Efficiency Analysis of Hybrid Fuzzy C-Means Clustering Algorithms and their Application to Compute the Severity of Disease in Plant Leaves

B.K. Tripathy[1], Anmol Agrawal[2]

School of Information Technology and Engineering[1], School of Computer Science and Engineering[2], VIT University, Vellore – 632014, India

Abstract

Data clustering has a wide range of application varying from medical image analysis, social network analysis, market segmentation, search engines, recommender systems and image processing. A clustering algorithm should be fast as well accurate. Some applications give priority to the speed of the clustering algorithms while some emphasize more on the accuracy rather than speed. A number of clustering algorithms have been proposed in the literature. Some of these include Fuzzy C-Means (FCM), Intuitionistic Fuzzy C-Means (IFCM), Rough Fuzzy C-Means (RFCM) and Rough Intuitionistic Fuzzy C-Means (RIFCM). In this paper, we compare the accuracy and execution time of the fuzzy based clustering algorithms. The clustering algorithms are applied on an image dataset and their running time as well as accuracy is compared by varying the number of clusters. Our results show that there is a clear trade-off between execution time and accuracy of these clustering algorithms. Algorithms having higher accuracy (lower DB and higher DUNN) have take more time to execute (measured in seconds) and vice versa. Also, we apply these algorithms on two different diseased leaf images and compute the severity of the disease of the leaves.

Keywords:Data Clustering; Fuzzy C-Means; Intuitionistic Fuzzy C-Means; Rough Fuzzy C-Means; Rough Intuitionistic Fuzzy C-Means; Leaf Disease

1. Introduction

A clustering algorithm involves segregation of data elements into groups (clusters). Members of the same cluster are similar to one another while different from members of different cluster. Clustering algorithms are one of the most important unsupervised learning algorithms and hence, do not require a labelled data set. Over the last few decades, clustering algorithms have been used widely in a number of fields such as machine learning, information retrieval, image processing, and bio-informatics. Each application has their specific requirements and so a number of clustering algorithms have been introduced to cater different needs.

The most basic and commonly used clustering algorithm is the K-means algorithm. Although it is fast, it loses out on accuracy when compared to soft clustering algorithms. These soft clustering algorithms include the Fuzzy C-Means (FCM) [1] which is based on the concept of Fuzzy sets [2], Intuitionistic Fuzzy C-Means (IFCM) [3] which uses the intuitionistic fuzzy set theory [4], Rough Fuzzy C-Means (RFCM) [5-6] that uses the rough set [7] model and Rough Intuitionistic Fuzzy C-Means (RIFCM) [8] which is a combination of RFCM and IFCM clustering algorithms.

A comparison of K-Means and Fuzzy C-Means has been made [9] where it was shown that K-Means is faster than FCM, but is less accurate and easily susceptible to outliers, local optima and has uncertainty in the number of iterations required to form the cluster. In this paper, we compare the efficiency of FCM, IFCM, RFCM and RIFCM by taking two parameters – (i) Execution time and (ii) Accuracy. Execution time is measured as the time required to form the clusters. Accuracy measures the quality of clusters formed. Two performance indices, namely, Davis Bouldin [10] and Dunn [11] have been used for this purpose. The parameters have been chosen to establish the superiority of different clustering algorithms for different requirements, i.e., to determine the clustering algorithms that should be used when (a) the time-constraint is most important (b) the accuracy is of utmost priority and (c) both accuracy and speed are significant.

In the later part of the paper, we show how these algorithms can be efficiently used to calculate the percentage of disease in plant leaves. Knowing the severity of disease is important because over the last few decades, there has been a reduction of 20-40% of the total agricultural productivity across the globe [12]. Several fungus, bacteria *(eg Thiobacillus* denitrificans and *Micrococcus denitrificans*) and viruses are responsible for this reduction. This calls for urgent and efficient steps to address this problem. Although the disease may be identified visually, the problem of estimating the severity of the disease on leaves still needs to be addressed. Several visual techniques exist to address this issue [13-14], but they are often time-consuming, inconsistent and in-accurate. To resolve this issue, several segmentation and image processing techniques have been employed, such as, segmentation of infected pixels on the basis of gray levels [15], identification of symptom edges using Sobel operator [16], Histogram of intensities [17] and triangle thresholding [18]. However, the techniques discussed above use discrete boundaries or thresholds to segment the diseased area from the healthy area. Thus, in this paper we use soft clustering algorithms to solve the issue. With respect to the position of elements in various clusters, clustering techniques can be divided into two types: (a) Hard Clustering and (b) Soft clustering. In hard clustering, one data point can belong to at most one cluster i.e. they either belong to the cluster or not. Soft clustering algorithms, on the other hand, allow data points to belong to more than one cluster simultaneously, thereby increasing accuracy. Some of the soft clustering techniques have been used for image segmentation [19, 20, 21]. The soft clustering algorithms that are analysed in this paper are then used to segment out the diseased area from the plant leaf which is used to measure the severity. A more accurate clustering algorithm signifies more precise results. Thus, using the results from the first part of the paper, we find the most accurate measure of the severity of disease in the plant leaves.

2. Definitions and Notations

In this section, we present some of the algorithms and definitions used in the paper:

2.1. Clustering Algorithms

Some of the clustering algorithms used in the paper are as follows:

2.1.1. Fuzzy C-Means

In FCM, cluster centroids are initialized randomly. The distance d_{ik} between every cluster center i and every pixel of the image k is calculated using the Euclidean distance. The Membership Matrix is calculated according to the equation:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}} \tag{1}$$

Where, c denotes the total number of clusters and m represents the fuzzifier. The cluster centres are calculated using the following formula:

$$v_i = \frac{\sum_{j=1}^{N}(\mu_{ij})^m x_j}{\sum_{j=1}^{N}(\mu_{ij})^m} \tag{2}$$

This method is important because it provides a solution to the limitations faced by the infinite solution space. This was achieved by transforming the original problem to the minimization of the objective function J given by:

$$J = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik})^m d^2(x_k, v_i) \tag{3}$$

2.1.2. Intuitionistic Rough Fuzzy C-Means

Chaira, in her paper [4] uses Yager's Fuzzy Complement (Yager, 1980):

$$M(\mu(x)) = f^{-1}(f(1) - f(\mu(x))) \tag{4}$$

Here, f is an increasing function between 0 and 1. Yager's intuitionistic fuzzy complement is given by:

$$f(x) = (1 - x^{\alpha})^{\frac{1}{\alpha}} \tag{5}$$

where M (1) = 0, M (0) = 1. Non-membership values are calculated from Yager's complement *M(x)*. Thus, the hesitation degree of *x* is:

$$\pi_A(x) = 1 - \mu_A(x) - (1 - \mu_A(x)^{\alpha})^{\frac{1}{\alpha}} \tag{6}$$

The modified membership function μ' is given by $\mu_A^{'}(x) = \mu_A(x) + \pi_A(x), \forall x$.

The modified cluster centre is:

$$v_i = \frac{\sum_{k=1}^{n} \mu_{ik}^{'} x_k}{\sum_{k=1}^{n} \mu_{ik}^{'}} \tag{7}$$

The objective function of IFCM is given as:

$$J = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik}^{'})^m d(x_k, v_i)^2 + \sum_{i=1}^{c} \pi_i^{'} e^{1-\pi_i^{'}} \tag{8}$$

2.1.3. Rough Fuzzy C-Means

This algorithm combines both Rough sets [9] and fuzzy sets [10]. The major difference is that instead of checking for the closest and the next closest distance between a pixel and the cluster centres, the maximum and next to maximum membership values of the pixel to all the clusters is considered. The membership values are computed using equation (1). The pseudo-code of RFCM is:

Step 1. Assign initial means v_i for c clusters

Step 2. Compute μ_{ik} using (1)

Step 3. Let μ_{ik} and μ_{jk} be the maximum and next to maximum values of the object x_k to the clusters with centroids v_i and v_j respectively among all the clusters

Step 4. If $\mu_{ik} - \mu_{jk} < \varepsilon$ (for some preassigned value ε) then $x_k \in \overline{B}U_i$ *and* $x_k \in \overline{B}U_j$ and x_k cannot be a member of any lower approximation

Else $x_k \in \underline{B}U_i$

Step 5. Compute the new cluster centres by the following formula (8), where $0 \le w_{low} + w_{up} \le 1$

$$v_i = \begin{cases} w_{low} \dfrac{\sum_{x_k \in \underline{B}U_i} x_k}{|\underline{B}U_i|} + w_{up} \dfrac{\sum_{x_k \in \overline{B}U_i \setminus \underline{B}U_i} \mu_{ik}^m x_k}{\sum_{x_k \in \overline{B}U_i \setminus \underline{B}U_i} \mu_{ik}^m}, if\ \underline{B}U \ne \phi \wedge \overline{B}U_i \setminus \underline{B}U_i \ne \phi; \\ \dfrac{\sum_{x_k \in \overline{B}U_i \setminus \underline{B}U_i} \mu_{ik}^m x_k}{\sum_{x_k \in \overline{B}U_i \setminus \underline{B}U_i} \mu_{ik}^m}, \quad if\ \overline{B}U_i \setminus \underline{B}U_i \ne \phi; \\ \dfrac{\sum_{x_k \in \underline{B}U_i} x_k}{|\underline{B}U_i|}, \quad ELSE. \end{cases} \tag{9}$$

where, $0 \le w_{low}, w_{up} \le 1$ such that $w_{low} + w_{up} = 1$

Step 6. Repeat steps 2 to 5 until the difference between two successive values of U is less than a preassigned value.

2.1.4. Rough Intuitionistic Fuzzy C-Means

All the clustering algorithms discussed so far have some or the other major drawback. In RFCM, the fuzzy part does not handle the error efficiently resulting in lack of accuracy. IFCM does not deal with incompleteness due to the absence of rough set theory. In order to overcome these challenges, RIFCM was developed [2]. In RIFCM, any cluster can be uniquely identified by:

- Centroid
- Crisp lower approximation
- Intuitionistic fuzzy boundary

The pseudocode for RIFCM is given below.

Step 1. Select c objects from the data set and assign one each to the c clusters as initial centroids

Step 2. Compute d_{ik} the distance between the object x_k and the centroid v_k by using the Euclidean distance formula $d(x_i, v_k) = \left(\sum_{j=1}^{n} (x_{ij} - v_{kj})^2\right)^{1/2}$.

Step 3. Compute the initial matrix U.

Step 4. If d_{ik} = 0 or $x_k \in \underline{B}U_i$ then $\mu_{ik} = 1$. Else μ_{ik} is computed by using the formula (1).

Step 5. Compute π_{ik} by using the formula

$$\pi_A(x) = 1 - \mu_A(x) - \frac{1-\mu_A(x)}{1+\lambda\mu_A(x)}, \forall x \quad (10)$$

Step 6. Compute $\mu_{ik}^{'}$ by the formula (11) and normalize

$$\mu_{ik}^{'}(x) = \mu_{ik}(x) + \pi_{ik}(x), \forall x \quad (11)$$

Step 7. Let $\mu_{ik}^{'}$ and $\mu_{jk}^{'}$ be the maximum and next to maximum values of the object x_k to the clusters with centroids v_i and v_j respectively among all the clusters

Step 8. If $\mu_{ik}^{'} - \mu_{jk}^{'} < \varepsilon$ (for some preassigned value ε) then $x_k \in \overline{B}U_i$ *and* $x_k \in \overline{B}U_j$ and x_k cannot belong to any lower approximation

Else $x_k \in \underline{B}U_i$

Step 9. Calculate the new cluster centres by using the following formula (12), where $0 \le w_{low}, w_{up} \le 1$ such that $w_{low} + w_{up} = 1$

$$v_i = \begin{cases} w_{low} \dfrac{\sum_{x_k \in \underline{B}U_i} x_k}{|\underline{B}U_i|} + w_{up} \dfrac{\sum_{x_k \in \overline{B}U_i \setminus \underline{B}U_i} (\mu_{ik}^{'})^m x_k}{\sum_{x_k \in \overline{B}U_i \setminus \underline{B}U_i} (\mu_{ik}^{'})^m}, & if\ \underline{B}U \ne \phi \wedge \overline{B}U_i \setminus \underline{B}U_i \ne \phi; \\ \dfrac{\sum_{x_k \in \overline{B}U_i \setminus \underline{B}U_i} (\mu_{ik}^{'})^m x_k}{\sum_{x_k \in \overline{B}U_i \setminus \underline{B}U_i} (\mu_{ik}^{'})^m}, & if\ \overline{B}U_i \setminus \underline{B}U_i \ne \phi; \\ \dfrac{\sum_{x_k \in \underline{B}U_i} x_k}{|\underline{B}U_i|}, & ELSE. \end{cases} \quad (12)$$

Step 10. Repeat steps 2 to 9 until the difference between two successive values of U is less than a preassigned value.

From the pseudo-code, it can clearly be observed that the major difference between RFCM and RIFCM is that the hesitation degree is added to the membership matrix before the Lower and Upper approximations are computed. Thus, RIFCM is a hybridization of IFCM and RFCM.

2.2. Performance Indices

Performances indices are used to evaluate the performance of clustering algorithms. There are several performance indices available in the literature. The Davis-Bouldin (DB) index [6] and Dunn index (D) [7] are some of the most common efficiency analysis indices.

2.2.1. Davis-Bouldin (DB) index

The DB index is defined as the ratio of the sum of distance within the cluster to distance between cluster. It is formaulated as follows (eq. 13),

$$DB = \frac{1}{c}\sum_{i=1}^{c} max_{k \ne i}\left\{\frac{S(v_i)+S(v_k)}{d(v_i,v_k)}\right\} for 1 < i,k < c \quad (13)$$

The DB index aims to minimize the separation within cluster and maximize the distance between clusters. Hence, a low DB value indicates good clustering.

2.2.2. Dunn (D) index

The D index is used to identify the compact and separated clusters. It is calculated as follows (eq. 14):

$$Dunn = min_i \left\{ min_{k \neq i} \left\{ \frac{d(v_i, v_k)}{max_l S(v_l)} \right\} \right\} for 1 < k, i, l < c \qquad (14)$$

The aim of Dunn index is to maximize the distance between cluster and minimizing the distance within-cluster. Thus, a high D index signifies better clustering.

3. Efficiency Analysis of The Clustering Algorithms

Implementations have been carried out in Python 3.6 using Spyder IDE as it provides a number of useful libraries that are needed for computation and plotting of figures. Sypder IDE is used as it is a very efficient tool for writing and debugging Python programs. The programs have been executed on a Lenovo Ideapad machine running on Intel® Core™ i5 6th generation processor, 8 GB memory and 1TB hard disk. The matplotlib library has been used to plot the resultant figures.

Figure 1. Damaged Leaf

Figure 1 represents the image of a damaged leaf of dimensions 256 x 256 pixels. Hence, the total number of data objects in the image dataset is 65536. The above image has been considered as an image dataset to study the performance of the various clustering algorithms. The clustering algorithms FCM, IFCM, RFCM, RIFCM have been applied on the image. The image is segmented into 2, 3 and 4 clusters in each case and the execution time for each case is recorded. Also, the performance indices, DB and DUNN are computed for each case.

Table 1. Execution Time (in seconds) of each algorithm for different number of clusters

Number of Clusters	FCM	IFCM	RFCM	RIFCM
1	0.0052	0.0167	0.8447	0.9411
2	0.0268	0.1189	2.2330	5.8950

3	0.1837	0.6197	8.6272	11.0020
4	0.4373	1.4350	14.7647	23.1873

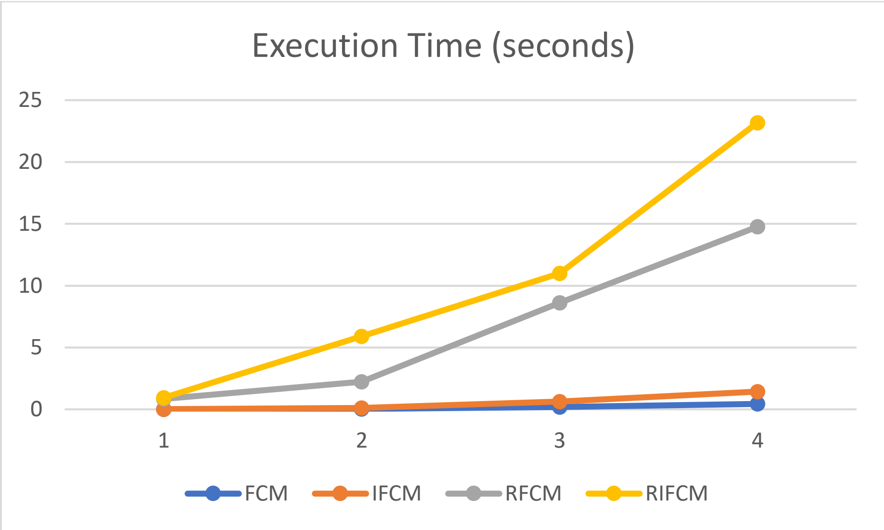

Figure 2. Execution Time of each algorithm for different number of clusters

From the above graph, we can observe that execution time of FCM and IFCM are roughly the same and are much lower than those of RFCM and RIFCM. The execution time for RIFCM is the highest, while that of FCM is the lowest. The execution time of RFCM and RIFCM rises significantly as the number of clusters increase.

Table 2. DB Index for different algorithms for different number of clusters

Number of Clusters	**FCM**	**IFCM**	**RFCM**	**RIFCM**
2	6.6695	6.5748	3.6028	1.6653
3	6.3667	6.1570	2.2081	0.8839
4	6.1087	5.9167	0.6028	0.5835

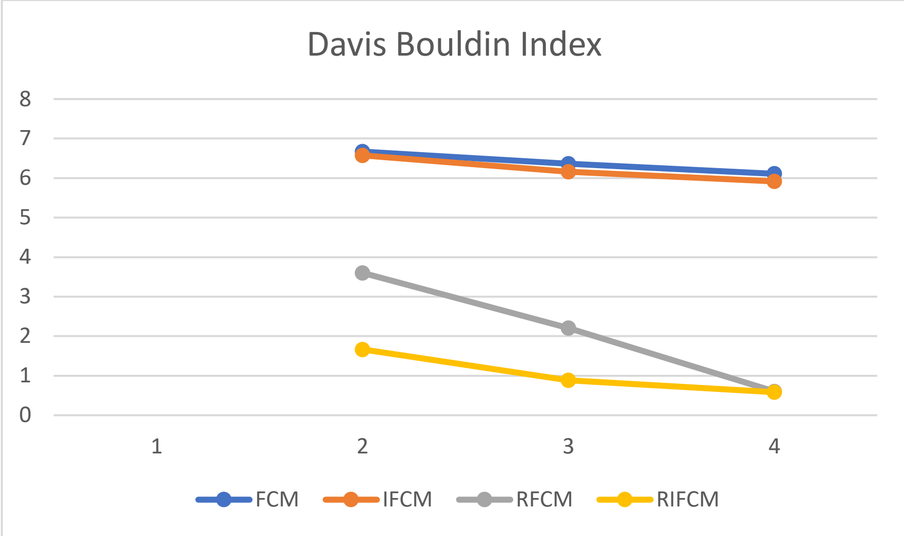

Figure 3. DB Index for different algorithms for different number of clusters

The above graph shows the relationship between the DB index of the various clustering algorithms. A lower value of DB index implies better clustering. It is clearly evident from the figure 3 that the DB values of FCM and IFCM is almost the same, with IFCM slightly outperforming FCM. The performance of RFCM and RIFCM is also similar, with RIFCM giving the best results in terms of accuracy closely followed by RFCM.

Table 3. DUNN index of various algorithms for different number of clusters

Number of Clusters	FCM	IFCM	RFCM	RIFCM
2	0.1767	0.1769	0.4297	0.7815
3	0.1389	0.1431	0.5187	0.8203
4	0.1422	0.1476	0.8113	1.1199

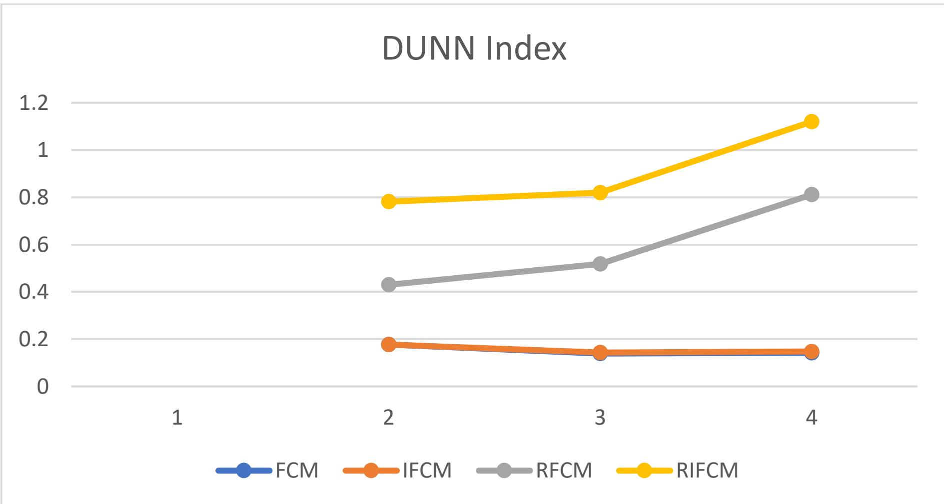

Figure 4. DUNN index of various algorithms for different number of clusters

Figure 4 shows the graphical representation of the DUNN values of various clustering algorithms. A higher DUNN value implies more accurate clustering. Even here, it is clear that the performance of FCM and IFCM is similar while the performance of RFCM and RIFCM are much better, with RIFCM outperforming RFCM.

Table 4. Number of iterations required to converge for various algorithms with different number of clusters

Number of Clusters	FCM	IFCM	RFCM	RIFCM
2	7	7	5	6
3	23	20	18	15
4	39	33	34	31

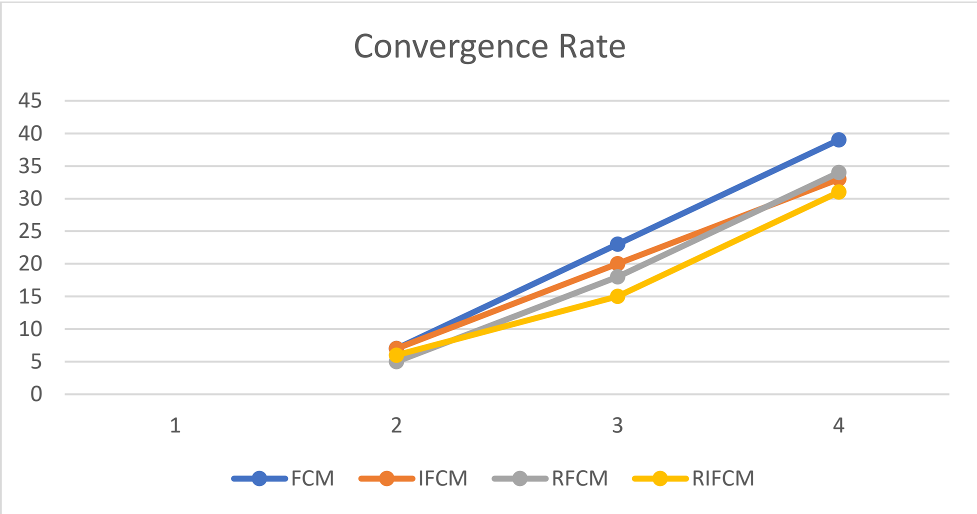

Figure 5. Number of iterations required to converge for various algorithms with different number of clusters

Figure 5 plots the number of iterations that are required for the algorithms to converge. It is clear that FCM requires the maximum number of iterations to converge, while RIFCM required the least number of iterations. The difference broadens as the number of clusters increase.

4. Methodology for Computing The Severity of Disease in Plant Leaves

4.1. Image Dataset

The following two leaf images have been taken from the PlantVillage dataset [22]. The database contains 54,309 images of leaves spanning 14 species affected by fungal, bacterial and several other diseases.

Figure 6(a). Leaf suffering from early blight and Figure 6(b). Leaf suffering from late blight

Figure 6(a) shows the image of a potato leaf suffering from early blight (*Alternaria solani)*. Figure 6(b) corresponds to late blight (*Phytophthora Infestans*) disease on a potato leaf. The images have been separated from the background and colour corrected.

4.2. Segmentation

The fuzzy based clustering algorithms analysed earlier, namely, Fuzzy C-Means (FCM), Intuitionistic Fuzzy C-Means (IFCM), Rough Fuzzy C-Means (RFCM) and Rough Intuitionistic Fuzzy C-Means (RIFCM) have been applied to the above leaf images to segment out the disease affected area from the healthy area. This is the most important step that determines the accuracy of the result obtained. A more accurate clustering algorithm will provide better results. Since RIFCM performed best in the above analysis, it will give the best results.

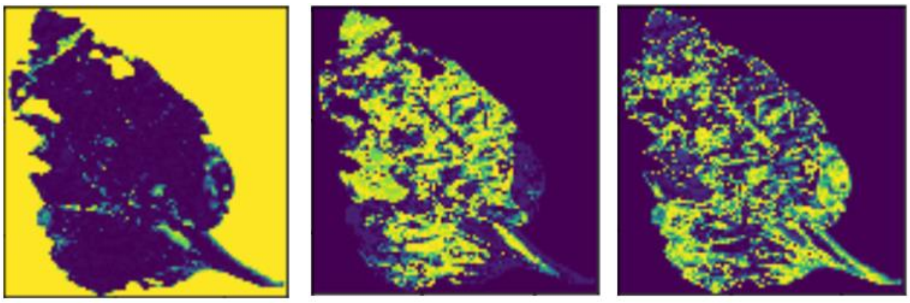

Figure 7. Results obtained after segmenting the leaf in figure 6(a) into three clusters using FCM

4.3. Conversion of RGB To Binary Image

The segmented image is then converted to a binary image using Otsu's binarization algorithm. The purpose of converting the images obtained after segmentation into binary is to simplify the area calculation process. The image now consists of pixels having values either 0 (black) or 1 (white). Therefore, the area covered by the white region can be computed easily.

Figure 8. Result obtained after converting the image in figure 7 into binary by Otsu's method.

4.4. Calculating the percentage of leaf area affected by the disease

The percentage area of the affected region is finally calculated using the following formula:

$$Severity\ (\%) = \frac{Area\ of\ the\ disease\ affected\ surface}{Total\ area\ of\ the\ leaf} X\ 100 \quad (15)$$

5. Results

The above method has been applied using FCM, IFCM, RFCM and RIFCM and the following results were observed.

Table 5: Severity percentage obtained after applying the various clustering algorithms on fig 6(a)

ALGORITHM	DB	DUNN	SEVERITY (%)
FCM	9.6329	0.1577	22.56%
IFCM	9.3866	0.1629	27.07%
RFCM	2.1767	0.4588	36.42%
RIFCM	1.0970	0.8930	37.90%

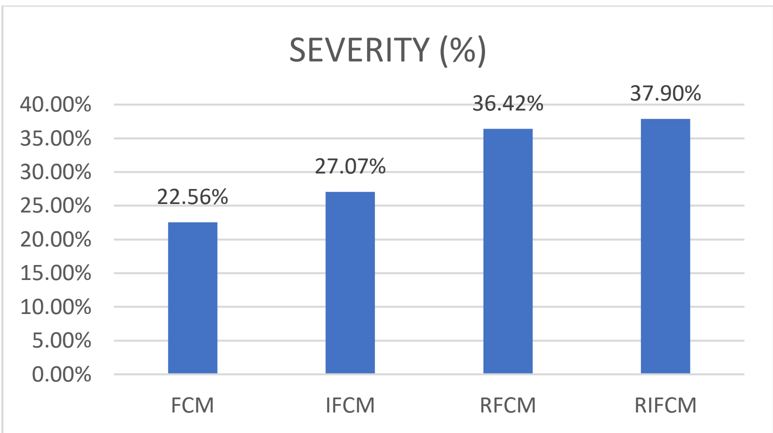

Figure 9. Graph showing the severity for the leaf in figure 6(a) using various algorithms.

The severity percentage of the disease calculated for the leaf in fig. 6(a) and the performance of the clustering algorithms is tabulated in table 5. The severity percentage of the damage caused by the disease is plotted in fig. 9. Since it has been established above that the accuracy of RIFCM is superior than all other algorithms hence, the severity percentage of the disease calculated by using this algorithm is the most precise.

Table 6: Severity percentage obtained after applying the various clustering algorithms on fig 6(b)

ALGORITHM	DB	DUNN	SEVERITY (%)
FCM	10.3077	0.1526	63.73%
IFCM	10.0627	0.1571	63.48%
RFCM	2.3275	0.4896	59.88%
RIFCM	1.0755	0.9096	60.22%

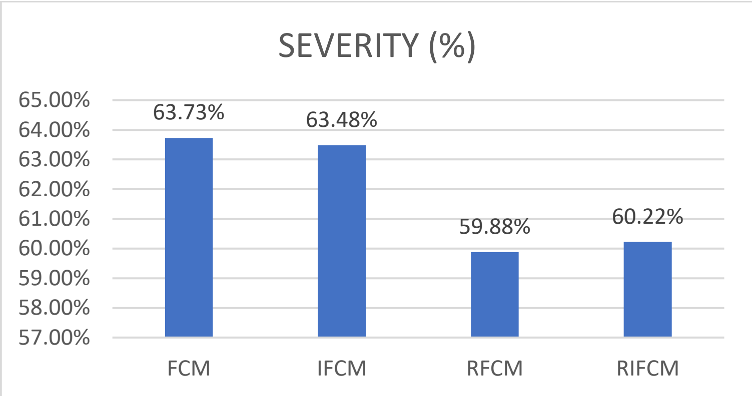

Figure 10. Graph showing the severity percentage of the leaf in figure 6(b) using various algorithms

The severity percentage of the disease calculated for the leaf in fig. 6(b) and the performance of the clustering algorithms is tabulated in table 6. Just like the previous case, since the accuracy of RIFCM clustering algorithm is the highest, the severity percentage obtained using RIFCM is the most accurate.

6. Conclusion

From the results obtained above, the following points can be concluded:

- FCM takes the least execution time and thus can applied where time constraint is of utmost significance.
- RIFCM has the best accuracy and can be applied in situations where accuracy is the most important factor.
- There is a clear trade-off between speed and accuracy of the clustering algorithms. FCM which has the lowest execution time also has lowest accuracy, while RIFCM which has the highest execution time also has the highest accuracy.
- RFCM suits the best when both speed and accuracy are important, as it lies roughly in between the two extremes (FCM and RIFCM) both in terms of speed and accuracy.

- There is no direct relation between convergence rate and execution time of the clustering algorithms. FCM requires the highest number of iterations to form the final clusters but takes the least time. On the other hand, RIFCM takes the least number of iterations to form the clusters but has the longest running time. This is because RIFCM involves a greater number of computationally expensive steps in each iteration when compared to FCM.

- In applications like computing the severity of disease in plant leaves, since accuracy is of utmost importance, RIFCM is most suitable for this purpose.

- From the results obtained above, we can conclude that the leaf in fig.6(a) is 37.90% infected by early blight disease (*Alternaria solani)*. Similarly, the leaf in fig. 6(b) is 60.22% infected from late blight disease (*Phytophthora Infestans*).

The method proposed in the paper can be applied to several important commercial applications. There are a number of edible leaves such as betel leaf (*Piper betle*), basket vine (*richostigma octandrum*), Lagos spinach (*Celosia argentea*) that are sold commercially. Similarly, there are a number of leaves such as neem (*Azadirachta indica*), tulsi *(Ocimum tenuiflorum)* that are used for medicinal purposes and are important ingredients of Ayurvedic medicine. The above proposed method provides an automatic method to compute the severity of damage on the leaf so that the infected leaves can be rejected before being sold or being used as an ingredient in medicines, cosmetics and chemicals.

Conflicts of Interest:

The authors declare that there is no conflict of interest regarding the publication of this article.

Funding Statement:

No funding was required or granted for the preparation of the article.

7. References

1. Bezdek, J. C., Ehrlich, R., and Full, W., "FCM: The fuzzy c-means clustering algorithm", Computers & Geosciences, Vol. 10 No. 2-3, 1984, pp.191-203
2. Zadeh, L.A., "Fuzzy Sets", Information and Control, Vol. 8 No.3, 1965. pp.338 – 353.
3. Chaira, T., "A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images", Applied Soft Computing, Vol 11 No.2, 2011, pp.1711-1717.
4. Atanassov, K.T., "Intuitionistic Fuzzy Sets", Fuzzy sets and Systems, Vol. 20 No.1, 1986, pp.87-96.
5. Maji, P. and Pal, S.K, "RFCM: A Hybrid Clustering Algorithm using rough and fuzzy set", Fundamenta Informaticae, Vol. 8 No. 4, 2007, pp.475-496.
6. Mitra, S., Banka, H. and Pedrycz, W., "Rough-Fuzzy Collaborative Clustering", IEEE Transactions on System, Man, and Cybernetics, Part B: Cybernetics, Vol. 36 No. 4, 2006, pp.795-805.
7. Pawlak, Z., "Rough sets", International Journal of Parallel Programming, Vol.11 No.5, 1982, pp.341-356.
8. Bhargava, R. Tripathy, B. K., Tripathy, A., Dhull, R., Verma, E. and Swarnalatha, P., "Rough intuitionistic fuzzy C-means algorithm and a comparative analysis", Proceedings of ACM Compute-2013, International Conference, SITE, VIT University, 21-22 August.
9. Ghosh, Soumi & Kumar, Sanjay. (2013). "Comparative Analysis of K-Means and Fuzzy C-Means Algorithms", International Journal of Advanced Computer Science and Applications.
10. Davis, D.L. and Bouldin, D.W., "A cluster separation measure", IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI Vol 1 No.2, 1979, pp.224 – 227.

11. Dunn, J. C., "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters", Journal of Cybernetics, Vol 3 No. 3, 1974, pp.32-57.

12. Dangl, J.L., Horvath, D.M., & Staskawicz, B.J. (2013). Pivoting the plant immune system from dissection to deployment. Science, 341(6147), 746-751

13. Bock, C.H., Parker, P.E., Cook, A.Z., & Gottwald, T.R. (2008). Visual rating and the use of image analysis for assessing different symptoms of citrus canker on grapefruit leaves. Plant Disease, 92(4), 530-541.

14. Bock, C.H., Cook, A.Z., Parker, P.E., & Gottwald, T.R. (2009). Automated image analysis of the severity of foliar citrus canker symptoms. Plant disease, 93(6), 660-665.

15. Price, T.V., Gross, R., Ho, W.J., & Osborne, C.F. (1993). A comparison of visual and digital image-processing methods in quantifying the severity of coffee leaf rust (Hemileia vasta-trix). Australian Journal of Experimental Agriculture, 33(1), 97-101.

16. Weizheng, S., Yachun, W., Zhanliang, C., & Hongda, W. (2008). Grading method of leaf spot disease based on image processing. In Proceedings of the IEEE international conference on Computer Science and Software Engineering (Vol. 6, pp. 491-494).

17. Camargo, A., & Smith, J.S. (2009). Image pattern classification for the identification of dis-ease causing agents in plants. Computers and Electronics in Agriculture, 66(2), 121-125.

18. Patil, S.B., & Bodhe, S.K. (2011). Leaf disease severity measurement using image processing. International Journal of Engineering and Technology, 3(5), 297-301

19. Chinta, Srujan & Tripathy, B.K. &, K. Govindarajulu. (2017). Kernelized Intuitionistic Fuzzy C-Means algorithms fused with Firefly Algorithm for Image Segmentation. 10.1109/ICMDCS.2017.8211702.

20. Chinta, Srujan & Jain, Abhay & Tripathy, B.K. (2017). Image Segmentation using Hybridized Firefly Algorithm and Intuitionistic Fuzzy C-Means.

21. Tripathy, B.K. and Agrawal, A, "A Comprehensive Analysis of Kernelized Hybrid Clustering Algorithms with Firefly and Fuzzy Firefly Algorithms", in 5th International Conference on Computational Intelligence in Data Mining (ICCIDM 2018), 2018. (Accepted).

22. Hughes, D., & Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060.