URSI AP-RASC 2019, New Delhi, India, 09 - 15 March 2019

Accuracy Assessment for Microwave TerraSAR-X SAR Dataset Using Supervised Classification

S. B. Sayyad *(1), M. A. Shaikh (2), S. B. Kolhe (3) and P. W. Khirade (4)
(1) Department of Physics, Milliya Arts, Science & Management Science College, Beed 431 122 Maharashtra, India
(2) Department of Electronic Science, New Arts, Commerce & Science College, Ahmadnagar 414 001 Maharashtra, India
(3) Department of Physics, Shivaji College, Kannad Dist. Aurangabad 431 103 Maharashtra, India
(4) Department of Physics, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad 431 004 Maharashtra, India

Abstract

The microwave Synthetic Aperture Radar (SAR) is a type of active remote sensing. It has its own energy source for illumination. It receives the radiation reflected from the target on the ground surface. It generates a very high resolution imagery of the Earth or planetary bodies. It enables observation in all types of weather condition, day and night capabilities. The classification analysis has become one of the very important task, after the availability of microwave SAR datasets from satellite. The one of the major challenges faced is the accuracy regarding classification analysis. In the present paper the two supervised classification techniques used i.e., Wishart and Support Vector Machine (SVM) classifier. The accuracy results of both classifiers are analyzed on the basis errors calculation like omission and commission. The overall process is applied on microwave X-band TerraSAR-X SAR dataset of Pangkalan Bun, Indonesia. The four major classes studied is Water, Trees, Vegetation and Open Land. The present work focus on the agricultural application. From the overall work it is found that the accuracy of the Wishart supervised classifier is 71.94% and for SVM supervised classifier it is 58.36%. There is very huge difference of 13.58% between these two classifiers. Hence the Wishart classifier has better accuracy compare to SVM classifier. The overall work done by using PolSARPro Ver. 5.0 and NEST Ver. 5.0.16 software

Keywords: SAR, Wishart, SVM, Confusion Matrix, Accuracy Assessment.

1. Introduction

In remote sensing the microwave Synthetic Aperture Radar (SAR) is an active type of system, which acquired very high resolution images of the Earth or planetary bodies. It has the capacity to sense the objects during the day as well as at night, though there is change in environmental conditions, also it penetrate through clouds, smoke, fog etc. [1,2] The classification is one of the important tasks after the availability of microwave SAR datasets from satellite [3]. Since, last few decades many researcher work on accuracy assessment of microwave SAR image using classification techniques. The supervised algorithm based on complex Wishart distribution developed by (Lee et. al., 1994) [4]. The L band ALOS PALSAR L-1.1 fully polarimetric dataset for land cover classification has been used by (Mishra et. al., 2011) [5]. They analyased and compared supervised classification results using five classes of land like minimum distance, maximum likelihood, parallelepiped, based on Pauli decomposition and Wishart classification based on Eigen value decomposition by using ENVI and PolSARpro software [5]. The overall study suggest that the most of class having classification accuracy less than 85%. In the present paper work two supervised classification technique used i.e., Wishart and Support Vector Machine (SVM). The microwave SAR dataset selected is the X-band TerraSAR-X satellite SAR dataset. The objective of these works is to classify microwave SAR image using the above said supervised classification techniques and accuracy is estimated on the basis of error calculation from the classified image. This paper will provide comparative simulation model results of both supervised classified SAR images using PolSARPro Ver. 5.0 and NEST Ver. 5.0.16 software. The both software's are freely available on the internet developed by European Space Agency (ESA).

2. Classification Techniques

The classification is a technique used to identify the different objects in the remotely sensed imagery. It helps to understand image information in more depth. The class is nothing but the group of pixels, or digital numbers (DNs) having same spectral properties. Hence, in any multispectral image the pixel having its characteristics is expressed in the form of vector, i.e., spectral properties of the pixel. These numbers of vector determine the specific class in the objects. The number of classes can be made by using certain indices. The process involves labelling of each class entity using DNs. The spectral pattern recognition is one of the key points in the classification that can be utilized pixel by pixel spectral information. Hence, with the help of spatial and temporal pattern recognition the features of the land cover are identified [6,7]. The classification technique is based upon polarimetric decomposition parameters such as Entropy (H), Anisotropy (A) and Alpha (α). The classification procedure is

carried out using decomposition theorem and the H/A/α set of the coherency matrix defined by Cloude et. al. 1996 [8]. In classification assessment the coherency matrix is calculated on the basis of eigenvalue and eigenvector *[T]*. The eigenvalue of *[T]* have direct physical significance in terms of the scattered component's power into a set of orthogonal unitary scattering mechanism. It can be given by the eigenvectors of *[T]*, where the radar backscatter themselves form the column of 3x3 matrix. Hence, the arbitrary coherency matrix is written as,

$$\langle [T] \rangle = [U_3][\sum][U_3]^{-1} \tag{1}$$

$$= \sum_{i=1}^{i=3} \lambda_i u_i u_i^{*T} \tag{2}$$

where $_{[\sum]}$is a 3x3 diagonal matrix with nonnegative real elements and *[U3]* is a unitary matrix [9].

2.1 Supervised Classification

The supervised classification technique needs prior knowledge of the study area, which is under studied. The results of computer generated unsupervised classification are also helpful in this study. In a study region the each area is known as a training site. The each training site having different spectral characteristics of DNs. In this classification, basic three steps are involved. First in the training stage identifies the training areas and develop a numerical description of the spectral attributes of each land cover type in the selected study region. In second step, each land cover class in a specific class and labeled it with specific land cover type name. While making class takes care of selecting same pixels in the group. If pixel are not able to understand then make an unknown class for that. In these ways after entire study area is categories, then it is forwarded towards the output stage. After successful selection of training area and categories into different class, at last run the whole process. The accuracy of creating class depends upon selecting the training area and identifying the same pixel from the dataset. Then in the final stage classified output image is used for further analysis [10].

2.1.1 Wishart Classifier

The Wishart H A Alpha classification is a special type of H A Alpha classification. Here the coherency matrix $_{\langle T_i \rangle}$ of a pixel *i* of a multilook image knowing the class *ωi*, the Wishart complex distribution is given by,

$$p(\langle T_i \rangle / \omega_m) = \frac{N^{-qN} \exp(-tr(N[\sum_m]^{-1} \langle T_i \rangle))}{K(N,q)\left|\sum_m\right|^N} \tag{3}$$

Since, $\sum_m = E(\langle T_i \rangle \left| \langle T_i \rangle \varepsilon \omega_m \right.)$

$$\sum_m = \frac{1}{N_m} \sum_{i=1}^{N_m} \langle T_i \rangle \tag{4}$$

where *Nm* is the pixel number of *ωm, K (N, q)* is the factor of standardization [11]. Using Wishart classification method there is significant improvement in each iteration. When the number of pixel switching classes becomes smaller than a predetermined number the iteration end. After applying Wishart method the original class boundaries in the H and the alpha plane become less distinct with considerable overlap. The advantage of using Wishart method is its effectiveness in automated classification. It provides the interpretation based on scattering mechanism of each class [12-14].

2.1.2 SVM Classifier

The Support Vector Machines (SVMs) have been introduced as one of the powerful tools for performing supervised classification. These methods can also be applied to various advanced applications like data mining, regression analysis etc. The advantage of the SVMs for image classification is their ability to classify nonlinear data. Particularly, the dataset like fully Polarimetric SAR dataset [15].

Suppose the *N* training samples are available and it can be represented by a set of pairs *{(yi,xi), i=1,.....,N}* , where *yi* are the class labels of value ± 1 and $x_i \varepsilon R^n$ are the feature vectors with *n* components. The classifier is represented by a function, such as where *α* indicates the parameters of the classifier. The SVM method consists of finding the optimum separating hyperplane such that, the samples with labels $y = \pm 1$ are located on each side of the hyperplane; and the distance of closest vectors from the hyperplane in each side of the plane must be maximized.

These closest vectors are the support vectors, and the distance between them is the optimal margin shown in figure 1, the hyperplane is defined by *w.x+b=0*, where w and b are the parameters of the hyperplane. The vectors that are not on this hyperplane lead to $w.x+b \neq 0$, allowing the classifier to be defined as *f(x;a)*=sign *(w.x+b)*. The support vectors lie on two

hyperplanes that are parallel to the optimal hyperplane with the equation $w.x+b=\pm 1$. The maximization of the margin with the equations of the two support vector hyperplanes leads to the following constrained optimization problem:

$$\min\left\{\frac{1}{2}\|w\|^2\right\} \tag{5}$$

With $y_i(w.x+b)>=1_i, i=1,.....,k$

If the training samples are not linearly separable (Figure 1 a), a regularization parameter is introduced and initialized with a large value corresponding to assigning a higher penalty for errors. Therefore, the constrained optimization problem becomes,

$$\min\left\{\frac{1}{2}\|w\|^2+C\sum_{i=1}^{k}\xi_i\right\} \tag{6}$$

With $y_i(w.x+b)>=1-\xi_i i$,

where C is the regularization parameter.

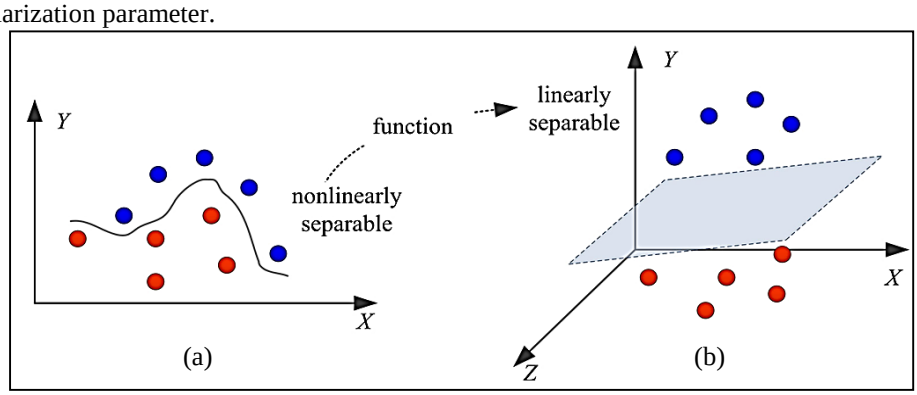

Figure 1. Support Vectors Machine, (a) nonlinearly separable case, (b) linearly separable case

The error variables (ξ_i), addressed in equation (6), are introduced to reduce the weights of misclassified vectors. This optimization problem can be solved using Lagrange multipliers:

$$\min: \sum_i \alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i,\alpha_j y_i, y_j(x_i, x_j) \tag{7}$$

$$\text{Subject to } \sum_{i=1}^{n}\alpha_i, y_j = 0 \text{ and } 0<=\alpha_i<=C\ i=1,.....,k \tag{8}$$

where the a_i represents Lagrangian multipliers, which are nonzero only for support vectors.

Thus, the hyperplane parameters (w and b) and the classifier function, $f(x; a)$, can be estimated by an optimization process. Consequently, the nonlinear classifier can be expressed as,

$$f(x) = sign(\sum_{SV} y_i\alpha_i^0(x_i.x)+b^0) \tag{9}$$

SVMs can be generalized to compute nonlinear decision surfaces in n-dimensional space. The method consists of projecting the data in a higher dimensional space, where the SVMs is considered to become linearly separable. The SVMs applied in this space lead to the determination of nonlinear surfaces in the original space [16-21].

3. Error and Accuracy Assessment

The error and accuracy of classification are determined empirically by corresponding reference ground data.

3.1 Confusion Matrix

The results are tabulated in the form of a square matrix known as confusion matrix. This matrix helps in summarizing the performance of classification and corresponding categorized pixels to each land cover type class. The results help to

summarized sample results. This is also called as error matrix. With the help of this the omission and commission error is calculated as below.

3.2 Commission Error

It is refers as to the samples of a certain class of the reference data that were not classified [22]. The error depends upon selection of pixel class from the area which can be under studied.It uses the non-diagonal row elements. It can be calculated as,

$$\left(\frac{a+b+c}{T_R} \right) \tag{10}$$

3.3 Omission Error

It is refers as to the samples of a certain class of the classified data that were wrongly classified [22]. It uses the non-diagonal column elements. The error depends upon the proper selection of certain area which can be classified. It can be calculated as,

$$\left(\frac{d+e+f}{T_C} \right) \tag{11}$$

It can be useful when evaluating the effectiveness of a discrete classification of remotely-sensed data [23]. This derived from a comparison of classified map pixels and actual land cover map pixel. It is organized as a two dimensional matrix. In this matrix columns representing the reference data by category and rows representing the classification by category.

Table 1. Confusion matrix for four class

	Reference Data					
Classified Data	C1	C2	C3	C4	Row Total	**Commission Error (%)**
C1	w	a	b	c	T_R	
C2	d	x				
C3	e		y			
C4	f			z		
Column Total	T_C					
Omission Error (%)						

Table 1 shows four classes are labelled as C1, C2, C3, C4. Matrix column shows reference data class. Matrix row shows classified data class. The diagonal values w, x, y, z shows the proportion of correctly classified pixels. The row total (T_R) can be calculated by adding the values of non-diagonal elements a, b and c except w in the first row. Similarly the column total (T_C) calculated by adding the values of non-diagonal elements d, e and f except w. The same procedure obeyed for all rows and columns. Using this value T_C and T_R omission and commission error is calculated [22-24].

3.4 Accuracy Assessment

Accuracy assessment have two types of calculations i.e., user's accuracy and producers accuracy. The user's accuracy provides the user information about the accuracy of the land cover data. It can be calculated as the number of correctly classified samples divided by total row (T_R).

The producer's accuracy indicates the percentage samples of a correctly classified reference class. It can be calculated as dividing the number of correctly classified samples by the total column (T_C). The both accuracy depends upon the amount of omission error and commission error. The relation between accuracy and error,

$$\text{User's accuracy (\%)} = (100 - \text{commission error}) \% \tag{12}$$

$$\text{Producer's accuracy (\%)} = (100 - \text{omission error}) \% \tag{13}$$

The overall accuracy can be calculated by dividing the number of correctly classified samples positioned in diagonal of confusion matrix by the total number of reference pixels checked.

4. Result and Discussion

In the present work the microwave X band TerraSAR-X satellite SAR dataset of Pangkalan Bun, Indonesia with latitude 2^0 26' 58.44'' S to 2^0 59' 02.51'' S and longitude 111^0 42' 14.22'' E to 111^0 57' 04.67'' E used. The dataset is in SSC (Single look Slant Range Complex), i.e. level 1 format. It is obtained on 13/03/2008. The SAR sensor satellite having an incident angle of 33.7^0 with dual polarization, i.e., HH and HV [25]. The field of dataset is situated on 79 feet above the Sea level. The major area covered by Trees and Vegetation. The river present in the study area is connected to Java Sea. The study area is used for interpretation of agricultural applications. The SAR dataset outline map and the region of the study area is shown in figure 2.

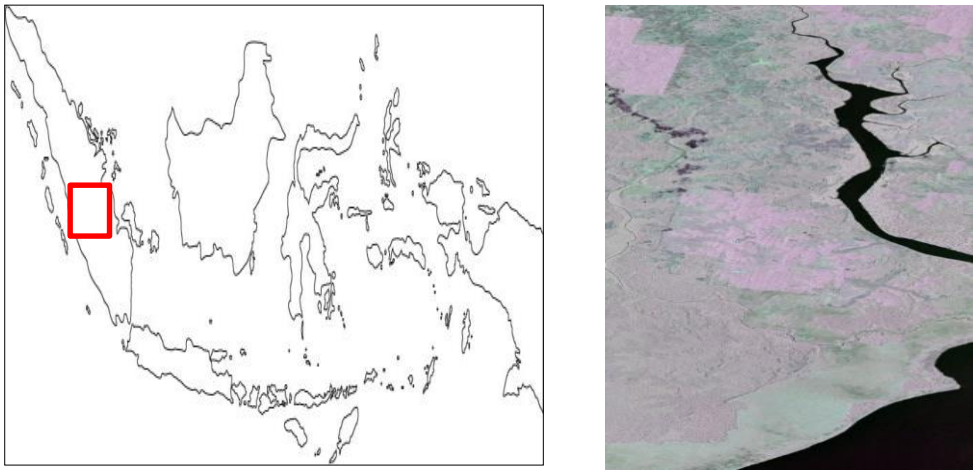

Figure 2. Location of Study Area (a) Map of Indonesia (b) Selected Area of Pangkalan Bun.

Here in the study the four classes like Water, Trees, Vegetation and Open land are made using classification techniques. In supervised classification the prior knowledge of the object present on the field area essential. This can be obtained by manual class generation with the help of the computer generated results called unsupervised classification and the scattering effect profile of the object present on the selected study region. In supervised classification two classification techniques i.e., Wishart and SVM classification is used. In both the classification 4 testing samples are selected and for every testing site 5 training samples are given. The figure 3 (a) and (b) shows supervised Wishart and SVM classifier results.

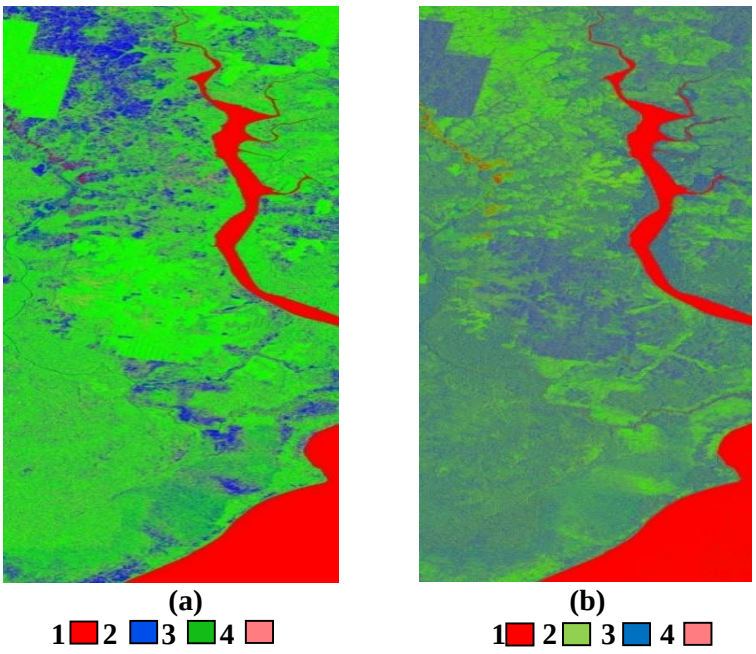

Figure 3. Supervised classification for Pangkalan Bun, Indonesia TerraSAR-X SAR image (a) Wishart (e) SVM Classifier

The coherency matrix given by equation (1) generates the confusion matrix. This matrix helps to calculate the accuracy of correct class made by users. It shows the accuracy of four classes and the class population. The class population generated from the number of pixels, or DN's contain in each class. It obtained from the supervised classified results and the number of training sample selections. The table 2 shows the class population for both the Wishart and SVM supervised classification. The confusion matrix for Wishart and SVM classification is shown table 3 and 4 respectively, where the row indicates user defined classes and the column represents the producer's classes.

Table 2. Class Population for Wishart and SVM Supervised Classification

Class	Name of Class	Class Population Wishart	Class Population SVM
C1	Water	29275	39303
C2	Trees	29003	24302
C3	Vegetation	16585	19406
C4	Open Land	16557	1486

Table 3. Confusion Matrix for Wishart Supervised Classification

Class	C1	C2	C3	C4
C1	100.00	00.00	00.00	00.00
C2	000.11	76.44	09.79	13.65
C3	000.02	10.33	72.65	17.00
C4	000.04	17.39	43.90	38.67

Table 4. Confusion Matrix for SVM Classification

Class	C1	C2	C3	C4
C1	99.58	00.42	00.00	00.00
C2	02.73	55.22	17.97	24.08
C3	00.41	31.28	40.15	28.16
C4	01.35	40.11	20.05	38.49

The accuracy of each class is calculated with the help of error finding in each class. The two error calculation is necessary, i.e. omission error and commission error.

The omission error and commission error is calculated using equation (10) and (11). In the omission error calculation, the non-diagonal column elements usedand in the commission error calculation, the non-diagonal row elements used. The omission error indicates the producer's accuracy and commission error indicates the user's accuracy. The table 5 and table 6 shows the error and accuracy assessment for Wishart supervised classification and SVM supervised classification respectively. Using this the accuracy for each class and overall accuracy is calculated.

Table 5. Error and Accuracy Assessment for Wishart Supervised Classification

Class	Total Population	Omission Error (%)	Total Population	Commission Error (%)	Accuracy (%)
C1	00000.00/29275	0.00	41.8431/29316.84	0.0014	100.0
C2	06830.20/29003	0.23	4592.49/26762.38	0.1700	76.44
C3	04536.00/16585	0.27	10107.91/22156.91	0.4500	72.65
C4	10154.40/16557	0.61	6778.36/13180.96	0.5100	38.67
Overall Accuracy					**71.94**

Table 6. Error and Accuracy Assessment for SVM Supervised Classification

Class	Total Population	Omission Error (%)	Total Population	Commission Error (%)	Accuracy (%)
C1	00165.07/39303	0.0041	0000763.06/39901	0.019	99.58
C2	10882.43/24302	0.4400	06831.30/20250.86	0.330	55.22
C3	11614.49/19406	0.5900	046650.01/12456.52	0.370	40.15
C4	00914.03/1486	0.6100	113166.50/11888.61	0.950	38.49
Overall Accuracy					**58.36**

The comparison of error and accuracy between Wishart and SVM classification is shown by graphical representation. The figure 4 shows omission error, figure 5 shows commission error and the figure 6 shows accuracy assessment for both the classification.

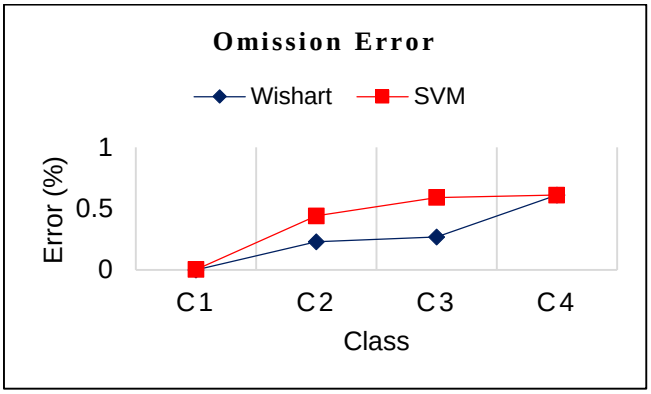

Figure 4. Graph of Omission Error

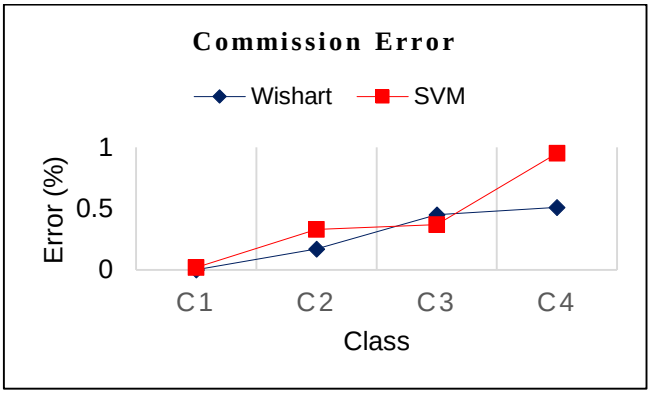

Figure 5. Graph of Commission Error

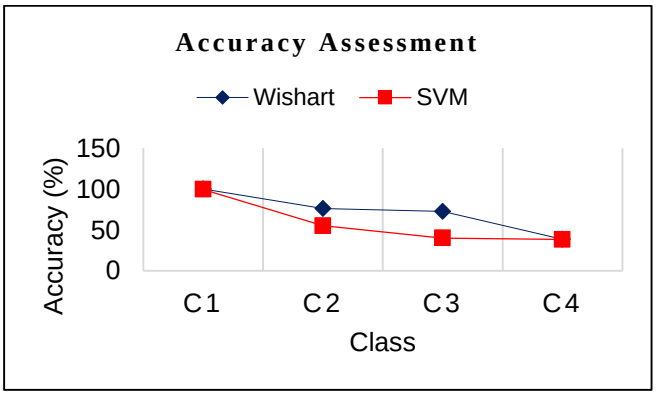

Figure 6. Graph of Accuracy Assessment

From the simulation result, it is found that the overall accuracy of the Wishart supervised classifier is 71.94% and for SVM supervised classifier it is 58.36%. There is very huge difference of 13.58% between these two classifiers. Hence, from the overall study it is found that the results of both Wishart and SVM supervised classifier using microwave X band SAR dataset for agricultural applications is not that much accurate compare to microwave L and C band satellite SAR datasets.

5. Conclusion

The present paper work microwave X band TerraSAR-X dataset is successfully used for agricultural application. The accuracy assessment based on error calculation is found to be the realistic method for microwave SAR data analysis. In the two types of

supervised classifier, Wishart and SVM results the accuracy for classes like trees and vegetation is found to be poor. From the literature, it was found that the high frequency X band SAR datasets having a lower wavelength, which is not capable of retrieving information from tree and vegetation variation. Hence, from the overall study it is concluded that the analysis of X band TerraSAR-X satellite SAR dataset is not better for agricultural application using supervised classification.

References

1. R. J. Jensen. "Remote Sensing of the Environment an Earth Resource Perspective," (2nd Ed.) Pearson, 2014, pp.12-35.
2. OPN. Calla. "Microwave Remote Sensing," Director, DESIDOC, Metcalfe House: Delhi, 2009, pp. 1-25.
3. M. A. Shaikh, P. W. Khirade, S. B. Sayyad. "Classification of Polarimetric SAR (PolSAR) Image Analysis Using Decomposition Techniques," *International Journal of Computer Application (IJCA) Proceedings on National Conference on Digital Image & Signal Processing (NCDISP 2016),* 1, 2016a, pp. 20-23.
4. J. S. Lee, M. R. Grunes, R. Kwok. "Classification of Multi-Look Polarimetric SAR Imagery Based on the Complex Wishart Distribution," *International Journal of Remote Sensing,* 15(11), 1994, pp. 2299-2311.
5. P. Mishra and D. Singh. "Land Cover Classification of PALSAR Images by Knowledge Based Decision Tree Classifier and Supervised Classifiers Based on SAR Observables," *Progress in Electromagnetics Research*, B 30, 2011, pp. 47-70.
6. M. A. Shaikh, P. W. Khirade, S. B. Sayyad. "Unsupervised and Supervised Classification of PolSAR Image Using Decomposition Techniques: An Analysis from L- band SIR-C Data," *Asian Journal of Multidisciplinary Studies (AJMS)*, 4(8), 2016, pp. 140-144.
7. V. Turkar and Y. S. Rao. "Supervised and Unsupervised Classification of PolSAR Images from SIR-C and ALOS/PALSAR using PolSARPro," *Proceeding of Conference*, 2009.
8. S. R. Cloude and P. Eric. "A review of Target Decomposition Theorems in Radar Polarimetry," *IEEE Transactions on Geoscience and Remote Sensing,* 34(2), 1996, pp. 498-518.
9. M. Ouarzeddine, B. Souissi, A. Belhadj-Aissa. "Classification of polarimetric SAR Images Based on Scattering Mechanisms," *Spatial Data Quality,* 2007.
10. S. R. Cloude and E. Pottier. "An Entropy Based Classification Scheme for Land Applications of Polarimetric SAR." *IEEE IGRS.* 35(1), 1997, pp.68-78.
11. M. Ouarzeddine, B. Souissi, "Unsupervised Classification using Wishart Classifier," *USTHB, F.E.I, BP No 32 EI Alia Bab Ezzouar, Alger.*
12. J. S. Lee, M. R. Grunes, T. L. Ainsworth, L. J. Du, D. L. Schuler, S. R. Cloude. "Unsupervised Classification Using Polarimetric Decomposition and Complex Wishart Distribution," *IEEE Transactions Geoscience and Remote Sensing*, 37/1(5), 1999, pp. 2249-2259.
13. G. Shenglong, T. Yurun, L. Yang, C. Shiqiang, H. Wen. "Unsupervised Classification Based on H/Alpha Decomposition and Wishart classifier for Compact Polarimetric SAR," *IEEE IGARSS 2015,* 2015, pp. 1614-1617.
14. J. S. Lee, M. R. Grunes, E. Pottier, L. Ferro-Famil. "Unsupervised Terrain Classification Preserving Polarimetric Scattering Characteristics," *IEEE Transactions on Geoscience and Remote Sensing*, 42(4), 2004, pp. 722-731.
15. R. S. Hosseini, I. Entezari, S. Homayouni, M. Motagh and B. Mansouri. "Classification of Polarimetric SAR Images Using Support Vector Machines," *Can. J. Remote Sensing*, 37(2), 2011, pp. 220-233.
16. C. Lardeux, P. L. Frison, J. P. Rudant, J. C. Souyris, C. Tison, B. Stol. "Use of the SVM Classification with Polarimetric SAR Data for Land Use Cartography," *Proceeding of IEEE,* 2006, pp. 497-500.
17. S. Fukuda, H. Hirosawa. "Polarimetric SAR Image Classification Using Support Vector Machines," *In IEICE Transactions on Electronics*, E84-C, 12, 2001a, pp. 1939-1945.
18. S. Fukuda, H. Hirosawa. "Support Vector Machine Classification of Land Cover: Application to Polarimetric SAR Data," *Geoscience and Remote Sensing Symposium (IGARSS), IEEE,* 2001b, pp. 187-189.
19. C. Lardeux, P. L. Frison, J. P. Rudant, J. C. Souyris, C. Tison, B. Stoll. "Use of the SVM Classification with Polarimetric SAR Data for Land Use Cartography," *In IGARSS 2006, Denver, Colorado,* 2006, pp. 497-500.
20. M. Ramakalavathi, A. James, H. Nicolas, L. Younan, M. Bruce. "Supervised Classification Using Polarimetric SAR Decomposition Parameters To Detect Anomalies On Earthen Levees," *IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 983-986.
21. Q. Zhao, J. C. Principe. "Support Vector Machines for SAR Automatic Target Recognition," *IEEE Transactions on Aerospace and Electronic Systems*, 37 (2), 2001, pp. 643-654.
22. L. F. Janssen, J. M. Frans, V. D. Wel. (1994). "Accuracy Assessment of Satellite Derived Land-Cover Data: A Review," *Photogrammetric Engineering and Remote Sensing*, 60 (4), 1994, pp. 419-426.
23. M. Story and R. G. Congalton. "Accuracy Assessment At: User's Perspective," *Photogrammetric Engineering and Remote Sensing*, 52 (3), 1986, pp. 397-399.
24. N. Kumar. Lecture Notes on Remote Sensing-Digital Image Processing Information Extraction. *IISc Banglore*, MSL2, 1.
25. TerraSAR-X Dataset. Retrieved from http://www.dlr.de/dlr/en/desktopdefault.aspx/tabid-10081/151_read-7429/year-all/#/gallery/11378.