INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, ALLAHABAD

Project Report of

Polarimetric Image Classification using Unsupervised And Supervised ML

Semester-6

Group Members:

IEC2019050	AKASH BARAD
IEC2019076	VIGNESHWARAN
IEC2019086	UDHAV RANA

PROJECT INSTRUCTOR : Dr. HIMANSHU MAURYA

Table of Contents

1. INTRODUCTION AND MOTIVATION

Polarimetry images has been highly in use for monitoring of Earth's surface. Concepts of polarimetric imaging and its diverse applications such as Disaster Preparation & Response, Search & Rescue Operations , Counter Drug Trafficking Operations, Military Surveillance, Border Security and Infrastructure Monitoring.

The theme of this polarized radar image is to deal with the science of acquiring, processing and analyzing the polarization state of radar images. Radar images are formed by radar echoes of various combinations of transmit and receive polarizations from the scattering medium. Obviously, the development of polarized radar imaging goes back to optical polarization measurement theory and optical remote control Lee / polarized radar imaging.

Sensor technology is directly applied or extended to polarized radar imaging. For example, the Stokes vector is used for description Partially polarized electromagnetic waves in terms of degree of polarization Keno's Muller matrix has been extended to study radar Backscatter from target, polarized Faraday rotation.

The plane in the magnetic field has the ionospheric effect Poincaré sphere remains a powerful graphical visualization of the polarized state.

Advances in polarized radar imaging for the last two remote sensing
Decades have stimulated several universities to establish research and education program. Two of them deserve special mention here. With the university
Germany trained DLR PhD students in the Microwave and Radar Institute
Since the late 1980s, he has specialized in SAR, radar polarization measurement, and interference measurement.

2. OBJECTIVES

Comparison Analysis Of The Existing Decomposition Techniques Based on Machine Learning.

3. EXISTING DECOMPOSITION METHODS AND LITERATURE SURVEY

- Coherent Decomposition technique:

  - Pauli's decomposition
  - Lexicographic decomposition

- Incoherent Decomposition Technique:

  - Eigenvector-based decomposition
  - Model-based decomposition

Radar Image Formation and signal Interaction:

Radar can measure amplitude and phase. Radar can only measure the part of the echo reflected back towards the antenna (backscatter).

Parameter :

1) Wavelength
2) Incident Angle
3) Polarization.

POLSAR Data Representation: Covariance(lexicographic)

Deterministic scattering: $$[S]=\begin{bmatrix} S_{HH} & S_{HV} \\ S_{VV} & S_{VV} \end{bmatrix} \quad \Rightarrow K_L = V([S]) = \frac{1}{2}Tr\left([S]\Psi_L\right)$$

$$\{\Psi_L\} = \left\{2\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} 2\sqrt{2}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} 2\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right\}$$

$$K_L = \left[S_{HH} \quad \sqrt{2}S_{HV} \quad S_{VV}\right]^T$$

Statistical Description: $$\langle [C] \rangle = \left[\frac{1}{N}\sum_{i=1}^{N} K_{L_i}.K_{L_i}^{*T}\right] \qquad \in \mathbb{R}^3 \times \mathbb{C}^3$$

Diagonal Elements: $$C_{11} = \left\langle |S_{HH}|^2 \right\rangle \quad C_{22} = 2\left\langle |S_{HV}|^2 \right\rangle \quad C_{33} = \left\langle |S_{VV}|^2 \right\rangle$$

Off-Diagonal Elements: $$C_{12} = \sqrt{2}\left\langle \left(S_{HH}S_{HV}^*\right)\right\rangle \quad C_{13} = \left\langle \left(S_{HH}S_{VV}^*\right)\right\rangle$$
$$C_{23} = \sqrt{2}\left\langle \left(S_{HV}S_{VV}^*\right)\right\rangle$$

Eigenvector-based decomposition:

$$[T] = [U][\Sigma][U]^{*T}$$

$$[U] = \begin{bmatrix} \cos\alpha_1 & \cos\alpha_2 & \cos\alpha_3 \\ \sin\alpha_1\cos\beta_1 e^{\delta_1} & \sin\alpha_2\cos\beta_2 e^{\delta_2} & \sin\alpha_3\cos\beta_3 e^{\delta_3} \\ \sin\alpha_1\sin\beta_1 e^{\gamma_1} & \sin\alpha_2\sin\beta_2 e^{\gamma_2} & \sin\alpha_3\sin\beta_3 e^{\gamma_3} \end{bmatrix}; \qquad [\Sigma] = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

$$[T] = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{12}^* & T_{22} & T_{23} \\ T_{13}^* & T_{23}^* & T_{33} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}|S_{HH}+S_{VV}|^2 & \frac{1}{2}(S_{HH}+S_{VV})(S_{HH}-S_{VV})^* & (S_{HH}+S_{VV})(S_{HV})^* \\ \frac{1}{2}(S_{HH}+S_{VV})^*(S_{HH}-S_{VV}) & \frac{1}{2}|S_{HH}-S_{VV}|^2 & (S_{HH}-S_{VV})(S_{HV})^* \\ (S_{HH}+S_{VV})^*(S_{HV}) & (S_{HH}-S_{VV})^*(S_{HV}) & 2|S_{HV}|^2 \end{bmatrix}$$

POLSAR Data Representation: Coherency(Pauli)

Deterministic scattering: $[S]=\begin{bmatrix} S_{HH} & S_{HV} \\ S_{VV} & S_{VV} \end{bmatrix} \Rightarrow Kp = V([S]) = \frac{1}{2}Tr([S]\Psi p)$

$$\{\Psi p\} = \left\{ \sqrt{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \sqrt{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \sqrt{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\}$$

$$Kp = 1/\sqrt{2}\left[S_{HH}+S_{VV} \quad S_{HH}-S_{VV} \quad 2S_{HV}\right]^T$$

Statistical Description: $\langle [T] \rangle = \frac{1}{2}\left[\frac{1}{N}\sum_{i=1}^{N} K_{P_i}.K_{P_i}^{*T}\right] \quad \in \mathbb{R}^3 \times \mathbb{C}^3$

Diagonal Elements: $T_{11} = \left\langle |S_{HH}+S_{VV}|^2 \right\rangle \quad T_{22} = \left\langle |S_{HH}-S_{VV}|^2 \right\rangle \quad T_{33} = \left\langle 2|S_{HV}|^2 \right\rangle$

Off-Diagonal Elements: $T_{12} = \langle (S_{HH}+S_{VV})(S_{HH}-S_{VV})^* \rangle \quad T_{13} = 2\langle (S_{HH}+S_{VV})(S_{HV})^* \rangle$

$T_{23} = 2\langle (S_{HH}-S_{VV})(S_{HV})^* \rangle$

4. UNSUPERVISED METHODOLOGY

Unsupervised learning is a machine learning technique that does not monitor the model using training datasets. Instead, the model itself finds hidden patterns and insights from the given data. It can be compared to the learning that takes place in the human brain when learning new things.

- What are clustering algorithms?

Clustering is an unsupervised machine learning task. Depending on how this method works, this is sometimes called cluster analysis.

- Types of clustering algorithms:

There are different types of clustering algorithms that handle all kinds of unique data.

1. Density based:

In density-based clustering, data is grouped by areas of high concentrations of data points surrounded by areas of low concentrations of data points. Basically the algorithm finds the places that are dense with data points and calls those clusters.

2. Distribution-based:

With a distribution-based clustering approach, all of the data points are considered parts of a cluster based on the probability that they belong to a given cluster.

3. Centroid-based:

Centroid-based clustering is the one you probably hear about the most. It's a little sensitive to the initial parameters you give it, but it's fast and efficient.

4. Hierarchical-based:

Hierarchical-based clustering is typically used on hierarchical data, like you would get from a company database or taxonomies. It builds a tree of clusters so everything is organized from the top-down.

5 K-MEANS MODEL

K-Means clustering is an unsupervised learning algorithm used to solve clustering problems in machine learning or data science. In this topic, you will learn what the k-means clustering algorithm is, how the algorithm works, and the Python implementation of k-means clustering.

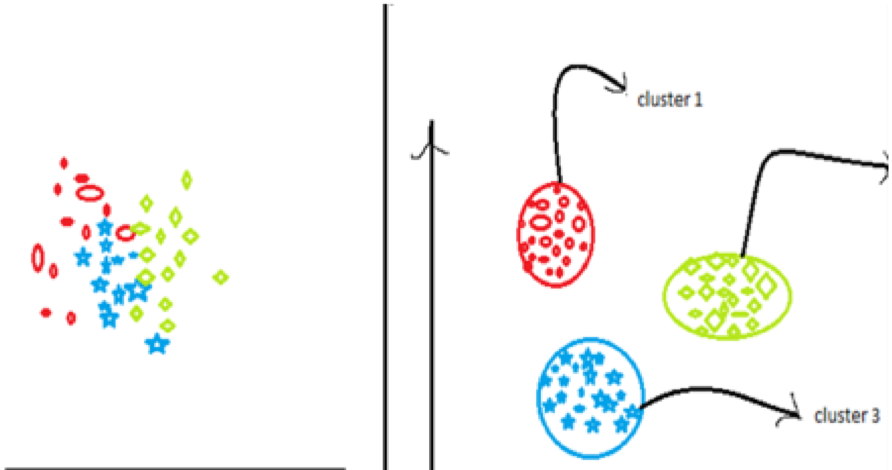


fig 1: before applying k-means clustering

fig 2: After applying K-means clustering

This allows you to group your data into different groups and find yourself the categories of groups in your unlabeled dataset without the need for training. This is a centroid-based algorithm, with each cluster associated with a centroid. The main purpose of this algorithm is to minimize the total distance between the data points and their corresponding clusters.

- Steps for K-means Algorithm :

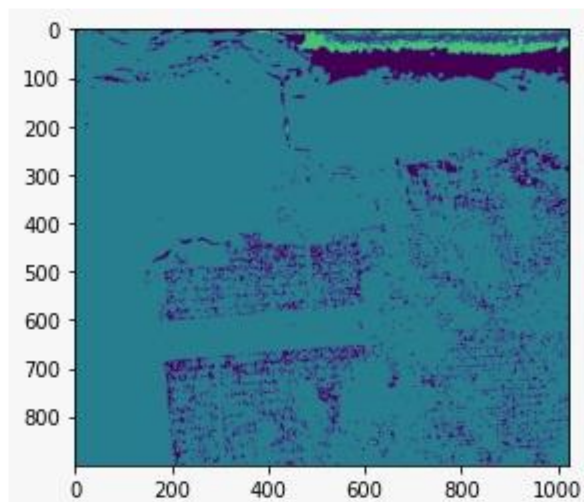
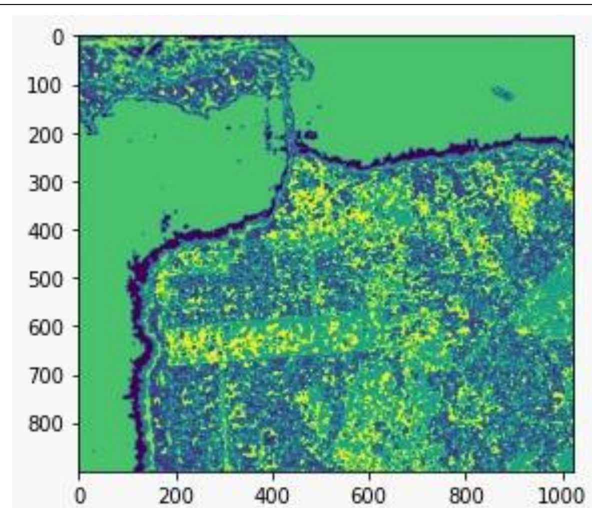
Phase1: Select the number K to decide the number of clusters.
Tree.
Phase 2: Select random K points or centroids.
Phase 4: Calculate the variance and place a new centroid of each cluster.
Phase 5: Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.

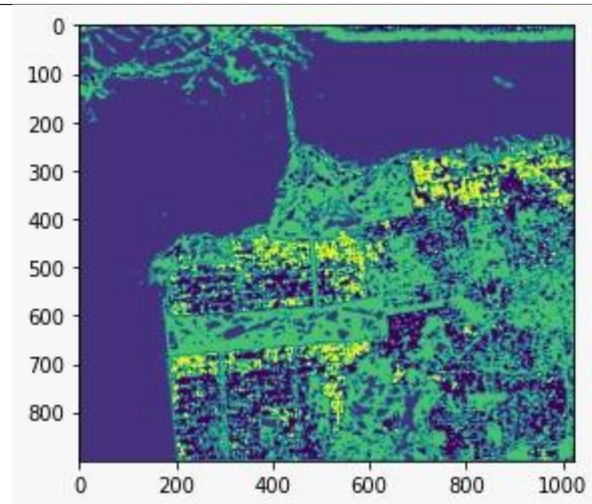
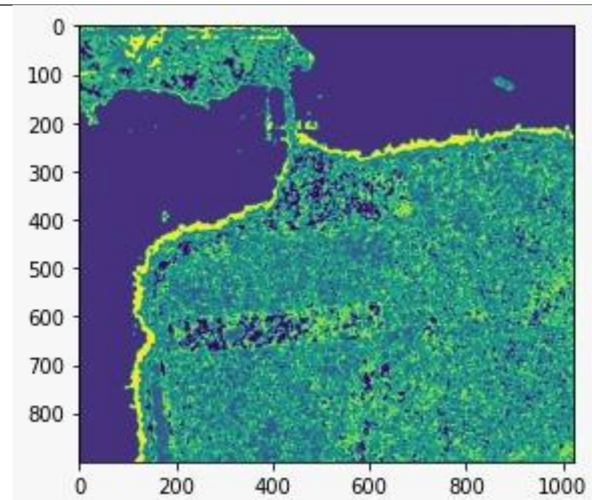
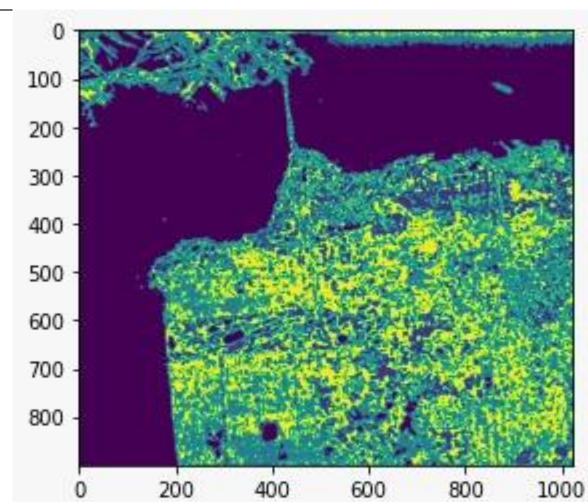
- Normalization of dataset for K-means:

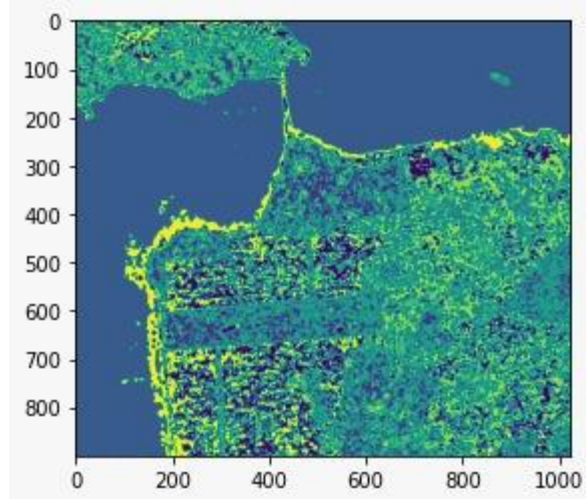
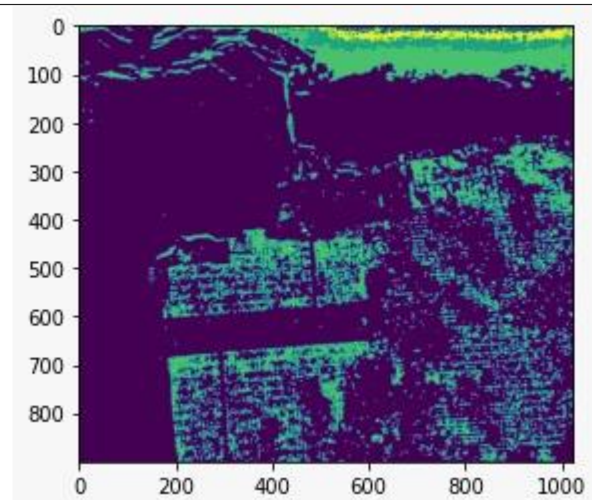
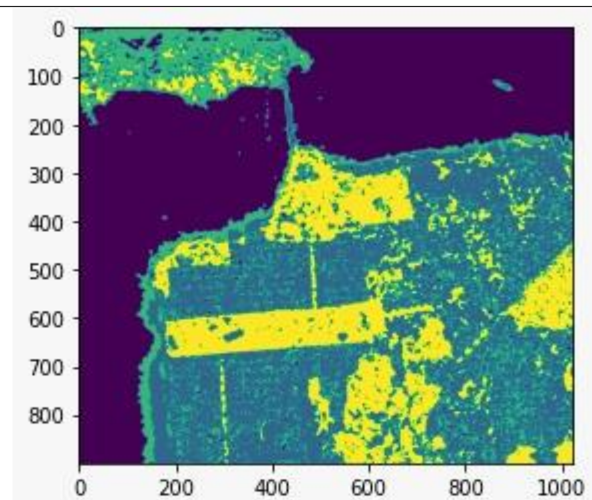
  K-means algorithm can generate better results after the modification of the databases.

  The K-means algorithm can produce effective results when Normalization is applied to the dataset. that is process of standardizing all attributes of record and give them equal weights, eliminates redundant or noisy objects there is valid and reliable data to improve its result accuracy.

6 RESULTS FOR UNSUPERVISED LEARNING

the image show the output is for feature 1 dataset for k-means algo	
the image show the output is for normalized feature 1 dataset for k-means algo.	

the image show the output is for feature 2 dataset for k-means algo.	
the image show the output is for normalized feature 2 dataset for k-means algo.	
the image show the output is for feature 3 dataset for k-means algo.	

the image show the output is for normalized feature 3 dataset for k-means algo.	
the image show the output is for combined feature123 dataset for k-means algo.	
the image show the output is for combined and normalized feature123 dataset for k-means algo.	

7 SUPERVISED METHODOLOGY

Supervised learning is a subcategory of machine learning and artificial intelligence. It is defined by training algorithms that use labeled datasets to classify data and accurately predict results. As input data is entered into the model, its weights are adjusted until the model fits properly. This is done as part of the cross-validation process.

Supervised learning helps organizations solve a variety of real-world problems on a large scale, including: B. Place spam in a different folder than your inbox normalization is applied to the dataset. that is process of standardizing all attributes of record and give them equal weights, eliminates redundant or noisy objects there is valid and reliable data to improve it result accuracy. Supervised learning uses a training set to train the model to achieve the desired output. This training dataset contains inputs and correct outputs that allow the model to train over time. The algorithm measures its accuracy via a loss function and adjusts until the error is sufficiently minimized.

- Types of Supervised learning:
  - Classification:

    Classification uses an algorithm to accurately assign test data into specific categories. It recognizes specific entities within the dataset and attempts to draw some conclusions on how those entities should be labeled or defined

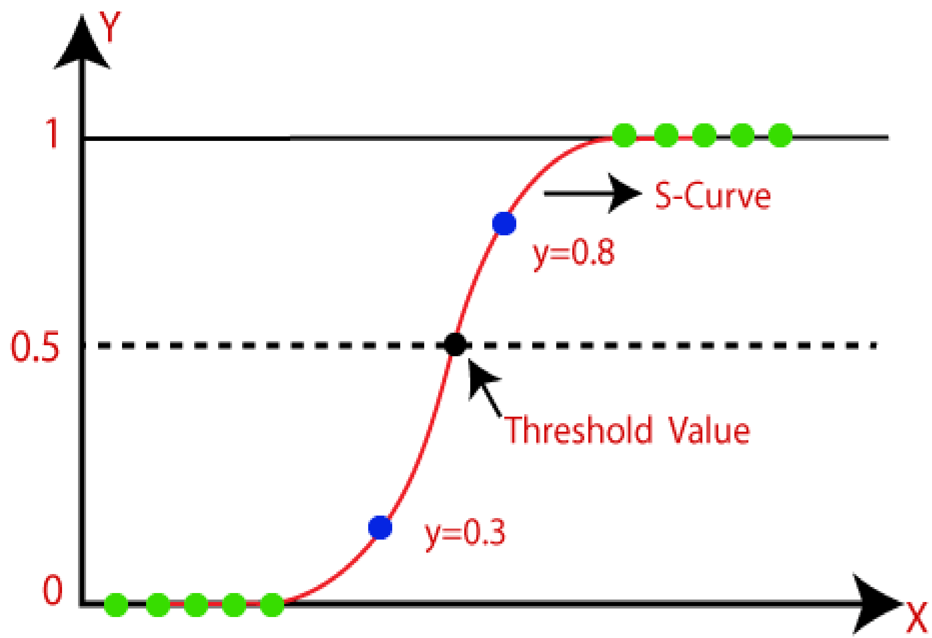
  - Regression:

    Regression is used to understand the relationship between dependent and independent variables. Often used to make predictions. B. Sales of a particular company. Linear regression, logistic regression, and polynomial regression are common regression algorithm.

8. LOGISTIC REGRESSION:

Logistic regression is one of the most common machine learning algorithm for supervised learning techniques. It is used to predict categorical dependent variables using a specific set of independent variables.

Logistic regression is a classification algorithm used to assign observations to individual sets of classes. Examples of classification issues are email spam or non-spam, online transaction fraud or non-fraud, malignant or benign tumors. Logistic regression uses a logistic sigmoid function to transform its output and return a probability value.

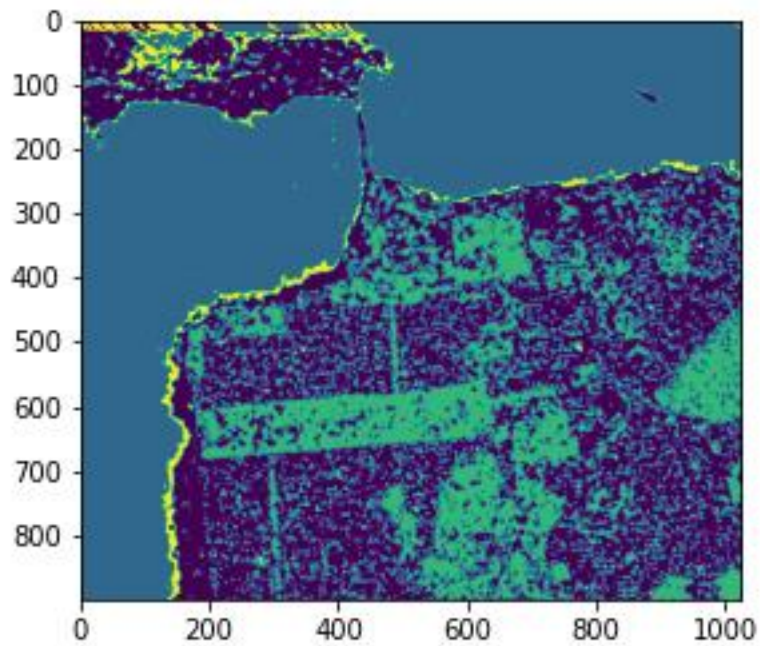


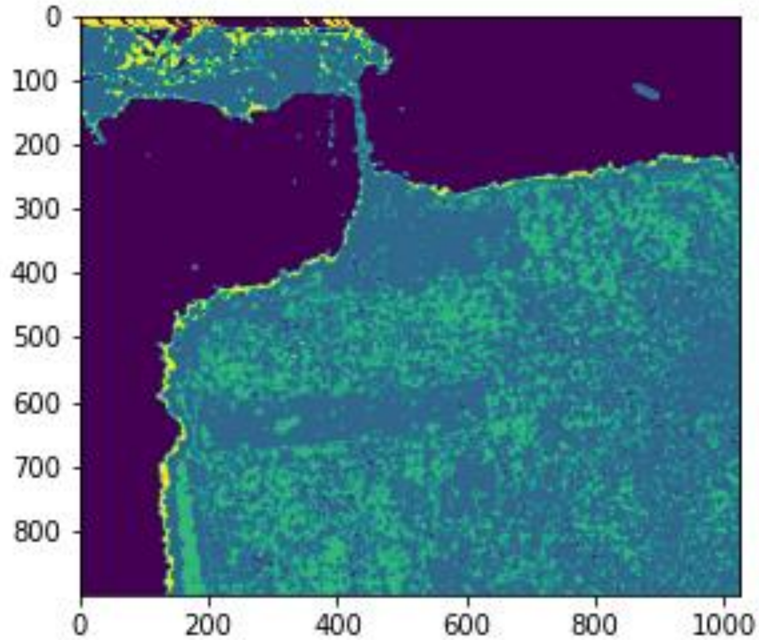
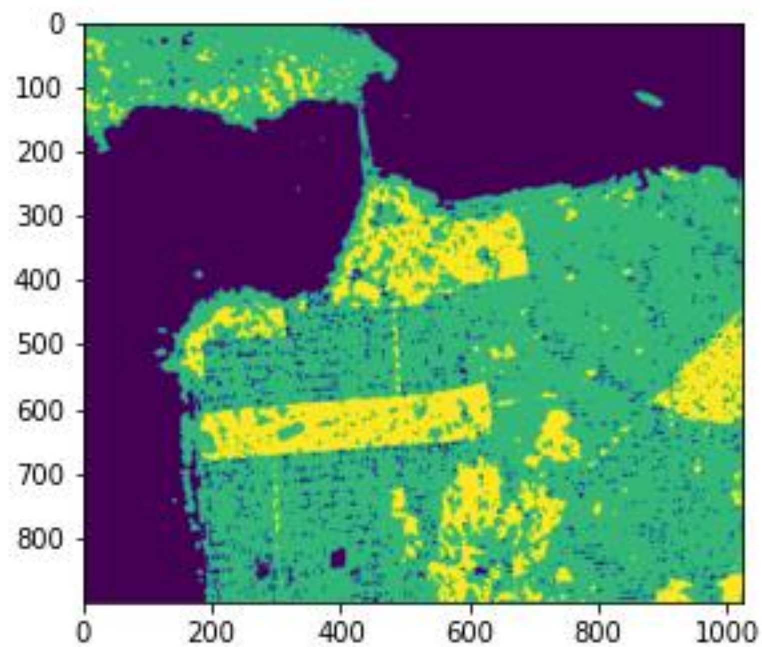
- Algorithm for Logistic Regression:

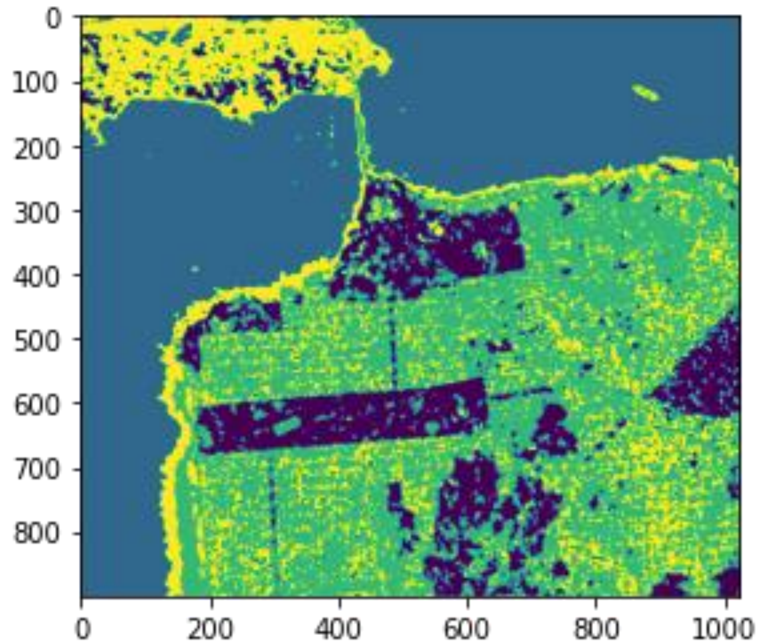
Phase 1: Train the model using the train data and train output.

Phase 2: Using the trained model calculate the Prediction for the remaining test data. (note: here train data is a matrix of size (100*128) and the trained output is the labels obtained using k-means of size (100*128).)

9. RESULTS FOR SUPERVISED LEARNING

the image show is for normalized feature 1 dataset for supervised algorithm.	

the above image show is for normalized feature 2 dataset for supervised algorithm.	
the above image show is for normalized feature 3 dataset for supervised algorithm.	

the above image show is for combined and normalized feature123 dataset for supervised algorithm.	

10. CONCLUSIONS

- From the given raw data set we have improved the results by performing normalisation and then further improved using supervised machine learning.

- We have successfully classified the polarimetric image into land , water and vegetation using unsupervised machine learning with a accuracy of 91.125 %

- After applying supervised machine learning to the same data set we get an accuracy of 82.02 %. We can also improve the supervised machine learning results if the features are uniformly distributed in the data set.}

11. FUTURE SCOPE

- The model we have created has a low accuracy for volume scattering data.

- By exploring different machine learning algorithms we can further improve the accuracy depending on the type and size of the data set.

- Similar to eigenvalue decomposition we can also explore model based decomposition like freeman and Yamaguchi.

12. REFERENCES

1) An Entropy Based Classification Scheme for Land Applications of Polarimetric SAR Shane Robert Cloude , Senior Member, IEEE and Eric Pottier, Member, IEEE}

2) Review of Target Decomposition theorems in radar Polarimetry Sham Robert Cloude, Member, IEEE, and Eric Pottier, Member, IEEE }

3) Polarimetric Radar Imaging Founding Editor Brian J. Thompson University of Rochester Rochester, New York.

4) Freeman, A. and S. L. Durden, “A three-component scattering model for polarimetric SAR data," IEEE Transactions on Geoscience and Remote Sensing, Vol. 36, No. 3, 963-973, 1998.