VOL. 10, NO. 4, MARCH 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

FEATURE SELECTION BASED HYBRID CLASSIFICATION ALGORITHM WITH EMBEDDED ZERO TREE WAVELET

T. Karthikeyan and P. Manikandaprabhu
Department of Computer Science, PSG College of Arts & Science, Coimbatore, India
E-Mail: manipsgphd@gmail.com

ABSTRACT

With the development of the remote sensing imaging systems and hyperspectral sensors, the use of hyperspectral image is becoming more interesting. Remote sensing classification is a difficult procedure and requires thoughtfulness of lots of factors. The most important process of image classification may comprise resolve of a fitting classification system, selection of training samples, image pre-processing, feature extraction and accuracy assessment. Land cover information plays an important role in sustainable management, development and exploitation of resources, environmental protection, scientific analysis, modeling, monitoring and planning. Feature extraction recognizes and extracts remarkable features for a challenging task in order to decrease the complication of processing. Embedded Zero tree wavelet (EZW) are used for feature extraction process. Multi-layer perceptron (MLP) is recognized as the best ANN used in classification. The main aim of feature selection is to determine a minimal feature subset from a problem domain while retaining a suitably high accuracy in representing the original features. Particle Swarm Optimization (PSO) is a population based optimization technique used for feature selection. This paper mainly focuses on land cover image classification using EZW algorithm. Here, MLP is a classifier and PSO used for Feature selection process.

Keywords: embedded zero tree wavelet, feature extraction, feature selection, hyper spectral image, image classification, multi-layer perceptron, particle swarm optimization, remote sensing.

1. INTRODUCTION

Hyperspectral imagery has very high spectral resolution, and the plentiful spectral information can be used to provide finer classification that could not be achieved by traditional multispectral imagery [1, 2]. With the development of the remote sensing imaging systems and hyperspectral sensors, the use of hyperspectral image is becoming more interesting. A hyperspectral image consists of a large number of very narrow contiguous spectral bands. The number of bands can vary from tens to several hundreds and usually cover visible through middle infrared spectral images.

However, hyperspectral remote sensing image has many common problems such as large amount of data, the higher proportion of mixed pixels, and low spatial resolution, thus the improvement of classification speed and accuracy will be limited if ground objects are classified simply using hyperspectral images. New challenges and opportunities in Remote Sensing Image (RSI) classification have emerged due to the recent advances in sensor technologies [3]. Hyper spectral images suffer from the disadvantages of requiring very high computational and storage/transmission bandwidth because of the large quantity of data involved.

Land cover is the physical material at the surface of the earth. It includes grassland, asphalt, trees, bare soil, concrete, etc. Land cover information plays an important role in sustainable management, development and exploitation of resources, environmental protection, scientific analysis, modeling, monitoring and planning. The data become even more essential when there are rapid changes on the Earth's surface due to dynamic human activities as well as natural factors. Remotely sensed data in particular satellite images, among different advantages such as huge repetitive competencies, several spectral bands or multiple frequency/polarization are more effective tools for land cover mapping and they have been applied extensively for land cover monitoring and classification. Therefore, the challenging tasks are to understand the contribution of each dataset to select the most useful input features and to determine the combined datasets which can maximize the benefits of multi-source remote sensing data and to give the highest classification accuracy [4]. However, limited research has explored ways to determine variables from multi-source data in order to increase the classification accuracy [5].

Image classification automatically assigns an unknown image to a category according to its visual content, which has been a major research direction in computer vision. Classification is a decision making task of human activity. Classification problems occur when an object is assigned to a predefined group/class depending on many related observed attributes. Remote sensing classification is a difficult procedure and requires thoughtfulness of lots of factors. The most important process of image classification may comprise resolve of a fitting classification system, selection of training samples, image pre-processing, feature extraction and accuracy assessment. Before classification, the land cover types in the study area were defined with the help of a land use map produced. The main land cover types are forest, oil palm, urban area, rubber and water bodies. Two types of classification techniques have been applied for analyzing the accuracy of land use map by using corrected satellite imagery. Training phase is the mainly essential component in supervised classification as it might manipulate the absolute classification results.

ARPN Journal of Engineering and Applied Sciences

Land use and land cover are the basic assumption to identify the global ecology or environmental changes. Land use/ land cover mapping is a vital component, which has various parameters that are integrated on the basis of requirement. Land cover refers to a water body, cultivated land, built-up, natural vegetation, fallow land, glacial, rock/soil, artificial cover and others observed on the land. Global Information monitoring of land use and land cover is possible due to remote sensing technology in the form of spatial, spectral and temporal resolution. Remote sensing technology has many important roles, like reduction of survey time, latest map availability, more economic, digital image classification (pixels), spectral information etc. In remote sensing applications, Image classification is an important part. The most important task is Classification of land cover and one the analysis of remotely sensed data is one of the primary objectives [6].

This paper is organized as follows. The section 2 deals with the related work. The section 3 elaborates with feature extraction process followed by the Multilayer Perceptron classification algorithms. The section 5 deals with Feature Selection method. The Section 6 discusses with the experimental results and followed by a conclusion.

2. RELATED WORK

Hyperspectral remote sensing provides very high spectral resolution image data and the potential for discrimination of subtle differences in ground covers [7]. However, the high-dimensional data space generated by hyperspectral sensors introduces new challenges in the development of data analysis techniques [8], [9]. Improving land use/cover classification accuracy is an important issue in remote sensing literature. Moreover, advanced classification algorithm including Neural Network and Support Vector Machine approach instead of conventional classification method has been developed recently. The SVM algorithm has been used widely for pattern recognition applications. Many researchers in this field have found that a higher level of accuracy can be achieved by SVM than other processes of classification like MLC, artificial neural network (NN) etc. [10-13].

Ongoing research in the application of ensemble classification to land cover mapping has focused on the different ways ensembles can be constituted [14-16]. Some of the common approaches [17-21] have involved constituting ensembles using different classification algorithms, constituting base classifiers from using different training data, or deriving base classifiers using different band combinations (ensemble feature selection) [22]. Bischof et al. [23] describes multi spectral classification of land-sat images using neural networks. Heerman et al. [24] has been used the back propagation Neural network for the classification of multispectral remote sensing data. Hepner et al. [25] have given a comparison of conventional supervised classification by using minimal training set in Artificial Neural Network.

Previous research has demonstrated that high-dimensional data spaces are mostly empty, indicating that the data structure involved exists primarily in a subspace [26]. As a result, there is a need for feature extraction methods that can reduce the dimensionality of the data to the right subspace without losing the original information that allows for the separation of classes. In other words, dimension reduction is the transformation that brings data from a high order dimension to a low order dimension, thus overcoming the “curse” of dimensionality [27].

But classification accuracy poses serious challenge and this is due to, the design procedure of classifier, choice of training sets from dataset and information conveyed to the algorithm [28]. Statistical based classifiers have been successfully applied to multispectral data but are not effective for hyperspectral data [29]. The major reason is the fact that the number of spectral bands in hyperspectral data is too large, relative to the training samples. An effective way to solve this problem is to reduce the dimension of the hyperspectral images. This can be done by extracting a number of salient features of the hyperspectral data [30, 31].

However, a hyperspectral image has very strong spectral correlation, so dimensionality reduction through spectral feature selection is a necessary preprocessing step to obtain an informative but a compact set of spectral features. The main purpose of this study is to observe the accuracy improvement of algorithm hybridization and to select which combinations among candidate algorithms can provide the best improvement in land cover image data.

3. PROPOSED METHODOLOGY

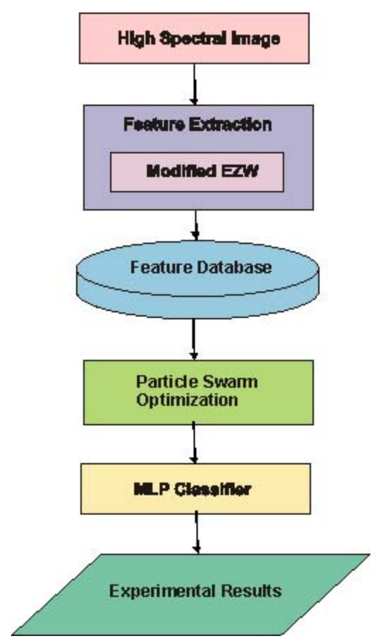

Figure-1. Proposed methodology.

Here, we list out the Proposed Methodology shown in Figure-1.

VOL. 10, NO. 4, MARCH 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

1. Get the Image Dataset like Satellite Land Cover Image.
2. To extract the Image Features like Texture, Shape and Spectral features data using Embedded Zero Tree Wavelet.
3. To store these feature into a dataset.
4. Make both training and testing set phases for data discrete.
5. Reduce the Feature Data set using Particle Swarm Intelligence.
6. Apply Multilayer Perceptron classification algorithms.
7. Find the Experimental Accuracy Result Based on Classification.

4. FEATURE EXTRACTION

Image contains information in a very dense and complex form, which a human eye, after years of training, can extract and understand. The main goal is to extract from an image a set of composing objects or real life attributes. This information is inferred from low level physical and mathematical properties of the image using a complex model of the reality the image reproduces. Image is described by several image features such as color, texture, shape or combination of these features with appealing tie frequency localization and multi-scale properties. A feature is a characteristic element that differentiates one class from other and the method of transforming the input data into the set of features is called feature extraction.

Feature extraction recognizes and extracts remarkable features for a challenging task in order to decrease the complication of processing. It is basically a case of reducing the dimension of the object. By reducing the dimension means, that the data in higher dimensional space are transformed into data in a few lower dimensions. Once a feature extraction is performed the input data is transformed to give a set of features collectively called as the feature vector. This feature vector is supposed to extract the relevant information from the data set to serve our needs.

4.1. Embedded zero tree wavelet (EZW)

The Embedded Zero tree Wavelet (EZW) coding is a simple, effective progressive image coding algorithm and can be worn for both lossless and lossy compression systems. This algorithm works well with the proposed coding scheme because the zero tree structure is effective in describing the significance map of the transform coefficients, as it exploits the inherent self-similarity of the subband image over the range of scales, and the positioning of majority of zero valued coefficients in the higher frequency subbands. The EZW algorithm applies Successive Approximation Quantization in order to provide multi-precision representation of the transformed coefficients and to facilitate the embedded coding. The algorithm codes the transformed coefficients in decreasing order in several scans. Each scan of the algorithm consists of two passes: significant map encoding and refinement pass.

The dominant pass scans the subband structure in zigzag, right-to-left and then top-to-bottom within each scale, before proceeding to the next higher scale of subband structure as presented in Figure-2. For each and every pass, a threshold (T) is chosen against which all the coefficients are measured and encoded as one of the following four symbols,

- Significant positive – If the coefficient value is greater than threshold T
- Significant negative – If the magnitude of the coefficient value is greater than threshold T
- Zero tree root – A coefficient is encoded as zero tree root if the coefficient and all its descendants are insignificant with respect to threshold T
- Isolated zero – If the coefficient is insignificant but some of its descendants are significant.

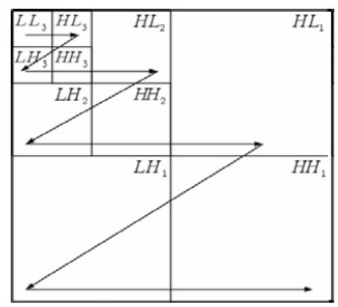

Figure-2. EZW subband structure scanning order.

In the embedded zero tree wavelet coding strategy, developed by Shapiro [54], a wavelet/subband decomposition of the image is performed. The wavelet coefficients/pixels are then grouped into Spatial Orientation Trees. The magnitude of each wavelet coefficients/pixels in a tree, starting with the root of the tree, is then compared to a particular threshold T. If the magnitude of all the wavelet coefficients/pixels in the tree are smaller than T, the entire tree structure (that is the root and all its descendant nodes) is coded by one symbol, the zero tree symbol ZTR. If however, there exit significant wavelet coefficients/pixels, then the tree root is coded as being significant or insignificant, if its magnitude is larger than or smaller than T, respectively. The descendant nodes are then each examined in turn to determine whether each is the root of a possible sub zero tree structure, or not. This process is carried out such that all the nodes in all the trees are examined for possible sub zero tree structures.

The significant wavelet coefficients/pixels in a tree are coded by one of two symbols, POS or NEG, depending on whether their actual values are positive or negative, respectively. The process of classifying the pixels as being ZTR, IZ, POS, or NEG is referred to as the dominant pass in [55]. This is then followed by the subordinate pass in which the significant wavelet coefficients/pixels in the image are refined by determining whether their magnitudes lie within the intervals (T, 3T/2) and

VOL. 10, NO. 4, MARCH 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

(3T/2,2T). Those wavelet coefficients/pixels whose magnitudes lie in the interval (T, 3T/2) are represented by a 0 (LOW), whereas those with magnitudes lying in the interval (3T/2,2T) are represented by a 1 (HIGH). Subsequent to the completion of both the dominant and subordinate passes, the threshold value T is reduced by a factor of 2, and the entire process repeated. This coding strategy, consisting of the dominant and subordinate passes followed by the reduction in the threshold value, is iterated until a target bit rate is achieved. A feature is nothing but the significant representative of an image which can be used for classification, since it has a property which distinguishes one class from other. The extracted features provide characteristics of input pixel to the classifier [56].The spatial features can be extracted by statistical and co-occurrence methods.

5. MULTI-LAYER PERCEPTRON (MLP)

Neural networks are used as statistical tools in a variety of fields, including Psychology, Statistics, Engineering, Economics and even Physics. They are used also as models of cognitive processes by neuro and cognitive scientists. Basically, neural networks are built from simple units, sometimes called neurons or cells, by analogy with the real thing. These units are linked by a set of weighted connections. Learning is usually accomplished by modification of the connection weights. Each unit codes or corresponds to a feature or a characteristic of a pattern that we want to analyze or that we want to use as predictor.

The neural networks usually organize their units into several layers. The first layer is called the input layer; the last one is the output layer. The intermediate layers are called the hidden layers. The information to be analyzed is fed to the neurons of the first layer and then propagated to the next layer and so on until the last layer. Each unit receives some information from other units and processes this information, which will be converted into the output of the unit. The goal of the network is to learn or to discover some association between input and output patterns, or to analyze, or to find the structure of the input patterns. The learning process is achieved through the modification of the connection weights between units.

The main neural networks types based on their structures are Single layer perceptron, Multi-layer perceptron, Backpropagation net, Hopfield net and Kohonen feature map. Multi-layer perceptron (MLP) is recognized as the best ANN used in classification from examples [32]. In this work, the multi-layer perceptron with back-propagation supervised learning algorithm is used for experimentation. Due to its extended structure, MLP is able to solve every logical operation, including XOR problem. The back-propagation algorithm in MLP is the solution of choice for many machine learning tasks [33], [34]. An advantage of supervised learning is the minimization of error between the desired and computed unit values.

5.1. Multilayer perceptron classification using back propagation

In order to classify the satellite images, the first step is feature extraction. In feature extraction, certain features are calculated for each pixel. Then, the network is trained by computing the input matrix and the target vector. The input matrix is obtained from the features and the target vector is manually calculated. Once training is completed, the network is simulated with an input image, for which classification should take place, to specify agriculture, urban and water body.

5.2. Back propagation two phases

- **Forward pass phase:** computes 'functional signal', feed forward propagation of input pattern signals through network
- **Backward pass phase:** computes 'error signal', *propagates* the error *backwards* through network starting at output units (where the error is the difference between actual and desired output values)

5.3. MLP algorithm

Step-1: Initialize all weights at random. Choose a learning rate η

Step-2: Propagated the input forward through the network for every layer in the network and each node of the layer.

Step-3: Evaluate the Weight Sum of the inputs to the node.

Step-4: Multiplied by Weights and summized.

Step-5: Calculate the Sigmoid Activation function.

Step-6: Output passed to each Neuron in next layer.

Step-7: Propagated the error backwards through the network for every node of the output layer.

Step-8: Precise the output layer of weights using Equation (1).

$$w_{ho} = w_{ho} + \left(\eta\, \delta_o o_h\right) \quad (1)$$

where w_{ho} is the weight connecting hidden unit h with output unit o, η is the learning rate, o_h is the output at hidden unit h. δ_o is given by the following.

$$\delta_o = o_o(1 - o_o)(t_o - o_o) \quad (2)$$

where o_o is the output at node o of the output layer, and t-o is the target output for that node.

Step-9: Modifiy the input weights using Equation (3).

$$w_{ih} = w_{ih} + (\eta\delta_o o_h) \quad (3)$$

where w_{ih} is the weight connecting node i of the input layer with node h of the hidden layer, o_i is the input at node i of the input layer, η is the learning rate. δ_h is calculated as follows.

$$\delta_h = o_h\left(1 - o_h\right)\sum_o \left(\delta_o w_{ho}\right) \quad (4)$$

ARPN Journal of Engineering and Applied Sciences

Step-10: Compute the error, by taking the average difference between the target and the output vector using Equation (5).

$$E = \frac{\sqrt{\sum_{n=1}^{p} (t_0 - o_0)^2}}{p} \quad (5)$$

Where p is the number of units in the output layer.

Step-11: Repeat from 2 for each pattern in the training set to complete one epoch.
Step-12: Shuffle the training set randomly. This is important so as to prevent the network being influenced by the order of the data.
Step-13: Repeat from step 2 for a set number of epochs, or until the error ceases to change.

6. FEATURE SELECTION METHODS

Feature selection (FS) is a process which attempts to select more informative features. In some cases, too many redundant or irrelevant features may overpower main features for classification. Feature selection can remedy this problem and therefore improve the prediction accuracy and reduce the computational overhead of classification algorithms. The main aim of feature selection is to determine a minimal feature subset from a problem domain while retaining a suitably high accuracy in representing the original features.

Feature selection can be accomplished through wrapper and filter methods. Wrappers depend heavily on classification algorithm to measure the prominence of a feature to be included in the model. Feature selection through wrappers generally performs better than filters because the filter selection is optimized for the particular learning algorithm to be used [35]. Wrapper methodss are very time taking and they are computationally expensive.

Filters based feature selection evaluate the usefulness of features in prediction independent of any learning algorithm. Filters are fast and are computationally more efficient but totally ignore the dependency of features' worthiness on learning algorithms [35]. Most attribute evaluation filters work in conjunction with rank searching. Features are ranked and a specific number of features falling below the user specified threshold are discarded from the feature set included for the purpose of model building.

In order to decrease the feature space and preselect the foremost vital features for a particular classification task, dataset and classifier, few feature selection methods are projected within the literature [36, 37]. Most generally established technique in remote sensing applications could be a manual feature selection with conventional knowledge exploration tools like histograms or scatter-plots. This methodology needs better understanding of the classification strategy and also the aspect of the features in study. Conversely, increasing the quantity of features of manual methods becomes unfeasible and additional quantitative feature selection techniques are needed [38]. One of the most widely used dimension reduction techniques in remote sensing is the Partial Swarm Optimization (PSO).

6.1. Particle Swarm Optimization (PSO) algorithm

Particle swarm optimization (PSO) is a stochastic, population based optimization technique aiming at finding a solution to an optimization problem in a search space. This algorithm was first described by Kennedy and Eberhart in 1995 [39]. PSO algorithm, which is tailored for optimizing difficult numerical functions and based on metaphor of human social interaction, is capable of mimicking the ability of human societies to process knowledge [40, 41]. It has roots in two main component methodologies: artificial life (such as bird flocking, fish schooling and swarming); and, evolutionary computation. Although the PSO algorithm is initially developed as a tool for modeling social behavior, it has been applied in different areas [39, 41-46]. Moreover, it has been recognized as a computational intelligence technique intimately related to evolutionary algorithms. The main objective of PSO is to optimize a given function called fitness function. PSO is initialized with a population of particles distributed randomly over the search space. In Figure-3 shows the flow chart of PSO process.

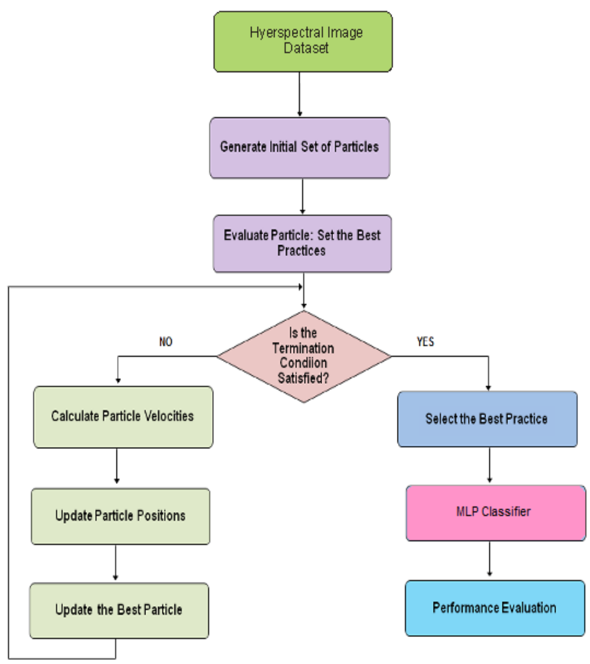

Figure-3. Flowchart of Particle Swarm Optimization algorithm.

6.2. PSO in remote sensing

Artificial Intelligence (AI) techniques have been increasingly applied in the classification of remote sensing images [47]. Swarm Intelligence (SI) is actually a complex multi-agents system, consisting of numerous simple individuals (e.g., ants, birds, etc.), which exhibit their swarm intelligence through cooperation and competition

VOL. 10, NO. 4, MARCH 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

among the individuals. Although there is typically no centralized control dictating the behavior of the individuals, the accumulation of local interactions in time often gives rise to a global pattern, SI has currently become a hot topic in artificial intelligence research, and it has succeeded in solving problems such as traveling salesman problems, data clustering, combination optimization, network routing, rule induction, and pattern recognition [48-53]. However, using SI in remote sensing classification is a fairly new research area and needs much more work to do.

Particle swarm optimization (PSO) is a stochastic optimization method [39] loosely based on the behavior of swarming animals such as birds and fish. Originally, these two started out developing computer software simulations of birds flocking around food sources and then later realized how well their algorithms worked on optimization problems. A number of particles, representing potential solutions to the problem, are released in the search space of potential solutions. Each particle has a position and a velocity, and is free to fly around the search space. The movement is controlled, however: the particles accelerate towards the position of the best performing particle as well as towards each particle's personal best previous position. The PSO algorithm is governed by a set of rules describing how each particle's position and velocity changes over time.

Each particle consists of:

- Data representing a possible solution
- A Velocity value representing how much the Data can be modified
- A pBest value signifying the closest the particle's Data has yet come up to the Target

6.3. PSO evolution steps

Step-1: Initialization phase, Initialize the *swarm*
Evolution phase repeat
Step-2: Evaluate *fitness* of each particle
Step-3: Update personal best position for each particle
Step-4: Update global best position for entire population
Step-5: Update each particle's velocity
Step-6: Update each particle's position until (termination criteria are met or stopping condition is satisfied)

Since the stochastic PSO algorithm has been found to be able to find the global optimum with a large probability and high convergence rate [15, 16], it is adopted to train the multi-layer perceptrons in this case study.

7. EXPERIMENTAL RESULTS

The following figure and table shows the results of classification accuracy. Table-1 and Figure-4 shows that comparison of before feature selection process and after feature selection using PSO. Table-1 and Figure-4 shows that Multilayer Perceptron algorithm get better accuracy than other classification algorithm in both before and after feature selection. After using feature selection using PSO gives better accuracy results.

Table-1. Comparison of classification results.

Classification	Before feature selection	PSO
RBF	68.6	75.23
Naïve Bayes	70.93	78.23
SMO	71.51	79.65
J48	73.26	77.91
MLP	79.65	80.23

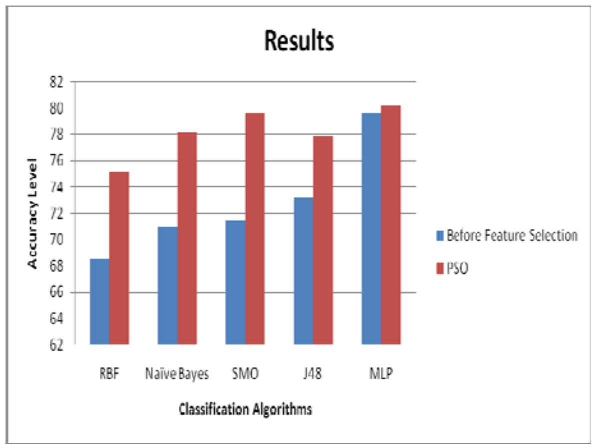

Figure-4. Comparison of classification accuracy results.

8. CONCLUSIONS

The main objective of this research was to present an image classification strategy in the problem of urban land-cover data. In this proposed hybrid classification algorithm, efficient classification accuracy then other classification algorithms. Based on the experimental result, the proposed work gets better classification accuracy. Here, the EZW algorithm are used for feature extraction process. MLP classification algorithm with PSO based feature selection get the better classification results. In future enhancement of this work will be texture based classification, fuzzy and rough set based approach.

REFERENCES

[1] Tan. K. and Du. P. 2011. Combined multi-kernel support vector machine and wavelet analysis for hyperspectral remote sensing image classification. Chinese Optics Letters. 9(1): 011003–110006.

[2] Dobigeon. N., Tourneret. J.Y. and Chang. C. 2008. Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery. IEEE Trans. Signal Process. 56(7): 2684-2695.

[3] Benediktsson J.A., Chanussot J. and Moon W.M. 2012. Very high-resolution remote sensing: Challenges and opportunities [point of view]. Proceedings of the IEEE. 100(6): 1907–1910.

[4] Peddle D.R. and Ferguson D.T. 2002. Optimization of multisource data analysis: an example using evidence reasoning for GIS data classification. Comp. & Geosci. 28: 45-52.

[5] Li G., Lu D., Moran E. and Hetrick S. 2011. Land-cover classification in a moist tropical region of Brazil with Landsat TM imagery. Int. J. Remote Sens. 32(23): 8207-8230.

[6] Roy P. and Giriraj A. 2008. Land use and land cover analysis in Indian context. Journal of Applied sciences. 8: 1346-1353.

[7] Arditi D., Oksay F.E., Tokdemir O.B. 1998. Predicting the Outcome of Construction Litigation Using Neural Networks. Computer-Aided Civil and Infrastructure Engineering. 13(2): 75-81.

[8] Thirumalaiah K., Deo M.C. 1998. River Stage Forecasting Using Artificial Neural Networks. Journal of Hydrologic Engineering. ASCE. 3(1): 26-32.

[9] Chau K.W., Cheng C.T. 2002. Real-time Prediction of Water Stage with Artificial Neural Network Approach. Lecture Notes in Artificial Intelligence. 2557: 715-715.

[10] Mondal A., Kundu S., Chandniha S. K., Shukla R. and Mishra P. 2012. Comparison of Support Vector Machine and Maximum Likelihood Classification Technique using Satellite Imagery.

[11] Naguib A., Farag M., Yahia M., Ramadan H. and Abd Elwahab M. 2009. Comparative study between support vector machines and neural networks for lithological discrimination using hyperspectral data. Egypt Journal of Remote Sensing and Space Science. 12: 27-42.

[12] Oommen T., Misra D., Twarakavi N. K., Prakash A., Sahoo B. and Bandopadhyay S. 2008. An objective analysis of support vector machine based classification for remote sensing. Mathematical Geosciences. 40: 409-424.

[13] Pal M. and Mather P. 2005. Support vector machines for classification in remote sensing. International Journal of Remote Sensing. 26: 1007-1011.

[14] Chan J. C. and Paelinckx D. 2008. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sensing of Environment. 112: 2999 – 3011.

[15] Chen Y., Crawford M. and Ghosh J. 2007. Knowledge based stacking of hyperspectral data for land cover classification. In Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining. pp. 316-322.

[16] Udelhoven T., Waske B., Linden S. and Heitz S. 2009. Land-cover classification of hypertemporal data using ensemble systems. IEEE Transaction on Geoscience & Remote Sensing, pp. III-1012 – III - 1015.

[17] Breiman L. 1996. Bagging predictors. Machine Learning. 24(2): 123 – 140.

[18] Opitz D. 1999. Feature Selection for Ensembles. In Proceedings of the 16th National Conference on Artificial/ Intelligence (AAAI). pp. 379 – 384.

[19] Pal M. 2003. Random forests for land cover classification. In Proceedings Geoscience and Remote Sensing Symposium. 6:3510 – 3512.

[20] Tsymbal A., Pechenizkiy M. and Cunningham P. 2005. Diversity in search strategies for ensemble feature selection. Information Fusion. 6(1): 83 – 98.

[21] Polikar R. 2006. Ensemble based systems in decision making. IEEE Circuits and Systems Magazine. Third Quarter. pp. 21 – 45.

[22] Gidudu A., Abe B.T. and Marwala T. 2008. Ensemble feature selection for hyperspectral imagery. In Proceedings of the 19th Annual Symposium of the Pattern Recognition Association of South Africa, Cape Town, South Africa, 27th – 29th November. pp. 27-32.

[23] Bischof . H. 1992. Multispectral classification of Landsat images using Neural Network. IEEE Transactions on Geo Science and Remote Sensing. 30(3): 482-490.

[24] Heerman P.D. and khazenie. 1992. Classification of multispectral remote sensing data using a back propagation neural network. IEEE trans. Geosci. Remote Sensing. 30(1): 81-88.

[25] Hepner G.F. 1990. Artificial Neural Network classification using mininmal training set:comparision to conventional supervised classification. Photogrammetric Engineering and Remote Sensing. 56: 469-473.

[26] Durairaj M. and Meena K. 2008. Application of Artificial Neural Network for Predicting Fertilization

Potential of Frozen Spermatozoa of Cattle and Buffalo. International Journal of Computer Science and System Analysis. pp. 1-10.

[27] Yegnanarayana B. 2001. Artificial Neural Network, Eastern Economy Edition.

[28] Oza C. N. and Tumer K. 2008. Classifier ensemble: Select real – world applications. Information Fusion. 9:4 – 20.

[29] Hsu Pai-Hui. 2007. Feature extraction of hyperspectral images using wavelet and matching pursuit. In ISPRS Journal of Photogrammetry and Remote Sensing. 62: 78 – 92.

[30] Keshava N. and Mustard J. F. 2002. Spectral unmixing. IEEE Signal Processing Magazine. 19(1): 44-57.

[31] Sanchez S., Martin G., Plaza A. and Chang C. 2010. GPU implementation of fully constrained linear spectral unmixing for remotely sensed hyperspectral data exploitation. In Proceedings SPIE Satellite Data Compression, Communications and Processing VI. 7810: 78100G-1 – 78100G-11.

[32] Durairaj M. and Meena K. 2007. Intelligent Classification Using Rough Sets and Neural Networks. The ICFAI Journal of Information Technology. 3:75-85.

[33] Cheng C.T., Chau K.W., Sun Y.G., Lin J.Y. 2005. Long-term prediction of discharges in Manwan Reservoir using artificial neural network models. Lecture Notes in Computer Science. 3498:1040-1045.

[34] Lin J.Y., Cheng C.T., Chau K.W. 2006. Using support vector machines for long-term discharge prediction. Hydrological Sciences Journal. 51(4):599-612.

[35] Abe N. and M. Kudo M. 2005. Entropy criterion for classifier-independent feature selection. Lecture Notes in Computer Science. 3684:689-695.

[36] Langley P. 1994. Selection of Relevant Features in Machine Learning. In: AAAI Fall Symposium. pp.127-131.

[37] Koprinska I. 2009. Feature Selection for Brain-Computer Interfaces. In: 13th Pacific-Asia Int. Conference on Knowledge Discovery and Data Mining: New Frontiers in Applied Data Mining. Pp.106-117.

[38] Witten I. H., Frank E., Hall M. A. 2011. Data Mining: Practical Machine Learning Tools and Techniques. Elsevier.

[39] Kennedy J., Eberhart R. 1995. Particle Swarm Optimization. Proceedings of the 1995 IEEE International Conference on Neural Networks. Perth. 1942-1948.

[40] Kennedy J., Eberhart R., Shi Y. 2001. Swarm Intelligence. Morgan Kaufmann Publishers, San Francisco.

[41] Clerc M., Kennedy J. 2002. The Particle Swarm—Explosion, Stability, and Convergence in a Multidimensional Complex Space. IEEE Transactions on Evolutionary Computation. 6(1):58-73.

[42] Kennedy J. 1997. The Particle Swarm: Social Adaptation of Knowledge. Proceedings of the 1997 International Conference on Evolutionary Computation. Indianapolis. pp. 303-308.

[43] Chau K.W. 2004. River Stage Forecasting with Particle Swarm Optimization. Lecture Notes in Computer Science. 3029:1166-1173.

[44] Chau K.W. 2004. Rainfall-Runoff Correlation with Particle Swarm Optimization Algorithm. Lecture Notes in Computer Science. 3174:970-975.

[45] Chau K.W. 2005. Algal bloom prediction with particle swarm optimization algorithm. Lecture Notes in Artificial Intelligence. 3801 : 645-650.

[46] Chau K.W. 2006. Particle swarm optimization training algorithm for ANNs in stage prediction of Shing Mun River. Journal of Hydrology. 329(3-4): 363-367.

[47] Stathakis D. and Vasilakos A. 2006. Comparison of computational intelligence based classification techniques for remotely sensed optical image classification. IEEE Transactions on Geoscience and Remote Sensing. 44(8):2305-2318.

[48] Machado Th. R. and Lopes H. S. 2005. A Hybrid Particle Swarm Optimization Model for the Traveling Salesman Problem. Proceedings of the International Conference in Coimbra, Portugal. Springer. pp. 255-258.

[49] Kennedy J. 2000. Stereotyping: improving particle swarm performance with cluster analysis. In Proceedings of the 2000 Congress on Evolutionary Computing. 2:1507-1512,.

[50] Ting T. O., Rao M. V. C., Loo C. K., Ngu. S.S. 2003. Solving unit commitment problem using hybrid particle swarm optimization. Journal of Heuristics. 9(6): 507-520.

VOL. 10, NO. 4, MARCH 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

[51] Kwang M. S. and Weng H. S. 2002. Multiple ant-colony optimization for network routing. In Proceedings of the First International Symposium on Cyber Worlds. pp. 277-281.

[52] Liu X. P., Li. X., Yeh. A. G. O., *et al.* 2007. Discovery of transition rules for geographical cellular automata by using ant colony optimization. Science in China Series D-Earth Sciences. 50(10):1578-1588.

[53] Omran M. 2005. Particle Swarm Optimization methods for pattern Recognition and Image Processing. Dissertation for the Doctoral Degree. University of Pretoria.

[54] Haralick R. and Shapiro L. 1985. Image segmentation techniques. Comput. Vis., Graph., Image Process. 29: 100-132.

[55] Reichel J., Menegaz G., Nadenau M.J. and Kunt M. 2001. Integer Wavelet Transform for Embedded Lossy to Lossless Image Compression. IEEE Transactions on Image Processing. 10(3):383-392.

[56] Kavitha K., Arivazhagan S., Kayalvizhi N. 2010. Wavelet Based Spatial - Spectral Hyperspectral image Classification Technique Using Support Vector Machines. International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE, pp. 1-6.