Land Usage Analysis: A Machine Learning Approach

Kishwar Ali
Department of Computer Science Lahore Leads University Lahore Pakistan

Hidayat Ur Rahman
Department of Computer Science Lahore Leads University
Lahore Pakistan

Rehanullah Khan
Department of Computer Science Lahore Leads University
Lahore Pakistan

Muhammad Yasir Siddiqui
Department of Computer Science
Lahore Leads University
Lahore Pakistan

Fahad Najeeb
Department of Computer Science
Lahore Leads University
Lahore Pakistan

Sadiq Amin
Department of Computer Science
Lahore Leads University
Lahore Pakistan

Rafaqat Alam Khan
Department of Computer Science
Lahore Leads University
Lahore Pakistan

Sanaullah Manzoor
Department of Computer Science
Lahore Leads University
Lahore Pakistan

Nadeem Jabbar Chaudhry
Department of Computer Science
Lahore Leads University
Lahore Pakistan

ABSTRACT
In this article, machine learning based land usage analysis has been investigated. The objective is twofold: Firstly, the analysis and usage of simple pixel based features from the more complex Hyper Spectral images to land cover recognition. Secondly, an investigation into the parametric and non-parametric machine learning algorithms for the pixel based land cover analysis. For an experimental evaluation, we use the SPOT-5 satellite imagery having resolution of 2.5m. From the machine learning set, we select Support Vector Machine (SVM), Maximum Likelihood Estimator (MLE) and Artificial Neural Network (ANN). These algorithms are selected based on their superior performance in pattern recognition tasks. We distribute the feature space in seven classes i.e. Roads, Settled Areas, Tobacco, Sparse Vegetation, Sugar Cane, Barren Land and water. From the extensive experimentation, and in the current setup, it is concluded that SVM is best suited to the land cover analysis.

Keywords
SVM,MLE, ANN, remote sensing, land cover classificaiton, SPOT-5

1. INTRODUCTION
Remote sensing is used to obtain the information from the earth surface by using satellite imaging system. To classify the satellite imagery usually a supervised classifier is trained on the training sets. These supervised classifier are either parametric which requires the prior knowledge about the pixel distribution such as MLE, SVM and Naïve Bayesian or non-parametric such as ANN, K-Nearest Neighbor and Multi-layer Perceptron. Recent studies have been carried out in comparing the performance of various parametric and non-parametric classifiers such as MLE, SVM and ANN. MLE is mostly preferred in land cover classification using high resolution imagery. However, it is found that when applied on low resolution imagery such as Landsat Satellite imagery, the detection probability of MLE reduces. In contrast, the non-parametric classifier ANN results in better classification performance [1] for low resolution of Landsat Satellite imagery. SVM are always very effective in low resolution imagery classification. In such scenarios, it has been shown to outperform the supervised classifiers such as; MLE, ANN and Multi-layer Perceptron [2]. SVM uses different kernels and hence the performance of the SVM varies with the selection of the kernel functions and its parameters [3]. SVM has been very useful in classification of coastal areas and settlements, compared to ANN and ML. In general, the SVM has good classification accuracy for surfaces having the same spectral reflectance [4]. The use of SVM using Aster imagery has also reported the highest accuracy as compared to MLE [5]. In [6], the comparison of parametric supervised classifiers has been carried out for an urban environment. The results show that SVM outperforms other classifiers such as MLE and ANN. SVM was originally designed for binary linear classification, with mostly linear problems that can be separated by linear hyper plan. However, in real world problems, we mostly deal with the non-linear problems. In order to enhance the SVM to fit the non-linear problems domain, kernels such as Radial Basis Function (RBF), Polynomial and Sigmoid were introduced [7]. With the addition of Non-linear kernels, SVM can optimally solve non-linear and complex classification problems [1, 2 and 8]. In [9], Keramitsoglou et. al. used SVM Radial Basis kernel to accurately distinguish the spectral signatures of different vegetation using their spectral characteristics. For [9], the dataset used IKONOS imagery. It was found that RBF kernel proves useful to differentiate between different vegetation's. In [10], the comparison of SVM and MLE is done using Multi-angle Imaging Spectro Radiometer (MISR) to distinguish different vegetation category. From the experimentation setup, it was found that SVM outperforms the MLE classifier.

In this article, as a continuation of the comparative analysis of classifiers for land usage analysis, we have compared the performance of various supervised classifiers: The set includes SVM, Maximum Likelihood Estimator (MLE) and ANN on high resolution SPOT-5 imagery. These algorithms are selected based on their superior performance in pattern recognition tasks. For analysis, we distribute the feature space in seven classes i.e. Roads, Settled Areas, Tobacco, Sparse Vegetation, Sugar Cane, Barren Land and water. For each of the classifiers, we computed the evaluation parameters: Overall Accuracy (OA), User Accuracy (UA), Producer Accuracy (PA) and Kappa Statistics (KS). From the extensive experimentation, and in the current setup, we found that SVM is best suited to the land cover analysis outperforming ANN and MLE.

2. SITE DESCRIPTION

The site used as a dataset for the experimental evaluation is located at 34°12'0N 72°1'60E. The site is 47 Km away from the Peshawar located in the province of Khyper Pakhtoonkhwa (KPK) Pakistan. The land is very fertile in growth of vegetation such as Tobacco and Sugarcane. It is the 2nd biggest city of the Province and hence it has an enormous population. Various streams and canals are used for the purpose of irrigation and livelihood. The imagery we have used for this analysis is obtained in July 2013 from the SPOT-5 which has the resolution of 2.5m (see Figure 1).

3. FEATURE SET

The dataset of SPOT-5 imagery used for an experimental evaluation is obtained from the Suparco[1] on July 2013. The imagery consists of four bands i.e. Red, Green, Blue and the Infra-red. The imagery has a resolution of 2.5m which is a high quality imagery compared to the Landsat imagery. We have divided our dataset into seven classes. The seven classes include Roads, Settled Areas, Tobacco, Sparse Vegetation, Sugar Cane, Barren Land and Water. Table 1 shows further detail of the classes. The Region Of Interests (ROI) for classification task consist of 4737 instances of Roads, 3121 instances of Settled Areas, 4948 instances of Tobacco, 4442 instances of Sparse Vegetation, 2848 samples of Sugar Cane, 13989 instances of Barren Land and 2764 pixels of water.

4. METHODOLOGY

After acquisition of the dataset, we selected the training data and created the ROI. Based on the train data, classifier on these training set is trained and performance evaluated. For performance evaluation, we use the Over-all Accuracy (OA), User Accuracy (UA), Producer Accuracy (PA) and Kappa Statistics (KS). Fig-2 describes the flow chart of our methodology. The rest of the section provides the classification and experimental evaluation.

4.1 Machine Learning

From the machine learning set, we use MLE, SVM ANN. These are selected based on their superior performance in the state-of-the-art. Further explanation of these approaches is as follows:

4.2 Support Vector Machine (SVM)

SVM is a binary classification algorithm which allows the training data to fit in a hyper plane. The hyper plane is a space in which maximum separation between the classes take place [14]. Separation between the training classes will be Maximum if the distance between the neighboring pixels is high. SVM generates feature vector from training sets. From feature vector the hyper plane is created which provides maximum separation amongst all the class. SVM can be further enhanced and it can be used as multi-class SVM, since a multi-class problem can be resolved by treating a multi-class as a series of binary combination. For this purpose SVM has nonlinear kernels such as polynomial sigmoid and radial basis function. In our experiment we have used radial basis function (RBF) which has shown good results in [13], [15], [16], and [17]. Consider the training sets as (x_i, y_i) where i=1…n. x$\in R$ and y $\in \{1, -1\}$. The equation for SVM is given as [18].

$$y_i = (W^T \emptyset(x_i + \mathrm{b})) \geq 1 - \xi_i$$

In the above equation x_i is the training vector, which is converted to higher dimensional feature space by the kernel $\emptyset$. W is the normal vector to the hyper plane and W^T is the transpose of W. RBF is one of the nonlinear kernels which we have used for our classification. RBF kernel is chosen for the classification of the imagery because it has been advocated by various researchers in [5,6 and 9]. The equation of the RBF is below.

$$K(x_i, y_i) = \exp(\gamma ||x_i - y_i||^2) \text{ where } \gamma > 0.$$

The parameter of the RBF kernel γ and the Penalty parameter plays important role in the hyper plane creation. Generally γ is set to the inverse of the number of spectral bands, penalty parameter is set to its Maximum 100 and the threshold probability it is set to zero to remove unclassified pixels.in our experiment we have used $\gamma = .25$ since we have four bands, penalty is maximized i.e. 100 and pyramids is set to 0.

4.3 Maximum Likelihood (MLE)

Maximum likelihood is the parametric classifier which uses statistical approach; it uses the normal Gaussian distribution of each pixels of the class. Based on the normal distribution it calculates the probability of pixel belonging to the class [11]. If the pixel has highest probability for a particular class it will be assigned to that class. E.g. the probability of Tobacco class c_i and a random pixel x can be calculated as:

$$P(x \mid c_i) = \frac{1}{(2\pi)^{1/2} \partial_i} \exp\left[\frac{(x - \mu_i)^2}{2(\partial_i^2)}\right]$$

In the above equation μ_i is the mean of the tobacco class and ∂^2 is the variance of the class. Since here we have to deal with more than one class, we have used seven categories in our research, so for multiple classes we have to calculate n-dimensional multivariate density as shown in the following equation.

$$P(X \mid c_i) = \frac{1}{(2\pi)^{\frac{n}{2}} |V_i|^{1/2}} \, exp \left[-\frac{1}{2}(X - M_i)^T Vi^{-1} (X - M_i)^T\right]$$

Where $|V_i|$ is the absolute of the Co-variance matrix, Vi^{-1} is the inverse of the covariance matrix and (X- $\mu_i)^T$ is the transpose of the matrix (X - μ_i). Assume n=7 as in our case, to calculate (X | c_i) where X is an unknown pixel from c_i . The unknown pixel will be classified in the class c_i if it satisfies the maximum likelihood equation as follows.

$$P(X \mid c_i).p(c_i) \geq P(X \mid c_j).p(c_j)$$

From the above equation it is clear that the classification of a pixel to a class depends upon its prior probability. If the prior probability of a pixel is high for a certain class it will be assigned to that class, in the above equation shows the prior probability of class tobacco denoted by P (c_i). Recent study carried out using MLE for land cover classification showed better results in the classification of high resolution imagery

[1] http://www.suparco.gov.pk/

[12]. The disadvantage of using MLE is the time consumption involved in the computation of the probabilities for each class. In our experiment we noticed that MLE has near accuracy to the SVM kernel although SVM has the highest accuracy compared to MLE.

4.4 Artificial Neural Network (ANN)

ANN is non-parametric supervised classifier composed of small processing units; these small processing units operate in parallel and hence form a central processing unit. In other words ANN consists of small units called functions connected with the weights [19]. A typical neural network consists of input, hidden and output layer. The input layer is composed of variables which are the inputs to the network. The input could be either pixel data or texture etc. in our experiment the inputs are pixels and the outputs are the number of classes which correspond to the categories we have used. The hidden layer is comprised by series of function or kernels which are linked to the previous function and hence each function is a layer connected to the other layer. Due to this chain of layers the ANN learns by adjusting the weight and minimizes the error, this error is feedback to the network. ANN is also driven by various parameters such as RMS value, momentum, training rate and threshold. The training rate is used to adjust the weights and in our experiment we have used training rate of 0.2, the training threshold is used to determine the node internal weight in our experiment we have used threshold of .9, Training momentum helps to adjust the step size used for adjusting internal weights. We have used training momentum of .9 and RMS of 0.1 which determines the exit criteria for the training. In ANN the hidden layer defines the numbers of hyper planes if the hidden layer is 0 then all the pixels are classified in a single hyper plane. Recently researchers was able to find that the number of hidden layer has the significant effect on the performance of ANN and has used two layers in [20] and reported better accuracy than using a single layer.

5. EXPERIMENTAL EVALUATION

For performance evaluation, we have used (Overall Accuracy) OA, (User Accuracy) UA, (Producer Accuracy) PA and KS (Kappa Statistics). These metrics are selected based on their wide usage in the remote sensing domain. OA provides the probability of unknown pixel to be correctly classified. The UA determines the probability of a pixel that is correctly mapped in the given class to which it belongs. The parameter PA defines the probability of correctly identified pixels. The KS is the ratio of the actual agreement with the reference data against the chance agreement. All these parameters are reported based on their percentage values. For performance evaluation, we have used a total of 36849 pixels for training data collected for all the classes. For this training data, the classifiers were trained and for each of the classifier, the four parameters OA, UA, PA and KS are calculated. The classifiers selected are based on their performance in the land usage analysis and general classification tasks. These include SVM, MLE and ANN.

Table 2 only shows the comparison of UA and PA for each ROI and for each of the classifiers; SVM, MLE and ANN. Table 2 shows that for SVM, the lowest UA is reported for water class which is 53.8% (shown bold in Table 2). The water class has UA of 49.35 for ANN and 58.94 for MLE. The same scenario was observed with the classification of Sparse Vegetation regarding the parameter PA. Using SVM, the Sparse Vegetation has a PA value of 71.74 while MLE has PA of 81.18 using the same. For the classification of Settled Areas, the same trade-off was found in which SVM has PA of 93.2, ANN has PA of 92.3 and MLE has PA of 94.9. For the above three ROI, we have noticed that MLE has considerably good recognition performance (shown in bold) for Water, Sparse Vegetation and Settled Areas. Compared to the MLE, the ANN only showed the highest accuracy (UA) for Tobacco ROI, it has PA of 94.83 (shown in bold) whereas, the MLE has UA of 91.65 and the SVM have UA of 92.89. Regarding other ROI, SVM has the highest UA for the Barren Land, Roads and Sugar Cane as shown in the Table 2.

To evaluate the overall performance of the three classifiers SVM, ANN and MLE, we report the overall accuracy and kappa statistics. Table 3 shows the OA and KS with the corresponding classifier. From table 3, we have noticed that with the OA of 84.8% and KS of 0.8, the SVM has outperformed all the classifiers. MLE performs better than ANN having OA of 83.2% and kappa statistics of 0.78.Table 3 shows that ANN has the lowest OA of 80.4% and KS of 0.78.

For probabilities class distribution point of view, for each of these classifiers, we also have examined the range of probabilities for different classes. For SVM, the UA ranges from 53.8% to 95.2%, for the MLE the probability lies in between 53.2% to 94.5% for different classes. For the ANN, the UA lies in the range 49.3% to 94.8%. The better distribution is provided by the SVM in which it has the lowest detection probability (UA) of 53.8% and highest probability of 95.26%. The MLE also reports satisfactory detection performance. However, contrary to the state-of-the-art, the ANN reports the lowest accuracy compared to MLE and SVM in our experimentation setup. Further comparison of the probabilities of detection and error are shown in Figure 3(a) and Figure 3(b). Figure 3(a) compares the probability of detection of classifiers for the seven ROI, whereas; Figure 3(b) shows the probability of error of all the classifiers for the seven classes. In Figure 3(b), the highest error probability is noticed for Sparse Vegetation when classified with ANN. For Sugarcane, the probability of error was highest when classified with ANN as shown in figure 3(b). For visual analysis of the detection algorithms, see Figure 4. From Table 3, and Figure 3(a) and Figure 3(b), we notice that the SVM exhibits the highest overall performance in terms of OA and KS. However, SVM still has the deficiency of accurately classifying the Vegetation such as Sugar cane, Sparse Vegetation and Tobacco. We assume that this confusion is due to the same spectral signature of vegetation color i.e. green color. The highest OA and KS for SVM depict that the SVM can optimally differentiate between various classes and outperform ANN and MLE. To visualize strength of supervised classifier in classification of land cover we have plotted classified image shown in Figure-4

Fig-4 shows the visualize result of classifying the image. 4(a) shows the original image 4(b) shows the image classified with ANN, 4(c) shows the classified image using MLE and 4(d) shows the SVM results.

6. CONCLUSION

In this paper, we analyzed the land usage based on machine learning algorithms using the SPOT-5 imagery. From the machine learning set, we used SVM, ANN and the MLE. We used seven classes i.e. Roads, Settled Areas, Tobacco, Sparse Vegetation, Sugar Cane, Barren Land and water. From the extensive experimentation, it is concluded that the SVM is best suited to the land cover analysis. Our results agree with the state-of-the-art [13, 16, 17] advocating the superior performance of the SVM.

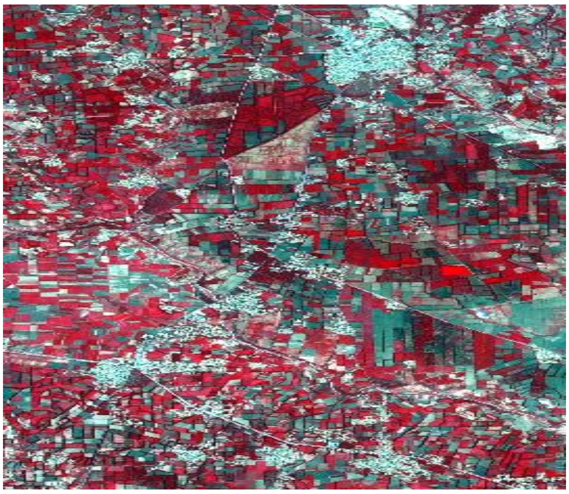

Fig:1 SPOT-5 Imagery used for Classification.

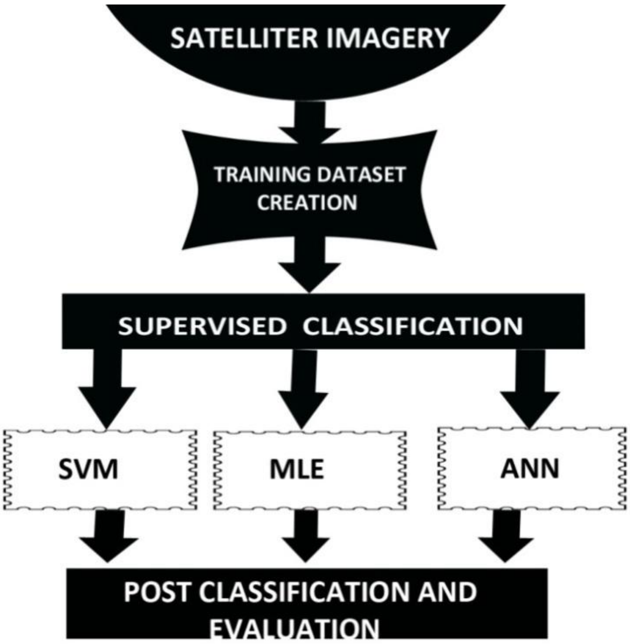

Fig-2: Flow Chart diagram of Methodology.

Table: 1 Training set details for each ROI's.

Classifier	OA	KS
SVM	84.8	0.8
MLE	83.2	0.78
ANN	80.4	0.78

Table: 3 Comparisons of OA and KS. (OA: Overall accuracy, KS: Kappa Statistics)

Classes	Description
Roads	Highways, Small Roads, other Routes used for Land Communication
Settled Areas	Buildings, Homes, Restaurants, Shops, Industries, Schools and Hospitals
Tobacco	Crop fields which are rich in tobacco crop.
Sparse Vegetation	Contains of sparse vegetation such as Persimmon, leeches, Peaches, apples and trees found on road side.
Sugar Cane	Fields which contains the Sugar Cane crop
Barren Land	Contains of the land which is not able to be cultivated and hence is not used for vegetation.
Water	Streams, Lakes, Canals, Channels, Drains, Rivers, Ponds and Reservoirs.

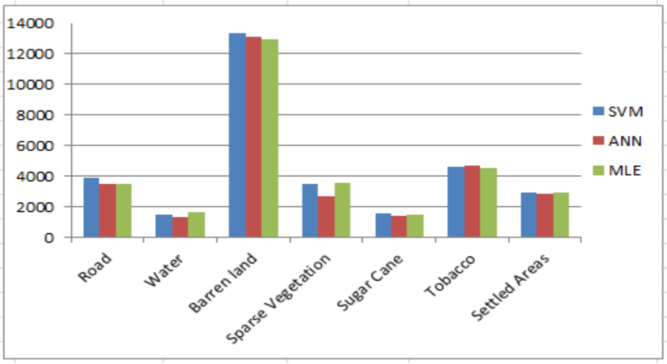

Fig-3(a) Comparison of Probability of Detection of SVM,ANN and MLE

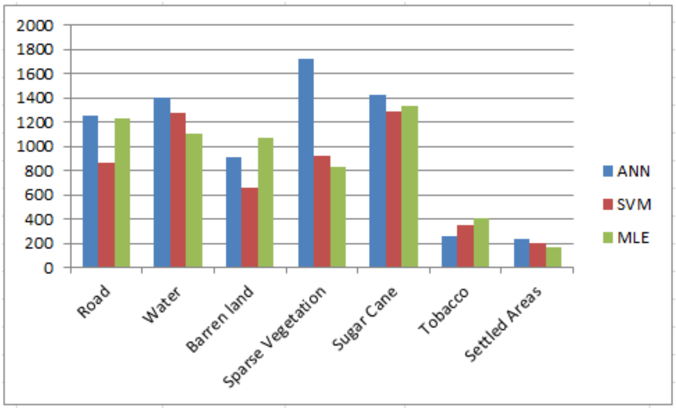

Fig-3(b) Comparison of Prob. of Error.

Table: 2 Performance evaluations. (UA: User Accuracy, PA: Producer Accuracy)

	SVM		ANN		MLE	
	UA	**PA**	**UA**	**PA**	**UA**	**PA**
Road	**81.74**	70.95	73.44	59.92	73.99	72.75
Water	53.8	90.45	49.35	66.9	**58.94**	77.09
Barren Land	**95.26**	94.86	93.49	95.87	92.36	97.69
Sparse Vegetation.	79.11	71.74	61.26	75.5	**81.18**	68.1
Sugar Cane	**54.71**	68.76	49.75	60.87	53.27	56.63
Tobacco.	92.89	85.05	**94.83**	78.33	91.65	84.39
Settled Areas.	93.53	93.2	92.31	83.77	**94.49**	88.19

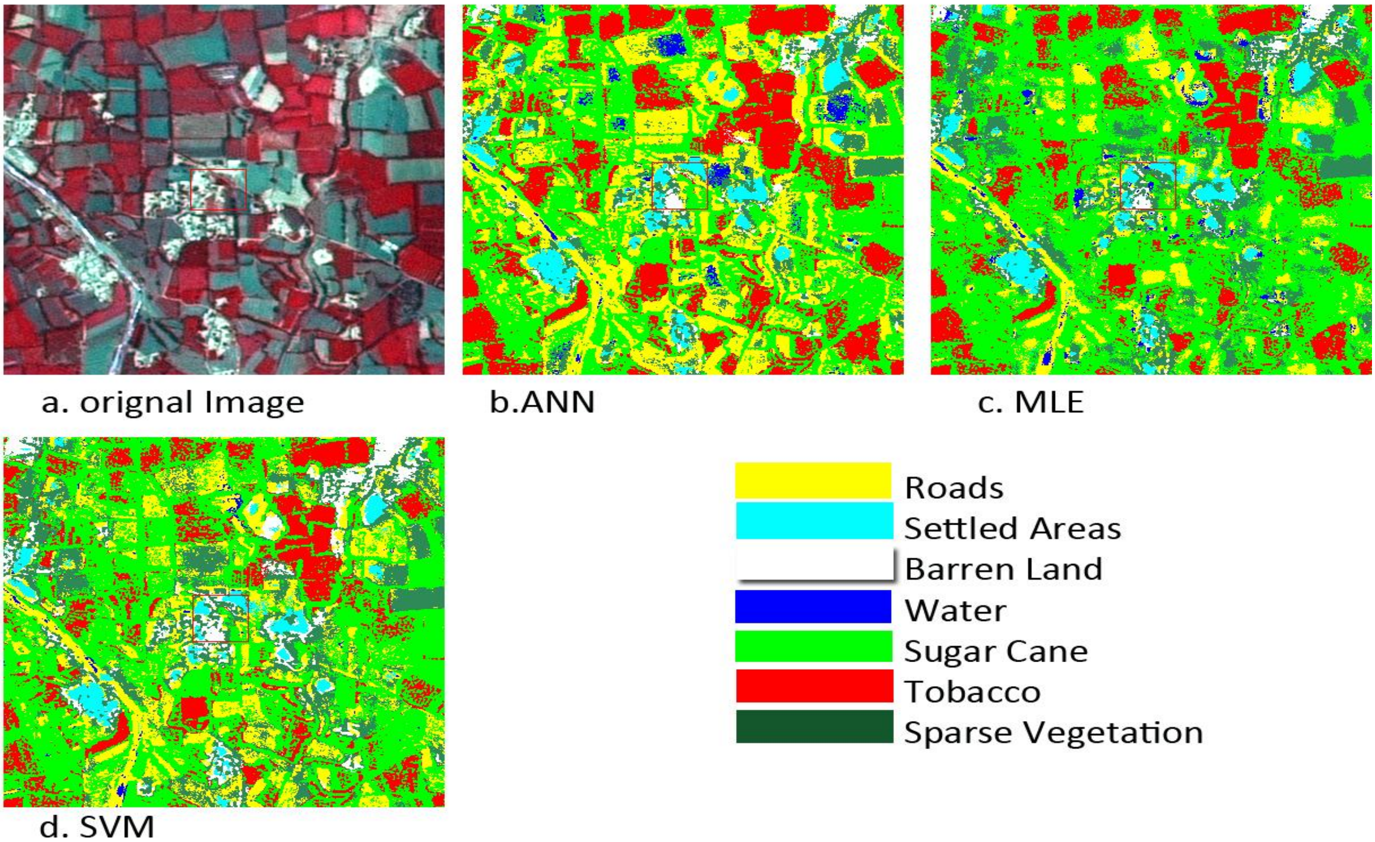

Figure 4: Detection of land use using ANN, SVM and MLE for corresponding classes.

7. REFERENCES

[1]. Kavzoglu, T., Reis, S. (2008). Performance analysis of maximum likelihood and artificial neural network classifiers for training sets with mixed pixels. GIScience and Remote Sensing, 45, 330–342.

[2]. Huang, C., Davis, L.S., Townshed, J.R.G. (2002). An assessment of support vector machines for land cover classification. International Journal of Remote Sensing, 23, 725–749.

[3]. Kavzoglu, T., Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification. International Journal of Applied Earth Observation and Geo information, 11, 352–359

[4]. Szuster, B.W., Chen, Q. and Borger, M. (2011). A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. Applied Geography, 31, 525-532.

[5]. Yu, L., Porwal, A., Holden, E.J. and Dentith, M.C. (accepted in 2011). Towards automatic lithological classification from remote sensing data using support

vector machines. Computers & Geosciences, Article in Press.

[6]. Peijun Du, Pei Liu, Junshi Xia, Li Feng, Sicong Liu, Kun Tan and Liang Cheng (2014) Remote Sensing Image Interpretation for Urban Environment Analysis: Methods, System and Examples.

[7]. Haykin, S. (1999) Neural Networks: A Comprehensive Foundations, 2ed. Upper Saddle River: Prentice Hall.

[8]. Yang, X. (2011) Parameterizing Support Vector Machines for Land Cover Classification, Photogrammetric Engineering and Remote Sensing, 77, 1, pp. 27-37.

[9]. Keramitsoglou, I. Sarimveis, H. Kiranoudis, C.T. Kontoes, C. Sifakis and N. Fitoka, E. (2006) The Performance of Pixel Window Algorithms in The Classification of Habitats Using VHSR Imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 60, 4, pp. 225-238.

[10]. Su, L. and Huang, X. (2009) Support Vector Machine (SVM) Classification: Comparison of Linkage Techniques Using a Clustering-Based Method for Training Data Selection. GIScience& Remote Sensing, 46, 4, pp. 411-423.

[11]. Scott, A.J. and Symons, M.J. (1971). Clustering methods based on likelihood ratio criteria. Biometrics 27(2), 387–397

[12]. Al-Ahmadi F.s, Hames A.S(2009) Comparison of Four Classification Methods to Extract Land Use and Land Cover from Raw Satellite Images for Some Remote Arid Areas, Kingdom of Saudi Arabia. Journal of King Abdulaziz University-Earth Sciences vol 20 No. 1 pg 167-191

[13]. G.P. Petropoulos, K. Arvanitis, and N. Sigrimis, "Hyperion hyperspectral imagery analysis combined with machine learning classifiers for land use/cover mapping", ;presented at Expert Syst. Appl., 2012, pp.3800-3809.

[14]. Foody and Marther, 2004. A relative evaluation of multiclass image classification by support vector machine. IEEE Transactions on Geoscienes and Remote Sensing. Vol 42.

[15]. Haung et al, Use of dark object concept and support vector machines to automate forest cover change analysis. Remote Sensing of Environment (2008). v112. 970-985.

[16]. Petropoulos et al., 2011. Burnt Area Delineation from a uni-temporal perspective based on Landsat TM imagery classification using Support Vector Machines. International Journal of Applied Earth Observation and Geoinformation. 70-80.

[17]. Petropoulos et al., 2010a. ASTER multispectral imagery analysis and support vector machines for rapid and cost-effective post-fire assessment: a case study from the Greek Fires of 2007. Combining Natural Hazards and Earth Systems Science. v10. 305-317.

[18]. Boser, B.E., Guyon, I. and Vapnik, V. (1992). A training algorithm for optimal margin classi_ers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, pages 144-152. ACM Press.

[19]. Haykin, 1994.Neural networks. Macmillan College Publishing Company, New York, USA

[20]. J.F.Mas, J.J. Flores, The application of artificial neural networks to the analysis of remotely sensed data, International Journal of Remote sensing, Vol 29,No 3, PG 617-663.Feb 2008.