International Journal of Advances in Engineering Research http://www.ijaer.com

(IJAER) 2022, Vol. No. 23, Issue No. III, March **e-ISSN: 2231-5152, p-ISSN: 2454-1796**

SEGMENTATION OF MICROARRAY IMAGES USING MULTIPLE FEATURE CLUSTERING ALGORITHMS

***Ulli Moulali, **Dr. Syed Umar**
**Research Scholar, **Supervisor,*
Department of Computer Science, Faculty of Computing & Information Technology (Computer Science), Himalayan University, Itanagar

ABSTRACT:

In a single experiment, utilizing Microarray Technology, one may examine the function of thousands of genes simultaneously. Medical diagnostics, drug development, and biomedical research all rely on microarrays. Numerous copies of a single DNA sequence may be found in every single one of the microarray's many thousands of locations. There are three steps to the microarray image analysis process: gridding, segmentation, and extraction of information (or data). Data from microarray image analysis may be used to classify and identify differentially expressed genes by creating spot metrics from spot intensities. Each spot's intensity shows the gene's expression level. Multi-feature clustering techniques for microarray image segmentation are presented in this study, which expands the single-feature (pixel intensity) k-means clustering approach to multiple features. Multi-feature clustering methods are more efficient than single-feature k-means clustering algorithms in sizing the spot area, thereby generating more accurate expression-ratio results.

Keywords: Image processing, Microarray Image Analysis, Multi-feature Clustering methods

1. INTRODUCTION:

Microarray image processing relies heavily on picture segmentation in order to concurrently capture the emotions of individual spots. Disturbances may affect the performance of modern clustering-based segmentation algorithms. In order to deal with this issue and enhance the precision of segmentation, several modifications are made to the rapid and basic clustering algorithms in this study. Microarray image segmentation relies only on pixel intensity information when performing clustering procedures. Nevertheless, in the microarray image segmentation process, depending on where a pixel is located and how close it is to its neighbor’s median values, the clustering and segmentation results will differ. Numerous feature-clustering algorithms were developed as a result of this finding. For microarray image segmentation, k-means clustering with a single feature (pixel intensity) may be improved using the multiple feature clustering algorithms discussed in this paper. Clustering algorithms are the primary focus of this investigation. They don't need a starting condition of pixels and don't require any post-processing since these algorithms aren't restricted to a certain size and form of spot. These algorithms were created only using information about the pixel intensities (one feature). Microarray image segmentation, however, takes into account not just the brightness of individual pixels, but also their distance from the median intensity of the pixels around them, all of which have an effect on the final result. It's critical to use feature selection when dealing with huge datasets that include a significant number of features and metrics. A fraction of the components could include the high point of characteristics, and they might corrupt the execution and impair the many-sided computational quality of the networks. So the efficiency may be increased by the selection of striking characteristics, by disposing of minor features, from huge datasets. These picked parts may enhance the execution of the networks and can also reduce the complicated computational nature. Several analysts have provided unique methods for feature selection that are worth mentioning. Selecting the most appropriate collection of characteristics may be accomplished via the use of several techniques such as classification and grouping, sequential feed forward selection and reverse search, and successive scanning point procedures, among others. In addition, efficient methods for feature selection depends on the similarity measure and fuzzy rules are classified and discussed. As an additional point of interest, Since 1997, the extraction of gene expression levels from microarray images has attracted growing attention due to the need of segmenting the images. Many segmentation methods have been suggested to meet this requirement. These segmentation techniques may be divided into seven groups, which are as follows:

1. For example, the Snake Fisher model [18] and active contours [15–17] use shape-based segmentation, as does fixed circles as well as adaptive circle segmentation [14, 15–17]. Information may be gleaned by looking at the target shape's relativity.
2. To aid in the model-based segmentation process, three-dimensional spot modelling [20] and regularisation based on total variation (TV) [21] have been used. Modeling the picture as a function of several parameters is used to segment places in the image.
3. Using the seeded region growth (SRG) & watershed algorithms [22] [23] are used for region-based segmentation. Depending on the picture topology, spots are segmented by separating the image into regions.
4. Soft-thresholding [25] and threshold-based segmentation are examples of this kind of approach. Fluorescence intensities are used to distinguish between the foreground and background.
5. A mathematical morphological approach to segmentation [26–27]. It uses morphological processes and a variety of structural elements to segment locations.
6. Segmentation using NN and SVMs under the guidance of a trained expert On the basis of a pre-prepared training dataset, these algorithms separate the spots.
7. Three types of clustering methods: K-means, enhanced K-means, and Fuzzy C-means are only a few examples of unsupervised learning-based segmentation techniques. Segmenting spots with this technique relies on the inherent link between the intensities of the individual pixels.

The clustering-based approaches, on the other hand, are more susceptible to being influenced by noise than the other methods. The genuine microarray pictures, on the other hand, often exhibit low contrast, noise, artefacts, shape-variable spots and other problems of poor quality. All of these flaws combine to make clustering-based segmentation a difficult process to do successfully.

2. LITERATURE

By Karypis et al, the process of aggregating data points into significant subgroups (clusters) in a manner that increases the link between clusters and lowers the similarity between two separate groups is called clustering. Numerous clustering methods have been developed in a variety of fields. In contrast, the majority of conventional clustering algorithms use either numerical or categorical data as the basis for their clustering methods. It is also known as unsupervised learning and segmentation, wherein similar-looking items are grouped together. Clustering algorithms are used to get an understanding of a dataset's overall structure. When it comes to arranging data into groups, clustering is the primary goal. Statistical data analysis often uses the data clustering technique.

Osmar et al 2003 [22] propose its usage in a variety of domains, including pattern recognition, image analysis, and bioinformatics. Outliers, or data values that differ drastically starting the rest of the dataset, are common in datasets that are not regularly distributed. Data values that may be generated from probability density functions are known as Inliers. Furthermore, as Williams et al. 2002 points out [23], managing outliers is a challenging process since they don't fit the rest of the data items. Managing outliers in data mining is a difficult task. Since categorization relies on accurate data points, improving them is vital.

Following the development of an iterative approach, Bradley and Fayyad and et al. 1998 [24] utilized K-means clustering. Enhances Different sorts of numerical and categorical variables may be clustered using K-means clustering. K-means is an efficient and well-known clustering technique because of its high computing efficiency. A fitness function is constructed to cluster the data in genetic algorithms, which are derived from natural evolution. Fitness functions are used to apply crossover & survival of the fittest on the data. Clustering is simulated using an iterative simulation approach, although this is a computationally intensive process.

A clustering technique that can distinguish between outliers and non-outliers was developed by Jaing et al. 2001 [25], which combines the K-means approach with the Minimum Spanning Tree. Clustering data using a point symmetry distance has been proposed as an alternative to K-means. The computational burden is minimized by using a K-means clustering approach that is increased at each round of clustering. Because of this, this approach is often used in the clustering of huge datasets.

By using both local search and incrimination, Kuo et al. 2005 [21] were able to find suitable solutions. Outlier removal techniques were also suggested and created a method that concurrently eliminates both clusters and cluster data from the data set. Since it is straightforward and quick to implement, K-means clustering is really the most often used clustering approach. They have created two-approximation techniques that may be utilised to choose the first K locations for clustering in their study.

It is possible to discover outliers by using PAM clustering and then putting these individuals together in outlier clusters. According to Yinghua et al.2009 [27], these strategies perform better than K-means clustering because of the PSO algorithm's ability to find a global minimum. This approach is best suited for integrating features from datasets since it selects a minimal number of K beginning points.

A novel clustering approach, Affinity Propagation clustering, has been developed that may build clusters quickly. K-means clustering and affinity propagation characteristics that have low squared errors may be combined. Xuhui et al. 2009 [28] introduced the cluster evaluation technique, which aims to find the first cluster Center. Barakbah et al. have developed a novel method for maximizing the initial cluster centroids, which is critical to the success of clustering. The longest cumulative distance between the centroids is used to locate them in this method.

With the help of Alhadidi et al.2006 [29], the unpredictability in initial cluster selection may be eliminated. The user does not have to provide a value for the number of clusters at the beginning when using this approach. Using either Euclidean distance or Manhattan distance, a new K-means method has been devised that is more accurate.

Data may be divided into pieces using the median and then clustered using the approach proposed by Bozinov et al. 2006 [30]. Median-based clustering is superior than K-means clustering since it doesn't need a starting clustering number. This methodology was developed by Chen et al., which uses accuracy and discrepancy distance to identify outliers in both K-means and K-median clustering methods.

3. SINGLE FEATURE CLUSTERING ALGORITHMS

3.1 Algorithm K-means

It is an Unsupervised Learning technique that uses K-Means Clustering to partition a huge unlabeled dataset into a number of unique groups, referred to here as clusters. It's possible to produce two predefined clusters if K=2, three clusters if K=3, and so on. The constant K determines how many pre-defined clusters will be generated throughout the method. With this method, you don't need any further training or preparation to figure out what groups are in an unlabeled dataset.

In this approach, each cluster is linked with a centroid, and the algorithm is based on centroid-based data. Data points with their clusters should be as close together as feasible in order to get the smallest possible sum of distances. Input data is provided by an unlabeled dataset. When the approach is unable to find the best clusters, it stops the operation and moves on to the next k-clusters. The value of k must be known in advance with this procedure.

One of the key responsibilities of the k-means clustering approach is the creation of two distinct clusters.

- Optimization is a method for iteratively determining the exact value for each of K centroids.
- The nearest k-center is allocated to each data point. The data points that are most closely linked to a certain k-center create a cluster.

The following is a description of the clustering method with k-means [13] used for microarray image segmentation:

1. K initial clusters $\{C_1, C_2, \ldots\ldots, C_k\}$ from the image pixels $\{I_1, I_2, I_3, \ldots\ldots, I_{m*n}\}$ are randomly selected.
2. If the following condition is true, then each pixel should be assigned towards the cluster C_j $\{j=1,2,\ldots..K\}$

$$D(I_i, C_j) < D(I_i, C_q), q = 1, 2, ..., K$$
$$j \neq q \qquad (1)$$

D(.,.) specifies the dissimilarity metric in this instance.

3. In order to find a better cluster centre, perform these instructions:

$$C_i^{\wedge} = \frac{1}{n_i} \sum_{I_j \in C_i} I_j, i = 1, 2, ...K \qquad (2)$$

Where n_i is the number of pixels belonging to cluster C_i.

4. If $C_i^{\wedge} = C_i, i = 1, 2, ..K$ (3)

Then stop.

Else continue from step 2.

3.2 K-medoids Algorithm

The following is a description of the k-medoids clustering method [14] for segmentation of microarray images:

1. Randomly consider K initial medoids {M_1, M_2,……,M_k} for the clusters {C_1, C_2,……,C_k} from the m*n image pixels {I_1, I_2, I_3,……,I_{m*n}}.
   A mediod occupies a central position in a cluster. It is the cluster node with the lowest sum of distances to all other nodes.
2. Pixel assignment to clusters is done using the condition given by
   $$D(I_i, M_j) < D(I_i, M_q), q = 1, 2, ..., K$$
   $$j \neq q \qquad (1)$$
   D(.,.) specifies the dissimilarity metric in this instance.
3. Find out new medoids $M_i^{\wedge}$ belonging to clusters C_i , $i = 1, 2, ...K$. As the name suggests, it's made up of pixels that are as distinct to one another as possible.
4. If

$$M_i^{\wedge} = M_i, i = 1, 2, ..K \qquad (2)$$

Then stop.
Else continue from step 2.

3.3 K-modes Algorithm

Modes are employed instead of medoids in the clustering method with k-modes [15], which is comparable to the k-medoids algorithm [14]. In order to segment a microarray picture, the K-modes method is stated in the following way:

1. Consider randomly the initial K modes {MO_1, MO_2,……,MO_k} for the clusters {C_1, C_2,……,C_k} from the m*n image pixels {I_1, I_2, I_3,……,I_{m*n}}.

Among the many pixel values in a cluster, a mode is a value that is repeated more often than any other value.

The assignment of pixels to clusters is accomplished by the use of the condition provided by

$$D(I_i, MO_j) < D(I_i, MO_q), q = 1, 2, ..., K \quad j \neq q \tag{1}$$

D(.,.) specifies the dissimilarity metric in this instance.

2. Find out new medoids $MO_i^{\wedge}$ denoting to clusters C_i , $i = 1, 2, ...K$.
3. If

$$MO_i^{\wedge} = MO_i, i = 1, 2, ..K \tag{2}$$

Then stop.

If not, go on to step 2.

3.4 Algorithm Fuzzy c-means

According to the below given description, the clustering method Fuzzy c-means [16] for segmentation of microarray images is used:

1. K initial clusters $\{C_1, C_2, \ldots, C_k\}$ from the image pixels $\{I_1, I_2, I_3, \ldots, I_{m*n}\}$ are randomly selected.
2. u_{ij} is initialized with values from 0 to 1 and m=2 in the membership matrix. If a cluster's pixel memberships are summed together, the result should be 1.
3. Pixel assignment to clusters is done using the condition given by

$$u_{ij}^m D(I_i, C_j) < u_{iq}^m D(I_i, C_q), q = 1, 2, ..., K \quad j \neq q \tag{1}$$

D(.,.) specifies the dissimilarity metric in this instance.

4. Follow these directions to determine new members and cluster centroids:

$$u_{ik} = \frac{1}{\sum_{j=1}^{K} \left(\frac{D(C_i, I_k)}{D(C_j, I_k)}\right)^{\frac{1}{m-1}}}, \; for 1 \leq i \leq K$$

u_{ik} denotes the kth object in the ith cluster.

$$C_i^{\wedge} = \frac{\sum_{j=1}^{n} u_{ij}^m I_j}{\sum_{j=1}^{n} u_{ij}^m} \tag{2}$$

Where n is the number of pixels in the cluster Ci, which is the total number of pixels.

5. Repeat until each item has been allocated to the cluster with the most members [17].

3.5 Fuzzy K-medoids Algorithm Segmentation

The Fuzzy k-medoids [18] clustering technique is a variation of the fuzzy K-means approach, in which the means are replaced with medoids instead of k-medoids. The following is a description of the algorithm used for segmenting the microarray image:

1. Consider random K initial medoids $\{M_1, M_2,\ldots\ldots,M_k\}$ for the clusters $\{C_1, C_2,\ldots\ldots,C_k\}$ from the m*n image pixels $\{I_1, I_2, I_3,\ldots\ldots,I_{m*n}\}$.
2. Pixel assignment to clusters is done using the condition given by

$$u_{ij}^{m} D(I_i, M_j) < u_{iq}^{m} D(I_i, M_q), q = 1, 2, ..., K \quad j \neq q \qquad (1)$$

D(.,.) specifies the dissimilarity metric in this instance.

3. Find additional members and medoid values to cluster.

$$u_{ik} = \frac{1}{\sum_{j=1}^{K} \left(\frac{D(M_i, I_k)}{D(M_j, I_k)}\right)^{\frac{1}{m-1}}}, \; for 1 \leq i \leq K \qquad (2)$$

u_{ik} represents the ith cluster's kth object.

For each item, repeat steps 2-3 until it is allocated to a cluster with the most members.

4. MULTIPLE FEATURES CLUSTERING ALGORITHM

Unlike other image segmentation techniques, Microarray image segmentation relies only on information regarding pixel intensities when using clustering techniques. Microarray picture segmentation, on the other hand, requires that the pixel's position be known as well as the median value of the surrounding pixels have an impact on the outcome of clustering and, as a consequence, the result of segmentation. An method for segmentation of microarray pictures based on multiple feature clustering has been developed as a result of this discovery. First, we collect all of the pixels that are included inside the spot area and aggregate them into a dataset $D = x_1, x_2, x_3, x_4,...x_n$, where $xi = [x_i^{(1)}, x_i^{(2)}, x_i^{(3)}]$ represents the ith location in the spot area as a three-dimensional vector. Then, we apply clustering methods to this dataset. In order to do this, we use the following three characteristics:

$x_i^{(1)}$: The value of pixel intensity is represented by this value.

$x_i^{(2)}$: This number represents the distance in between pixel as well as the weighted centre of the spot region.

The following is the formula for calculating the spot centre:

1. Using the canny approach, find edges in the spot area picture using the canny method.
2. A flood fill may be achieved by using the imfill method.
3. An eight-by-eight matrix of labelled objects may be created using bw label function
4. The centroid of each named area (connected component) should be calculated using the region props approach.

A two-dimensional vector $p_i=[p_x, p_y]^t$, then $x_i^{(2)}= ||p_i-c||$ representing the ith pixel's coordinates is used to calculate xi(2), where the weighted centroid of the spot area is represented by c.

$$c = \frac{1}{\sum_{i=1}^{n} x_i^{(1)}} \sum_{i=1}^{n} x_i^{(1)} p_i \, . \qquad (1)$$

$x_i^{(3)}$: This value depicts the mean, median, or variation of the intensity of the pixels in the immediate vicinity. In this case, we have utilised the variance of the pixels in the surrounding area.

Once the characteristics for each pixel in the spot area have been acquired and combined to generate the dataset D, the clustering methods are performed to each pixel in the spot region.

5. RESULTS AND DISCUSSIONS

The suggested clustering technique produces the following outcomes when applied to two microarray pictures taken from the stand alone microarray database and corresponding to breast category aCGH tumour tissue: There are 38808 pixels in total on Image 1, whereas there are 64880 pixels on Image 2. As explained in [13], gridding is used for picture segmentation, with each compartment comprising just a single spot area and a single back-ground colour. Quantitative reasoning is another name for quantitative analysis. Quantitative analysis use numerical approaches and algorithms to determine algorithm performance without the usage of human error. Measures such as Mean Square Error (MSE) [19] are useful in evaluating the image's overall quality. What's being measured here is the variance in square error between the pixels in the original image and the images that come from the image. Mathematics may be used to express the MSE as

$$\text{MSE} = \frac{1}{N} \sum_{j=1}^{k} \sum_{i \in c_j} ||v_i - c_j||^2 \qquad (13)$$

xi signifies the pixel in the jth cluster of an image, and N depicts the total amount of pixels in the image. Because all of the data is concentrated in the picture's centre, the final image differs less from the original.

Image 1	Gridded Image

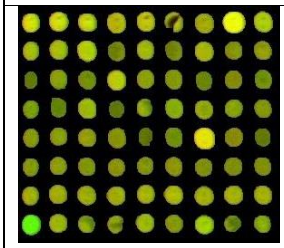
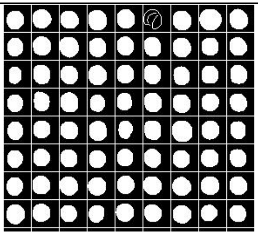

	

Image 2	Gridded Image

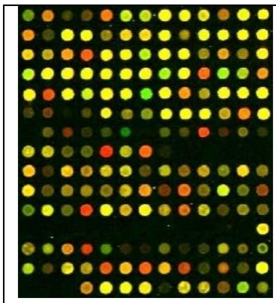
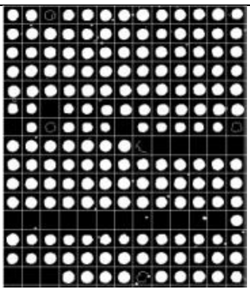
	

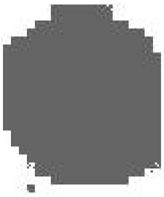
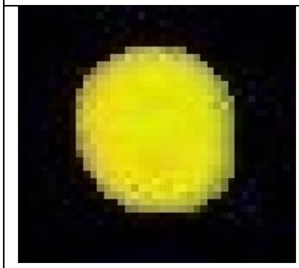
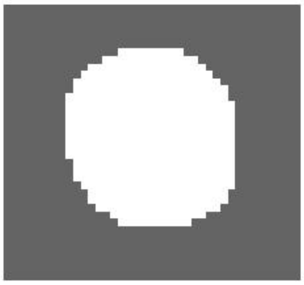
Image 1 of Compartment No 1	Segmentation with multiple feature fcm
	
Image 2 of Compartment No 8	Segmentation using multiple feature fcm
	

Figure 5.1 : Gridding and segmentation results

Figure 5.1 represents the outcome of the gridding algorithm. After dividing the picture into compartments in such a way that each and every compartment has a single spot and a single backdrop, compartment Number 1 from image 1 and compartment no 8 from image 2 are recovered from the grid. Multiple feature clustering methods are used to partition the image compartments into subsets. Figure 5.2 depicts the segmentation result obtained by the use of multiple feature fuzzy c-means.

Table 5.1: Comparison of MSE values

	Normal Clustering		**Multiple feature clustering**	
Method	**Compartment No 1**	**Compartment No 8**	**Compartment No 1**	**Compartment No 8**
K-means	96.2	95.8	**95.4**	**93.8**
K-medians	95.1	94.2	**94.2**	**92.2**
K-mode	95.2	94.3	**94.4**	**92.6**
Fuzzy c-means	90.7	90.2	**88.7**	**88.5**
Fuzzy K-medoids	91.8	91.4	**89.2**	**89.9**

The quantitative assessments of clustering techniques are shown in Table 5.1. For segmenting the microarray picture, the findings demonstrate that the multiple feature fuzzy c-means method delivers the lowest mean square error (MSE) value.

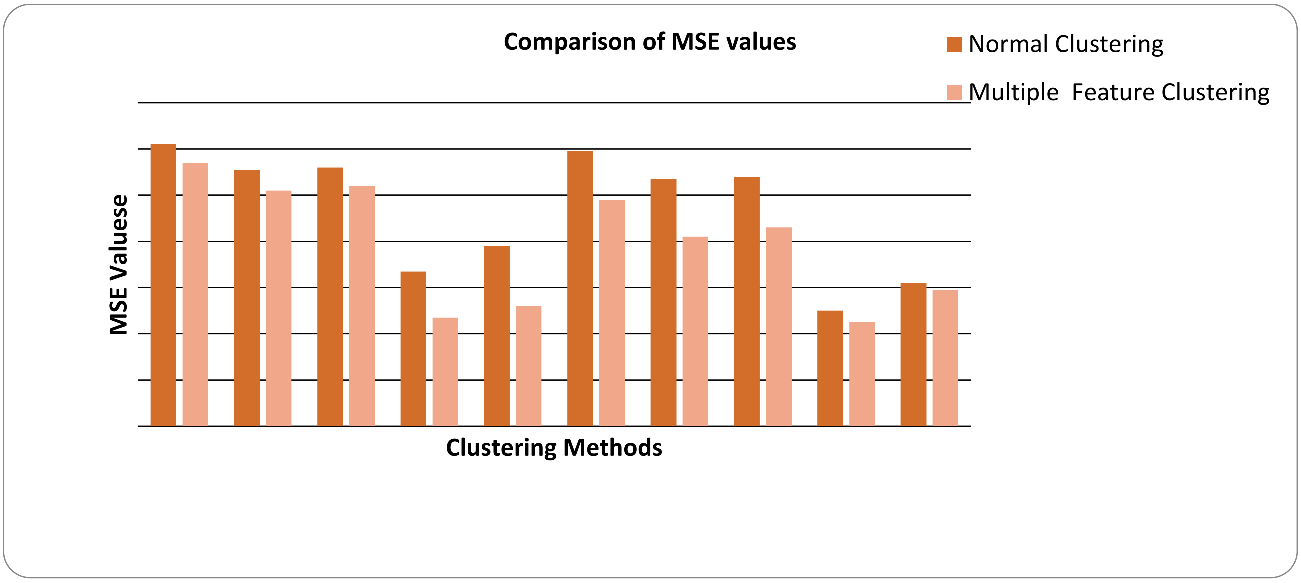


Figure 5.2: Comparative analysis of proposed method.

6. CONCLUSION

In real time, hundreds of gene expression levels may be monitored using microarray technology. Data extraction, segmentation, and gridding all fall under the umbrella term "microarray image analysis" (MIA). In this Paper multiple feature clustering techniques are employed to separate the microarray image. K-means clustering seems to be the basis for these methods. Only the intensity of the pixels has been taken into account in the development of these algorithms (one feature). Micro array image segmentation, however, is affected by not only the intensity of the pixels, but also their distance from the spot centre and the median intensity of a particular number of nearby pixels. This study presents a variety of feature clustering techniques based on these findings. The multiple feature fuzzy c-means clustering technique is the best of the bunch and delivers the lowest mean square error (MSE) value. The abundance of expression level of the relevant gene is determined by the logarithmic ratio of R/G, which is expressed as a percentage of R.

REFERENCES

[1] M.Schena, D.Shalon, Ronald W.davis and Patrick O.Brown, “Quantitative Monitoring of gene expression patterns with a complementary DNA microarray”, Science, 270,199, pp:467-470.
[2] Wei-Bang Chen, Chengcui Zhang and WenLin Liu, “An Automated Gridding and Segmentation method for cDNA Microarray Image Analysis”, 19th IEEE Symposium on Computer-Based Medical Systems.
[3] Tsung-Han Tsai Chein-Po Yang, WeiChiTsai, Pin-Hua Chen, “Error Reduction on Automatic Segmentation in Microarray Image”, IEEE 2007.

[4] J.Harikiran, et.al. “Vector Filtering Techniques for Impulse Noise Reduction with Application to Microarray images”, International Journal of Applied Engineering Research”, volume 10, Number 3, pp. 7181-7193, 2015.

[5] J.Harikiran, A.Raghu, Dr.P.V.Lakshmi, Dr.R.Kiran Kumar, “Edge Detection using Mathematical Morphology for Gridding of Microarray Image”, International Journal of Advanced Research in Computer Science, Volume 3, No 2, pp.172-176, April 2012.

[6] B.Sivalakshmi, N.Nagamalleswara rao,”Microarray Image Analysis Using Genetic Algorithm”, IAES Indonesian Journal of Electrical Engineering and Computer Science, Volume 4, No. 3, pp.561-567, 2016.

[7] J.Harikiran, Dr.P.V.Lakshmi, Dr.R.Kirankumar, “Automatic Gridding Method for Microarray Images”, Journal of Theoretical and Applied Information Technology, Volume 65, Number 1, pp.235-241, 2014.

[8] J.Harikiran, D.Ramakrishna, B.Avinash, Dr.P.V.Lakshmi, Dr.R.Kiran Kumar, “A New Method of Gridding for Spot Detection in Microarray Images”, Computer Engineering and Intelligent Systems, Volume 5, No 3, pp.25-33, 2014.

[9] M.Eisen, ScanAlyze User’s manual, 1999,

[10] J.Buhler, T.Ideker and D.Haynor, “Dapple:Improved Techniques for Finding spots on DMA Microarray Images”, Tech. Rep. UWTR 2000-08-05, University of Washington, 2000.

[11] R.Adams and L.Bischof, “Seeded Region Growing”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 16,no. 6, pp.641-647, 1994.

[12] D.RamaKrishna, J.Harikiran et.al.” “Various Versions of K-means Clustering Algorithm for Segmentation of Microarray Image”, International Journal of Electronics Communication and Computer Engineering, Volume 4, Issue 1, pp.1554-1558, 2012.

[13] J.Harikiran, Dr.P.V.Lakshmi, Dr.R.Kiran Kumar, “Fast Clustering Algorithms for Segmentation of Microarray Images”, International Journal of Scientific & Engineering Research, Volume 5, Issue 10, pp 569-574, 2014.

[14] M.Anirban, et.al. “Multiobjective Genetic Algorithm based Fuzzy Clustering of Categorical Attributes”, IEEE transactions on Evolutionary Computing, volume 13, number 5, pp.991-1005.

[15] J.Harikiran, P.V.Lakshmi,” Extensions to the K-means Algorithm for Segmentation of cDNA Microarray Images”, CSI Communications, December 2015

[16] Dr.R.Kiran Kumar et.al. “Dimensionality Reduction and Classification of Hyperspectral Images using Genetic Algorithm”, IAES Indonesian Journal of Electrical Engineering and Computer Science, Volume 3, No 3, pp.503-511, September 2016.

[17] J.Harikiran et.al. “Fuzzy C-means with Bi-dimensional empirical Mode decomposition for segmentation of Microarray Image”, International Journal of Computer Science Issues, volume 9, Issue 5, Number 3, pp.273-279, 2012.

[18] B.Saichandana et.al. “Image Fusion in Hyperspectral Image Classification using Genetic Algorithm”, IAES Indonesian Journal of Electrical Engineering and Computer Science, Volume 2, No. 3, June 2016, pp.703-711.

[19] B.Saichandana, et.al, “Clustering Algorithm combined with Hill climbing for Classification of Remote Sensing Image”, IAES International Journal of Electrical and Computer Engineering, volume 4, No 6, December 2014, pp.923-930.

[20] J. Lyons-Weiler, S. Patel, and S. Bhattacharya. A classification-based machine learning approach for the analysis of genome-wide expression data. Genome Research, 13(3):503– 512, 2003.

[21] H. H. Milioli, R. Vimieiro, C. Riveros, I. Tishchenko, R. Berretta, and P. Moscato. The discovery of novel biomarkers improves breast cancer intrinsic subtype prediction and reconciles the labels in the metabric data set. PLOS One, 10(7):e0129711, 2015.

[22] Osmar, R. and Zaïane. "Principles of Knowledge Discovery in Databases - Paper 8 Data Clustering" and ShantanuGodbole data mining Data mining Workshop. 2003.

[23] Williams, G., Baxter, R., He, H., Hawkins, S. and Gu, L. "A Comparative Study for RNN for Outlier Detection in Data Mining", In Proceedings of the 2nd IEEE International Conference on Data Mining, Maebashi City, Japan, pp.709, 2002.

[24] Bradley, P. S. and Fayyad, U.M. "Refining Initial Points for K-Means Clustering", ACM, Proceedings of the 15th International Conference on Machine Learning, pp. 91-99,1998.

[25] Jaing, M., Tseng, S. and Su, C. "Two-phase Clustering Process for Outlier Detection", Pattern Recognition Letters, Volume 22, pp: 691-700, 2001.

[26] Kuo, R. J., Liao, J. L. and Tu, C. "Integration of ART2 neural network and genetic Kmeans algorithm for analyzing Web browsing paths in electronic commerce", Decision Support Systems, Vol 40, pp: 355-374, 2005.

[27] Yinghua Zhou., Hong Yu. andXuemeiCa. "A Novel k-Means Algorithm for Clustering and Outlier Detection", Second International Conference on Future Information Technology and Management Engineering (FITME '09), 476–480, 2009.

[28] Xuhui Chen. and Yong Xu. "K-Means Clustering Algorithm with Refined Initial Center", 2nd International Conference on Biomedical Engineering and Informatics (BMEI '09), pp. 1–4, 2009.

[29] Alhadidi, H. N. Fakhouri, and O. S. Al Mousa, "cDNA Microarray Genome Images Processing Using Fixed Spot Position", American Journal of Applied Sciences volume :3, pp 1730-1734, 2006.

[30] Bozinov, D., &Rahnenführer, J. Unsupervised Technique for Robust Target Separation and Analysis of DNA Microarray Spots through Adaptive Pixel Clustering. Bioinformatics, 18(5), 747-756, 2002.

[31] M. T. Miller, A. K. Jerebko, J. D. Malley, and R. M. Summers, "Feature selection for computer-aided polyp detection using genetic algorithms," in Proc. of SPIE, Santa Clara, pp. 102-110, 2003.

[32] P. O'Neill and G.D. Magoulas, "Improved processing of microarray data using image reconstruction techniques", IEEE Transactions on Nano bioscience. Volume 2, pp 176–183, 2003.