



Histological Image Segmentation with Fuzzy Clustering Method

Amani Abraheem Salim Alshoul¹, Fatma Kandemirli², Javad Rahebi^{3*}

¹ Department of Biomedical Engineering, Kastamonu University, Kastamonu, Turkey, albragate83@yahoo.com

² Department of Biomedical Engineering, Kastamonu University, Kastamonu, Turkey, (ORCID: 0000-0001-6097-2184), fkandemirli@yahoo.com

^{3*} Department of Software Engineering, Istanbul Ayvansaray University, Istanbul, Turkey, (ORCID: 0000-0001-9875-4860), cevatrahebi@ayvansaray.edu.tr

(First received December 2020 and in final form January 2021)

(DOI: 10.31590/ejosat.836329)

ATIF/REFERENCE: Alshoul, A. A. S., Kandemirli, F. & Rahebi, J. (2021). Histological Image Segmentation with Fuzzy Clustering Method. *European Journal of Science and Technology*, (22), 393-399.

Abstract

In this study, a method for nuclei image segmentation in histopathological images is proposed. This method is based on a fuzzy clustering method, which is pre-trained on an supplementary domain with very large labeled images, and coupled with an additional network composed of fully connected layers. In this study, Fuzzy Clustering Mean (FCM) was used for clustering and segmentation and the effective ways for breast cancer nuclei detection were obtained. Wherefore, fuzzy clustering means have been used to detect the centers of breast cancer nuclei, then the extracted centers were compared with the ground truth samples. It is worth mentioning, that this work passes through many experimental stages, of detection and segmentation by applying a combination of more than one effective method.

Keywords: Image segmentation, Fuzzy clustering mean, Nuclei image.

Bulanık Kümeleme Yöntemi ile Histolojik Görüntü Segmentasyonu

Öz

Bu çalışmada, histopatolojik görüntülerde çekirdek meme kanseri tespiti ve segmentasyonu için bir yaklaşım önerilmektedir. Bu yaklaşım, çok büyük etiketli görüntülere sahip bir yardımcı alan üzerinde önceden eğitilmiş ve tamamen bağlı katmanlardan oluşan ek bir ağ ile birleştirilen bulanık bir kümeleme yöntemine dayanmaktadır. Bu çalışmada, Fuzzy Clustering Mean (FCM) kümeleme ve segmentasyon için kullanılmış ve meme kanseri çekirdek tespiti için etkili yollar elde edilmiştir. Bu nedenle, göğüs kanseri çekirdeklerinin merkezlerini tespit etmek için bulanık bir kümeleme ortalaması kullanılmış, daha sonra çıkarılan merkezler kesin gerçek örnekleriyle karşılaştırılmıştır. Bu çalışmanın, birden fazla etkili yöntemin bir kombinasyonunu uygulayarak birçok deneysel, algılama ve bölümlenme aşamasından geçtiğini belirtmekte fayda var.

Anahtar Kelimeler: Görüntü bölümlenme, Bulanık kümeleme ortalaması, Çekirdek görüntüsü.

* cevatrahebi@ayvansaray.edu.tr

1. Introduction

Histopathology is the botanical or medical discipline intended to make a diagnosis by microscopic study of tissues (living or dead). It is still the most widely The word histopathology comes from the Greek histos (tissues) and pathos (suffering). It refers to the observation on a microscopic scale of living or dead tissue to identify the mechanisms, traces, or histological indices of manifestations of diseases (viral or not). In clinical medicine, histopathology refers more particularly to the examination of biopsies or surgical samples. The tools used in pathology are routinely for clinical diagnosis of cancer and other diseases [1][2][3].

2. Material and Method

2.1. Fuzzy Clustering Mean

After the introduction of fuzzy logic by Zadeh in 1965 a solution for this problem emerged, characterizing the similarity of each element to each of the groups. This is achieved by representing the similarity between an element and a group by a function, called a membership function, which takes values between zero and one. Values close to one indicate greater similarity, while values close to zero indicate less similarity. Therefore, the problem of fuzzy clustering boils down to finding such an optimal characterization [4][5].

Fuzzy clustering algorithms have been widely applied in different areas such as image processing, engineering systems, parameter estimation, among others.

2.1.1. Fuzzy partition

A fuzzy partition is a partition that characterizes the participation of each sample in all groups using membership functions that take values between zero and one. Also, they satisfy that for each sample the sum of their participation in each group is one. In this way, it is possible to translate the fuzzy clustering problem into finding an optimal fuzzy partition. Below is a more formal definition of this concept [6][7].

Let

$$M_{fc} = \{U \in \mathfrak{R}^{c \times n} | U = [u_{ij}]; u_{ij} \in [0,1] \forall i, j; \sum_{i=1}^c u_{ij} = 1 \forall j; \sum_{j=1}^n u_{ij} > 0 \forall i\}$$

be a subset of a Euclidean space of dimension s and c a positive integer greater than one. A fuzzy partition of X into c groups is a tuple of c membership functions $\mu = (\mu_1, \dots, \mu_c)$ that fulfill that:

$$0 \leq \mu_i(x) \leq 1, \forall i = 1, \dots, c \tag{1}$$

$$0 < \sum_{j=1}^n \mu_i(x_j) < 1, \forall i \tag{2}$$

$$\sum_{i=1}^c \mu_i(x_j) = 1, \forall j \tag{3}$$

Fuzzy partitions are represented as a matrix associating each row to one of the c groups and each column to one of the elements of X , in such a way that the value in row i and column j indicates the membership of element j to the group i . More formally, the set of fuzzy partitions can be defined as:

$$M_{fc} = \{U \in \mathfrak{R}^{c \times n} | U = [u_{ij}]; u_{ij} \in [0,1] \forall i, j; \sum_{i=1}^c u_{ij} = 1 \forall j; \sum_{j=1}^n u_{ij} > 0 \forall i\} \tag{4}$$

2.2. Fuzzy c -Means algorithms

The procedure, in a general sense, consists of iteratively minimizing this function until an optimal fuzzy portion is obtained [8].

Various grouping criteria for optimum fuzzy partition for X are proposed, but the most popular so far is associated with the least square error function:

$$J_m(U, v) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m d_{ik}^2 \tag{5}$$

The value d_{ik}^2 indicates the square distance between the elements of X and the centers of the groups and can be calculated using the following formula:

$$d_{ik}^2 = \|x_k - v_i\|_A^2 = (x_k - v_i)^T A (x_k - v_i) \tag{6}$$

where:

$X = (x_1, x_2, \dots, x_n) \subset \mathfrak{R}^s$ is the data,

$v_i = (v_{i1}, v_{i2}, \dots, v_{ic})$ is the center vector of group i ,

$\|A\|$ is the norm induced by A

A is a positive definite matrix of weights of dimensions $(n \times n)$.

In particular, if A is the identity matrix, d_{ik}^2 is the square of the Euclidean distance.

The weight associated with each square distance, $(u_{ik})^m$, is the m -th power of the degree of belonging of the k -th data to group i . When $m \rightarrow 1$ the optimal partition is getting closer to an exclusive partition, whereas when $m \rightarrow \infty$ the optimal partition approximates the matrix with all its values equal to $(1 / c)$. The values of m that are normally used are values in the interval $[1,30]$. Each selection of a particular value of m marks a specific Fuzzy c -Means algorithm.

With this in mind, the general procedure of the Fuzzy c -Means algorithms can be formalized in the following steps:

Set c, m, A and $\|k\|_A$. Choose an initial matrix $U^{(0)} \in M_{fc}$.

Calculate the centers of the groups with the formula

$$v_i = \frac{\sum_{k=1}^n (u_{ik})^{m x_k}}{\sum_{k=1}^n (u_{ik})^m}; 1 \leq i \leq c \quad (7)$$

Update the fuzzy partition matrix $U = [u_{ik}]$ with

$$u_{ik} = \left(\sum_{j=1}^c \left(\frac{d_{ijk}}{d_{jk}} \right)^{\frac{2}{m-1}} \right)^{-1}; 1 \leq k \leq n; 1 \leq i \leq c \quad (8)$$

If the stop criterion was met, terminate. If not, return to step 2.

Some of the stop criteria most used are:

- A maximum number of iterations
- That the variation in matrix U is very small:

$$\|U^{k+1} - U^k\| < \epsilon.$$

2.3. Possibilistic c -Means algorithms

The Possibilistic c -Means algorithms appear intending to solve the bad behavior of the Fuzzy c -Means algorithms when used in data sets with a lot of noise. 3 These algorithms are characterized by interpreting u_{ij} values as degrees of compatibility with the groups, rather than as probabilities of membership. To do this, the restriction of fuzzy partitions is relaxed, which forces the sum of the degrees of membership of an element to all groups to be one, requiring only that at least one of the degrees of membership be positive.

Therefore the constraints on the fuzzy partition definition could be rewritten as:

$$0 \leq \mu_i(x) \leq 1, \forall i = 1, \dots, c \quad (9)$$

$$0 < \sum_{j=1}^n \mu_i(x_j) < 1, \forall i \quad (10)$$

$$\max_j \mu_i(x_j) = 1, \forall j \quad (11)$$

One of the objectives most commonly used functions in these algorithms is as follows:

$$J_m(U, v, \eta) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m d_{ik}^2 + \sum_{i=1}^c \eta_i \sum_{k=1}^n (1 - u_{ik})^m \quad (12)$$

This is the same objective function of the Fuzzy c -Means algorithms with an added term that prevents the partition obtained from being the trivial solution where all membership values are equal to zero. Vector $\eta = (\eta_1, \eta_2, \dots, \eta_c)$ is a vector of positive values, where its values η_i denote the distance from the center of group i to which the degree of membership of an element is 0.5. These values determine the size and shape of their corresponding group and usually calculated using the following formula:

$$\eta_i = K \frac{\sum_{j=1}^n u_{ij}^m d_{ij}^2}{\sum_{j=1}^n u_{ij}^m} \quad (13)$$

where K is normally one.

The general procedure for these algorithms is:

Set c, m, A and $\|k\| A$. Choose an initial matrix $U^{(0)} \in M_{fc}$. Estimate the values of η .

Calculate the centers of the groups with the formula

$$v_i = \frac{\sum_{k=1}^n (u_{ik})^{m x_k}}{\sum_{k=1}^n (u_{ik})^m}; 1 \leq i \leq c \quad (14)$$

Update the fuzzy partition matrix $U = [u_{ik}]$ with

$$u_{ik} = \left(1 + \left(\frac{d_{ik}^2}{\eta_i} \right)^{\frac{1}{m-1}} \right)^{-1}; 1 \leq k \leq n; 1 \leq i \leq c \quad (15)$$

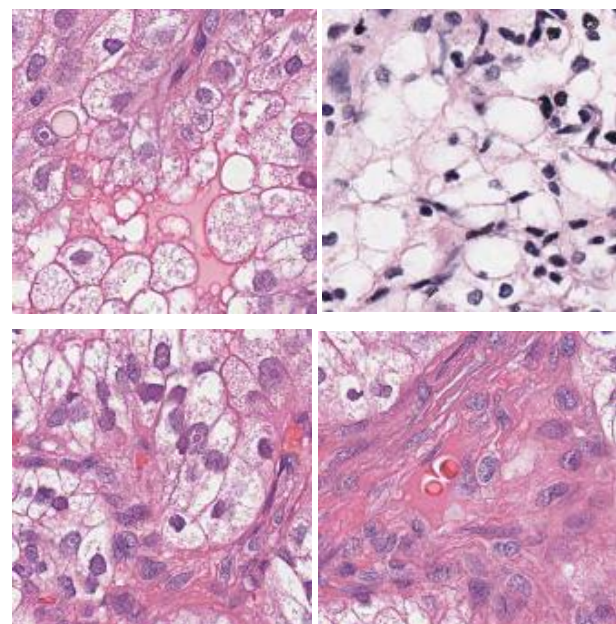
If the stop criterion was met, terminate. If not, return to step 2.

The stop criteria used by these algorithms are similar to those used by the Fuzzy c-Means algorithms.

3. Experimental Result

3.1. Database

In this study, we use the 810 microscopic slides of hematoxylin and eosin-stained samples of breast biopsy from PSB 2015 crowdsourced nuclei dataset. Breast tissue samples stained with hematoxylin and eosin hematoxylin and eosin. Each tile comes from a different patient. Some samples of these datasets are shown in Figure 1. The data set contains nuclei detection, nuclei segmentation, and original images. This data set contains 811 images from 811 peoples. Samples of this dataset images are shown in figure 1.



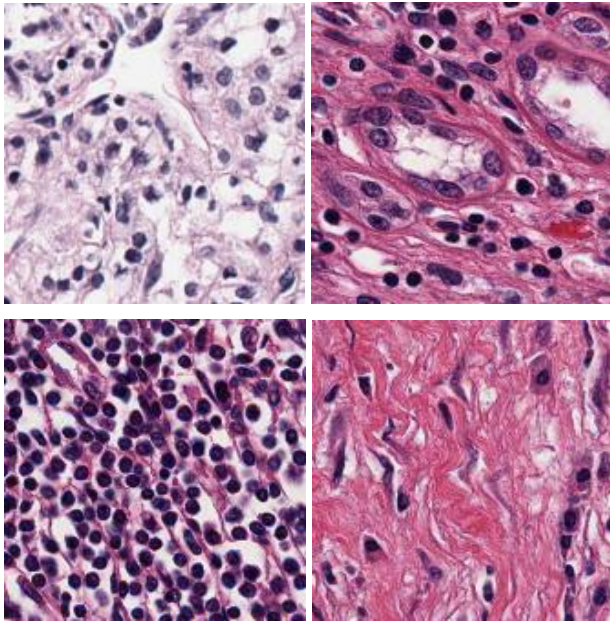


Figure 1. Sample of the dataset [9]

A colored image was not used in this study and the gray level image was used. Because color increases the complexity of the model and the processing of color images requires a long time. Color images usually have three channels. We have used grayscale images that have one channel.

3.2. Summary of proposed work

The flowchart in Figure 2 shows the steps of applying FCM.

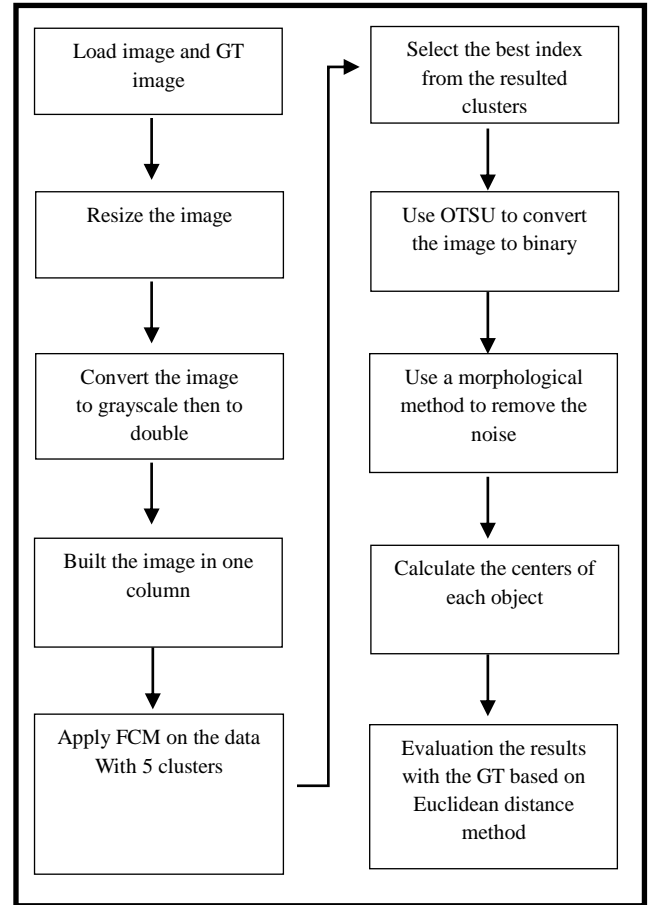


Figure 2. Flowchart of applying FCM

3.3. Binarization in the Proposed Method

The segmentation output of the proposed method is a grayscale image with 256 light intensities and needs to be binarized. For binarization, a light intensity threshold between 0 and 255 can be selected to binarize the segmentation image like Equation (16):

$$img = \begin{cases} 1 & \text{if } img(i,j) \geq \text{Threshold} \\ 0 & \text{if } img(i,j) < \text{Threshold} \end{cases} \quad (16)$$

Where MRI is a segmented image and the output of the HHO segmentation and FCM algorithms, while B is a binary image with only two light intensities. In the proposed method, the serious challenge is to find the appropriate threshold, which can be calculated through the Otsu algorithm. The light intensity histogram of image segmentation can be used for finding the Otsu threshold. In MATLAB, there is a command (multithresh) for this purpose. The tumor area can be extracted through image binarization. Then it can be compared with the real area.

3.4. Implementation and Analysis

The proposed method evaluated on data set explained in [10], which consists of T1 weighted, T2 weighted, proton

density (PD) MRI image. This dataset contains 10 cases and the proposed method tested for each case separately. For implementing of proposed method MATLAB R2019a version was used with Intel core i7 6 GHz processor and 4 GB RAM with Windows 10 operating system.

Several grayscale MRIs of the brain and nuclei were used to evaluate the proposed algorithm. These images included the areas of the brain affected by nuclei. The proposed method was employed to detect and extract the edges of these areas and compared them with manual areas, showing the areas with damaged brain tissues that were extracted by a physician or researchers from brain MRIs. First, the implementation and simulation parameters were adjusted in MATLAB. Then the input images were preprocessed and their noise was reduced through a filter such as the median. Then the proposed segmentation method was applied to them to calculate the average values of indicators such as accuracy and similarity for all images. After that, the outputs were determined and compared with other methods. Figure 3 shows three output samples of the proposed method for the diagnosis of nuclei on brain MRIs. In this test, it was decided to use three clusters like the previous test. The HHO population size and iteration were 10 and 50, respectively. The dimensions of the grayscale images were 512×512 pixels. Other parameters of the HHO algorithm were considered in the same way as the original paper. First, the image was preprocessed to eliminate possible noises. Then the brightness level was improved, and the image was segmented. After binarization of the image, Nuclei lesions appeared as white stains on the binary image.

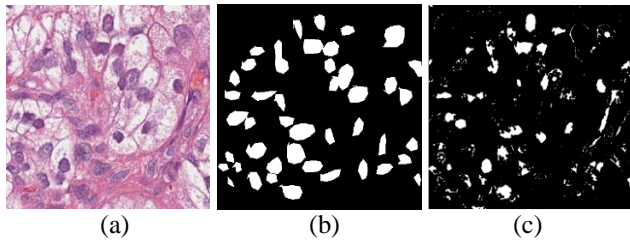


Figure 3. Various examples of segmentation, (a) original nuclei image, (b) Ground truth image, (c) Proposed method.

Different indices can be employed to evaluate the proposed algorithm. The most famous indices include similarity index (SI), accuracy, sensitivity, and specificity, for everyone of which it is necessary to calculate initial indices such as true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Then SI, accuracy, sensitivity, and specificity can be obtained from Equations (23), (24), (25), and (26) [11][12][13][14][15]:

$$SI = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (17)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (18)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (19)$$

$$Specificity = \frac{TN}{TN + FP} \quad (20)$$

The proposed indices are of the classification type and range between zero and one. If the value of an index is one, it shows the proper quality of classification. If it is expressed by percentage, proximity to 100% shows the accuracy of the proposed method in the segmentation and extraction of Nuclei areas. TP, TN, FP, and FN have different meanings and count the true or false pixels about Nuclei tissues or healthy tissues. Table 1 shows the meanings of these indices:

Table 1. Meanings of TP, TN, FP, and FN in evaluating the proposed method

Criterion	Definition
TP	Nuclei pixels that are correctly detected.
TN	Healthy pixels are correctly detected as healthy tissues.
FP	Pixels incorrectly detected as Nuclei pixels.
FN	Nuclei pixels incorrectly detected as healthy pixels.

To evaluate the proposed method, it can be analyzed based on similarity, accuracy, sensitivity, and specificity . For this purpose, 35 images were used. Figure 4 shows the average values of these indicators.

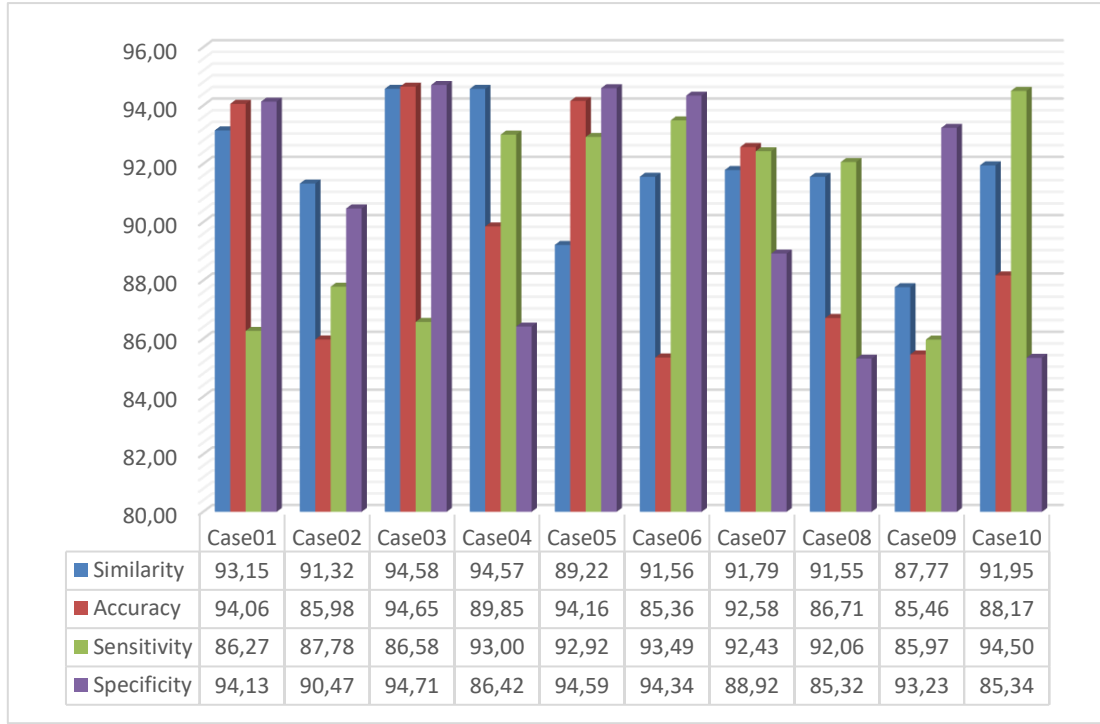


Figure 4. The average values for similarity, accuracy, sensitivity, and specificity in the proposed method for nuclei segmentation

According to the test results, it is fair to state that the similarity, accuracy, sensitivity, and specificity indicators of the proposed method were 94.28 %, 94.23 %, 89.56 %, and 93.34 %, respectively. Therefore, the extracted area resembled the manually detected area. Sensitivity shows what percentage of nuclei pixels were detected correctly, and specificity indicates what percentage of healthy pixels were detected correctly in nuclei. The test results were analyzed in MATLAB for the detection of nuclei lesions with the following findings:

- The proposed method properly detected the nuclei area and distinguished it from brain tissues.
- The proposed method properly analyzed the damaged areas of the brain using three clusters.

In this study, we studied nuclei center detection and classification methods in histopathological images. Nuclei center detection and nuclei classification system is designed for solving cluttering problem in pathological images diagnosis.

The fuzzy clustering method is frequently used in pattern recognition. It is based on minimization of the objective function:

$$J_{fcm} = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, 1 \leq m \leq \infty \quad (21)$$

The algorithm is composed of the following steps:

1. Initialize $U=[u_{ij}]$ matrix, $U^{(0)}$
2. At k-step: calculate the centers vectors $C^{(k)}=[c_j]$ with $U^{(k)}$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m} \quad (22)$$

3. Update $U^{(k)}, U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (23)$$

4. If $\|U^{(k+1)} - U^{(k)}\| < \mathcal{E}$ then STOP; otherwise return to step 2 where \mathcal{E} is a termination criterion between 0 and 1.

4. Conclusion

In this paper we applied the fuzzy clustering method to PH2 dataset containing hundreds of samples that provided the hospital dataset. For nuclei classification in histopathological images, in this work, we tested the performance of FCM to classify nuclei or non-nuclei in histopathological data set. Our proposed method observed that the match class works better in case of using FCM, where the highest class matches classification performance in this paper is reached to 95.3 %s and 96%. The future work can consider combining convolutional neural network with k-NN method and get the performance of classification between Euclidian distance method and another one established in machine learning algorithms.

References

- [1] M. Ilse, J. M. Tomczak, and M. Welling, “Deep multiple instance learning for digital histopathology,” in *Handbook of Medical Image Computing and Computer Assisted Intervention*, Elsevier, 2020, pp. 521–546.
- [2] A. ter Telgte *et al.*, “Histopathology of diffusion-weighted imaging-positive lesions in cerebral amyloid angiopathy,” *Acta Neuropathol.*, pp. 1–14, 2020.

- [3] R. B. Dettmeyer, *Forensic histopathology: fundamentals and perspectives*. Springer, 2018.
- [4] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The Fuzzy c-Means Clustering Algorithm Computer & Geosciences," *Volume*, vol. 10, pp. 2–3, 1984.
- [5] E. H. Ruspini, J. C. Bezdek, and J. M. Keller, "Fuzzy clustering: A historical perspective," *IEEE Comput. Intell. Mag.*, vol. 14, no. 1, pp. 45–55, 2019.
- [6] M.-S. Yang, "A survey of fuzzy clustering," *Math. Comput. Model.*, vol. 18, no. 11, pp. 1–16, 1993.
- [7] R. Bhukya and J. Gyani, "Survey on Fuzzy Associative Classifications Techniques and Their Performance Evaluation with Different Fuzzy Clustering Techniques Over Big Data," in *ICDSMLA 2019*, Springer, 2020, pp. 420–431.
- [8] X. Zhu, S. Zhang, Y. Zhu, W. Zheng, and Y. Yang, "Self-weighted multi-view fuzzy clustering," *ACM Trans. Knowl. Discov. from data*, vol. 14, no. 4, pp. 1–17, 2020.
- [9] H. Irshad *et al.*, "Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: evaluating experts, automated methods, and the crowd," in *Pacific symposium on biocomputing Co-chairs*, 2014, pp. 294–305.
- [10] S. Roy, D. Bhattacharyya, S. K. Bandyopadhyay, and T.-H. Kim, "An effective method for computerized prediction and segmentation of multiple sclerosis lesions in brain MRI," *Comput. Methods Programs Biomed.*, vol. 140, pp. 307–320, 2017.
- [11] H. E. Atlason, A. Love, S. Sigurdsson, V. Gudnason, and L. M. Ellingsen, "Unsupervised brain lesion segmentation from MRI using a convolutional autoencoder," in *Medical Imaging 2019: Image Processing*, 2019, vol. 10949, p. 109491H.
- [12] M. Sahnoun, F. Kallel, M. Dammak, C. Mhiri, K. Ben Mahfoudh, and A. Ben Hamida, "A comparative study of MRI contrast enhancement techniques based on Traditional Gamma Correction and Adaptive Gamma Correction: Case of multiple sclerosis pathology," in *2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 2018, pp. 1–7.
- [13] A. S. Abdullah, Y. E. Özok, and J. Rahebi, "A novel method for retinal optic disc detection using bat meta-heuristic algorithm," *Med. Biol. Eng. Comput.*, vol. 56, no. 11, pp. 2015–2024, 2018.
- [14] A. A. I. Mohamed, M. M. Ali, K. Nusrat, J. Rahebi, A. Sayiner, and F. Kandemirli, "Melanoma skin cancer segmentation with image region growing based on fuzzy clustering mean," *Int. J. Eng. Innov. Res.*, vol. 6, no. 2, p. 91C95, 2017.
- [15] I. A. Masoud Abdulhamid, A. Sahiner, and J. Rahebi, "New Auxiliary Function with Properties in Nonsmooth Global Optimization for Melanoma Skin Cancer Segmentation," *Biomed Res. Int.*, vol. 2020, 2020.