Journal of Medical Systems

Automated Classification of Liver Disorders using Ultrasound Images

--Manuscript Draft--

Manuscript Number:	JOMS1339R1
Full Title:	Automated Classification of Liver Disorders using Ultrasound Images
Article Type:	Research Paper
Section/Category:	
Keywords:	Ultrasound, Fatty liver disease, Heterogeneous Liver, Wavelet Packet Transform, Support Vector Machines.
Corresponding Author:	Fayyaz ul Amir Afsar Minhas Colorado State University Fort Collins, Colorado PAKISTAN
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Colorado State University
Corresponding Author's Secondary Institution:	
First Author:	Fayyaz ul Amir Afsar Minhas
First Author Secondary Information:	
All Authors:	Fayyaz ul Amir Afsar Minhas
	Durre Sabih
	Mutawarra Hussain
All Authors Secondary Information:	
Abstract:	This paper presents a novel approach for detection of Fatty liver disease (FLD) and Heterogeneous liver using textural analysis of liver ultrasound images. The proposed system is able to automatically assign a representative region of interest (ROI) in a liver ultrasound which is subsequently used for diagnosis. This ROI is analyzed using Wavelet Packet Transform (WPT) and a number of statistical features are obtained. A multi-class linear support vector machine (SVM) is then used for classification. The proposed system gives an overall accuracy of ~95% which clearly illustrates the efficacy of the system.

Suggested Reviewer List

Dr Muhammad Arif
Professor, Dept. of Computer Science
College of Computer and Information system,
Al-Abidiyya Building
Um-Alqura University, PO Box 715
Makkah, Saudi Arabia
Webpage: http://syedmarif.webnode.com/
Email: syedmarif2003@yahoo.com

Dr. Charles Anderson
Professor, Dept. of Computer Science
Colorado State University
1873 Campus Delivery
Fort Collins, CO 80523
USA
Web: http://www.cs.colostate.edu/~anderson/
Email: anderson@cs.colostate.edu

Dr. AN Khan
Consultant Radiologist
The Pennine Acute Hospitals NHS Trust
North Manchester General Hospital
Delaunays Road
Crumpsall
Manchester M8 5RB
Tel 00441617954567 Ext3897
00441617202270
(Formerly: Chairman Imaging. King Fahd Hospital NGHA, King Abdulaziz Medical City Riyadh Saudi Arabia)
Email: drkhan1966@msn.com

Automated Classification of Liver Disorders using Ultrasound Images

Fayyaz ul Amir Afsar Minhas[1], Durr-e-Sabih[2], Mutawarra Hussain[3]

[1]*Department of Computer Science*

Colorado State University, Fort Collins, CO 80523, USA

fayyazafsar@gmail.com

[2] *Multan Institute of Nuclear Medicine and Radiotherapy (MINAR)*

Multan, Pakistan

dsabih@yahoo.com

[3] *Department of Computer Science*

Pakistan Institute of Engineering and Applied Sciences

Nilore, Islamabad, Pakisan

mutawarra@pieas.edu.pk

Abstract: This paper presents a novel approach for detection of Fatty liver disease (FLD) and Heterogeneous liver using textural analysis of liver ultrasound images. The proposed system is able to automatically assign a representative region of interest (ROI) in a liver ultrasound which is subsequently used for diagnosis. This ROI is analyzed using Wavelet Packet Transform (WPT) and a number of statistical features are obtained. A multi-class linear support vector machine (SVM) is then used for classification. The proposed system gives an overall accuracy of ~95% which clearly illustrates the efficacy of the system.

Keywords: *Ultrasound, Fatty liver disease, Heterogeneous Liver, Wavelet Packet Transform, Support Vector Machines.*

INTRODUCTION

Fatty liver disease (FLD) is caused by the accumulation of fat in the liver. It affects about 15% of the world population. Common causes include alcohol overuse, Insulin resistance, obesity and hyperlipidemia [1]. FLD has a strong association with chronic liver disease [1, 2] therefore the correct diagnosis of FLD plays an important role in the long term health of a subject.

Liver biopsy is considered the diagnostic reference standard for the assessment of FLD; but some authors have noted a variation in the histological interpretation of liver biopsies specially in diagnosing inflammation and fibrosis but good correlation is found in fatty changes[3]; however, the invasiveness of the procedure limits its widespread use to diagnose this condition. Noninvasive diagnostic techniques for FLD include ultrasound, Computerized Tomography (CT) and Magnetic Resonance Imaging (MRI) [1]. Ultrasound imaging offers a inexpensive diagnostic tool for the detection of FLD. Ultrasound liver texture can be used to classify a liver as normal, fatty, fibrofatty, fibrotic or inflammatory [3]. FLD can be diagnosed by ultrasound by observing the echogenicity of the liver [1, 2]. The echogenicity of a normal liver equals or minimally exceeds

that of the renal cortex or spleen and the intrahepatic vessels are clearly demarcated. In FLD ultrasound images show increased echogenicity in comparison to renal cortex or spleen, there is attenuation of the ultrasound energy with loss of definition of the diaphragm and poor delineation of the intrahepatic architecture. There are difficulties with this approach due to variation in image quality in different machines and patient-to-patient variation [4]. Accurate use of visual criterion for the diagnosis of FLD requires experience on the part of the medical expert. Moreover, visual assessment, due to its subjective nature is inherently variable and non-reproducible. The reported sensitivity and specificity for FLD detection using ultrasound are 60%-100 and 77%-95% respectively [1]. The current paper presents a statistical machine learning approach for performing accurate and efficient diagnosis of FLD based on ultrasound images with minimal user input.

Liver texture heterogeneity has been shown to have a correlation with fibrosis in chronic liver disease and has been assessed as a separate parameter from increased texture of FLD[5]. There is considerable overlap and subjectivity involved in diagnosing borderline cases of both FLD and heterogeneity of liver, the situation becomes even more complex when the two findings coexist. The approach presented in this paper allows detection of heterogeneity in the liver texture along with FLD.

Automatic detection of FLD or other diffuse liver disorders has been investigated by a number of researchers. The most common features employed for diagnosis of FLD are based upon texture analysis of the ultrasound image. These include mean gray level, gray level variance and descriptors of the gray level histogram such as selected percentiles [6, 7]. Apart from these, features based upon the co-occurrence matrix such as contrast, entropy, correlation [8], fractal parameters [9] and grayscale change rate [10] have also been reported for the analysis of ultrasound images. Ribiero et al. and Li et al. [11, 12] have used radio frequency and speckle features for FLD detection. Li et al. [12] suggested a Near-Far-Field Grayscale Ratio (NFFGR) to model the increased attenuation caused by presence of fat in the liver. To calculate this feature, two separate rectangular regions of interest (ROI) (one in the near field of the ultrasound with the other in the far field) are selected and the ratio of the sum of grayscale values in the two sections is calculated. Our proposed approach uses a similar criterion for measuring attenuation but uses only a single ROI. In this paper, we have used simple statistical features for modeling the change in echogenicity, homogeneity and attenuation characteristics caused by FLD. We have employed wavelet packet transform (WPT) [13] for performing a multi-resolution analysis of sections taken from an ultrasound image. The use of wavelet packet transform allows a rich representation and allows analysis of the granularity of the liver due to its multi-scale nature. Both WPT and Discrete Wavelet transform (DWT) [14] have been extensively used in texture analysis [15] and some applications of wavelet methods for ultrasound image analysis also exist [16, 17]. In the absence of a standard dataset, a fair comparison of the performance of existing methods of FLD detection is difficult as the accuracy for FLD classification in the literature range from 85-95%. The proposed scheme is compared against several existing approaches in the results section.

MATERIALS & METHODS

Different steps involved in the proposed approach are depicted in figure 1. These steps are detailed henceforth.

[FIGURE 1 GOES HERE]

Image Acquisition

Ultrasound images used in this work were acquired at Multan Institute of Nuclear Medicine and Radiotherapy (MINAR), Multan, Pakistan by professional medical experts. All images were acquired on a Toshiba SSA 550 digital ultrasound machine with a convex probe; all images were acquired at a 5MHz Tissue Harmonic Imaging frequency. The images are of size 560x450 and are saved as bitmap files. These images were annotated by a medical expert (DS) as 'Fatty', 'Normal' or 'Heterogeneous' and these labels are used as the gold standard in this preliminary study. Ultrasound images of 88 subjects (30 with FLD, 39 normal, 19 heterogeneous) were acquired. Image acquisition settings for the ultrasound machine such as focus, frequency, zoom and probe location were kept the same for all images to minimize unwanted variations. An example image from the dataset used in this work is shown in Figure 2.

During data collection, a 64x64 sized region (inter-costal section showing segment 7) was manually selected by our medical expert as the region of interest (ROI) in each ultrasound images. In order to avoid distortion, these sub-images were taken near the center lobe of the ultrasound image. Care was taken to avoid hepatic vessels, bile stores in the liver and areas of echo non-homogeneity while selecting the ROI. The white rectangle in Figure 2 indicates a selected sub-image. These regions are representative regions on the basis of which diagnosis can be done. These manually obtained regions of interest are used for training in our segmentation procedure.

[FIGURE 2 GOES HERE]

Segmentation

The purpose of segmentation is to automatically discover regions of interest (without the assistance of a medical expert) in a liver ultrasound which can then be used for performing diagnosis. As stated earlier the ROI should be near the central lobe and should not include hepatic vessels, bile stores and other anomalies. For performing segmentation, we have developed a novel approach based upon Continuous Wavelet Transform (CWT) and one-class support vector machine (SVM) classifiers.

Since our medical expert has manually extracted ROIs from our training ultrasound images therefore we do have positive training examples (regions that should be taken as the ROI) but we do not have any negative training examples (corresponding to areas in the image which are to be avoided in our search for ROI). Moreover the types of anomalies in a liver ultrasound can be quite diverse. This makes the use of a conventional classifier for finding appropriate regions of interest impractical. We, therefore, resorted to one-class classification (or novelty detection) in

segmentation, i.e. we would be looking for regions which are the most similar to the ones selected by our medical expert. Figure 3 shows different steps involved in segmentation.

[FIGURE 3 GOES HERE]

A given ultrasound image is first cropped to extract a wide region near the central lobe. This is done in order to reduce the computational cost of the process. The continuous wavelet transform (CWT) of the cropped image is then taken. The CWT of an image *f(x,y)* at scale *'s'* using wavelet *Ψ(x,y)* is given mathematically as:

$$cwt(s,a,b) = \frac{1}{\sqrt{s}} \int\int f(x,y)\psi(\frac{x-a}{s}, \frac{y-b}{s})dxdy \quad (1)$$

In frequency domain,

$$CWT(s,\omega_1,\omega_2) = \sqrt{s}F(\omega_1,\omega_2)\Psi(s\omega_1, s\omega_2) \quad (2)$$

CWT allows a multi-resolution analysis of the given image. We used a Gaussian wavelet with 8 different scale parameter values (1.0, 1.6, 2.6, 3.9, 4.0, 5.0, 5.4 and 7.0). Figure 4 shows the 8 images obtained from CWT (of the image in Figure 2). Thus a single pixel of the original image can now be represented as a 9 dimensional (one value from original image and 8 from CWT) feature vector.

[FIGURE 4 GOES HERE]

Each of the images in the training set is analyzed in this manner and the pixels in regions selected manually by our medical expert are used to train a one-class support vector machine. As each of the manually selected ROI is 64x64 pixels, therefore the total number of pixels available for training becomes 4096 for a single ROI. Thus, the total number of examples for training the one class SVM for all the 88 images is 360,448. For efficient training of the one-class SVM, the number of examples was reduced from 360,448 to 4096 using k-means clustering (with k=4096). The cluster centers from the k-means clustering algorithm are then used to train the one class SVM. 4096 training examples proved to be sufficient for training the one class SVM for segmentation.

Conventional SVM requires both positive and negative examples for training whereas one-class SVM is used where data for only the positive (true) class is available for training. It uses the kernel trick to first project the data points into a high dimensional space in which the positive class is expected to become linearly separable from the outliers that are placed close to the origin (see Figure 5). Lagrange optimization is then used to find a hyper-plane that best separates the data points from the origin. The one class SVM can be put in the form of a constrained optimization problem given below.

$$\begin{aligned} \min_{\mathbf{w},\xi,\rho} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\nu l}\sum_{i=1}^{l}\xi_i - \rho \\ \text{subject to} \quad & \mathbf{w}^T\Phi(\mathbf{x}_i) \geq \rho - \xi_i, \qquad \xi_i \geq 0 \end{aligned} \quad (3)$$

In this optimization problem, *l* is the number of training examples (*l=4096*), and $\mathbf{x_i}$ are the training vectors. The kernel function can be written as $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$ with $\Phi(\mathbf{x}_i)$ being the function which transforms $\mathbf{x_i}$ to high dimensional space. This optimization leads to a

hyper-plane ($\mathbf{w}^T\Phi(\mathbf{x}_i)=\rho$) in the high dimensional space. This optimization attempts to place all the training data points above the hyper-plane but whenever this is impossible some points are placed below the hyper-plane and those points have their corresponding $\xi_i > 0$. The parameter ν (with range 0 to 1) controls the number of outliers: a small value of ν reduces the number of training examples which will be classified as outliers as the cost of constraint violation becomes large and vice-versa. We have used the Radial Basis Function (RBF) kernel $K(\boldsymbol{u},\boldsymbol{v})=exp(-\gamma||\boldsymbol{u}-\boldsymbol{v}||^2)$ with γ=1 and ν=0.5 is used in this work. The decision for a given **x** is made as follows:

$$g(\mathbf{x}) = \operatorname{sgn}\left(\mathbf{w}^T\Phi(\mathbf{x}) - \rho\right) \quad (4)$$

[FIGURE 5 GOES HERE]

For segmentation, the CWT of a given ultrasound image is taken and the 9-dimensional representation **x** of each pixel is subjected to the one-class SVM. The following parameter for each **x** is computed from the trained one-class SVM:

$$z(\mathbf{x}) = \mathbf{w}^T\Phi(\mathbf{x}) - \rho \quad (5)$$

The Z-image obtained using the above equation is then normalized as follows:

$$z'(\mathbf{x}) = \frac{z'(\mathbf{x}) - z_{\min}}{z_{\max} - z_{\min}} \quad (5)$$

Morphological opening is then applied to the normalized image using a square (2x2) structuring element in order to remove isolated pixels. A 64x64 averaging filter is then applied on the output of the opening procedure and the maximum value in the averaged image is then selected as the center of the 64x64 ROI. These steps are depicted in figure 6 below.

[FIGURE 6 GOES HERE]

Wavelet Packet Transform

2D Discrete Wavelet Packet Transform (WPT) [13] is a well known method of multi-resolution image analysis. In both the conventional Discrete Wavelet Transform (DWT) [14] and WPT, the image $I(x,y)$ is broken down into approximation (*A*) and details (*D*) coefficients in three directions (horizontal (*h*), vertical (*v*) and diagonal (*d*)). In DWT only the approximation coefficients are analyzed at the next level of decomposition whereas for the WPT, each of approximation and details coefficients are broken down further and the process is repeated to the desired level of decomposition. WPT for 2 levels is shown in Figure 7. WPT offers multiple ways to represent the same signal and the WPT tree includes the DWT representation as well. The optimal representation of the signal is selected through a Shannon entropy-based criterion at each node of the decomposition tree to quantify the information to be gained by performing each split. Therefore the WPT offers a richer analysis of the signal in comparison to the DWT.

Figure 7 illustrates the procedure for obtaining WPT coefficients for the $(j+1)^{th}$ level given the (approximation or details) coefficients of the previous level in cX_j. The rows of the approximation

coefficients are passed through a low pass decomposition filter L_D and the columns of the output are down sampled by a factor of 2 (indicated by the downward arrow block in which only even indexed columns are retained). The columns of down sampled output are passed through L_D and the high pass decomposition filter H_D to produce the approximation coefficients for the $(j+1)^{th}$ level and the horizontal details cD_{j+1}^{h} after down sampling rows of the filtered output. The vertical and diagonal details coefficients are obtained through a similar procedure (lower half tree of figure 7b) which starts with passing the rows of input through H_D. The low and high pass filters are dependent upon the choice of the wavelet used for analysis.

Figure 8 shows the WPT of an ultrasound image ROI at level 2 using the Daubechies 3 wavelet. The impetus of using WPT for analysis of ultrasound images stems from the ability of wavelet analysis to capture the granularity in ultrasound images through its multi-scale nature. Granularity is a potent indicator of presence of different types of liver disorders e.g. FLD causes a fine granular pattern in the ultrasound image [18]. Also textural properties of the ultrasound image can be analyzed easily in the decomposed image at different frequency levels.

[FIGURE 7a and 7b GOES HERE]

[FIGURE 8 GOES HERE]

The following statistical features are extracted from each of the WPT sub images:

a. Median (m)
b. Standard deviation (s)
c. Inter-quartile range (q)

The median (m) of the image represents the median value of intensity of the image which is higher for FLD images due to their increased echogenicity caused by fat deposition. The standard deviation (s) and the Inter-quartile range (q) of the image measure the regularity or smoothness of the ultrasound texture. The ratio of the maximum to the minimum value in the original image is also used as a feature. The optimum wavelet was found to be Daubechies 3 through cross validation analysis and the number of levels was chosen to be N=2. These features are extracted for each of the WPT coefficient images and the original ROI image resulting in a total of L=61 features. Figure 9 shows the scatter plot of first five principal components of the 61 features and these plots show a clear discrimination amongst the three classes indicating the effectiveness of the features chosen.

[FIGURE 9 GOES HERE]

Classification

For the purpose of classification we have used a ν-linear support vector classifier (ν-LSVC) [19]. ν-LSVC finds the optimal linear boundary between two classes by maximizing the margin between them by solving the following constrained optimization problem given a total of M training examples denoted by $\boldsymbol{x}_i$ each having L-dimensions with a corresponding class label $y_i=\{+1,-1\}$.

Type equation here.

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^L,\xi\in\mathbb{R}^M,\rho,b} \quad & \tau(\mathbf{w},\xi,\rho)=\frac{1}{2}\|w\|^2-\nu\rho+\frac{1}{M}\sum_{i=1}^{M}\xi_i \\ \text{subject to} \quad & y_i\left(\mathbf{w}^T\mathbf{x}+b\right)\geq\rho-\xi_i,\ \ i=1...M \\ & \xi_i\geq 0,\rho\geq 0 \end{aligned} \tag{7}$$

The parameter ν in the LSVC acts as a lower bound on the fraction of support vectors and an upper bound on the fraction of training errors. It was set at 0.15 in this work. SVM is a binary classifier, however we have used the one-against-all heuristic for performing multi-class classification.

RESULTS AND DISCUSSION

The results of automated segmentation (ROI selection) are shown for some representative images in Figure 10. This shows that the automated segmentation method is able to avoid all the problem areas in the images (bile ducts, regions of non-homogeneous reflectance etc.) while selecting the ROI.

[FIGURE 10 GOES HERE]

As the number of training examples is small, therefore for evaluation of classification, Leave-One-Out cross validation is used. In order to quantify the performance of the proposed method the following measures were employed:

a. **Sensitivity for each class c**: It is defined as follows

   Se=TP/(TP+FN)

   TP (true positives) is the number of instances of class c which have been correctly labeled to belong to class c by the classifier. FN (False Negatives) is the number of instances of class c which have been labeled to belong to any other class by the classifier.

b. **Positive Predictive value for each class c**: It is defined as follows

   PPV=TP/(TP+FP)

   FP (False Positives) is the number of instances of other classes that have been labeled to belong to class c by the classifier.

c. **Accuracy**: It is defined as the ratio of number of correctly classified images to the total number of images tested.

The confusion matrix for different classes (Fatty (F), Normal (N) and Heterogeneous (H)) is given in Table-1. This shows that both the sensitivity and positive predictive values for all the classes are quite high and the overall high accuracy advocates the use of the proposed system in a practical scenario. Figure 11 shows the four misclassifications made by the classifier.

[TABLE 1 GOES HERE]

The images miss-classified by the proposed system are shown below.

[FIGURE 11 GOES HERE]

In order to analyze the effect of increase in number of test samples relative to the size of the training data, we performed two separate cross validation analyses as follows:

a. 10 Fold Cross-Validation

In this analysis the given data was randomly partitioned into K=10 folds and the system was trained on 9 of the folds (90% of the data) at a time and testing was performed on the left out fold. This procedure was repeated 10 times, once for each fold. Thus, for a single random partition of the date, results from 10 Fold Cross-Validation indicate the performance of the system when 90% of the data is used for training and the remaining 10% for testing.

b. 5 Fold Cross-Validation

This analysis procedure is similar to 10 Fold Cross-Validation except with K=5. This cross validation allows us to analyze the performance of the system when 80% of the given data is used as training and the remaining 20% as testing.

Each of the above two experiments was repeated with 10 different random partitions in order to marginalize any data bias associated with partitioning. This allows us to estimate the average and standard deviation of the sensitivity, positive predictive value and accuracy of the system which are shown in the table below.

[TABLE 2 GOES HERE]

The results show that as the number of test examples is increased relative to the number of training examples, the accuracy decreases by only a small amount indicating the robustness of the system.

One of the novelties of the scheme proposed in this work in comparison to existing methods is that our approach considers classification in to three different classes whereas most of the existing research focuses on two class classification (normal versus FLD) only. In practice, it is useful to be able to discriminate heterogeneous liver patterns from normal and fatty liver especially for cases on the borderline between these conditions. As a consequence, the scheme proposed in this paper is more usable in a practical scenario than other existing approaches.

A fair comparison between different existing approaches and the proposed scheme is complicated by the differences in evaluation data sets, the number of classes considered, experiment design and the performance metrics reported in the papers. Latest methods in the literature e.g. [11, 12] perform classification between Normal and FLD only. The results given in [11] show that their method gives a sensitivity of 100% for detecting FLD with an overall accuracy of 95%. Their approach uses only 20 images from 10 different subjects (10 images from 5 Normal subjects and 10 images from 5 FLD subjects). The approach in [12] shows that their method gives sensitivity of 97.1% and 84.0% in detecting FLD and normal liver respectively. The approach given in [20] gives area under the Receiver Operating Characteristics Curve (auROC) to be 0.93. The approach given in [21] shows a classification accuracy of 80.68% over 88 subjects.

For the classification between Normal Liver and FLD only, our method gives a leave-one-out cross validation accuracy of 95.6% with sensitivity values for normal and FLD classes being 97.4% and 93.3% respectively. The auROC is 0.96. This clearly shows the efficacy of the proposed scheme in comparison to existing approaches. Our approach also gives the added benefit of detecting heterogeneous liver pattern. Moreover, our approach offers a completely automated solution since the selection of region of interest from the images is also done automatically.

CONCLUSIONS AND FUTURE WORK

This paper presents an effective method of using the multiscale analysis capability of WPT to detect FLD using features which utilize changes in echogenicity, granularity and homogeneity of ultrasound due to the incidence of FLD. The classification accuracy of the proposed scheme clearly shows its suitability in assisting medical experts by performing automated and accurate diagnosis of FLD and heterogeneous liver. This work can be extended further by enlarging the available database of liver images. Moreover, true class labels can be derived from needle biopsy over a population set to label the training examples more accurately. Another interesting extension of the work can be to use features based upon the differences in echogenicity of the liver from that of spleen and renal cortex.

The authors do not have any conflict of interest.

REFERENCES

1. Hamer, O.W., et al., *Fatty Liver: Imaging Patterns and Pitfalls.* Radiographics, 2006. **26**(6): p. 1637-1653.
2. Tchelepi, H., et al., *Sonography of Diffuse Liver Disease.* Journal of Ultrasound in Medicine, 2002. **21**(9): p. 1023-32.
3. Kutcher, R., et al., *Comparison of sonograms and liver histologic findings in patients with chronic hepatitis C virus infection.* J Ultrasound Med, 1998. **17**(5): p. 321-5.
4. Nicolau, C., L. Bianchi, and R. Vilana, *Gray-scale ultrasound in hepatic cirrhosis and chronic hepatitis: diagnosis, screening, and intervention.* Semin Ultrasound CT MR, 2002. **23**(1): p. 3-18.
5. Nishiura, T., et al., *Ultrasound evaluation of the fibrosis stage in chronic liver disease by the simultaneous use of low and high frequency probes.* Br J Radiol, 2005. **78**(927): p. 189-97.
6. Wun, Y.T. and R. Chung, *Ultrasound characterization by stable statistical patterns.* Computer Methods and Programs in Biomedicine, 1998(2): p. 117-126.
7. Badawi, A.M., A.S. Derbala, and A.-B.M. Youssef, *Fuzzy logic algorithm for quantitative tissue characterization of diffuse liver diseases from ultrasound images.* International Journal of Medical Informatics, 1999. **55** (2): p. 135-147.
8. Yeh, W.C. and S.W. Huang, *Liver fibrosis grade classification with B-mode ultrasound.* Ultrasound in Medicine and Biology, 2003. **29**: p. 1229-1235.
9. Liu, M., *Feature extraction and quantitative analysis of liver ultrasound images*. 2004, Hebei University.
10. Lupsor, M., et al. *Ultrasonography Contribution to Hepatic Steatosis Quantification. Possibilities of Improving this Method through Computerized Analysis of Ultrasonic Image*. in *2006 IEEE International Conference on Automation, Quality and Testing, Robotics*. 2006.
11. Ribeiro, R. and J. Sanches, *Fatty Liver Characterization and Classification by Ultrasound* in *Pattern Recognition and Image Analysis*. 2009, Springer Berlin / Heidelberg. p. 354-361.
12. Li, G., et al. *Computer aided diagnosis of fatty liver ultrasonic images based on support vector machine*. in *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008(EMBS 2008)*. 2008.
13. Coifman, R.R. and Y. Meyer, *Orthonormal wave packet bases*. 1989, Yale University.
14. Gonzalez, R.C. and R.E. Woods, *Wavelets and Multiresolution Processing*, in *Digital Image Processing*. 1992, Addison-Wesley. p. 349-408.
15. Laine, A. and J. Fan, *Texture classification by wavelet packet signatures.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993. **15**(11): p. 1186-1191.
16. Yoshida, H., et al., *Wavelet-packet-based texture analysis for differentiation between benign and malignant liver tumours in ultrasound images.* Physics in Medicine and Biology, 2003. **48**(22).
17. Huang, Y., L. Wang, and C. Li. *Texture Analysis of Ultrasonic Liver Image Based on Wavelet Transform and Probabilistic Neural Network*. in *2008 International Conference on BioMedical Engineering and Informatics*. 2008.
18. Machi, J. and E.D. Staren, *Ultrasound for Surgeons*. 2004: Lippincott Williams & Wilkins.
19. Chen, P.-H., C.-J. Lin, and B. Scholkopf, *A tutorial on v-support vector machines.* Applied Stochastic Models in Business and Industry, 2004. **21**: p. 111-136.
20. Thijssen, J.M., et al., *Non-Invasive Staging of Hepatic Steatosis using Computer Aided Ultrasound Diagnosis* in IEEE Ultrasonics Symposium 2008, Beijing, China, 2-5 Nov. 2008 p. 1987-1990.
21. Ribeiro, R. et al., Diffuse Liver Disease Classification from Ultrasound Surface Characterization, Clinical and Laboratorial Data in Pattern Recognition and Image Analysis, p. 167-175, 2011.

Fig. 1 Different steps in the proposed method

Fig. 2 An example liver ultrasound image and the region of interest (marked square)

Fig. 3 Different steps in Segmentation

Fig. 4 CWT of the ultrasound image at different scales

Fig. 5 Concept learning in one-class SVM

Fig. 6 Different steps in ROI detection (segmentation). Row-wise from the top-left: Original Image, the corresponding Z-image, normalized Z-image (note the change change in the range on the colorbar), after morphological opening, after averaging filter and finding the point with the maximum value (indicated by the star) and the resulting ROI square.

Fig. 7 (a) The 2 level wavelet packet decomposition of image I(x,y) into approximation (cA_1) and details coefficients in horizontal, vertical and diagonal directions (cD_1^h, cD_1^v, cD_1^d). Each of these coefficients is then decomposed further to level 2 yielding a total of 16 sets of coefficients at level 2. (b) The coefficients of the j^{th} level are decomposed further to produce the corresponding approximations and details. The symbols (r) and (c) indicate whether the down-sampling (by 2, indicated by the downward arrow) and thefiltering operation in the indicated box is performed on the rows (r) or columns (c) of the input.

Fig. 8 level WPT decomposition tree of a fatty liver ultrasound ROI (top) using Daubechies 3 wavelet.

Fig. 9 The scatter-matrix plot of the first five principal components (obtained only for visualization) of the 61 features extracted from each sub-image

Fig. 10 Examples of automated ROI selection (segmentation)

Figure 11 The misclassified instances

Table 1 Confusion Matrix for Leave-One-Out Cross Validation

		Classifier Labels			Sensitivity
		Fatty	Normal	Heterogeneous	
True Labels	Fatty	28	2	0	93.3%
	Normal	1	38	0	97.4%
	Heterogeneous	1	0	18	94.7%
Positive Predictive Value		93.3%	95.0%	100.0%	
Accuracy		95.4%			

Table 2 Results for K=5 and K=10 Fold Cross Validation (CV). Average results over multiple cross validation runs are shown. The numbers in the parenthesis indicate the standard deviation.

Performance Metric (%)		K=5 Fold CV	K=10 Fold CV
Positive Predictive Value	Fatty	93.8 (4.5)	92.4 (3.1)
	Normal	87.2 (3.8)	92.3 (2.8)
	Heterogeneous	92.9 (7.3)	100 (0.0)
Sensitivity	Fatty	87.3 (4.1)	92.4 (2.5)
	Normal	94.9 (2.7)	96.3 (2.0)
	Heterogeneous	85.3 (7.8)	91.0 (5.0)
Accuracy		90.2 (2.6)	93.7 (2.0)

*Figure 1
Common.Links.ClickHereToDownloadHighResolutionImage

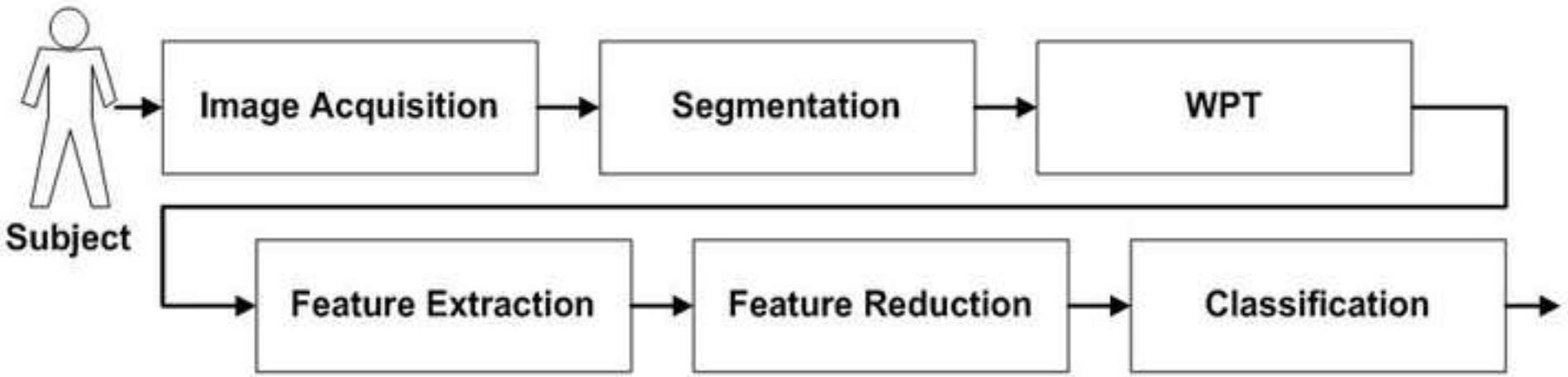


Figure 2
Common.Links.ClickHereToDownloadHighResolutionImage

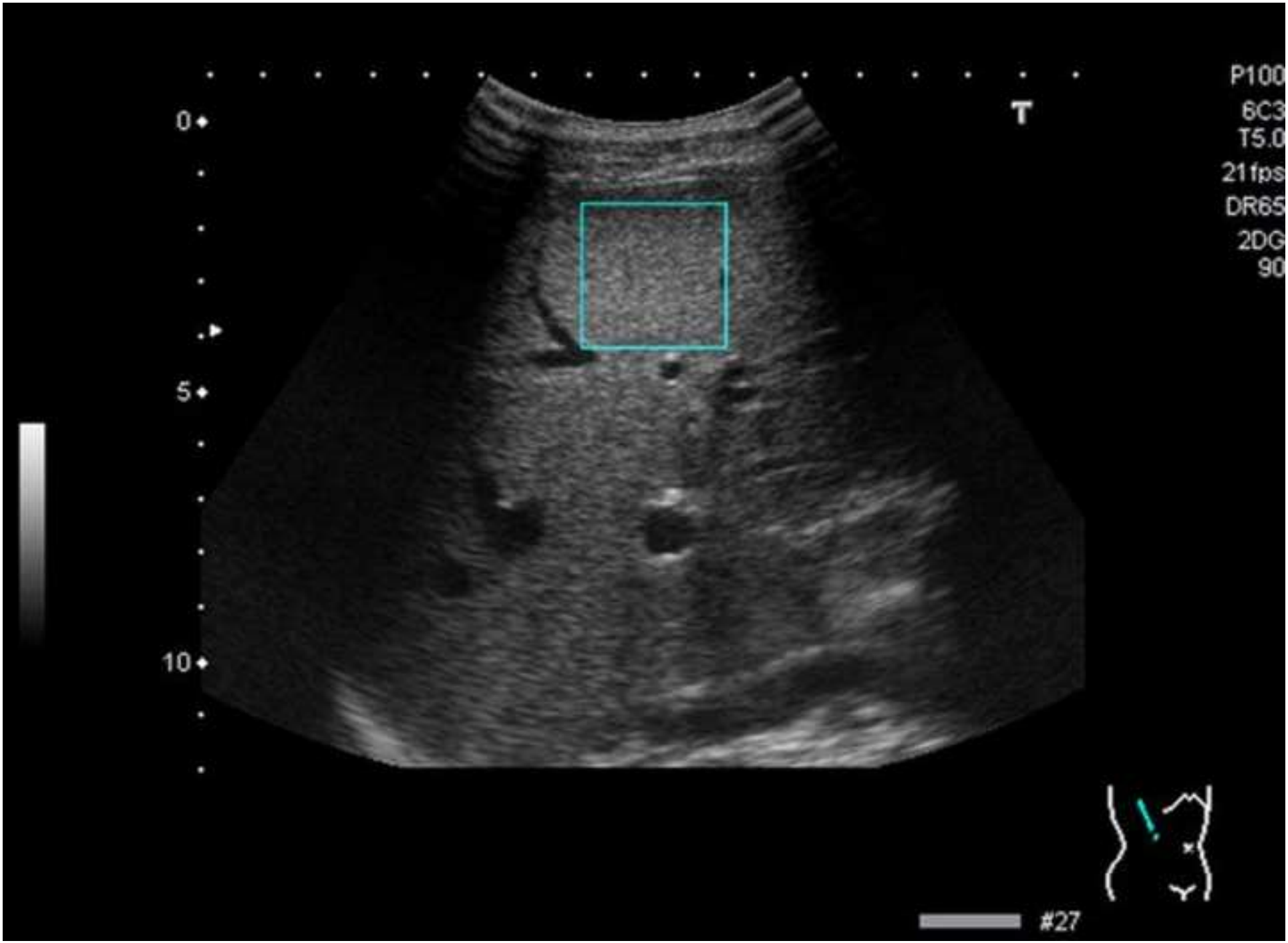


Figure 3
Common.Links.ClickHereToDownloadHighResolutionImage

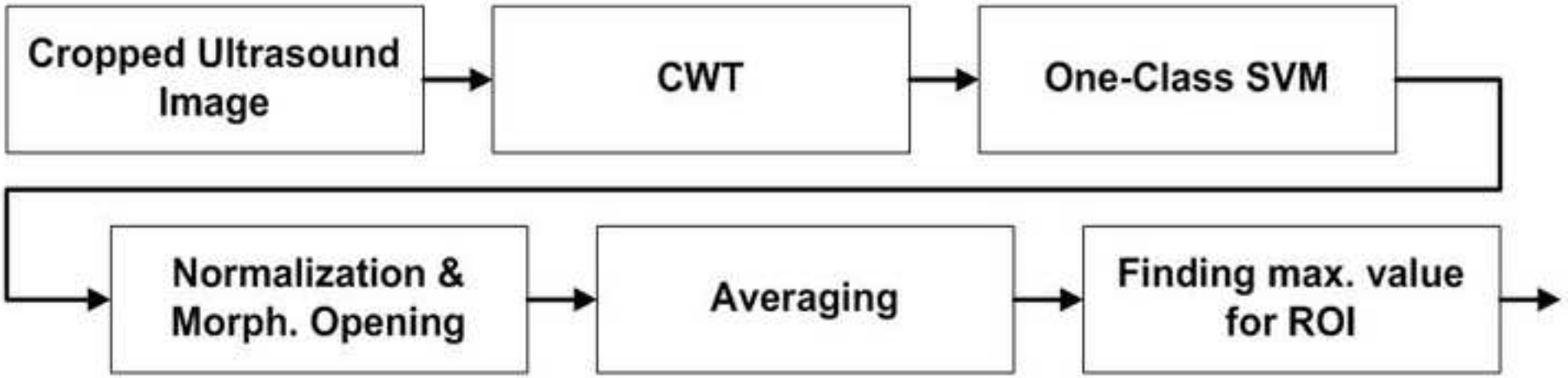


Figure 4
Common.Links.ClickHereToDownloadHighResolutionImage

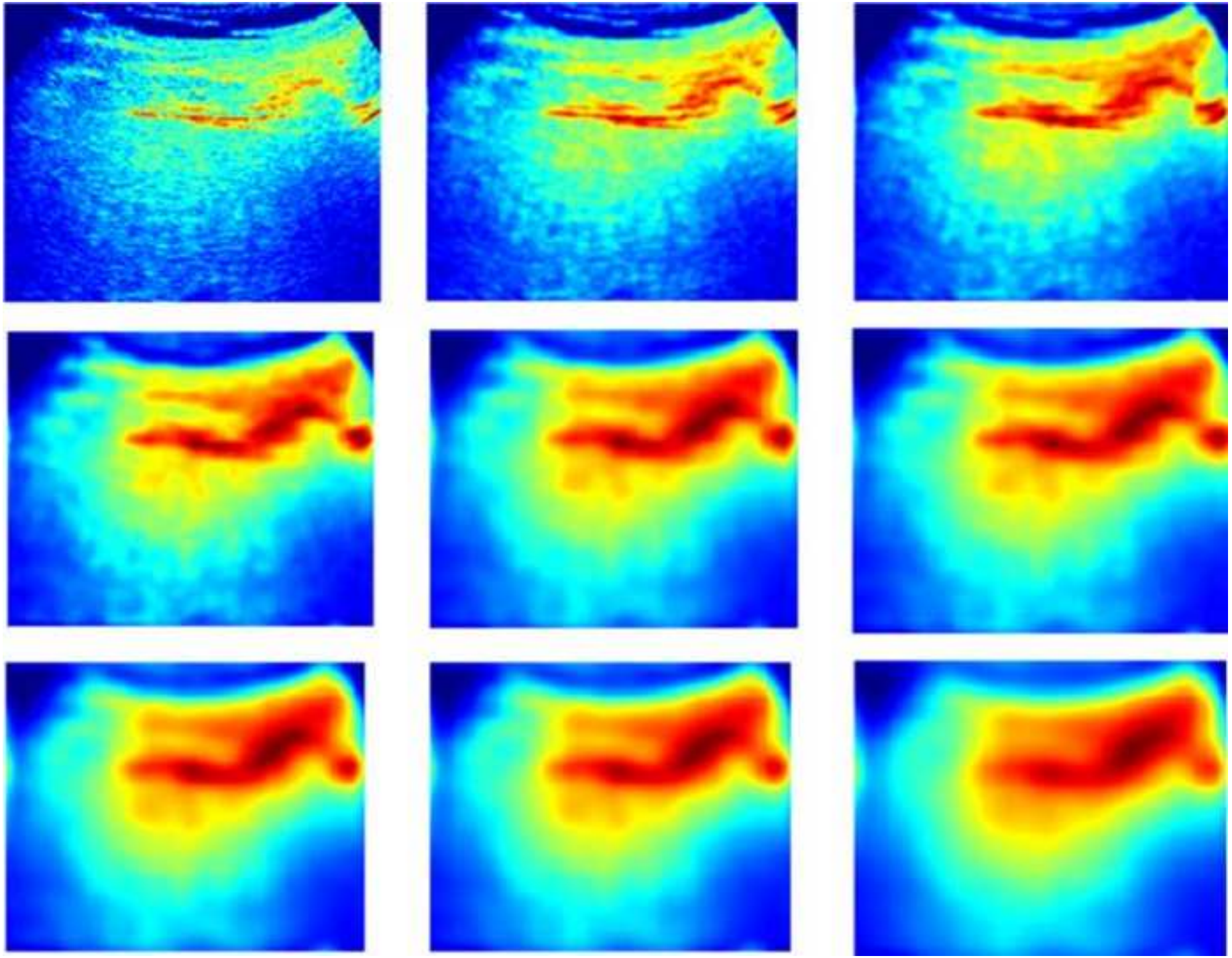


Figure 5
Common.Links.ClickHereToDownloadHighResolutionImage

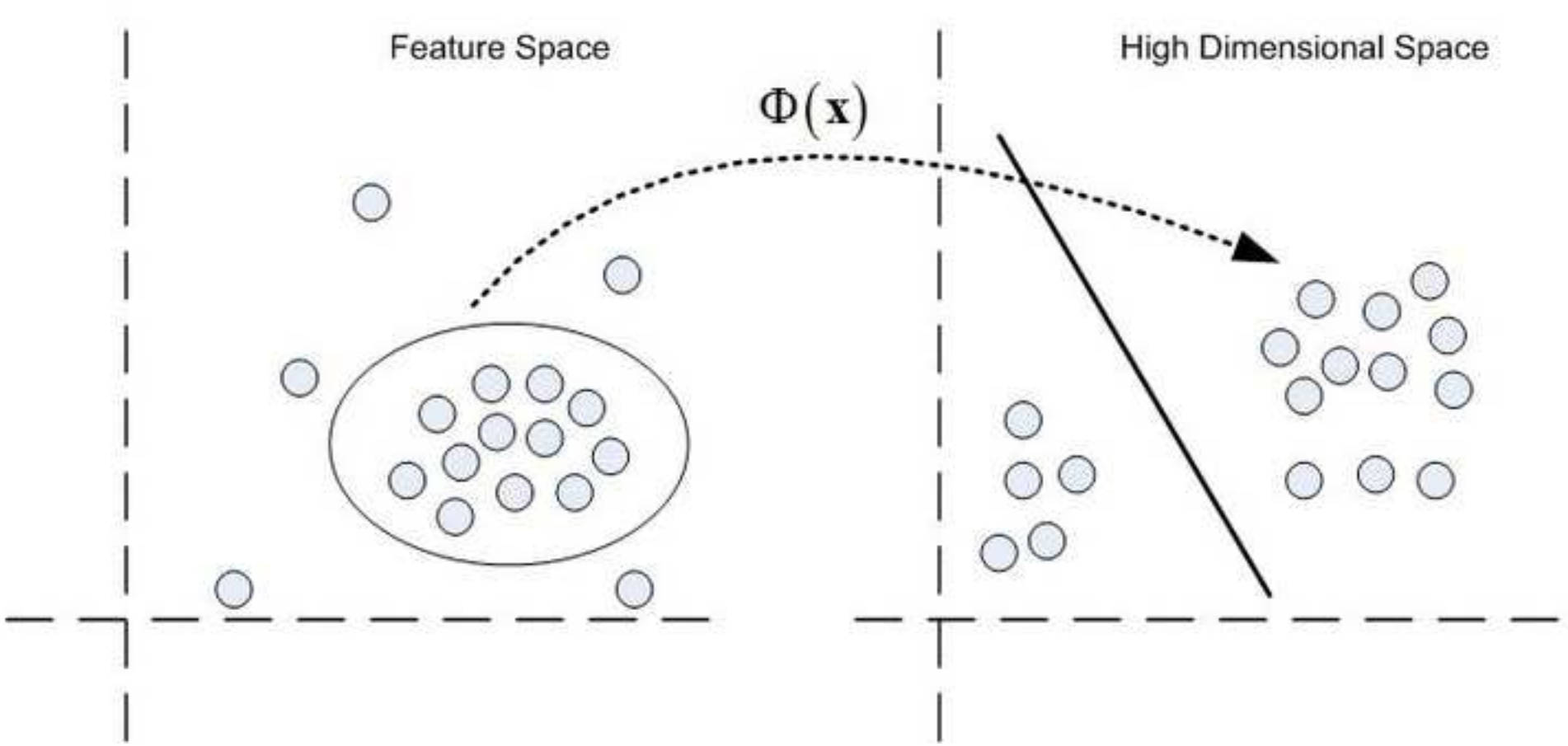


Figure 6

Common.Links.ClickHereToDownloadHighResolutionImage

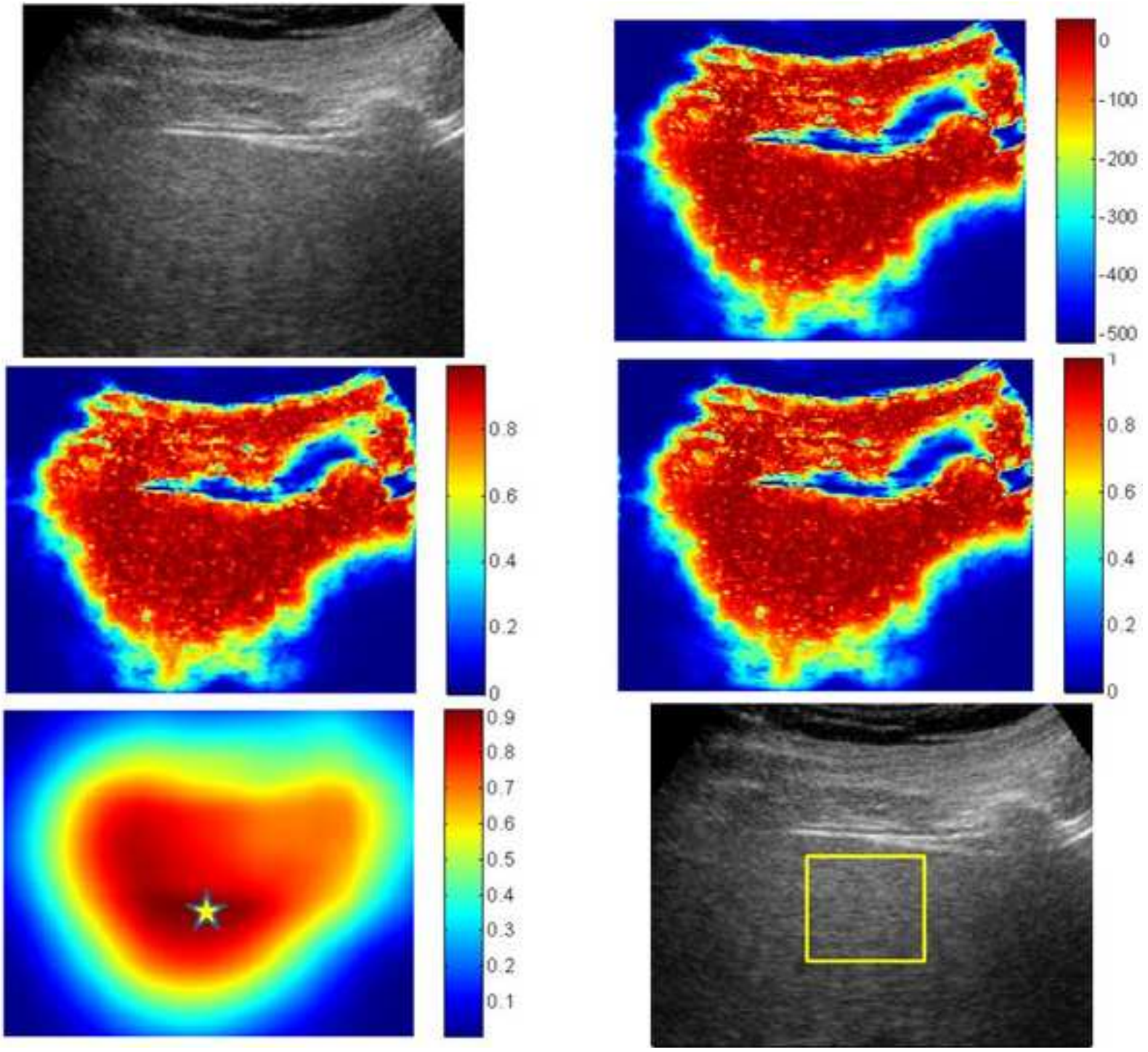


Figure 7a
Common.Links.ClickHereToDownloadHighResolutionImage

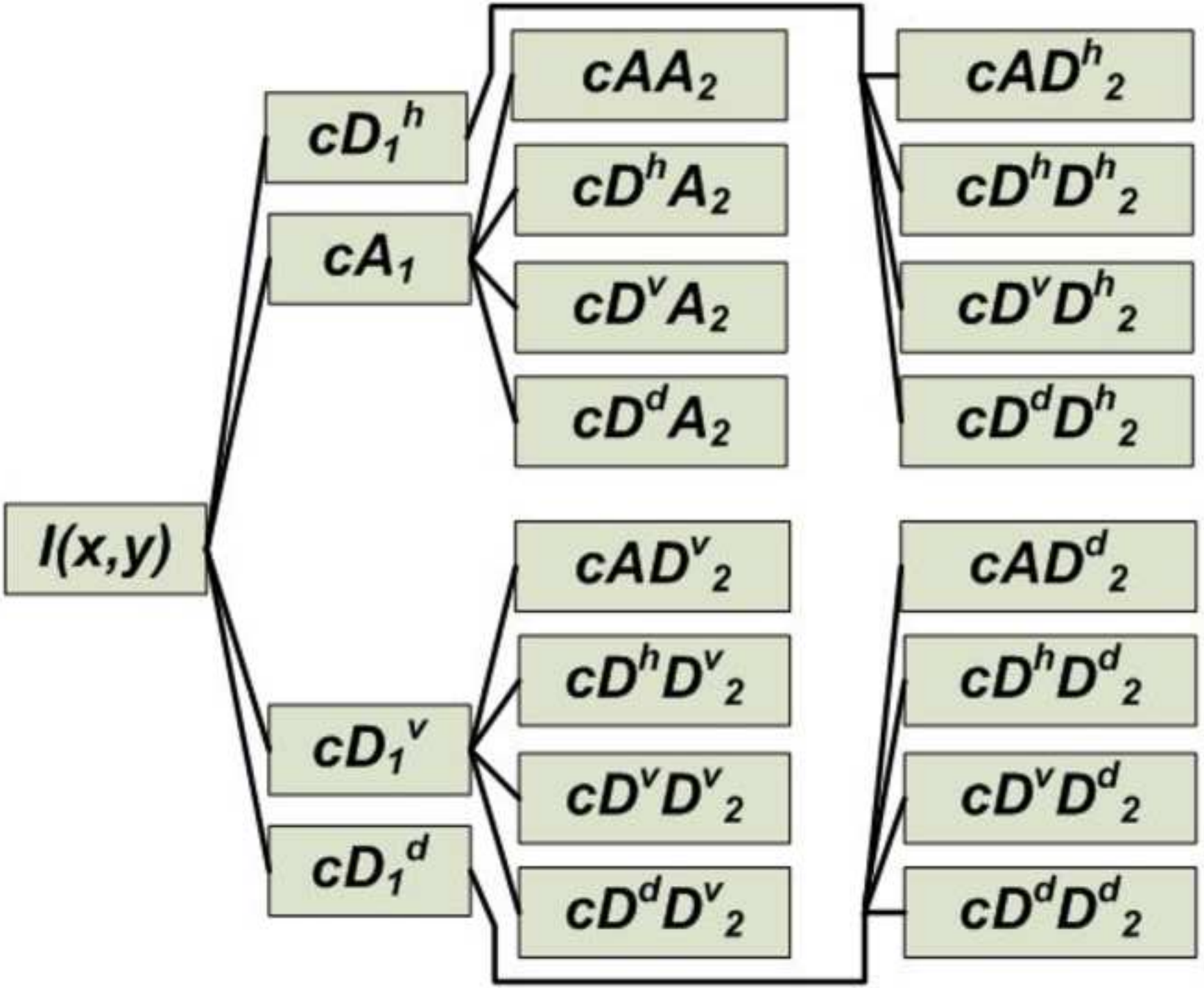


Figure 7b

Common.Links.ClickHereToDownloadHighResolutionImage

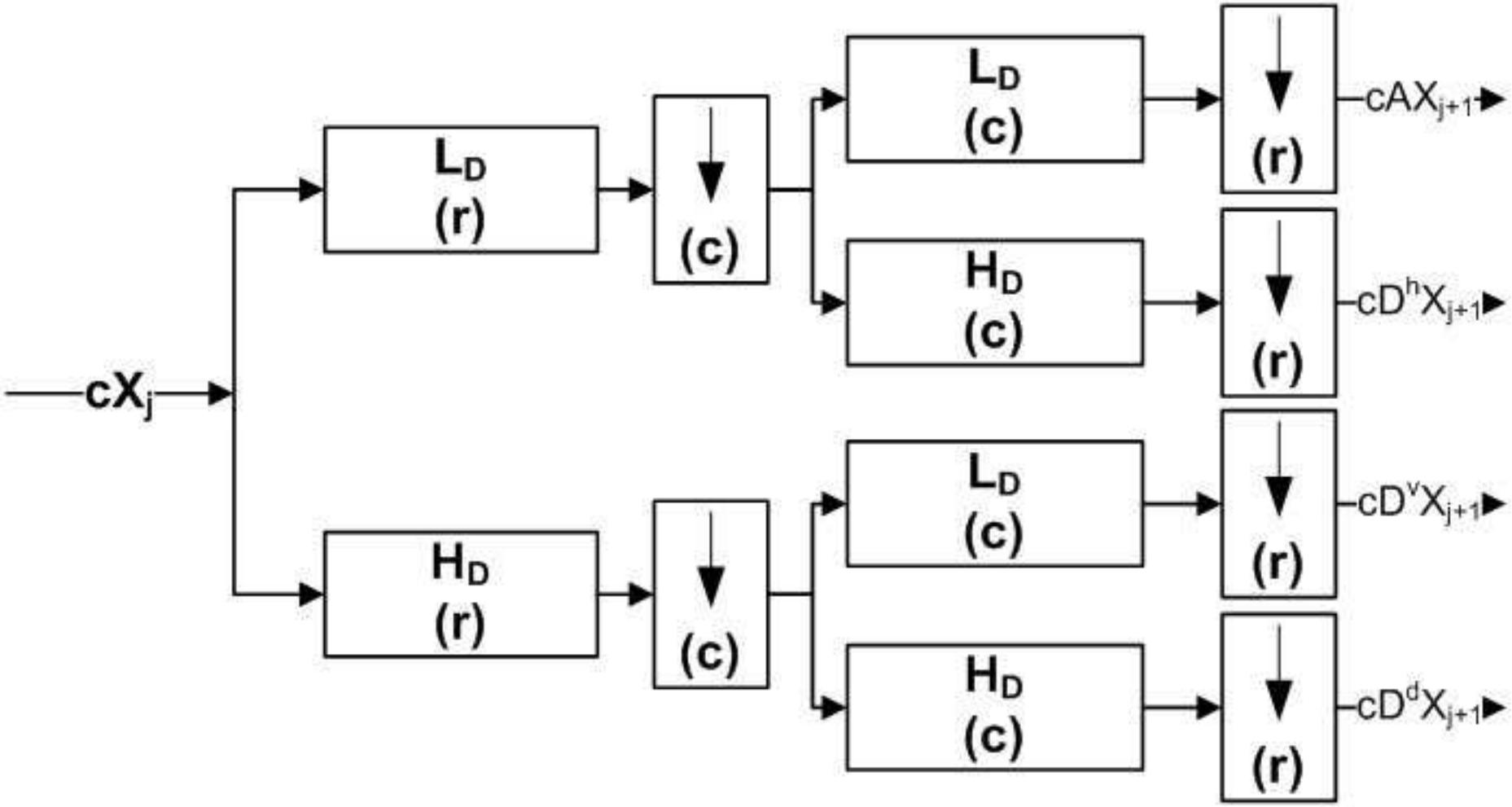


Figure 8

Common.Links.ClickHereToDownloadHighResolutionImage

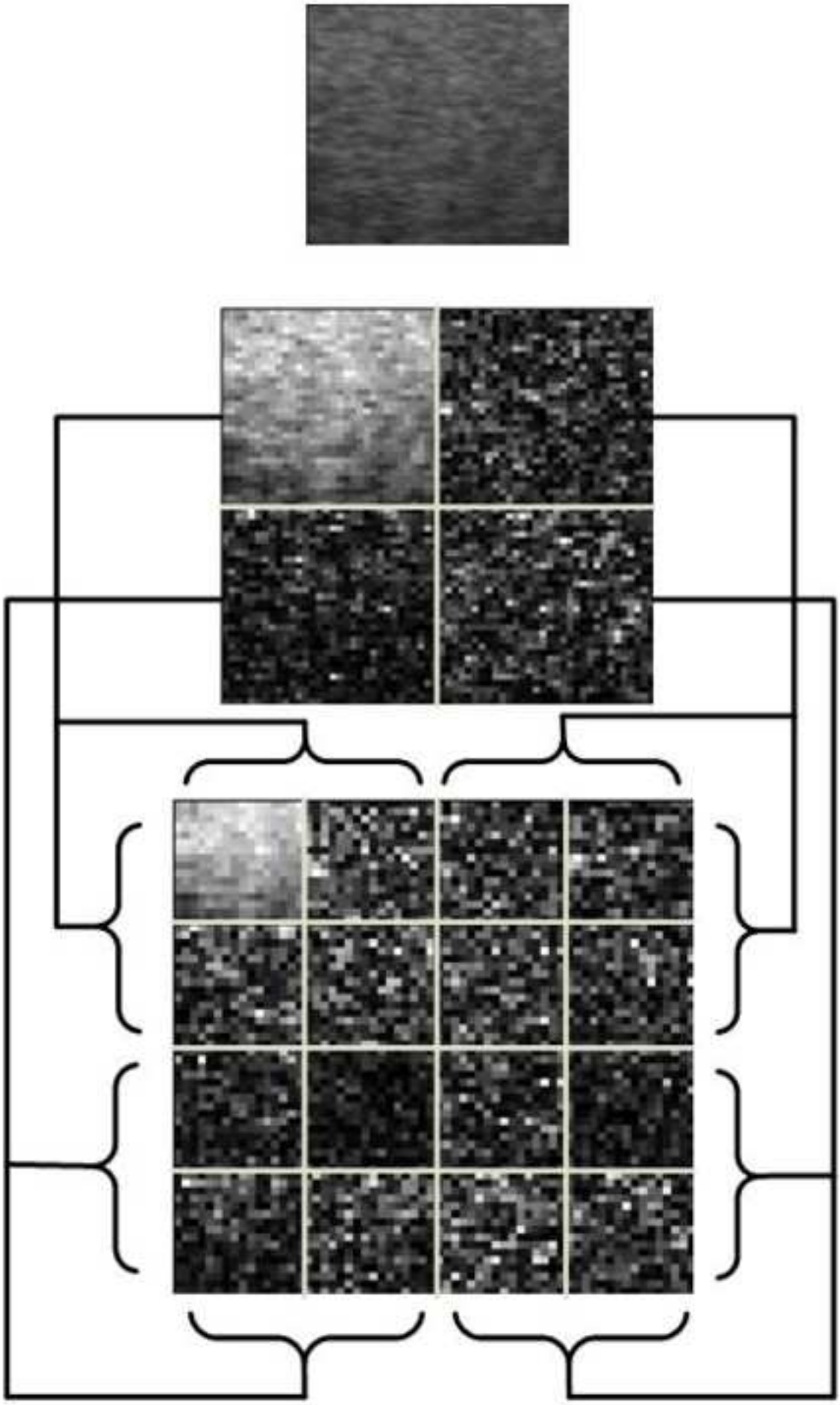


Figure 9
Common.Links.ClickHereToDownloadHighResolutionImage

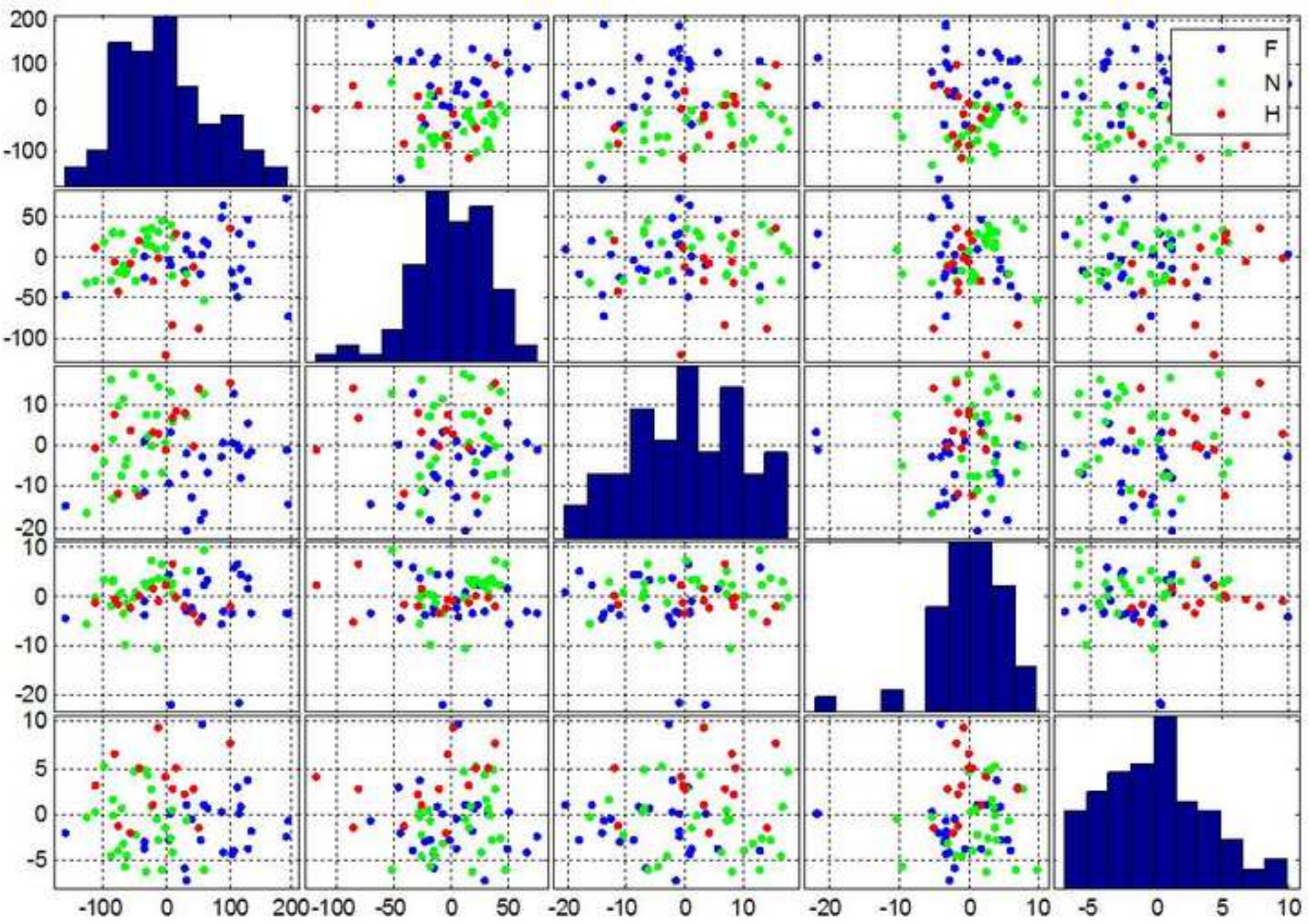


Figure 10

Common.Links.ClickHereToDownloadHighResolutionImage

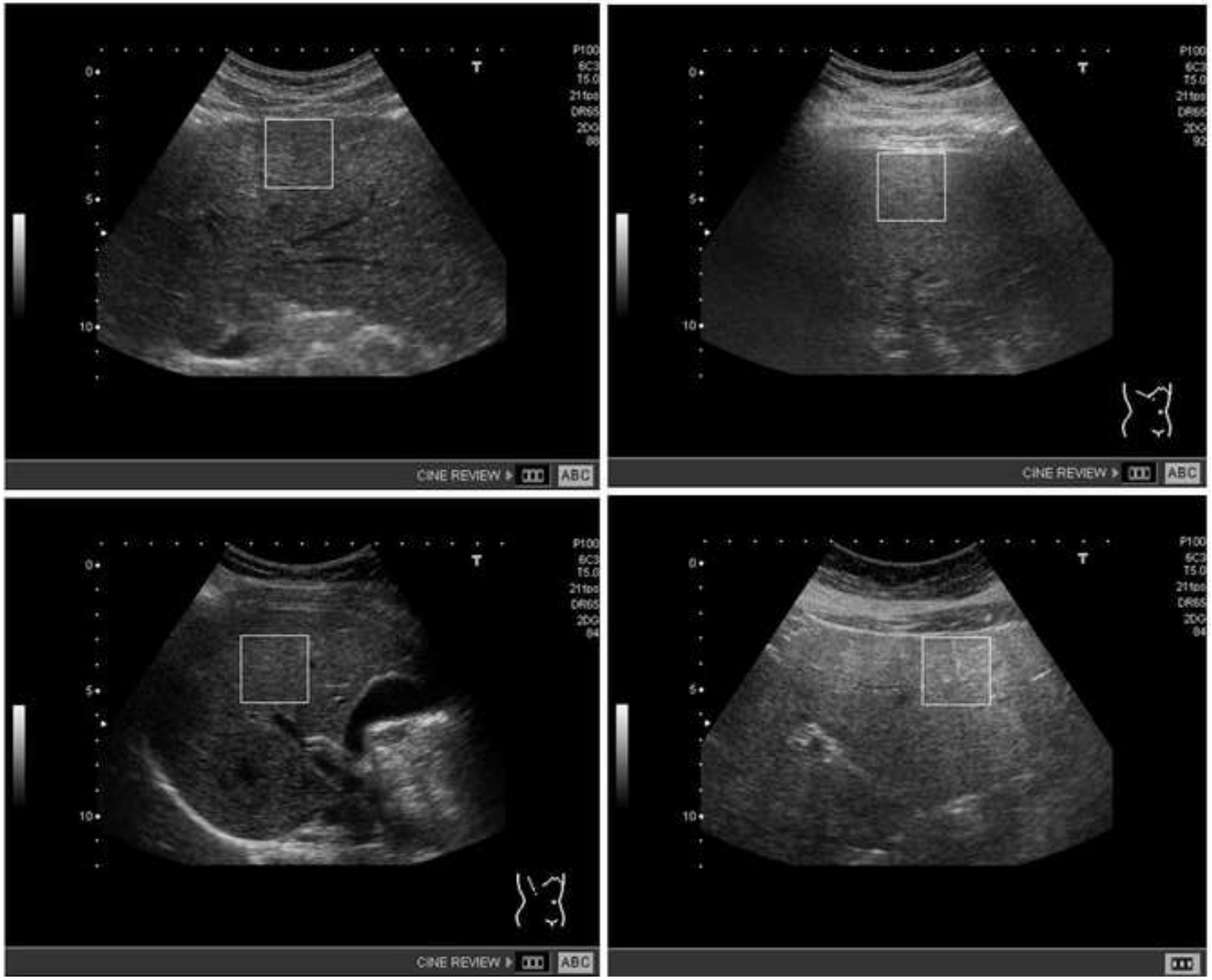
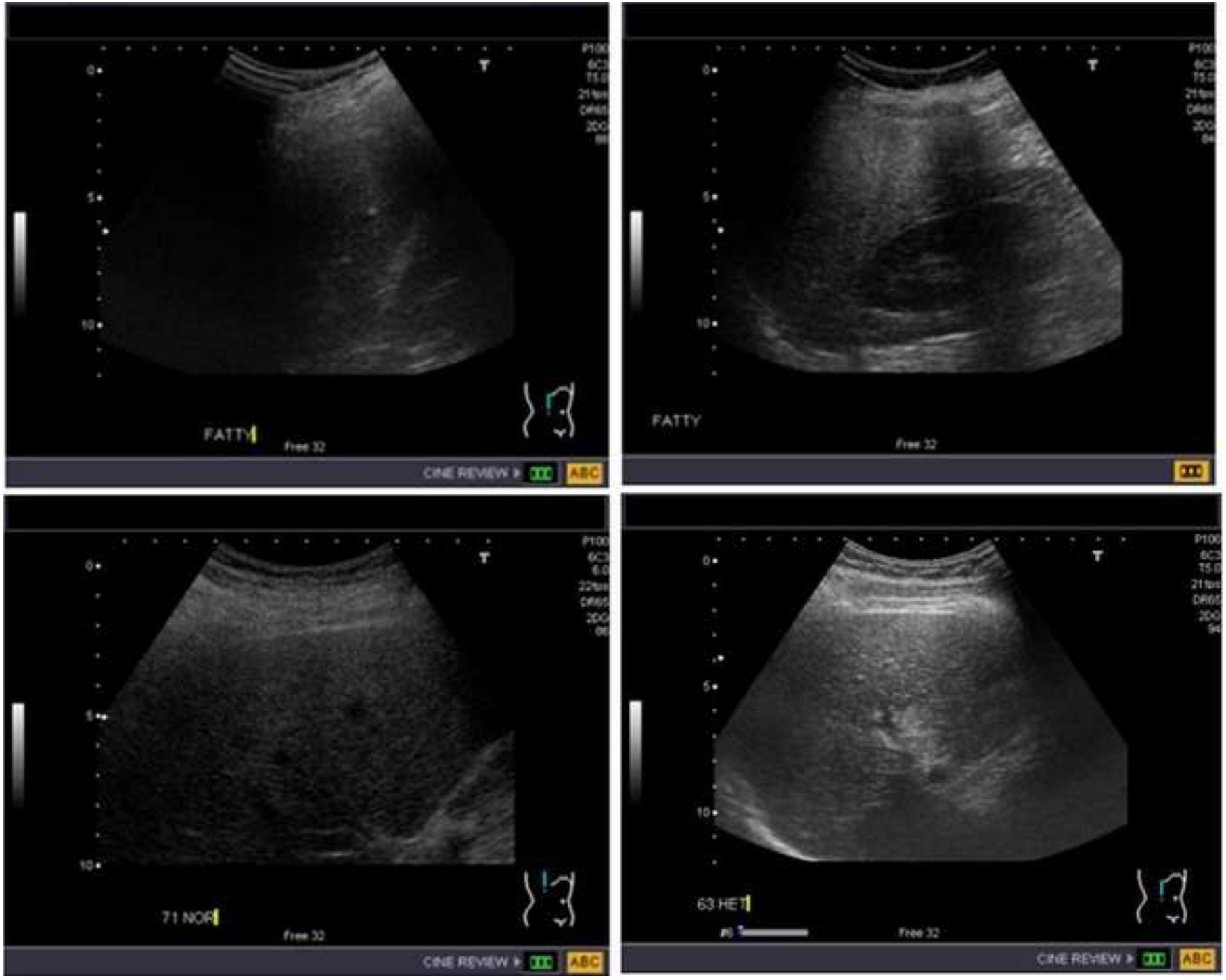


Figure 11
Common.Links.ClickHereToDownloadHighResolutionImage



Dear Editor(s),

Thank you for your thorough and quick review. The manuscript has been updated in accordance with comments by the reviewers. The response to reviewer comments is given below:

Response to Comments by Reviewer #1

Thank you for your comments and encouragement. They really helped us in improving the quality of the paper.

1. Literature review may be updated to include more recent papers.

The paper now contains relevant literature about automated classification of FLD and heterogeneous liver available till 2011.

2. On line 33 page 4, It is reported that 4096 9 dimensional vectors as cluster centers out of all images? In leave one out validation, does k-mean clustering is done for every set of images after leaving one image out? If so then centers are changing during validation.

This is related to the training of the one class SVM which is used for segmentation of the region of interest (ROI) form a given image. Training ROIs of size 64x64 are obtained from a medical expert which are then used to train the one-class SVM in the automatic segmentation algorithm. There are 4096 pixels in each of the 64x64 image and each pixel has a 9 dimensional feature vector (obtained from the Continuous Wavelet Transform) associated with it. The data from all 88 training regions of interest is combined and clustering is performed on it to reduce the amount of data used for training the one class SVM. Once the one class SVM has been trained (using cluster centers obtained from all data points), it is used to automatically select regions of interest in the images. These automatically selected ROIs are then used by the multi-class SVM to generate classification labels (Normal, FLD and heterogeneous).

If the clustering algorithm is run multiple times, it can potentially generate different cluster centers since the initialization of the cluster centers is done randomly. However, it was observed (through visual inspection and by input from our medical expert) that this variation does not affect the performance of the resulting one-class SVM used for segmentation. No significant change in performance was noted by leaving a few images out of the training process as well. A plausible reason for this observation is that a single image contributes only by 1/88 to the clustering algorithm and the concept (of selecting proper ROIs) being learnt is not specific to a small subset of images in the training set. Based upon these observations, the one-class SVM was not re-trained during the cross-validation process in order to save computational time.

3. On line 41, page 6 what are the five classes?

Thanks for pointing this out. It is a typing mistake and has been corrected. The intention was to show a discrimination between the 3 classes in the principal component scatter plot.

4. In the result section, leave one out validation include all the available diversity in the training set

and hence shows good accuracy. Classification results for 90%-10% or 80%-20% (training-testing) will be interesting to illustrate the effect of diversity. These results may also be included in the study.

The results for these analyses have been added to the manuscript.

5. A comparison of reported accuracies in the literature may also be included.

Comparison with the latest available methods (from 2008 onwards) has been added in the paper.

6. Some figures are not labeled properly.

Labeling of the figures has been corrected.

Response to Comments by Reviewer #4

Thank you for your comments and encouragement.