

# Automated Focal Liver Lesion Staging Classification based on Haralick Texture Features and Multi-SVM

Ali.A.Sakr  
Information Systems dept.  
Faculty of Engineering  
KFS University

Magdi Elias Fares  
Math dept.  
Faculty of Science  
Mansoura University

Mai Ramadan  
Statistics and Computer  
Science dept. Faculty of  
Science, Mansoura University

## ABSTRACT

This paper proposes automated identification and classification of various stages of focal liver lesions based on the Multi-Support Vector Machine (Multi-SVM). The proposed system can be used to discriminate focal liver diseases such as Cyst, Hemangioma, and Hepatocellular carcinoma along with normal liver. The multi-class scenario is a composition of a series of two-class problems, using one-against-all which is the earliest and one of the most widely used implementations. We formulate the discrimination between cysts, cavernous hemangioma, hepatocellular carcinoma, and normal tissue as a supervised learning problem, and apply Multi-SVM to classify the diseases using Haralick local texture descriptors and histogram based features calculated from Regions Of Interest (ROIs), as input. Selection of ROI significantly impact the classification performances, thus we propose an automatic ROI selection using Fuzzy c-means initialized by level set technique. For multi-class classification, we adopt the One-Against-All (OAA) method. The proposed Multi-SVM based CAD system using 10-fold cross validation yielded classification accuracy of 96.11% with the individual class accuracy of 97.78%, 95.56%, 93.33% and 97.78% for NOR, Cyst, HEM and HCC cases respectively. The proposed Multi-SVM based system is compared with the K-Nearest Neighbor (KNN) based approaches. Experimental results have demonstrated that the Multi-SVM based system greatly outperforms KNN-based approaches and other methods in the literature. The good performance of the proposed method shows a reliable indicator that can improve the information in the staging of focal liver lesion diseases.

## General Terms

Medical Image Processing, Image Features and Analysis, and Liver diagnosis.

## Keywords

Support Vector Machine, Haralick, Classification, Diagnosis, Focal Liver Lesions.

## 1. INTRODUCTION

Follow-up stages of liver diseases are essential because they are rapidly increases and sometimes up to late stages such as the hepatocellular carcinoma. Different to the diffused liver diseases i.e. fatty and cirrhotic livers, the focal lesion diseases such as cyst, Hemangioma (HEM) and Hepatocellular Carcinoma (HCC) lesions are concentrated in a very small area of the tissue which is difficult to identify the affected part of the liver from the image alone; hence a biopsy test is needed to confirm the disease. These focal lesions are (i) Cyst (liquid nodule), (ii) HEM (non-cancerous solid nodule) and (iii) HCC (primary cancerous solid nodule). The hem-appearances/features in a focal liver lesion may lead to

misdiagnosis and confusion with other lesions even for the expert radiologists. In most cases, the sonographic appearance of Cysts, HEM and HCC overlaps sufficiently to make a distinction difficult or impossible. This research proposes automated identification and classification of various stages of focal liver lesions based on the Multi-Support Vector Machine (Multi-SVM).

The crucial step is in extracting appropriate features enables accurate distinction. Feature extraction techniques include local texture descriptors and histogram based features. Further, to extract the sonographic information, the contrast of the US image is enhanced by using a new methodology which is a spatial filter but nonlinear one called bilateral filter. The bilateral filter is a local filter, but non-linearly considers both gray level similarities and geometric closeness of the neighboring pixels without smoothing edges [14]. Selection of ROI significantly impact the classification performances, thus ROI selection was done in an automatic way with the aid of Fuzzy c-means and level set. The combination of GLCM features by haralick and multi-SVM classifier is used to characterize focal liver lesion diseases. Selection of ROI significantly impact the classification performances, thus we propose an automatic ROI selection using Fuzzy c-means with the aid of level set. For the liver case, the most commonly used features are related to textural measures by constructing spatial gray-level dependence matrices [1-2, 7, 11-12], also termed as co-occurrence matrices that were introduced by haralick. In the present work, large numbers of features are extracted by using statistical, textural and histogram-based features to maximally discriminate between focal lesions mentioned by developing a Multi-SVM classification system. These features are used in training and testing Multi-SVM using k-fold cross validation.

SVMs aim to minimize the upper bound of generalization error through maximizing the margin between the separating hyperplane and the data. Unlike traditional methods (e.g. Neural Networks), which minimize the empirical training error. SVM has achieved superior performance in a wide range of applications [4, 11, 21, 23, 27]. SVM is designed for binary classification, but in our study, the classification problem is multiple. In order to effectively extend it for multiclass classification, we adopt the OAA method. The proposed Multi-SVM based system is compared with the K-Nearest Neighbor (KNN) based approaches. Experimental results have demonstrated that the Multi-SVM based system greatly outperforms KNN-based approaches and other methods in the literature. The rest of this paper comprises four sections. Section 2 review today's literature in interpretation of ultrasound images. Section 3 describes the material used in this study and the proposed methodology. Section 4, explores the experimental evaluation and the classification results

obtained. A discussion among, with a summary of conclusions is provided in the last section.

## **2. RELATED WORK**

Several methods have been proposed for diagnosis and classification of different soft tissues. Each method has its own benefits and limitations. Chen, et. al. [10] presented an effective and efficient diagnosis system using fuzzy KNN (FKNN) for Parkinson's disease. It uses concepts from fuzzy logic to assign degree of membership to different classes while considering the distance of its KNN. They reported (96.07% for mean accuracy) by 10-fold CV method.

Wen-Jie, et al. [26] proposed an approach which simultaneously combines feature selection and parameter setting, to promote the classification accuracy and decrease the time of extracting features and finding (near) optimal classification model of an ultrasound breast tumor image computer-aided diagnosis system. They reported 95.24% accuracy. However, the proposed GA-based approach may be extended to simultaneously determine the SVM-type, kernel function, the parameter values, and feature selection.

Dimitris et al. [8] proposed a noise-resistant coding of both texture and echogenicity, based on combination of a fuzzy distribution of local binary patterns, and ultrasound echogenicity represented by the fuzzy grey-level histogram (FGLH), in the representation of thyroid ultrasound patterns. They achieved 97.5% accuracy. However the investigation of this issue requires extensive experimentation with images degraded with various levels of artificial speckle noise.

Liu et al. [5] proposed an automatic classification method for BUS images, which was divided into two steps: ROI generation step and ROI classification step. A well-trained Kernel Support Vector Machine (KSVM) is employed for classifying normal tissue and tumor. They reported 93.75% accuracy.

Sohail et al. [1] presented a combined method of content-based retrieval and classification of ultrasound medical images representing three types of ovarian cysts: Simple Cyst, Endometrioma, and Teratoma. Features used were a combination of: Histogram moments and Gray Level Co-Occurrence Matrix (GLCM) based statistical texture descriptors. Classification has been performed applying FKNN classification technique. They reported average of classification accuracy as 88.12%. However, this approach needs minimum interaction from a user for the segmentation step.

Chaudhry et al. [4] developed an automatic technique for carotid artery ultrasound image segmentation, based on active contour. They recorded 98.8 % classification accuracy using jackknife validation technique. However, the proposed approach needs minimum interaction from a user for an early detection of plaque in carotid artery.

For liver diseases, this section presents previous classification and clustering techniques for liver grade prediction.

Mittal et al.[7], proposed a computer-aided diagnostic system to assist radiologists in identifying focal liver lesions in B-mode ultrasound images. The images are first enhanced by using a new methodology which handles simultaneously an enhancement and speckle-reduction process; subsequently 208-texture based features are extracted using statistical, spectral and TEM methods to feed to a two-step neural network classifier. They recorded classification accuracy of 86.4%. They employed traditional methods (e.g. ANNs),

which aim to minimize the empirical training errors. Unlike SVMs which have achieved superior performance in a wide range of applications by minimizing the upper bound of the generalization error through maximizing the margin between the separating hyper-plane and the data.

Garla et al. [25] employed semi-supervised learning algorithms using both labeled and unlabeled data to train classifiers and compared linear and Laplacian SVMs on abdominal CT, MRI, and ultrasound reports; to evaluate the effect of unlabeled training data on Laplacian SVM performance. They reported Sensitivity 94.3% vs. 91.1%, Positive Predictive value 87.7% vs. 88.3%. However, they employed a semi-supervised method in classification.

Ribeiro et al. [22] segmented and quantified liver contour in semi-automatic way in order to help the identification and diagnosis of diffuse liver disease. The features extracted from the liver contour are jointly used with clinical and laboratorial data in the staging process. Features extracted from the liver contour detected in the de-speckled field, are used to train supervised classifiers to detect the disease. Using the k-nearest neighbor classifier, they achieved overall accuracy of 80.68% using Leave-one-out cross-validation strategy. However, the method needs a user interaction to select four points of the liver surface; to initialize the snake.

Virmani et al. [11] proposed a Computer-Aided Diagnostic (CAD) system for FLLs based the texture inside and outside the region in classification of FLLs. Texture features were computed from inside and outside of the regions of lesions, two types of features were considered for analysis, i.e. texture features and texture ratio features. Feature set consisted of 208 texture features. Principal Component Analysis (PCA) was used for finding the optimal number of principal components to train a SVM classifier for the classification task. They obtained classification accuracy of 87.2% with the individual class accuracy of 85%, 96%, 90%, 87.5% and 82.2% for NOR, Cyst, HEM, HCC and MET cases respectively using 10 fold cross validation. However, this system does not have image preprocessing or image segmentation components. In such systems, only some texture features obtained directly from the images or ROIs are used as inputs of the classifiers; that the features extracted directly from ROIs may not provide robust and accurate performance.

Andreia et al. [2] presented a semi-automatic classification approach to evaluate steatotic liver tissues using B-scan ultrasound images. Several features have been extracted and used in three different classifiers, such as ANNs, SVM and KNN. Twenty-two descriptors have been extracted from each ROI, producing 198 features (22 GLCM features  $\times$  9 different GLCM). Using the 10-cross validation they suggested that SVM generally performs KNN and ANN better due this good accuracy 76.92%. They used all the displacements and all orientation which may lead to redundant features. According to [12], only the "near" and "far" displacements are enough to capture the spatial properties of the texture for the liver.

Ribeiro et al. [21] addressed identification and diagnosis of various stages of chronic liver disease. They employed ultrasound image intensity and texture features with clinical and laboratorial data in the staging process. They used a set of features from speckle and despeckled image fields. Using the support vector machine with a radial-basis kernel, they obtained 73.20% of overall accuracy using leave-one-out, cross-validation strategy. However, a multi classifier can be used to obtain the classification at various stages.

Ribeiro et al. [23] proposed a semi-automatic procedure to stage Chronic Liver disease based on ultrasound liver images, clinical and laboratorial data. They used a set of features from US, laboratorial and clinical. The results obtained were achieved with the SVM classifier with polynomial kernel of fourth degree, achieving a sensitivity of 91.67%. However, a multi classifier can be used to improve the accuracy.

The aforementioned studies show SVM and other methods in classification of tissue ultrasound images. However, some CAD systems do not have image preprocessing and image segmentation components. Thus the features extracted directly from ROIs may not provide robust and accurate performance. On the other hand the others employed traditional methods (e.g. ANNs), which aim at minimizing the empirical training error. Different to SVMs which have achieved superior performance in a wide range of applications, by minimizing the upper bound of the generalization error through maximizing the margin between the separating hyperplane and the data. The proposed Multi-SVM based CAD system outperforms these studies yielded classification accuracy of 96.11% with the individual class accuracy of 97.78%, 95.56%, 93.33% and 97.78% for NOR, Cyst, HEM and HCC cases respectively. The good performance of the proposed method shows a reliable indicator that can improve the information in the staging of focal liver lesion diseases.

### 3. MATERIALS AND METHODS

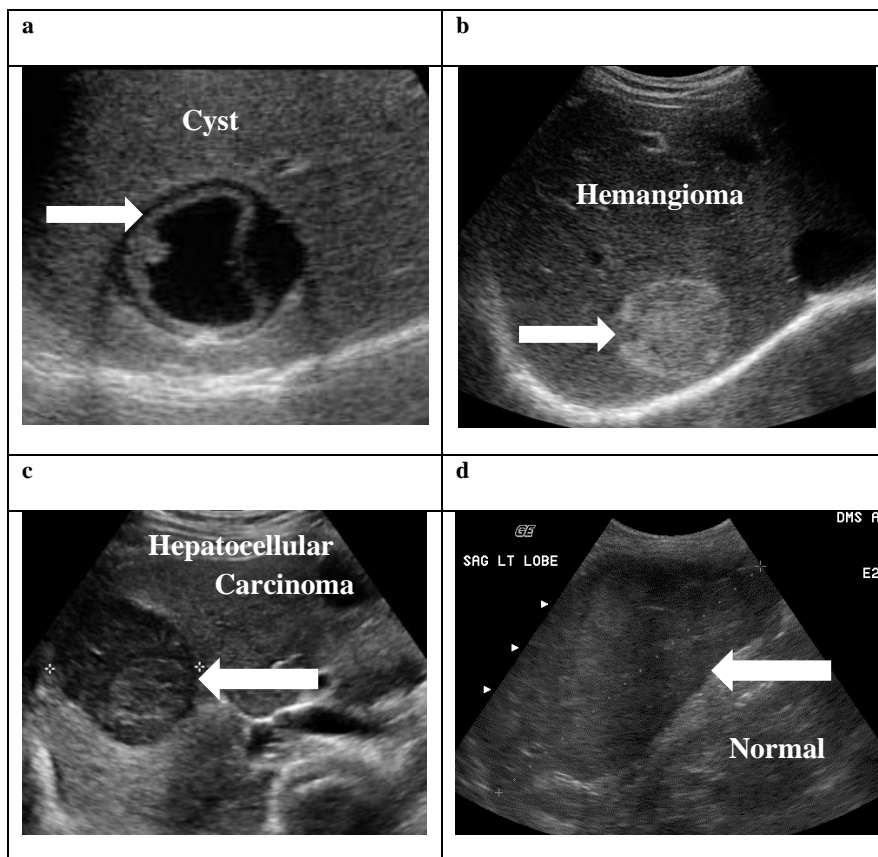
#### 3.1. Data Acquisition

To the best of our knowledge, till date the research in the area of liver disease diagnosis using conventional B-Mode liver US images has been carried out using individual databases collected by the efforts of individual research groups due to non-availability of benchmark image database.

From among 150 different liver diseases pictures we have chosen 94 with best quality and best applicability in pattern recognition area. The age of the patients in the image dataset ranges from 38 to 78 years. Patient's privacy has been protected by labeling the data with numeric dummy values and keeping patients' credential undisclosed. Data was collected from the Egyptian Liver Research Institute and General Hospital in Sherbin, Dakhalia Governorate. Those pictures represent 3 different focal liver lesion diseases (see Fig. 1) which are shortly described in this section.

These focal lesions are (i) Cyst (liquid nodule), (ii) HEM (non-cancerous solid nodule) and (iii) HCC (cancerous solid nodule). The appearances/features in a focal liver lesion may lead to misdiagnosis and confusion with other lesions even for the expert radiologists.

Liver cysts are the earliest stage of the liver disease. The liver cyst contour is smooth, but contrast between normal liver and cyst tissue is high. i.e., the gray level of cyst tissue is much darker than that of normal tissue as shown in Fig. 1(a). This dark tissue represents presence of fluid. Hemangioma (HEM) is the most common benign tumor of the liver. It affects the liver's blood vessels. HEM can have a wide range of sonographic appearance. Typical HEMs are hyperechoic to normal liver and their texture patterns are homogeneous as shown in Fig. 1(b). The most malignant liver tumor is Hepatocellular carcinoma (HCC) which in the end-stage of every chronic liver disease. These lesions appear hypoechoic with heterogeneous texture on US as shown in Fig. 1(c). NOR liver tissues have isoechoic appearance and homogeneous texture and it is shown in Fig. 1(d).



**Fig. 1. Ultrasound liver images with typical appearances of (a) Cyst, (b) Hemangioma, (c) Hepatocellular carcinoma and (d) Normal liver.**

In most cases, the sonographic appearance of Cysts, HEM and HCC overlaps sufficiently to make a distinction difficult or impossible. The present classification study is concentrated on classification among four liver image classes including three common focal liver lesions along with NOR liver tissues. In order to classify focal liver lesions, it has been widely accepted that the characteristics of the lesions have to be well represented. To represent the characteristics of lesions a combination of haralick descriptors, statistics and histogram features were extracted and feed to Multi classifier SVM.

### 3.2 Pre-Processing Steps

The next two subsections will explore tools for image enhancement and noise removal, and segmenting the ROI.

#### 3.2.1 Image enhancement and noise removal

The contrast of the US image is enhanced by using a new methodology which is a spatial filter but nonlinear one called bilateral filter. The bilateral filter is a local filter, but non-linearly considers both gray level similarities and geometric closeness of the neighboring pixels without smoothing edges [14].

#### 3.2.2 Segmentation of Region-Of-Interest

Selection of ROI significantly impacts the classification performances, thus ROI selection was done in an automatic way with the aid of Fuzzy c-means and level set techniques. The integration between fuzzy c-means and 'level set' technique assists to extract contours of liver tumors from ultrasound images with a very high reliability.

### 3.3 Feature extraction and selection

Feature extraction and selection is one of the critical steps in ultrasound CAD systems [1-2, 5, 7,11-12]. For the liver case, the most commonly used features are related to textural measures by constructing spatial gray-level dependence matrices [2, 7, 11-12], also termed as co-occurrence matrices that were introduced by haralick [20]. These features are normalized in the range of [0,1] and then used as input to the SVM classifier.

In the proposed CAD system, we analyze and extract haralick and histogram based features from the suspicious areas and ROIs. Furthermore, based on a literature review [6, 13, 17, 19, 24] it was found that the most common parameters were applied are distance length from  $d = \{1, 2, 3 \text{ and } 4\}$  pixels with orientations of  $\theta = (0^\circ, 45^\circ, 90^\circ \text{ and } 135^\circ)$  shown if Fig. 2. However, most of these research works calculate the co-occurrence matrix either across certain displacements and angles with no justification on the selection of the specific values of the two parameters [12] investigated the usefulness of using all possible angles and all displacements for capturing the texture properties of an organ of interest, specifically, the liver. Based on the Analysis of Variance (ANOVA) technique and multiple pair-wise comparisons [6], they found that using only the "near" and "far" displacements is enough to capture the spatial properties of the texture for the liver.

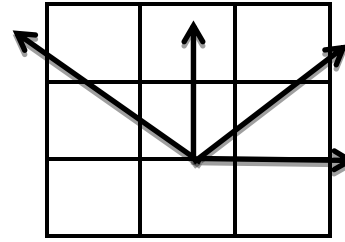


Fig. 2. The four directions of adjacency for calculating the Haralick texture features.

The Haralick Texture Features comprises 22 features [3,15]. A co-occurrence matrix  $C(\Delta x, \Delta y)$  with the offset  $(\Delta x, \Delta y)$  is defined over an  $n \times m$  Image  $I$ :

$$C_{\Delta x, \Delta y}(i, j) = \sum_{p=1}^n \sum_{q=1}^m \begin{cases} 1, & I(p, q) = i \text{ and } I(p + \Delta x, q + \Delta y) = j \\ 1, & I(p, q) = j \text{ and } I(p - \Delta x, q - \Delta y) = i \\ 0, & \text{Otherwise} \end{cases} \quad (1)$$

The co-occurrence matrix can be interpreted as a matrix of frequency from neighboring pixels in image  $I$  with an offset  $(\Delta x, \Delta y)$  where a pixel has the gray level  $i$  and the other pixel a gray level  $j$ . Note that this matrix is symmetric.

Let

$$P_{\Delta x, \Delta y} = \frac{1}{R} C_{\Delta x, \Delta y} \quad (2)$$

be the normalized co-occurrence matrix from  $C(\Delta x, \Delta y)$  with

$$R = \sum_{p=1}^n \sum_{q=1}^m C_{\Delta x, \Delta y}(i, j), \quad (3)$$

being the normalized factor.

Let  $p_{ij}$  be the  $(i, j)$ -th element of matrix  $P(\Delta x, \Delta y)$ , with  $N_g$  being the number of distinct gray levels in the image  $I$ . The  $i$ -th entry in the marginal probability matrix obtained by summing the rows of  $p_{ij}$  is

$$P_x(i) = \sum_{j=1}^{N_g} p_{ij} \text{ and } P_y(i) = \sum_{i=1}^{N_g} p_{ij} \text{ for the lines, respectively}$$

Further we are calculating:

$$P_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_{ij}, \quad i + j = k, \quad k = 2, 3, \dots, 2N_g \quad (4)$$

$$P_{x-y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_{ij}, \quad |i - j| = k, \quad k = 0, 1, \dots, N_g \quad (5)$$

since  $P_{(\Delta x, \Delta y)}$  is also a symmetric matrix  $p(i) = p_x(i) = p_y(i)$ .

This results in:

$$\mu = \mu_y = \mu_x = \sum_{k=1}^{N_g} k P(k) \quad (6)$$

$$\sigma = \sum_{k=1}^{N_g} P(k) (k - \mu)^2 \quad (7)$$

The Haralick Texture Features are defined as follows [3,15]:

#### Energy

This parameter is a measure of image homogeneity; it also called Angular second moment or Uniformity, it reflects pixel-pair repetitions. Homogeneous images have very few dominant gray tone transitions, which result into higher energy. Energy is defined as follows:

$$Energy = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij}^2 \quad (8)$$

#### Entropy

The feature entropy is a measure of non-uniformity (randomness) of a gray-level distribution in the image or region of interest. If the image is heterogeneous, many

elements on the co-occurrence matrix have small values, which imply that entropy is very large. Entropy is inversely correlated to energy, it is given by the following expression:

$$Entro = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij} \log P_{ij} \quad (9)$$

Sum Entropy:

$$SEnt = - \sum_{k=2}^{2N_g} P_{x+y}(k) \log P_{x+y}(k) \quad (10)$$

Difference Entropy

$$DEnt = - \sum_{k=0}^{N_g-1} P_{x-y}(k) \log P_{x-y}(k) \quad (11)$$

### Contrast

This parameter also called Inertia. It is a measure of the local variations of gray levels present in an image. Images with large neighboring gray level differences are associated with high contrast. This parameter can also characterize the dispersion of the matrix values from its main diagonal. Contrast is defined as follows:

$$Contr = \sum_{k=0}^{N_g} k^2 \left\{ \begin{array}{l} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij} \\ |i - j| = k \end{array} \right. \quad (12)$$

Where  $g(i, j)$  corresponds to the elements of co-occurrence matrix, ie the probability of moving from a pixel with gray level  $i$  to a pixel with gray level  $j$ .

### Homogeneity

This parameter, called also Inverse Difference Moment, measures the local homogeneity of an image. It assigns larger values to smaller gray level differences within pixel pairs. This parameter has opposite behavior of the contrast. More the texture has homogeneous regions, more the parameter is high. Homogeneity is written as:

$$Homo = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{1}{1+(i-j)^2} P_{ij} \quad (13)$$

### Sum Average

Provide the average of the gray levels in the image. The Sum Average is expected to be large if the sum of the gray levels of the image is high, and is written as:

$$SAvg = \sum_{k=2}^{2N_g} k P_{x+y}(k) \quad (14)$$

### Sum of Squares

Variance: It is a measurement of heterogeneity and was correlated strongly with standard deviation. It characterizes the distribution of gray levels around the mean. Therefore, variance increased when the gray levels values differed from their means. The expression of the Sum of Squares: Variance is:

$$SOSV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 P_{ij} \quad (15)$$

### Sum Variance

$$SVar = \sum_{k=2}^{N_g} (i - SAvg)^2 P_{x+y}(k) \quad (16)$$

### Difference Variance

$$DVar = \text{variance of } p_{x-y} \quad (17)$$

### Correlation

Measures a correlation of pixel pairs on gray-levels. Correlation is defined as follows:

$$corr = \frac{1}{\sigma_x \sigma_y} \left( \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ij P_{ij} - \mu_x \mu_y \right) \quad (18)$$

### Autocorrelation

$$Autoc = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (ij) P_{ij} \quad (19)$$

### Dissimilarity

$$Dissi = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} k P_{ij} \quad (20)$$

### Maximum Probability (MP)

Determines the most predominant pixel pair in an image. Result is retrieved from maximum value in the pixel pair that is most predominant in the image. The maximum probability is expected to be high if the occurrence of the most predominant pixel pair is high. Maximum probability is given by:

$$MaxPr = \max_{\substack{i \text{ to } N_g \\ j \text{ to } N_g}} P_{ij} \quad (21)$$

### Cluster Tendency (CT)

Measures the grouping of pixels that have similar gray-level values. Cluster Tendency is defined as:

$$CTend = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - 2\mu)^m P_{ij} \quad (22)$$

### Cluster Prominance, in case m=3

$$CPro = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - 2\mu)^3 P_{ij} \quad (23)$$

### Cluster Shade, in case m=4

$$CShad = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - 2\mu)^4 P_{ij} \quad (24)$$

### Information Measure of Correlation 1

$$Inf1 = \frac{Entro - HXY1}{H} \quad (25)$$

### Information Measure of Correlation 2

$$Inf2 = \sqrt{1 - e^{[-2(HXY2 - Entro)]}} \quad (26)$$

Where

$$HXY1 = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij} \log P_x(i) P_y(j)$$

$$HXY2 = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_x(i) P_y(j) \log P_x(i) P_y(j)$$

$$H = \sum_{k=1}^{N_g} P \log P$$

### Inverse difference normalized (INN)

$$Indn = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{1}{\frac{1+(i-j)^2}{N_g}} P_{ij} \quad (27)$$

### Inverse difference moment normalized (IDN)

$$Idmn = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{1}{\frac{1+(i-j)^2}{N_g * N_g}} P_{ij} \quad (28)$$

## 3.4. Histogram Based-Features.

Histogram based approach to texture analysis is based on the intensity value concentrations on all or part of an image represented as a histogram. The intensity-level histogram is a function showing (for each intensity level) the number of pixels in the whole image, which have this intensity:

$$h(i) = \sum_{x=1}^N \sum_{y=1}^M \delta(f(x, y), i) \quad (29)$$

where  $\delta(j, i)$  is the Kronecker delta function.

$$\delta(j, i) = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases}$$

Dividing the values  $h(i)$  by the total number of pixels in the image one obtains the approximate probability density of occurrence of the intensity levels.

$$P(i) = h(i)/NM, \quad i=1,2,\dots,Ng$$

The histogram can be easily computed, given the image. The shape of the histogram provides many clues as to the character of the image. Most often the so-called central moments are derived to characterize the texture [14], as defined by Equations (30-35) below.

**Mean**

$$\mu = \mu_y = \mu_x = \sum_{k=1}^{Ng} k P(k) \quad (30)$$

**Variance**

$$\sigma^2 = \sum_{k=1}^{Ng} P(k) (k - \mu)^2 \quad (31)$$

**Skewness**

$$\mu_3 = \sigma^{-3} \sum_{k=1}^{Ng} P(k) (k - \mu)^3 \quad (32)$$

**Kurtosis**

$$\mu_4 = \sigma^{-4} \sum_{k=1}^{Ng} P(k) (k - \mu)^4 \quad (33)$$

**Energy**

$$Energy = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} P_{ij}^2 \quad (34)$$

**Entropy**

$$Entro = - \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} P_{ij} \log P_{ij} \quad (35)$$

The mean takes the average level of intensity of the image or texture being examined, whereas the variance describes the

variation of intensity around the mean. The skewness is zero if the histogram is symmetrical about the mean, and is otherwise either positive or negative depending whether it has been skewed above or below the mean. Thus  $\mu_3$  is an indication of symmetry [27]. The kurtosis is a measure of flatness of the histogram. The entropy is a measure of histogram uniformity. Other possible features derived from the histogram are the minimum, the maximum, the range and the median value [27].

**3.5. Classifying by SVM**

The criterion used by SVMs is based on margin maximization between the two data classes. The margin is the distance between the hyper planes bounding each class. By maximizing the margin, we search for the classification function that can most safely separate the classes of normal from infected tissue. The threshold separating normal and infected tissue is the line in the middle between the two margin boundaries, which are represented as  $x^T w + b = 1$  and  $x^T w + b = -1$ . Then the margin is  $2 / \|w\|$ , where  $\|w\|$  is the norm of the vector  $w$ . The separating hyperplane has to be determined in such a way that the margin between positive class and a negative class is maximized to produce good generalization ability [9].

**4. RESULTS AND DISCUSSION**

Fig. 3 shows the images of all three liver image classes and their enhanced counterparts accompanied with segmentation step. Fig. 3 A1, B1, and C1, are original images of HCC, HEM and Cyst liver diseases, respectively. Fig.3 A2, B2, and C2, are corresponding enhanced images using bilateral non-linear filter and the surrounded regions using level set method in Fig. 3 show the lesions. These contours are used to initialize FCM segmentation technique. This resulted in a reliable segmentation result using FCM initialized by level set shown by fig. 3. A3, B3 and C3.

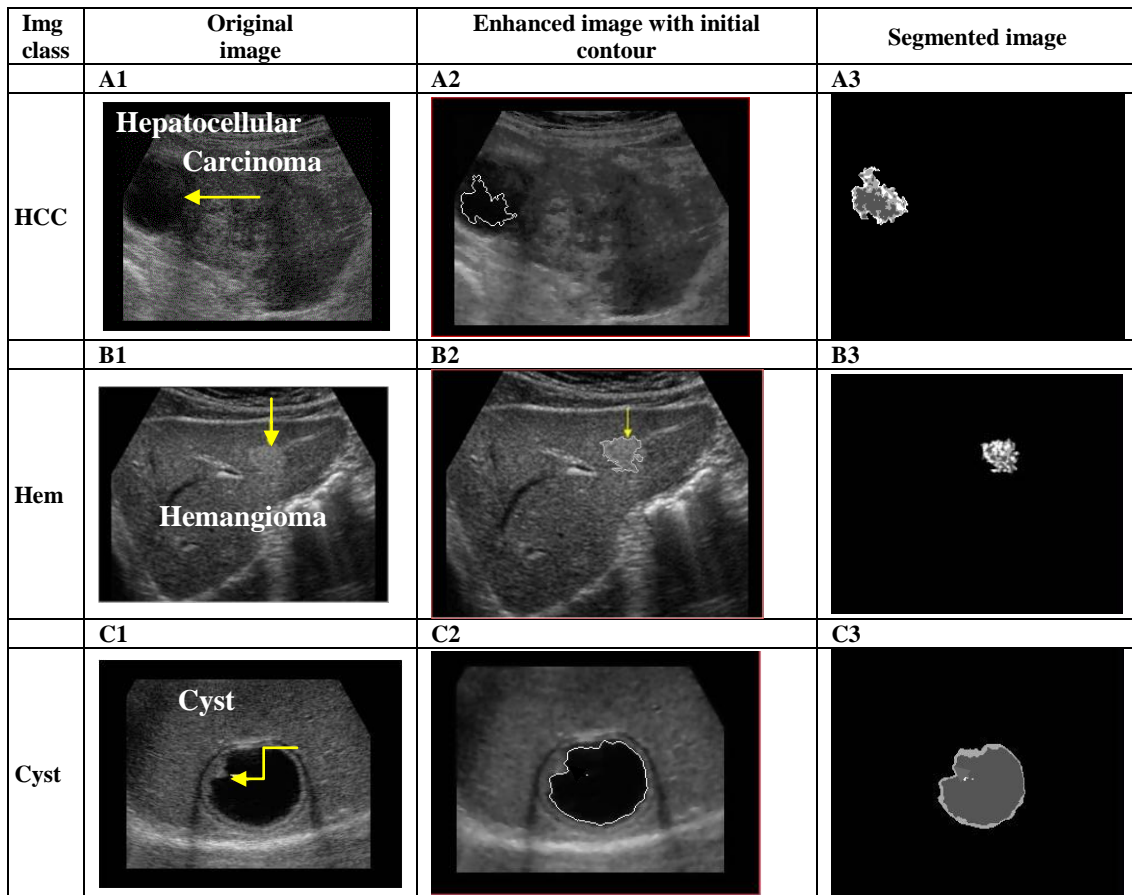


Fig. 3. Original, enhanced with initial contour, and segmented ultrasound images of three liver image categories.

In this work, a total of 94 features have been extracted from each ROI of liver ultrasound images, namely 6 histogram features and 88 GLCM based texture features. We have measured these features from ROI for every image of three categories along with normal. After dimensionality reduction using rough sets we obtain 38 features. These features are used to train and test the Multi-SVM classifier. The multi-class classification problem is commonly solved by decomposition to several binary problems for which the standard SVM can be used. For instance, one-against-all (O-A-A) decomposition is often applied. In this case the classification problem to  $k$  classes is decomposed to  $k$ -dichotomic decisions  $f_m(x)$ ,  $m \in K = \{1, \dots, k\}$ , where the rule  $f_m(x)$  separates training data of the  $m$ -th class from the other training patterns. The classification of a pattern  $x$  is performed according to maximal value of functions  $f_m(x)$ ,  $m \in K$ , i.e., the label of  $x$  is computed as  $\text{argmax}_{m \in K} f_m(x)$ . The OAA method [16] is the earliest used method for SVM multiclass classification. For an  $m$ -class classification problem, the OAA approach constructs  $m$  SVM classifiers. The  $i$ th SVM separates class  $i$  with positive labels from all the others with negative labels. For a test sample  $x$ , the final prediction is based on the integration of  $m$  classification results. Each  $x$  is classified as in class  $i$  when the  $i$ th classifier has the highest value of the decision function. In our study, given 4 classes, 4 independent binary SVM classifiers are constructed where the first classifier discriminates primary cyst (class label assigned to 1) from hemangioma, hepatocellular carcinoma and normal liver (class label assigned to 0), the second discriminates hemangioma (class label assigned to 1) from cyst, hepatocellular carcinoma and normal liver (class label assigned to 0), the third discriminates hepatocellular

carcinoma (class label assigned to 1) from cyst, hemangioma and normal liver (class label assigned to 0) and the Fourth discriminates normal liver (class label assigned to 1) from cyst, hemangioma and hepatocellular carcinoma (class label assigned to 0).

Table 1 shows Individual class validity measures for the classification using proposed approach. It can be observed from Table 1 that focal liver lesion image classification using these features shows significant performance in terms of the four objective measurements accuracy, sensitivity, specificity, and Negative Predictive value (NPV).

Table 1. Individual class validity measures for the classification using proposed approach.

		Cysts	Heman	HCC	Normal
Validity Measures	Accuracy	95.56	93.33%	97.78%	97.78%
	Sensitivity	97.3%	93.75%	97.3%	100%
	Specificity	87.5%	92.31%	100%	87.5%
	NPV	87.5%	85.71%	88.88%	100%

Furthermore, to check the performance of our proposed approach, we have applied KNN on the given dataset. Table 2 shows Individual class validity measures for the classification using K-Nearest Neighbor approach.

**Table 2. Individual class validity measures for the classification using KNN approach**

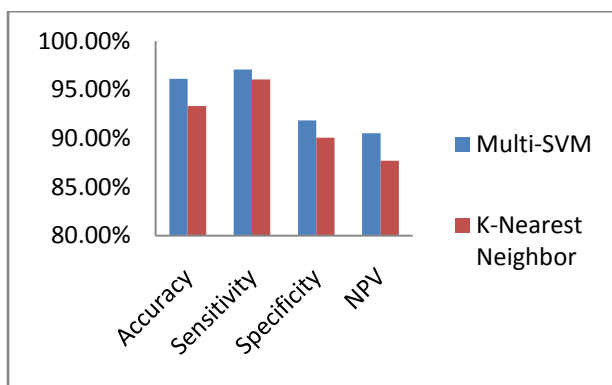
		Cysts	Heman	HCC	Normal
Validity Measures	Accuracy	91.11%	91.11%	95.5%	95.5%
	Sensitivity	89.66%	100%	94.6%	100%
	Specificity	93.75%	69.23%	100%	97.3%
	NPV	83.33%	100%	80%	87.5%

It can be observed from Table 3 that focal liver lesion image classification using these features shows significant performance in terms of the four objective measurements accuracy, sensitivity, specificity, and Negative Predictive value (NPV).

**Table 3. Performance comparison of the proposed approach with KNN approach**

		Accuracy	Sensitivity	Specificity	NPV
Classifier	Multi-SVM	96.11%	97.08%	91.83%	90.5%
	KNN	93.3%	96.06%	90.07%	87.71%

Table 3 shows comparison between the proposed approach and KNN at different quality overall measures. From Table 1, it is clear that Multi-SVM offers superior results compared to other techniques. Statistical analysis of results represents the usefulness of the proposed approach. Figure 4 shows the graphical performance comparison of KNN, and Multi-SVM at various classification validity measures. Classification using Multi-SVM offers superior performance compared to K-NN technique at all validity measures. The good performance of the proposed method shows a reliable indicator that can improve the information in the staging of focal liver lesion diseases.



**Fig. 4. Performance comparison between the proposed multi SVM and KNN classification techniques at different classification validity measures.**

## 5. CONCLUSION

The paper uses the Multi-SVM in classification of various stages of focal liver lesions. Another major contribution of this research work is development of an automatic classification system for focal liver lesions in ultrasound images. Therefore, it is useful to discriminate between focal liver diseases such as Cyst, Hemangioma, and Hepatocellular carcinoma along with Normal liver. The multi-class scenario is a composition of a series of two-class problems, using one-

against-all which is the earliest and one of the most widely used implementations. The selection of ROI was done with the aid of Fuzzy c-means and level set. The estimated features extracted using haralick local texture descriptors and histogram based features. Different types of validity measures are calculated to show the effectiveness of the proposed approach. Using 10-fold cross validation to train and test Multi-class classifier system, we have obtained 96.11% with the individual class accuracy of 97.78%, 95.56%, 93.33% and 97.78% for NOR, Cyst, HEM and HCC cases respectively. A numerical comparison is made with KNN at different validity measures. Statistical analysis shows that the proposed approach outperforms other techniques in literature. The good performance of the proposed method shows a reliable indicator that can improve the information in the staging of focal liver lesion diseases.

## 6. REFERENCES

- [1] Abu Sayeed Md. Sohail Prabir Bhattacharya, S. P. M. S. K. & Gilbert, L. Content-Based Retrieval and Classification of Ultrasound Medical Images of Ovarian Cysts Springer-Verlag Berlin Heidelberg, pp.173–184, 2010.
- [2] Andreia Andrade Jose, Silvestre Silva. Classifier Approaches for Liver Steatosis using Ultrasound Images Procedia Technology, vol.5, pp.763–770, 2012.
- [3] Anja Attig, P. P. A Comparison between Haralick’s Texture Descriptor and the Texture Descriptor Based on Random Sets for Biological Images Springer, Machine Learning and Data Mining in Pattern Recognition, Vol. 6871, 524-538, 2011.
- [4] Asmatullah Chaudhry, Mehdi Hassan, Asifullah Khan, Jin Young Kim. Automatic Active Contour-Based Segmentation and Classification of Carotid Artery Ultrasound Images J Digit Imaging, Society for Imaging Informatics in Medicine, Springer, 2013.
- [5] Bo Liu H.D. Cheng, J. H. Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images Pattern Recognition, vol.43, pp.280–298, 2010.
- [6] Clausi, D. A. An analysis of co-occurrence texture statistics as a function of grey level quantization Can. J. Remote Sensing, Vol. 28, No. 1, pp.45–62, 2002.
- [7] Deepti Mittal, Vinod Kumar. Neural network based focal liver lesion diagnosis using ultrasound images Computerized Medical Imaging and Graphics, vol.35, pp.315–323, 2011.
- [8] Dimitris K. Iakovidis, Eystratios G. Keramidas, Fusion of fuzzy statistical distributions for classification of thyroid Artificial Intelligence in Medicine, vol.50, pp.33–41, 2010.
- [9] Heena Farooq Bhat, M. A. W. Modified One-Against-All Algorithm Based on Support Vector Machine International Journal of Advanced Research in Computer Science and Software Engineering, Vol.3, Issue 12, December 2013.
- [10] Hui-Ling Chen, Chang-Cheng Huang, Xin-Gang Yu, Xin Xu, Xin Sun, Gang Wangd, Su-Jing Wang. An efficient diagnosis system for detection of Parkinson’s disease using fuzzy k-nearest neighbor approach Expert Systems with Applications, Vol. 40, pp.263–271, 2013.



- [11] Jitendra Virmani Vinod Kumar, N. K. & Khandelwal, N. PCA-SVM based CAD System for Focal Liver Lesions using B-Mode Ultrasound Images Defence Science Journal, vol.63, pp.478-486, 2013.
- [12] Johnson, R. S. D. J. F. T. B. A co-occurrence texture semi-invariance to direction, distance, and patient size Proc. SPIE 6914, Medical Imaging, Image Processing, 69141Y, March 11, 2008.
- [13] Jong Kook Kim ; Seoul, S. K. ; J. M. P. ; K. S. S. ; H. P. Texture analysis and artificial neural network for detection of clustered microcalcifications on mammograms Neural Networks for Signal Processing Proceedings of ANN, IEEE Workshop, V.28, pp.199 – 206, 2007.
- [14] Kumar, B. K. S. Image denoising based on gaussian/bilateral filter and its method noise thresholding Springer-Verlag,SIVIP, London, vol. 7, pp.1159–1172, 2013
- [15] Leen-Kiat Soh & Costas Tsatsoulis Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 37, NO. 2, pp.780-795, 2009.
- [16] Liu, Y. & Zheng, Y. F. One-Against-All Multi-Class SVM Classification Using Reliability Measures IEEE International Joint Conference on Neural Networks, Vo.2, pp.849–854, Aug. 2005.
- [17] M.VASANTHA DR.V.SUBBIAH BHARATHI, R. Medical Image Feature, Extraction, Selection And Classification International Journal of Engineering Science and Technology, Vol. 2(6), pp.2071-2076 ,2010.
- [18] Materka, A, M. S. Texture Analysis Methods–A Review University of Lodz, Institute of Electronics, COST B11 report, Brussels, 1999.
- [19] Olfa Ben Sassi Lamia sellami, M. B. S. K. C. A. B. H. Improved Spatial Gray Level Dependence Matrices For Texture Analysis, Int. Journal of Computer Science & Information Technology (IJCSIT), Vol 4, No 6, pp.209-219, 2012.
- [20] R. Haralick K. Shanmugam & I. Dinstein Textural Features for Image Classification, IEEE Trans. on Systems, Man and Cybernetics), SMC–vol.3, issue 6, pp. 610–621, 2003.
- [21] Ricardo Ribeiro Rui Marinho, J. V. F. R. & Sanches, J. M. Chronic Liver Disease Staging Classification Based On Ultrasound, Clinical And Laboratorial Data, IEEE International Symposium, Biomedical Imaging: From Nano to Macro, pp.707-710, 2011.
- [22] Ricardo Ribeiro Rui Marinho, J. V. F. R. & Sanches, J. M. Diffuse Liver Disease Classification from Ultrasound Surface Characterization, Clinical and Laboratorial Data Springer-Verlag Berlin Heidelberg, pp.167–175, 2011.
- [23] Ricardo Ribeiro, Rui Tato Marinho, The Usefulness of Ultrasound in the Classification of Chronic Liver Disease IEEE Annual International Conference, Engineering in Medicine and Biology Society, EMBC, pp.5132-5135, 2011
- [24] Shichong Zhou Jun Shi, J. Z. Y. C. R. W. Shearlet-based texture feature extraction for classification of breast tumor in ultrasound image ELSEVIER, Science Direct, Biomedical Signal Processing and Control, pp.688– 696, vol.8, 2013.
- [25] Vijay Garla Caroline Taylor, C. B. Semi-supervised clinical text classification with Laplacian SVMs: An application to cancer case management Journal of Biomedical Informatics, vol.46, Issue 5, pp.869-875, 2013.
- [26] Wen-Jie Wu Shih-Wei Lin, W. K. M. Combining support vector machine with genetic algorithm to classify ultrasound breast tumor images Computerized Medical Imaging and Graphics, vol.36, pp.627– 633, 2012.
- [27] Xiangjun Shi, H.D. Cheng, Liming Hu, Wen Ju, Jiawei Tian. Detection and classification of masses in breast ultrasound images Digital Signal Processing, vol.20, pp.824–836, 2010.