

## COMPARATIVE ANALYSIS OF SVM, ANN AND CNN FOR CLASSIFYING VEGETATION SPECIES USING HYPERSPECTRAL THERMAL INFRARED DATA

Mehmood ul Hasan<sup>1,\*</sup>, Saleem Ullah<sup>2</sup>, Muhammad Jaleed Khan<sup>1</sup>, Khurram Khurshid<sup>1</sup>

<sup>1</sup>Vision Lab, Department of Electrical Engineering, Institute of Space Technology, Islamabad - \*akhunzada33@gmail.com  
mjk093@gmail.com, khurram.khurshid@ist.edu.pk

<sup>2</sup>Department of Space Science, Institute of Space Technology, Islamabad - sirifsaleem@gmail.com

**KEY WORDS:** Hyperspectral imaging, thermal infrared data, machine learning, classification, vegetation species

### ABSTRACT:

Vegetation includes a significant class of terrestrial ecosystem. Information on tree species categorization is important for environmentalists, foresters, agriculturist, urban managers, landscape architects and biodiversity conservationist. The traditional methods of measuring and identifying tree species (i.e., through field-based survey) are time taking, laborious and costly. Remote sensing data provides an opportunity to identify and classify vegetation species over a large spatial extent. Hyperspectral remote sensing can detect the subtle spectral details among species classes and thus make it possible to differentiate vegetation species based on these subtle variations. This research examines the thermal infrared (2.5 to 14.0  $\mu\text{m}$ ) hyperspectral emissivity spectra (comprised of 3456 spectral bands) for the classification of thirteen different plant species. The use of thermal infrared hyperspectral emissivity spectra for the identification of vegetation species is very rare. Three different machine learning methods including support vector machine (SVM), artificial neural network (ANN) and convolutional neural network (CNN) are used to classify thirteen vegetation species and their performance is assessed based on their overall accuracy. The accuracy obtained by CNN, ANN and SVM is 99%, 94% and 91%, respectively. Each classifier was also tested for the advantage associated with increase in training samples or object segmentation size. Increase in the training samples improved the performance of SVM. In a nutshell, all comparative machine learning methods provide very high classification accuracy and CNN outperformed the comparative methods. This study concludes that thermal infrared hyperspectral emissivity data has the potential to discern vegetation species using state of the art machine learning and deep learning methods.

## 1. INTRODUCTION

### 1.1 Background

Identifying tree species through statistical classification is an essential step to manage, store, and guard forestry resources. Thorough and precise forestry maps are vital for the prevention and sensing of fire, water scarcity and a variety of other forest disorders triggered by change of climate. Satellite images comprise pixels showing different ground objects with identifiable brightness values, letting the statistical classification of objects like vegetation and shrubs, due to their spectral signs. Classification of forestry has extended from more general classifications (like classifying deciduous and coniferous trees) to more deeper classifications (like species within a tree type). Usually, the capability to distinguish species is inadequate, because of low spectral differences, which aids in distinguishing the minor spectral variations among species. [1].

Imagery type is a key feature in classification because the spectral and spatial resolutions can affect the classification accuracy. Three to around eight image stacks of multispectral bands are usually utilized for distinguishing land covers or forest cover (broadleaf, conifer) classification. Hyperspectral stack of data comprises many (usually around 64 to 256) successive thinner bands, giving more details that permits the classification of small spectral differences among forest covers. Even with more quantity of information present in hyperspectral imagery, discerning the identical genus species might be tough, often reducing the classification accuracy. Still, Clark et al. observes that the use of hyperspectral data essentially performs better than the use of multispectral data.

Many high-level features in raw data are captured by the Convolutional Neural Networks (CNN) right away. This study provides machine learning approach for identifying certain vegetation species using hyperspectral data in the thermal infrared band. Utilizing the modern AI based methods like CNN to statistically classify vegetation has improved the overall accuracy. Statistical classification is being used in numerous applications that includes a choice mechanism to allocate observations to a set of ground cover types. [3] The data-intensive hyperspectral stack of images enhances the classification accuracy by giving better spectral resolution to distinguish the land cover types spanning one band in a multispectral dataset. Hence, the hyperspectral imagery is widely employed for discerning the tree species [4].

### 1.2 Significance of Vegetation Classification in a Forest

Familiarity with the vegetation state of a forest cover is significant for both management of protected parts (Nagendra et al., 2013) and approximating the worth of forestry (Ashutosh, 2012). Centralized monitoring necessitates multidimensional approach and accuracy in extracting forest statistics. The forest ecosystem cover information is a basic and significant constituent (Shen, Sakai, Hoshino, 2010). (Jadczyk, 2009) studies the areas threatened by human intrusions in Karkonosze Mountains, Poland. A section of the Karkonosze National Park forms part of the valuable Karkonosze Mountains ecosystem. Industrial expansion surrounded the mountains around 30 years ago, due to which drought, pollution, and acid rains occurred and the ecosystem of the area was damaged. The lack of foresight leading to the unplanned decisions caused the ecosystem's disturbance (Raj, 2014). Therefore, exhaustive vegetation information extraction is important for the management of

\* Corresponding author

national parks as well as forests. Before the damage occurred, the low-resolution Landsat satellite images were only available (Bochenek, Ciołkosz, & Iracka, 1997; Jarocińska et al., 2014). The advent of hyperspectral sensing using, for example, the Airborne Prism Experiment (APEX), an in-depth vegetation study can be conducted (Jarocińska et al., 2016; Peerbhay, Mutanga, & Ismail, 2013; Masaitis & Mozgeris, 2013; Thenkabail, Lyon, & Huete, 2012).

## 2. MATERIALS AND METHODS

### 2.1 Leaf sampling

The data were used in this research study in Enschede, Netherlands, between July and September 2010. Leaves were mixed to a total of thirteen plant species (Table 1), eleven of which were local and two were tropical (*Asplenium nidus* and *Spathiphyllum*, local nursery). The species was identified by an expert taxonomist. In order to sample the positional variation, the leaves were collected from the lower, middle and upper part of the tree, both on the sun and the shaded side, to the laboratory within 5 minutes, and placed in moist cotton to avoid desiccation. Spectral measurements were recorded as soon as possible.

#	Species Name	Common Name	Code	Sample size
1	<i>Acer platanoides</i>	Norway maple	AP	35
2	<i>Asplenium nidus</i>	Hawai'i birdnest fern	AN	35
3	<i>Cornus sericea</i>	Redosier dogwood	CS	35
4	<i>Fallopia japonica</i>	Japanese knotweed	FJ	35
5	<i>Ginkgo biloba</i>	Maidenhair tree	GB	35
6	<i>Hedera helix</i>	English ivy	HH	35
7	<i>Ilex opaca</i>	Ilex	IL	35
8	<i>Liquidambar styraciflua</i>	Sweetgum	LS	35
9	<i>Platanus orientalis</i>	Oriental planetree	PO	35
10	<i>Prunus laurocerasus</i>	Cherry laurel	PL	35
11	<i>Rhododendron caucasicum</i>	Rhododendron	RH	35
12	<i>Spathiphyllum cochlearispathum</i>	Peace Lily	SP	35
13	<i>Tilia platyphyllos</i>	Largeleaf linden	TP	35

**Table 1. Thirteen Plant Species Common Name, Code, And the Number of Leaves Per Species Measured Using A Laboratory Spectroradiometer (2.5 Mm-14 Mm).**

### 2.2 Leaf spectral measurement

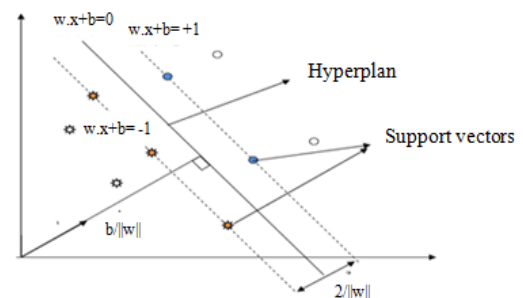
The Directional Hemispherical Reflectance (DHR) spectrum is a spectrometer of Bruker VERTEX 70 FTIR (Fourier transform infrared, Bruker Optics GmbH, Ettlingen, Germany). The spectrometer was continuously purged with water vapor and carbon dioxide using nitrogen (N<sub>2</sub>) gas. The measured leaves

were collected from at least 10 different plants, with high reflectance (about 0.96). The DHR spectra of the plant were measured between 2.5 and 14 μm, with a spectral resolution of 4 cm. Leaf sample was placed on a black surface to absorb transmitted radiation through the leaf and to minimize transparency in the 3.5 and 5.7 μm range, as suggested by Gerber et al. 2011. Each leaf spectrum was derived from the averaging of 1000 scans. The spectral region was 6-8 μm was excluded from the analysis because the spectra were noisy in the region. The applicability of. The Kirchoff's law is based on a black surface.

### 2.3.1 Support Vector Machine (SVM)

Support vector machine (SVM) is a supervised classifier, it has been proposed by Vapnik. This classifier has been introduced to solve two-class pattern recognition problems using the Structural Risk Minimization principle. Given a training set in a vector space, this method finds. The best decision hyper-plane that separates a set of positive examples from a set of negative examples with maximum margin.

A. Optimal hyper-plane in the linear separable case  
 Considering the training data  $(x_i, y_i)$ , for all  $1 \leq i \leq n$ ,  $y_i \in \{-1, 1\}$ ,  $x_i \in \mathbb{R}^d$ . The points  $x$  which belongs to the hyperplane satisfy  $(w \cdot x) + b = 0$ , where  $w$  is normal to the hyperplane,  $\frac{b}{\|w\|}$  is the perpendicular distance from the hyperplane to the origin, and  $\|w\|$  is the Euclidean norm of  $w$  (Fig.1). Define the decision function[42].



**Figure 1: A Linear SVM Two-Dimensional Space**

$$f(x) := \text{sign}((w \cdot x) + b). \quad (3)$$

For the linearly separable case, the support vector algorithm simply looks for the separating hyperplane with largest margin. This can be formulated as follows: Suppose that all the training data satisfy the following constraint.

$$w \cdot x_i + b \geq +1 \quad \text{for } y_i = 1; \quad (1)$$

$$w \cdot x_i + b \leq -1 \quad \text{for } y_i = -1. \quad (2)$$

These can be combined into one set of inequalities:

$$y(w \cdot x_i + b) \geq +1 \quad (3)$$

The optimal hyperplane is the hyperplane that maximizes the margin between the samples and the separating hyper-plane which is equal to  $2/\|w\|$ .

Lagrange multipliers; in this case, the decision function is defined by:

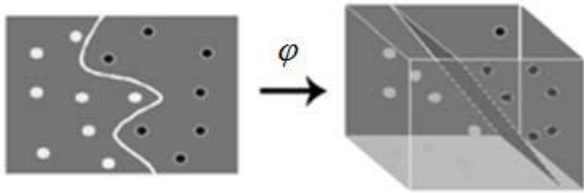
$$f(x) = \text{sign}(\sum_{i=1}^n a_i y_i (x, x_i) + b) \quad (4)$$

by satisfying This problem can be solved by the use of their conditions:

$$a_i [y_i (w \cdot x_i + b) - 1] = 0 \quad \forall i = 1 \dots n, a_i \neq 0. \quad (5)$$

B. Optimal hyper-plane in the nonlinear separable case

To handle nonlinearly separable classes, a nonlinear transformation  $\phi$  is used to map the original data points into a higher dimensional space, in which the data points are linearly separable, called re-description space Figure 2.



**Figure 2: Linear Discrimination in Re-Description Space**

In this case, the equation of the separator hyperplane is  $h(x) = w \varphi(x) + b$  and the decision function is defined by:

$$f(x) = \text{sign}(\sum_{i=1}^n a_i y_i \langle \varphi(x) \varphi(x_i) \rangle + b), \quad (6)$$

The determination of  $\langle \varphi(x_i) \varphi(x_j) \rangle$  is computationally expensive and sometimes impossible to calculate. To solve this problem, several kernel functions have been used, which we can cite as

- Linear kernel:  
 $K(x, x_i) = x^t * x_i. \quad (9)$

- Polynomial kernel: for degree  $d$  polynomials, the polynomial kernel is:

$$K(x, x_i) = (x^t * x_i + b)^d, \quad b > 0. \quad (10)$$

- Radial Basis Function kernel (RBF):

$$K(x, x_i) = \exp\left(-\frac{|x - x_i|^2}{2\sigma^2}\right), \quad \sigma \neq 0. \quad (11)$$

SVMs have shown a good performance for classifying high dimensional data when a limited number of training samples are available. Thus, they have been used for the classification of hyperspectral images. Indeed, many classification approaches have been proposed. In authors developed spectral-spatial classification techniques capable to consider spatial dependences between pixels. It is illustrated with two different classifiers: multinomial logistic advanced spectral-spatial classification techniques. In the first step, a probabilistic SVM pixel-wise classification of the hyperspectral image has been applied. In the second step, spatial contextual information has been used for refining the classification results obtained in the first step. This is achieved by means of the MRF regularization. Experimental results prove that the proposed techniques yield good classification accuracies. In authors developed a new spatio-spectral classification method[1]. The parameter C specifies the allowable classification of non-separable training data, allowing training rigidity to be remote sensing. The gamma parameter is a kernel display parameter that specifies the smoothing of the extra plane shape that divides the class.

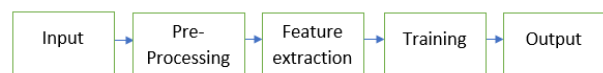
### 2.3.2 Artificial Neural Network

Artificial Neural Networks (ANN) is a nonlinear model that is similar to a human neural system. Each ANN is a series of nodes and weighted connections between them (Carvajal et al., 2006). One of the privileges or ANN method in comparison with comparison Traditional statistical methods The training and recalling are dependent on the linear relationship between data patterns and independent input data (Jayas and Paliwal et al., 2006, Civco and Waug, 1994). However, the reasons for the success of the ANN in the classification Can be Summarize as: there is no need for pre-assumption in data distribution, it allows the user to use initial knowledge about classes and possible limits . the method allows the management of the spatial data from Multiple sources and can achieve their classification wheat. Results equally

(Carvajal et al., 2006). are assigned Out of 826 detected pixels in the previous stage, 124 were used as training data and the rest were left for algorithm evaluation. He classification algorithm was run once for each of the three species (barley, alfalfa, and wheat). At the end, the output images were composed in an RGB image.

### 2.3.3 Convolutional Neural Network (CNN)

Deep learning is a very effective method for learning optimum features from large amount of training datasets automatically. Trend towards deep learning in computer vision applications is increasing tremendously because modern deep learning methods are more accurate than humans in classifying images. Furthermore, the use of Graphical Processing Units (GPUs) have decreased the training time of deep learning methods. Large databases of labelled data and pre-trained networks are now publicly available.

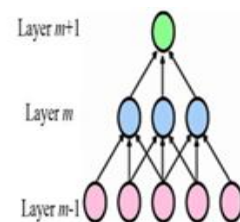


**Figure 3: Traditional image recognition methods model**



**Figure 4: Convolutional neural networks model.**

Three crucial characteristics make the convolutional neural networks different from other neural networks. Generally, these three characteristics are all aiming at dimensionality reduction to reduce calculation time. The first characteristic is sparse connectivity, which means that there are only local connections between neurons of adjacent layers. In other words, the inputs of hidden units in layer  $m$  are from a subset of units in layer  $m-1$ , units that have spatially contiguous receptive fields. The sparse connectivity can be illustrated in figure 5. However, in the traditional neural network structure, every neuron in layer  $n$  connects with all the neurons in layer  $m-1$ . Therefore, by sparse connectivity, parameters for neurons in layer  $m$  decrease a lot. The theory for supporting sparse connectivity is that local pixels have strong correlations while long-distance pixels have weak correlations.



**Figure 5: Sparse Connectivity**

The second characteristic is shared weights. In neural networks, neurons in layer  $m-1$  are the variables or inputs to neurons in layer  $m$ . And neurons in layer  $m$  can be seen as outputs. Every input has a weight to determine the output. As show in figure 6, each neuron in layer  $n$  shares the same weights and bias. All neurons in the same layer form a feature map. It reduces the number of parameters for the network.

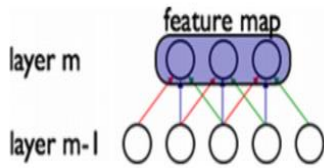


Figure 6: Shared weights Colour identifies the weight

The last characteristic is pooling, which is a form of down sampling. Pooling means aggregate features in a rectangle neighborhood into one feature, as show in figure 7. The figure on the left in figure 11 is a feature map, pooling reduces the dimension and avoids overfitting.

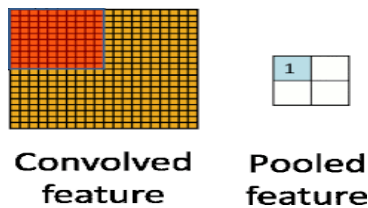


Figure 7: Pooling

Krizhevsky et al. achieved excellent recognition rates on Large Scale Visual Recognition Challenge dataset using standard backpropagation for training a deep CNN.

### 3. RESULT AND DISCUSSION

#### 3.1 Support Vector Machine Line Kernel Classifier (SVM-Line Kernel):

One of the main results of Statistical Learning Theory is that the error probability of a classifier is upper bounded by a quantity depending not only on the error rate achieved on the training set, but also on an intrinsic property of the classifier, which is a measure of the “richness” of the set of decision functions it can implement (Roliand Fumera, 2000) . In the two-class case, a support vector classifier attempts to locate a hyper plane that maximizes the distance from the members of each class to the optimal hyper plane figure 5. Assume that the training data with  $k$  number of samples is represented by where  $X$  is an  $n$ -dimensional vector and  $Y \in \{-1, +1\}$  is the class label. These training patterns are said to be linearly separable if a vector  $w$  (which determines the orientation of a discriminating plane) and a scalar  $b$  (determines offset of the discriminating plane from origin) can be defined so that inequalities (12) and (13) are satisfied.

$$W.X_i + b \geq +1 \quad \text{for } Y_i = 1 \quad (12)$$

$$W.X_i + b \leq -1 \quad \text{for } Y_i = -1 \quad (13)$$

$$Y(W.X_i + b) \geq +1 \quad (14)$$

The aim is to find a hyper plane which divides the data so that that all the points with the same label lie on the same side of the hyper plane. This amounts to finding  $w$  and  $b$  so that This results in a quadratic optimization problem which is solved by a standard numerical optimization package. The minimization procedure uses Lagrange multipliers and Quadratic Programming (QP) optimization methods. If  $\lambda_i, i = 1, \dots, k$  are the non-negative Lagrange multipliers associated with constraint (7), The training vectors  $x$  are solely used in inner products which can be replaced by a kernel function  $K(x, y)$  that obeys Mercer’s condition. This is equivalent to mapping the feature vectors into a high-dimensional feature space before using a hyper plane classifier there. The use of kernels makes it possible

to map the data implicitly into a feature space and to train a linear machine in such a space, potentially side-stepping the computational problems inherent in evaluating the feature map (Cristianni and Shawe-Taylor, 2000). In this research, three kinds of kernels are studied. These kernels are mathematically defined as below (Chang and Lin, 2003).

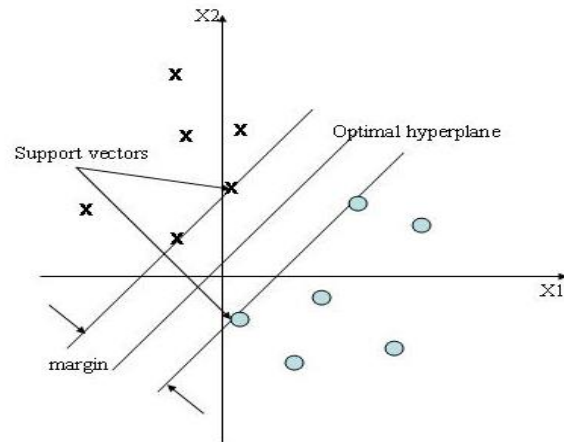


Figure 8: SVM classification with a hyper plane that maximizes the separating margin between the two classes (indicated by data points marked by "x"s and "O"s). Support vectors are elements of the training set that lie on the boundary hyper planes of  $t$ .

It is possible to rescale  $w$  and  $b$  so that

$$y(w.x + b) \geq 1 \quad (15)$$

and minimizing  $1/w$  under these constraints. This concept can be extended to the case when the classes are not linearly separable, i.e. when (3) has no solution. A slack variable,  $\xi_i, i = 1, \dots, k$  can be introduced such that can be written as while the objective function is supplemented to keep the constraint violation as small as possible:

#### 3.1.1 Kernel selection

Three kernels were studied and the results summarized in Table 1. The Polynomial kernel (Equation 10) of degree 3 ( $d=3$ ) and cost ( $C=1000$ ) showed the best accuracy on test data.  $C$  refers to the cost or error penalty. A high value of error penalty will force the SVM training to avoid classification errors. A large value of  $C$  will result in a larger search space for the QP optimizer. However, some experiments fail to converge for  $C > 1000$ . In the kernels under study, a value of  $C=1000$  was optimum. For the RBF kernel (Equation 11) gamma (radius) was set to 1. This gives the area of influence the particular support vector has over the data space. The RBF

kernel was experimented with different values of  $\nu$ . Nu-SVC is the same as C-SVC except that the range of  $\nu$  is always between  $[0, 1]$  while  $C$  is from zero to infinity.  $\nu$  is related to the ratio of support vectors and the ratio of the training error. Three SVM models have been used for further study: The degree 3 polynomial kernel, RBF kernel with  $\nu=0.01$  and the linear kernel (Equation 11); the cost for all the kernels set to 1000. Table 2 shows that the linear kernel highest accuracy of 91% while RBF ( $\nu=0.1$ ) is 87% and poly kernel is 84%.

#### 3.2 Evaluation of signatures

Confusion matrices: The confusion matrices for the four classifiers with User Accuracy and Producer Accuracy are as shown in Tables 3-6. These matrices help evaluate the classifiers based on their performance on a class-by-class basis. For

example, they show the number of AP species misclassified as LS, AN etc, and also give the count of other classes misclassified as AP. This leads us to find the Users and Producers Accuracy. Users accuracy calculates correctly classed from the trace variable over the row total and provides an indication of errors of case omission. Producers accuracy is the calculation of correctly classed from the trace value over the column total. (Congalton,1991). Producers' accuracy gives an indication of the accuracy of what the model was able to itself predict, whereas User accuracy relates how well the training data was discerned. A look at these matrices suggests that AP and LS have confusing signatures and so do ilex and LN. Clearly, the CNN outperformed the SVM classifier in terms of testing accuracy.

### 3.3 Discussion

In comparing the overall correctacies of the CNN and SVM classifier, CNN was determined to have a static-significant advantage over SVM when the pixel-based reflectance samples used, without the segmentation size. The most misclassified class for both classifiers was *Fallopia japonica*, with class correctacies all above 70%. The *Fallopia japonica* class was misclassified as multiple different species and had fewest number of training samples and consistently low class correctacies. *Acer platanoides* were properly recognized most consistently. A noticeable trait of both classifiers was the highest overall correctacies despite abundant misclassifications found during visual inspection. Duro et al. Experienced similar issues with their classification, highlighting that limited test samples can result in incorrect classification accuracies in the same way, Congalton explains that a large quantity of zeroes within the confusion matrix could mean that the test sample size is inadequate or classification very successful. Considering the number of zeroes on the confusion matrix for the small classes, it is quite possible that insufficient test samples are responsible for some species of high classification correctacies. For this research, the stratified sampling method utilizes single species to ensure that the correct sample identification; However, this prevented test samples are being selected near standing boundaries, in mixed forest areas or inaccessible areas. As a result, the test samples did not represent all the areas of the study site, and thus, the accuracy assessment failed to reflect those performance in classifier. In some cases, an individual species was misclassified as one or two other classes, as shown by the *Acer platanoides* class generally being incorrectly classified as bay or Douglas fir.

As a matter of fact, it's a common tree type (conifers and broadleaf). Some of the broadleaf tree classes were misclassified. The fact that Douglas has been misclassified as to whether there are any misclassifications. Leckie et al. highlighting Shadowing may well explain the misclassification of Douglas fir as coast live throughout the upper portions of the mosaic. Also, it can be used to create a new window. Although a normalization was a normalization, the normalization was unlikely. Because of the results of all classes This may have caused spectral overlap.

The effect of pixel-based training samples played a significant role in SVM and CNN classifications. The difference between object and pixel-based training samples can be seen as an increase in sample size. Although the actual number of samples are both the sample types for the same, pixel-based reflectance samples provide additional spectral reflectance values to classifiers trained.

The current literature generally acknowledges the support of the machine classifier for its ability to work well with limited training samples. SVM only uses the subset of training samples that uses the location of SVM hyperplane. SVM classifier. In the same way, Zhang and Xie found SVM and CNN both trained with a sensitive size and a high spatial and spectral variable. This study may be complexity and heterogeneity of forest and insufficient object-based samples. Considering the effect of training samples on SVM hyperplanes, eliminating object-based samples (based on the mean values of the pixel) may not be able to capture high spectral variations within thirteen tree species in a highly heterogeneous forest as compared to a less complex forest with less classes or less spectral overlap species.

	Ap	An	Cs	Fj	Gb	Hh	Il	Ls	Po	Pl	Rh	Sp	Tp	Total	UA
Ap	32	0	0	0	0	0	0	0	0	1	0	0	0	33	96
An	0	30	0	2	2	0	0	1	0	0	0	0	0	35	85
Cs	0	0	28	0	0	0	0	0	1	0	0	1	0	30	93
Fj	0	0	0	30	0	0	4	0	0	0	1	0	0	35	85
Gb	0	0	0	1	29	0	0	0	0	0	2	0	0	32	90
Hh	0	0	0	0	0	32	0	1	0	2	0	0	0	35	91
Il	0	0	0	0	2	0	30	0	0	0	0	0	0	32	93
Ls	1	0	0	0	0	0	0	32	0	0	0	0	0	33	96
Po	0	0	1	0	0	0	0	0	29	0	1	2	0	32	90
Pl	1	1	0	0	0	0	0	0	0	33	0	0	0	35	94
Rh	0	0	0	0	0	0	1	0	0	0	31	1	0	33	93
Sp	0	0	0	0	0	1	0	0	0	0	0	32	0	33	96
Tp	0	0	0	0	0	0	0	2	0	0	4	1	28	35	80
Total	34	31	29	33	33	33	35	34	30	38	39	37	28	433	
PA	94	96	96	90	87	96	85	94	96	86	79	86	100	OA=91%	

Table 2: Confusion Matrix of Support Vector Machine Used Line Kernel Classifier (SVM-Line Kernel)

	Ap	An	Cs	Fj	Gb	Hh	Il	Ls	Po	Pl	Rh	Sp	Tp	Total	UA
Ap	31	0	1	0	0	1	0	0	0	2	0	0	0	35	88
An	1	27	0	0	0	0	1	1	0	0	1	0	0	31	87
Cs	0	0	28	0	0	0	1	0	2	0	1	0	0	32	87
Fj	1	0	1	26	2	2	0	0	2	1	0	0	0	35	74
Gb	0	0	0	1	20	0	0	0	0	0	0	2	0	23	86
Hh	0	1	0	0	0	26	0	0	0	1	0	0	0	28	92
Il	0	0	0	0	0	0	31	0	0	0	0	0	0	31	100
Ls	0	0	0	0	0	0	0	30	1	0	1	1	0	33	90
Po	1	0	0	1	2	0	0	0	29	0	0	1	0	34	85
Pl	0	0	1	0	0	0	0	0	0	28	0	0	0	29	96
Rh	0	0	0	1	0	1	0	0	0	0	31	0	0	33	93
Sp	0	4	0	0	1	0	0	4	0	2	0	12	0	23	52
Tp	0	0	2	0	0	0	0	0	0	0	1	0	31	33	93
Total	34	32	33	29	25	30	33	35	34	34	35	16	31	399	
PA	91	84	84	89	80	86	93	88	85	82	88	75	31	OA=87%	

Table 2: Confusion Matrix of Support Vector Machine Used Radial Base Function Classifier (SVM-RBF Kernel)

	Ap	An	Cs	Fj	Gb	Hh	Il	Ls	Po	Pl	Rh	Sp	Tp	Total	UA
Ap	35	0	0	0	0	0	0	0	0	0	0	0	0	35	100
An	0	29	0	2	0	0	0	2	0	2	0	0	0	35	82
Cs	0	0	30	0	1	0	0	0	2	0	0	1	0	34	88
Fj	0	0	0	25	0	0	1	0	4	2	2	0	0	34	73
Gb	0	0	0	0	28	1	0	1	0	1	0	1	0	32	87
Hh	0	0	1	0	0	27	0	0	0	0	4	0	0	32	84
Il	0	0	0	0	0	0	21	0	0	4	0	2	0	27	77
Ls	0	0	1	0	0	0	0	30	0	0	2	1	0	34	88
Po	1	1	0	0	1	0	0	0	27	2	0	1	0	33	81
Pl	0	0	0	1	0	1	2	0	0	26	0	2	0	32	81
Rh	0	0	0	2	0	1	1	0	1	0	30	0	0	35	85
Sp	0	0	0	0	2	0	1	0	0	1	0	31	0	35	88
Tp	0	0	0	4	0	2	0	0	2	2	0	0	24	34	70
Total	36	30	32	34	32	32	26	33	36	38	38	39	24	432	
PA	97	96	93	73	87	84	80	90	75	68	78	79	100	OA=84%	

Table 3: Confusion Matrix of Support Vector Machine used Poly Kernel Classifier (SVM-Poly kernel)

	Ap	An	Cs	Fj	Gb	Hh	Il	Ls	Po	Pl	Rh	Sp	Tp	Total	UA
Ap	10	0	0	0	0	0	0	0	0	0	0	0	0	10	100
An	0	12	1	0	0	0	0	0	0	0	0	0	0	13	92
Cs	0	0	4	0	0	0	0	0	0	0	0	0	0	4	100
Fj	0	0	0	3	0	0	1	0	0	0	2	0	0	4	75
Gb	0	0	0	0	2	0	0	0	0	0	0	0	0	2	100
Hh	0	0	0	0	0	6	0	0	0	0	4	0	0	6	100
Il	0	0	0	0	0	0	2	0	0	0	0	0	0	2	100
Ls	0	0	1	0	0	0	0	9	0	0	2	0	0	10	90
Po	0	0	0	0	0	0	0	0	8	0	0	0	0	8	100
Pl	0	0	0	0	0	0	0	0	0	5	0	0	0	5	100
Rh	0	0	0	0	0	0	0	0	1	0	6	0	0	6	100
Sp	0	0	0	0	2	0	0	0	0	0	0	15	0	17	88
Tp	0	0	0	0	0	0	0	0	0	0	0	0	6	6	100
Total	10	12	5	3	2	6	3	9	9	5	6	15	8	93	
PA	100	100	80	100	100	100	66	100	88	100	100	100	75	93	OA=94%

Table 4: Confusion Matrix of Artificial Neural Network

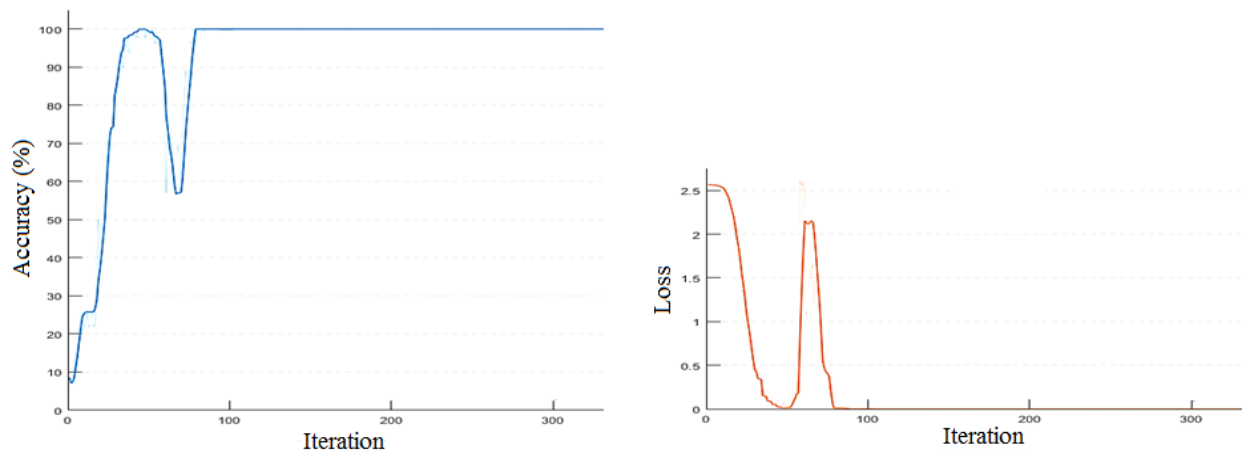


Figure 9: CNN Training Accuracy plot

	Ap	An	Cs	Fj	Gb	Hh	Il	Ls	Po	Pl	Rh	Sp	Tp	Total	UA
Ap	33	0	0	0	0	1	0	0	0	0	0	0	0	33	97
An	0	35	0	0	0	0	0	0	0	0	0	0	0	35	100
Cs	0	0	34	0	0	0	0	0	0	0	0	0	0	34	100
Fj	0	0	0	32	0	0	1	0	0	0	0	0	0	33	96
Gb	0	0	0	0	34	0	0	0	0	0	0	0	0	34	100
Hh	0	0	1	0	0	33	0	0	0	0	0	0	0	34	97
Il	1	0	0	0	0	0	30	0	0	0	0	0	0	31	96
Ls	0	0	0	1	0	0	0	34	0	0	0	0	0	35	97
Po	0	0	0	0	0	0	0	0	30	0	0	0	0	30	100
Pl	0	0	0	0	0	0	0	0	0	29	0	0	0	29	100
Rh	1	0	0	0	0	0	0	0	0	0	34	0	0	35	97
Sp	0	0	0	0	0	0	0	0	0	0	0	29	0	29	100
Tp	0	0	0	0	0	0	0	0	0	0	0	0	34	34	100
Total	35	35	35	33	34	34	31	34	30	29	34	29	34	426	
PA	94	100	97	96	100	97	96	100	100	100	100	100	100	100	OA=99%

Table 5: Convolutional Neural Network Classifier (CNN)

#### 4. CONCLUSION

This study compares the classification accuracy of convolution neural network artificial neural network and support vector machine on thirteen forest-vegetation specie. SVM classifier is experimented with linear, RBF and polynomial kernels. CNN has higher overall accuracy as compared to the three SVM types.

And artificial neural network. We can conclude this things CNN is better then conventional morphological methods for feature extraction in plant species.

The data taken from Bruker VERTEX 70 FTIR spectrometer (Bruker Optics GmbH, Ettlingen, Germany) was used to acquire the Directional Hemispherical Reflectance (DHR) spectrum of each leaf. Experimental results indicate that CNN based approach is significantly effective with an overall accuracy of about 99%. Upon increasing the training data, the classification accuracy of both SVM and CNN improves. However, the size of

objects in the object-based classification has no significant impact on both classifiers' performance.

Due to the statistically significant classification of thirteen vegetation species, further depth in the study would give more insights. The deep learning classification and pre-processing methods will be further explored to enhance the classification performance.

## REFERENCES

- H. Zhou, C. Yan, H. Huang, Tree Species Identification Based on Convolutional Neural Networks, in: 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2016, pp. 103-106.
- J. Leng, T. Li, G. Bai, Q. Dong, H. Dong, Cube-CNN-SVM: A Novel Hyperspectral Image Classification Method, in: 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI), 2016, pp. 1027-1034.
- H. Yalcin, S. Razavi, Plant classification using convolutional neural networks, in: 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics), 2016, pp. 1-5.
- R. Pu, An exploratory analysis of in situ hyperspectral data for broadleaf species recognition, Publisher, City, 2008.
- A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Publisher, City, 2004.
- J.A. Gualtieri, R.F. Crompton, Support vector machines for hyperspectral remote sensing classification, in: 27th AIPR Workshop: Advances in Computer-Assisted Recognition, International Society for Optics and Photonics, 1999, pp. 221-233.
- E. Raczko, B. Zagajewski, Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images, Publisher, City, 2017.
- S. Ullah, M. Schlerf, A.K. Skidmore, C. Hecker, Identifying plant species using mid-wave infrared (2.5–6  $\mu\text{m}$ ) and thermal infrared (8–14  $\mu\text{m}$ ) emissivity spectra, Publisher, City, 2012.
- N.I.S. Bahari, A. Ahmad, B.M. Aboobaidar, Application of support vector machine for classification of multispectral data, in: IOP Conference Series: Earth and Environmental Science, IOP Publishing, 2014, pp. 012038.
- M.D. Atkinson, A.P. Jervis, R.S. Sangha, Discrimination between *Betula pendula*, *Betula pubescens*, and their hybrids using near-infrared reflectance spectroscopy, Publisher, City, 1997.
- P. Gong, R. Pu, B. Yu, Conifer species recognition: An exploratory analysis of in situ hyperspectral data, Publisher, City, 1997.
- M. Ge, R. Du, G. Zhang, Y. Xu, Fault diagnosis using support vector machine with an application in sheet metal stamping operations, Publisher, City, 2004.
- S. Ullah, A.K. Skidmore, A. Ramoelo, T.A. Groen, M. Naeem, A. Ali, Retrieval of leaf water content spanning the visible to thermal infrared spectra, Publisher, City, 2014.
- M.J. Khan, K. Khurshid, F. Shafait, INK MISMATCH DETECTION IN HYPERSPECTRAL DOCUMENT IMAGES USING DEEP LEARNING, Publisher, City, 2017.
- P.E. Dennison, K.Q. Halligan, D.A. Roberts, A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper, Publisher, City, 2004.
- P.E. Dennison, D.A. Roberts, Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE, Publisher, City, 2003.
- A.N. Schaaf, Using hyperspectral data to classify vegetation at the plant functional type-level in mountain terrain at three spatial resolutions, The University of Utah, 2010.
- P.R. Lawrence, J.W. Lorsch, Organization and environment: managing differentiation and integration (Harvard Business School Classics), Publisher, City, 1986.
- D.L. Rodland, D.J. Bottjer, Biotic recovery from the end-Permian mass extinction: behavior of the inarticulate brachiopod *Lingula* as a disaster taxon, Publisher, City, 2001.
- P.S. Thenkabail, E.A. Enclona, M.S. Ashton, B. Van Der Meer, Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications, Publisher, City, 2004.
- R.E. Clark, N.J. Broadbent, L.R. Squire, Impaired remote spatial memory after hippocampal lesions despite extensive training beginning early in life, Publisher, City, 2005.
- M. Sano, T. Minamino, H. Toko, H. Miyauchi, M. Orimo, Y. Qin, H. Akazawa, K. Tateno, Y. Kayama, M. Harada, p53-induced inhibition of Hif-1 causes cardiac dysfunction during pressure overload, Publisher, City, 2007.
- E.L. Hestir, S. Khanna, M.E. Andrew, M.J. Santos, J.H. Viers, J.A. Greenberg, S.S. Rajapakse, S.L. Ustin, Identification **There are no sources in the current document.** of invasive vegetation using hyperspectral remote sensing in the California Delta ecosystem, Publisher, City, 2008.
- A. Baldridge, S. Hook, C. Grove, G. Rivera, The ASTER spectral library version 2.0, Publisher, City, 2009.
- C. Wong, W. Blevin, Infrared reflectances of plant leaves, Publisher, City, 1967.
- C.R. de Almodovar, P.J. Fabre, E. Knevels, C. Coulon, I. Segura, P.C. Haddick, L. Aerts, N. Delattin, G. Strasser, W.-J. Oh, VEGF mediates commissural axon chemoattraction through its receptor Flk1, Publisher, City, 2011.
- N. Jafar, S.A. Thamrin, A. Lawi, Multiclass classification using Least Squares Support Vector Machine, in: Computational Intelligence and Cybernetics (CYBERNETICSCOM), 2016 International Conference on, IEEE, 2016, pp. 7-9.
- M. J. Khan, A. Yousaf, A. Abbas, K. Khurshid, "Deep Learning for Automated Forgery Detection in Hyperspectral Document Images", Journal of Electronic Imaging, 27(5): 053001, 2018.
- M. Arshad, S. Ullah, K. Khurshid, A. Ali, "Estimation of Leaf Water Content from Mid and Thermal Infrared spectra by coupling Genetic Algorithm and Partial Least Squares Regression", Journal of Applied Remote Sensing, 12(2): 022203 1