A PROJECT REPORT

ON

CAD system for Benign and Malignant Focal Liver Lesions using GLCM Descriptors and SVM Classifier

SUBMITTED IN PARTIAL FULFILLMENT OF THE CREDITS REQUIRED FOR THE SUBJECT

DIGITAL IMAGE PROCESSING

SUBMITTED BY (GROUP-20)

ANSHUL VIRMANI (ROLL NO. 2210110172)

KUSH SAHNI (ROLL NO. 2210110371)

AINESH DAS (ROLL NO. 2210110130)

AKSHAT (ROLL NO. 2210110135)

SUBMITTED TO

DR. SAURABH JANARDAN SHIGWAN

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

SHIV NADAR

INSTITUTION OF EMINENCE DEEMED TO BE UNIVERSITY

DELHI NCR

Introduction

Why Liver Tissue: Liver is the most important tissue which performs over 500 functions for our human body. It stores vitamins and carbohydrates. It produces bile (a digestive fluid). It cleans poisons from the blood. It has ability to repair itself. It is vital organ.

Liver Anatomy: Liver is the largest gland. It weighs from 1.1 to 1.6 Kgs.

Liver Appearance: It has triangular shape and normal liver is pinkish brown in color.

Other Characteristics: It is soft in nature and has four lobes.

Location: It is located in right upper quadrant of abdominal cavity just below the diaphragm to the right of the stomach and overlying the gallbladder.

Vascular Structure: It has hepatic arteries and portal veins

Liver Diseases:

(a) Diffused Liver Diseases: affect the entire liver for example, chronic active hepatitis, Fatty Liver and Cirrhosis.

(b) Focal Liver Diseases: affect a localized region of the liver (a) Benign Focal liver lesions (FLLs)[6], (b) Malignant FLLs

Common Benign FLLs: Hemangioma (HEM), Focal Nodular Hyperplasia (FNH)

Common Malignant FLLs: Hepatocellular carcinoma (HCC)

Why Liver Ultrasound Images? Why Not Liver CT or MRI?

Ultrasound is considered as first examination for characterization of FLLs because (a) Non-ionizing, non-invasive, inexpensive nature (b) Real time imaging capabilities

In comparison contrast enhanced US (CEUS), contrast enhanced computed tomography and magnetic resonance imaging (MRI) offer higher sensitivity for characterization of FLLs at the same time these modalities are expensive and pose greater operational inconvenience and not widely available

Therefore an efficient CAD system for characterization of FLLs using B-Mode US is highly desired.

Sonographic Appearances of HEM

Typical HEM

Typical HEMs are well circumscribed uniformly hyperechoic lesions. In 70% of cases the sonographic appearance is typical**.** The sample images of HEM are shown in Fig 1.

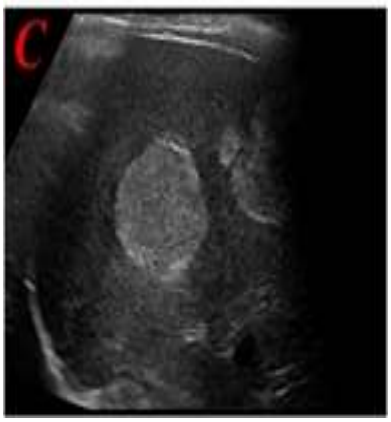
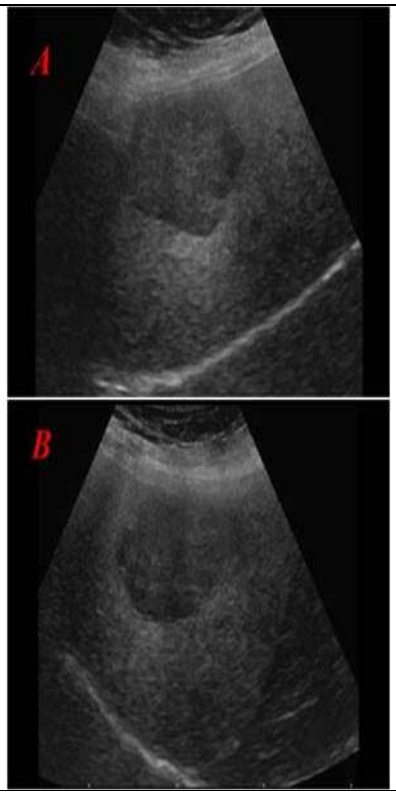


Fig 1. Sample images of HEM class from ultrasoundcases.info [2] (a), (b) contains solitary lesions and (c) contains 2 lesions.

Fig2. Sample images of FNH from ultrasoundcases.info. [2] (a), (b) and (c) contains Solitary Lesions.

Sonographic Appearances FNH

The sample images of FNH are shown in Fig 2.



Sonographic Appearances HCC

Small HCC (<=2cms) or Large HCC

The sonographic appearances of small HCC (SHCC) <2cms vary from hypoechoic to hyper echoic. The sample images of small HCC and Large HCC are shown in Fig 3 and Fig 4

Fig 3. Sample images of small HCC from ultrasoundcases.info [2] (a), (b) and (c) contains Solitary Lesions.

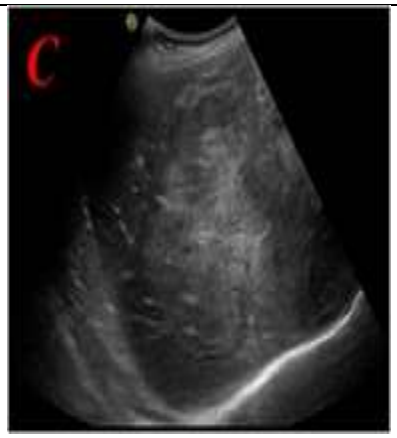
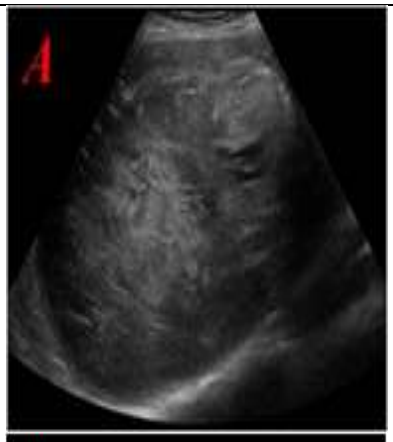
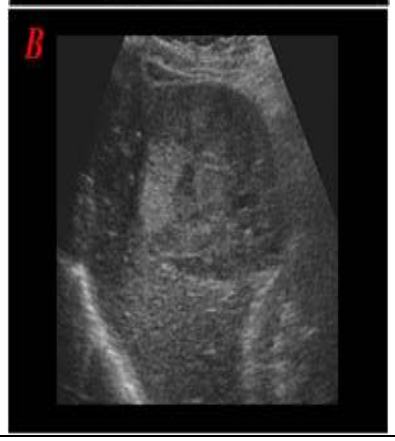


Fig 4. Sample images of large HCC from ultrasoundcases.info [2] (a), (b) and (c) contains Solitary Lesions.

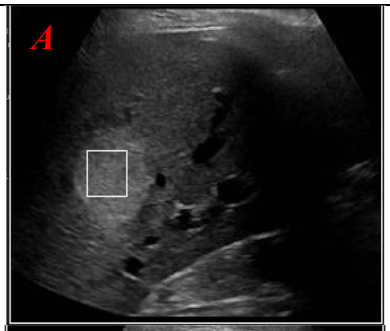
Large HCC appear frequently with mixed echogenicity. Heterogeneous echotexture represents complex and chaotic structure exhibited by LHCC due to coexistence of areas of necrosis, fibrosis and active growth areas.

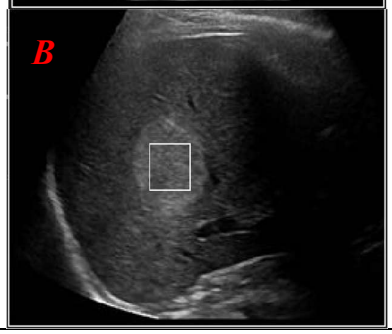


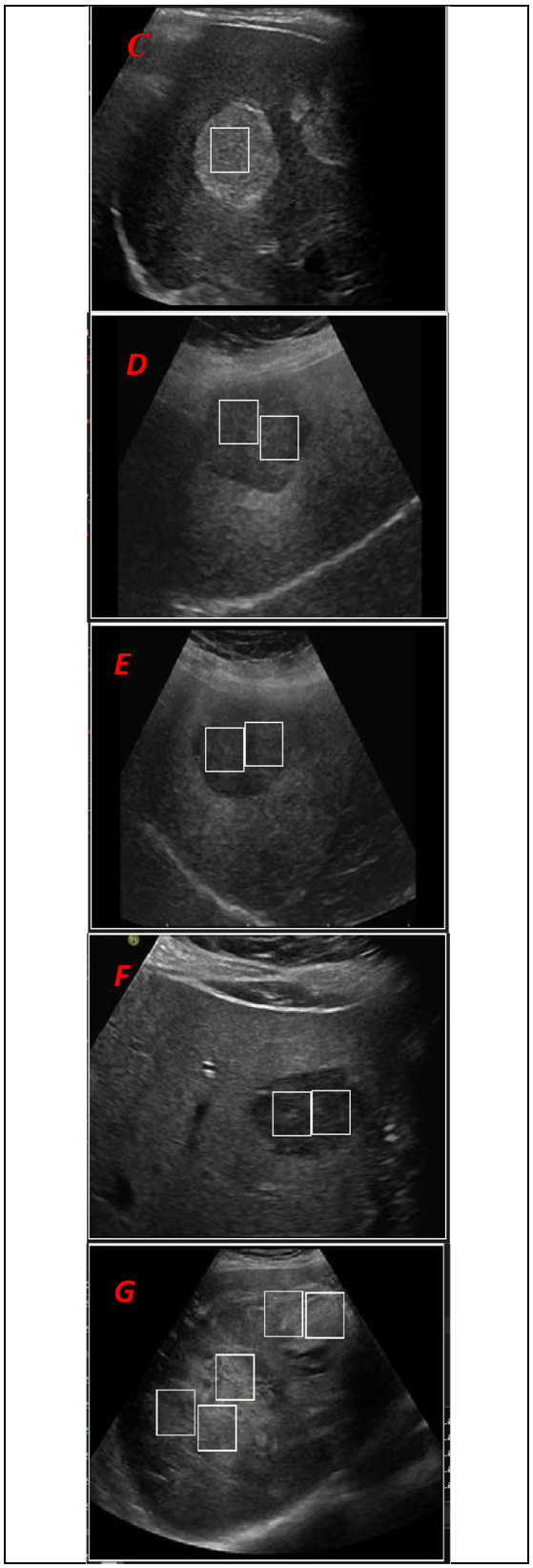


Images with ROI marked

The sample Images with ROIs marked are shown in Fig 5.







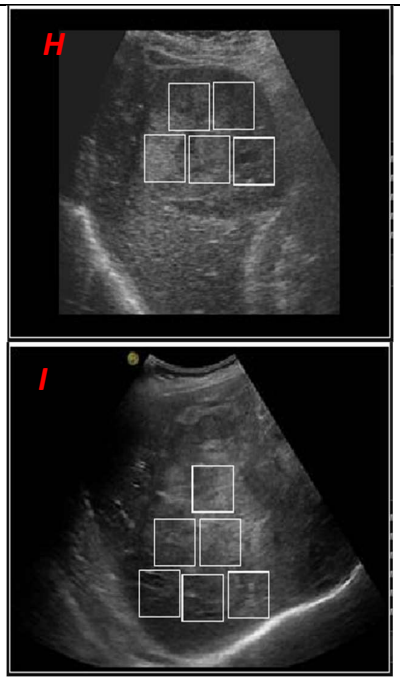


Fig 5. Sample images with Marked ROIs (a),(b) and (c) contains marked ROIs of HEM (d),(e) and (f) contains marked ROIs of FNH and (g),(h) and (i) contains marked ROIs of HCC

Why CAD system using Ultrasound Images?

Irrespective of the known advantages of using conventional gray scale ultrasound imaging modality , there are certain limitations of using it for differential diagnosis between FLLs using B-Mode US images, these are (a) Limited sensitivity for detection of small FLLs (<2cm) developed on cirrhotic liver which is already nodular and coarse-textured (b) Overlapping sonographic appearances of HEM , FNH and HCC lesions and (c) Limited sensitivity for detection of isoechoic lesions with the slim lesion-to-liver contrast.

Therefore it is highly desired to minimize these limitations and design an efficient CAD system for characterization of FLLs using B-Mode US images.

Acquisition of Image database

Source: https://www.ultrasoundcases.info/cases/abdomen-and-retroperitoneum/liver/ [2]

The number of images considered for Developing CAD System is shown in Fig 6.

In order to develop an efficient and robust classifier design, it is necessary to train the classifier using a comprehensive image database with representative images from each sub class. Thus comprehensive image database (ultrasoundcases.info) with representative cases from each class including, (1) Haemangioma, (2) Focal Nodular Hyperplasia and (3) Hepatocellular Carcinoma was considered for analysis.

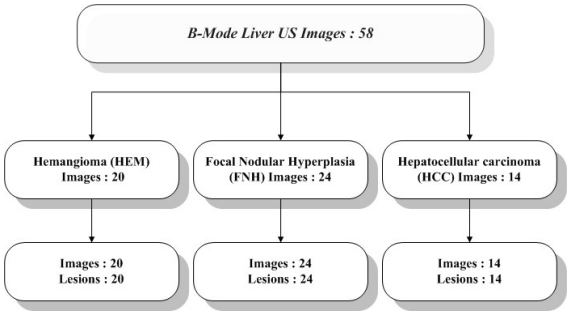


Fig 6. No of Images considered for developing a CAD system

Data Augmentation

The Sample images obtained in Testing Data were then augmented using Geometric transformation (Rotation, Flipping). Each individual image is rotated at certain angles.

In case of HEM, image was rotated on the angles 25, 50, 75, 100. The rotated images were then flipped horizontally. Both the rotated and horizontally flipped images were then flipped vertically.

In case of FNH, image was rotated on the angles 25, 50, 75, 100. The rotated images were then flipped horizontally. Both the rotated and horizontally flipped images were then flipped vertically.

In case of HCC, image was rotated on the angles 7,14 ,21 ,28 ,35 ,42 ,49 ,56 ,63 ,70 ,77 ,84 ,91 ,98 ,105 ,112 ,119. The rotated images were then flipped horizontally. Both the rotated and horizontally flipped images were then flipped vertically. The sample augmented images are shown in Fig 7.

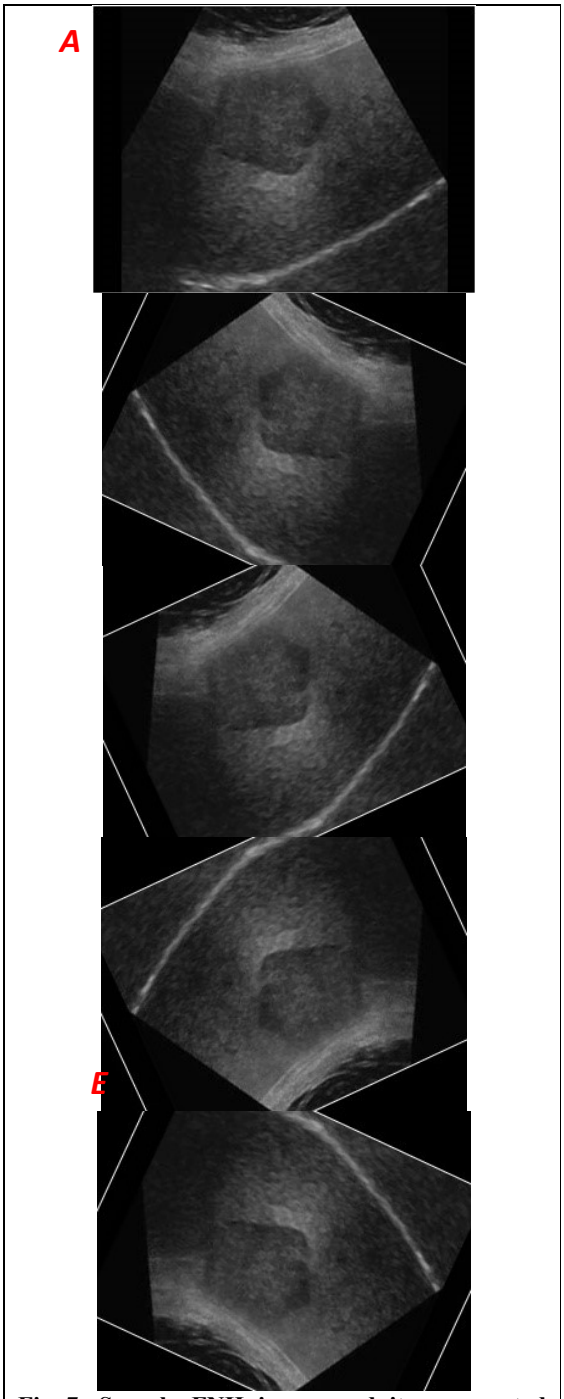


Fig 7. Sample FNH image and its augmented images (a) Original FNH image (b) Original image rotated by 25º. (c) Original image rotated by 25deg then flipped horizontally(d) Original image rotated by 25deg then flipped horizontally , then flipped vertically (e) Original image rotated by 25deg then flipped vertically

The Data set description and Training and Testing Data Spilt is shown in Fig 8.

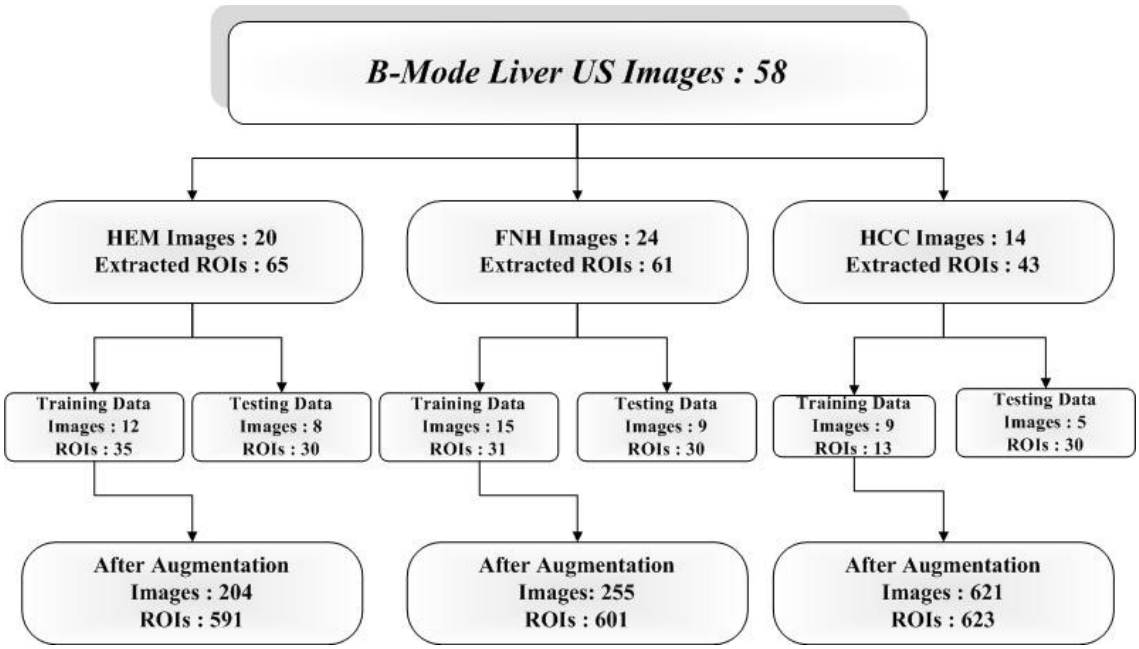


Fig 8. Data set Description

Training and Testing Data Split:

To avoid any biasing, the ROIs from one image set are used for training and the ROIs from the other image set are used for testing.

Extraction of ROIs

The ROIs were cropped by using Python software The image was first loaded and maximum non-overlapping ROIs of size 32 x 32 are extracted from well within the boundary of each lesion. Necrotic areas within the lesion are avoided while extracting ROIs.

CAD system for classification of HEM, FNH and HCC liver

Methodology:

The methodology adopted is shown in Fig 9.

GLCM-M and GLCM Range Features and SVM Classifier

Feature Extraction:

Texture features extraction methods can be classified as shown in Fig 10.

In the present study the SOS GLCM features [1,3,4] are considered for analysis as shown in Table 1. The methodology adopted and its work flow is shown in Fig. 9.

Table 1. F1 to F13: GLCM Features

Feature ID	Feature Name
1	$\mathbf{F_1}$: Angular Second Moment (ASM)
2	$\mathbf{F_2}$: Contrast
3	$\mathbf{F_3}$: Correlation
4	$\mathbf{F_4}$: Sum of Squares-Variance
5	$\mathbf{F_5}$: Inverse Difference Moment (IDM)
6	$\mathbf{F_6}$: Sum Average
7	$\mathbf{F_7}$: Sum Variance
8	$\mathbf{F_8}$: Sum Entropy
9	$\mathbf{F_9}$: Entropy
10	$\mathbf{F_{10}}$: Difference Variance
11	$\mathbf{F_{11}}$: Difference Entropy
12	$\mathbf{F_{12}}$: Information Measures of Correlation-1
13	$\mathbf{F_{13}}$: Information Measures of Correlation-2

Formulas to compute GLCM (mean) and GLCM (range) descriptors:

$$\mathbf{GLCM(mean)}_{\mathbf{IDM\ (d=1)}} = \left(\frac{\mathbf{A+B+C+D}}{\mathbf{4}}\right)$$

$$\mathbf{GLCM(range)}_{\mathbf{IDM\ (d=1)}} = \max(\mathbf{A,B,C,D}) - \min(\mathbf{A,B,C,D})$$

$$\mathbf{A = GLCM}_{\mathbf{IDM(\theta=0^o,d=1)}}, \mathbf{B = GLCM}_{\mathbf{IDM(\theta=45^o,d=1)}},$$

$$\mathbf{C = GLCM}_{\mathbf{IDM(\theta=90^o,d=1)}}, \mathbf{D = GLCM}_{\mathbf{IDM(\theta=135^o,d=1)}}$$

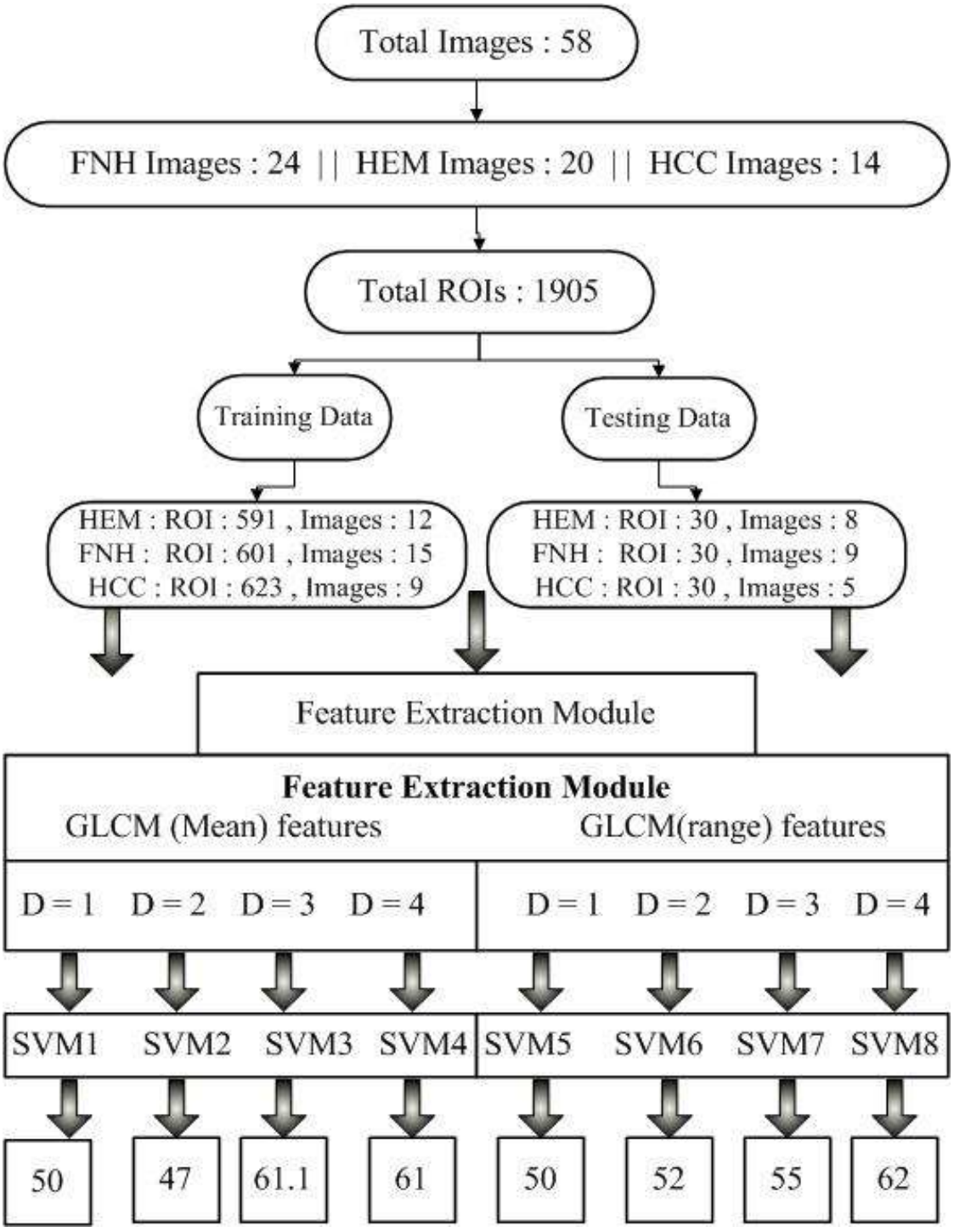


Fig 9. Methodology Adopted to develop the CAD system.

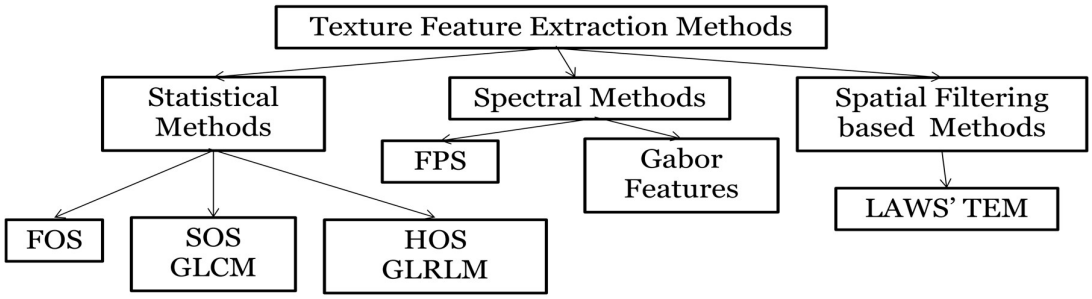


Fig 10. Broad classification of texture features

SVM Classifier:

SVM – Support Vector Machines

It is one of the most popular machine learning algorithms. It falls under the category of supervised learning. SVM can be used for both "classification" as well as "regression" problem. However it is primarily used for the classification task in Machine learning.

It is based on the Logistic regression but with some added functionalities to improve performance. The logistic regression algorithm gives a best fit line that separates the data points; similarly SVM provides a best fit line that separates the data points. This best fit line in SVM is called "hyperplane". But additionally along with the hyperplane SVM also gives Marginal planes that separate the data points.

The main aim of this algorithm is to provide the Hyperplane and the marginal planes with the maximum margin. There are two types of marginal planes, namely "hard marginal plane" and "soft marginal plane". The marginal plane that completely differentiates the data points is called hard marginal plane, while the marginal plane that doesn't completely differentiates the data points and few are left to be separated is called soft marginal plane.

SVM chooses the extreme points that help in creating the hyper plane. These extreme cases are called "support vectors". There can be many lines that can separate the support vectors but the line with maximum marginal distance from the support vectors is the Hyper plane as shown in Fig 11.

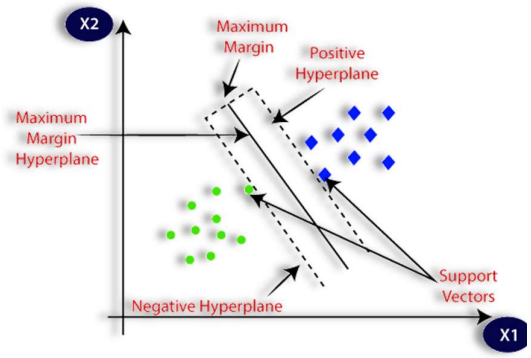


Fig 11. Support Vector Machine Classifier.

In real world scenario we may not always have linearly separable data. In order to separate the non linear support vectors, we use the kernel trick to achieve accuracy separating the support vectors. In this the lower dimensional feature space is converted into higher dimensional feature space in order to easily separate the support vectors. There are various kernels used like RBF (radial bias function), Polynomial and Sigmoid etc. The sample image of data points before applying RBF kernel is shown in Fig 12 and after applying RBF kernel is shown in Fig 13.

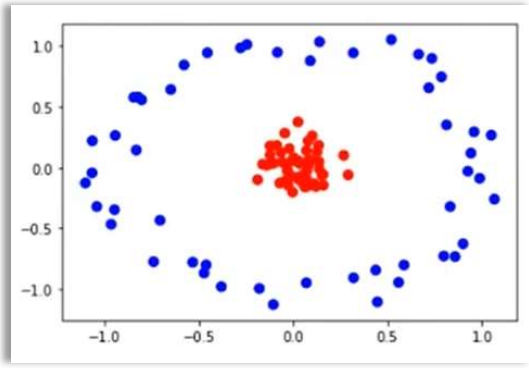


Fig 12. Data points before applying "RBF" kernel

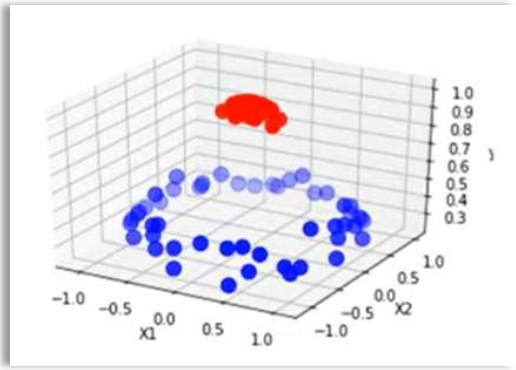


Fig 13. Data points before applying "RBF" kernel

Hyper parameters for tuning the SVM classifier- C and Gamma (γ):

C - It is a hypermeter in SVM to control error. It tells about how many points can be avoided in misclassification of data points. It is used to overcome the condition of over fitting. Gamma – Gamma (γ) is used when we use the Gaussian RBF kernel. if you use linear or polynomial kernel then you do not need gamma only you need C hypermeter.

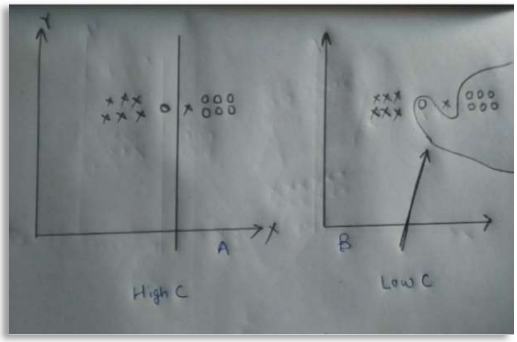


Fig 14. The effect of hyper parameter tuning [C]

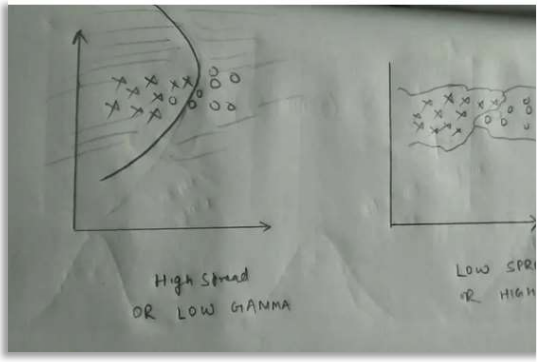


Fig 15. Hyperparmeter tuning effect of Gamma

Gamma is a Hyper parameter which we have to set before training model. Gamma decides that how much curvature we want in a decision boundary. Gamma high means more curvature. The effect of hyper parameter tuning is shown in Fig 15.

The optimal values for C and γ for design of SVM model are obtained by extensive search, carried out in the parameter space for the values of C ϵ $\{2^{-4}, 2^{-3}\ldots 2^{15}\}$, $\gamma \epsilon \{2^{-12}, 2^{-11}\ldots 2^{4}\}$ using 5 fold cross validation on training data. The LibSVM library was installed for implementing SVM classifier [5].

Experiments, Results and Discussion

Experiment 1: To observe the performance of GLCM-M features for differential diagnosis between HEM, FNH and HCC cases using SVM Classifier (Table 2)

Experiment 2: To observe the performance of GLCM-R features for differential diagnosis between HEM, FNH and HCC cases using SVM Classifier (Table 3)

Table 2. Classification Performance of SVM with GLCM-M features

Features	Confusion Matrix				Accuracy (%)	ICA(HEM) (%)	ICA(FNH) (%)	ICA (HCC) (%)
GLCM-M (d=1)		**HEM**	**FNH**	**HCC**	**50**	**70**	**56**	**23**
	HEM	**21**	**4**	**5**				
	FNH	**13**	**17**	**0**				
	HCC	**11**	**12**	**7**				
GLCM-M (d=2)		**HEM**	**FNH**	**HCC**	**47**	**60**	**56**	**26**
	HEM	**18**	**4**	**8**				

	FNH	**11**	**17**	**2**				
	HCC	**8**	**14**	**8**				
GLCM-M (d=3)		**HEM**	**FNH**	**HCC**	**61.1**	**73.3**	**80**	**30**
	HEM	**22**	**3**	**5**				
	FNH	**1**	**24**	**5**				
	HCC	**8**	**13**	**9**				
GLCM-M (d=4)		**HEM**	**FNH**	**HCC**	**61**	**73**	**80**	**30**
	HEM	**22**	**3**	**5**				
	FNH	**1**	**24**	**5**				
	HCC	**8**	**13**	**9**				

Table 3: Classification Performance of SVM with GLCM-R features

Features	**Confusion Matrix**				**Accuracy (%)**	**ICA(HEM) (%)**	**ICA(FNH) (%)**	**ICA (HCC) (%)**
GLCM-R (d=1)		**HEM**	**FNH**	**HCC**	**50**	**60**	**60**	**23**
	HEM	**19**	**5**	**6**				
	FNH	**7**	**19**	**4**				
	HCC	**8**	**15**	**7**				
GLCM-R (d=2)		**HEM**	**FNH**	**HCC**	**52**	**73**	**70**	**13**
	HEM	**22**	**3**	**5**				
	FNH	**6**	**21**	**3**				
	HCC	**12**	**14**	**4**				
GLCM-R (d=3)		**HEM**	**FNH**	**HCC**	**55**	**70**	**63**	**33**
	HEM	**21**	**3**	**6**				
	FNH	**6**	**19**	**5**				
	HCC	**12**	**8**	**10**				
GLCM-R (d=4)		**HEM**	**FNH**	**HCC**	**62**	**70**	**66**	**50**
	HEM	**21**	**3**	**6**				
	FNH	**4**	**20**	**6**				
	HCC	**10**	**5**	**15**				

Note: Highest Accuracy of 62% is achieved by GLCM-R features at d=4

Conclusion:

The highest classification accuracy 62% is achieved by GLCM range features at inter pixel distance d =4 , however it is worth mentioning that these features yield reasonable ICA for HEM and FNH cases but not for HCC cases. So in future the performance of other features for differential diagnosis between HEM, FNH and HCC cases shall be explored.

Key References:

[1] Mryka Hall-Beyer, GLCM Texture Tutorial v. 3.0 March 2017, University of Calgary-Calgary, Alberta T2N 1N4 Canada.

[2] Image Database Link: https://www.ultrasoundcases.info/cases/abdomen-and-retroperitoneum/liver/

[3] Robert M. Haralick, K. Shanmugam, Iis'hak Dinstein, Textural Features for Image Classification, IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-3, Issue-6, November-1973., pp. 610-621.

[4] Mittal D, Kumar V, Saxena SC, Khandelwal N, Kalra N. Neural network based focal liver lesion diagnosis using ultrasound images. Comput Med Imaging Graph. 2011 Jun;35(4):315-23. doi: 10.1016/j.compmedimag.2011.01.007. Epub 2011 Feb 18. PMID: 21334176.

[5] LibSVM: Library of Support Vector Machine downloaded from: https://www.csie.ntu.edu.tw/~cjlin/libsvm/

[6] B. Schmauch et. Al. Diagnosis of focal liver lesions from ultrasound using deep learning, Diagnostic and interventional imaging (2019), 100, pp. 227-233. Elsevier.