Clustering Amelioration and Optimization with Swarm Intelligence for Color Image Segmentation

Kiranpreet and Prince Verma

M.tech CSE Dept., CT group of engg., mgmt.& tech., Asst.proff. M.tech CSE Dept.,CT group of engg.,mgmt.&tech.
Jalandhar(India)
preetk822@gmail.com, prince.researchwork@gmail.com

Abstract

Cluster examination is data mining task for the assignment of collection a set of items in such a path, to the point that questions in the same gathering (called a cluster) are more like one another than to those in different gatherings (clusters). K-means grouping is a technique for group investigation which intends to parcel n perceptions into k groups in which every perception fits in with the cluster with the closest mean. This paper, decided the aftereffect of standard parameter estimations of shading picture division with k-means and the modified k-means with ABC and ACO algorithms. The paper demonstrates that division of color picture with modified k-mean consolidated with swarm Intelligence calculations for color image segmentation gives preferable results over simple k-means and Modified k-means with Ant colony optimization gives better results than modified k-means with Artificial bee colony .

Keywords: *Data mining, clustering, k-means algorithm, swarm intelligence, artificial bee colony, ant colony*

1. Introduction

Data mining (in some cases called information or learning revelation) is the procedure of breaking down information from alternate points of view and compressing it into helpful information. It permits clients to examine information from a wide range of measurements or edges, order it, and compress the connections identified. It is a valuable system needed for the convergence of machine learning, dbase frameworks. The real information mining undertaking as in [5] is the programmed or self-loader examination of substantial amounts of information to concentrate already obscure fascinating examples, for example, gatherings of information records (cluster investigation), strange records (abnormality discovery) and conditions (affiliation principle mining).

Cluster examination or clustering is the errand of accumulation an arrangement of things in such a way, to the point that inquiries in the same social event (called a cluster) are more like each other than to those in diverse get-togethers (clusters). It is an essential undertaking of exploratory information mining, and a regular framework for verifiable data examination, used as a piece of various fields, including machine learning, illustration recognitions, picture examination and bioinformatics. Clustering is a philosophy of allocating arrangement of data (or articles) into an arrangement of vital sub-classes, called clusters. The objective of clustering is to section the practically identical thing are assembled in one group and distinctive are amassed in other gathering. Clustering uses in various fields *e.g.* simulated neural network, digital picture processing, networking .Clusters can be classified by dividing the objects into different classes having the similarities and it can be differentiate with the some different colors or different names to the clusters. In the given Figure1 there are large number of objects and objects are

divided into three different clusters on the basis of the similarities and clusters are differentiate by having the three different colors *i.e*. red, blue and green color to each color.

2. Preliminaries

2.1. K-Means Clustering

The crucial venture of k-means grouping is fundamental. At first we center number of K cluster and we acknowledge the centroid or center of these groups. K-means clustering is a framework for clusters examination which hopes to fragment n recognition into k group in which each recognition has a spot with the cluster with the nearest mean. By then the K means count will do the three stages underneath until combining Iterate until stable (= no article move group):
   1. Center the centroid coordinate
   2. Center the partition of every one article to the centroids
   3. Cluster the thing considering minimum detachment.
Algorithm 1: k-mean calculation
Let X = {x1,x2,x3,… … ..,xn} be the arrangement of data centers and V = {v1,v2,… … .,vc} be the arrangement of core interests.
   1. Subjectively select "c" gathering core interests.
   2. Learn the detachment between every data point and cluster centers.
   3. Name the data point to the gathering center whose division from the group center is in particular the cluster centers.
   4. Recalculate the new gathering center using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

Where, "ci" speaks to the quantity of information focuses in ith cluster.
   5. Recalculate the separation between every information point and new acquired group focuses.
   6. On the off chance that no information point was reassigned then stop, otherwise repeat from step3.

2.2. Artificial Bee Colony

Artificial bee colony algorithm is a standout amongst the latest advancement calculations that are composed concurring nourishment pursuit methodology of bumble bees. [2, 7].

In the ABC model, the settlement comprises of three gatherings of bees: employedbees, onlooker and scouts. It is expected that there is one and only counterfeit utilized honey bee for every sustenance source. As it were, the quantity of utilized employedbees in the province is equivalent to the quantity of sustenance sources around the hive. Utilized employedbees go to their sustenance source and return to hive and move on this territory. The utilized employedbee whose nourishment source has been surrendered turns into a scout and begins to look for discovering another sustenance source. Onlooker watches the moves of utilized honey bees and picks sustenance sources relying upon moves. The principle steps of the calculation are given underneath:
   1. Starting sustenance sources are created for all utilized employed bees
   2. Repeat UNTIL (prerequisites are met)
       2.1 Each one utilized employedbee goes to a sustenance source in her memory and decides a neighbor source, then assesses its nectar sum and moves in the hive

2.2 Every passerby watches the move of utilized employed bees and picks one of their sources relying upon the moves, and afterward goes to that source. In the wake of picking a neighbor around that, she assesses its nectar sum.

2.3 Deserted nourishment sources are dead set and are supplanted with the new sustenance sources found by scouts.

2.4 The best sustenance source discovered so far is enlisted.

2.3. ANT Colony Optimization

The Ant Colony Optimization [27] is the optimization strategy which runs on Ant System. In the Ant System approach the ant behaviour is studied when they are in search of feeding sources. The problem here studied was to understand howl most blind ants manages to establish the shortest route path from their colony to feeding sources and come back. And it was found that to communicate the information among individuals regarding paths, and decide the feasible path is due to pheromone trails [27].A moving ant lays some pheromone on their way and thus creating the trail. An isolated ant moves at random, and when encounters a previously laid trail then it decides to follow the path with dense pheromone trail as it is more probable path (*i.e.* shortest path) to follow.
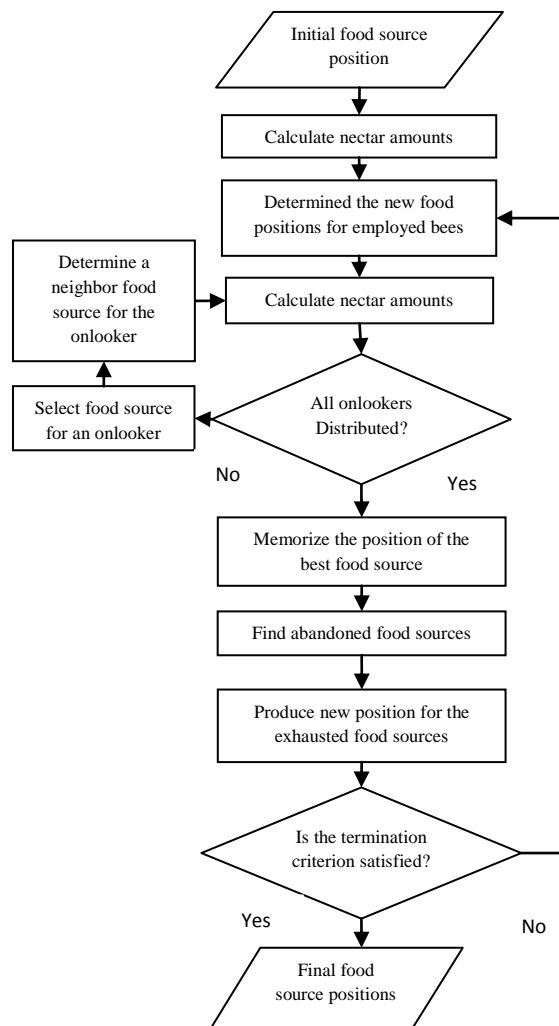

Figure 1. Artificial Bee Colony

The pheromone density is different for every path as pheromone gets updated every time and the pheromone trail is updated due to trail evaporation [27] which evaporates

those paths which are not followed rapidly. The trail evaporation with time results with only a single path that should be followed.The principle steps of the calculation are:

1. Ants(blind)navigate from home to sustenance source
2. Shortest way is found by means of pheromone trails
    2.1 Each burrowing little creature moves aimlessly
    2.2 Pheromone is saved on way
    2.3 When way is suitable, more ants take after that way and more is pheromone affidavit
    2.4 More pheromone further builds the likelihood of way being taken after
3. Virtual "trail" amassed on way portions
4. Starting hub is chosen aimlessly
5. Ant when reaches next hub, chooses next way
6. Continues until achieves beginning hub
7. Finished "visit" is an answer
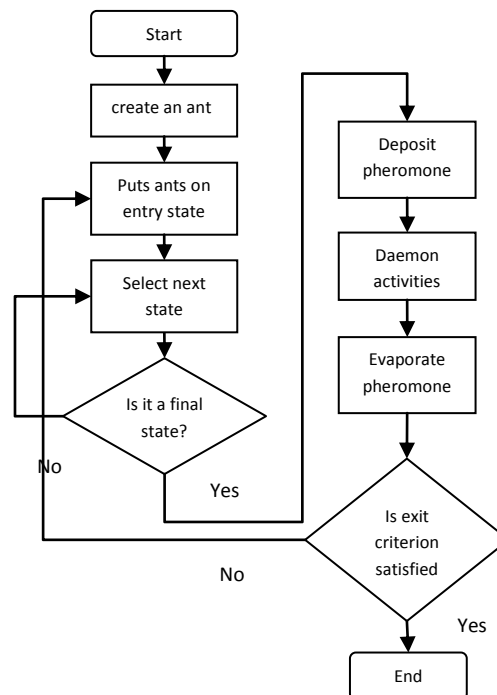8. Can be utilized for both Static and Dynamic Combinatorial streamlining issues.

Figure 2. ANT Colony Optimization

3. Proposed Approach

3.1. Modified K-Means

K-meansalgorithm chooses the initial clusters as randomly. The choice of initial cluster defines the efficiency of this algorithm.Concerning effectiveness purposes, first all information articles are clustered utilizing k-means calculation to discover the starting group focuses to be utilized as a part of the arrangements taking into account all their characteristics. Based on produced cluster the example for an item is created from each one characteristic at any stage. Objects with the same examples are found in one group and subsequently all articles are clustered. The obtained clusters in this stage will be more than the first number of clusters. In this approach, Clustering is done in two stages. The

main stage is executed as above and in the second stage; comparable clusters are incorporated with one another until attaining to a given number of groups. Algorithm 2 demonstrates the proposed methodology for introductory clustering of information articles and the accomplished cluster focuses are called seed focuses.

Algorithm 2.MK-mean algorithm

1. Input data set (d1, d2, …..dn) its attributes (a1,a2,……aq) and K number of clusters.
2. Start with ($\forall$ aj$\in$a) and find the standard deviationand mean
3. Find the center of K number of cluster using above standard deviation and mean
4. Apply k means algorithm on this attribute and give name to obtained clusters.

3.2. ABCMK-Means

Artificial bee colony is a populace based optimization technique or it's a swarm based algorithm.ABC algorithm is used to solve a huge set of numerical test functions.And Modified K-means clustering algorithm is a unsupervised learning algorithm .It is used to solve a fundamental clustering problems. ABCMK-means is applied for the color image segmentation. ABCMK-means is applied on the color images by the following steps:

1. Select the image and initialize variables of ABC.
2. Double the image which represents heuristic parameters
3. Divide the image into R,G,B channels
4. Determine and Apply adaptive masking and thresholding of red.
5. Apply morphological operations on R, G, B image into small segments to reduce artifact of image segmentation.
6. Apply convolution operations which divide the image into three channels *i.e.* hue,saturation and intensity for segmentation
7. Select the color intensity value max. High value is 255; medium is 100 and low is 0.
8. Apply ABC on each plane by the following steps:
9. Initial sustenance sources are delivered for every single utilized employed bees.
10. REPEAT UNTIL (prerequisite are met)
    10.1 Each utilized employed bee goes to a sustenance source in her memory and decides a neighbor source, assesses its nectar sum and moves in the hive.
    10.2 Each onlooker watches the move of utilized employed bees and pick one of their sources relying upon the moves, and afterward goes to that source.
    10.3 After picking a neighbor around that, she assesses its nectar amount.
11. Abandoned nourishment sources are dead set and are supplanted with the new sustenance sources found by scouts.
12. The best nourishment source discovered so far is enrolled.

At last the final segmented image of ABC-MK means shows as an output or as a result.

3.3. ACOMK-Means

The algorithmincludes a population of ants. Each ant is operating as an autonomous agent thatreorganizes data patterns during exploration to reach an optimal clustering. Modified k-means algorithm is powerful algorithm to find the initial cluster.ACOMK-means is applied on the color images by the following steps:

1. Select the image and initialize variables of ACO.
2. Double the image which represents heuristic parameters
3. Divide the image into R,G,B channels
4. Determine and Apply adaptive masking and thresholding of red.
5. Apply morphological operations on R, G, B image into small segments to reduce artifact of image segmentation.
6. Apply convolution operations which divide the image into three channels *fig* hue, saturation and intensity for segmentation

7. Select the color intensity value max. High value is 255; medium is 100 and low is 0.
8. Apply ACO on each plane by the following steps:
    8.1 Initialize variables of ACO *i.e.* lambda
    8.2 Double the segmented image which represents heuristic parameters.
    8.3 Initialize 2-D array X and Y *i.e.* distance between vertex X and Y.
    8.4 Select the size of the image in row, column and dimension.
    8.5 Initialize the pheromone parameters *i.e.* α and β.
    8.6 Initialize the pheromone, on the edges and ant are going to search the best path.
    8.7 Inverted distance will help the ants to find the new path.
    8.8 Update the pheromones, which will be an adjacency matrix.
    8.9 Repeat the entire process until there are no changes in the best path for X iterations.

At last the final segmented image of ACOMK- means shows as an output or as a result.
Advantage of ACO over ABC

o Ant colony optimization on modified k-means for proficient color image segmentation is proposed as it performs better against other global optimization techniques as:
o It holds memory of whole colony rather than past generation only.
o It is less influenced by poor introductory arrangements (because of mix of irregular way choice and province memory)
o Can be utilized as a part of element applications (adjusts to changes, for example, new separations, and so on.)
o It has more exactness of the ideal worth.
o It is nearly quick to meeting.

Table 1. I mages Used for Implementation of ABC-MK Means & ACO-MK Means

Image1: Satellite Image

Image2: Underground Image1

Image3: Pebble Image1

Image4: Butterfly Image

Image5: Underwater Image

Image6: Horses Image

4. Implementation and Results

Color images are used on different conditions *i.e.* first image is of satellite image, second image is of underground image, third image is of pebbles image, fourth image is of underground image, fifth image is of pebbles image, sixth image is of horse image. The suggested algorithm is coded by an appropriate programminglanguage and is run in an i5 computer with 2.60 GHz microprocessor speed and 4GB main memory. Window7 home basic is used for doing the research work. MATLAB tool is utilized for implementation and evaluation of the results.MS office and Microsoft excel is also used in it.To evaluate the results of the ABCMK-means and ACOMK-means for the color image segmentation parameters used are *i.e.* accuracy,sensitivity,specificity,F-measure,Bit error rate and execution time.ACOMK-means gives better results than the ABCMK-means because ACOMK-means has high accuracy, sensitivity, specificity, F measure and less bit error rate and execution time.Color images are used on the different conditions *e.g.* satellite image, pebbles images, underground images, horse images because the accurate results will be find out rather than on the same conditions. The implementation of ABCMK-means and ACOMK-means are use for the segmentation of different color images. With the help of readings and graphs it is prove that the ACOMK-means gives better results than the ABCMK-means in the segmentation of the color images. And ACOMK-means a powerful algorithm than the ABCMK-means. MATLAB 2013 tool is used for implementing the results.

4.1. Accuracy

AC is the proportion of the total number of predictions that were correct. It is determined using the equation:

$$Accuracy = A+D/ A+B+C+D$$

Where

A refers to True positive, B refers to False negative, C refers to False positive and D refers to True negative.

Table 2. Accuracy of ABC-MK Means & ACO-MK Means

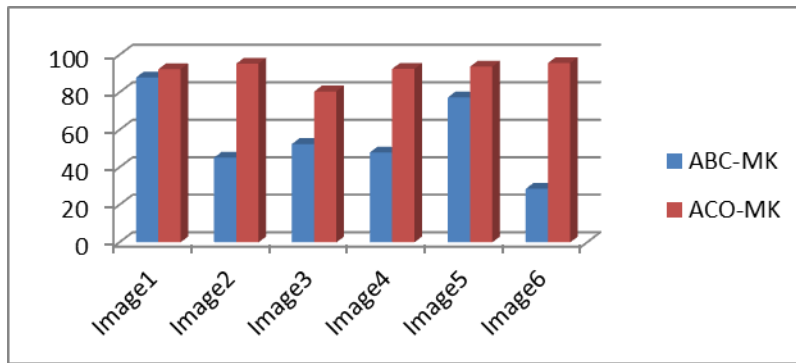| Algorithms	Image1	Image2	Image3	Image4	Image5	Image6
ABC-MK	87.987	45.160	52.470	63.582	21.353	28.581
ACO-MK	92.334	95.354	80.519	80.058	94.702	95.695

Figure 3. Graphical Representation of Accuracy

4.2. Sensitivity

It is the proportion of positive cases that were correctly identified, as calculated using the equation:

$$\text{Sensitivity} = A/A+B$$

Table 3. Sensitivity of ABC-MK Means &ACO-MK Means

Algorithms	Image1	Image2	Image3	Image4	Image5	Image6
ABC-MK	0.0029	0.0513	0.2652	0.4729	0.0771	0.8127
ACO-MK	0.7187	0.9710	0.9051	1.000	0.8774	0.9712
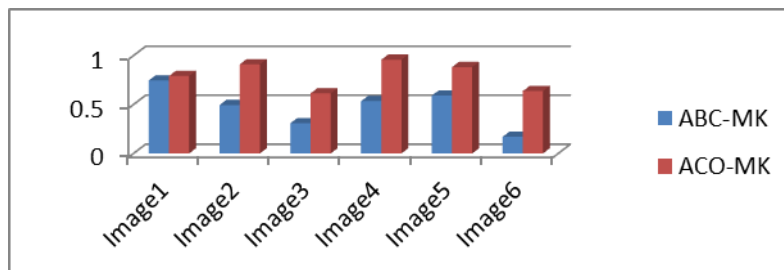
Figure 4. Graphical Representation of Sensitivity

4.3. Specificity

It is defined as the proportion of negatives cases that were classified correctly, as calculated using the equation:

$$\text{Specificity} = D/C+D$$

Table 4. Specificity of ABC-MK Means &ACO-MK Means

Algorithms	Image1	Image2	Image3	Image4	Image5	Image6
ABC-MK	0.7458	0.4943	0.3079	0.3305	0.2952	0.1702
ACO-MK	0.7904	0.9099	0.6136	0.7239	0.8853	0.6368
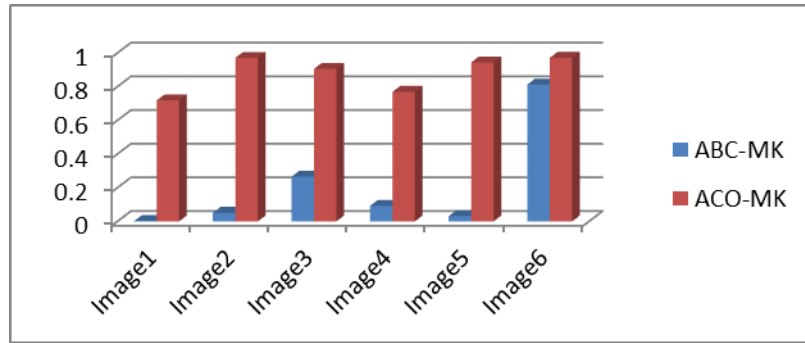
Figure 5. Graphical Representation of Specificity

4.4. F-Measure

The F-Measure computes some average of the information retrieval precision and recall metrics. The F-measure of cluster j and class i is defined as follows:

$$Fij = 2*Recall(i,j)*precision(i,j)/precision(i,j)+Recall(i,j)$$

Table 5. F-Measure of ABC-MK Means &ACO-MK Means

Algorithms	Image1	Image2	Image3	Image4	Image5	Image6
ABC-MK	84.437	39.549	42.968	46.374	20.008	28.367
ACO-MK	87.142	94.909	72.290	83.978	91.973	73.632
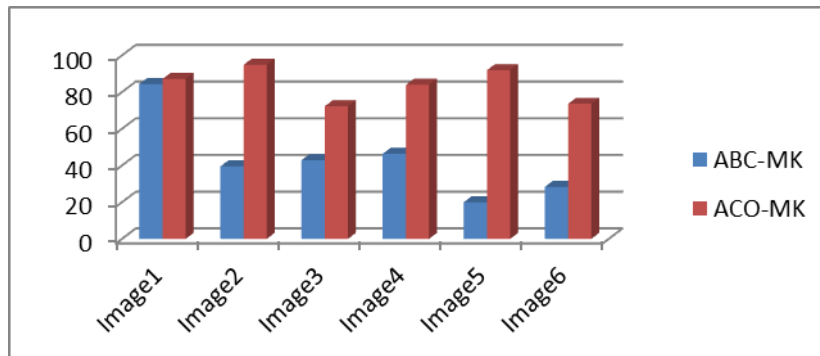
Figure 6. Graphical Representation of F-Measure

4.5. Bit Error Rate

The bit error is the no. of bit error per unit time. The bit error ratio is the no. of bit errors divided by the total number of transferred bits during a studied time interval. Bit error rate will calculated as:

$$BER = 1-Accuracy$$

Table 6. Bit Error Rate of ABC-MK Means & ACO-MK Means

Algorithms	Image1	Image2	Image3	Image4	Image5	Image6
ABC-MK	1.3355	1.8330	1.7540	1.2447	2.6857	5.0854
ACO-MK	0.6627	0.5316	0.6584	0.5801	0.5673	0.6219
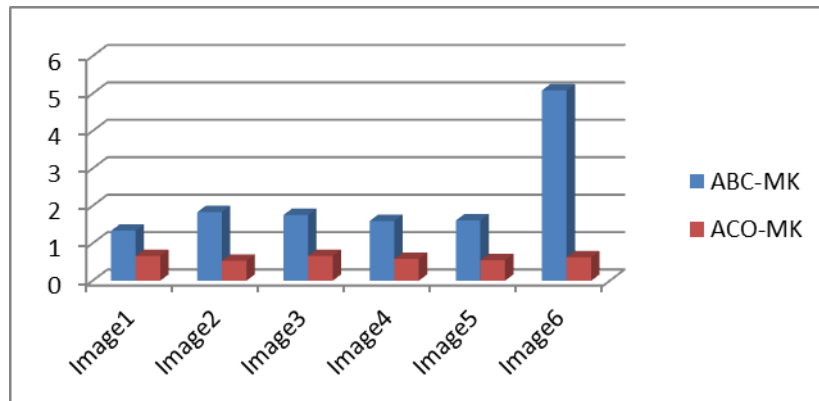
Figure 7. Graphical Representation of Bit Error Rate

4.6. Execution Time

Execution time is the time acquired in which the given color image is segmented and the appropriate results are provided.

Table 7. Execution Time of ABC-MK Means & ACO-MK Means

Algorithms	Image1	Image2	Image3	Image4	Image5	Image6
ABC-MK	2.9185	18.614	9.7959	13.745	120.64	13.422
ACO-MK	1.1815	4.4721	2.4289	4.1351	31.379	3.5171
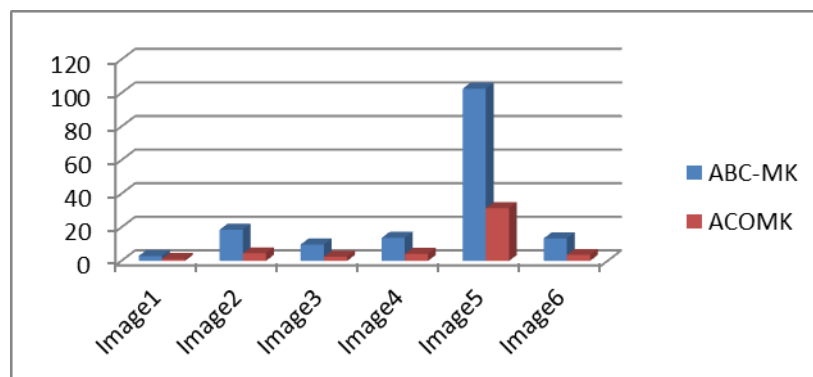
Figure 8. Graphical Representation of Execution Time

Table 8. Segmented Images of Color Images after Applying the ABC-MK Means
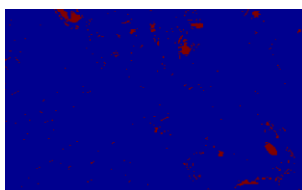
Image1: Satellite Image
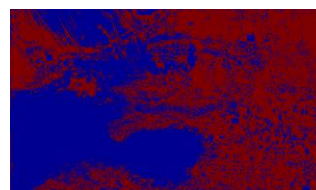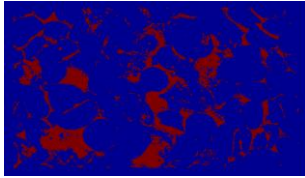
Image2: Underground Image1
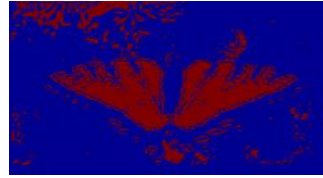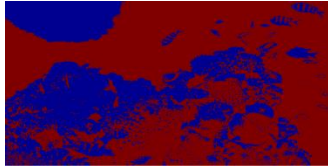
Image3: Pebble Image1

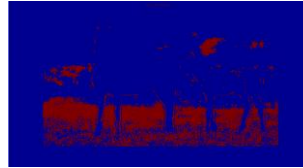Image4: Butterfly Image

Image5: Underwater Image

Image6: Horses Image

Table 9.Segmented Images of Color Images after Applying the ACO-MK Means
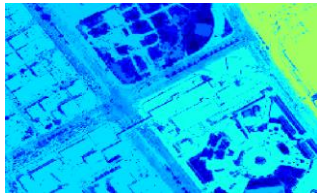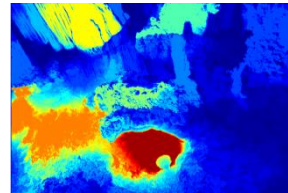

Image1Satellite Image
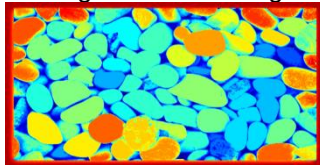

Image2: Underground Image1
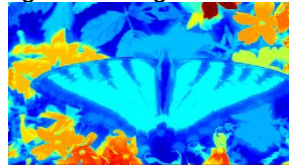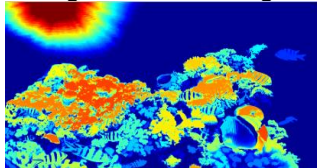

Image3: Pebble Image1

Image4: Butterfly Image

Image5: Underwater Image

Image6:Horses Image

5. Conclusion and Future Work

In this paper, modified k-means algorithm with compelling joined ANT colony optimization and artificial bee colony algorithm is proposed for ideal arrangements in color images. By utilizing these calculations, images can be differentiated into distinctive parts of images called fragments with high accuracy, sensitivity, specificity & f-measure and less bit error rate& execution time. This research work is utilized to focus the best cluster centers of images with modified k-mean clustering. The ideal cluster centers are gotten by combination of ANT colony optimization with modified k-means clustering algorithm and Artificial Bee Colony algorithm with modified k-means clustering algorithm. The results of comparison shows that proposed algorithm with ANT colony optimization for color images with optimum clusters has the capacity better segmentation and give preferable results over Artificial bee colony algorithm utilizing the parameters.I n future work some other optimization techniques *i.e.* genetic algorithm or fuzzy logic

may be applied to the modifiedalgorithm (MK) or the optimization strategies ABC and ACO can be applied to some other clustering algorithm.

References

[1] Jain A., Rajavat A. and Bhartiya R., "Design, Analysis and Implementation of Modified K-Mean Algorithm for Large Data-Set to Increase Scalability and Efficiency", Fourth International conference on computational intelligence and communication network, IEEE (2012), pp. 627-631.

[2] Akhtar N., Agarwal N. and Burjwal A., "K-mean algorithm for Image Segmentation using Neutrosophy", International conference on advances in computing communications and informatics, IEEE, (2014), pp. 2417-2421.

[3] C. Jacob and K. A. Abdul Nazeer, "An improved ICPACA based K-means algorithm with self-determined centroids", International conference on data science and engineering, IEEE, (2014), pp. 89-93.

[4] H. Parvin and B. M. Bidgoli, "A New Space Defined by Ant Colony Algorithm to Partition Data", Adaptive and Intelligent Systems, springer, vol. 6943, (2011), pp. 272-284.

[5] M. BabrdelBonab, S. Z. M. Hashim, A. K. Z. Alsaedi and U. Raba'ah Hashi, "Modified K-means Combined with Artificial Bee Colony Algorithm and Differential Evolution for Color Image Segmentation", Computational Intelligence in Information Systems Advances in Intelligent Systems and Computing, springer, vol. 331, (2015), pp. 221-231.

[6] S. Saatchi and C. C. Hung, "Hybridization of the Ant Colony Optimization with the K-Means Algorithm for Clustering", 14th Scandinavian Conference, springer, vol. 3540, (2005), pp. 511-520.

[7] Y. Yang, M. Kamel, and F. Jin, "Clustering Ensemble Using ANT and ART", studies in computational intelligence, springer, vol. 34, (2006), pp. 243-264.

[8] V. R. Patel and R. G. Mehta, "Modified k-Means Clustering Algorithm", Communications in Computer and Information Science, Springer, vol. 250, (2011), pp. 307-312.

[9] Y. J. Wei, C. J. Ming, X Bai-li and Z. Jian, "An Enhancing K-Means Algorithm Based on Sorting and Partition", Springer, vol. 391, (2013), pp. 365-374.

[10] Visalakshi N. K., "K-means clustering using Max-min distance measure", Annual meeting of the north America on fuzzy information processing society, IEEE, (2009), pp. 1-6.

[11] W. Z. Feng, M. X. Fan, L. Qiao and Q. Z. Guang, "Logical Symmetry Based K-means Algorithm with Self-adaptive Distance Metric", Advances in Computer Science and its Applications, Springer, vol. 279, (2014), pp. 929-936.

[12] Shah S. and Singh M., "Comparison of a Time Efficient Modified K-mean Algorithm with K-Mean and K-Medoid Algorithm", International conference on communication system and network technologies, IEEE, (2012), pp. 435-437.

[13] Singh R. V. and Bhatia M. P. S., "Data clustering with modified K-means algorithm", International Conference on recent trends in information technology, IEEE, (2011), pp. 717-721.

[14] J. R. Souza, T. B. Ludermir and L. M. Almeida, "A Two Stage Clustering Method Combining Self-Organizing Maps and Ant K-Means", Artificial Neural Networks – ICANN, Springer, vol. 5768, (2009), pp. 485-494.

[15] J. L. Hsu and H. X. Yang, "A Modified K-means Algorithm for Sequence Clustering", International conference on Hybrid intelligent systems, IEEE, (2009), pp. 287-292.

[16] N. Sharma, A. Bajpai and Mr. R. Litoriya, "Comparison the various clustering algorithms of wekatools", International Journal of Emerging Technology and Advanced Engineering, vol. 2, no. 5, (2012), pp. 73-80.

[17] A. Joshi and R. Kaur "Comparative Study of Various Clustering Techniques in Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 3, (2013), pp. 55-57.

[18] B. Chaudhari and M. Parikh, "A Comparative Study of clustering algorithms Using wekatools", International Journal of Application or Innovation in Engineering & Management (IJAIEM), vol. 1, no. 2, (2012), pp. 154-158.

[19] G. Singh, N. Kaur, "Hybrid Clustering Algorithm with Modifications Enhanced K-Means and Hierarchal Clustering", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 5, (2013), pp. 166-170.

[20] W. MaseriBinti Wan Mohd, A. H. Beg*, T. Herawan and K. F. Rabbi, "An Improved Parameter less Data Clustering Technique based on Maximum Distance of Data and Lioyd k-means Algorithm", Elsiever, (2012), pp. 367-371.

[21] Y. Xia and B. Xi, "Conceptual Clustering Categorical Data with Uncertainty Indiana University", Purdue University Indianapolis Indianapolis, IN 46202, USA.

[22] XuR. Survey of clustering algorithms. IEEE Trans. Neural Networks, (2005).

[23] Bohm C., Kailing K., Kriegel H. P. and Kroger P., "Density connected clustering with local Subspace preferences", In Proceedings of the 4th International Conference on Data Mining (ICDM), (2004).

[24] R. Ng and J. Han, "Efficient and effective clustering method for spatial data mining", In: Proceedings of the 20th VLDB Conference, Santiago, Chile, **(1994)**, pp. 144-155.

[25] F. Shao and Y. Cao, "A New Real-time Clustering Algorithm Department of Computer Science and Technology", Chongqing University of Technology Chongqing 400050, China.

[26] S. Kisilevich, F. Mansmann and D. Keim, "P-DBSCAN: A density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos", University of Konstanz.

[27] M. Dorigo, V. Maniezzo and A. Colorni, "The ant system: optimization by a colony of cooperating agents", IEEE Trans. Systems Man Cybernet. B 26, **(1996)**, pp. 29–42.