

# Clustering the Vegetation Areas using Fuzzy C-Means Algorithm

N. Saranya, N. Kanthimathi, A. Shyamalaprasanna, S. Vidhya, S. Dharani



**Abstract:** To detect the vegetation land from google earth image and clustered that vegetation land to get the different clusters and so the area of clustered land is calculated. The detection is done by land cover classification using fuzzy C-means clustering because it overcomes the disadvantage of K-means clustering algorithm because that clustered land is based on the land attributes not a particular distance. The exhibition of the FCM algorithm relies upon the choice of the primary cluster focus and the primary enrollment esteem. On the off chance that best primary cluster focus that is near the real, last cluster focus can be discovered, the FCM algorithm quickly cover the particular area and the preparing time can be radically decreased. Which altogether diminishes the calculation time required to segment a dataset into desired clusters?

**Keywords:** Fuzzy C means clustering, vegetated land, empty land, buildings and paths.

## 1. INTRODUCTION

### A. Clustering

An effective clustering (bunching) algorithm is essentially utilized in an unaided way. They are given a lot of information occurrences that ought to be gathered by some standard way of similitude. The essential calculation has at long last access just to the especially gathering of highlights portraying every single item; This isn't given any data about where every one of the occasions must be put inside the specific segment. Data bunching is an information investigation system that enables objects with comparable attributes to be assembled so as to encourage their further preparing. Bunching is one of the most precious undertakings inside information digging process to measuring a gatherings as well as distinctive fascinating conveyances and examples with regards to the hidden information.

The problem of bunching is related to the distribution of certain data index to gathering (groups) almost to an extreme that the range of data in a category seems to be more like each other instead of separate bunches. Clustering might be characterized as an information decrease apparatus for example used to make subgroups that are increasingly more sensible than singular datum. Essentially, bunching is legitimized as a procedure utilized for gathering countless information into important gatherings or groups dependent on some closeness between information. Bunches are the gatherings that have information comparable on premise of regular highlights and unlike information in different groups. The idea of a cluster, as found by divergent calculations, fluctuates fundamentally into their properties. Understanding these "bunching models" are the basic for accepting a varieties among the varying calculations. Bunch examination is an iterated strategy for information revelation and it's a variable measurable system that distinguishes groupings of the data items upheld the between object likenesses registered by a chose separation metric. Bunching calculations might be characterized into four classifications: various leveled grouping, Partitioned bunching, thickness based grouping and self-sorting out maps (SOM).

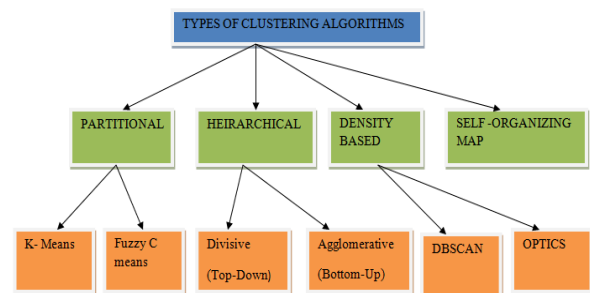


Fig.1 Types of Clustering Algorithm

### B. Types of Clusters

#### (i) Well-isolated groups:

A group is an accumulation of focuses to such an extent that some other point in a bunch nearer or progressively like every single other point in the group than to any point not in the bunch.



Fig.2 Well-isolated groups

Revised Manuscript Received on February 05, 2020.

\* Correspondence Author

**Mrs. N. Saranya M.Tech\***, Faculty of Electronics and communication Engineering, Bannari Amman Institute of Technology, Sathiyamangalam, Erode. Email: saranya@bitsathy.ac.in

**Mrs. Dr. N. Kanthimathi\***, Faculty of Electronics and communication Engineering, Bannari Amman Institute of Technology, Sathiyamangalam, Erode. Email: kanthimathi@bitsathy.ac.in

**Ms. A. Shyamalaprasanna.M.E\***, Faculty of Electronics and communication Engineering, Bannari Amman Institute of Technology, Sathiyamangalam, Erode. Email: shyamalapasanna@bitsathy.ac.in

**Ms. S. Vidhya**, PG Scholar of Electronics and communication Engineering, Bannari Amman Institute of Technology, Sathiyamangalam, Erode.

**Ms. S. Dharani**, PG Scholar of Electronics and communication Engineering, Bannari Amman Institute of Technology, Sathiyamangalam, Erode.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## (ii) Centre based bunches:

A bunch is a gathering of articles with the goal that an item in a group is all the more closer to the focal point of a group, than to the focal point

of other bunch. The focal point of a bunch is known as a centroid, the normal of the considerable number of focuses in the group, or a medoid, the most delegate purpose of a bunch.



Fig.3 Centre based bunches

## (iii) Contiguous groups: (Nearest neighbor or transitive)

A bunch is a gathering of focuses with the end goal that a solitary point in a group is more like at least one different indicates in the bunch than some other point not in the bunch.



Fig.4 Contiguous groups

## (iv) Density-based groups:

A group is thick district of focuses, which is individual isolated by low-thickness locales, from different areas of high thickness areas. It utilized when the bunches are sporadic, and when commotion and anomalies are accessible.



Fig.5 Density based groups

that have a base Euclidean separation. The bunch in the parcel is characterized by the ancient rarities of its part and the centroid. The centroid is where the aggregate of good ways from every one of the focuses in that group is diminished for every bunch. So K-means grouping is an iterative calculation in which over all bunches it limits the quantity of separations between each point and its group centroid. Euclidean separation is determined between the inside and every pixel of a picture is given by utilizing the formula,

$$d_i = ||s(m, n) - r_p||$$

after finding the Euclidean distance every one of the pixels is doled out to the closest focus dependent on separation d. hereafteragain calculate latest position centre by using these relation,

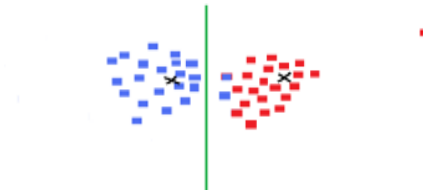
$$r_p = 1/p \sum_{m \in r_p} s(m, n)$$

## Advantages

- It's simple and durable.
- If there are a lot of variables, then K-means are determined more easily than hierarchical clusters.
- K-means produce tighter clusters than hierarchical clusters when the clusters are globular.

## Disadvantages

- It can classify the area only with Euclidean distance and cannot specify the individualities of vegetated plant.
- Compare difficulties with the quality of the produced clusters
- It can not be possible to predict a certain number of clusters
- Does not work well with non-globular clusters
- Various initial partitions can produce various final clusters.
- Sensitive to outliers.



## B. Proposed Methodology

Fuzzy C mean clustering is used in the vegetation land to categorize the difference of land according to their variety of crops cultivated. in this proposed methodology after clustering of the cultivated land and the area is calculated. i.e, vegetated land, empty land and its path is calculated. Fuzzy grouping is indeed a dominant unattended scheme for data analysis as well as model design. Fuzzy c means grouping are more common than other complicated clusters. Substances on multiclass borders are not pushed into total classes. Fuzzy C-means (FCM is an extreme technique for grouping that grants one piece of information to have a place with multiple bunches. It is regularly utilized in design acknowledgment strategy. It relies upon minimization of the accompanying target work:

$$V_n = \sum_{X=1}^M \sum_{Y=1}^C W_{xy}^n$$



## II.METHODOLOGY

### A. Existing Methodology

K-means clustering is used for segmenting the vegetation land which is one of the popular clustering method for dividing the collection of particular group of data. Through the clustering of K-means, a collection of data is partitioned into a k-number data set. Classifies the group of data into the disparate set of k-number. Here the two different stages of the K-means algorithms are presented. It calculates the k centroid in the first phase and in the second phase it takes that point from the respective data point to the cluster that has the nearest centroid. There are different strategies to decide the separation of the closest centroid and Euclidean separation is one of the most generally utilized techniques. Upon grouping the new centroid of each group is recalculated and fixated on that centroid, another Euclidean separation is resolved between each inside and every datum point and apportioned focuses inside the bunch

Where  $n$  is some genuine number which has more prominent than one ( $>1$ ),  $w_{xy}$  is the level of enrollment of

$E_x$  in the group,  $E_x$  is the  $x^{\text{th}}$ -dimensional estimated information,  $F_y$  is the  $d$ -measurement focus of the bunch, and  $\|*\|$  is any norm demonstrating the similitude between any deliberate information and the inside.

### 2.1.1 Steps for The Fuzzy C Means Clustering Algorithm

#### • Step:1 Initialization

First Select the initial membership functions its denoted as,

- (i)  $0 < W_{xy}^n(E_x) < 1$  for  $X=1,2,\dots,M$
- (ii)  $\sum_{X=1}^M W_{xy}^n(E_x) = 1$  for  $Y=1,2,\dots,C$
- (iii)  $0 < \sum_{y=1}^C W_{xy}^n(E_x) < C$  for  $X=1,2,\dots,M$

This is equivalent to specifying fuzzy clusters  $F_1, F_2, \dots, F_C$ . One Method to accomplish this selection is to chose  $r_{yx}$  randomly from the open interval (0, 1) and then Normalize,

$$W_{yx}^n(E_x) = r_{yx} / \sum_{Z=1}^M r_{yz}$$

Where,  $X=1,2,\dots,M$  and  $Y=1,2,\dots,C$

#### • Step:2 Computation of Fuzzy Centroids

Compute the Fuzzy Centroids as,

$$J_x = \sum_{y=1}^C W_{xy}^n E_x(E_x) / \sum_{y=1}^C W_{xy}^n(E_x)$$

Where,  $X=1,2,\dots,M$  and  $Y=1,2,\dots,C$

#### • Step:3 Compute New Fuzzy Membership Functions

Using the  $J_x$ ,  $X=1,2,\dots,M$  from step 2,  $W(E_x)$  is denoted as,

$$W_{xy}^n(E_x) = (1/d^2(E_x - F_y))^{1/n-1} / \sum_{Z=1}^M (1/d^2(E_x - F_Z))^{1/n-1}$$

Alternative method for computing the Membership Functions,

$$W_{xy}^n(E_x) = 1 / \sum_{Z=1}^M ((d^2(E_x - F_y) / (d^2(E_x - F_Z)))^{1/n-1}$$

Where,  $X=1,2,\dots,M$  and  $Y=1,2,\dots,C$

#### • Step:4 Check for Convergence

If algorithm converges then the  $W^n$  represent the fuzzy

clusters and we Stop. If Convergence has not occurred and the number of iterations is less than some pre assigned maximum values (MAXIT) then return to step 2. If otherwise then Stop with no solution.

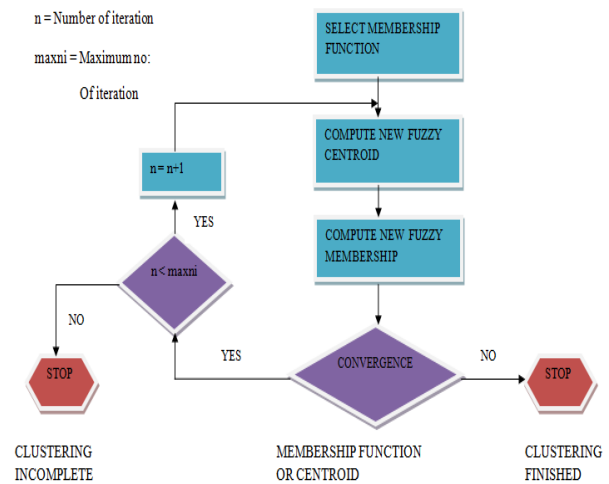


Fig.6 Flow Chart of Fuzzy C Means Clustering Algorithm

### III. RESULT AND DISCUSSION

FCM Grouping is an unsupervised technique of clustering which is useful to segmented images keep on to the cluster by way of closely related spectral properties. To calculate the membership function, it uses the cluster center as well as distance between each pixels in the spectral domain.

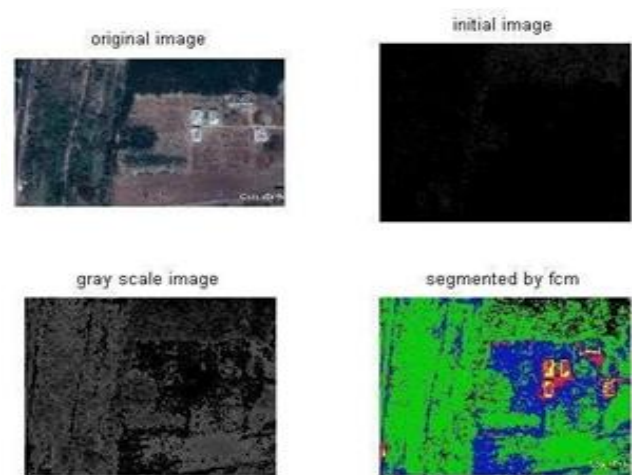


Fig.7 Google Earth Image1 clustered by Fcm

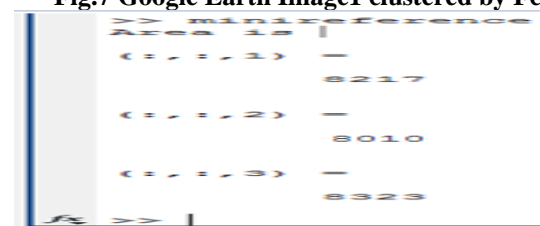
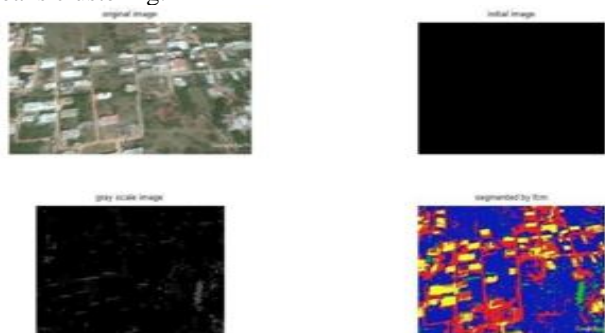
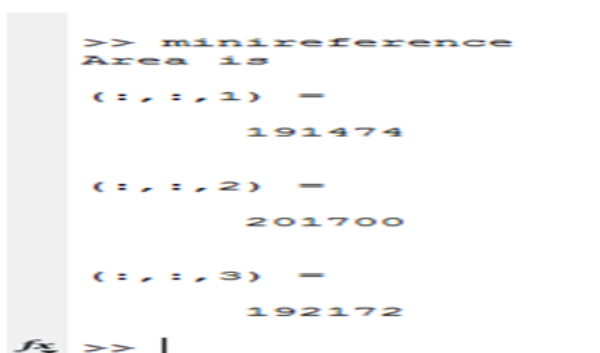


Fig.8 Area of the Google Earth Image1

In some images, are profoundly associated, and spatial data is a fundamental feature that can be utilized to help their marking. The spatial connection between pixels is barely utilized in fcm. This article is to find the actual cluster by using a simplified set of the original complete dataset. Area is calculated for each cluster (i.e.) vegetated land, empty land and area of the path is calculated by using fuzzy c means clustering.



**Fig.9 Google Earth Image2 clustered by Fcm**



**Fig.10 Area of the Google Earth Image2**

In Fig.7, The original image is initially converted into zeros and ones pixels (B&W). B&W image is converted into Grey scale image for the purpose of discriminating the pixels ranging from 0 to 255. The Grey scale image is converted into RGB to add Hue and Saturation information by eliminating luminance to perform Fuzzy C means clustering. After land is clustered by using FCM, Red color denotes the Path, Green color denotes the vegetated crops and The Blue denotes empty land. Area is calculated for each land (no. of pixels are calculated for each land).

## IV.CONCLUSION

In various applications clustering has a crucial part to play K means, Which is commonly utilized inefficient clustering algorithm. K means clustering is now a foremost research topic in data mining for a few days but there is a problem of improving a cluster creation, cluster quality and algorithm efficiency. In order to overcome the disadvantage of K means clustering algorithm, The Fuzzy C Means algorithm is adapted. It is used to find the vegetated crops, waste lands, buildings, path and area of each land is calculated.

## REFERENCES

1. Chen yanyun, Qiu jianlin, Gu xiang and Chen jianping, "Advancesinvestigate of Fuzzy C-means clustering Algorithm" IEEE, vol.12, no.6, 2018.
2. Namierakpm Dhamachandra, Kumanthem Manlem and Yambems Jiena Chanu, " Image Segmentation utilizingK means Clustering

Algorithm and Subtractive Clustering Algorithm" ELSEVIER, vol.6, no.13, 2017.

3. Jayanti Mehra, RS Thakur" Probability density based Fuzzy C Means clustering for web usage mining" Internal Journal of Innovative Technology and Exploring Engineering (IJITEE), vol.8, no.4, 2019.
4. James C. Bezdek, Robert Enrich, William Full "Survey of Fuzzy C means clustering data mining" ELSEVIER, vol.10, no.23, 2016.
5. M.E Jopin Christi, R.PParvathidevi "Fuzzy C means algorithm image division in medical field" IEEE, vol.17, no.5, 2017.
6. Shi nau, Li Xumina, Guna Yongi " Advance Research on K means clustering algorithm: An improved k means grouping technique" IEEE, vol.4, no.16, 2016.
7. Xini Zhengge, Qiyi Lein, Runa yao and Qiani Yina " Image subdivision Based on the adaptive K means algorithm" Springer, vol.26, no.68, 2018.
8. Huiyi zhoi, Geralad Schaefare, Chuinmei Shye "Fuzzy C means technique for medical Image subtraction" Springer, vol.13, no.41, 2015.
9. S.S. Tharieb, R.Bendy, A.Lbdelfattah "Image Segmentation Using Fuzzy C means Clustering fusing weighted neighborhood supplement enrollment and nearby information separations" IEEE vol.5, no.17, 2016.
10. Hu Xini Peyi, Rongu rong zhengi, Chanwang, Chun ra Gi "D-FCM : Density based fuzzy C means clustering algorithm with application of identify the abnormal tissue of our body" ELSEVIER, vol.122, no.26, 2017.
11. H.P.Ng, Sim hen gong, Kelvin w.c.fong "Medical Image Subtraction using K means technique and enhanced watershed algorithm" IEEE, vol.12, no.43,2016
12. Ajala funmilola alaba, Poh-sun goh " Comparision of K maeans algorithm and Fuzzy C means clustering algorithm for segmentation of image in medica field" Springer, vol 34, no.22, 2015
13. Anil chitade, chunmei shi "Image text based image segmentation using K-means clustering" ELSEVIER, vol.19, no.8, 2016.
14. Mythili S, Madiya E "An Analysis of Clustering Algorithm in Data mining" IJITEE, vol.3, no.12, 2015.
15. V.Guau, G.T.J. Metes, T.Hlcanize, Y.Kikinise "Based on Prior Data to enhance the Watershed algorithm for Image Segmentation in medical field" IEEE, vol.24, no.4, 2014.

## AUTHORS PROFILE



**Mrs.N. Saranya. MTech**, Faculty of Electronics and communication Engineering , Bannari Amman Institute of Technology, Sathiyamangalam, Erode. Email: [saranyan@bitsathy.ac.in](mailto:saranyan@bitsathy.ac.in)



**Mrs. Dr.N.Kanthimathi**, Faculty of Electronics and Communication Engineering, Bannari Amman Institute of Technology, Sathiyamangalam, Erodec. Email: [kanthimathi@bitsathy.ac.in](mailto:kanthimathi@bitsathy.ac.in)



**Ms. A. Shyamalaprasanna. M.E**, Faculty of Electronics and communication Engineering ,Bannari Amman Institute of Technology, Sathiyamangalam, Erode. Email: [shyamalaprasanna@bitsathy.ac.in](mailto:shyamalaprasanna@bitsathy.ac.in)