



# Diagnosis of Liver Cancer by Fuzzy Kmeans Clustering Based on Evidence Theory

Babak Fouladi Nia<sup>1</sup>, Abbas Karimi<sup>1\*</sup>, Faraneh Zarafshan<sup>2</sup>, Manochehr Kazemi<sup>3</sup>

<sup>1</sup>Department of Computer Engineering, Arak Branch, Islamic Azad University, Arak, Iran

<sup>2</sup>Department of Computer Engineering, Ashtian Branch, Islamic Azad University, Ashtian, Iran

<sup>3</sup>Department of Mathematics, Ashtian Branch, Islamic Azad University, Ashtian, Iran

[bf93@iau-arak.ac.ir](mailto:bf93@iau-arak.ac.ir), [akarimi@iau-arak.ac.ir](mailto:akarimi@iau-arak.ac.ir), [fzarafshan@aiau.ac.ir](mailto:fzarafshan@aiau.ac.ir), [m.kazemi@aiau.ac.ir](mailto:m.kazemi@aiau.ac.ir)

## Abstract

Liver cancer is one of the most common cancers that causes many deaths every year. In recent years, the risk of men and women getting liver cancer has increased by 40% and 23%, respectively. In order to identify a tumor in the liver, segmentation is performed on CT images. The use of data fusion methods in data mining techniques is one of the most practical methods to improve accuracy, which also has many applications in the field of medical image processing. Correct and efficient diagnosis of liver abnormalities leads to a significant reduction in human error and a more accurate diagnosis by physicians. This requires the use of methods based on automatic and semi-automatic detection. Combining clustering methods and considering cluster uncertainty is an appropriate tool in solving clustering problems in medical image processing, especially cancer diagnosis. The proposed method, in addition to having high accuracy, has a high convergence speed.

**Keywords:** Liver cancer, Evidence theory, Fuzzy system, KMEANS clustering

**Article history:** Received 25-Dec-2021; Revised 20-Jan-2022; Accepted 22-Feb-2022. Article Type: Research paper

© 2022 IAUCTB-IJSEE Science. All rights reserved

<https://dx.doi.org/10.30495/ijsee.2022.1948054.1163> & <https://dorl.net/dor/20.1001.1.22519246.2022.11.03.2.2>

## 1. Introduction

With the introduction of incorrect methods in human lifestyle, the use of inorganic and unhealthy foods and the increase in pressure and stress caused by daily activities in stressful environments, the number of patients with incurable diseases has increased. Tumors and cancers make up a large part of these diseases. Fortunately, with the advancement of medical science in today's societies and with the rapid and correct diagnosis of various diseases, many of these diseases can be treated and prevented. Human errors are an inseparable part of medical errors and have always caused problems. In some cases, the mistake is irreparable and causes irreversible damage, including a human life[1]. Therefore, eliminating human error is on the agenda of researchers. Over the past decade, expert systems in medicine have been introduced to minimize human error by utilizing artificial intelligence and machine vision. Medical equipment is one of the most vital items in the world and has grown significantly. This equipment is divided into two general parts, which are: diagnosis and treatment.

Most diagnostic devices use imaging[1]. These images can be formed by passing X-rays through the patient's body or by sound reflection and other things. The cost of this equipment is very high and in fact they are known as a significant trade in the world. In recent years, manufacturers of this equipment have provided diagnostic software. Liver cancer is one of the most common cancers and causes many deaths in the world. Medical image analysis is an application in this field that has a high computational complexity and requires automated and semi-automated systems. Image analysis methods for detecting abnormalities in the human body are performed by scanning images. In the field of computer-aided diagnosis (CAD), accurate and robust segmentation of liver tissue is used in medical imaging, which is a prerequisite for tumor identification and diagnosis [2]. Due to the various problems that exist in the structure of the edges of the liver and other adjacent organs (including the heart, stomach, and muscles), liver segmentation is a challenging task that has recently received a great

deal of attention in research. In addition, low contrast between the liver and adjacent organs impedes complete and accurate segmentation. In some cases, the liver is shown in various dimensions, which also causes problems in diagnosis and segmentation. Imaging techniques such as computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) are standard tools in the diagnosis of liver diseases such as cirrhosis, liver cancer, and liver failure. Among these techniques, CT images are often preferred by experts because they show more accurate and comprehensive information about the structural tissue of the liver. The cancer detection system consists of two parts: image segmentation and image classification[3]. The process of liver segmentation begins with a preprocessing stage and continues by dividing it into constituent regions. There are several methods for liver segmentation. Some of these methods are semi-automatic and some are fully automatic. Semi-automatic methods interact with the user and the desired area is identified by the user and used for subsequent processes, while in fully automatic methods, segmentation is done without user interaction. The various methods used for liver segmentation are based on threshold, statistical model, level set, region growing, active contour, and clustering. Due to the fact that manual liver segmentation is very time consuming, so in this article, methods based on automatic and semi-automatic segmentation are considered. The structure of the article is as follows: In Section 2, the types of methods available in preprocessing and segmentation of liver cancer are discussed. Section 3 describes the methods of diagnosing liver cancer and also in this section these methods are compared with each other. Conclusions are also given in Section 4. In the next section, the existing methods in the diagnosis of liver cancer are reviewed and the pros and cons of the methods are discussed.

## 2. Materials and methods

### A) Literature review

Pre-processing is an essential step required to achieve high accuracy in liver segmentation. CT images include salt and pepper noise, which reduces accuracy. The preprocessing process is such that the CT image is converted to a gray scale image and then 3\*3 median filter is used to reduce CT image noise[1]. In the method proposed in[2], three-dimensional anisotropic diffusion is used as a preprocessing step to eliminate image noise. In addition, this method preserves important parts of the image, such as edges, lines, or other details that are important for interpreting the image. The disadvantage of this method is that it requires a lot

of computational time due to its iterative process. In[3] in order to reduce noise, curvature anisotropic diffusion image filter is used to smooth CT images. Different dimensions of the filter for this purpose are presented in this article. In[4], image sampling, image smoothing and edge detection are defined as a preprocessing step. This article uses canny edge detection. The use of mean filter to reduce staircase edges and image smoothing and the use of Gaussian functions in image preprocessing to smooth images and obtain possible liver images have been suggested in[5]. Image segmentation is done by dividing the image into different regions according to a specific application. In medicine, liver segmentation is difficult because of the similarity of the various organs of the body, such as the kidneys, spleen, and pancreas, to the liver in the picture. General methods in liver image segmentation are: region growing, threshold-based, level set method, statistical model, active contour, clustering algorithm method, histogram-based method and gray level based method. The following is a description of this method. The region growing method produces reliable results on high-contrast images. These methods generally start with a small region called the seed point and continue by adding neighboring points that are similar to each other. This process continues until the segmented regions are accessible with high accuracy. This method has been used in[6]. The similarity value of each adjacent pixel with the seed point region is calculated. Pixels with the least difference are added to the growing region. This process is repeated until all adjacent unallocated pixels are compared with the growing region. The disadvantage of this method is that the wrong choice of seed point leads to an error in segmentation. In the fully automatic liver segmentation for high-contrast CT images, the use of the region growing is defined in such a way that the seed point is first determined based on prior information and knowledge and to prevent the growth of the heart region (Which is usually almost similar in intensity to the liver), using the lower connection of the left and right lung lobes to a surface, the heart is separated from the liver[7]. The liver is then separated from the heart to prevent wrong segmentation in this region. The disadvantage of this method is that if the liver has large lesions and as long as there are large lesions within this division, this method will have an unfavorable result in all criteria. An adaptive and automated region growing area algorithm is presented in[8]. Criteria of similarity are trained from the characteristics of the regions to be segmented. The parameters of this criterion are estimated from the candidate positions in this region. These regions are alternately selected in a random walks starting from the seed point, and

similarity criteria are constantly updated. The most important advantage of this method is the automatic detection of the seed point. The segmentation of the Liver Vasculature using the context based voting method is presented in[9]. To segment and identify liver Vasculature, regions are identified based on available characteristics. Quad-Tree analysis (QT) was used to identify soft tissue regions determined using the expectation minimization (EM) algorithm. Then, a primary region of the liver uses the classification and regression model (C&R) and also the threshold-based method is used to distinguish the liver region from the Initial intensity. The semi-automated method based on two-dimensional region growing with constraint-based knowledge is presented in[10]. Seed point and feature vectors calculated were used to identify the growing region. The authors of this article ensure the correctness of the obtained regions after using the constraint-based knowledge. The main advantage of the constraint-based knowledge method is the reduction of computational complexity. The method presented in[11] is based on the analysis of liver Vasculature. This method consists of several steps. In the first stage, the threshold-based region growing method was used to segmenting the Vasculature. The second stage involves the analysis of the structure of Vasculature such as the hepatic vein and the portal vein, which is done using the methods of theoretical diagrams. In the third step, the nearest neighbor approximation and the Laplace approximation of the segment are used. Finally, the shape of the Vasculature is obtained based on the skeleton and information about the diameter of the Vasculature. Liver segmentation has been used using the adaptive threshold method in[12]. In this method, morphological operations are used to obtain the final segmented liver region.

One of the advantages of this method is the different thresholds used for different regions in the image. This method uses a neural network to analyze liver tumor tissue on a CT image. This method has been followed by morphological operations such as erosion and dilation in order to preserve the structure of the liver and remove small tissues of other organs close to it in [13]. In the method based on automatic threshold segmentation of liver lesions in (2D-CT) abdominal images, pre-contrast and post-contrast images, the Intensity Brightness of the whole liver region varies according to the liver lesions and the type of lesion. CT images require at least two thresholds for segmentation of the liver and lesions. The average value of the image is used to determine LT (low threshold) and another statistical measure is the standard deviation for determining HT (high threshold). Usually, in pre-contrast images, the difference between the gray level in the liver and the liver lesion is much weaker and less than in post-

contrast images, which makes the segmentation more difficult. However, in this method, the boundaries of the lesion in pre-contrast images can be determined with acceptable accuracy. One of the advantages of this method is that it is able to segment different types of lesions in different sizes and shapes in both pre-contrast and post-contrast images, and it has also improved radiological analysis and diagnosis [14]. Adjacent abdominal organs can be removed using the histogram threshold sequence method. Then, using a binary morphological filter, the quality of liver segmentation is increased[15]. The advantages of this method are achieving topological changes and defining the problem in a higher dimension. One of the most important disadvantages of this method is its time consuming and large number of production regions[16]. Segmentation has been used by the authors using the level set method[17]. The main constraint of this method is the adjustment of the first Segmentation using a fast function obtained from a pixel classification algorithm. At the core of the algorithm is a level- set function that normally manages the separation and join of the liver border. Liver level set (LLS) is divided into two stages of preprocessing and level set with a hybrid energy minimization algorithm. In another method for liver segmentation using level set performed by Pan, a time-varying speed cumulative level set function was introduced to improve the sensitivity of detecting weak edges [18]. The method proposed by Pan is a semi-automatic method. In this method, the previous region of the liver based on anatomical knowledge is used to assist the segmentation process. Pan's two-dimensional algorithm places a small circle inside the liver region for each segment by initializing the curve. Therefore, if there is Separate point (not connected) in this curve, a user is required to initialize a circle for each of the disconnected regions. The model-based method is one of the most reliable methods in segmentation[19]. Many model-based methods use the Statistical Shape Model (SSM), which includes shape matching, shape display, and search algorithms. The process of making the shape model creates the corresponding points of the mark among all the shapes of the training sets [20]. The Statistical Shape Model of the liver is composed of 20 individual segmented CT databases. They have defined a geometric method based on distortion minimization. They define features as patches. The boundaries of these patches are created by specifying only a few points on each surface, and then the shortest path between them is calculated. Principal component analysis (PCA) has been used to analyze changes in a training data set to the corresponding levels of the liver[21]. The disadvantage of the SSM method is that it does not

give an acceptable result if the number of training datasets is very small. In the statistical metamorphosis model for liver CT segmentation, by Hermann, SSM trained 35 training databases to model the expected shape. The SSM context consists of 2500 Landmarks. They also propose a graph-based optimal surface detection. But this method is image dependent. Segmentation of liver vessels from high-contrast CT images using context-based voting (CBV) is presented in[9]. The region and nature of the liver vessels are described using context features based on that region. GVF Snake has been proposed for semi-automatic segmentation of the liver using the active contour method in[22]. This method consists of several steps. The first step is to reduce image noise using histogram equalization and isotropic diffusion filtering. In the next step, several manually selected points are connected together to form a hermite splines curve for the initial boundaries of the Snake. Finally, the best segmentation is based on the developed GVF Snake. GVF uses a clever edge detector to segment the liver to produce an edge map. Segmentation has been used by the Slice-by-Slice method in[23]. This method is based on unsupervised learning and consists of three steps. In the preprocessing stage, the input CT images are detected by estimating the liver position (ELP) and thresholding is performed by multilevel execution on the estimated position of the liver. The proposed method uses a multi-scale morphological filter recursively with region-labeled and clustered. The final contour is formed by a Gradient-label Map search algorithm[17]. Platero used a method based on three-dimensional isotropic diffusion processing to liver segmentation to diagnose liver lesions[24]. In this paper, histogram analysis, morphological processing and gradual evolution have been used to form an active contour. Sudhamani et al. showed that segmentation of CT images of the liver helps to analyze the occurrence of cancerous tumors and classify tumors from images[20]. They used the region growing method to diagnose the cancerous region. They also used the pool segmentation method to segment the contours of the tumor region. Then, tissue features are extracted by GLCM from the segmented region. Finally, cancerous and non-cancerous images are classified using support vector machine (SVM). In[25], various machine learning algorithms such as KNN, MLP neural network, SVM have been used. It has been shown that among these algorithms, the SVM method has a better performance and has an accuracy of 95.29% in cancer diagnosis. Masuda et al. introduce a new method that automatically detects cancer in CT images. Their method consists of two steps. In the first stage, the regions where there is a possibility of a tumor are extracted by the EM / MPM algorithm. The EM / MPM algorithm

uses the Intensity brightness values of the pixels and the labeling of the neighboring pixels in that region for classification. In the second stage, regions that have been misdiagnosed as cancer are removed by shape information. In this article, structural shape information has been used to correctly diagnose cancerous regions. Due to the fact that the structural shape of the tumor is almost spherical, so it is necessary to perform the pattern matching operation, in other words, this spherical structure is compared with the selected regions. In this article, the border points of the liver are not considered as tumors. The method used in this article has less error in diagnosing cancerous regions than the convolution method[26]. Rodrigues et al. proposed a user-interactive method based on segmenting an image into geometric shapes using a pool-like algorithm. In this paper, CT images are evaluated using the DSC criterion with an average of 87% [27]. In [28], a semi-automated algorithm is proposed that has an effective trend in the segmentation of CT images. The advantages of this method are reduced decisions to diagnose cancer in the region, reduced computational time and resistance to noisy images. The proposed method first converts the image to a gray scale, then preprocessing methods are applied to the images to reduce noise and increase image quality. K-means clustering is then used and the appropriate clusters are selected according to a threshold limit. Now for these clusters, haar wavelet transform is used. Finally, the pixels of these clusters as well as their standard deviation are used. In[29], to increase the quality of CT images, two different algorithms have been used, which are: Contrast Limited Adaptive Histogram Equalization (CLAHE) and the Constrained Variable Histogram Equalization (CVHE). CLAHE enhances the quality of the tumor region and detects a normal liver in this algorithm. Liver cancer is also diagnosed by the CVHE algorithm. SVM was also used to classify CT images of the liver. The accuracy in this method is 97%. In [30], a method is presented that in the MRF segmentation section, the process of assigning labeling to image pixels is performed. In other words, pixels that are more similar to each other are labeled with the same value. This will make it easier for the proposed method to work in the next steps. In other words, preprocessing is done on the image. In the Shape Analysis & Correction step, in order to find the tumor regions, different views of (spherical) shapes are matched on the image. Therefore, regions are found that are very similar to geometric shapes. According to these regions, a pyramid is built. In order to have the correct segmentation of the liver tumor part, in the tumor segmentation and classification stage, liver tumors based on graph cutting method have been used. The histogram is then calculated for each region. The support vector

machine is also used to train the algorithm. In [31] Ahmadi et al., Segmentation of liver blood vessels was performed using a combination of fcm and ga algorithms on 20 ct images and the values of accuracy, sensitivity, specificity, and cpu time were 91%, 83.62% and 94.11%, And 27.17 respectively was obtained. High convergence speed is another achievement of this paper. In [32]Fouladi et al., A combined segmentation algorithm using fcm and cuckoo optimization algorithm was presented and compared with fcm-ga hybrid algorithm. The results showed that the values of sensitivity, specificity and accuracy were 84.24%, 94.30% and 91.36%, respectively. This method solves the low sensitivity of the fcm method to the initial values of class centers. High convergence speed is another advantage of this method and it does not have the problem of threshold based algorithms in selecting suitable threshold values. A semi-automatic method consisting of three steps to overcome leakage and over-segmentation problems is used in the paper [33] in which the main input images are pre-processed by a series of methods to obtain binary images. A number of seed points are then set in the binary image for the region growing method to obtain the initial contour of the liver. The active contour model of the new level set is finally made to obtain the initial segmentation of the liver. In the first step, the slope gradient magnitude of original images is smoothed by the Gaussian function to eliminate noise and increase the edges. In general, a binary method is presented which, based on the information gradient of image, obtains the optimal threshold value without artificially adjusting the parameters. In the third step, a new signed pressure function integrates local and global information and automatically assigns local and global information rates based on image gradient information. In Article ([34]), in the preprocessing phase, image noise is removed and then the edge of the input image is sharpened using a frequency-based edge sharpening method. In the image segmentation phase, the output of the previous steps is considered as the input of the Outline Preservation Based Segmentation (OPBS) algorithm for segmentation. In the next step, the classification is done based on the features extracted from the previous step. The method used in this section is a similarity search based hybrid classification method. In paper [35], the combination of seed point selection and connected component with watershed algorithm is used in order to achieve better accuracy and robustness in CT scan images. The results show a high accuracy of 98.6%. In article [36], a novel external force method called gradient vector flow over manifold (GVFOM) is used. The model uses GVF vector as a two-dimensional manifold embedded in a four-dimensional Euclidean space,

which leads to the generalization of the Laplace operator in Euclidean space, and two external force components have been used to improve the GVF snake properties.

#### B) Clustering by kmeans method

The developed k-means clustering algorithm is used for liver segmentation in[37] . The k-means algorithm is one of the easiest unsupervised learning algorithms that categorizes a given image into a certain number of clusters. The disadvantage of this clustering method is that the cyst region is not extracted properly. To improve performance, morphological operators such as erosion and dilation at the output of the K-means algorithm are used. In the segmentation of cyst regions in liver images, the K-means clustering method has a better performance than the region growing method. Fuzzy C-means (FCM) clustering method is not very effective for liver tumors segmentation with noisy or outlier points and different clusters with different sizes[6] . To overcome these problems, an alternative FCM clustering algorithm (AFCM) has been used[6] . AFCM is a clustering-based segmentation algorithm that modifies similar pixels that are repeated at the centers of the clusters for defined iterations. This makes segmentation more effective.[38] Presents a multi-instance learning (MLI) method for the diagnosis of liver cancer using abdominal CT images that emphasizes instance optimization (IO) and SVD parameters by a hybrid particle swarm optimization (PSO) algorithm And local optimization, optimized. The MIL method in the diagnosis of liver cancer can solve the problem of multiple regions of the desired classification. The proposed method consists of two main steps. The first step is to obtain key instances through IO by tissue features and classification threshold in instance classification and the second step is to predict unknown instances with key instances and classification threshold.

#### C) FCM clustering algorithm

Clustering algorithms are known as unsupervised methods because they do not require training data. Similar to image segmentation, clustering involves dividing a set of objects into groups (clusters) so that objects in the same group are more similar than objects in other groups. Often, the similarity is determined by a distance scale such as the Euclidean distance or the Manhattan distance. With a definite number of branches  $k$  and a number of data  $N$ , the matrix

$$U_{K \times N} = [u_{ik}], k = 1, \dots, C, \text{ and } i = 1, \dots, N \quad (1)$$

Displays the data set, as  $u_{ik}$  describes the  $x_i$  data in the  $ck$  cluster. Clustering is hard if  $u_{ik}$  has

only two values 1 (belongs to) or 0 (does not belong to) that are set by Boolean membership functions, and clustering is fuzzy if  $u_{ik}$  has values between zero and one that is valued by continuous membership functions. If  $v_k$  is the center of the  $c_k$  cluster then  $v_k$  is calculated as follows:

$$v_k = \sum_{i=1}^N u_{ik} x_i / \sum_{i=1}^N u_{ik}, k = 1, \dots, C \quad (2)$$

The most famous hard clustering algorithm is the K-Means algorithm. In this algorithm, the  $u_{ik}$  membership value must meet the following expression:

$$\begin{aligned} & \forall k, \forall i, u_{ik} \\ & = \{0,1\}, \forall i, \sum_{k=1}^c u_{ik} \\ & = 1, \text{ and } \forall k, 0 < \sum_{i=1}^n u_{ki} < N. \end{aligned} \quad (3)$$

Using the search of a  $d_{ki}$  distance function for example, the Euclidean distance is equal to:

$$d_{ik} = |X_i - V_k| \quad (4)$$

Then we need to find a  $U$  that satisfies the membership constraint in Equation (5) and minimizes the following function.

$$J(U, V) = \sum_{k=1}^c \sum_{x_i \in c_k} d_{ki}^2. \quad (5)$$

A common way to find  $U$  is to use the iterative method proposed by Lloyd. The Fuzzy c-means algorithm is an extension of k-Means. Unlike hard membership in k-Means, this algorithm allows data points to belong to more than one cluster, which indicates a degree of membership confidence in each cluster.  $u_{ik}$  membership must meet:

$$\forall k, \forall i, 0 < u_{ik} < 1, \forall i, \sum_{k=1}^c u_{ik} = 1, \text{ and } \forall k, 0 < \sum_{i=1}^n u_{ki} < N. \quad (6)$$

Where  $\{v_k\}_{k=1}^c$  are the centers or initial instances of the clusters, and the array  $\{u_{ki}\}_{k=1}^c (= U)$  represents a satisfactory matrix clustering:

$$U \in \{u_{ik} \in [0,1] \mid \sum_{k=1}^c u_{ik} = 1, \forall k \text{ and } 0 < \sum_{i=1}^n u_{ik} < N, \forall i\} \quad (7)$$

So that  $m \geq 1$  controls the degree of fuzziness. If  $m$  is close to 1, then it gives the cluster with the closest center to the point more weight than the other clusters. The smaller  $m$  we are approaching the fuzzy state, and the larger  $m$  we have a hard state. Unlike hard membership in k-means, membership amounts and cluster centers are calculated as follows:

$$u_{ik}^{(t+1)} = \frac{d_{ik}^{-2/(m-1)}}{\sum_{j=1}^c d_{ij}^{-2/(m-1)}} \quad (8)$$

$$c_k^{(t+1)} = \frac{\sum_{i=1}^n (u_{ik}^{(t+1)})^m x_i}{\sum_{i=1}^n (u_{ik}^{(t+1)})^m} \quad (9)$$

The FCM clustering algorithm was investigated in this section. This algorithm does not work well for noisy data. In the following, we will review the proposed method.

#### D) PCM clustering algorithm

A fuzzy PCM approach is proposed to make FCM resistant to Outlier and noisy data. The objective function of the algorithm is defined as relation (10).

$$\begin{aligned} & J_f(X, U_f, C) \\ & = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m d_{ik}^2 \\ & + \sum_{i=1}^c \eta_i \sum_{j=1}^n (1 - u_{ij})^m \end{aligned} \quad (10)$$

Adding the second phrase to the above relation prevents all memberships from tending to zero. Because by decreasing the value of  $u_{ij}$  in the first expression, it's inverse, ie  $(1 - u_{ij})$  in the second expression will increase. In the above relation  $\eta_i > 0$ , ( $i = 1, \dots, c$ ) is a constant that must be defined for each cluster to establish a balance between the two expressions. That is,  $\eta_i$  specifies the last fuzzy boundary point of each cluster. So with this parameter, the extent of cluster expansion can be controlled. To determine the shape of the cluster, the value of  $\eta_i$  must be estimated:

$$\eta_i = \frac{\sum_{i=1}^n u_{ik}^m d_{ij}^2}{\sum_{i=1}^n u_{ik}^m} \quad (11)$$

In the PCM method, by removing the normalization constraint, data belonging  $i$  depends only on the distance between the same data and cluster  $j$  (and is independent of the distance from the data to other clusters). The amount of data belonging to each cluster is also controlled by the  $\eta_i$  parameter. We will use this idea in our proposed method below.

#### E) KPFCM clustering algorithm

In the previous sections, the FCM and PCM clustering algorithms were investigated. These algorithms are in the category of kernelless clustering algorithms. The following is a brief overview of kernel-based clustering methods. An important category of new methods in machine vision and artificial intelligence in general is kernel-based methods. The effectiveness of these methods is well known when combined with other machine vision algorithms and artificial intelligence. In fuzzy clustering methods, the use of kernel trick methods enables the algorithm to remove noise and analyze the data. The KPFCM algorithm is a generalization of the FCM and PCM algorithms. The KPFCM

algorithm is the same as the kernel-based fuzzy PCM algorithm. The objective function of the PFCM algorithm is as follows.

$$J_{m,\eta}(u, T, V) = \sum_{i=1}^c \sum_{k=1}^n (au_{ik}^m + bt_{ik}^n) D_{ik}^2 + \sum_{i=1}^c \gamma_i \sum_{k=1}^n (1 - t_{ik})^m \quad (12)$$

$$\begin{aligned} \|\Phi(x_k) - [\Phi(v)_i]\|^2 &= (\Phi(x_k) - [\Phi(v)_i])^T (\Phi(x_k) \\ &\quad - [\Phi(v)_i]) \\ &= \Phi(x_k)^T \Phi(x_k) - \Phi(v)_i^T \Phi(x_k) \\ &\quad - \Phi(x_k)^T \Phi(v)_i + \Phi(v)_i^T \Phi(v)_i \\ &= K(x_k, x_k) + K(v_i, v_i) - 2K(x_k, v_i) \end{aligned} \quad (13)$$

Only the Gaussian kernel is used in fuzzy clustering, because for the Gaussian kernel  $K(x, x) = 1$ , a Gaussian kernel can be defined as follows:

So we have:

$$\|\Phi(x_k) - \Phi(v_i)\|^2 = 2(1 - K(x_k, v_i)) \quad (14)$$

By placing the kernel function (14) in relation (12) we have:

$$J_{m,\eta}(U, T, V) = 2 \sum_{i=1}^c \sum_{k=1}^n (au_{ik}^m + bt_{ik}^n) (1 - k(x_k, v_i)) + \sum_{i=1}^c \gamma_i \sum_{k=1}^n (1 - t_{ik})^m \quad (15)$$

In this relation  $a, b > 0$  as well as  $m, \eta > 1$  and  $\gamma_i > 0$  with given constraints,  $J_{m,\eta}^{\phi}$  can be solved using Lagrange multiplication. Here the relation is omitted due to Article space constraints. Therefore, the reduction of relation (15) leads to the following necessary conditions, which of course are not sufficient, and for  $J_{m,\eta}^{\phi}$  a local minimum is obtained according to the following relations.

$$u_{ik} = \frac{(1 - K(x_k, v_i))^{-1/(m-1)}}{\sum_{j=1}^c (1 - K(x_k, v_j))^{-1/(m-1)}} \forall i, k \quad (16)$$

$$t_{ik} = \left[ 1 + \frac{2b}{\gamma_i} (1 - k(x_k, v_i))^{-1/(m-1)} \right]^{-1} \forall i, k \quad (17)$$

And the relation of calculating the centers is as follows:

$$v_i = \frac{\sum_{k=1}^n (au_{ik}^m + bt_{ik}^n) k(x_k, v_i) X_k}{\sum_{k=1}^n (au_{ik}^m + bt_{ik}^n) k(x_k, v_i)} \quad (18)$$

Finally, the value of  $\gamma_i$  in the KPFCM objective function of Equation (15) is calculated as follows.

$$\gamma_i = \frac{2 \sum_{k=1}^n (1 - k(x_k, v_i))}{\sum_{k=1}^n u_{ik}^m} \quad (19)$$

The KPFCM algorithm was investigated in this section.

### 3. The results and Discussion

#### A) Proposed clustering algorithm

In clustering, it is usually assumed that an instance is a member of a cluster or not. In clustering with two classes, we have two colors, black and

white. In cancer pixel clustering, two classes of Belonging to a cancer clusters and healthy clusters are considered for each pixel. If it belongs to a cancerous cluster, it is marked in white, and if it belongs to a healthy cluster, it is marked in black. But in the world of digital image processing, they are associated with high ambiguity and uncertainty. Therefore, there are situations where although two Cluster heads are considered for the clustering problem, there is so much ambiguity in the instance that it practically does not belong to any set. Under these circumstances, this paper proposes a solution to this problem and that is Unknown hypothesis. In other words, for the space of clustering in the case of liver cancer, three hypotheses are considered for one pixel: cancerous pixel, healthy pixel, unknown pixel.

In this paper, a fuzzy system-based clustering method is proposed in which clustering is performed based on the Three hypotheses expressed in the article. The probability of belonging to each cluster is also calculated in this method. The proposed clustering algorithm is based on kmeans clustering. In this method, belonging to each cluster head is determined based on Euclidean distance from each cluster head. The basis of the clustering algorithm is based on kmeans clustering. First, two head clusters are randomly selected in the instances space. The distance from each instance to the head clusters should be divided by the maximum distance of instance from the head cluster to normalize the distance values between zero and one. In the next step, for fuzzy clustering, it is necessary to determine the inputs and outputs of the fuzzy system. The inputs of the fuzzy system are the distance from the first and second headings, respectively. The output of the system is Possibility of belonging to headings one and two as well as the unknown value.

In this paper, a proposed fuzzy algorithm is used to combine clustering. Where according to the design of the fuzzy system, the degree of confidence in each hypothesis is determined. In this method, the probability of belonging to each cluster is considered as a parameter to reject or confirm the hypothesis and the fuzzy system is designed. For this purpose, the following fuzzy system is used to trust the probability of occurrence of a hypothesis. The output of a fuzzy system is the probability of being rejected, confirmed, or ignored in relation to a hypothesis. To validate the proposed algorithm, the dataset images available on the website (<https://sliver07.grand-challenge.org/>) have been used. The designed fuzzy system is shown in the figure below:

In the proposed fuzzy system, the inputs of the fuzzy system (Figure 1) are obtained according to the probability percentage of each cluster. The

output of the fuzzy system (Figure 2) determines the proof or rejection of a hypothesis and can use the raw data obtained from the basic clustering to determine the degree of belonging of each pixel to each cluster based on the probability percentage between 0 and 100, so that the probability What hypothesis does each observed pixel belong to and also the degree of this probability is determined.

The output values of the fuzzy system are considered as the input of the Dempster Shafer algorithm. In the Dempster Shafer algorithm, according to the relations 8 to 22, the values of the information combination as well as the unknown hypotheses and the new hypothesis are made. And then, based on the point obtained in the fuzzy set, the values belonging to each heading are calculated by interpolating. The rules of the fuzzy system are based on try and error and the logical rules governing clustering. Euclidean distance is the criterion for belonging to a head cluster. If an instance is closer to one head cluster than another, it belongs to that header more.

Practically, we can refer to the problem of clustering for the diagnosis of liver tumors, whose common assumptions are that each pixel is cancerous or healthy. With the help of the proposed method, the unknown hypothesis can be added to the space of the initial hypotheses. Raw data obtained from different clustering methods must be converted to a format that can be used for the data fusion unit. Raw data needs to be converted to probability format (confirmation or rejection of a hypothesis) in order to be used in the proposed method. In the following, the methods of preparing raw data and converting it into a probabilistic format for use in information fusion algorithms are discussed. The black part means that the tissue is healthy; the gray part means ignorance, the white part means the patient.

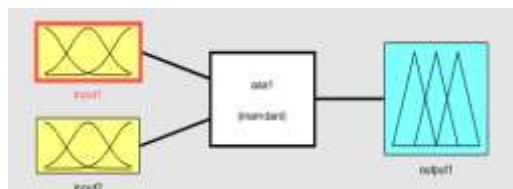


Fig. 1. Proposed fuzzy system

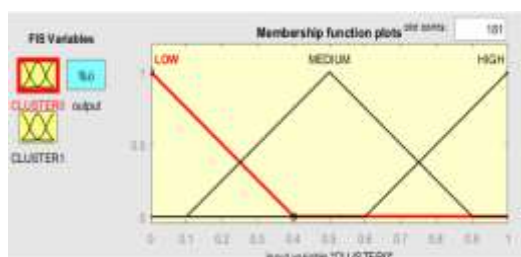


Fig. 2. The first input of fuzzy system membership functions

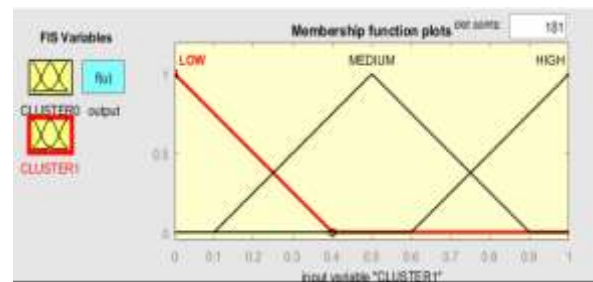


Fig. 3. The second input of fuzzy system membership functions

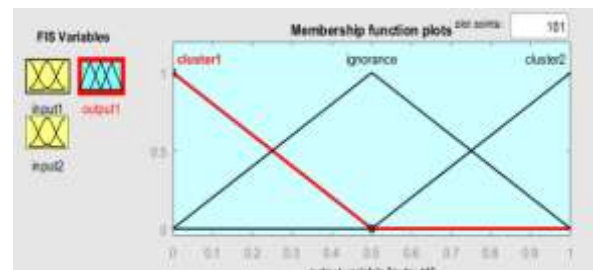


Fig. 4. The output of fuzzy system membership functions

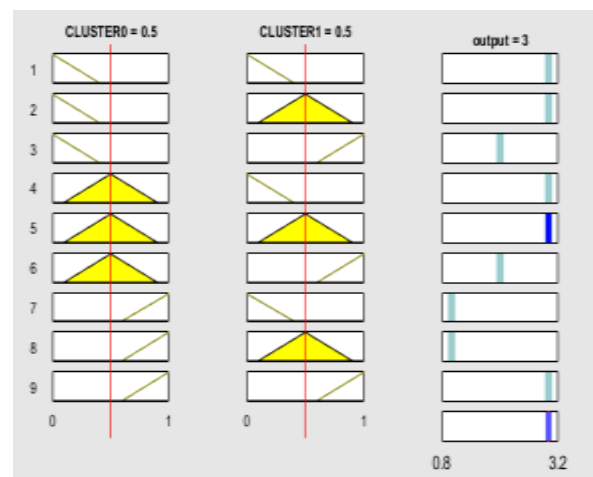


Fig. 5. fuzzy system rules

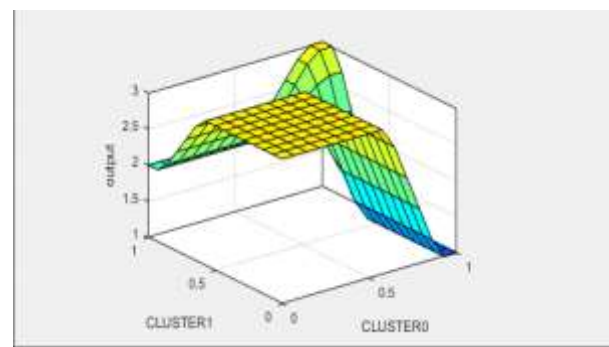


Fig. 6. fuzzy system graph



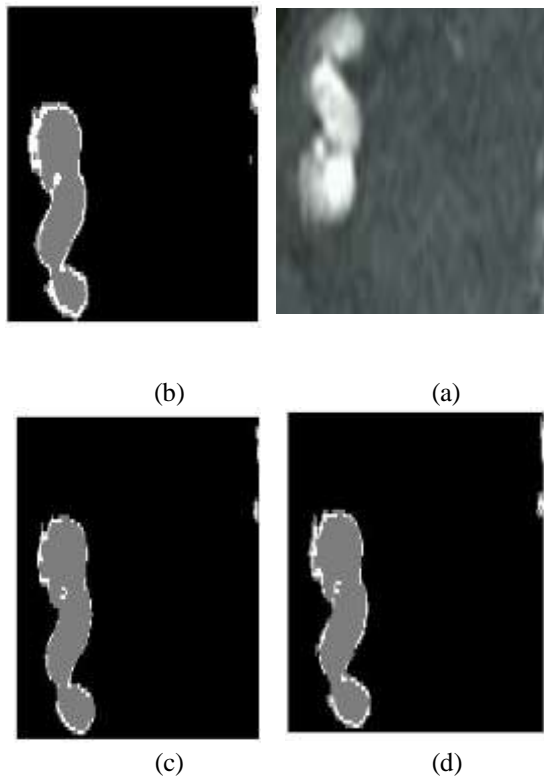


Fig. 7. Sample image (a) Cancer part clustering, (b) The proposed fuzzy kmeans method with Euclidean distance, (c) Proposed fuzzy kmeans method with Manhattan distance, (d) Proposed fuzzy kmeans method with Euclidean-Manhattan combined distance

#### B) Evaluation by other methods

In this section, experiments have been performed on two types of real clinical and virtual images. To validate the proposed algorithm, the proposed algorithm is compared with six fuzzy clustering algorithms. The accuracy of segmentation in this article is based on the JS criterion, which is as follows:

$$JS(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} \quad (20)$$

Where  $S_1$  represents a set of pixels belonging to a class that are found by a particular method. And  $S_2$  represents a set of pixels in a class very similar to a reference segmented image.

The proposed algorithm can be used for other clustering methods. Different distances in kmeans algorithm can be used and clustering can be done based on it. Thus, kmeans clustering methods with Manhattan and Euclidean distances are compared and shown by FCM method. The advantage of the proposed method over other methods of clustering algorithm is in reducing uncertainty and clustering error in medical images. It can also be used with the help of the proposed algorithm for hybrid clustering methods. This method is especially suitable for the Schaffer-Dempster method because it needs an unknown hypothesis.

Table.1.

Comparing the accuracy of segmentation of the proposed algorithm with other methods

Row	Algorithm	segmentation rate for Gaussian noise 2%			Average accuracy
		White matter	Gray matter	Cerebro Spinal Fluid	
1	K-Means	65	60	63	62.66
2	FCM	69	65	66	66.66
3	PCM	67	68	60	65.00
4	PFCM	74	76	71	73.66
5	SFCM	92	86	73	83.66
6	KPFCM	95	88	83	88.66
7	IKPFCM	95	91	89	91.66

#### 4. Conclusion

Correct and efficient diagnosis of liver cancer results in successful treatment. In this paper, fuzzy clustering methods based on evidence theory in the diagnosis of liver cancer are studied. The method presented in this paper, in addition to having high accuracy, also has a high convergence speed. In semi-automatic methods, user interaction is required and the desired area is identified by the user and used for subsequent image processing processes. While in fully automatic methods, segmentation is done without user interaction. Finally, the features and disadvantages of these methods are evaluated. Also, Kmeans clustering with Euclidean and Manhattan distance criteria was used for the final hybrid clustering based on evidence theory, which led to the improvement of clustering accuracy in the uncertainty of the issue under discussion. The advantage of the proposed method over other methods of clustering algorithm is in reducing uncertainty and clustering error in medical images. It can also be used with the help of the proposed algorithm for hybrid clustering methods. This method is especially suitable for the Dempster-Shafer method because it needs an unknown hypothesis.

#### References

- [1] M. U. Akram, A. Khanum, and K. Iqbal, "An automated system for liver ct

- enhancement and segmentation," *Graphics, Vision and Image Processing GVIP*, vol. 10, no. 4, pp. 17-22, 2010.
- [2] W. Nural, J. Yussof, and H. Burkhardt, "3D anisotropic diffusion for liver segmentation. World Acad Sci," *Eng Technol*, vol. 57, 2009.
- [3] Y. Qi *et al.*, "Semi-automatic segmentation of liver tumors from CT scans using Bayesian rule-based 3D region growing," in *MICCAI workshop*, 2008, vol. 41, no. 43, p. 201.
- [4] X. Zhang, J. Tian, K. Deng, Y. Wu, and X. Li, "Automatic liver segmentation using a statistical shape model with optimal surface detection," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 10, pp. 2622-2626, 2010.
- [5] Y. Chen, Z. Wang, and W. Zhao, "Liver segmentation in CT images using Chan-Vese model," in *2009 First International Conference on Information Science and Engineering*, 2009: IEEE, pp. 3669-3672.
- [6] S. Kumar, R. Moni, and J. Rajeesh, "Automatic segmentation of liver and tumor for CAD of liver," *Journal of advances in information technology*, vol. 2, no. 1, pp. 63-70, 2011.
- [7] L. Rusko, G. Bekes, G. Nemeth, and M. Fidrich, "Fully automatic liver segmentation for contrast-enhanced CT images," *MICCAI Wshp. 3D Segmentation in the Clinic: A Grand Challenge*, vol. 2, no. 7, 2007.
- [8] R. Pohle and K. D. Toennies, "Segmentation of medical images using adaptive region growing," in *Medical Imaging 2001: Image Processing*, 2001, vol. 4322: International Society for Optics and Photonics, pp. 1337-1346.
- [9] Y. Chi *et al.*, "Segmentation of liver vasculature from contrast enhanced CT images using context-based voting," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 8, pp. 2144-2153, 2010.
- [10] D. Wong *et al.*, "A semi-automated method for liver tumor segmentation based on 2D region growing with knowledge-based constraints," in *MICCAI workshop*, 2008, vol. 41, no. 43, p. 159.
- [11] D. Selle, B. Preim, A. Schenk, and H.-O. Peitgen, "Analysis of vasculature for liver surgical planning," *IEEE transactions on medical imaging*, vol. 21, no. 11, pp. 1344-1357, 2002.
- [12] KumarSS and D. Moni, "Diagnosis of Liver Tumor from CT Images Using Fast Discrete Curvelet Transform," *International Journal of Computer Applications*, vol. CASCT, 08/20 2010, doi: 10.5120/999-34.
- [13] K. Mala, V. Sadasivam, and S. Alagappan, "Neural network based texture analysis of liver tumor from computed tomography images," *International Journal of Biomedical Sciences*, vol. 2, no. 1, pp. 33-60, 2006.
- [14] A. A. Moghe, J. Singhai, and S. Shrivastava, "Automatic threshold based liver lesion segmentation in abdominal 2D-CT images," *International Journal of Image Processing (IJIP)*, vol. 5, no. 2, p. 166, 2011.
- [15] O. Zayane, B. Jouini, and M. A. Mahjoub, "Automatic liver segmentation method in CT images," *Canadian Journal on Image Processing & Computer Vision*, vol. 2, no. 8, pp. 92-85, 2011.
- [16] J. A. Sethian, *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*. Cambridge university press, 1999.
- [17] S.-J. Lim, Y.-Y. Jeong, and Y.-S. Ho, "Segmentation of the liver using the deformable contour method on CT images," in *Pacific-Rim Conference on Multimedia*, 2005: Springer, pp. 570-581.
- [18] S. Pan and B. M. Dawant, "Automatic 3D segmentation of the liver from abdominal CT images: a level-set approach," in *Medical Imaging 2001: Image Processing*, 2001, vol. 4322: International Society for Optics and Photonics, pp. 128-138.
- [19] D. Kainmüller, T. Lange, and H. Lamecker, "Shape constrained automatic segmentation of the liver based on a heuristic intensity model," in *Proc. MICCAI Workshop 3D Segmentation in the Clinic: A Grand Challenge*, 2007, vol. 109, p. 116.
- [20] M. Sudhamani and G. Raju, "Segmentation and Classification of Tumour in Computed Tomography Liver Images for Detection, Analysis and Preoperative Planning," *International Journal of Advanced Computer Research*, vol. 4, no. 1, p. 166, 2014.
- [21] H. Lamecker, T. Lange, and M. Seebass, "A statistical shape model for the liver," in *International conference on medical image computing and computer-assisted intervention*, 2002: Springer, pp. 421-427.
- [22] T. Gui, L.-L. Huang, and A. Shimizu, "Liver segmentation for CT images using an improved GGVF-snake," in *SICE*

- Annual Conference 2007, 2007: IEEE, pp. 676-681.
- [23] S. Huang, B. Wang, and X. Huang, "Using GVF snake to segment liver from CT images," in *2006 3rd IEEE/EMBS International Summer School on Medical Devices and Biosensors*, 2006: IEEE, pp. 145-148.
- [24] C. Platero *et al.*, "Liver segmentation for hepatic lesions detection and characterisation," in *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2008: IEEE, pp. 13-16.
- [25] L. Ali *et al.*, "Intelligent image processing techniques for cancer progression detection, recognition and prediction in the human liver," in *2014 IEEE Symposium on Computational Intelligence in Healthcare and e-health (CICARE)*, 2014: IEEE, pp. 25-31.
- [26] Y. Masuda, A. H. Foruzan, T. Tateyama, and Y. W. Chen, "Automatic liver tumor detection using EM/MPM algorithm and shape information," in *The 2nd International Conference on Software Engineering and Data Mining*, 2010: IEEE, pp. 692-695.
- [27] P. Rodrigues, J. L. Vilaça, and J. Fonseca, "An image processing application for liver tumour segmentation," in *1st Portuguese Biomedical Engineering Meeting*, 1-4 March 2011 2011, pp. 1-6, doi: 10.1109/ENBENG.2011.6026097.
- [28] P. Anisha, C. K. K. Reddy, and L. N. Prasad, "A pragmatic approach for detecting liver cancer using image processing and data mining techniques," in *2015 International Conference on Signal Processing and Communication Engineering Systems*, 2015: IEEE, pp. 352-357.
- [29] A. Krishan and D. Mittal, "Detection and classification of liver cancer using CT images," *International Journal on Recent Technologies in Mechanical and Electrical Engineering*, vol. 2, no. 5, pp. 93-98, 2015.
- [30] A. Raj and M. Jayasree, "Automated liver tumor detection using markov random field segmentation," *Procedia Technology*, vol. 24, pp. 1305-1310, 2016.
- [31] K. Ahmadi, A. Karimi, and B. Fouladi Nia, "New Technique for Automatic Segmentation of Blood Vessels in CT Scan Images of Liver Based on Optimized Fuzzy C-Means Method," (in eng), *Comput Math Methods Med*, vol. 2016, p. 5237191, 2016, doi: 10.1155/2016/5237191.
- [32] B. F. Nia and A. Branch, "A hybrid automatic segmentation method of blood vessels in CT scan images of liver," *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 4, pp. 799-804, 2017.
- [33] L. Xu, Y. Zhu, Y. Zhang, and H. Yang, "Liver segmentation based on region growing and level set active contour model with new signed pressure force function," *Optik*, vol. 202, p. 163705, 2020.
- [34] B. Sakthisaravanan and R. Meenakshi, "OPBS-SSHC: outline preservation based segmentation and search based hybrid classification techniques for liver tumor detection," *Multimedia Tools and Applications*, vol. 79, pp. 22497-22523, 2020.
- [35] A. L. K. JE and S. V. Jinny, "Automatic Extraction of Lesions and Hepatic Structures in Liver using Segmentation Techniques," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2020: IEEE, pp. 114-117.
- [36] Z. Zhang, C. Duan, T. Lin, S. Zhou, Y. Wang, and X. Gao, "GVFOM: a novel external force for active contour based image segmentation," *Information Sciences*, vol. 506, pp. 1-18, 2020.
- [37] R. Kaur, L. Kaur, and S. Gupta, "Enhanced K-mean clustering algorithm for liver image segmentation to extract cyst region," *IJCA Special Issue on Novel Aspects of Digital Imaging Applications*, vol. 1, pp. 59-66, 2011.
- [38] H. Jiang, R. Zheng, D. Yi, and D. Zhao, "A novel multiinstance learning approach for liver cancer recognition on abdominal CT images based on CPSO-SVM and IO," *Computational and mathematical methods in medicine*, vol. 2013, 2013.