Technium Vol. 4, No.10 pp.1-10 (2022)
ISSN: 2668-778X
www.techniumscience.com

Increasing The Taxonomic Accuracy of Remote Sensing Data Using Traditional Pattern Recognition Methods

Alhan Anwer Younis Alsafar

Education College for Girls, University of Mosul, Mosul, Iraq
alhan.alsafar@uomosul.edu.iq

Abstract. The most significant outcomes of remote sensing are maps of Land Use and Land Cover (LULC), which can be controlled by a procedure known as image classification. Using image processing methods, we are developing a system in this effort to categorize satellite photos and extract information. Satellite pictures have been classified into usable and unused areas, as well as sub-classifying each class into numerous further classes. Used satellite photos are further divided into residential, commercial, highway, and agricultural areas, while unused images are divided into forest, river, desert, and beach areas. Since K-means classifiers analyze images features and Otsu's approach for multilevel image thresholding is efficient for automatic classification of satellite images, that is the main topic of this research.

Keywords. Remote Sensing, unsupervised Classification algorithm, k-Means Classifier, Otsu's Threshoding.

1. Introduction

Many applications, including image classification, object detection in industrial production, medical image analysis, action recognition, and remote sensing, use deep learning and computer vision [1]. There are several uses for satellite image analysis, which is regarded as the primary method of acquiring geographic information [2]. Design, construction, urban planning, and water resource management are all examples of civil engineering. There is a need for effective techniques for data extraction because the data obtained from satellite sources is enormous and is growing exponentially[3]. These Numerous Satellite Images Can Be Arranged In Semantic Orders Through Image Classification. The process of classifying satellite images is multi-layered and begins with the extraction of features from the images before categorizing them [4].When talking about the remote sensing data, where the distribution of the variety in the parameter space is affected by many conditions, including the degree of illumination during photographing the land, the roughness of the ground cover in terms of topography and its disparity within the space represented by the discriminating power of the sensor used in imaging. For example, the discriminatory capacity of the Thematic Mapper is 30 meters, while the spot sensor is 20 meters, Therefore, the probability of the purity of a single pixel in the image, which is equivalent to the discriminating power of the sensor, increases with the increase in the discriminatory power, meaning that the probability of the reflective purity of an area of 5 square meters is more than the purity of the reflective area of an area of 30 square meters. The location of the item's distribution in the parameter space.[5].

2. Literature Survey

This section offers a selection of earlier methods that dealt with the clustering of satellite images and proposes a number of methods for doing so:

Using k- means to categorize blocks of high-resolution satellite photos was suggested by researcher [2]. Authors [3] developed a segmentation and clustering technique with two phases for classifying land cover in satellite pictures. In order to categorize remote sensing photos, [4] suggested a double feature extraction highbrid deep learning strategy. A new approach is suggested by researchers [5] to automatically discover the true value of k, which depends on the central values of the clusters. This method uses the k-means clustering algorithm to determine multiple cluster counts for each value of k in order to verify the uniqueness of various cluster centroids. Researchers [6] conduct a comparison study using unsupervised classification methods on the classification features of high-resolution and polar metric SAR images. Additionally, Support Vector Machine (SVM) and (K-means) were developed and utilized by [7] to the categorization of high resolution and low-resolution satellite pictures. In order to separate the roadways and residential areas from the vegetative areas in remote sensing photos, researchers [8] studied the (Otsu's) thresholding technique. Additionally, [9] constructed algorithm that we provide a novel multilevel Rényi's entropy function and MCS algorithm for color image thresholding. This technique for segmenting remote sensing images combines Rényi's entropy model and the MCS algorithm. Authors [10] developed, two unsupervised classification algorithms—k-means clustering and fuzzy c-means—will be applied to Landsat-8 satellite pictures. A new study is suggested by researchers [11] concentrates on using UNet and Tensorflow to find water bodies. To lessen the losses, a Nadam optimizer is used. Additionally, it determines the number of nodes in each layer and the activation function's best-optimized parameters.

3. Proposed algorithm

3.1. Data Acquisition

The most proposed classification system has been applied to multispectral images captured by the TM (Thematic Mapper) sensor Aboard the satellite (Landsat5) for a number of different areas in Nineveh Governorate, including a view of the city of Mosul, another view of the Mosul Dam, and the last of Mount Sekh Ibrahim, with different bands.

3.2. Pre- processing

This is the initial stage implemented on satellite image which includes of converting the input image (satellite image) into another form (grayscale image) at first then removing noise by using wiener filter. The Wiener filter determines the variance and means corresponding to every pixel as follows[12]:

Mean:

$$\mu_{(A)} = \frac{1}{mn} \sum_i^m \sum_j^n ai,j \tag{1}$$

And variance

$$\sigma^2(A) = \frac{1}{mn} \sum_i^m \sum_j^n a_i^2,j - \mu^2(A) \tag{2}$$

Devising a Wiener filter expressing using a combination of variance and mean equations [17]:

Wi,j=

$$\mu_{(A)} + \frac{\sigma^2(A) - v^2}{\sigma^2(A)} \left(ai,j - \boldsymbol{\mu}_{(A)}\right) \tag{3}$$

where:

Wij is Wiener.

μ is the mean.

σ is the variance around each pixel.

v 2 is the noise variance.

3.3. Implement k-means clustering method

The N data objects are divided into K clusters using the k-means algorithm. To reduce the square error, it attempts to divide the data into K pieces from N bits [13]. The basic outcome of the partitioning of the k-means algorithm maximizes the similarity of the data in the set and minimizes the similarity between the sets. To identify the best k-clustering, seeding numbers are a big challenge. As to how it operates [14]:

Number of cluster (k)
Input: Set of features.
Output: Clustered image.
Steps:
1: Begin
2: Choose number of clusters k.
3: Choose the initial cluster centroids ci, i = 1... k randomly.
4: Calculate Euclidian distance

$$\boldsymbol{ED} = \sqrt{(x1 - c1) + (x2 - c2)} \dots (xn - ck) \qquad .(4)$$

6: Update new centroid by mean of each clusters point.
7: Calculate ED for next observation.
8: Assign data with minimum distance to nearest class.
9: Repeat 6 to 8 until no change in centroid.
10: Return clustered image.
11: End algorithm.

3.4. Otsu's method for image thresholding

Thresholding is the technique used to distinguish foreground pixels from background pixels. The Otsu's method, put out by Nobuyuki Otsu, is one of the various methods for obtaining optimal thresholding. The threshold value where the weighted variance between the foreground and background pixels is the least is found using Otsu's method, which is a variance-based technique[15]. The important thing is to measure the distribution of background and foreground pixels while iterating over all conceivable threshold settings. Locate the threshold at which the dispersion is the smallest. The algorithm iteratively seeks the threshold that reduces the within-class variance, which is determined by the weighted sum of the variances for the two classes (background and foreground). Grayscale typically has hues between 0-255. (0-1 in case of float). Therefore, if we select a threshold of 100, all pixels with values below 100 form the image's background, while all pixels with values at or above 100 become its foreground [16]. The following equation can be used to calculate the within-class variance at any threshold t[17]:

$$\sigma^2(t)=\omega bg(t)\ \sigma^2 bg(t)+\omega fg(t)\ \sigma^2 fg(t) \qquad (5)$$

where the probability of the number of pixels for each class at threshold t is represented by bg(t) and fg(t), respectively, and the variance of color values is represented by 2.

3.5. Algorithm

The following steps might be used to summarize the methodology employed for this study:
1. Enter the data set's images.
2. Convert the source photos to a representation of the appropriate grey level.
3.. Present the original images.
4. To get rid of extra noise and artifacts, use a [3,3] wiener filter.
5. Display the edited image.
6.. To categorize an image, use the k-means clustering technique.
5. implement the Otsu's method for image thresholding with k=5.
6. Display the categorized photo.

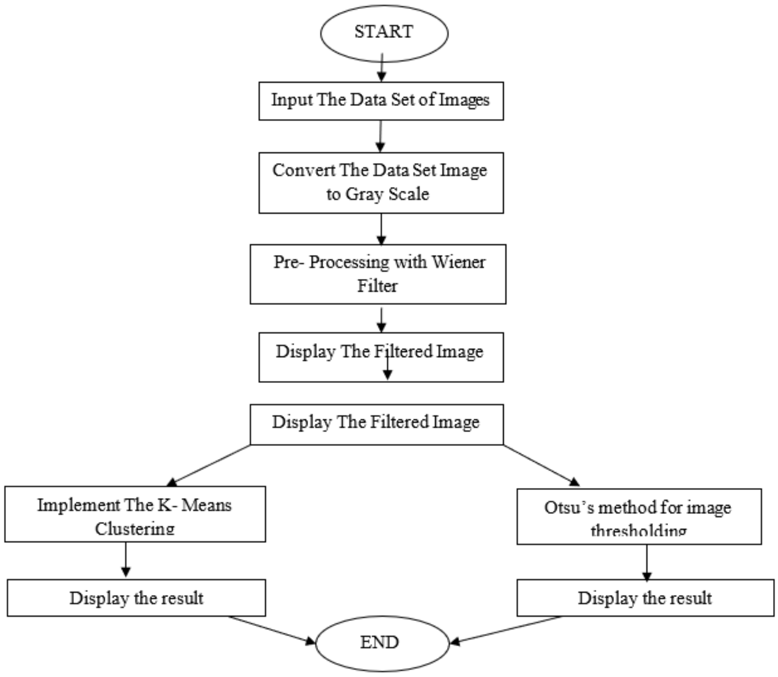

Figure 1. The flowchart of the proposed algorithm

4. Results

To implement the proposed algorithm, a software package was designed in the Matlab language and applied to classify the remote sensing data captured by the object map sensor (Thematic Mapper) stationed on the satellite (Landsat5). Three different areas were adopted for the study, namely, the center of the city of Mosul, the Sheikh Ibrahim area located southwest of the city of Mosul, and the Mosul Dam area.

4.1. The results of applying the system to Mashhad in the Jabal Sheikh Ibrahim region Within the band (0.45-0.52) micron:

This area is characterized by the presence of vegetation cover with different degrees of density and includes irrigated agricultural lands, natural plants and harvested agricultural fields. The area also contains dry lands ranging from dry and wet soils. The number of output classes was determined by nine classes. Since in all types of undirected classification, the identification of the varieties is not determined until after the classification process, that is, the process of interpreting the results comes after the (Possvier Classification) stage original study areas. Table (1) Euclidean distance between alternating pairs of classes of K-means algorithm.

4.2. The results of applying the system to a scene in the city of Mosul within (0.52-0.59) micron:

This part of Mosul city center is characterized by the presence of agricultural lands and residential areas, as well as the presence of dry lands and harvested agricultural fields, in addition to the water represented by part of the Tigris River. The number of output classes is seven. Table (4-1) Euclidean distance between alternating pairs of classes of K-means algorithm.

Table 1. A: The matrix of Euclidean distances between each alternating pair of classes output from the. k-means algorithm.
B: The minimum values associated with each of the k-means, which represent the distance between each category and its closest.

(A)

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9
Class 1	0	5470	5446	8548	17417	8226	7496	4628	4860
Class 2	5470	0	5086	4116	3411	1740	11792	3500	13786
Class 3	5446	5086	0	918	11627	4696	2578	544	4254
Class 4	8548	4116	918	0	7535	2456	4310	732	8186
Class 5	17417	3411	11627	7535	0	4279	21967	9733	27657
Class 6	8226	1740	4696	2456	4279	0	8584	2374	12972
Class 7	7496	11792	2578	4310	21967	8584	0	2596	1560
Class 8	4628	3500	544	732	9733	2374	2596	0	4850
Class 9	4860	13786	4254	8186	27657	12972	1560	4850	0

(B)

distance to the nearest class (d)	Class Number
4628	1
1740	2
544	3
732	4
3411	5
1740	6
1560	7
544	8
1560	9

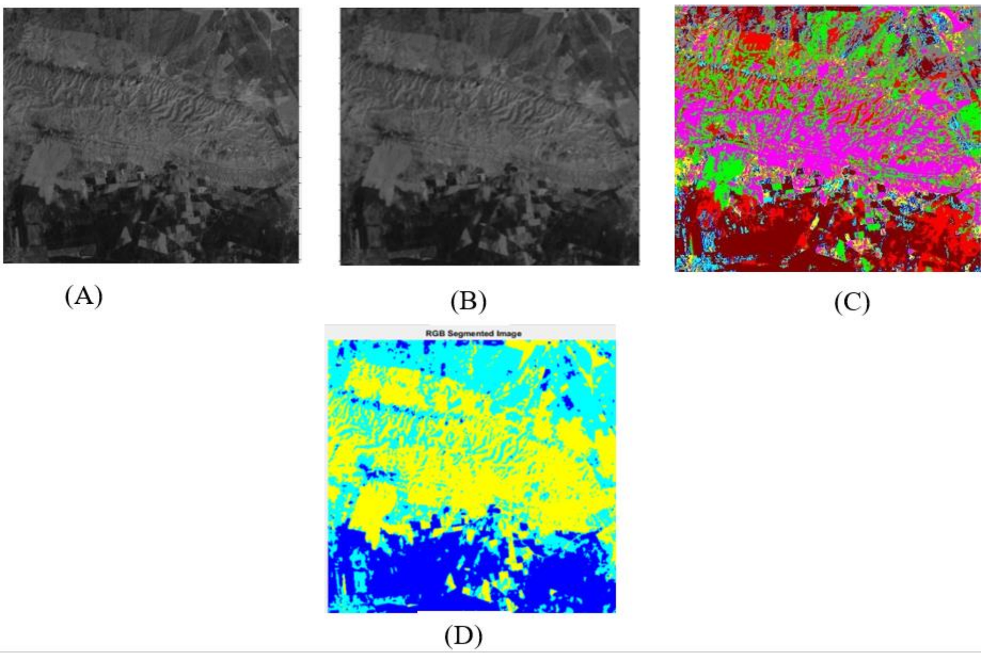

Figure 2. A: The Original Image, (B): Filtered Image with wiener Filter, (C): Classified Image with K- Means Clustering, (D): Image with Otsu's Method for Image Thresholding

Technium Vol. 4, No.10 pp.1-10 (2022)
ISSN: 2668-778X
www.techniumscience.com

Table 2. A: The matrix of Euclidean distances between each alternating pair of classes output from the k-means algorithm.
B: The minimum values associated with each of the k-means, which represent the distance between each category and its closest.

(A)

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Class 1	0	457	41	293	2889	131	533
Class 2	457	0	238	580	1110	750	1910
Class 3	41	238	0	266	2276	222	824
Class 4	293	580	266	0	2354	806	1218
Class 5	2889	1110	2276	2354	0	3674	5828
Class 6	131	750	222	806	3674	0	338
Class 7	533	1910	824	1218	5828	338	0

(B)

distance to the nearest class (d)	Class Number
41	1
238	2
41	3
266	4
1110	5
131	6
338	7

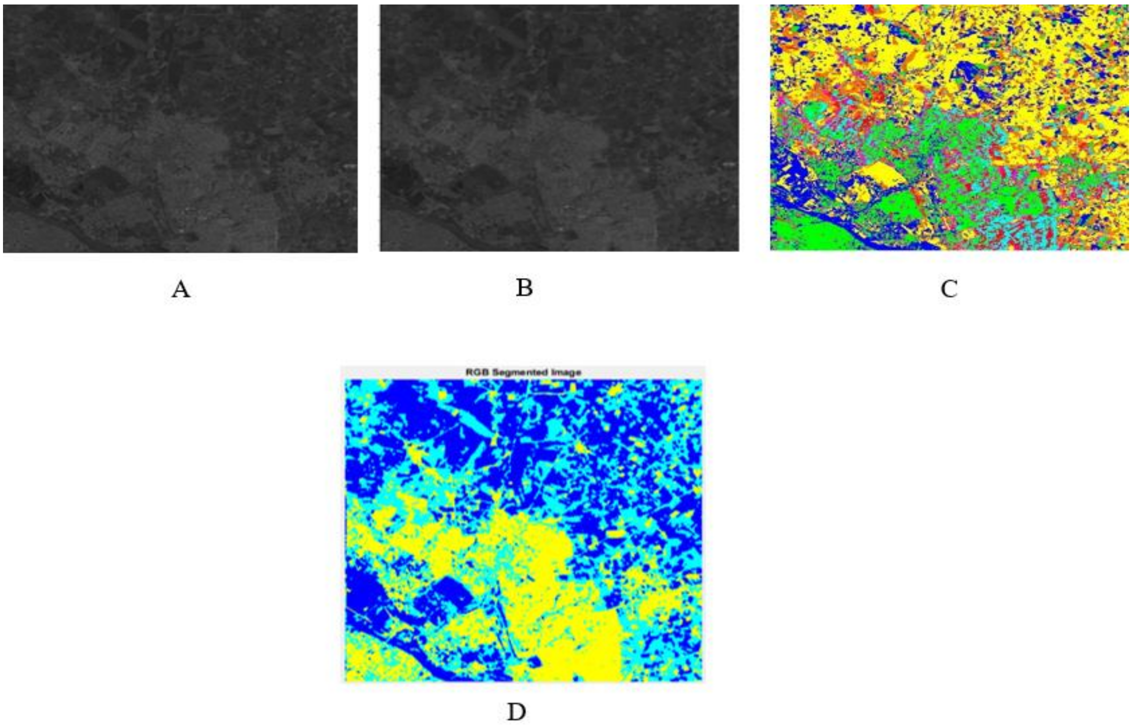

Figure 3. A: The Original Image, (B): Filtered Image with wiener Filter, (C): Classified Image with K- Means Clustering, (D): Image With Otsu's Method For Image Thresholding

4.3. The results of the application of the system to the scene of the Mosul Dam within (0.52- 0.59) micron:

This area is characterized by the presence of dam water, agricultural areas and natural vegetation, and there are some quarries in the area. The number of classes was limited to seven classes. Table (3) Euclidean distance between alternating pairs of classes of K-means algorithm.

Table 3. A: The matrix of Euclidean distances between each alternating pair of classes output from the k-means algorithm.
B: The minimum values associated with each of the k-means, which represent the distance between each category and its closest.

(A)

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Class 1	0	749	1531	1513	314	199	234
Class 2	749	0	1046	1656	475	469	425
Class 3	1531	1046	0	194	689	637	1761
Class 4	1513	1656	194	0	699	717	2157
Class 5	314	475	689	699	0	198	646
Class 6	199	469	637	717	198	0	402
Class 7	234	425	1761	2157	646	402	0

(B)

distance to the nearest class (d)	Class Number
199	1
425	2
194	3
194	4
198	5
198	6
234	7

Table 4. Demonstrates the accuracy of the two methods

Images	parameters	Otsu's method	k- means method
Image1	Accuracy	87.7	90.11
Image2	Accuracy	84.5	93.32
Image3	Accuracy	82.2	94.55

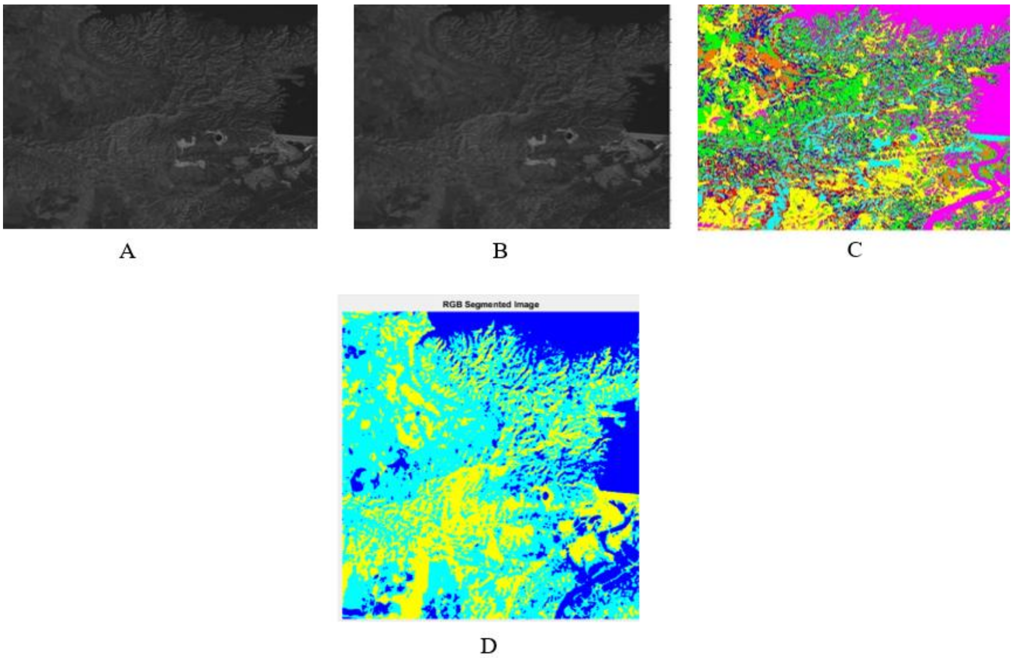

Figure 4. A: The Original Image, (B): Filtered Image with wiener Filter, (C): Classified Image with K- Means Clustering, (D): Image With Otsu's Method For Image Thresholding

5. Discussion:

Based on the analysis, it can be concluded that the Otsu method requires picture pre-processing in order to achieve high accuracy. frequently Otsu method shifted to a greater class variance that may result in missing of Image [8] contains weak objects. Good Otsu techniques results if the histogram displays a bimodal distribution and is suggested using a narrow valley to connect two tops rather than the the fact that the object is smaller than the background size When this happens, the histogram loses its bimodality [3]. In If the image is distorted because of noise, Otsu thresholding is used. Histogram became distorted, and Otsu technique terminates owing to a segmentation mistake.

One of the greatest drawbacks of the Otsu approach is that it only takes into account two classes in the histogram, but a digital image typically contains more than two classes of pixels, so under those circumstances the Otsu method is not appropriate for segmentation.

The noise brought into the image can be avoided by using the K-means method of image segmentation, which does not require histograms for the segmentation process. K-means groups together similar pixels or data sets after evaluating each pixel individually.

K-means is effective and adaptable to change since it operates on spherical clusters and is appropriate for huge data sets. The K-means approach to image segmentation is quick and economical in terms of computational cost when compared to other segmentation techniques. The uniform effect, which refers to the production of clusters that are the same size even while the input data is of varied sizes, is one of the main shortcomings of the K-means approach. It can be challenging to estimate the value of k, which must be given during the beginning stages of the algorithm, in K-means.

6. Conclusions:

1. The study proved that undirected classification algorithms, whether traditional clustering algorithms or other algorithms, each depend on a special mechanism for dividing the parameter space, and that this mechanism is determined through the process of determining the centers of the classes.

2. The proposed algorithm for classification is flexible in terms of implementation, as two or more algorithms can be used for the purpose of integration without being limited to specific algorithms. Cyclic.
3. 3.The Otsu algorithm and the k-means method are examined together with their benefits and drawbacks. Otsu algorithm is popular due to its simplicity, but it also has significant drawbacks, which is why several refinements have been produced No single image segmentation technique is adequate for all types of digital photos, and certain techniques are not appropriate for certain image types. Hazardous photos are incompatible with the Otsu technique. Similar to this, the K-means method is not appropriate for many image types. For example, if the image does not form spherical clusters, the method may not function effectively and the results may be impacted. As a result, numerous improvements to the Otsu algorithm have been developed, and more work can yet be done to address these shortcomings.

7. References

[1] Mehmood. Maryam, Shahzad. Ahsan, Zafar. Bushra, "Remote Sensing Image Classification: A Comprehensive Review and Applications", Hindawi . Mathematical Problems in Engineering, Volume 2022.

[2] Rashmi. C, Chaluvaiah. Sunil, Kumar. G, " An Efficient Parallel Block Processing Approach for *K*-Means Algorithm for High Resolution Orthoimagery Satellite Images", Procedia Computer Science 89 (2016),Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016), pp. 623 – 631.

[3] vishwanath. N, Ramesh. R, p. Sreennivasa, "unsupervised classification of remote sensing images using k – means algorithm", international journal of lasted trends in engineering and technology", vol. 7, issue 2, July 2016.

[4] Sungheetha. Akey, R . Rajesh, "Classification of Remote Sensing Image Scenes Using Double Feature Extraction Hybrid Deep Learning Approach", Journal of Information Technology and Digital World, Vol. 03, No. 02, 2021.

[5] Ali. Ishfaq, Rehman. Atiq, Khan. Dost, "Model Selection Using K-Means Clustering Algorithm for the Symmetrical Segmentation of Remote Sensing Datasets", Symmetry, Academic Editors: Palle E. T. Jorgensen and Mihai Postolache, 2022.

[6] Qu. Junrong, Qiu. Xiaolan, Wang. Wei, Wang. Zezhong, " A Comparative Study on Classification Features betweenHigh-Resolution and Polarimetric SAR Images through Unsupervised Classification Methods ", MDPI,Remote sensing,2022.

[7] Ali. Fadi, " Urban classification by pixel and object-based approaches for very high resolution imagery", thesis, Faculty Of Engineering And Sustainable Development, Department of Industrial Development, IT and Land Management, 2015.

[8] Srinivas. CH, Prasad. M, Sirisha. M, "Remote Sensing Image Segmentation using OTSU Algorithm",*International* Journal of Computer Applications, Volume 178, No. 12, May 2019.

[9] Khairuzzaman. Abdul Kayom, Chaudhury. Saurabh, " Masi entropy based multilevel thresholding for image segmentation", Multimedia Tools and Applications, springer,2019.

[10] wang. Qunming, shi. Wenzhong, " Unsupervised classification based on fuzzy c-means with uncertainty analysis", remote sensing letters, vol. 4, 2013.

[11] Ch. Anusha, "ECDSA-Based Water Bodies Prediction from Satellite Images with UNet", MDPI, water, vol. 14.2022.

[12] Jadwa. Sana'a, " Wiener Filter based Medical Image De-noising", International Journal of Science and Engineering Applications, Vol. 7, Issue 09, pp. 318-323, 2018.

[13] Ali. Ishfaq, Rehman. Atiq, " Model Selection Using K-Means Clustering Algorithm for the Symmetrical Segmentation of Remote Sensing Datasets", MDPI, Symmetry 2022.

[14] Ot. Eric, Olusola. Michael, " Comprehensive Review of K-Means Clustering Algorithms", International Journal of Advances in Scientific Research and Engineering, Vol. 7, Issue 8, 2021.

[15] kumar. Arpan, Tiwari. Anamika, " A Comparative Study of Otsu Thresholding and K means Algorithm of Image Segmentation", International Journal of Engineering and Technical Research, Vol.9, Issue.5, 2019 .

[16] Sun. Yu, Yang. Yingying, " An Adaptive Bi-Mutation-Based Differential Evolution Algorithm for Multi-Threshold Image Segmentation", MDPI, 2022.

[17] Zhao. Qi, Lyu. Shuchang, " A Multi-Modality Ovarian Tumor Ultrasound Image Dataset for Unsupervised Cross-Domain Semantic Segmentation", IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS,2022.