



An Improved Dual Clustering Method for Classification of Microarray Image Segmentation

D. Ravi Kumar, V. Devi, M. Balasaraswathi, B. Karthik

Abstract Microarray is a fast and rapid growing technology which plays dynamic role in the medical field. It is an advanced than MRI (Magnetic Resonance Imaging) and CT scanning (Computerised Tomography). The purpose of this work is to make fine perfection against the gene expression. In this study the two clustering are used which fuzzy c means and k means and also it classifies with better results. The microarray data base indicates the classification in support vector machine. Segmentation is most important step in microarray image. The classification in support vector machine is compared with other two classifiers which means the k nearest neighbour and with the Bayes classifiers.

Keywords: Segmentation, Microarray image, classification (SVM – Support Vector Machine), dual clustering.

I. INTRODUCTION

A. Microarray

A high throughput technology that allows detection of thousands of genes simultaneously. Principle: basepairing hybridization. Much rely on computer aids Central platform for functional genomics DNA microarray technology may be defined as a high throughput and adaptable technology used for parallel gene expression analysis for thousands of genes of known and unknown functions. It's used for detection of polymorphisms and mutations in genomic DNA. ADNA microarray is a gathering of infinitesimal DNA spots on strong surface. Each spot contains picomoles of a particular DNA arrangement, known as tests or columnists. DNAmicroarray are strong backings, ordinarily of glass or silicon, upon which DNA is attached in an organised grid fashion. Each spot of DNA, called a probe, represents a single gene. There are several synonyms of DNA microarrays such as DNA chips, gene chips, DNA arrays, gene arrays and biochips.

The thousands of genes in a single experiment is allowed with the simultaneous measurement by using this technology. One important task involved in the analysis of cDNA microarray images is the spots, addressing and gridding, can be divided into three main types like manual, semi- automated, and automated. The other important task in the analysis of cDNA microarray images is a microarray image segmentation process, which characterizes the pixels into background and foreground pixels.

B. Segmentation

The reason for picture division is to segment a picture into important locales regarding a specific application.

The division depends on estimations taken from the picture and may be dark level, shading, surface, profundity or movement. Typically picture division is an underlying and essential advance in a progression of process went for general picture understanding. Utilizations of picture division incorporate distinguishing objects in a scene for question-based estimations, for example, size and shape. It also identifies objects in a moving scene for object-based video compression. Its also identifies the objects which are at different distances from a sensor using depth measurements from a laser range finder enabling path planning for a mobile robot.

C. Clustering

A group is in this way a gathering of articles which are "comparable" amongst them and are "divergent" to the items having a place with different bunches. Various things of the same kind, growing or held together, also refers to a group of things or persons close together. Clustering can be considered the most important unsupervised learning problem. so, as every other problem of this kind, it deals with finding a structure in a collection of unlabelled data. A main definition of clustering could be "the process of organizing objects into groups whose members are similar in some way". Here the objects are grouped according to their fit to descriptive concepts, not according to simple similarity measures. The goal of clustering is to determine the intrinsic grouping in s set of unlabelled data. The main requirements that a clustering algorithm should satisfy are its scalability, should able to deal with noise and outliers, high dimensionality etc., In this paper two of the most used clustering algorithms they are,

- K-means
- Fuzzy c-means

Therefore, clustering here is a grouping of an objects into a more similar and the objects in the same group which is called as a cluster.

Manuscript published on November 30, 2019.

* Correspondence Author

Dr. D. Ravi Kumar*, Head, Department of Electronics and Communication Engineering, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India.

Dr. V. Devi, Head, Department of Computer Applications, Gurunanak College, Chennai, India.

Dr. M. Balasaraswathi, Associate Professor, Department of Electronics & Communication Engineering, Saveetha School of Engineering, SMITS, Chennai, India.

Dr. B. Karthik, Associate Professor, Department of Electronics and Communication Engineering, Bharath Institute of Higher Education & Research, Chennai, India.

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

D. Classification

Classification is created on the SVM (Support Vector Machines) this classifier is further compared and experimented with the other two classifiers namely KNN (K Nearest Neighbours) and Bayes. In machine learning, support vector machines which also support vector networks are the supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. When data are not labelled, supervised learning is not possible, and an unsupervised learning approaches is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. More exactly a support vector machine which constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space which can be used for classification or other tasks like outlier's detection. Therefore, these three classifiers are compared and experimented for further analysis.

II. RELATED WORKS

A. Literature Survey

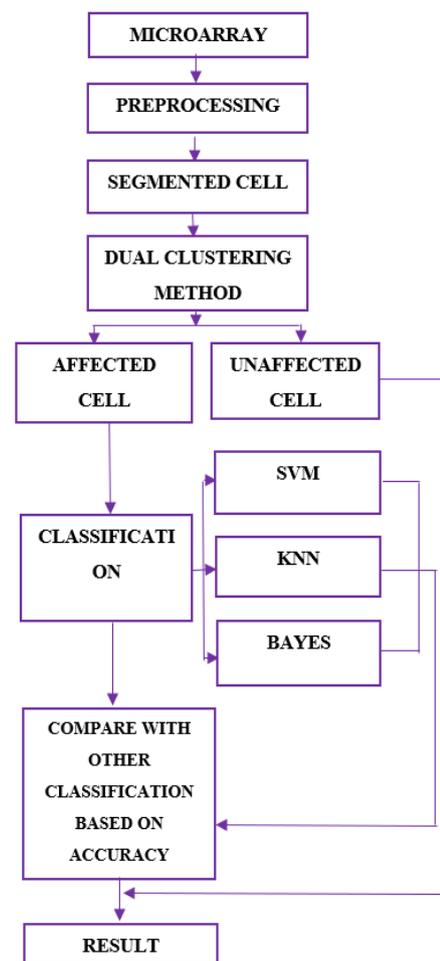
This proposes an acquisition system for microarray image on a FPGA base platform, together with a hardware implantation of image segmentation for cDNA microarray images [1]. The foremost works of clustering-based segmentation have shown that k-means clustering algorithm and also moving of k-means algorithm are the two commonly used methods in microarray image processing. To get the segmentation accuracy, we combine clustering-based segmentation which may be more reliable and also able to segment its spots spontaneously. Finally, to replace its false segmentation and the inseparable ones of missing spots are done with the help of refinement step has been used [2]. One of the key issues in the experimental approaches that utilize microarrays is to extract quantitative information from the spots, which represents genes in a given experiments. For this process, the initial stages are quite important, they have identified the spots and separating background from the foreground. And also, they presented an overview of the state-of-the art methods for image segmentation. They discussed about histogram – based methods, and recently introduced clustering-based techniques [3]. In this paper, a set of experiments conducted to obtain results for several popular image coding techniques, by including the most recent coding standards. They used prediction – based schemes CALIC and JPEG-LS and JPEG2000 using zero wavelet decomposition levels are the best performing standards compressors but are all outperformed by the best microarray specific technique, CNN-based schemes [4]. Microarray image analysis spot segmentation is one of the important heavy tasks. A great solution to this problem is to provide an algorithm which can be used to find its spots within the microarray image. Therefore, in the image analysis, Circular Hough Transformation method is used and also for computer vision and digital image processing. The algorithm is applied on

the cDNA microarray images which is to improve its exact accuracy and the effectiveness of the spots area, addressing and segmentation process. Therefore, the main purpose of Applying this method is to find the imperfect instances of spots within a certain class of circles [5].

III. MATERIALS AND METHODS

3.1 Microarray image

A microarray which is defined as an array ordered of microscopic elements on a planar substrate that allows the specific gene products and its binding of genes. The double standard DNA results of length 500 to 2500 base pairs which is printed in the substrate for the spotted cDNA microarrays. Therefore, the printed spots which have their diameter in some ranges.



Flowchart of Classification In Microarray Image Segmentation Using Dual Clustering Method

Some of the steps in the microarray is labelling and the sample preparations, then next process is the hybridisation, washing, image processing, biological interpretation, data normalization and data analysis. Other than this some of the methods involved in the microarray image processing are,

- **Gridding method**
- **Segmenting method and**
- **Spot extraction method**

3.2 Gridding

Gridding is the method in the microarray image processing that is used to detect the position of its spots centres and identifies their important coordinates in that image. In the gridding the image begins with the background disposition with the help of pre-processing of an image. In this most of the microarray images the noise appears in that is because of the slide emission.

Then the entire image is affected. when compared to the signal pixels the number of the background pixels is larger. In an image despite locating of an object using the template matching we perform an initial segmentation which is equal to a spot.

3.3 Segmentation

Image segmentation is another process which is involved in the microarray image processing. Segmentation which is usually based on greyscale. It is very simple model of grey scale which leads to error in object labelling. It is also based on the texture which enables the surfaces objects with contrast figure of grey to be segmented. It is also based on motion. The segmentation must be based on the evaluate and not in general, the true flow. The two simple image segmenting techniques that are based on the grey level histogram of an image

- **Thresholding**
- **Clustering**

3.3.1 Thresholding

One of the most important methods used in the image segmenting also its useful in differential foreground from the background. By selecting the threshold vale T, the image is converted into the binary image from the grey level image.

3.3.2 Clustering

Clustering which is defined as the process of identifying its groups of similar images low. Based on the credit this process organizes its objects. Cluster quality measures are having a high correspondence with the cluster also has low across clusters.

3.4 Spot Intensity extraction

An image is constructed by standard, the first n images its checks the first n frames then it is used for the spot finding. This similar image which is used to calculate the background and the exact spot is calculated in an image and its accuracy is measured.

3.5 Dual Clustering

Due to the various applications spatial clustering has fascinate a lot of research attention. To correct this, we explore and propose a new clustering on two domains in this paper which is called as the dual clustering. dual clustering is means k-means clustering and the fuzzy c means clustering. Therefore, only clustering which is divided into k and fuzzy, where k and fuzzy which combine together which is called as the dual clustering.

3.5.1 K-Means clustering

Its main advantages in this is its simple, understandable and its items spontaneous assigned to clusters. Where the k-means is based on the colour biases. In this if the number of

the clustering is given it will exhibit the exact number of classification according to the clustering.

3.5.2 Fuzzy c-means clustering

In fuzzy c means bunching, each point has a level of having a place with groups, as in fuzzy c means rationale, instead of having a place totally with only one group. Along these lines, focuses on the edge of a group, might be in the bunch to a lesser degree than focuses in the focal point of group. A review and correlation of various fuzzy c means grouping calculation is accessible. Parameters of the FCM algorithm must be specified. they are the number of clusters c and the fuzziness exponent, m, where m is any real number greater than 1. In the fuzzy c means it permits information focuses to be relegated into in excess of one group. Every datum point has a level of participation of having a place with each group.

3.6 Support Vector Machine classification

SVM are learning systems that uses a hypothesis space of linear functions. In a high dimensional feature space -kernel function. It trained with a learning algorithm from optimization theory. Support vectors are data points that lie closes to the decision surface. But they are difficult to classify as well. They also have a direct bearing of optimum location on the surface. Its features of SVM are do not suffer from the curse of dimensionality and do not affected by the local minima. They are various properties of the SVM solution help avoid over fitting, even in every high dimensional feature spaces.

3.6.1 k-Nearest Neighbour classification

For a neighbouring classis a new instance is classified by a majority votes. It is measured by a distant function where its object is assigned to the common class among its k nearest neighbours. Numerical pattern is given to the each of the instances is p represented with a set. All the instance corresponds to points in an n-dimensional feature space. Feature vectors of different k nearest points is compared and classification is done. Its strengths are its very simple when compared to the others and from the any distribution it can be applied to the data. And it's a good classification. A major weakness in this is it takes the more time to classify and give a new example. Here choosing the value k may be tricky and for its accuracy it needs many samples.

3.6.2 Bayes classifiers

In machine learning, Bayes classifiers are a family of simple classifier which is based on applying Bayes theorem with effective self-dependence as self-sufficiency between the features. Its main advantage is computational efficiency.

IV. MICROARRAY IMAGE DATABASE

Microarray image database taken from NIH (National Institute of Health) and NIC (National Institute of Cancer). It has extracted to pixels by pixels information.



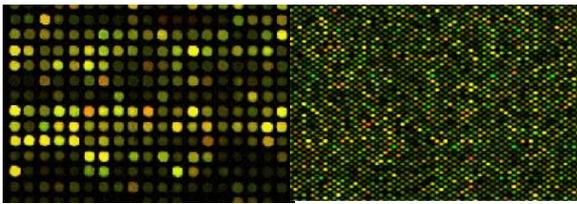


Fig:1 Bladder cancer

Fig:2 Blood cancer

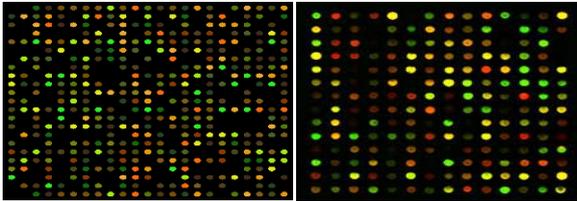


Fig:3 Skin cancer

Fig:4 Lung cancer



Fig:5 Blood cancer

Each spot is generated by the binary map. Signal pixels where pixels inside the circle is characterized and the outside circle of the pixels is as background.

V. EXPERIMENT AND RESULT

Matlab is used to create and applied the advance technique.

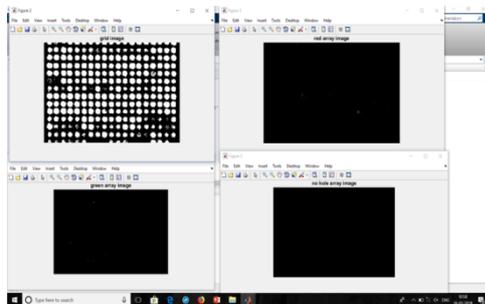


fig:6.(a) gridding image of bladder cancer
6.(b) red array image of bladder cancer
6.(c) green array image of bladder cancer
6.(d) no hole array image of bladder cancer

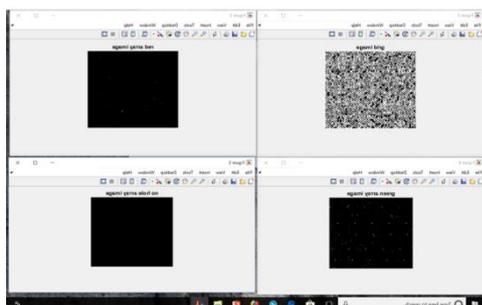


fig:7.(a) gridding image of blood cancer
7.(b) red array image of blood cancer
7.(c) green array image of blood cancer
7.(d) no hole array image of blood cancer

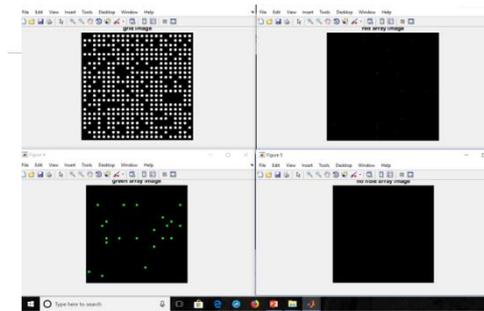


fig:8.(a) gridding image of skin cancer
8.(b) red array image of skin cancer
8.(c) green array image of skin cancer
8.(d) no hole array image of skin cancer

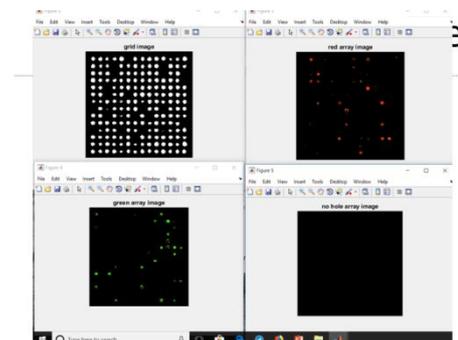


fig :9(a). gridding image of lung cancer
9(b).red array image of lung cancer
9(c).green array image of lung cancer
9(d). no hole array image of lung cancer

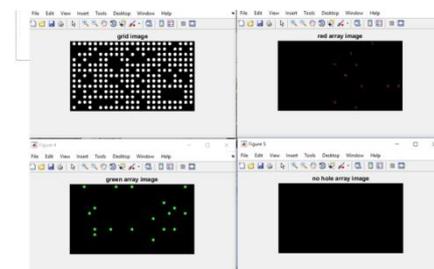


Fig :10.(a) grdding image of blood cancer
10.(b) red array image of blood cancer
10.(c) green array image of blood cancer
10.(d) no hole array image of blood cancer

The microarray images used in the experiments is described below table. Here, the foreground is perfect and therefore background is separated from it for all the important spots.

VI. DISCUSSION

Biological discovery are new and better molecular diagnostics. The new molecular targets for therapy finding and refining biological pathways. Mutation and polymorphism detection. Recent examples of molecular diagnosis of leukemia, breast cancer and so on. The appropriate treatment for genetic signature potential new drug targets. Cross-hybridization of sequences with high identity.



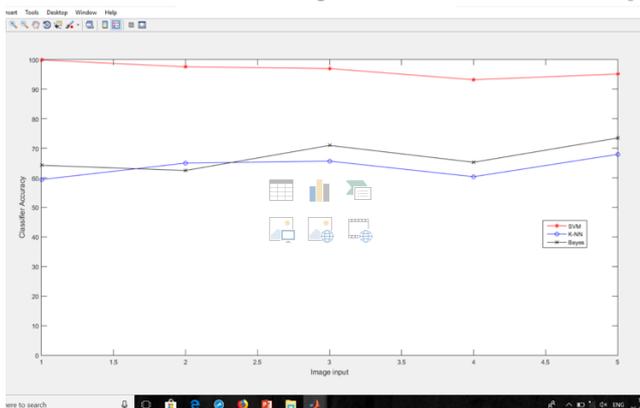
Chip to chip variation. Expensive not every lab can afford experiment repeat. Microarray image database taken from NIH (National Institute of Health) and NIC (National Institute of Cancer). It has extracted to pixels by pixels information.

6.1 Comparison of SVM, KNN and BAYES classifiers based on accuracy using Dual Clustering

Table .1 Comparison of SVM, KNN and BAYES classifiers based on accuracy

| | SVM | KNN | BAYES | DUAL CLUSTERING (affected cell) |
|----------------|-------|-------|-------|---------------------------------|
| Bladder cancer | 99.87 | 59.42 | 64.24 | 74 |
| Blood cancer | 97.57 | 64.99 | 62.47 | 33 |
| Skin cancer | 96.98 | 65.65 | 70.97 | 415 |
| Lung cancer | 93.19 | 60.37 | 65.26 | 555 |
| Blood cancer | 95.13 | 67.95 | 73.46 | 128 |

6.2 Comparison of classification



The microarray data base indicates the classification in support vector machine. Segmentation is most important step in microarray image. The classification in support vector machine is compared with other two classifiers which means the k nearest neighbour and with the Bayes classifiers.

VII. CONCLUSION

In this broadside dual grouping method is been used to make minute qualification against the quality articulations in a microarray picture preparing. The bunching strategies, for

example, Fuzzy c-means and k-means. The portioned image and calculated values have been obtained and compared. It is concluded that fuzzy c-means is more reliable than the k-means has guaranteed a touchy grouping when thought about to k-means. This has concluded to more accurate calcification of the week spots.

REFERENCES

1. BogdanBelean ,Albert Fazakas, Raul Malutan , Monica Borda “ Hardware implementation of microarray image segmentation” communication Department , IEEE International Conference On Automation , Quality And Testing, Robotics Vol 3 pp:22-25, May 2008.
2. Guifang Shao, TiejunLi ,WangdaZuo, Shunxiang Wu, Tundong Liu, “ A Combinational Clustering Based Method for cDNA Microarray Image Segmentation ,August 4,2015.
3. Li Qin, Luis Rueda, Ali Adnan, AliouneNgom, University of Windsor “Spot Detection and Image Segmentation in DNA Microarray Data-2005.
4. Miguel Hernandez Cabronero, Ian Blanes , Joan Serra-Sagrsta , Michael W.Marcellin, Department of Information and Communication Engineering, “A review of DNA microarray image compression”,IEEE-2011.
5. R.M.Farouk and M.A.SayedElahl, “Robust cDNA microarray image segmentation and analysis technique based on Hough circle transform”, Department of Mathematics, Faculty of Science, Zagazig University, Egypt
6. R.Lukac,K.N. Platanniotis ,B.Smolka, and A.N Venetsanopoulos, “cDNA microarray image processing using fuzzy vector filtering framework,” fuzzy Sets and Systems,vol. 152(1),pp.17-35,2005.
7. N. Giannakeas, P.S. Karvelis and D.I. Fotiadis, “ A Classification – Based Segmentation of cDNA microarray images using Support Vector Machines,30th Annual International IEEE EMBS Conference,Canada(2008).
8. N.Giannakeas, and D.I. Fotiadis, “Multichannel Segmentation of cDNA Microarray Images using the Bayes Classifiers,” 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,2007.
9. M. Nykter, T. Aho, M.Ahdesmaki, P.Ruusuvuori, A.Lehmussola and O. Yli-Harja, “Simulation of microarray data with realistic characteristics,” BMC Bioinformatics, vol.7, 2006.
10. D.Ravi Kumar, Dr.ArunRaaza, Dr.V.Devi,” Computational Analysis of microarray image” in journal of advanced research in dynamical and control systems, Vol:09, Issue 5, ppt 67-75 July 2017.

AUTHORS PROFILE



Dr. D.Ravikumar. has received the Ph.D degree in Electronics & Communication Engineering from Vels Institute of Science, Technology &Advanced Studies (VISTAS), Chennai, India. Currently he is working as a Head in the Electronics & Communication Engineering, Department at VISTAS. His main research Interest includes Embedded systems, Image and Signal processing, Linear and Digital Integrated systems, Computer communications and Applied Electronics. His 8 research articles have been anthologized widely.



Dr. V.Devi, has received her UG Degree, PG Degree and PhD Degree in the Department of Computer Applications from Madras University. Currently she is working as a Professor and Head in the Department of Computer Applications At Gurunanak College, Chennai , India. Her main research Interest and specialization includes Image and Signal processing, Computer communication and Networks, Problem solving & optimization methods in Computing systems. In addition, she published several research papers in those areas.



An Improved Dual Clustering Method for Classification of Microarray Image Segmentation



Dr.M.Balasaraswathi, has obtained the Ph.D degree in Electronics&Communication Engineering from Anna University, Chennai, India. Currently she is working as a Professor in the Department of Electronics &Communication Engineering at Saveetha School of Engineering, Chennai. Her main research Interest includes Antenna and Radiation systems, Signal processing and Applied Electronics.

Dr.B.Karthik, has obtained his Ph.D degree in Electronics and Communication Engineering from Bharath Institute of Higher Education and Research (BIHER), Chennai, India. Currently he is working as a Professor in the Department of Electronics and Communication Engineering and Research Coordinator at BIHER Chennai. Member in IEEE. His specialization includes VLSI systems, Medical Image processing, Wireless communication systems and networks. He has published several research papers.