ISSN: 0974-2115
www.jchps.com Journal of Chemical and Pharmaceutical Sciences

Aerial Image Classification Using Regularized Least Squares Classifier

Sharanya Chandran, Sajith Variyar V V, Nidhin Prabhakar T V, Soman K P
Centre for Excellence in Computational Engineering and Networking
Amrita Vishwa Vidyapeetham, Coimbatore, Tamil Nadu
***Corresponding author: E-Mail: sharanyachandranz@gmail.com, nidhin89@gmail.com**

ABSTRACT

The land cover classification and urban analysis of remotely sensed images has become a challenging problem, hence efficient classifiers are required in order to combat the problem of classifying the huge remote sensing aerial datasets. In this paper we have proposed the use of Random Kitchen Sink (RKS) algorithm and Regularized Least Squares (RLS) classifier for the classification of aerial image. The new machine learning algorithm RKS, primarily engages in mapping the feature data to a higher dimensional space and thereby generates random features. These randomized data are then adopted by RLS classifier for the classification task. It is observed that the randomization of the data reduces the computation time needed for training. The experiment is performed on five classes of the UC Merced Land Use Aerial Imagery Dataset. The efficiency of the proposed method is estimated by comparing the accuracy results with the conventional classifier namely, Support Vector Machine (SVM). Experimental result shows that the proposed method produces a high degree of classification accuracy i.e. 94.4%, when RBF kernel with LOO (Leave One Out) cross-validation was used, when compared to SVM. In this paper, statistical features show better precision and accuracy in classifying different set of classes, compared to textural features in both the classification approaches. Hence, better accuracies could be attained for multi class classification when compared to other classification technique like, SVM since, the random features reduces computation time and enhance the performance of kernel machines.

KEY WORDS: Classifier, Random Kitchen Sink, Regularized Least Squares, Aerial Imagery, Support Vector Machines, Kernel.

1. INTRODUCTION

Remote sensing imagery is an inevitable tool for studying the earth's environment, wherein remotely sensed aerial images are an exemplary data source for many applications including land use surveys and habitat analysis. Acquisition of this huge volume of aerial imagery usually surpasses the capacity of manual processing to carry out timely and cost-effective feature extraction process. In this context, the important task of classifying aerial images into various classes is crucial. A convenient classification system and a competent number of training samples are mandatory for a successful classification. Hence, a reliable and effective solution is supervised classification, where the user should have prior knowledge of where the classes of interest are located, or should be able to identify them visually.

In the near future, numerous experts have made efforts in improving the accuracy of the classification results by developing new techniques for classification. The segmentation and classification of aerial images based on pixel level have been carried out and comparisons of different classifiers k-Nearest Neighbors (k-NN), SVM and a sparse based classifier have been carried out (Ahmadi, 2013). A variety of feature extraction techniques namely, statistical, texture, Discrete Cosine Transform (DCT)and Discrete Wavelet Transform (DWT) have been employed, and these features have been evaluated on different classifiers like SVM, k-NN, Decision Tree, Nave Bayesian and Ensemble based classifiers for the classification of crops using aerial data set (Ul Qayyum, 2013). A supervised classification for aerial photographs and lidar data was performed using general Bayesian learning (Juho Lumme, 2006).

Normalized Difference Vegetative Index (NDVI) and Green Normalized Difference Vegetation Index (gNDVI) approach have been used for land cover classification of aerial images (Maciej, 2012). Another paper describes modified co-occurrence matrices combined with neural network of the type adaptive threshold learning for the purpose of exploiting spatial information and performing land-use classification (Bock, Steffen, 1996). A probabilistic neural network (PNN) to classify the aerial images was introduced in (Sheikh, 2012)which included a three-stage feature extraction involving edge extraction from input gray image, Gabor filter applied for the computation of Gabor energy feature and wavelet decomposition in order to extract the feature vector.

In the last few years, kernel based classification and regression approaches such as SVM have been extensively used in the field of remote sensing (Soman, 2009). The spectral and spatial information were used together for the purpose of classification with the aid of kernel methods through SVM formulation (Fauvel, 2012).

The use of kernel methods for classification, regression and ways to reduce the training to optimization of a convex cost function has been discussed in (Colin Campbell, 2000), here authors have also em-phasized the use of RBF (Radial Basis Function) kernels. A texture classification based on random projection approach for large texture database applications was accomplished in (Liu, 2012). The paper (Cesa-Bianchi, 2007), aimed at analyzing the fundamental properties of RLS for the purpose of classification, and due to the tractability and empirical behavior, RLS has been proved good method for solving learning problems. In (Rahimi Ali and Benjamin Recht, 2009), it was denoted that the same accuracy could be obtained when many random non-linearities are selected as in the case of greedy algorithm which selects the non-linearities in significantly less training time. Thereby, the role of the non-linearities in a shallow network is to transform the inputs so that they are linearly separable by the upper layer of the network. Hence, random features when combined with simple linear learning techniques compete favorably with kernel-based classification and regression algorithms (Rahimi Ali, Benjamin Recht, 2009).

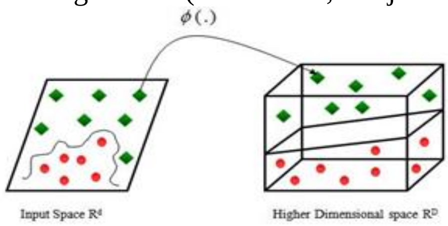

Fig.1. Transformation of the input space to a higher dimensional space

Random Kitchen Sink: Random Kitchen Sink algorithm is a new machine learning algorithm introduced for classification of non-linearly separable data set. This method is suitable for learning and classifying large data sets. Usually large proportion of data points become support vectors when large data set is involved in training non-linear SVM, thus it must be stored for classifying new data points which in turn demands more space and time for classification. In such a context, Random Kitchen Sink algorithm is a suitable alternative. Rather than finding a function curve to separate the data, in this algorithm the data is mapped to a higher dimensional space and a decision surface is derived in that space.

The inner products of the transformed data are approximated to obtain the kernel function (Rahimi Ali and Benjamin Recht, 2007). The choice of the kernel function is made according to the application performed by the end user.

$$K(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle \qquad (1)$$

Where, $K(x_1, x_2)$ represents the positive definite kernel function, $\phi(x_1)$ and $\phi(x_2)$ represent the mapping function. Kernel function widely accepted in remote sensing applications is the radial basis function (RBF). The RKS approximates an RBF kernel of the form,

$$e^{\frac{-1}{\sigma}\| x_1 - x_2 \|_2^2} = e^{\frac{-1}{\sigma}(x_1 - x_2)^T (x_1 - x_2)} \qquad (2)$$

$$\text{Where, } \Sigma = \begin{bmatrix} 2\sigma & 0 & \dots & 0 \\ 0 & 2\sigma & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 2\sigma \end{bmatrix} \qquad (3)$$

The RBF kernel is a Gaussian function. This kernel is expressed as a Gaussian probability density function. Since covariance matrix, $\sum$ is diagonal, the Gaussian density function can be expressed as product of n Gaussian function and hence the associated variables are independent when kernel is interpreted as probability density function.

$$z = x_1 - x_2 \qquad (4)$$

z is a single vector variable function, here x_1, x_2 signifies a data pair. Therefore the kernel function can be written as

$$f(z) = e^{\frac{-1}{2} z^T \Sigma^{-1} z} \qquad (5)$$

Let the Fourier Transform of $f(z)$ be $F(\Omega)$. That is

$$F(\Omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(z) e^{-jz^T \Omega} dz \qquad (6)$$

$F(\Omega)$ can be interpreted as a Gaussian (multivariate) density function.

Now as

$$F^{-1}(\Omega) = \langle \phi(x_1), \phi(x_2) \rangle = \int_{-\infty}^{\infty} F(\Omega) e^{jz^T\Omega} dz \qquad (7)$$

Hence, according to the Bochners theorem, the inverse Fourier transform of an RBF kernel can be illustrated as the expected value of the quantity $e^{jz^T\Omega}$. That is

$$E(e^{jz^T\Omega}) = \int_{-\infty}^{\infty} F(\Omega_i) e^{jz^T\Omega} dz \qquad (8)$$

Assuming the above integration as sampling several Ω_i vectors from Gaussian (multivariate) density function and there by taking average of the functions inside the integral over k samples.

$$\begin{aligned} E(e^{jz^T\Omega}) &= \frac{1}{k}\sum_{i=1}^{k} e^{jz^T\Omega_i} \Rightarrow \frac{1}{k}\sum_{i=1}^{k} e^{j(x-y)^T\Omega_i} \\ &= \frac{1}{k}\sum_{i=1}^{k} e^{jx^T\Omega_i}\overline{e^{jy^T\Omega_i}} \Rightarrow \langle \phi(x_1), \phi(x_2) \rangle \\ &= K(x_1, x_2) \qquad (9) \end{aligned}$$

Here, $K(x_1, x_2) = \frac{1}{k}\begin{bmatrix} e^{j(x_1-x_2)^T\Omega_1} \\ e^{j(x_1-x_2)^T\Omega_2} \\ \vdots \\ e^{j(x_1-x_2)^T\Omega_k} \end{bmatrix}$

$$= \frac{1}{k}\left\langle \begin{bmatrix} e^{jx_1^T\Omega_1} \\ e^{jx_1^T\Omega_2} \\ \vdots \\ e^{jx_1^T\Omega_k} \end{bmatrix}, \begin{bmatrix} e^{jx_2^T\Omega_1} \\ e^{jx_2^T\Omega_2} \\ \vdots \\ e^{jx_2^T\Omega_k} \end{bmatrix} \right\rangle \qquad (10)$$

$$= \left\langle \begin{bmatrix} \sqrt{1/k}e^{jx_1^T\Omega_1} \\ \sqrt{1/k}e^{jx_1^T\Omega_2} \\ \vdots \\ \sqrt{1/k}e^{jx_1^T\Omega_k} \end{bmatrix}, \begin{bmatrix} \sqrt{1/k}e^{jx_2^T\Omega_1} \\ \sqrt{1/k}e^{jx_2^T\Omega_2} \\ \vdots \\ \sqrt{1/k}e^{jx_2^T\Omega_k} \end{bmatrix} \right\rangle \qquad (11)$$

From the above equation it is inferred that,

$$\phi(x_1) = \begin{bmatrix} \sqrt{1/k}e^{jx_1^T\Omega_1} \\ \sqrt{1/k}e^{jx_1^T\Omega_2} \\ . \\ \sqrt{1/k}e^{jx_1^T\Omega_k} \end{bmatrix}, \; \phi(x_2) = \begin{bmatrix} \sqrt{1/k}e^{jx_2^T\Omega_1} \\ \sqrt{1/k}e^{jx_2^T\Omega_2} \\ . \\ \sqrt{1/k}e^{jx_2^T\Omega_k} \end{bmatrix}$$

Inorder to avoid the complex number computation while creating finite k dimensional $\phi(x)$, we can equivalently take sines and cosines of the higher dimensional mapped data.

$$\phi(x) = \sqrt{1/k}\begin{bmatrix} Cos(x^T\Omega_1) \\ \vdots \\ Cos(x^T\Omega_k) \\ Sin(x^T\Omega_1) \\ \vdots \\ Sin(x^T\Omega_k) \end{bmatrix} \quad (12)$$

Thenceforth obtaining $\phi(x)$ for all the data points, any linear classifier could now be used for the classification task.

Regularized Least Square Regression: Regularized least squares method is a dynamic, computationally efficient and elementary tool for supervised learning (Tacchetti, 2013). Since, RLS reduces to solving a linear system, this method also exploits recent advances of linear algebra. The fundamental properties of RLS that are particularly conformed for high dimensional learning are: it has natural primal and dual formulation; efficient parameter selection; natural and efficient extension to multiple outputs.

A multi-class classification problem can be formulated as,

$$W^* = \arg\min_{W^{m\times n}} \| Y - XW \|_F^2 \quad (13)$$

Consider the data set containing k classes. Total number of objects is m and the number of features is n .The size of data matrix is m x n, where, each data present in rows corresponds to an object. Let the class value be k tuple. Each row of matrix

Y is of size m x k , which consists of the corresponding label vectors. A weighted matrix W of size n x k maps n-tuple data vector to corresponding label vector that contains n x k is

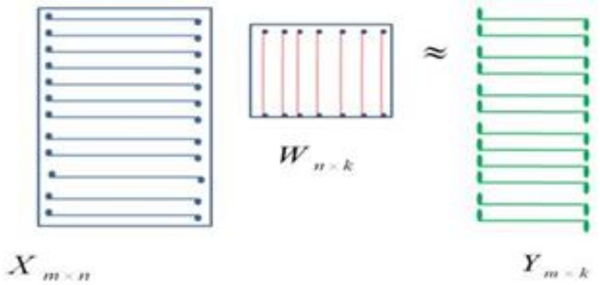

Fig.2. Data Matrix X; Weighted matrix W,Y matrix comprising of corresponding label vectors

Often over-fitting of the data occurs due to approximation of the function using higher degree polynomials, resulting in meaningless prediction. This over-fitting problem can be avoided by adding a regularization term to the above minimization expression.

$$W^* = \arg\min_{W^{m\times n}} \| Y - XW \|_F^2 + \lambda \| W \|_F^2 \quad (14)$$

λ is the control parameter, changing the value of λ changes the amount of the over-fitting. An advisable λ value can be obtained through trial and error method.

$$f(W) = \| Y - XW \|_F^2 + \lambda \| W \|_F^2 = Tr\left(\left(Y - XW\right)^T\left(Y - XW\right) + \lambda W^T W\right) \quad (15)$$

Further solving by taking derivative and equating to zero,

$$\frac{\partial f(W)}{\partial W} = 2\left(X^T X\right)W - X^T Y - X^T Y + 2\lambda W \Rightarrow 0,\; W^* = \left(X^T X + \lambda I\right)^{-1} X^T Y \quad (16)$$

The output class labels are predicted, with the help of weight matrix, W calculated above.

2. METHODOLOGY

The study proposes an evaluation to exploit the use of Random Kitchen Sink and Regularized Least Squares for aerial image classification on the UC Merced dataset. The proposed method involves feature extraction, storage of extracted features, labeling these features, classification, and evaluation of the best classification method.

The feature extraction technique implicates extracting the attributes that best describes the image, here statistical and textural features were evaluated. The statistical features like Mean, Variance, Standard Deviation, Median, Range, Kurtosis, Maxima, Minima and the textural features like GLCMS(Contrast, Correlation, Energy, Homogeneity) are derive from gray level co-occurrence matrices (Haralick, 1973) Entropy and threshold value of the image were extracted from the whole image. The extracted features were stored and labeled according to the respective classes. The aerial data is then transformed to a higher dimension feature space using randomized feature map i.e., $z: R^d \rightarrow R^D$ using Random Kitchen Sink, and then the inner product between a pair of transformed points approximates the kernel evaluation,

$$K(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle \approx z(x_1)^{'} z(x_2) \qquad (17)$$

Here, $\phi(x_1)$ and represents the mapping function and $K(x_1, x_2)$ represents the positive definite kernel function (Rahimi Ali and Benjamin Recht, 2009). This sort of transformation to a higher dimensional space helps in better identification of the data points, and there up on increasing the overall accuracy. Subsequently, the randomized data is given as training input to the RLS classifier. X_{tr} is here the randomized training data and Y_{tr} is the training label. The weight matrix, W is evaluated from the training data X_{tr}. The predicted labels, Y_{pr} are realized by multiplying randomized testing data X_{te} with W, the weight matrix. In this paper randomly 50% features have been used for training and the rest 50% for testing. Hence, the performance analysis of the classifier is carried out by observing the accuracy obtained by using different types of kernels from Grand Unified Regularized Least Squares (GURLS) toolbox and then evaluating by cross validation. GURLS is a software for efficient supervised learning which exploits all the favorable properties of RLS (Tacchetti, 2013). At last, a comparative study is made between the proposed Regularized Least Squares classifier and the conventional SVM classifi

a) Dataset Description: The UC Merced Land Use Dataset consists of aerial images obtained from the United States Geological Survey (USGS) National Map Urban Area Imagery collection for various urban areas in the United States (Yang, 2010).The dataset comprises of manually classified images of 21 various land use and land cover classes namely: agricultural, airplane, baseball diamond, beach, chaparral, buildings, dense residential, forest, golf course, mobile home park, harbor, intersection, tennis courts, medium density residential, parking lot, overpass, river, runway, freeway, storage tanks, and sparse residential; selected from aerial orthoimagery with a image resolution of one foot per pixel. Each class incorporates 100 RGB images each of 256 256 pixels. In order to exploit a variety of statistical properties, some analogous with respect to color, some analogous with respect to texture, and the rest not at all analogous, thus providing a rich set of data for the investigation. The training and testing has been performed on five classes. Samples from each class have been extracted from the main dataset are presented in Fig.3.

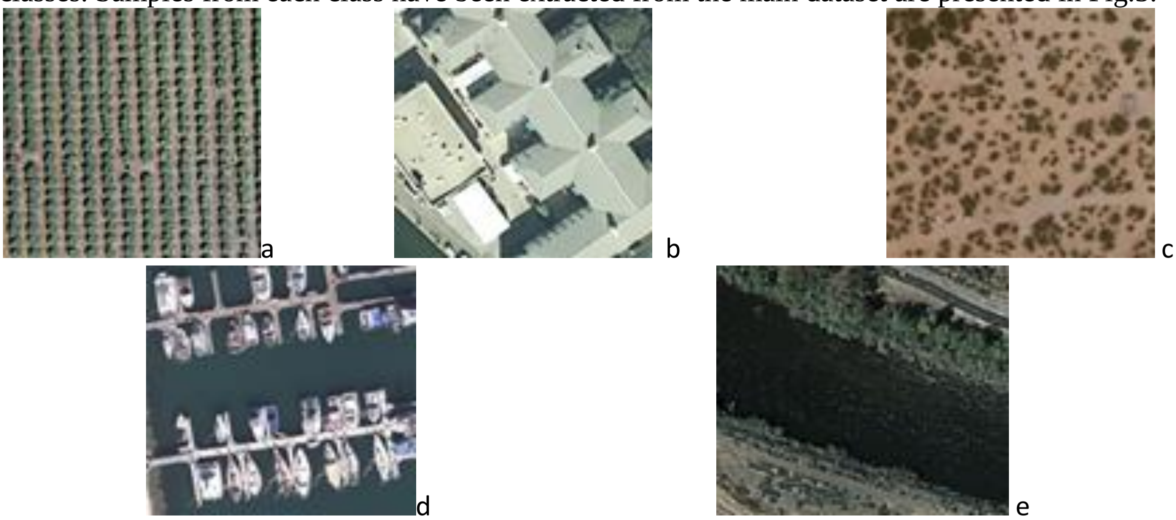

Fig.3.Samples from each class (a) agricultural, (b) buildings, (c) chaparral, (d) harbor, (e) rive

3. Results and Discussion

In this section an experiment was conducted to evaluate the performance of the proposed classifier. The system was realized in MATLAB environment and the performance was evaluated on the UC Merced Land use dataset by comparing the accuracy measures for the different classification algorithms.

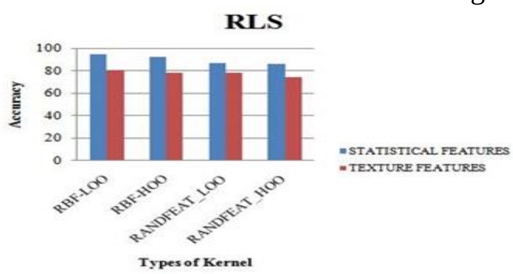

Fig.4. Accuracy measures of RLS classifier with each kernel for different features

Feature selection had been a crucial part during the classification process as selection of the good features which lead to higher accuracy and elimination of the features with little or no predictive information was important. Hence statistical and textural features were selected and made to classify. The training set was generated by randomly choosing 50% data from of the feature set of each class and the rest 50% was used as testing set. Ensemble features were evaluated; it was found that different levels of accuracy were obtained while textural features, statistical features were used individually for classification. Encouraging results with better performance in terms of classification accuracy were set out when statistical features were used.

The kernel RBF-LOO gave the best accuracy with 94.4%, followed by RBF-HOO with 92.4% accuracy when ensemble statistical features were used. Table I illustrates the classification accuracies of RLS classifier when different RBF kernels and RANDFEAT kernels with Leave One out (LOO) and Hold out (HO) cross validation were used, and classification accuracies of SVM classifier when Linear and RBF kernels were used. Fig.4 shows the plot of accuracy against different kernels used in RLS classifier.

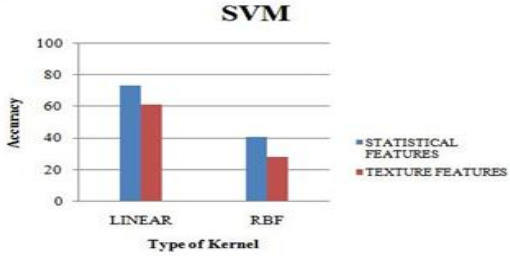

Fig.5. Accuracy measures of SVM classifier with each kernel for different features

The obtained results were compared with the SVM classifiers, the linear and non-linear kernel (RBF) were used for performance analysis. LIBSVM package (Chang, 2011) was used to implement SVM. A variety of different combinations of features were tested in order to determine the best combinations that lead to increased accuracy rate. Fig.5 shows the plot of accuracy against different kernels used in case of SVM classifier.

From the experimental results, it is clear that RLS classifier produces 94.4% accuracy when the non-linear kernels were used which is much higher than the accuracy attained by SVM. It was observed that the proposed RLS classifier produces better classification results when compared to SVM. Hence it can be stated that randomized features when incorporated with the elementary linear learning techniques compete favorably with state-of-the-art regression algorithms and kernel-based classification

Table.1. Classification Results (%) for RLS and SVM classifier

Sl.No.	**Classifier**	**RLS**				**SVM**	
	Kernel function	RBF-LOO	RBF-HO	RANDFEAT-LOO	RANDFEAT-HO	LINEAR	RBF
	Features						
1	Statistical	92.4	92.4	86.4	86	83.2	43.2
2	Textural	80.4	78	78	74	71	28

As the RKS transforms the input data to a higher dimensional feature space, randomized features are derived from the transformed data (Rahimi, Ali, and Benjamin Recht, 2007). Due to this randomization of the data in case of RLS, there exists better data separability and thereby significant reduction in the computation time needed for training, and obtain better classification results with improved accuracy.

4. CONCLUSION

In this paper, novel and powerful RKS and RLS classifiers were used for the classification of aerial image. The system was applied for the land use classification of various classes. The feature extraction provided good accuracy when ensemble statistical features were used. The generality of the proposed system was demonstrated by comparing with the conventional SVM algorithm. Due to the generation of random features in of case RLS, the data separability is increased and the computation becomes cheaper. The experimental analysis concludes that the proposed method is more efficient and produces higher accuracy for classification when compared to SVM hence shows that this approach improves the performance of supervised kernel machines. However much more work needs to be done in this field, as the scope for classifying the bulky aerial imagery is everlasting and the demand for such work is increasing day by day.

5. ACKNOWLEGMENT

We would like to thank Neethu Mohan for giving useful comments and insights, and all those who have directly and indirectly guided us for writing this paper.

REFERENCES

Ahmadi P, A simple and efficient method for segmentation and classification of aerial images, 6th International Congress on Image and Signal Processing (CISP), 01, 2013, 566 – 570.

Bock, Steffen, A region-based approach to land-use classification of remotely-sensed image data using artificial neural networks, International Archives of Photogrammetry and Remote Sensing, 31, 1996, 71-76.

Cesa-Bianchi, Nicolo, Applications of regularized least squares to pattern classification, Theoretical Computer Science, 382.3, 2007, 221-231.

Chang, Chih-Chung, and Chih-Jen Lin, LIBSVM: a library for support vector machines, ACM Transactions on Intelligent Systems and Technology (TIST), 2.3, 2011, 27.

Colin Campbell, Howlett R.J, Jain L.C, An introduction to Kernel methods in radial basis function networks, Design and applications, 2000.

Fauvel, Mathieu, Jocelyn Chanussot, and Jon Atli Benediktsson, A spatial spectral kernel-based approach for the classification of remote sensing images, Pattern Recognition, 45.1, 2012, 381-392.

Haralick, Robert M, Karthikeyan Shanmugam, and Hak Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man and Cybernetics, 6, 1973, 610-621.

Juho Lumme, Supervised classification using Aerial photographs and Lidar data, Proceedings of the 2nd Workshop of the EARSeL SIG on Land Use and Land Cover, Bonn, 2006, 28-30.

Liu, Li, and Paul W Fieguth, Texture classification from random features, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34.3, 2012, 574-586.

Maciej, Dzieszko, and Krlewicz Slawomir, Digital aerial images land cover classification based on vegetation indices, Quaestiones Geographicae, 31.3, 2012, 5-23.

Rahimi Ali, and Benjamin Recht, Weighted sums of random kitchen sinks: Replacing minimization with randomization in learning, Advances in neural information processing systems, 2009.

Rahimi, Ali, and Benjamin Recht, Random features for large-scale kernel machines, Advances in neural information processing systems, 2007.

Sheikh M.A.A, A noise-tolerant framework for aerial images classification based on Gabor energy feature, 5th International Conference on Computers and Devices for Communication (CODEC), 1-4, 2012, 17-19.

Soman K.P, Loganathan R, and Ajay V, Machine learning with SVM and other kernel methods, PHI Learning Pvt. Ltd., 2009.

Tacchetti, Andrea, GURLS: a least squares library for supervised learning, The Journal of Machine Learning Research, 14.1, 2013, 3201-3205.

Ul Qayyum Z, Akhtar A, Sarwar S, Ramzan M, Optimal Feature Extraction Technique for Crop Classification Using Aerial Imagery, International Conference on Information Science and Applications (ICISA), 2013, 1-5, 24-26.

Yang, Yi, and Shawn Newsam, Bag-of-visual-words and spatial extensions for land-use classification, Proceedings of the 18th SIGSPATIAL, International Conference on Advances in Geographic Information Systems, ACM, 2010.