



INTERNATIONAL JOURNAL OF  
RESEARCH IN COMPUTER  
APPLICATIONS AND ROBOTICS

ISSN 2320-7345

# A FICTION WORK ON FUZZY BASED CANCER SEQUENCE IDENTIFICATION THROUGH CLUSTER ALGORITHM

A.Maivizhi<sup>1</sup>, Mrs.C.Kalaiselvi<sup>2</sup>

<sup>1</sup>M.Phil Research Scholar, Tiruppur Kumaran College for Women, Tiruppur.

<sup>2</sup>Associate Professor & Head, Dept. of computer Application,

Tiruppur Kumaran College for Women, Tiruppur.

Maivizhi.baby@gmail.com, kalaic29@gmail.com

---

**Abstract:** - Cancer malady prediction is one in all the rising analysis areas. Several algorithms like K-Means, Kernel K-Means and Fuzzy C-Means area unit wont to notice the cancer affected genes within the sample dataset DLBCL, MLL, SRBTCM, and EWS. The top of specified algorithms is finding the cancer genes with additional accuracy. So, the changed Fuzzy C-Means algorithmic program is planned to understand the cancer genes. This paper is planned the cancer identification ways and evaluated them supported the computation time, classification accuracy and skill to reveal biologically meaning factor info. This paper highlights the cancer factor identification victimization the changed Fuzzy C-Means algorithmic program that is used expeditiously during this work and examined with higher results.

**Keywords:** Clustering, Fuzzy C-Means, Kernel.

---

## Introduction

Cancer natural phenomenon information is used to contain the keys for locating the foremost problems concerning cancer designation. The recent advent of Desoxyribo nucleic acid microarray technique may be a heap of powerful that is used to create continuous observation of thousands of issue expressions. With this vast quantity of natural phenomenon information, researchers began to look the chances of cancer detection victimization natural phenomenon information. There area unit several methods area unit projected in recent years with good results. But there area unit still many issues that need to be clearly explained.

In this paper, modified Fuzzy C-Means rule is used to hunt out the cancer affected genes among the sample dataset. Whereas examination the algorithms modified Fuzzy C-Means, Kernel based clump, the modified Fuzzy C-Means is healthier. The modified Fuzzy C-Means attains the deserves of it slow concern and correct issue identification rule. The performance of these algorithms are also evaluated victimization the results of the specified algorithms.

## Problem Definition

Two major issues connected the unattended analysis of organic phenomenon knowledge square measure diagrammatic by the accuracy and reliableness of the discovered clusters, and by the biological indisputable fact that the boundaries between categories of patients or categories of functionally connected genes square measure typically not clearly outlined. The first goal of this work consists the method concerned within the exploration of recent methods and within the development of recent agglomeration ways that is employed to enhance the accuracy and strength of agglomeration results. A cancer identification system is developed for identifying attainable sequence markers and applying the changed kernel primarily based fuzzy agglomeration formula to spot the illness.

## Contributions

The main contributions of the thesis are as follows. Investigation of recent ways and development of recent clump ways are accustomed improve the accuracy and strength of clump results. The planned formula is functioning proficiently once scrutiny the algorithms like Fuzzy -Means, Kernel primarily based Fuzzy clump and changed Kernel primarily based Fuzzy clump formula. The time concern and accuracy also are evaluated. Hence, the performance criteria also are relatively high. FUZZY LOGIC symbolic logic is one in every of the rising technique that's wide used. The modeling of inaccurate and qualitative data, further as handling of uncertainty at numerous stages is potential through the utilization of fuzzy sets. Symbolic logic is logic of fuzzy sets; a Fuzzy set has associate degree endless vary of truth values between one and 0. Propositions in symbolic logic have a degree of truth, and membership in fuzzy sets may be totally comprehensive, totally exclusive, or a point in between. FUZZY clump Addition of symbolic logic with data processing techniques has become one in every of the key constituents of soppo computing in handling the challenges posed by vast collections of natural information. The essential plan in fuzzy clump is that the non-unique partitioning of the info in a very assortment of clusters. The info points are assigned the membership values for every of the clusters. The fuzzy clump formula is employed to permit the clusters to grow into their natural shapes.

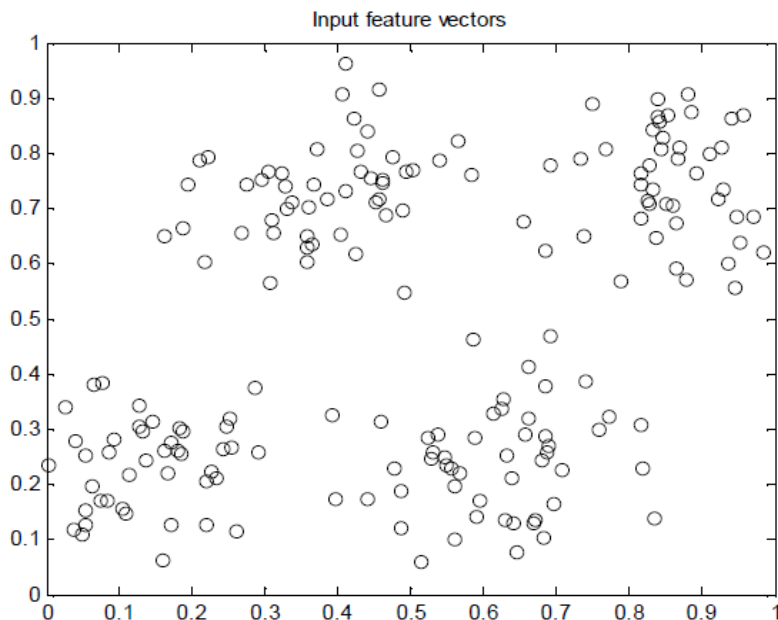
## Fuzzy C-Means

Fuzzy bunch may be a category of algorithms for cluster analysis within which the allocation of information points to clusters isn't hard (all-or-nothing) however fuzzy within the same sense as symbolic logic. Bunch may be a mathematical tool that makes an attempt to find structures or bound patterns in an exceedingly information set, wherever the objects within every cluster show a definite degree of similarity. Fuzzy bunch permits every feature vector to belong to over one cluster with completely different membership degrees (between zero and 1) and unclear or fuzzy boundaries between clusters.

## Algorithm

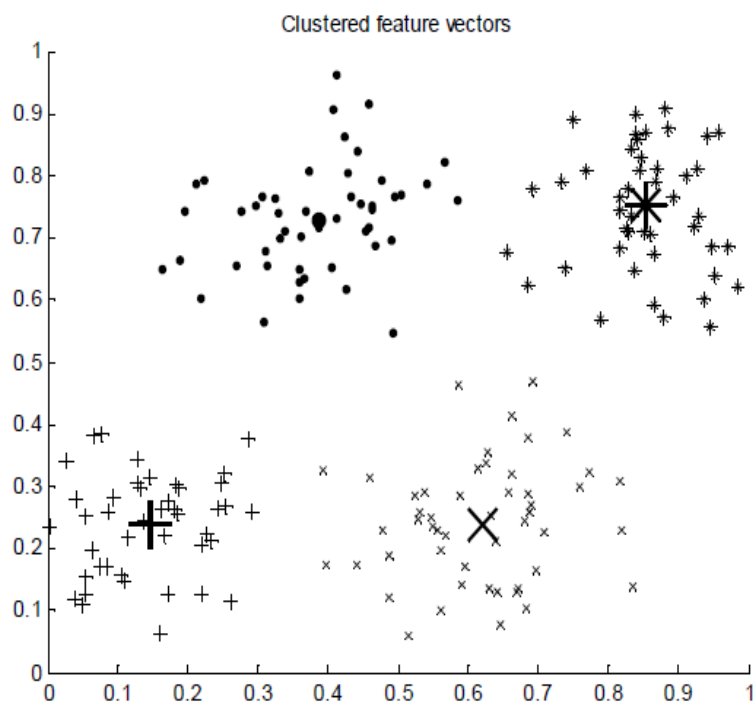
The FCM algorithm is implemented using the following steps: Let  $X = \{x_1, x_2, x_3 \dots, x_n\}$  be the set of data points and  $V = \{v_1, v_2, v_3 \dots, v_c\}$  be the set of centers.

- 1) Randomly select 'c' cluster centers.
- 2) calculate the fuzzy membership ' $\mu_{ij}$ ' using equation.
- 3) compute the fuzzy centers ' $v_j$ ' using



4) Repeat step 2) and 3) until the minimum ' $J$ ' value is achieved or  $\|U(k+1) - U(k)\| < \beta$ . where,  $_k^c$  is the iteration step.  $_b^c$  is the termination criterion between  $[0, 1]$ .  $_U = (\mu_{ij})_{n \times c}$  is the fuzzy membership matrix.  $_J^c$  is the objective function,  $_m^c$  is a fuzzifier;  $m > 1$

Conditions given in equations above require that the total membership of one data feature vector is normalized to 1 and that element cannot belong to more clusters then it exists. The fuzzifier,  $m$ , should be more then 1, since when  $m = 1$ , the optimization problem becomes a crisp case. In the literature,  $m = 2$  is the mostly used value for the fuzzifier.



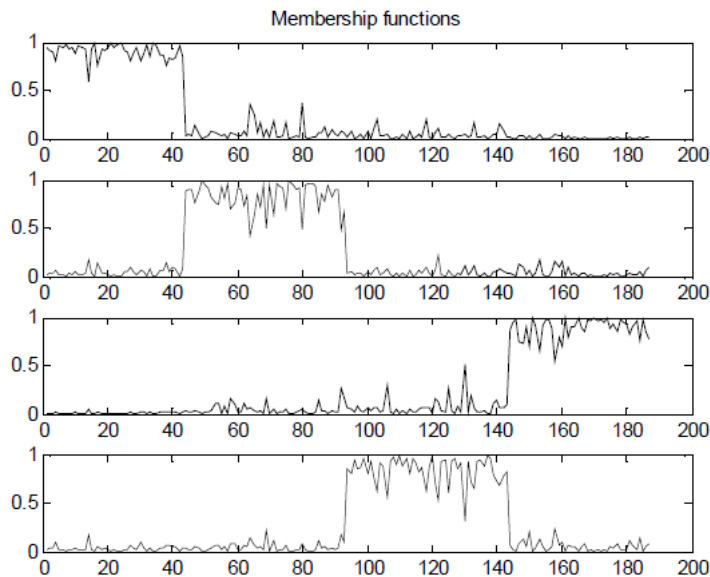


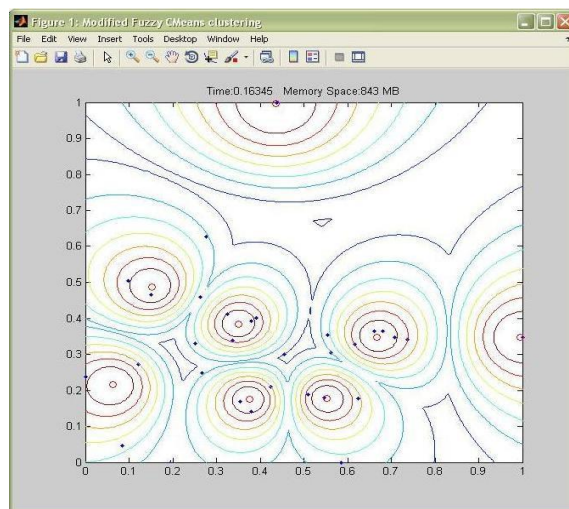
Fig.3.5 Membership Functions obtained by FCM

Thus the FCM algorithm yields the clustered image based on the number of clusters used. The abnormality region of the input image is grouped into one particular cluster which can be easily extracted.

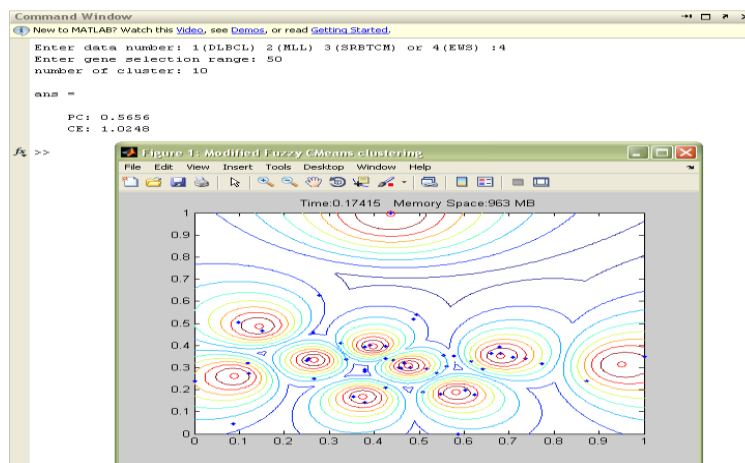
### Experimental Results:

The following results are happened with modified fuzzy c-means algorithm.

The MLL dataset is employed with thirty coaching sets and range of cluster is eight. The resultant cluster read is shown within the Fig.5.9. The earned the worth of laptop is zero.6428 and metallic element is zero.8100.



Using changed fuzzy c-means algorithmic rule, dataset taken for analysis is DLBCL, the coaching set is twenty and variety of clusters is three then the time taken for analysis zero.15686. The time taken for analysis in k-means is zero.633 times over fuzzy c-means.



DATA SET	GENE SELECTION RANGE	NO.OF CLUSTERS	TIME (SEC)	PC	CE
DLBCL	20	3	0.16422	0.7357	0.4661
		5	0.15686	0.6583	0.6799
	25	3	0.1531	0.7581	0.4503
		30	3	0.16858	0.7342
MLL	20	3	0.16116	0.7330	0.4915
		5	0.15759	0.6888	0.6185
	25	5	0.18612	0.6785	0.6473
		7	0.19544	0.6612	0.7419
SRBTCM	20	3	0.17268	0.6837	0.5655
		5	0.16916	0.6484	0.7222
	25	4	0.17164	0.6688	0.6526
		6	0.16094	0.6583	0.7411
EWS	30	3	0.16198	0.6346	0.6362
		5	0.1829	0.5512	0.896
	40	6	0.17642	0.6056	0.8417
		8	0.16982	0.6428	0.8100
40	7	0.1712	0.5757	0.9131	
	9	0.16805	0.5966	0.9286	

## Overall Performance

ALGORITHMS	PARAMETERS		
	TIME	ACCURACY	PERFORMANCE
K - means	2.889368	7.69	91.52533
Kernel Based Fuzzy Clustering	4.43915	8.58	94.186
Fuzzy C – Means	2.205545	6.65	93.36081
Modified Fuzzy C- Means	0.159944	9.59	95.95383

## Conclusion

In this paper, varied algorithms square measure accustomed realize the cancer affected genes within the sample dataset. That square measure K-Means, Kernel primarily based fuzzy cluster, fuzzy c-means and changed fuzzy c-means. The sample dataset that is employed for this analysis work square measure DLBCL, MLL, SRBTCM, and EWS. the required algorithms are well functioned, in a very cancer genes. So, the changed fuzzy c-means rule is projected to understand the cancer genes.

## REFERENCES

- [1] F. Chu and L. Wang, "Applications of Support Vector Machines to Cancer Classification with Microarray Data", International Journal of Neutral Systems. Vol.15.No. 6, 475-484, 2005.
- [2] Wang, Y., Tetko, I. -V., Hall, and M. -A., Frank, E., Facius, A., Mayer, K. -F., And Mewes H. -W., "Gene Selection From Microarray Data For Cancer Classification —A Machine Learning Approach", Comput Biol Chem, 29 (1): 37-46, 2005.
- [3] Huilin Xiong And Xue-Wen Chen,"Optimized Kernel Machines For Cancer Classification Using Gene Expression Data", Proceedings Of The 2005 IEEE Symposium On Computational Intelligence In Bioinformatics And Computational Biology, Pp.1-7, 2005.
- [4] L. Shen And E.C. Tan, "Dimension Reduction-Based Penalized Logistic Regression For Cancer Classification Using Microarray Data," IEEE/ACM Trans. Computational Biology And Bioinformatics, Vol. 2, No. 2, Pp. 166-175, Apr.-June 2005..
- [5] A. J. Gentles, S. K. Plevritis, R. Majeti, and A. A. Alizadeh, " Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia," J. Amer, Med, Assoc., vol. 304, pp. 2706-2715, 2010.
- [6] Mallika Rangasamy And Saravanan Venketraman,"An Efficient Statistical Model Based Classification Algorithm For Classifying Cancer Gene Expression Data With Minimal Gene Subsets", International Journal Of Cyber Society And Education, Vol. 2, No. 2, Pp.51-66, 2009.
- [7] Xin Jin, Anbang Xu, Rongfang Bie and Ping Guo, "Machine Learning Techniques and Chi-Square Feature Selection for Cancer Classification Using SAGE Gene Expression Profiles", Springer link, Data Mining for Biomedical Applications, 2006.

- [8] Kai-Lin Tang, Wei-Jia Yao, Tong-Hua Li, Yi-Xue Li and Zhi-Wei Cao, "Cancer Classification from the Gene Expression Profiles By Discriminant Kernel-Pls", *Journal Of Bioinformatics And Computational Biology*, Vol.8, Suppl.1 (2010) 147-160.
- [9] A. Bharathi and Dr.A.M.Natarajan,"Cancer Classification of Bioinformatics Data Using ANOVA ", *International Journal of Computer Theory And Engineering*, Vol. 2, No. 3, 1793-8201, 2010.
- [10] Huilin Xiong And Xue-Wen Chen, "optimized Kernel Machines For Cancer Classification Using Gene Expression Data" , *Proceedings Of The 2005 IEEE Symposium On Computational Intelligence In Bioinformatics And Computational Biology*,Pp.1-7, 2005.
- [11] G. Raju, A. Singh, Th. Shanta Kumar, Binu Thomas, "Integration of Fuzzy Logic in Data Mining: A comparative Case Study", *Proc. Of International Conf. on Mathematics and Computer Science*, Loyola College, Chennai, 128-136, 2008.
- [12] Wiest JS, Franklin WA, Drabkin H, Gemmill R, Sidransky D, Anderson MW, "Genetic markers for early detection of lung cancer and outcome measures for response to chemoprevention", *University of Cincinnati College of Medicine*, Ohio, USA, 2010.
- [13] U. Maulik, S. Bandyopadhyay, and A. Mukhopadhyay, *Multiobjective Genetic Algorithms for Clustering: Applications in Data Mining and Bioinformatics*. New York: Springer- Verlag, 2011.
- [14] A. Mukhopadhyay, S. Bandyopadhyay, and U. Maulik, "Multi-class clustering of cancer subtypes through SVM based ensemble of pareto-optimal solutions for gene marker identification," *PLoS ONE*, vol. 5, no. 11, pp. 1–14, 2010.
- [15] Wiest JS, Franklin WA, Drabkin H, Gemmill R, Sidransky D, Anderson MW, "Genetic markers for early detection of lung cancer and outcome measures for response to chemoprevention", *University of Cincinnati College of Medicine*, Ohio, USA, 2010.
- [16] Mathew J. Garnett, Patricia Greninger, I. Richard Thompson, Xi Luo, Jorge Soares, Qingsong Liu, Francesco Iorio, Didier Surdez, Li Chen, Randy J. Milano, Graham R. Bignell, Ah T. Tam, Helen Davies, Jesse A. Stevenson, Syd Barthorpe, Stephen R. Lutz, Fiona Kogera, Karl Lawrence, Anne McLaren-Douglas, Xenia Mitropoulos, "Systematic identification of genomic markers of drug sensitivity in cancer cells ", 483,570–575, 29 March 2012.
- [17] A. J. Gentles, S. K. Plevritis, R. Majeti, and A. A. Alizadeh, "Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia," *J. Amer. Med. Assoc.*, vol. 304, pp. 2706–2715,2010.