Automated Segmentation and Classification of Hepatocellular Carcinoma Using Fuzzy C-Means and SVM

Mai R. Ibraheem and Mohammed Elmogy

Abstract For successful classification of Hepatocellular Carcinoma (HCC) in ultrasound (US) images, effective preprocessing steps are highly desirable. Most of Computer Aided Diagnostic (CAD) systems miss the most important steps of image preprocessing and image segmentation. In such a framework, only some texture features, which are obtained directly from the images or ROIs, are used as inputs of classifiers. Image preprocessing and segmentation of US images are useful for better judgment of normal and cancerous cases. Although, there are many studies on the classification of medical images, the fully automatic classification is still a difficult task. In this work, we propose an automated classification of US liver tumors using SVM with the aid of Fuzzy c-means (FCM) and level set method. A large number of features were extracted by using statistical, textual, and histogram-based features to discriminate the HCC maximally by developing an SVM classification system. SVMs work on maximizing the margin between the separating hyperplane and the data to minimize upper bound of the generalization error. The proposed Fuzzy C-SVM based system is compared with the K-Nearest Neighbor (KNN) based approach. Experimental results demonstrated that the Fuzzy C-SVM based system greatly outperforms KNN-based approach.

Keywords Hepatocellular carcinoma (HCC) · Computer aided diagnostic (CAD) systems · Ultrasound liver images · Fuzzy c-means (FCM) clustering · Support vector machine (SVM) · K-nearest neighbor (KNN) classification

M.R. Ibraheem (✉)
Faculty of Science, Statistical and Computer Science Department, Mansoura University, Mansoura, Egypt
e-mail: dr_mai_ramadan@yahoo.com

M. Elmogy (✉)
Faculty of Computers and Information, Information Technology Department, Mansoura University, Mansoura, Egypt
e-mail: melmogy@mans.edu.eg; elmogy@gmail.com

1 Introduction

Disease diagnosis based on medical imaging is an invaluable tool for medical experts to plan a patients' rehabilitation process. Imaging modalities, including US imaging, MRI, CT, and digital mammography, help physicians diagnosing the disease accurately in a non-invasive way. However, as the volume of medical images has exponentially grown, manual analysis and interpretation of these images is no more feasible. A CAD system is highly desirable to provide additional support to the radiologists. Particularly, efficient computer algorithms are required to illustrate automatically the structures and region of interest (ROI) [12, 13]. Although, there are many studies on the classification of medical images, the fully automatic classification is still a difficult task. Most of the CAD systems require user intervention [1, 7, 8, 15, 25, 31]. Sometimes intervention from an inexperienced user may lead to false results.

For successful classification of HCC in liver US images, effective preprocessing steps are highly desirable. HCC is the most common liver malignancy [3]. It can be single or multiple, and it has a variable, imprecise delineation. It may have a very pronounced circulatory signal and has a heterogeneous structure. However, US images are contaminated with an inherent noise called 'speckle.' It tends to have a granular effect on the image, thereby degrading its visual quality [30]. The US images should have a minimum amount of noise for simplifying the therapeutic decision-making and diagnosis. It calls for the development of speckle filtering techniques over past decades.

To extract the subtle sonographic features, the contrast of the US image is enhanced by using a bilateral filter. It is a non-linearly filter that considers both similarities of gray level and geometric closeness of the neighboring pixels [16]. In some present work, a large number of features are extracted by using statistical, textual, and histogram-based techniques to discriminate the HCC maximally by developing an SVM classification system [19, 21–24]. SVMs work on maximizing the margin between the separating hyperplane and the data to minimize upper bound of the generalization error. To help radiologists to decide the normal and cancerous cases, a new approach to the two-phase FCM clustering is developed to classify the liver tumor with high performance. The proposed Fuzzy C-SVM based system is compared with the KNN based approach. Experimental results have demonstrated that the Fuzzy C-SVM based system outperforms KNN-based approach.

This paper is organized into six sections. Section 2 reviews the current literature on the interpretation of US images. Section 3 describes the used material and the proposed methodology. Section 4 reviews the proposed diagnostic system. Section 5 refers to the obtained experimental and the classification results. The conclusion and future work are discussed in Sect. 6.

2 Related Work

Many researchers proposed and implemented CAD systems for analyzing liver images. For example, Mittal et al. [18] proposed a CAD system to assist in identifying focal liver lesions in B-mode US images. The proposed system discriminated focal liver diseases, such as Cyst, Hemangioma, HCC, and Metastases, against the normal liver. The images were enhanced using a new methodology that simultaneously operates both an enhancement and speckle reduction process. Subsequently, 208-texture based features are extracted using spectral, statistical, and TEM methods. A two-step neural network classifier is designed for the classification of five liver image categories. Classification using two-step neural network showed correct decisions of 432 out of 500 images with a classification accuracy of 86.4 %.

Sohail et al. [25] presented a combined method of content-based retrieval and classification. The used medical US images include three different types of ovarian cysts: Simple Cyst, Endometrioma, and Teratoma. The features were a combination of histogram moments and Gray Level Co-Occurrence Matrix (GLCM) that based on statistical texture descriptors. The classification performed by using Fuzzy KNN classification technique. Features were extracted from 200 images of the databases. They used to train the classifier by applying "k-Fold Cross-Validation" technique with k = 5. The performance of the proposed method is compared with other popular classification techniques namely, SVM (with RBF, Sigmoid, and Polynomial kernels), ordinary KNN, and ANN.

Ribeiro et al. [24] proposed a semi-automatically liver segmentation system. It helped in identification and diagnosis of diffuse liver diseases. The extracted features from the liver contour were used with clinical and laboratory data. Using the despeckle image, the liver surface contour was obtained using a snake technique. The classification results were compared with the SVM, a Bayesian, and a KNN classifier. Using Leave-one-out cross-validation strategy, the best results are obtained using the KNN classifier, with an accuracy of 80.68 %.

Ribeiro et al. [21] addressed identification and diagnosis of the chronic liver diseases. They used commonest features described in the literature including first order statistics, co-occurrence matrices, wavelet transform, attenuation, and backscattering parameters and coefficients. The classification results of an SVM, a decision tree, and a KNN classifier are compared. The best results were obtained using the SVM with a radial basis kernel, with 73.20 % of overall accuracy.

Ribeiro et al. [23] proposed a semi-automatic procedure for the stage based on US liver images, clinical, and laboratory data. The proposed algorithm was based on a set of laboratory and clinical features from The US. The classification was tested using two classifiers: a KNN and an SVM, with different kernels. The SVM, polynomial kernel, outperformed the others classifiers in every class studied, achieving a sensitivity of 91.67 %. From the attained results, the SVM with polynomial kernel outperformed the KNN and the SVM with the radial basis kernel classifier.

The previously mentioned studies show that the SVM and other methods are used as classifiers for the tissue in US images. However, these systems miss the image preprocessing and image segmentation components. Thus, the features extracted from ROIs directly may not provide accurate performance. The proposed method outperforms these studies and the study of James and Skowron [14] which reported with 77.31 % accuracy, and Garla et al. [10] which reported sensitivity 94.3 %.

3 Materials and Methods

3.1 Data Acquisition

The research in the area of liver diseases using US images had been carried out using some collected individual databases due to non-availability of benchmark image database.

From 150 different liver diseases pictures, we have chosen 94 of the best quality and applicability in pattern recognition area. The age of the patients in the dataset ranges from 38 to 78 years. Further, patient's privacy has been protected by labeling the data with numeric dummy values and keeping patients' credential undisclosed. We have obtained data from the Egyptian Liver Research Institute and Hospital in Sherbin, Dakahlia Governorate, Egypt. Figure 1 shows some pictures for the chosen diseases with the different appearance of hepatocellular carcinoma.

It is critical to processing the liver tumor US images by image preprocessing and image segmentation stages for better judgment of normal and cancerous cases. These two stages will facilitate and increase the performance of the classification stage in liver tumor applications.
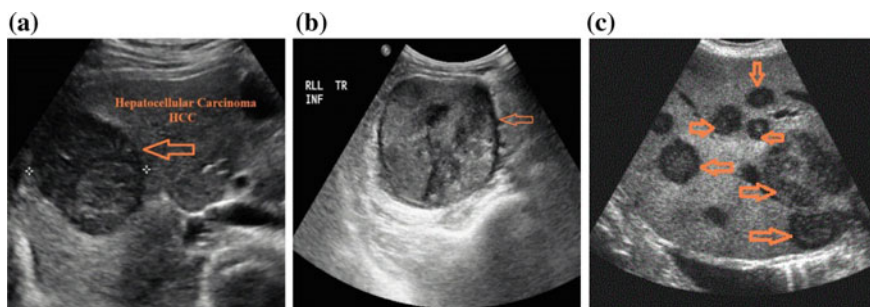

Fig. 1 Ultrasound liver images with different appearance of Hepatocellular carcinoma, **a** small HCC, **b** large HCC, and **c** multiple masses

3.2 Pre-processing Stage

3.2.1 Image Enhancement and Noise Removal

To improve the visualization of US images, the process of image enhancement is done in such a way that the visual details of the texture should remain preserved. Since image noise and artifacts often impair the performance of the classification, it would be attractive to incorporate spatial information into the classifier. Spatial filters have been widely used for noise elimination in the presence of additive noise. The Gaussian filter is a local and linear filter. It smoothes the whole image irrespective of its details or edges. Whereas the bilateral filter is also local but non-linearly that considers the gray level similarity and geometric closeness without smoothing edges [16].

Bilateral filtering is a non-linear filtering technique introduced by Tomasi et al. [29]. It extends the technique of Gaussian smoothing by weighting the coefficients of the filter with their corresponding relative pixel intensities. Bilateral Filter considers intensity and spatial information between each point and its neighboring points. It is unlike the conventional linear filtering where only spatial information is considered. It preserves the sharp boundaries well and averages the noise out as it average pixels belonging to the same region as the reference pixel. Mathematically, the bilateral filter output at a pixel location p is calculated as follows [16]:

$$I_F(p) = \frac{1}{W} \sum_{q \epsilon S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I(p) - I(q)|) I(q) \tag{1}$$

$$G_{\sigma_s}(\|p - q\|) = e^{-\frac{p-q^2}{2\sigma_s^2}} \tag{2}$$

$$G_{\sigma_r}(|I(p) - I(q)|) = e^{-\frac{|I(p)-I(q)|^2}{2\sigma_r^2}} \tag{3}$$

$$W = \sum_{q \epsilon S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I(p) - I(q)|) \tag{4}$$

Equation (2) represents a geometric closeness function, whereas Eq. (3) is a gray level similarity function. Equation (4) is a normalization constant, whereas $\|p - q\|$ is the Euclidean distance between p and q. The two parameters σ_s and σ_r control the behavior of the bilateral filter. Also the optimal σ_s value is relatively insensitive to noise variance compared to the optimal σ_r value and is chosen based on the desired amount of low-pass filtering. A large σ_s blurs more, i.e., it combines values from more distant image locations [28].

3.2.2 Segmentation of Region-of-Interest

FCM is an unsupervised fuzzy segmentation technique. The clusters are obtained iteratively by minimizing a cost function that depends on the distance of pixels to the cluster centers. Each data point may belong to more than one cluster with certain degrees of membership [9]. Therefore, it is especially useful for medical image segmentation where objects in the images do not have well-separated boundaries [9, 32]. FCM assigns pixels to clusters based on their fuzzy membership values. It strives to minimize the following cost function [9]:

$$J = \sum_{j=1}^{N} \sum_{i=1}^{c} u_{ij}^m \left\| x_j - c_i \right\|^2, \ 1 \le m \le \infty \tag{5}$$

where u_{ij} shows the membership of pixel x_j to ith cluster $\forall \ x_j \in \Omega$, where Ω represents the set of points that an image is composed. C and N represent a total number of clusters and data points in Ω, and vi is the centroid of the ith cluster. The constant m is also known as the degree of fuzziness and is usually set to 2 for most applications. The following mathematical expressions [4] are used to update the fuzzy membership functions and cluster centers, respectively:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\left\| x_j - c_i \right\|}{\left\| x_j - c_k \right\|} \right)^{\frac{2}{m-1}}} \tag{6}$$

$$c_i = \frac{\sum_{j=1}^{N} u_{ij} x_j}{\sum_{j=1}^{N} u_{ij}^m} \tag{7}$$

In this work, FCM image segmentation was employed to extract the contours of liver tumors automatically from US images. It integrates FCM with 'level set' technique to extract contours of liver tumors from US images with high reliability.

3.3 Feature Extraction and Selection

Feature extraction and selection are the most critical steps in CAD systems [2, 7, 17, 23–25, 31]. For the liver, the most commonly used features are textural measures by constructing spatial gray level dependence matrices, also termed as co-occurrence matrices that were introduced by Haralick et al. [11]. These features are normalized [0, 1] and then used as input to the SVM classifier.

In the proposed CAD system, five kinds of features (statistical, textural, run length, difference method, and histogram based features) were analyzed and extracted from the suspicious areas and ROIs. Usually, many features are extracted,

and we need to select the significant ones. In this paper, we apply the rough set theory [27] to select an optimal subset of features. These five sets of texture features were calculated for each ROI and combined into one set of features for image characterization. The five sets of texture features are listed as follows.

(1) **First order statistics (FOS)**: First order texture measures are statistically calculated from the original image values, such as variance. They do not consider pixel neighborhood relationships. Based on the image histogram, six features are used [26, 31]. Average gray level, standard deviation, entropy, the coefficient of variance, skewness, and kurtosis are obtained from each ROI [19].

(2) **Textural**: It based on co-occurrence matrices of the texture information also it is called the spatial gray-level dependence (SGLD) matrices. The gray level co-occurrence matrix (Second-order statistical model) gives relevant information about the inter-pixel relationship, periodicity, and spatial gray level dependencies. The analysis consists of the construction of sixteen different GLCM considering angles between pixels of 0°, 45°, 90° and 135°, for an inter-pixel distance equal to 1, 2, 3, and 4. Twenty-two descriptors have been extracted from each ROI, producing 325 features (22 GLCM features × 16 different GLCM). According to [24], it is using only the "near" and "far" displacements, which are enough to capture the spatial properties of the texture for the liver. It produces 88 features (22 GLCM features × 2 different GLCM).

(3) **Gray-Level Run Length Matrix Features**: Another measure of texture is based on run length. The image is scanned line by line and the length in pixels of each line is noted. Then, the relationship between each run length is identified. This relationship of all these statistical parameters run length makes a pattern. This pattern is a measure of the image texture. The average of all the line lengths (in pixels) in a region gives a measure of coarseness of the texture. GLRLM expresses the number of the consecutive image elements that have the same gray level (gray level run) [2]. Seven features are computed from GLRLM, as long run emphasis, short run emphasis, low gray level run emphasis, high gray level run emphasis, run length non-uniformity, gray level non-uniformity, and run percentage [7]. Feature values are averaged over four basic angular directions, 0°, 45°, 90°, 135°.

(4) **Gray-Level Difference Method (GLDM)**: The GLDM is based on the probability of occurrence that two pixels separated by a particular displacement vector have a given difference [26]. Four kinds of displacement vectors are considered, such as (0, d), (−d, d), (d, 0), (−d, −d), where d is the inter-sampling spacing distance. The five textural features used in the experiments are contrasted, angular second moment, entropy, mean, and inverse difference moment. In this study, the probability density functions are computed according to four kinds of vector displacements and textural features for each probability density function.

(5) **Histogram-Based Features**: It provides many clues to describe the charac-
teristics of the image. Six statistical features are extracted from the histogram.
They are mean, variance, skewness, kurtosis, energy, and entropy. The mean
is the average intensity level, whereas the variance implies the variation of
intensities around the mean. The skewness shows whether the histogram is
symmetric about the mean. The kurtosis is a measure of whether the data are
peaked or flat about a normal distribution. Entropy is a measure of the system
disorder.

Many features are extracted with a strong correlation with each other. To select
the significant ones from them. Feature selection is a very effective preprocessing
step to the data mining in reducing dimensionality, increasing learning accuracy,
and improving comprehensibility. To find the optimal subsets of the feature, the
Rough set theory provides a mathematical approach to finding optimal feature
subset. The Rough Set Theory (RST) was applied to reduce these features suc-
cessfully and decide the most efficient ones.

Most classification problems involve a large set of potential features that must
identify feature selection. Principal component analysis (PCA) is the most widely
adopted traditional statistical method. However, the features selected using PCA are
variable-independent but may not be the most benefit for a particular problem [32].
RST deals with the approximation of sets assembled from empirical data. RST is
helpful in discovering the decision rules and minimizing the conditional attributes
[27].

After extracting and selecting of the histogram features, FOS, GLDM, GLRM,
and GLCM based texture features are calculated for the processed image. They are
organized into a single feature vector. Each feature vector x_k consists of 112 fea-
tures. Then, it is normalized and used as an input to the SVM classifier.

3.4 SVM Classification

SVMs are a new type of pattern classifier based on a novel statistical learning
technique that recently proposed by Asa and Jason [2] and Ng [20]. Unlike (e.g.
Neural Networks), which works on minimizing the empirical training error [20].
The criterion used by SVMs is based on maximizing the margin between the
separating hyperplane and the data. The maximum margin classifier is the dis-
criminant function. It maximizes the geometric margin $1/\|w\|$, which is equivalent
to minimizing $\|w\|^2$. It leads to the following constrained optimization problem:

$$\text{Minimize} \quad \frac{1}{2}\|w\|^2$$
$$w; b \tag{8}$$
$$subject\ to:\ y_i(w^T xi + b) \geq 1\ i = 1, \ldots, n$$

Maximizing the margin, means searching for the classification function that can most safely separate the two data classes. The threshold separating two data classes is the line in the middle between the two margin boundaries, which are represented as $x^T w + b = 1$ and $x^T w + b = -1$. Thus, margin is $2/||w||$, where $||w||$ is the norm of the vector w. To determine the separating hyperplane, the margin between positive class and a negative class has to be maximized to produce good generalization ability [5]. In general, the RBF kernel maps samples into a higher dimensional space nonlinearly. Unlike the linear kernel, which handles the nonlinear case between class labels and attributes [6].

$$K(x_i, x_j) = \exp(-\gamma ||x_i - x_j||^2), \ \gamma > 0 \tag{9}$$

Gauss radial basis function SVM classification system was developed to discriminate the HCC. Five objective measurements (accuracy, sensitivity, specificity, positive predictive value, and negative predictive value) are used to evaluate the classification results. The higher the five measurements are, the more reliable and valid the CAD system is.

4　The Proposed Diagnostic System

The developed CAD system architecture is composed of four modules: preprocessing steps, feature extraction, feature selection and Multi-SVM classifier (Fig. 2).

The performance of the proposed CAD system was evaluated using the overall accuracy that expresses the correct percentage of classifier predictions. We have used the k-fold method to perform the cross-validation testing. The 10-fold cross-validation method randomly divides the dataset into ten groups. Nine groups of them are used for training and rest group for classifiers testing. This procedure is repeated until all groups have been used in the testing. The final result corresponds to the average accSX1uracy estimated for each iteration [17].

The classification performance is in terms of the four objective measurements Classification accuracy (ACC), sensitivity, specificity, and Negative Predictive value (NPV) were used to test the performance of classifiers models. ACC, sensitivity, specificity, and NPV are defined as follows according to the confusion matrix that is shown in Table 1 as follows [32]:

$$\text{Accuracy} = \frac{TP + TN}{(TP + FP + TN + FN)} \times 100 \tag{10}$$

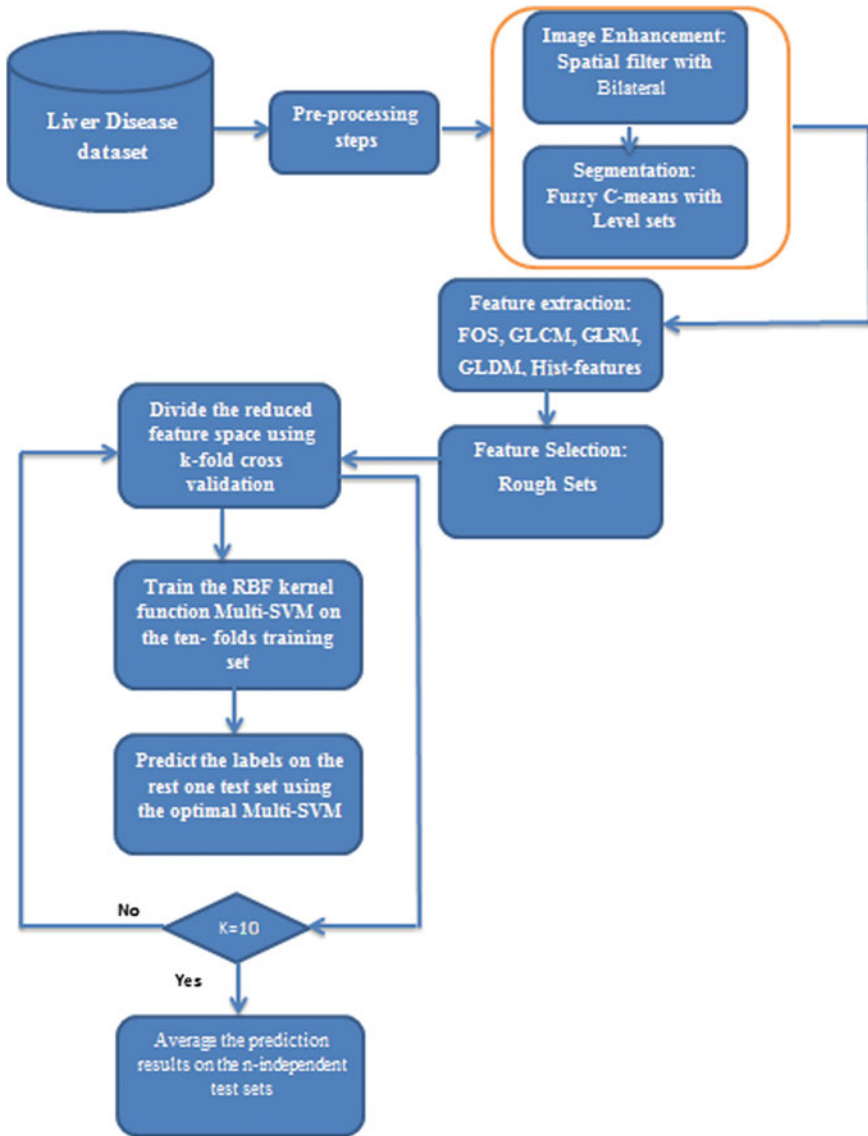**Fig. 2** Overall procedure of the proposed HCC multi-SVM-based diagnostic system

$$Sensitivity = \frac{TP}{TP + TN} \qquad (11)$$

$$Specificity = \frac{TN}{TN + FP} \qquad (12)$$

Table 1 Confusion matrix

		Predicted label	
		Positive	Negative
Known label	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

$$NPV = \frac{TN}{(TN + FN)} \times 100 \tag{13}$$

In the confusion matrix, TP is the number of true positives, which means that some cases with 'HCC' class are correctly classified as HCC. FN is the number of false negatives, which means that some cases with the 'HCC' class are classified as healthy persons. TN is the number of true negatives, which means that some cases with the 'Healthy' class are correctly classified as healthy persons. FP is the number of false positives, which means that some cases with the 'Healthy' class are classified as HCC [32].

5 Results and Discussion

The liver tumor US images are segmented and classified by our proposed technique. Figure 3a shows one of the original HCC ultrasound images. To reduce the speckle noise and improve the visualization of US images, bilateral filter, and non-linear noise reduction is applied to the original image, as shown in Fig. 3b.

As previously described, contours for tumor initialization has been determined automatically using the Level set, as shown in Fig. 3c. These contours are used for initializing FCM to obtain an accurate segmentation of HCC US images, as shown in Fig. 3d. The proposed approach of classifying the liver tumor US images is automatic, and hence no user intervention is required.

Higher classification accuracy has been obtained corresponding to images segmented by the mentioned technique that verify the positive influence of the proposed FCM-SVM algorithm for the classification process. For the liver case, the most commonly used features are related to textural measures by constructing spatial gray level dependence matrices also termed as co-occurrence matrices that were introduced by Haralick. These features are normalized in the range of [0, 1] and then used as input to the SVM classifier.

In this work, a total of 112 features has been extracted from each ROI of liver ultrasound images, namely six histogram features, 6 FOS, 5 GLDM, 7 GLRM and
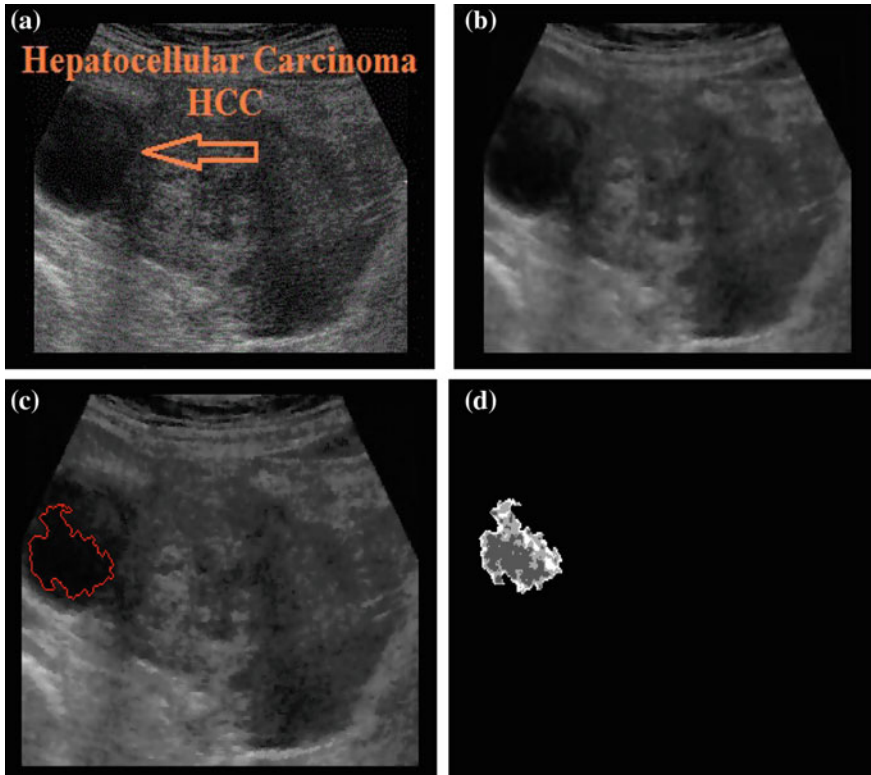
Fig. 3 The results using our proposed approach to automatic segmentation and classification system. **a** The original HCC image, **b** image enhanced by the bilateral filter, **c** initial contour of the level set, **d** segmented image by FCM

88 GLCM based texture features. These features have been measured from ROI for every normal and abnormal image. After dimensionality reduction using rough sets, 38 features were obtained. These features are used to train and test SVM classifier using k-fold cross validation (Table 2).

Table 3 shows the classification results based on 38 features obtained. It can be observed from Table 3 that classifying HCC using these features shows significant performance in terms of the four objective measurements accuracy, sensitivity, specificity, and Negative Predictive value (NPV). Furthermore, to check the performance of our proposed approach, KNN have been applied on the given dataset.

Table 3 shows a comparison between the proposed approach and KNN at different quality measures. From Table 2, it is clear that SVM gives superior performance compared to other techniques. Figure 4 shows the graphical performance comparison of KNN and Fuzzy C-SVM at various validity measures. Classification

Table 2 A comparison with some approaches to clustering and classification of Liver tissue

Authors	Liver image classes	Features	Pre-processing steps		Classifier	Dataset description	Evaluation method	Accuracy (%)
			Enhancement	Segmentation				
This study	Ultrasound focal liver lesions: Cyst, HEM, and HCC	FOS, GLCM, GLDM, GLRLM, histogram features	Spatial bilateral non-linear filter	FCM and level set	Multi-SVMRBF	Individual collected database, included 24Cyst, 33 HEM, 28 HCC and 11 normal	10-fold cross-validation	96.11
Virmani et al. [31]	Cyst, HEM, HCC, and MET	FOS, SGLDM, GLRLM, and Gabor wavelet methods	No pre-		PCA-SVM	Individual collected database, included 21 NOR images, 12 Cyst, 15 HEM, 28 HCC and 32 MET images with 35 MET	10-fold cross-validation	87.2
Andreia et al. [1]	Ultrasound steatosis liver	GLCM features × 9 different displacements 0°, 45° and 90°, for an inter-pixel distance equal to 1, 2 and 3	No pre-		SVM	Individual collected database, included 36 patients as normal and steatosis liver	10-fold cross-validation	76.92
Kumar and Shreyamsha [16]	Ultrasound focal liver lesions: Cyst, HCC, HEM, and MET	FOS, SGLDM, GLRLM, texture energy measures and Gabor wavelet methods	Using diamond shape template	Marked by an expert radiologist	A two-step neural network classifier	Individual collected database, included 16 normal, 17 Cyst, 15 HCC, 18 HEM and 45 MET	Manual	86.4

(continued)

Table 2 (continued)

Authors	Liver image classes	Features	Pre-processing steps		Classifier	Dataset description	Evaluation method	Accuracy (%)
			Enhancement	Segmentation				
Ribeiro et al. [19]	Chronic Hepatitis without cirrhosis, Cirrhosis, and HCC	Liver contour features with laboratory and clinical parameters	De-speckled	A snake technique	k-nearest neighbor	Individual collected database, included 88 image	Leave-one-out cross-validation	80.68
Ribeiro et al. [20]	Chronic Hepatitis without cirrhosis, Cirrhosis, and HCC	First order statistics, co-occurrence matrices, wavelet transform, with clinical and laboratory data	De-speckled	A snake technique	Kernel SVM	Individual collected database, included 97 patients	Leave-one-out cross-validation	73.20
Ribeiro et al. [21]	Chronic Hepatitis without cirrhosis, Cirrhosis, and HCC	First order statistics, co-occurrence matrices, wavelet transform, with clinical and laboratory data	De-speckled	Snake technique	Kernel SVM	Individual collected database, included 115 images	Leave-one-out cross-validation	81.25

Table 3 The performance of the proposed approach compared with KNN approach

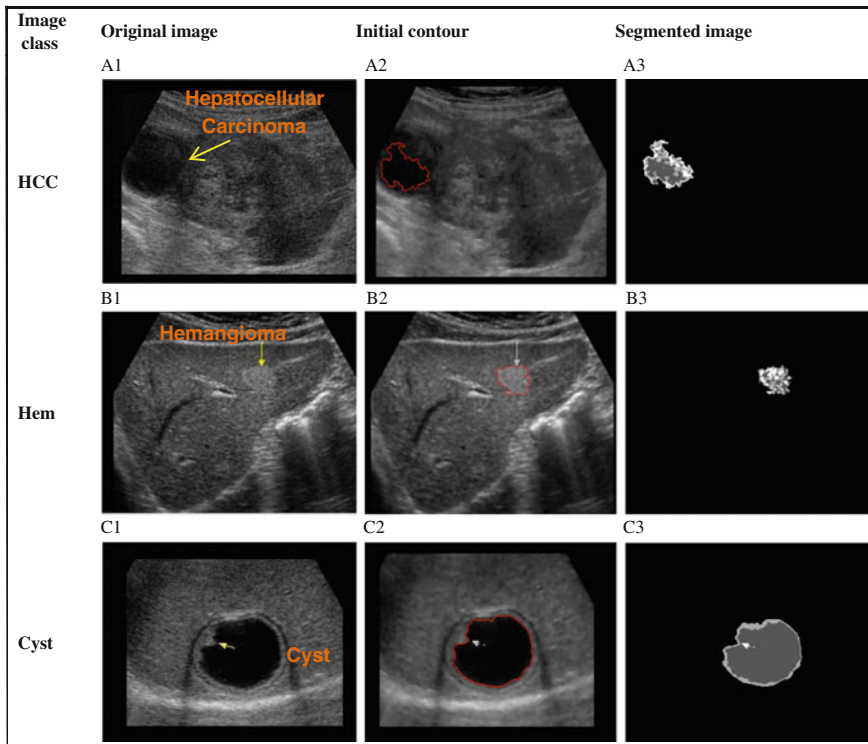|		Accuracy (%)	Sensitivity (%)	1-Specificity (%)	NPV (%)
Classifier	Fuzzy C-SVM	84.44	97.3	75.00	66.66
	K-nearest neighbor	71.11	75.68	50.00	30.76

Fig. 4 Original, initialized, and segmented ultrasound images of three liver image categories

using SVM outperforms KNN technique at all validity measures. Image classification is one of the most important steps to know about the presence of HCC in liver images. In the proposed approach, SVM has been used for the classification of normal and abnormal images (Fig. 5).

Fig. 5 Performance
comparison of SVM and
KNN classification techniques
at different classification
validity measures
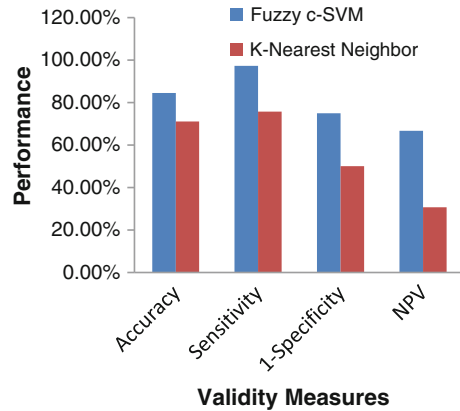
6 Conclusion

This paper proposes using the image processing and image segmentation compo-
nents prior to classification to improve decision accuracy. Another contribution of
this work is developing an automatic classification system for HCC ultrasound
images. Therefore, it is useful to discriminate normal and cancerous cases. The
proposed approach of automatic contour initialization by level set shows the
effectiveness of the method. The estimated features extracted from statistical, his-
togram, difference method, run length and textural approaches have led to
encouraging results as inputs of the used classifiers. In our proposed approach, we
have used SVM for the classification of normal and abnormal images. Different
types of validity measures are calculated to show the effectiveness of the proposed
approach. Using k-fold cross-validation to train and test Fuzzy C-SVM classifier,
we have obtained 84.44 % classification accuracy and sensitivity 97.3 % using a
10-fold cross-validation method. A numerical comparison is made with K-nearest
neighbor at different validity measures. In the future, we intend to extend the
automatic classification system for various types of medical images to show the
seriousness of other diseases. Further, exploring various types of features, which
may be used to improve the accuracy of the classifier.

References

1. Andreia, A., Jose, S., Santos, J., et al.: Classifier approaches for liver steatosis using ultrasound
   images. Proc. Technol. **5**, 763–770 (2012)
2. Asa, B., Jason, W.A.: User's guide to support vector machines. Springer, data mining techn
   for the life sciences. Methods Mol. Biol. **609**, 223–239 (2010)
3. Badea, R., Ioanitescu, S.: Ultrasound imaging of liver tumors - current clinic apps. Liver
   Tumors (2012)

4. Balafar, M., Ramli, A., Saripan, M., Mahmud, R., et al.: Medical image segmentation using fuzzy c-mean (FCM). In: Learning Vector Quantization (LVQ) and User Interaction, vol. 15, pp. 177–184. Springer, Heidelberg (2008)
5. Bhat, H., Wani, M.: Modified one-against-all algorithm based on support vector machine. Int. J. Adv. Res. Comput. Sci. Softw. Eng. 3(12) (2013)
6. Chih, W., Chih, C., Chih, J.: A Practical Guide to Support Vector Classification. Citeseerx (2010)
7. Dimitris, G., Ioannis, K., Theocharakis, P., et al.: A multi-classifier system for the characterization of normal, infectious, and cancerous prostate tissues employing transrectal ultrasound images. Comput. Methods Programs Biomed. 97, 53–56 (2010)
8. Dimitris, K., Eystratios, G., Dimitris, M.: Fusion of fuzzy statistical distributions for classification of thyroid. Artif. Intell. Med. 50, 33–41 (2010)
9. Dzung, L., Chenyang, X., Jerry, L.: A survey of current methods in medical image segmentation. Ann. Rev. Biomed. Eng. 2, 315–338 (2000)
10. Garla, V., Taylor, C., Brandit, C.: Semi-supervised clinical text classification with Laplacian SVMs: an application to cancer case management. J. Biomed. Inform. 46(5), 869–875 (2013)
11. Haralick, R., Shanmugam, K., Dinstein, H.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. SMC8211 3(6), 610–621 (1973)
12. Hassanien, A.E.: Classification and feature selection of breast cancer data based on decision tree algorithm. Int. J. Stud. Inform. Control J. SI 12(1), 33–39 (2003)
13. Hassanien, A.E., Moftah, H.M., Azar, A.T., Shoman, Mahmoud: MRI breast cancer diagnosis hybrid approach using adaptive ant-based segmentation and multilayer perceptron neural networks classifier. Appl. Soft Comput. 14, 62–71 (2014)
14. James, P., Skowron, A.: Transactions on Rough Sets. Springer, Heidelberg (2008)
15. Jeon, J., Choi, J., Sihyoung, L., et al.: Multiple ROI selection based focal liver lesion classification in ultrasound images. Expert Syst. Appl. 40, 450–457 (2013)
16. Kumar, B., Shreyamsha, K.: Image denoising based on gaussian/bilateral filter and its method noise thresholding. SIViP 7, 1159–1172 (2013). Springer, London
17. Liu, B., Cheng, H., Jianhua, H.: Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images. Pattern Recogn. 43, 280–298 (2010)
18. Mittal, D., Kumar, V., Suresh, C., et al.: Neural network based focal liver lesion diagnosis using ultrasound images. Comput. Med. Imaging Graph. 35, 315–323 (2011)
19. Namita, A., Agrawal, R.: First and second order statistics features for classification of magnetic resonance brain images. J. Signal Inf. Process. 3, 146–153 (2012)
20. Ng, A.: Support Vector Machines, Lecture notes. CS 229 Machine Learning, Course Materials, CS229 Lecture Notes (2008)
21. Ribeiro, R., Marinho, R., Velosa, J., et al.: The usefulness of ultrasound in the classification of chronic liver disease. In: IEEE Annual International Conference, Engineering in Medicine and Biology Society, EMBC, pp. 5132–5135 (2011)
22. Ribeiro, R., Marinho, R., Velosa, J., et al.: Diffuse Liver Disease Classification from Ultrasound Surface Characterization, Clinical, and Laboratorial Data, pp. 167–175. Springer, Heidelberg (2011)
23. Ribeiro, R., Rui, M., Velosa, J., Ramalho, F., et al.: Chronic liver disease staging classification based on ultrasound clinical and laboratorial data. In: IEEE International Symposium, Biomedical Imaging: From Nano to Macro, pp. 707–710 (2011)
24. Ruchaneewan, S., Daniela, R., Jacob, F., et al.: A co-occurrence texture semi-invariance to direction, distance, and patient size. In: Proceedings of SPIE, Medical Imaging: Image Proceedings, vol. 6914, p. 69141Y (2008)
25. Sohail, A., Bhattacharya, P., Sudhir, P., et al.: Content-based retrieval and classification of ultrasound medical images of ovarian cysts. Artif. Neural Netw. Pattern Recogn. 5998, 173–184. Springer, Heidelberg
26. Srinivasan, G.: Statistical texture analysis proceedings of world academy of science engineering, and technology. 36, 1264–1269 (2010)

27. Stoean, R., Stoean, C., Lupsorb, M., et al.: Evolutionary-driven support vector machines for determining the degree of liver fibrosis in chronic hepatitis c. Artif. Intell. Med. **51**, 53–65 (2011)
28. Sylvain, P., Pierre, K., Jack, T., Fredo, D.: Gentle introduction to bilateral filtering and its applications. In: Proceeding ACM SIGGRAPH '8, Article No. 1, New York, USA (2008)
29. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 839–46, New Delhi, India (1998)
30. Vikrant, B., Gopal, S., Atul, S.A.: Novel weighted diffusion filtering approach for speckle suppression in ultrasound images. In: Proceedings of (Springer) International Conference on Frontiers in Intelligent Computing Theory and Applications (FICTA 2013), AISC, vol. 247, pp. 459–466 (2013)
31. Virmani, J., Kumar, V., Kalra, N., Khandelwal, N.: PCA-SVM based CAD System for Focal Liver Lesions using B-Mode Ultrasound Images. Defence Sci. J. **63**(5), 478–486 (2013)
32. Wen, J., Shih, W., Lin, W., et al.: Combining support vector machine with genetic algorithm to classify ultrasound breast tumor images. Comput. Med. Imaging Graph. **36**, 627–633 (2012)