Defence Science Journal, Vol. 63, No. 5, September 2013, pp. 478-486, DOI : 10.14429/dsj.63.3951

PCA-SVM based CAD System for Focal Liver Lesions using B-Mode Ultrasound Images

Jitendra Virmani[*], Vinod Kumar, Naveen Kalra[#], and Niranjan Khandelwal[#]

Indian Institute of Technology Roorkee, Roorkee - 247 667, India
[#]Post Graduate Institute of Medical Education and Research, Chandigarh -160 012, India
[*]E-mail: jitendra.virmani@gmail.com

ABSTRACT

The contribution made by texture of regions inside and outside of the lesions in classification of focal liver lesions (FLLs) is investigated in the present work. In order to design an efficient computer-aided diagnostic (CAD) system for FLLs, a representative database consisting of images with (1) typical and atypical cases of cyst, hemangioma (HEM) and metastatic carcinoma (MET) lesions, (2) small as well as large hepatocellular carcinoma (HCC) lesions and (3) normal (NOR) liver tissue is used. Texture features are computed from regions inside and outside of the lesions. Feature set consisting of 208 texture features, (i.e. 104 texture features and 104 texture ratio features) is subjected to principal component analysis (PCA) for finding the optimal number of principal components to train a support vector machine (SVM) classifier for the classification task. The proposed PCA-SVM based CAD system yielded 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. The accuracy for typical, atypical, small HCC and large HCC cases is 87.5%, 86.8%, 88.8%, and 87% respectively. The promising results indicate usefulness of the CAD system for assisting radiologists in diagnosis of FLLs.

Keywords: Focal liver lesions, B-mode ultrasound, principal component analysis, support vector machine classifier, computer-aided diagnostic system

NOMENCLATURE

B-Mode US	B-Mode Ultrasound: Brightness-Mode ultrasound is a two dimensional representation of echo-producing interfaces in a single plane.
FLL	Focal Liver Lesion: Focal liver lesion refers to area of liver tissue damage.
Benign FLL	Benign Focal Liver Lesion: Non-cancerous focal liver lesion.
Malignant FLL	Malignant Focal Liver Lesion: Cancerous focal liver lesion.
HFLL	Hyperechoic Focal Liver Lesion: The focal liver lesion with more echogenicity as compared to the surrounding liver parenchyma.
Isoechoic FLL	Isoechoic Focal Liver Lesion: The focal liver lesion with same echogenicity as that of the surrounding liver parenchyma.
HFLL	Hypoechoic Focal Liver Lesion: The focal liver lesion with less echogenicity as compared to the surrounding liver parenchyma.
Anechoic FLL	Anechoic Focal Liver Lesion: The focal liver lesion which appears without echoes on ultrasound.
Typical FLL	Typical Focal Liver Lesion: Focal liver lesions with classic diagnostic sonographic appearance.
Atypical FLL	Atypical Focal Liver Lesion: Focal liver lesion with non-specific sonographic appearance.
NOR	Normal Liver: Normal liver has homogeneous texture with medium echogenicity (i.e. same or slightly increased echogenicity compared to the right kidney).
CYST	Liver Cyst: Abnormal fluid filled sacs in the liver.
Typical Cyst	Typical Cyst: Well defined, round, anechoic lesion with posterior acoustic enhancement and thin imperceptible wall.
Atypical Cyst	Atypical Cyst: Appear with irregular, thickened wall and internal echoes.
HEM	Hemangioma: The most common primary benign focal liver lesion.
Typical HEM	Typical Hemangioma: Appear as a well circumscribed uniformly hyperechoic lesion.
Atypical HEM	Atypical Hemangioma: Appear as isoechoic or even hypoechoic lesion.
HCC	Hepatocellular Carcinoma: The most common primary malignant focal liver lesion.
SHCC	Small Hepatocellular Carcinoma: HCC lesions (<2cms), appearance vary from hypoechoic to hyperechoic lesions.
LHCC	Large Hepatocellular Carcinoma: HCC

Received 28 February 2013, revised 6 July 2013, online published 25 September 2013

	lesions (>2cms), appear as lesion with mixed echogenicity.
MET	Metastasis: The most common secondary malignant focal liver lesion.
Typical MET	Typical Metastasis: Appears with 'target' or 'bull's-eye' appearance.
Atypical MET	Atypical Metastasis: Appearance extremely variable ranging from anechoic, hypoechoic, isoechoic, hyperechoic and even with mixed echogenicity.

1. INTRODUCTION

Typical focal liver lesions (FLLs) can be easily identified from their classic sonographic appearances but differential diagnosis between atypical FLLs is considered a confusing task for radiologists in routine practice, due to existence of large variety of sonographic appearances even with-in individual classes of atypical FFLs[1-6]. Even then, B-Mode ultrasound (US) is the first investigation for diagnosis of FLLs mainly due to its inexpensive, noninvasive, nonradioactive nature and real time imaging capabilities. Therefore, there is a need to develop an efficient computer-aided diagnostic (CAD) system with a comprehensive and representative database consisting of images with (1) typical and atypical cases of Cyst, hemangioma (HEM) and metastatic carcinoma (MET) lesions, (2) small as well as large hepatocellular carcinoma (HCC) lesions and (3) normal (NOR) liver tissue.

The echotexture of NOR liver tissue is homogeneous with medium echogenicity (i.e. same or slightly increased echogenicity compared to the right kidney)[5-7]. All the NOR cases are considered as typical as there is no atypical appearance for NOR liver tissue.

Typical Cyst appears as well defined, round, anechoic lesion with posterior acoustic enhancement and thin imperceptible wall. Atypical cysts appear with irregular, thickened wall and internal echoes[1-3,5-7].

HEM is the most common primary benign FLL. Typical HEM appears as a well circumscribed uniformly hyperechoic lesion. Atypical HEMs are a great mimic and definite diagnosis with B-mode US is difficult as these lesions can be isoechoic or even hypoechoic, mimicking the sonographic appearance of certain HCC and atypical MET lesions[1-3,6].

HCC is the most common primary malignant FLL, which mostly develops on top of coarse and nodular cirrhotic liver background. In radiology practice, the condition of cirrhosis is seen as a precursor to the development of HCC[1-3,7-9]. Thus cirrhotic liver background favors the possibility of a lesion being HCC in its differential diagnosis with other FLLs. The associated radiologists opined that 'no sonographic appearance can be considered typical for HCC' due to high degree of variability in sonographic appearances even within small HCC (SHHC) and large HCC (LHCC) lesions however, a representative dataset for designing the classifier should contain both SHCC and LHCC cases. The sonographic appearance of SHCC (<2cms) vary from hypoechoic to hyperechoic. LHCC appears frequently with mixed echogenicity[4,6,7].

MET is the most common secondary malignant FLL. Typical MET appears with 'target' or 'bull's-eye' appearance (i.e. hypoechoic center surrounded by a hyperechoic rim). The sonographic appearance of atypical MET lesions are extremely variable ranging from anechoic, hypoechoic, isoechoic, hyperechoic and even with mixed echogenicity[1-5,11].

The sample image variants of NOR liver tissue, typical Cyst, typical HEM, typical MET, SHCC, and LHCC cases are shown in Fig. 1.

There are certain disadvantages associated with use of B-Mode US for diagnosis of FLLs, (1) limited sensitivity for detection of small FLLs (< 2 cms) developed on cirrhotic liver which is already nodular and coarse-textured[1-4,7], (2) sonographic appearances of HCC, atypical HEM and atypical MET are highly overlapping, (3) sonographic appearances of cystic MET and atypical Cyst is often overlapping, and (4) in certain cases, it is difficult to characterize isoechoic lesions with very slim difference in contrast between regions inside and outside of the lesion[5]. It is very much desired to reduce these limitations and built an efficient CAD system by designing the classifier with a comprehensive and diversified dataset consisting of representative images for various sub-classes.

It is worth mentioning that till date the research in the area of liver disease diagnosis using conventional B-Mode liver US images have been carried out using individual databases collected by the efforts of individual research groups due to non-availability of benchmark image database. Consequently, there are very few studies reported in literature, the brief detail of these studies[4-6,12-15] is depicted in Table 1.

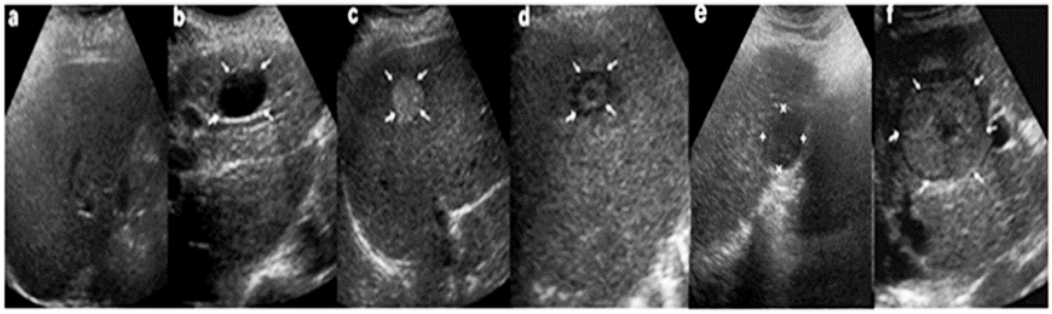

Figure 1. B-Mode ultrasound liver images with appearance of a. Normal liver (homogeneous echotexture with medium echogenicity) b. Typical cyst (thin walled anechoic lesion with posterior acoustic enhancement) c. Typical HEM (well circumscribed uniformly hyperechoic appearance) d. Typical MET ('target' or 'bull's-eye' appearance, i.e. hypoechoic center surrounded by a hyperechoic rim) e. Hypoechoic SHCC f. Heterogeneous echotexture represents complex and chaotic structure exhibited by LHCC due to coexistence of areas of necrosis, fibrosis and active growth areas.

Table 1. Brief details of studies for classification of FLLs using B-Mode US image

Authors	Image class	Dataset Description			
		Patients	**Images per class**	**No. of ROIs**	**ROI size**
Sujana[12], *et al*.	NOR, HEM, Malignant	-	-	113	10x10 pixels
	Classifier used	**Distribution of ROIs for classifier design**			
	Neural Network, LDA	**Training data**		**Testing Data**	
		NOR (40), HEM (15) and Malignant (30)		NOR (13), HEM (5) and Malignant (10)	
	Features used FOS, GLCM and GLRLM features				
Yoshida[13], *et al*.	HEM , Malignant (HCC+MET)	44	HEM(17),HCC(11) and MET(16)	193	64 x64 pixels
	Classifier used	**Cross validation procedure**			
	Neural Network	HEM (50) , HCC(87) and MET (56)			
	Features used Wavelet packet texture descriptors				
Poonguzhali[14], *et al*.	NOR, Cyst, HEM and Malignant	-	-	120	10x10 pixels
	Classifier used	**Cross validation procedure**			
	Neural Network	NOR (30) , Cyst (30) , HEM (30) and Malignant (30)			
	Features used Autocorrelation , Edge frequency, GLCM and Laws' TEM features				
Mittal[5], *et al*.	NOR, Cyst, HEM,HCC and MET	88	NOR(16),Cyst (17) HEM18) ,HCC(15) and MET(45)	800	25 x 25 pixels
	Classifier used	**Training data**	**Validation data**	**Testing data**	
	Neural Network	NOR(50),Cyst (50) HEM(50),HCC(50) and MET(50)	NOR(10),Cyst (10) HEM(10),HCC(10) and MET(10)	NOR (172),Cyst (6) HEM(30),HCC (167) and MET(125)	
	Features used FOS, GLCM, GLRLM, GWT and Laws' TEM features				
Virmani[5], *et al*.	HCC and MET	51	HCC(27),MET(24)	174 (120IROIs*, 54SROIs*)	32 x 32 pixels
	Classifier used	**Training data**		**Testing data**	
	SVM	HCC (30)and MET(30)		HCC (30) and MET(30)	
	Features used GLCM, GLRLM, FPS and Laws' TEM features				
Virmani[6], *et al*.	NOR, Cyst, HEM,HCC and MET	108	NOR(21),Cyst(12) HEM(15),HCC(28) and MET(32)	491 (380IROIs*, 111SROIs*)	32 x 32 pixels
	Classifier used	**Training data**		**Testing Data**	
	KNN, PNN, BPNN	NOR (30),Cyst (30), HEM(40), HCC (50) and MET(50)		NOR (40),Cyst (25),HEM (30), HCC (40) and MET(45)	
	Features used FOS, GLCM, GLRLM, FPS, GWT and Laws' TEM features				
Jeon[15], *et al*.	Cyst, HEM, and Malignant	102	Cyst(50), HEM(50), Malignant(50)	150	variable ROI sizes
	Classifier used	**Cross Validation Procedure**			
	SVM	Cyst(50), HEM(50), Malignant(50)			
	Features used FOS, GLCM, AMI, AC, GWT and Laws' TEM features				
Authors current study	NOR, Cyst, HEM,HCC and MET	108	NOR(21),Cyst(12) HEM(15),HCC(28) and MET(32)	491 (380IROIs*, 111SROIs*)	32 x 32 pixels
	Classifier used	**Training data**		**Testing Data**	
	SVM	NOR (30),Cyst (30), HEM(40), HCC (50) and MET(50)		NOR (40),Cyst (25),HEM (30), HCC (40) and MET(45)	
	Features used FOS, GLCM, GLRLM, FPS, GWT and Laws' TEM features				

Note: IROIs*: Inside ROIs, SROIs*: Surrounding ROIs, Texture features are computed from IROIs, SROIs are extracted to compute texture ratio features.

From Table 1, it is observed that statistical (FOS, GLCM, GLRLM), spectral (FPS, GWT) and spatial filtering based (Laws' TEM) texture features are important for classification of FLLs. The study[5] used, ROI size of 25x25 pixels for computing texture features however, in[12,14] the use of ROI size of 10x10 pixels is reported. The use of 10x10 pixels and even 25x25 pixels as ROI size yields smaller number of pixels in comparison to minimum 800 pixels required to estimate reliable statistics[4,7,16-18]. The study[13] reports the use of 64x64 pixels as ROI size extracted from high-resolution scanned images instead of real US images. It is otherwise difficult to select such a large ROI size keeping in view the size of small lesions and resolution of images obtained from US machines. The ROI size of 32x32 pixels was considered appropriate for the present study to estimate reliable statistics as well as to extract maximum ROIs from the acquired image database.

The researchers[13-15] experimented classification by considering malignant lesions as single class; however, the classification of malignant lesions as HCC or MET lesions is clinically significant for effective treatment of liver malignancies[4]. As per the best of the author's information, only two studies[5,6] are reported for five-class classification of FLLs using B-Mode US images. However, the proposed system design in[5] does not use any feature selection or dimensionality reduction methodology to deal with a large feature set consisting of 208 texture features. It is worth mentioning that some of these features may provide redundant information and therefore in the present work feature space dimensionality reduction is considered to design an efficient CAD system. The studies[5,12-14] experimented classification of FLLs by computing texture features from regions inside the lesions. In recent studies[4,6], it has been shown that the texture of the background liver contributes towards differential diagnosis between FLLs. The present work investigates the contribution of texture information from inside and outside the lesion in differential diagnosis between five liver image classes using PCA-SVM based CAD system design.

2. MATERIALS AND METHODS

2.1 Data Collection

It is worth mentioning that till date the research in the area of liver disease diagnosis using conventional B-Mode liver ultrasound images have been carried out on individual databases collected by the efforts of individual research groups due to non-availability of benchmark image database.

For the present work, 108 B-mode liver US images comprising of 21 NOR images, 12 Cyst images with 12 cystic lesions (8 typical cases and 4 atypical cases), 15 HEM images with 15 HEM lesions (8 typical cases and 7 atypical cases), 28 HCC images with 28 HCC lesions (13 SHCC cases and 15 LHCC cases) and 32 MET images with 35 MET lesions (12 typical cases and 23 atypical cases) were collected from the patients visiting the Department of Radiodiagnosis and Imaging, Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh, India over the time period from March 2010 to March 2012. Informed consent of patients for using these images for research was taken prior to recording. The medical ethics board of PGIMER, Chandigarh, granted the ethical clearance to carry out this research work. The direct digital images recorded by using Philips ATL HDI 5000 US machine equipped with multi-frequency transducer of 2-5 MHz range were used. The size of the images is 800 x 564 pixels with gray scale consisting of 256 tones and horizontal as well as vertical resolution is 96 dpi.

2.1.1 Liver Image Assessment Criteria

The gold standard for diagnosing liver malignancies is liver biopsy, and all the HCC and MET cases included in this study are biopsy proven cases, however, the typical and atypical cases of cysts and hemangioma have been verified by experienced participating radiologists with 13 and 23 years of experience in US imaging (co-authors of this paper) by using liver image assessment criteria, including (a) visualization of sonographic appearances, imaging features of FLLs based on their knowledge and expertise (b) follow-up of clinical history of the patient and other associated findings, and (c) imaging appearance on dynamic helical computed tomography (CT)/ magnetic resonance imaging (MRI) and (d) pathological examinations.

2.1.2 Data Collection Protocols

The protocols followed for data collection are (1). The diagnostic quality (free from artifacts) and representativeness of each image class was verified by participating radiologists (2). Only HCC lesions developed on top of cirrhosis were included as the existence of HCC on normal liver is very rare (3). Characterization of HCC lesions as SHCC or LHCC was made by observing the size of the lesion in transverse and longitudinal views (The HCC lesion less than 2cm in size is considered as SHCC). (4). The labeling of HCC lesions as SHCC or LHCC lesion and labeling of Cyst, HEM and MET lesions as typical or atypical lesion was done during data collection for training the classifier with representative data from each sub-class.

2.2 Selection of Regions of Interest (ROIs)

Two types of ROIs are used in this study, inside ROIs (IROIs) and surrounding ROIs (SROIs). These ROIs were cropped under the supervision of experienced participating radiologist. For each Cyst, HEM, HCC and MET lesion maximum non-overlapping IROIs were cropped from well within the boundary of each lesion. For each lesion a single SROI was cropped from surrounding liver parenchyma approximately at the same depth as that of center of the lesion by avoiding the inhomogeneous structures like, hepatic ducts and blood vessels etc. For each NOR image a single extreme ROI is considered as SROI and all other ROIs at the same depth are considered as IROIs. SROIs and IROIs for NOR image were cropped by avoiding the inhomogeneous structures like hepatic ducts and blood vessels etc. The NOR, Cyst, HEM HCC and MET B-mode US image variants from the acquired image database with ROIs marked are shown in Fig. 2.

In the present work, two types of features are considered for analysis, i.e. texture features computed from IROIs and texture ratio features computed by taking the ratio of texture feature computed from IROI and texture feature computed from corresponding SROI.

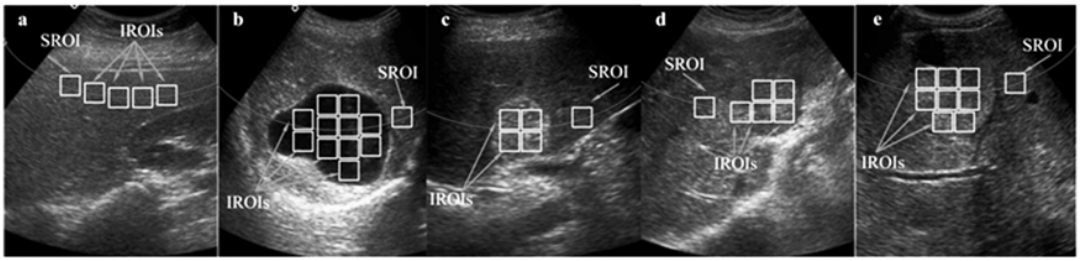

Figure 2. Figure (a) to (e) NOR, Cyst, HEM, HCC, and MET image with IROIs and SROI marked.

Note: As shown in (e) necrotic area within the lesions is avoided while cropping IROIs.

It can be noted that NOR liver image in Figure 2 (a) contains 4 IROIs and a corresponding SROI, thus 4 instances of texture feature set and 4 instances of texture ratio feature set are obtained. Similarly, from Cyst, HEM, HCC and MET lesions shown in Fig. 2(b)-2(e), 11, 4, 5 and 8 instances of texture feature set and 11, 4, 5 and 8 instances of texture ratio feature set are obtained.

2.3 Data Set Description

The distribution of acquired database with 108 B-mode liver US images among NOR, Cyst, HEM, HCC and MET image categories and the bifurcation of ROIs in training and test data set is depicted in Table 2.

The final data set consisting of total 111 SROIs and 380 IROIs was stored in a PC (Pentium Core-2-Duo, 2.67 GHz with 1.97 GB RAM).

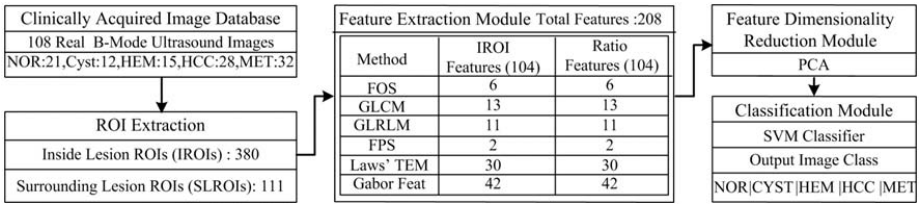

Figure 3. CAD system for classification of FLLs from B-mode US images.

Table 2. Data Set Description

Description: 108 B-Mode Liver US images									
NOR Images :21		Cyst Images:12		HEM Images:15		HCC Images: 28		MET Images: 32	
-		Cyst Lesions: 12		HEM lesions :15		HCC lesions:28		MET lesions: 35	
-		typ.:9	atyp.: 3	typ.:9	atyp.: 6	SHCC:13	LHCC:15	typ.:13	atyp.: 22
Description: 491 ROIs (SROIs: 111, IROIs: 380)									
Description: 111 SROIs									
NOR:21		Cyst:12		HEM:15		HCC:28		MET:35	
Description: 380 IROIs									
NOR:70		Cyst:55		HEM:70		HCC:90		MET:95	
typ.: 70	atyp.: Nil	typ.:35	atyp.:20	typ.:24	atyp.:46	SHCC: 19	LHCC:71	typ.:13	atyp.:82

Description: Training Data (SROIs: 59, IROIs:200)	**Description: Testing Data (SROIs:52, IROIs:180)**
NOR: Images: 10, SROIs: 10, IROIs:30	NOR: Images: 11, SROIs: 11, IROIs:40
Cyst: Images: 4, Lesions: 4, typ.:3, atyp.: 1 SROIs: 4, IROIs: 30, typ. IROIs: 22, atyp. IROIs:8	Cyst: Images: 8, Lesions: 8, typ.:6, atyp.: 2 SROIs: 8, IROIs: 25, typ. IROIs: 13, atyp. IROIs:12
HEM: Images: 9, Lesions: 9, typ.:6, atyp.: 3 SROIs: 9, IROIs: 40, typ. IROIs: 18, atyp. IROIs:22	HEM: Images: 6, Lesions:6, typ.:3, atyp.: 3 SROIs: 6, IROIs: 30, typ. IROIs: 6, atyp. IROIs:24
HCC: Images: 16, Lesions: 16, SHCC.:7, LHCC: 9 SROIs: 16, IROIs: 50, SHCC IROIs:10,LHCC IROIs:40	HCC: Images: 18, Lesions: 20, SHCC:8, LHCC: 12 SROIs: 12, IROIs: 40, SHCC IROIs:9, LHCC IROIs:31
MET: Images: 18, Lesions: 20, typ.:8, atyp.: 12 SROIs: 20, IROIs: 50, typ. IROIs: 8, atyp. IROIs:52	MET: Images: 14, Lesions: 15, typ.:5, atyp.: 10 SROIs:15, IROIs: 45, typ. IROIs: 5, atyp. IROIs:40

Note: typ.: Typical , atyp. : Atypical , The size of SHCC lesion varied from 1.5 to 1.9 cm and size of LHCC lesion varied from 2.1 to 5.6cm. **a.** Total lesions= 90 [(typ. lesions (31) + atyp. lesions (31) + SHCC lesions (13) + LHCC lesions (15)] **b.** Total SROIs = 111 [90 SROIs corresponding to 90 lesions + 21 SROIs corresponding to 21 NOR images] **c.** Total typ. IROIs= 142 [typ. lesion IROIs (72) + typ. NOR IROIs (70)] **d.** Total atyp. IROIs= 148 **e.** Total SHCC lesion IROIs= 19 **f.** Total LHCC lesion IROIs=71 **g.** Total IROIs= **c+d+e+f**=142+148+19+71=380.

2.4 Proposed Computer-aided Diagnostic System

The CAD system consisted of feature extraction, feature dimensionality reduction and classification module. The block diagram of the proposed CAD system is depicted in Fig. 3.

2.4.1 Feature Extraction Module

In the present work initially, a wide variety of visual and non visual echotexture features are extracted by using statistical, spectral and spatial filtering based feature extraction methods as depicted below:

2.4.1.1 FOS Features

Six FOS features, i.e. average gray level, standard deviation$_{FOS}$, smoothness, third moment, uniformity and entropy are considered for analysis.

2.4.1.2 GLCM Features

Thirteen GLCM mean features, i.e. angular second moment, contrast, correlation, variance, inverse difference moment, sum average, sum variance, sum entropy, entropy$_{GLCM}$, difference variance, difference entropy, information measures of correlation-1, and information measures of correlation-2 are considered for analysis[17,18].

2.4.1.3 GLRLM Features

Eleven GLRLM features, i.e. short run emphasis, long run emphasis, low gray level run emphasis, high gray level run emphasis, short run low gray level emphasis, short run high gray level emphasis, long run low gray level emphasis, long run high gray level emphasis, gray level non uniformity, run length non uniformity and run percentage are considered for analysis[21-23].

2.4.1.4 FPS Features

Two spectral features, i.e. radial sum and angular sum of the discrete Fourier transform are considered for analysis[24].

2.4.1.5 GWT Features

Two statistical features, i.e. mean and standard deviation of Gabor output images obtained by using a set of Gabor wavelets at 3 scales and 7 orientations results in a set of (2x3x7) forty two features considered for analysis[5].

2.4.1.6 Laws' Features

Special 1-D filters of length 5, i.e. L5= [1, 4, 6, 4, 1], E5= [-1, -2, 0, 2, 1], S5= [-1, 0, 2, 0, -1], W5=[−1, 2, 0, −2, 1] and R5 = [1, −4, 6, −4, 1] are used to compute Laws' features. Twenty five 2-D filters called Laws' masks are derived by outer vector product of these 1-D kernels with themselves or with each other. The texture images (TIs) are obtained by convolving the ROI with these 2D Laws' masks. A 15x15 square window is applied to TIs to obtain texture energy images (TEIs). Out of 25 2-D filters, 10 are identical to other 10, if rotated by 90°. TEIs which result from such pair of identical filters are combined to obtain a rotation invariant (TR) image, for example, TEI$_{E5L5}$ and TEI $_{L5E5}$ is combined to obtain TR image. Thus, total 15 TR images are obtained for each ROI. Statistics (mean and standard deviation) derived from these TR images (15 TR images x 2 statistical parameters) result in thirty Laws' texture features considered for analysis[25,26].

In all, 104 texture features (6 FOS, 13 GLCM, 11 GLRLM, 2 FPS, 42 Gabor and 30 Laws' features) are computed from each IROI and same 104 texture ratio features are computed from each IROI and a corresponding SROI.

2.4.2 Feature Dimensionality Reduction Module

In the present work, PCA is used for feature space dimensionality reduction and the optimal number of PCs to be retained for classification task is determined empirically by repeated experiments carried out by stepping through first 15 PCs to build the SVM model.

The four steps involved in PCA algorithm[27] are, (1) Standardize the values of all the features in the data set to zero mean and unit variance, (2) Obtain the covariance matrix from the training data set, (3) Obtain the Eigen values and Eigen vectors from the covariance matrix (Eigen vectors are the directions of the principal components), and (4) Obtain the projection of data points in the testing data set in the direction of principal components (PCs) of the training dataset.

2.4.3 Classification Module

2.4.3.1 SVM Classifier

SVM classifier is implemented using LibSVM library[28]. In kernel based classifiers such as SVM, kernel functions are used for nonlinear mapping of training data from input space to higher dimensional feature space. The performance of Gaussian radial basis function kernel is investigated for the present classification task. A crucial step for obtaining good generalization performance is correct choice of the regularization parameter C and kernel parameter γ. The optimal values for C and γ are obtained by extensive search, carried out in the parameter space for the values of C ϵ $\{2^{-4}, 2^{-3}\ldots 2^{15}\}$, γ ϵ $\{2^{-12}, 2^{-11}\ldots 2^{4}\}$ using 10 fold cross validation on training data.

3. RESULTS

The classification of FLLs from B-Mode US is experimented with the texture feature vectors (TFVs) depicted in Table 3.

Table 3. Description of TFVs

Texture Feature Vectors (TFVs)	(*l*)
TFV1: TFV consisting of 104 texture features (6 FOS, 13 GLCM, 11 GLRLM, 2 FPS, 42 Gabor and 30 Laws' features)	104
TFV2: TFV consisting of 104 texture ratio features (6 FOS, 13 GLCM, 11 GLRLM, 2 FPS, 42 Gabor and 30 Laws' features)	104
TFV3: Combined TFV consisting of 104 texture features (TFV1) and 104 texture ratio features (TFV2)	208
TFV4: Optimal reduced TFV consisting of first 5 PCs obtained by subjecting combined TFV (TFV3) to PCA	5

Note: *l*: Length of TFV.

PCA is applied to the feature set consisting of instances of TFV3, i.e. combined TFV consisting of 104 texture features (TFV1) and 104 texture ratio features (TFV2). It is observed that maximum classification accuracy is obtained by using the first five PCs which accounted for 90.17% of variance in the data. Thus, optimal reduced TFV, i.e. TFV4 is considered for analysis. The classification performance of feature sets consisting of instances of TFV1, TFV2, TFV3 and TFV4 is tested by using SVM classifier. The results obtained are reported in Table 4.

From Table 4, it can be observed that texture features computed from IROIs (TFV1) yields the lowest overall classification accuracy (OCA) of 66.1% with the individual class accuracy (ICA) values of 80%, 88%, 60%, 30% and 77.7% for NOR, Cyst, HEM, HCC and MET cases, respectively.

Table 4. Classification performance obtained with SVM classifier for different TFVs

TFV (*l*)		CM					OCA (%)	ICA (%)
		NOR	CYST	HEM	HCC	MET		
TFV1 (104)	NOR	32	0	0	5	3	66.1	80
	CYST	0	22	0	1	2		88
	HEM	0	0	18	8	4		60
	HCC	2	2	9	12	15		30
	MET	0	0	5	5	35		77.7
TFV2 (104)	NOR	32	0	0	4	4	70	80
	CYST	0	21	0	0	4		84
	HEM	0	0	22	0	8		73.3
	HCC	3	1	1	24	11		60
	MET	2	1	11	4	27		60
TFV3 (104)	NOR	35	0	0	2	3	86.1	87.5
	CYST	0	23	0	1	1		92
	HEM	0	0	26	1	3		86.6
	HCC	4	0	0	35	1		87.5
	MET	0	0	9	0	36		80
TFV4 (5)	NOR	34	0	1	3	2	87.2	85
	CYST	0	24	0	0	1		96
	HEM	0	0	27	0	3		90
	HCC	1	4	0	35	0		87.5
	MET	1	0	5	2	37		82.2

Note: TFV: Texture feature vector, *l*: Length of TFV, CM: Confusion matrix, OCA: Overall classification accuracy, ICA: Individual class accuracy.

Textural changes in the region surrounding the lesion are observed due to tumor aggression. Therefore the radiologists visualize the texture information of areas inside and outside of the lesion for differential diagnosis of FLLs. As TFV1 contains texture information of regions inside the lesions only, it yields lower classification accuracy.

The texture ratio features (TFV2) yield an OCA of 70% with the ICA values of 80%, 84%, 73.3%, 60% and 60% for NOR, Cyst, HEM, HCC and MET cases, respectively. It can be concluded that texture ratio features have more discrimination ability than texture features for classification of FLLs from B-mode US. It can be observed that texture features and texture ratio features combined (TFV3) yields the classification accuracy of 86.1% with the ICA values of 87.5%, 92%, 86.6%, 87.5% and 80% for NOR, Cyst, HEM, HCC and MET cases, respectively. In comparison to ICA values obtained by using texture ratio features only, the ICA values obtained by using texture features and texture ratio features collectively, show improvement by 7.5%, 8%, 13.3%, 27.5% and 20% for NOR, Cyst, HEM, HCC and MET classes, respectively. It can be concluded that both texture features and texture ratio features contribute towards effective classification of FLLs from B-mode US. The results are in agreement with the premise that radiologists visualize the textural characteristics of regions inside and outside the lesions for differential diagnosis between FLLs. Finally, it is observed that optimal reduced TFV (TFV4) with first five PCs yields the maximum overall classification accuracy of 87.2% with the ICA values of 85%, 96%, 90%, 87.5% and 82.2% for NOR, Cyst, HEM, HCC and MET cases respectively.

3.1 Misclassification Analysis

Analysis of 23 misclassified cases out of total 180 cases in the test data set as predicted by proposed PCA-SVM based classier is reported in Table 5.

From Table 5, it can be observed that the CAD system yields classification accuracy of 88.8%, 87%, 87.5% and 86.8% for SHCC, LHCC, typical and atypical cases, respectively.

Table 5. Misclassification Analysis of 23 out of 180 cases of test data set

Class	Misclassified Cases	Remarks
NOR: 40 (40 typ.)	6 (typ.)	typ. correctly classified: 34, ICA: 85
Cyst: 25 (13 typ., 12 atyp.)	1 (atyp.)	typ. correctly classified: 13, ICA: 100 atyp. correctly classified: 11, ICA: 91.6
HEM: 30 (6 typ., 24 atyp.)	3 (1 typ., 2 atyp.)	typ. correctly classified: 5, ICA: 83.3 atyp. correctly classified: 22, ICA: 91.6
HCC: 40 (9 SHCC, 31 LHCC.)	5 (1 SHCC, 4 LHCC)	SHCC correctly classified: 8, ICA: 88.8 LHCC correctly classified:27, ICA: 87.0
MET: 45 (5 typ., 40 atyp.)	8 (1 typ., 7 atyp.)	typ. correctly classified: 4, ICA: 80 atyp. correctly classified: 33, ICA: 82.5
typ:64(40NOR,13cyst,6HEM,5MET) atyp.:76 (12 cyst,24HEM,40MET)	8 (6NOR,1HEM,1MET) 10 (1cyst,2HEM,7MET)	typ. correctly classified: 56, ICA:87.5 atyp. correctly classified: 66, ICA:86.8

Note: typ.: Typical, Atyp. : Atypical, ICA: Individual class accuracy. ICA values are expressed in percentage.

4. CONCLUSION

It is observed that the performance of the CAD system, improves by including the texture ratio features for designing the classifier. It can be concluded that the texture of the background liver on which lesion has evolved significantly contribute for classification of FLLs. Further, by application of PCA to feature set consisting of 208 texture features (104 texture features and 104 texture ratio features), the information required for classification of FLLs is squeezed in first five PCs. Thus, it can be concluded that only five PCs are significant to account for textural variations exhibited by FLLs on B-mode US. The proposed CAD system obtained the OCA of 87.2% with ICA values of 88.8%, 87%, 87.5%, and 86.8% for SHCC, LHCC, typical and atypical cases, respectively. In practice, while scanning the liver the radiologist can always freeze the image of lesion so as to include the area surrounding it and then crop the IROIs and SROI which are used as input to the proposed CAD system. The promising results indicate the usefulness of the CAD system to assist radiologists in differential diagnosis between FLLs and thereby facilitate in better management of focal liver diseases.

ACKNOWLEDGMENTS

Author would like to acknowledge Ministry of Human Resource Development (MHRD), India for financial support. The authors wish to acknowledge the Department of Electrical Engineering, Indian Institute of Technology, Roorkee, India and Department of Radiodiagnosis and Imaging, Postgraduate Institute of Medical Education and Research, Chandigarh, India for their constant patronage and support in carrying out this research work. The authors would like to thank the anonymous reviewers for their substantive and informed review, which led to significant improvements in the manuscript.

REFERENCES

1. Harding, J & Callaway, M. Ultrasound of focal liver lesions. *Rad Magzine*, 2010, **36**(424), 33-34.
2. Jeffery, R.B & Ralls, P.W. Sonography of Abdomen. New York Raven Press, 1995.
3. Baert, A.L & Sartor, K. Focal Liver Lesions -Detection, Characterization, Ablation, Springer Berlin Heidelberg New York, 2005.
4. Virmani, J.; Kumar, V.; Kalra, N & Khandelwal, N. Characterization of primary and secondary malignant liver lesions from B-mode ultrasound. *J. Digit Imaging,* 2013, DOI: 10.1007/s10278-013-9578-7.
5. Mittal, D.; Kumar, V; Saxena, S.C.; Khandelwal, N & Kalra, N. Neural network based focal liver lesion diagnosis using ultrasound Images. *Int. J. Comput. Med. Imaging Graph.*, 2011, **35**(4), 315-323.
6. Virmani, J.; Kumar, V.; Kalra, N. & Khandelwal, N., A comparative study of computer-aided classification systems for focal hepatic lesions from B-mode ultrasound, *J. Med. Eng. Technol.*, 2013, **37**(4), 292-306.
7. Virmani, J.; Kumar, V.; Kalra, N & Khandelwal, N. SVM-Based characterization of liver ultrasound images using wavelet packet texture descriptors. *J. Digit Imaging,* 2013, **26**(3), 530-543.
8. Virmani, J.; Kumar, V.; Kalra, N & Khandelwal, N. Prediction of liver cirrhosis based on multiresolution texture descriptors from B-mode ultrasound. *Int. J. Convergence Comput.,* 2013, **1**(1), 19-37.
9. Virmani, J.; Kumar, V.; Kalra, N & Khandelwal, N. SVM-based characterization of liver cirrhosis by singular value decomposition of GLCM matrix. *Int. J. Artif. Intelligence Soft Comput.,* 2013, **4**(3), 276-296.
10. Virmani, J.; Kumar, V.; Kalra, N & Khandelwal, N. A rapid approach for prediction of liver cirrhosis based on first order statistics. *In* Proceedings of IEEE International Conference on Multimedia, Signal Processing and Communication Technologies, IMPACT-2011, Aligarh India, 212-215.
11. Scheible, W.; Gossink, B.B & Leopold, G. Gray scale echo graphic patterns of hepatic metastatic disease. *Am. J. Roentgenol*, 1997, 129, 983-987.
12. Sujana, S.; Swarnamani, S & Suresh, S. Application of artificial neural networks for the classification of liver lesions by image texture parameters. *Ultrasound Med. Bio.*,1996, **22**(9), 1177-1181.
13. Yoshida, H,; Casalino, D.D.; Keserci, B.; Coskun, A.; Ozturk, O &, Savranlar, A. Wavelet packet based texture analysis for differentiation between benign and malignant

liver tumors in ultrasound images. 2003, *Phys. Med. Bio.*, 2003, **48**(22), 3735–3753.
14. Poonguzhali, S.; Deepalakshmi & Ravindran, G. Optimal feature selection and automatic classification of abnormal masses in ultrasound liver images. *In* Proceedings of IEEE International Conference on Signal Processing, Communications and Networking, ICSCN'07, 2007, 503-506.
15. Jeon, J.H.; Cohi, J.Y.; Lee, S. & Ro, Y.M. Multiple ROI selection based focal liver lesion classification in ultrasound images. *Expert Systems with Applications*, 2013, **40**(2), 450-457.
16. Kadah, Y.M.; Farag, A.A.; Zurada, J.M.; Badawi, A.M & Youssef, A.M. Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images. *IEEE Trans. Med. Imaging,* 1996, **15**(4), 466-478.
17. Badawi, A.M.; Derbala, A.S & Youssef A.M. Fuzzy logic algorithm for quantitative tissue characterization of diffuse liver diseases from ultrasound images. *Int. J. Med. Inf.*, 1999, **55**(2), 135-147.
18. Fukunaga, K. Introduction to Statistical Pattern Recognition. Academic Press, New York, 1990.
19. Haralick, R.; Shanmugam, K. & Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man. Cyb.*, 1973, **SMC-3**(6), 610-121.
20. Virmani, J.; Kumar, V.; Kalra, N & Khandelwal, N. Prediction of cirrhosis based on singular value decomposition of gray level co-occurence matrix and a neural network classifier. *In* Proceedings of IEEE International Conference on Developments in E-systems Engineering, DeSe-2011, 146-151.
21. Galloway, R.M. Texture analysis using gray level run lengths. *Comput. Gr. Image Process.*, 1975, **4**(2), 172-179.
22. Chu, A.; Sehgal, C.M. & Greenleaf, J.F. Use of Gray value distribution of run lengths for texture analysis. *Pattern Recognition Lett,* 1990, **11**(6), 415-420.
23. Dasarathy, B.V & Holder, E.B. Image characterizations based on joint gray level-run length distributions. *Pattern Recognition Lett.,* 1991, **12**(8), 497-502.
24. Weszka, J.S.; Dyer, C.R & Rosenfeld, A. A comparative study of texture measures for terrain classification. *IEEE Trans. Syst. Man. Cyb.,*1976,**SMC-6**(4),269–285.
25. Laws, K.I. Rapid Texture Identification. *In* SPIE Proceedings of the Seminar on Image Processing for Missile Guidance, 1980, **238**, 376-380.
26. Virmani, J.; Kumar, V.; Kalra, N. & Khandelwal, N. Prediction of cirrhosis from liver ultrasound B-Mode images based on Laws' masks analysis. *In* Proceedings of IEEE International Conference on Image Information Processing, ICIIP-2011, 1-5, 2011.
27. Kadir, A.; Nugroho, L.E.; Susanto, A. & Santosa, P.I. Performance improvement of leaf identification system using principal component analysis. *Int. J. Adv. Sc.i Technol.*, 2012, **44**, 113–124.
28. Chang, C.C & Lin, C.J. LIBSVM, A library of support vector machines, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm. [Accessed on 15 December 2012]

CONTRIBUTORS

Mr Jitendra Virmani did his BTech (Hons.) in Instrumentation Engineering from Sant Longowal Institute of Engineering and Technology in 1999 and MTech in Measurement and Instrumentation from IIT-Roorkee in 2006. Currently pursuing his PhD at Biomedical Instrumentation Laboratory, Electrical Engineering Department, IIT-Roorkee as a full time Research Scholar under MHRD Assistantship. He is Life member of Institute of Engineers (IEI), India. His research interests include: Application of machine learning and soft computing techniques for analysis of medical images.

Dr Vinod Kumar obtained BSc (Electrical Engineering) with from Punjab University in 1973, ME (Measurement & Instrumentation) and PhD from University of Roorkee in 1975 and 1984, respectively. Currently working as a Professor, Department of Electrical Engineering, IIT Roorkee since 1995. He has guided 21 Doctoral and more than 80 Master's Thesis and has more than 150 research publications in internationally reputed journals and conference proceedings. He is a senior member of IEEE, USA. His research interests include : Medical image processing, digital signal processing and telemedicine.

Dr Naveen Kalra did MBBS and MD from Maulana Azad Medical College, New Delhi in 1992 and 1998 respectively. Currently working as a Additional Professor at Department of Radio diagnosis and Imaging, Post Graduate Institute of Medical Education and Research, Chandigarh. He has guided 30 MD Thesis and published more than 85 research publications and contributed 15 chapters in books.

Dr Niranjan Khandelwal did his MBBS from Calcutta University in 1980 and MD (Radiodiagnosis) from Postgraduate Institute of Medical Education & Research, Chandigarh in 1984. He is Professor & Head, Department of Radiodiagnosis and Imaging, PGIMER- Chandigarh. He has published more than 140 research papers and contributed 15 chapters in books.