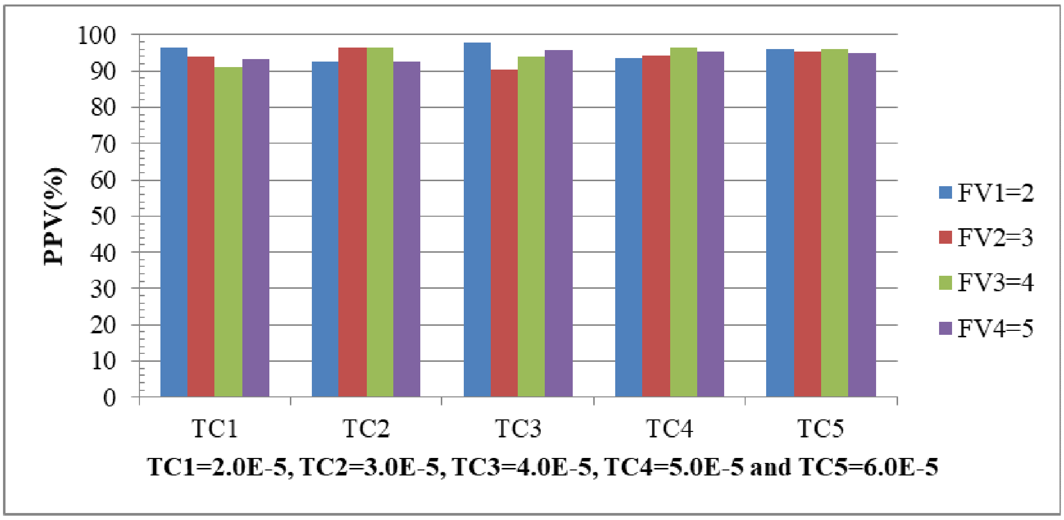

Fig. 5 PPV result based on degree of membership (random initialization) phase-3

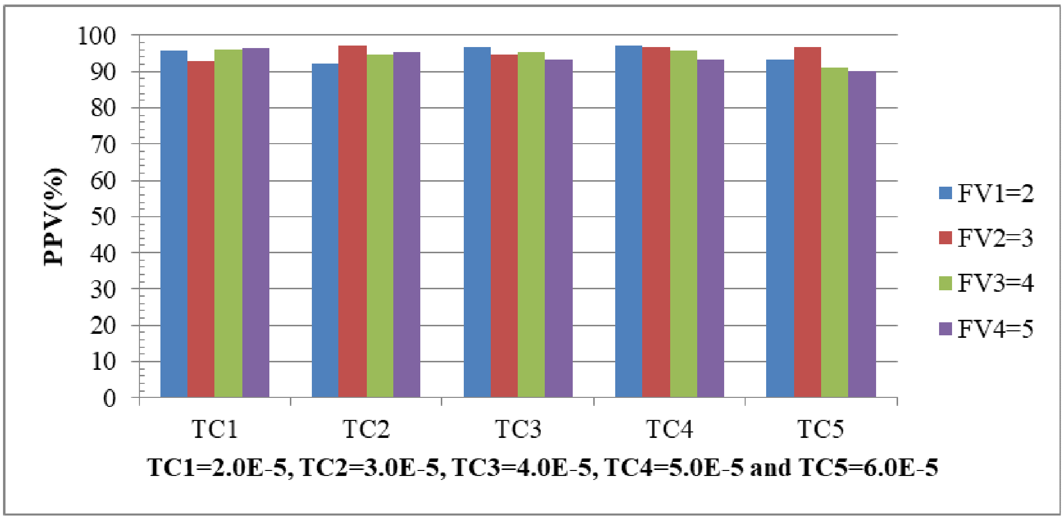

Fig. 6 PPV result based on degree of membership (random initialization) phase-4

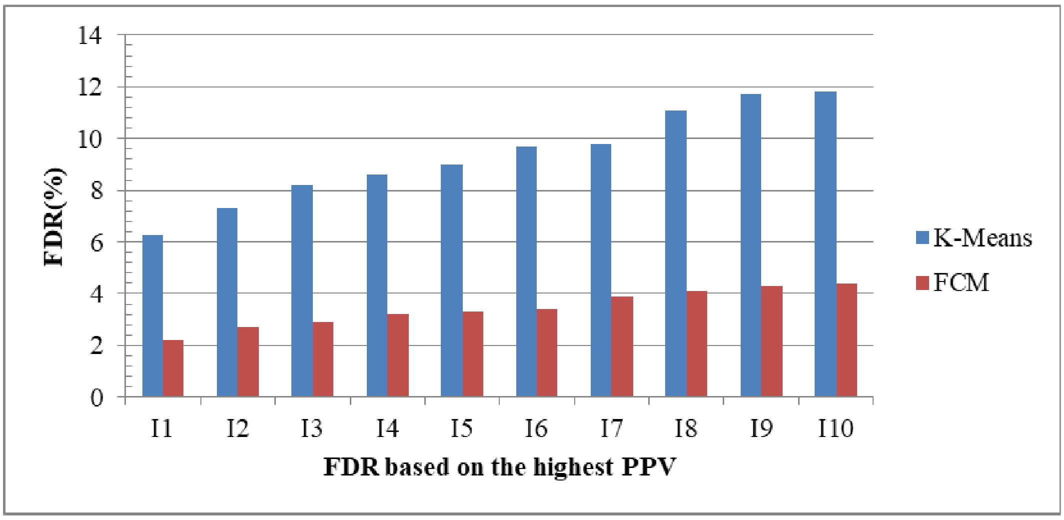

Fig. 7 FDR results based on the highest PPV achieved in case of k-means and FCM

So in all the cases, different parameters and populations were considered and it was found that FCM outperforms in comparison to k-means algorithm for the breast cancer data clustering.

Computation time comparisons with different FVs and TCs using FCM are shown in Table 13. The computation time, in case of FCM, is high due to the several iterations and fuzzy measures' calculations (Table 13). The calculative parameters, degree of membership evaluation and updating and the steps involve in comparing the termination criteria is extensive. So the time taken in the FCM process is more comparison to the k-means algorithm.

TABLE XIII
TIME VALUES CONSIDERING VARIOUS FUZZINESS VALUE AND TERMINATION CRITERIA USING FCM

S.no	Termination Criteria	Time (MS)	Time (MS)	Time (MS)	Time (MS)
		FV=2	FV=3	FV=4	FV=5
1	2.0E-5	928	1035	1538	1545
2	3.0E-5	1032	1506	1145	1521
3	4.0E-5	1548	1190	1464	889
4	5.0E-5	1339	1513	939	1461
5	6.0E-5	961	1232	1343	1473

In this study, the performance of k-means and FCM algorithms for BCW dataset was compared. BCW dataset is considered to delineate the impact of clustering on the basis of different parameters. The key findings are as follows:

1. The results obtained by k-means algorithm show that the variation in the total PPV is due to the random initialization in case of foggy centroid attribute values, as it is participating in determining the mode for the mean and variance in the centroid.
2. In the case of same centroid, the process is stopped if the means do not change anymore. Therefore, the process is not restricted to the epoch and the possibility of better cluster selection is improved. The combination of highest variance and the same centroid provides good results in case of k-means.
3. FCM algorithm produces better results in comparison to the k-means algorithm. The highest accuracy obtained is 97% and 92% respectively for FCM and k-means algorithms. FCM provides an iterative analysis so it gives better results for all selections.
4. The results are approximately same in the case of BCW dataset for a fuzziness value of 2-5. So for BCW dataset a fuzziness value 2-5 can be considered.
5. In case of computation time, k-means algorithm is far better than FCM algorithm. As the computation time is high in FCM algorithm. In case of k-means algorithm, if the number of iterations is more, the number of partitions is also more, which contain less attributes, so the computation time is low in case of high number of iterations compared to a low number of iterations. But in case of FCM algorithm, the computation time is high due to the various calculations, fuzzy measures, and a large number of iterations.
6. In k-means algorithm the data point should belong to one cluster but in case of FCM, it may belong to more than one cluster as membership is assigned to each data point.
7. As the computations are checked several times and the results obtained are uniform so FCM clustering is relatively efficient.

IV. CONCLUSIONS

K-means and FCM clustering algorithms were used in this study for the clustering of the BCW dataset. K-means algorithm is a simple and easy way to classify datasets through assuming k clusters with fixed apriori. FCM algorithm provides an iterative process with the update of cluster centers by updating and assigning membership values. In this work, a computational formulation is presented for integrative clustering with multi variant parameters including BCW data for obtaining good clustering accuracy. K-means algorithm is presented with foggy and random centroids considering the centroid, distance, split method, threshold, epoch, BCW attribute and number of iterations. This work is elaborated in several ways and makes certain important observations. The results of k-means algorithm indicated good accuracy in case of highest variance and same centroid. The consistency and uniformity in case of FCM algorithm is more prominent than k-means algorithm as the results of several repetitions suggest. The highest accuracy obtained is 97% and 92% for FCM and k-means algorithms, respectively. But the computation time is higher in FCM compared to k-means algorithm, thus k-means algorithm is efficient in terms of computation time. This implies that the FCM algorithm produces better results in comparison to the k-means algorithm but with the higher computation time in case of BCW dataset. In future the work is extended in the direction of clustering the non-clustered data obtained from k-means and FCM algorithms.

REFERENCES

[1] S.A. Eccles, E.O. Aboagye, S. Ali, A.S. Anderson, J. Armes, F. Berditchevski, J.P. Blaydes, K. Brennan, N.J. Brown, H.E. Bryant and N.J. Bundred, "Critical research gaps and translational priorities for the successful prevention and treatment of breast cancer," *Breast Cancer Res*, vol. 15, pp. 1-37, 2013.

[2] J. Ferlay, H.R. Shin, F. Bray, D. Forman, C. Mathers and D.M. Parkin, "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008," *Int J Cancer, vol.* 127, pp. 2893-2917, 2010.

[3] A.K. Dubey, U. Gupta and S. Jain, "Breast cancer statistics and prediction methodology: a systematic review and analysis," *Asian Pac J Cancer P*, vol. 16, pp. 4237-4245, 2015.

[4] A.K. Dubey, U. Gupta and S. Jain, "A survey on breast cancer scenario and prediction strategy," *In proceedings of the 3rd international conference on frontiers of intelligent computing: theory and applications (FICTA),* (pp. 367-375), Springer International Publishing, 2015.

[5] R. Jain, *Introduction to Data Mining Techniques*. Available: http://www.iasri.res.in/ebook/expertsystem/datamining.pdf.

[6] S. Ghosh and S.K. Dubey, "Comparative analysis of k-means and fuzzy c-means algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 4, pp.35-38, 2013.

[7] C. Mary and S.K. Raja, "Refinement of clusters from k-means with ant colony optimization," *J Theor Appl Inf Technol*, vol. 6, pp. 28-32, 2009.

[8] C. Wang, R. Machiraju and K. Huang, "Breast cancer patient stratification using a molecular regularized consensus clustering method," *Methods*, vol. 67, pp. 304-312, 2014.

[9] A.K. Dubey, U. Gupta and S. Jain, “Analysis of k-means clustering approach on the breast cancer Wisconsin dataset,” *Int J Comput Assist Radiol Surg, vol.* 11, pp. 2033-2047, 2016.
[10] S. Bhardwaj and V. Verma,” Improved k-means clustering algorithm using back propagation method,” *Int* J *Control Theor AP,* vol. 9, pp. 5169-5180, 2016.
[11] J. Lu, A. Hales and D. Rew, “Modelling of cancer patient records: a structured approach to data mining and visual analytics,” *In International Conference on Information Technology in Bio-and Medical Informatics,* (pp. 30-51), Springer, 2017.
[12] A. Rahideh and M.H. Shaheed, “Cancer classification using clustering based gene selection and artificial neural networks,” *In international conference on control, instrumentation and automation,* (pp. 1175-1180), IEEE, 2011.
[13] P. Festa, “A biased random-key genetic algorithm for data clustering,” *Math Biosci*, vol. 245, pp. 76-85, 2013.
[14] C.H. Chen, “A hybrid intelligent model of analyzing clinical breast cancer data using clustering techniques with feature selection,” *Appl Soft Comput,* vol. 20, pp. 4-14, 2014.
[15] D. Wei, Q. Jiang, Y. Wei and S. Wang, “A novel hierarchical clustering algorithm for gene sequences,” *BMC bioinformatics*, vol. 13, pp. 174, 2012.
[16] D. Vanisri and C. Loganathan, Fuzzy pattern cluster scheme for breast cancer datasets,” *In international conference on communication and computational intelligence,* (pp. 410-4), IEEE, 2010.
[17] F.K. Ahmad and N. Yusoff, “Classifying breast cancer types based on fine needle aspiration biopsy data using random forest classifier,” *In international conference on intelligent systems design and applications,* (pp. 121-5, IEEE, 2013.
[18] J. Yin, H. Sun, J. Yang and Q. Guo, “Comparison of k-means and fuzzy c-means algorithm performance for automated determination of the arterial input function”, *PloS one*, vol. 9(2), pp. e85884, 2014.
[19] Z. Zainuddin and O. Pauline, “An effective fuzzy c-means algorithm based on symmetry similarity approach”, *Appl Soft Comput*, vol. 35, pp. 433-448, 2015.
[20] B. Zheng, S.W. Yoon and S.S. Lam, “Breast cancer diagnosis based on feature extraction using a hybrid of k-means and support vector machine algorithms,” *Expert Syst Appl*, vol. 41, pp. 1476-1482, 2014.
[21] A.A. Rachman and Z. Rustam, “Cancer classification using fuzzy c-means with feature selection,” I*n international conference on mathematics, statistics, and their applications,* (pp. 31-34), IEEE, 2016.
[22] P.S. Deepthi and S.M. Thampi, “Predicting cancer subtypes from microarray data using semi-supervised fuzzy C-means algorithm,” *J Intell Fuzzy Sys,* vol. 32, pp. 2797-805, 2017.
[23] K. Bache and M. Lichman, *UCI machine learning repository,* University of California, School of Information and Computer Science, Available: http://archive. ics. uci. edu/ml.
[24] C.B. James, *Pattern recognition with fuzzy objective function algorithms*, New York: Plenum Press, 1981.
[25] N.R. Pal and J.C. Bezdek, “On cluster validity for the fuzzy c-mean model,” *IEEE Trans Fuzzy Syst,* vol. 3, pp. 370–379, 1995.
[26] J.C. Bezdek, “A physical interpretation of fuzzy ISODATA,” *IEEE Trans Syst Man Cy B*, vol. 6, pp. 387–389, 1976.
[27] K. Zhou, C. Fu and S. Yang, “Fuzziness parameter selection in fuzzy c-means: The perspective of cluster validation,” *Science China Information Sciences,* vol. 57, pp. 1-8, 2014.