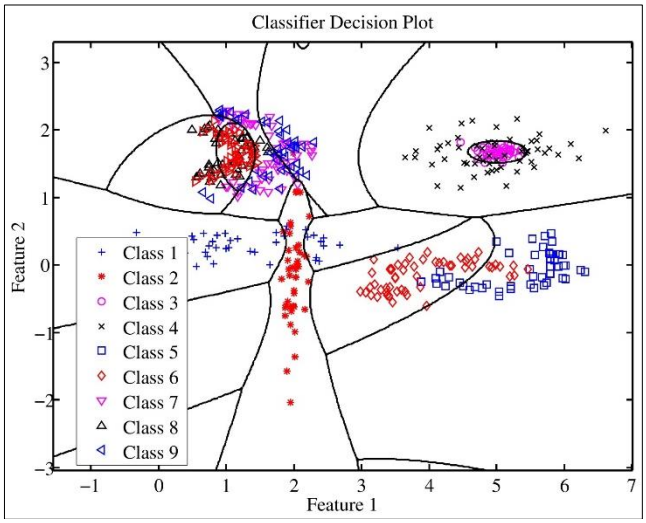


Fig 6: Decision Plot for Classification

The results from the AL process train the Relevance Vector Machine (RVM) to improve the classification performance. The probabilistic sparse kernel that adopts Bayesian approach for learning the over fitting samples refers Relevance Vector Machine (RVM). The number of predictions from the RVM is based on the function described as

$$y(x) = \sum_{n=1}^{N} \omega_n K(x, x_n) + \omega_0 \quad (15)$$

The function defined in (6) represents the relationship between the model weights (ω_n) and kernel function $K(.,.)$ in terms of input samples. The non-associated input samples with the non-zero weights are close to the decision boundary refers "relevance" vectors. The prediction of posterior membership for the given class and the optimal solution are the objectives of the RVM [31]. The logistic sigmoidal function generalizes the linear model and the corresponding likelihood is computed for the class instances (c) as

$$P(c/w) = \prod_{i=1}^{n} \sigma\{(y(x_i)\}^{c_i}[1 - \sigma\{(y(x_i)\}]^{1-c_i} \quad (16)$$

Where, $\sigma(y)$ is the logistic sigmoid function

$$\sigma(y(x)) = \frac{1}{1+\exp(-y(x))} \quad (17)$$

The most probable weights computation, iterative reweighted least square algorithm utilization and Gaussian approximation are repeated until the convergence criteria is satisfied.

Performance analysis

This section illustrates the performance analysis of the proposed EDP-AL regarding the sensitivity, specificity, accuracy, precision, and recall. Besides the comparative analysis of proposed EDP-AL with the existing SVM [32] and Class Level Joint Sparse Representation Classifier (CL-JSRC) [14] and Probabilistic weighed strategy [21] in hyper spectral image analysis.

Dataset

There are two data sets such as Pavia University and Indian Pines hyperspectral data sets are used to validate the performance of proposed EDP-AL. The Pavia University data set was collected by the Refl ective Optics Spectrographic Imaging System (ROSIS) sensor, which has 610 × 340 pixels and 103 spectral bands ranging from 0.43 to 0.86 μm. The spatial resolution is 1.3 m. The Indian Pines data set is acquired by the AVIRIS sensor. The data set has 145 × 145 pixels, with 220 spectral bands ranging from 0.4 to 2.5 μm and the spatial resolution of 20 m. the information classes and the labelled samples for Pavian university and Indian Pines are listed in Table II and Table III.

Table 2: Information classes and number of labeled samples (Pavia University)

Class	Train	Test
Asphalt	310	6206
Meadows	806	16123
Gravel	94	1880
Trees	146	2933
Metal	67	1345
Bare Soil	251	5029
Bitumen	66	1330
Bricks	184	3682
Shadow	47	947
Total	1971	39475

Table 3: Information classes and number of labeled samples (Indian Pines)

Class	Train	Test
Alfalfa	27	54
Corn-notill	50	1434
Corn-Min	50	834
Corn	50	234
Grass/Pasture	50	497
Grass/Trees	50	747
Grass-mowed	13	26
Hay-windrowed	50	489
Oats	10	20
Soybeans-notill	50	968
Soybeans-Min	50	2468
Soybeans-clean	50	614
Wheat	50	212
Woods	50	1294
Bldg-grass-drives	50	380
Stone-steel-towers	50	95
Total	700	10366

Performance metrics

The performance validation of the proposed EDP-AL on the basic parameters and the comparative analysis of the existing SVM [32] and the EDP-AL assures the effectiveness of the extended differential pattern-based classification. Table II presents the comparison of performance metrics for proposed EDP-AL and the existing method of SVM. The comparative analysis between SVM and EDP-AL shows that the sensitivity, specificity, accuracy, precision and recall values of proposed EDP-AL are 8.96, 0.6, 8.97, 9.07 and 8.97 % respectively.

Table 4: Performance Analysis

Parameters	SVM	EDP-AL
TP	2214	2432
TN	34777	34998
FP	256	35
FN	253	35
Sensitivity (%)	89.7446	98.5813
Specificity (%)	99.2693	99.9001
Precision (%)	89.6356	98.5813
Recall (%)	89.7446	98.5813
Jaccard Coeff	98.6427	99.8133
Dice Overlap	99.3167	99.9066
Kappa Coeff.	0.883	0.984
Accuracy (%)	89.76	98.6

Classification accuracy and kappa coefficient analysis

The comparative analysis between the proposed PCAL with the existing methods of SVM-DMP [14], SRC-DMP [14] JSRC-DMP [14], Raw [21], MNF [21] and VS-SVM [21] regarding the classification accuracy and Kappa coefficient as shown in Table V and VI illustrate the variations of both classification accuracy (for each class of the dataset) and Kappa coefficient. In existing methods, the JSRC-DMP and VS-SVM provided better results in classification accuracy and kappa coefficient for each class of image.

Table 5: Accuracy and Kappa Coefficient Analysis (Indian Pines)

Class	SVM-DMP [14]	SRC-DMP [14]	JSRC-DMP [14]	RAW [21]	MNF [21]	VS-SVM [21]	EDP-AL
1	82.75	83.14	85.1	82.93	68.85	100	97.92
2	83.48	87.85	90.92	60.66	73.99	94.26	97.8
3	87.83	89.18	86.74	41.07	53.99	91.39	99.96
4	91.35	88.92	87.34	31.82	55.76	82.65	99.96
5	92.22	93.41	91.36	59.13	80.39	96.77	100
6	96.11	94.36	92.98	88.29	96.3	99.59	100
7	92.5	97.08	81.67	96.3	100	100	100
8	97.16	97.13	95.54	97.1	99.35	100	99.96
9	51.58	56.32	48.95	63.64	100	100	100
10	71.64	83.48	86.83	61.32	61.83	88.54	100
11	90.22	90.51	96.17	78.29	83.24	97.42	100
12	73.46	78.78	79.78	45.29	56.86	97.93	100
13	97.61	97.91	98.61	88.44	97.14	99.66	100
14	97.99	98.19	98.9	89.99	93.34	100	100
15	94.93	96.45	88.53	56.28	70.36	94.88	100
16	78.11	79.56	74.67	98.89	95.7	97.84	100
Kappa Coeff	86.65	89.21	90.71	62.5	72.92	95.16	96.42

Table 6: Accuracy and Kappa Coefficient Analysis (Pavia University)

Class	SVM-DMP [14]	SRC-DMP [14]	JSRC-DMP [14]	RAW [21]	MNF [21]	VS-SVM [21]	EDP-AL
1	93.77	84.41	87.95	81.99	84.86	92.12	99.68
2	97.35	97.09	97.89	94.22	84.5	99.56	98.4
3	65.04	56.76	61.9	68.11	74.32	85.65	98.72
4	93.7	90.64	93.75	79.92	75.06	98.24	97.97
5	72.91	83.9	89.9	97.94	99.55	99.7	97.49
6	81.84	64.12	71.66	65.43	78.58	94.43	97.01
7	65.28	75.05	77.43	67.85	82.72	90.45	96.53
8	89.35	72.21	79.21	67.79	78.9	92.34	96.05
9	69.03	84.02	89.21	100	100	100	95.57
Kappa Coeff	86.7	80.28	84.65	76.76	77.89	94.72	95.87

But, the differential pattern in proposed work extract the relevant patterns from the diverse patterns that improve the classification accuracy and coefficient value further.

Acceptance /rejection rate analysis

The number of incorrect labeling for each unauthorized user attempt and the rejection are defined by two metrics called False Acceptance Rate (FAR) and False Rejection Rate (FRR). The mathematical formulations of FAR, FRR, and GAR for facial expression recognition is listed as follows:

$$FAR = (False\ claims\ acceptance)/(Total\ claims) \times 100 \quad (15)$$

$$FRR = (False\ claims\ rejection)/(Total\ claims) \times 100 \quad (16)$$

$$GAR = 100 - FAR \quad (17)$$

Table VII shows the variations of FAR, FRR and GAR for proposed EDP-AL on various classes.

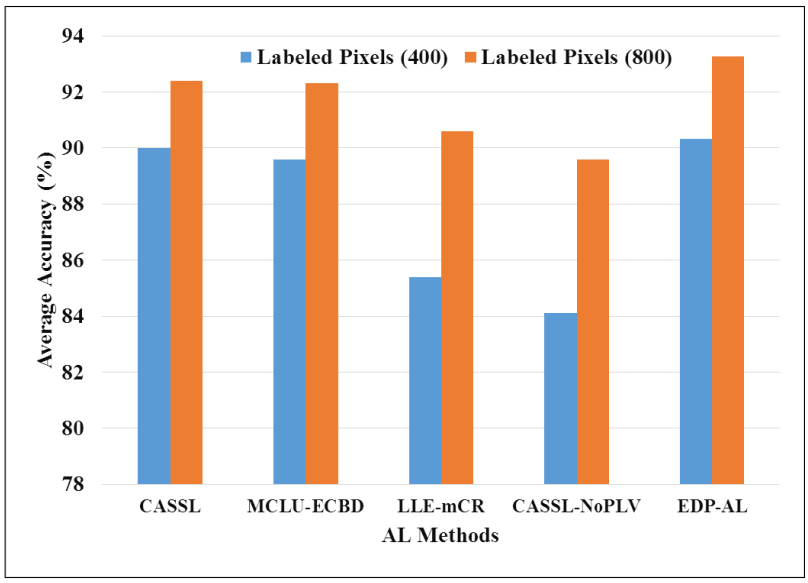
Table 7: Acceptance/Rejection Rate Analysis (Indian Pines)

Class	FRR	FAR	GAR
1	0.48	0.92	99.52
2	0	0.48	100
3	0.04	0	99.96
4	0	0	100
5	0	0	100
6	0	0	100
7	0	0	100
8	0.04	0	99.96
9	0	0	100
10	0	0	100
11	0.84	0	99.16
12	0	0	100
13	0	0	100
14	0	0	100
15	0	0	100
16	0	0	100

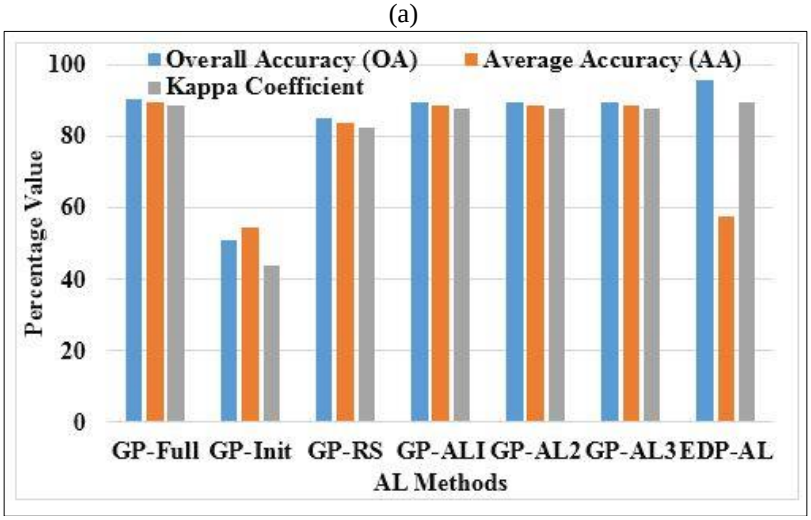
From the Table VI, it is observed that the proposed EDP-AL offers the improved GAR with minimum FAR and FRR rates due to the extended texture features.

Accuracy analysis with existing AL approaches

The analysis of Average Accuracy (AA) with the proposed EDP-AL and existing methods of CASSL [26], MCLU-ECBD [27], LLE-mCR [28] and CASSL-NoPLV [26] respectively. Besides, the comparative analysis of Overall Accuracy (OA), AA and Kappa coefficient for proposed EDP-AL with the existing GP-based AL versions [29] also suggest the effectiveness of proposed EDP-AL in remote sensing applications.



(a)



(b)

Fig 7: (a) Average Accuracy Analysis for Pavia University labelled Samples (400 and 800) and (b) Overall and Average Accuracy analysis for Indian Pines

The mathematical formulation of Kappa coefficient (in terms of %) is described as follows:

$$Kappa\ Coefficient\ (\%) = \frac{OA - AA}{100 - A}$$

Fig. 7 shows the variations of OA, AA and Kappa statistics for proposed and existing AL methods on the PU and IP. The AA of proposed EDP-AL is 93.2545 % which is better than the existing methods. Similarly, the OA, AA and Kappa statistics for EDP-AL are 95.71, 57.83 and 89.82 % which are better than the existing GP-based AL methods.

Conclusion and future work

This paper addressed the limitations in HSI classification and provided the solutions to them through the combination of texture-pattern and active learning framework. The major limitations in the HSI classification are spectrum variations, more number of rules and the multi-label segmentation problems. This paper proposed the novel methods to improve the classification performance against the various factors. Initially, the DIF removed the noise present in the image and the application of Histogram Equalization (HE) enhanced the image quality. The merging and classification of different labels for each image sample is performed through the EDP provided the textural information clearly. With these pattern set, the traditional active learning methods such as RVM and the multi-class SVM classified the HSI patterns that play the major role in remote sensing applications. The comparative analysis between the proposed EDP-AL with the existing algorithms regarding the various parameters conveyed the effectiveness of EDP-AL in remote sensing applications. The merging of textural patterns with the clear edge matching among the various bands through extended EDP will be

considered as the future work to improve the classification performance.

References

1. Vijayanarasimhan S, Grauman K. Large-scale live active learning: Training object detectors with crawled data and crowds. International Journal of Computer Vision. 2014; 108:97-114.
2. Li X, Guo Y. Adaptive active learning for image classification, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, 859-866.
3. Pasolli E, Melgani F, Tuia D, Pacifici F, Emery WJ. SVM active learning approach for image classification using spatial information. IEEE Transactions on Geoscience and Remote Sensing. 2014; 52:2217-2233.
4. Ma X, Wang H, Wang J. Semisupervised classification for hyperspectral image based on multi-decision labeling and deep feature learning. ISPRS Journal of Photogrammetry and Remote Sensing. 2016; 120:99-107.
5. Huang X, Zhang L. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. IEEE transactions on geoscience and remote sensing. 2013; 51:257-272.
6. Ohn-Bar E, Trivedi MM. Learning to detect vehicles by clustering appearance patterns. IEEE Transactions on Intelligent Transportation Systems. 2015; 16:2511-2521.
7. Wang Q, Lin J, Yuan Y. Salient Band Selection for Hyperspectral Image Classification via Manifold Ranking. IEEE transactions on neural networks and learning systems. 2016; 27:1279-1289.
8. Babaee M, Tsoukalas S, Rigoll G, Datcu M. Visualization-based active learning for the annotation of SAR images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2015; 8:4687-4698.
9. Satzoda RK, Trivedi MM. Multipart Vehicle Detection Using Symmetry-Derived Analysis and Active Learning. IEEE Transactions on Intelligent Transportation Systems, 2016; 17:926-937.
10. Polewski P, Yao W, Heurich M, Krzystek P, Stilla U. Combining Active and Semisupervised Learning of Remote Sensing Data Within a Renyi Entropy Regularization Framework.
11. Chatzilari E, Nikolopoulos S, Kompatsiaris Y, Kittler J. SALIC: Social Active Learning for Image Classification.
12. Li L, Sun C, Lin L, Li J, Jiang S. A dual-layer supervised Mahalanobis kernel for the classification of hyperspectral images. Neuro computing. 2016; 214:430-444.
13. Soomro BN, Xiao L, Huang L, Soomro SH, Molaei M. Bilayer Elastic Net Regression Model for Supervised Spectral-Spatial Hyperspectral Image Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2016; 9:4102-4116.
14. Zhang E, Jiao L, Zhang X, Liu H, Wang S. Class-Level Joint Sparse Representation for Multifeature-Based Hyperspectral Image Classification.
15. Liu T, Gu Y, Jia X, Benediktsson JA, Chanussot J. Class-Specific Sparse Multiple Kernel Learning for Spectral–Spatial Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing. 2016; 54:7351.
16. Persello C, Boularias A, Dalponte M, Gobakken T, Naesset E, Schoelkopf B. Cost-sensitive active learning with lookahead: Optimizing field surveys for remote sensing data classification. IEEE Transactions on Geoscience and Remote Sensing. 2014; 52:6652-6664.
17. Li W, Du Q, Zhang F, Hu W. Hyperspectral Image Classification by Fusing Collaborative and Sparse Representations. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2016; 9:4178-4187.
18. Arabi SYW, Fernandes D, Pizarro MA, da Silva Pinho M. Hyperspectral Images Classification with Typical Sequences associated to the Endmember. IEEE Latin America Transactions. 2016; 14:3102-3109.
19. Chen C, Li W, Tramel EW, Cui M, Prasad S, Fowler JE. Spectral–spatial preprocessing using multihypothesis prediction for noise-robust hyperspectral image classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2014; 7:1047-1059.
20. Polewski P, Yao W, Heurich M, Krzystek P, Stilla U. Active learning approach to detecting standing dead trees from ALS point clouds combined with aerial infrared imagery. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, 10-18.
21. Chunsen Z, Yiwei Z, Chenyi F. Spectral–Spatial Classification of Hyperspectral Images Using Probabilistic Weighted Strategy for Multifeature Fusion. IEEE Geoscience and Remote Sensing Letters. 2016; 13:1562-1566.
22. Zhang X, Song Q, Gao Z, Zheng Y, Weng P, Jiao L. Spectral–Spatial Feature Learning Using Cluster-Based Group Sparse Coding for Hyperspectral Image Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2016; 9:4142-4159.
23. Zhang E, Zhang X, Jiao L, Li L, Hou B. Spectral–spatial hyperspectral image ensemble classification via joint sparse representation. Pattern Recognition, 2016.
24. Zhou X, Prasad S. Crawford, Wavelet-Domain Multiview Active Learning for Spatial-Spectral Hyperspectral Image Classification, MM.
25. Zhang Y, Yang HL, Prasad S, Pasolli E, Jung J, Crawford M. Ensemble multiple kernel active learning for classification of multisource remote sensing data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2015; 8:845-858.
26. Wan L, Tang K, Li M, Zhong Y, Qin A. Collaborative active and semisupervised learning for hyperspectral remote sensing image classification. IEEE Transactions on Geoscience and Remote Sensing. 2015; 53:2384-2396.
27. Demir B, Persello C, Bruzzone L. Batch-mode active-learning methods for the interactive classification of remote sensing images. IEEE Transactions on Geoscience and Remote Sensing. 2011; 49:1014-1031.
28. Di W, Crawford MM. Active learning via multi-view and local proximity co-regularization for hyperspectral image classification. IEEE Journal of Selected Topics in Signal Processing. 2011; 5:618-628.
29. Sun S, Zhong P, Xiao H, Wang R. Active learning with Gaussian process classifier for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing. 2015; 53:1746-1760.
30. Alam MS, Islam MN, Bal A, Karim MA. Hyperspectral target detection using Gaussian filter and post-processing. Optics and Lasers in Engineering. 2008; 46:817-822.

31. Bai L, Hu D, Hui M, Li Y. RVM Classification of Hyperspectral Images Based on Wavelet Kernel Non-negative Matrix Fractorization. Telkomnika (Telecommunication Computing Electronics and Control). 2015; 13:976-984.
32. Cui Y, Wang J, Liu SB, Wang LG. Hyperspectral Image Feature Reduction Based on Tabu Search Algorithm. Journal of Information Hiding and Multimedia Signal Processing, 2015, 154-162.