The Application of Texture Analysis Pipeline on MRE imaging for HCC diagnosis

by

Gaurav Bansal

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2013 by the
Graduate Supervisory Committee:

Teresa Wu, Co-Chair
Ross Mitchell, Co-Chair
Jing Li

ARIZONA STATE UNIVERSITY

May 2013

ABSTARCT

Hepatocellular carcinoma (HCC) is a malignant tumor and seventh most common cancer in human. Every year there is a significant rise in the number of patients suffering from HCC. Most clinical research has focused on HCC early detection so that there are high chances of patient's survival. Emerging advancements in functional and structural imaging techniques have provided the ability to detect microscopic changes in tumor microenvironment and microstructure. The prime focus of this thesis is to validate the applicability of advanced imaging modality, Magnetic Resonance Elastography (MRE), for HCC diagnosis.

The research was carried out on three HCC patient's data and three sets of experiments were conducted. The main focus was on quantitative aspect of MRE in conjunction with Texture Analysis, an advanced imaging processing pipeline and multi-variate analysis machine learning method for accurate HCC diagnosis. We analyzed the techniques to handle unbalanced data and evaluate the efficacy of sampling techniques. Along with this we studied different machine learning algorithms and developed models using them. Performance metrics such as Prediction Accuracy, Sensitivity and Specificity have been used for evaluation for the final developed model.

We were able to identify the significant features in the dataset and also the selected classifier was robust in predicting the response class variable with high accuracy.

ACKNOWLEDGMENTS

I would like to express my profound gratitude to my advisors Dr. Teresa Wu and Dr. Ross Mitchell for their constant support, encouragement and valuable suggestions, without which this thesis would not have been possible.

My sincere thanks to Dr. Amy Hara, Dr. Alvin Silva and Dr. Wendy Stiles at Mayo Clinic for their valuable contribution to this thesis. Thanks to the member of my dissertation committee, Dr. Jing Li for her thoughtful comments.

I am also grateful to all my lab mates for their help and support: Miao He, Min Zhang and Can Cui. Special thanks to my Mayo Clinic colleagues Helene Hilaire and Gil Speyer, who had helped me a lot in extracting the data from Texture Analysis pipeline and working late hours in order to make the software work for me. It is for their invaluable help and assistance that made the completion of this project possible.

I would like to express my utmost gratitude to my parents, my brother and sister, and my friends without whom this whole journey of education would not even have started.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

1.1 Hepatocellular Carcinoma Disease Diagnosis

Hepatocellular carcinoma (HCC) is a malignant tumor and it is considered as the seventh most frequent occurring cancer in human [1]. An estimate incidence is between 250,000 and 1.2 million cases per year, worldwide [1]. In the United States, the rise in HCC has increased by 80 % in the past two decades [2]. The incidence of HCC is mainly due to Hepatitis C, which also leads to chronic liver complications such as fibrosis, cirrhosis. Nonalcoholic Steatohepatitis (NASH) & Non-Alcoholic Fatty Liver Disease (NAFLD) are also major concerns that are growing really fast and it is expected that they will be surpassing Hepatitis C as the major cause in the near future. The recent high rise of HCC is partially due to the increased obese in the U.S. population. Currently, NAFLD has been considered the most common liver abnormality in the U.S with an estimated prevalence of 25% of the total population. Simple fat (steatosis) is the first stage which leads to fat with inflammation and scarring (NASH) and ultimately to end stage liver fibrosis (cirrhosis).

Liver fibrosis, a symbol of structural liver damage, is an accumulation of Extracellular matrix (ECM) proteins in excess which includes collagen that is a characteristic of most types of liver diseases. It occurs from chronic damage to the liver along with the ECM proteins accumulation. This is an indication for most types of chronic liver diseases [3].

Advanced liver fibrosis leads to cirrhosis which in turn causes hepatocellular dysfunction and increased intrahepatic blood flow resistance. This results in hepatic insufficiency and

hypertension, respectively [4]. In the majority of patients, development to cirrhosis from fibrosis occurs after a span of 15-20 years [3]. Although many patients suffering from cirrhosis may remain free of any major problems for several years, it is a serious sign for developing HCC.

Most clinical research has focused on HCC early detection when the tumor might be curable by resection, liver transplantation, or ablation and a 5-year survival higher than 50% can be achieved [5]. In the U.S. a screening program has been implemented for HCC early detection. In the screening program, biopsy is not required before the treatment and diagnosis of HCC can be done effectively using imaging characteristics. This is probably due to the great promises the recent advanced developments of imaging techniques are showing. Emerging advancements in functional and structural imaging techniques have provided the ability to detect microscopic changes in tumor microenvironment and microstructure, thus allowing the assessment of tumor response after locoregional treatment by observing alterations in tumor viability, perfusion or vascularity.

Commonly used imaging techniques for HCC diagnosis include Ultrasound (US), Computed tomography (CT) and Magnetic resonance (MR). US has been largely replaced by CT and MRI due to its low sensitivity and specificity in HCC diagnosis [6]. As a result, liver disease diagnosis has mainly relied on CT and MRI imaging criteria. However, these two techniques are not without limitations. A retrospective analysis comparing the accuracy of radiologic staging with pathologic staging on liver disease patients found that imaging based diagnosis (both CT and MRI) resulted in a high

number of false positives labeling common benign focal abnormalities in the liver as malignant tissue [7].

One of the major difficulties in imaging cirrhosis is the detection of hyper vascular nodules which are smaller than 2 cm. These nodules often have nonspecific imaging characteristics making their detection highly complicated [8]. MR imaging technique is superior to CT in this area, even though the sensitivities for both of them remain highly disappointing. The MR imaging technique surpasses CT with 81 % pooled estimate of the sensitivity as compared to 68 % for CT as far as detection of HCC is concerned [8]. These research findings indicate there is great room for improving diagnosis accuracy (both sensitivity and specificity). In this research, our central hypothesis is: MR Elastography (MRE), a novel imaging technique developed by Mayo Clinic has the promise to improve HCC tissue characterization. In conjunction with advanced imaging process technique, such as texture analysis, to extract multiple biomarkers as imaging signature, the diagnosis power may be significantly improved. Texture analysis is a generalized tool applicable to varied imaging modalities including US, CT and MRI which are briefly reviewed in the next section.

1.2 Diagnostic Imaging for Hepatocellular Carcinoma

We will first review three common noninvasive imaging modalities namely Ultrasound (US), Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). In addition an advanced imaging modality, Magnetic Resonance Elastography (MRE) developed by Mayo Clinic is briefly reviewed. We further summarize some basic trend in the use of the modalities in HCC diagnosis.

1.2.1 Imaging Modalities

(a) Ultrasound

Ultrasonic Imaging is one of the most popular imaging techniques in medical field. Ultrasound Diagnosis working principle is based on the emission of sound waves with frequency far above the audibility threshold. Typical Ultrasonic frequency range lies from 2-15 MHz, sound frequency below such range is known as infrasound. At these frequencies the speed of sound travels in human body at 1540 m/s [9].

The interval of this wavelength is in the range from 0.75 - 0.1 mm. When sound travels at this wavelength it can be focused in a particular direction and the region of interest (ROI) can therefore be interrogated [9]. Similar to an electromagnetic wave, a sound wave can be reflected, retracted, diffracted, scattered, absorbed or shifted in frequency. In current practices, Ultrasound (US) imaging is being replaced by advanced imaging modalities such as CT scan and MRI owing to low sensitivity and positive prediction value with coexisting cirrhosis.

(b) Computed Tomography (CT)

A CT (computed tomography) scan is a noninvasive imaging technique that is largely replacing US imaging. It uses a special x-ray machine that produces multiple images or pictures of patient's internal organs and to join these images together a computational algorithm is applied thereby rendering cross-sectional views of the region of interest that is being studied. CT scans of internal organs, blood vessels, soft tissue and bones provides a superior clarity as compared to conventional x-ray exams.

CT imaging assessment of HCC suspected patients should be done using multiphase contrast imaging of the liver. The CT scans are carried out at different time intervals corresponding to the contrast enhancement phase. The accuracy of CT enhances with higher imaging speeds, which allows faster application of contrast media, thereby drastically improving contrast enhancement. The additional flexibility and speed of multidetector CT (MDCT) allows superior quality, thin-section imaging with 3D capabilities.

(c) Magnetic Resonance Imaging (MRI)

Another well used imaging for HCC diagnosis is MRI which uses similar concepts to those applied to CT imaging. MRI imaging utilizes magnetic field and energy of radio wave pulses to take pictures inside the body. In many cases the information that cannot be seen on an X-ray, US or CT scans is provided by MRI scans. Recent advancement in MR technology captures images within the time frame of one breath hold. However, MRI sensitivity when evaluating HCC tumors which are <2 cm in diameter, is far less than satisfactory.

(d) Magnetic Resonance Elastography (MRE)

The technique of MRE developed by Mayo Clinic is emerging to become a promising modality for HCC diagnosis. There is increasing evidence that HCC's behavior is mediated by its physical environment on a cellular level. Around 80% of HCCs is developed in the context of advanced liver fibrosis or cirrhosis where fibrosis is defined by changes in the physical and biochemical properties of the cellular microenvironment [10]. It is therefore obvious that development of HCC takes place in a niche with

mechanical properties different from those found in the normal liver. MRE is based on a transducer that transmits vibrations to the tissue that are of interest, a sequence of motion-encoding gradients to image tissue displacement and a mathematical reconstruction designed to reflect the mechanical properties of the tissue [11]. Therefore, there is growing interests in applying MRE for HCC diagnosis. MRE requires a wave driver to produce the mechanical excitation in the target tissues. It must be positioned as closely as possible to the target to be scanned on the body surface. During the scan the driver vibrates at controllable frequencies and amplitudes. Each cycle is synchronized with the motion encoding sequence that measures mechanical vibrations. An elastogram is calculated by the help of the mechanical parameters such as amplitude, frequency, wavelength and phase in the image volume [11]. Figure 1 below shows the workflow of MRE as just discussed.

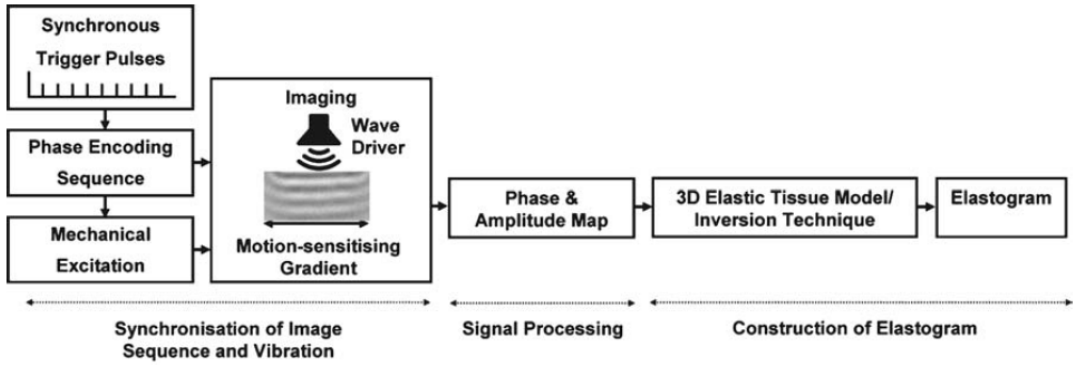


Figure 1: Workflow of MRE

(e) Summary on Imaging Techniques for HCC diagnosis

As each imaging modality is reviewed, we summarize the advantages and disadvantages of each modality in general medical practices in Table 1 below:

Table 1: Pros and Cons of Imaging techniques

Modality	Pros	Cons
US	▪ Inexpensive, quick and convenient. ▪ By the use of US no harmful effects have been observed, at the intensity levels that are being used for imaging and examinations [6].	▪ Images resolution is often limited. ▪ Cannot evaluate bone, lungs. ▪ Highly skilled specialists are required to interpret ultrasound images, especially when complicated procedures are involved [12]
CT	▪ CT can display bone changes much better than any other imaging methodology. ▪ CT can provide detailed images of bone, blood vessels and soft tissue. ▪ CT is widely available at healthcare institutions ▪ Reasonably priced [13]	▪ CT scans requires exposure to some radiation. This radiation exposure is likely to increase the risk of cancer in patients who are getting scanned. ▪ CT scan is not capable of identifying all diseases and medical conditions.
MRI	▪ MRI do not use ionized radiations and thereby avoiding patient's exposure to the potential harms of such radiation [12] ▪ MRI scans are useful for exhibiting soft tissue structures, such as cartilage, ligaments and organs such as the brain, eyes and heart [6]	▪ MRI scanners can be affected by movement therefore require patients to hold still for extended periods of time. ▪ People with pacemakers cannot have MRIs. [6]

MRE	▪ MRE measurements did not seem to be influenced by the presence of obesity or ascites ▪ MRE proved to be a more accurate and reproducible technique compared to conventional imaging features better diagnostic accuracy [11]	▪ The widespread use of MRE relies on the same technology and infrastructure as conventional MR imaging, and so claustrophobic or very obese patients and those with contraindications to MRI will be excluded from investigation due to an inability to enter the scanner [11]

Other than reviewing the technical pros and cons of each modality, we further explore the usage of each imaging technique in the medical practices. Specifically, we used "pubmed.gov", search from 2000-2013, using the keywords HCC and the related imaging modality. The results are shown in Figure 1(pie-chart) and Figure 2 (bar-graph). We observe:

(a) Ultrasonography (US) was most widely used for surveillance because it is not expensive, not invasive, well accepted by patients and can be repeated without risk. One of the major drawbacks if that it is difficult to distinguish small tumors from the nodularity of the cirrhotic liver by US.

(b) Computed tomography (CT) is a common imaging modality used in diagnosing HCC due to its widespread availability and short examination time, but the usefulness of CT in a surveillance program that requires periodical tests is limited by the patient's radiation exposure.

(c) Magnetic Resonance Imaging (MRI) has higher sensitivity than CT and US for HCC (89-100%) detection [14]. However, MRI is reserved for characterization

purposes, diagnostic confirmation and intrahepatic tumor staging because of its lower availability and high cost.

(d) MRI provides higher lesion-to liver contrast than CT, which is a significant advantage over CT.

(e) Several studies that have compared the accuracy of US, CT and MRI for HCC diagnosis show that MRI is better in the diagnosis of HCC when compared with CT and US. This is due to improved detection of small lesions 1–2 cm [15].

(f) In MR imaging, an emerging technique is Magnetic Resonance Elastography (MRE). Use of MRE has led to new quantitative tissue characterization parameters for differentiating benign and malignant hepatocellular nodules in a cirrhotic liver. MRE has proved to be a more accurate and reproducible technique compared to conventional imaging features better diagnostic accuracy. It is a relatively new technique and under considerations for improvement.

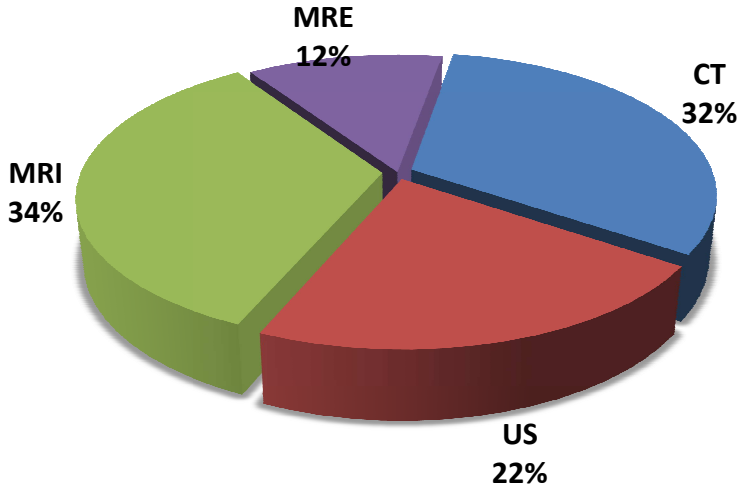


Figure 2: Use of imaging techniques in the HCC diagnosis during last ten years in medical field

It is evident from the pie-chart (Figure 2) that during last ten years CT and MRI imaging techniques have gained a lot of popularity in the medical field for detecting HCC. US used to be a popular diagnostic imaging technique earlier has been replaced by MRI and CT owing to high sensitivity and accuracy of the two techniques. MRE figures appear to be small as it is a relatively new technique and under considerations for improvement, but MRE has proved to be a more accurate and reproducible technique compared to conventional imaging features better diagnostic accuracy.

The following bar chart (Figure 3) is another representation of the trend followed in diagnosis of HCC using diagnostic imaging techniques. It is evident that US was a popular technique during early 2000 to 2004, but after 2005 CT and MRI has widely been used as the preferred imaging techniques. Also it can been seen that although MRE is fairly new imaging technique it has gained a significant amount of popularity during last few years and seems to be the future of the Diagnostic Imaging technique for detecting HCC.

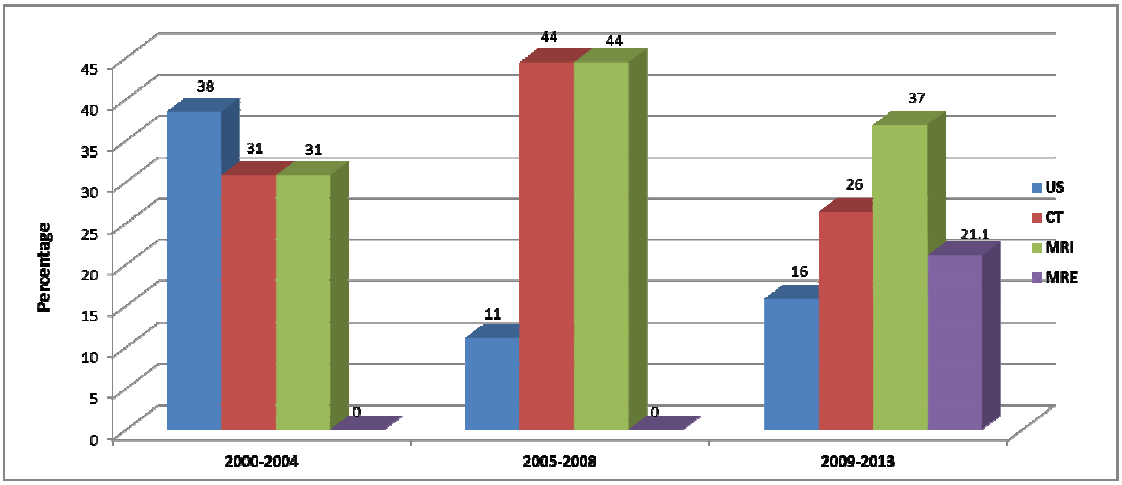


Figure 3: Year wise trend followed in diagnosis of HCC using diagnostic imaging techniques

1.3 Imaging Analytics

We have reviewed existing clinical literature on the usage of different modalities in HCC diagnosis. Other than the trend we observed, another interesting finding is that most clinical research has focused on only limited imaging biomarkers. In addition, most research published to date relies on subjective and variable assessment of imaging features. We believe Texture analysis is one potential technique to generate multiple, objective, reproducible, quantifiable features from medical images.

In the field of HCC, an early diagnosis will correspond to a more effective treatment if this can be done at an early phase. Surveillance aims to reduce disease-specific mortality by detecting HCC at a curable stage. The optimal profile for this endpoint is when the HCC is smaller than 2cm [16] while most conventional imaging techniques such as CT, MRI fails. Classification of healthy and diseased livers using modern imaging such as MRE in conjunction with Texture Analysis may address this clinical challenge [17].

Texture recognition is an important aspect of medical image analysis. In medical imaging, texture can be defined as the distinct image intensity pattern that is helpful in characterizing a tissue. Texture analysis also provides the local spectral or frequency content of an image; any change in the local texture should in turn cause changes in the local spatial frequency. Texture Analysis is of high importance in medical imaging analysis because, as the biological tissue becomes abnormal during a disease, its underlying texture could also change.

In Texture Analysis, the examined area of the sample (tissue) is represented by pixels. The intensity of the pixels is input information for classifying images because texture in an image refers to the distribution of brightness and darkness (gray tones) within the image. Texture Analysis evaluates the spatial location and signal intensity of each pixel in the examined area. There are a number of methods implemented for Texture Analysis. In general, we can divide them into four broad categories: statistical, structural, mathematical, transform based methods. These categories have been described in detail in the next chapter.

Texture Analysis has been successfully used for the separation of cirrhotic patients and healthy volunteers, and unknown patient data can be safely classified into the patient group using MRI, CT. While promising, the applicability of Texture Analysis on MRE images for HCC diagnosis is less studied. This is the focus of this thesis.

1.4 Research Objective

This research is to validate the applicability of MRE for HCC diagnosis. This study will in particular concentrate on quantitative aspect of MRE in conjunction with Texture

Analysis, an advanced imaging processing pipeline and multi-variate analysis machine learning method for accurate HCC diagnosis. To achieve this goal, we have acquired three patient data and three sets of experiments are conducted.

1.5 Thesis Organization

The overall thesis layout is shown in Figure 4. As seen, chapter 2 gives the in detailed description of the literature reviews with background and related work of Texture Analysis imaging techniques and briefs the initial work done and the challenges faced in the field of imaging technologies. This section also gives the in-depth knowledge about the current progress in Texture Analysis field and provides an in-sight of the future related works. Chapter 3 describes the datasets, experiments and results obtained. In this section the knowledge gained from previous chapters is applied in more practical manner. Here we do a retrospective comparison between the earlier techniques and the proposed techniques highlighting their pros and cons. Chapter 4 concludes the thesis with observations from current work with summary and giving a list of possible future work.

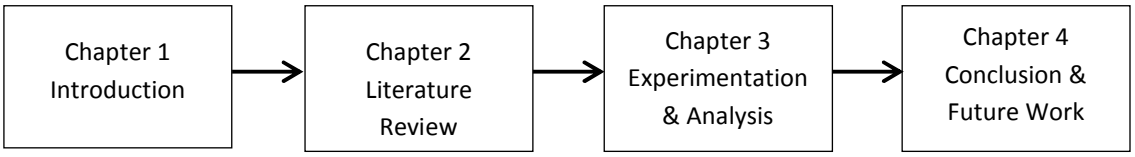


Figure 4: Thesis organization Flowchart

CHAPTER 2

LITERATURE REVIEW ON ADVANCED IMAGING ANALYTICS

2.1 Texture Analysis

Texture definition of an image is an important aspect in medical analysis. In medical imaging analysis, texture can be defined of as the local distinguish pattern of image intensity that helps radiologists in identifying a tissue. Texture analysis is important in medical image study primarily because there is always an underlying texture change associated with a tissue as soon as it becomes abnormal during a disease [18]. Current imaging techniques, such as MR, are not capable enough of providing microscopic information of tissue that can be assessed visually. However, tissue changes caused by some illnesses may bring about texture changes in the images that can easily be identified and quantified through texture analysis [19].

Thus, texture analysis is primarily a technique that evaluates the intensity and position of signal features, i.e. pixels, and also their corresponding grey level intensity in digital images. Texture features are therefore mathematical parameters that are calculated from the pixel distribution, which characterize the texture type and thus the underlying structure of the objects shown in the image [19].

Analyzing the texture parameters gives us highly useful information that is obtainable from medical images. In medical practices, the visual observation of imaging texture may be subjective. In addition, human observers may be able to observe only a limited part of the diagnostic information carried by an image. Texture analysis makes use of radiological images that are obtained in routine diagnostic practice, but involves

statistical analysis to be performed with the data contained within the images. As a result, it is becoming a tool that helps in reducing mistakes in clinical stage determination and also assists in unclear cases.

Texture Analysis provides innovation and diversity in many fields; it opens up a new era of analysis. During last three decades extensive research has been done on texture classification [20]. Few of the most popular texture approaches that were popular during 1980s, included gray level co-occurrence matrices (GLCM), Gauss–Markov random field, and local linear transform. These approaches had restriction that they could analyze the spatial relations between neighboring pixels for a small image region [20]. Texture analysis is most important for those cases in which change cannot be detected by direct inspection of the image. The approaches for analyzing texture are very diverse. We will review some common methods applied in the texture analysis in the following section.

2.1.1 Texture Analysis Pipeline

A computational pipeline is an integral part of Texture Analysis study in Medical Imaging. It was developed to combine texture analysis and pattern classification algorithms for investigating associations between high-resolution MRI / MRE features and clinical patient data, and also between MRI / MRE features and histological data [21]. A typical Pipeline design structure consists of three main stages i.e., Preprocessing, Feature extraction and Analysis. Figure 5 illustrates the pictorial representation of a medical imaging pipeline.

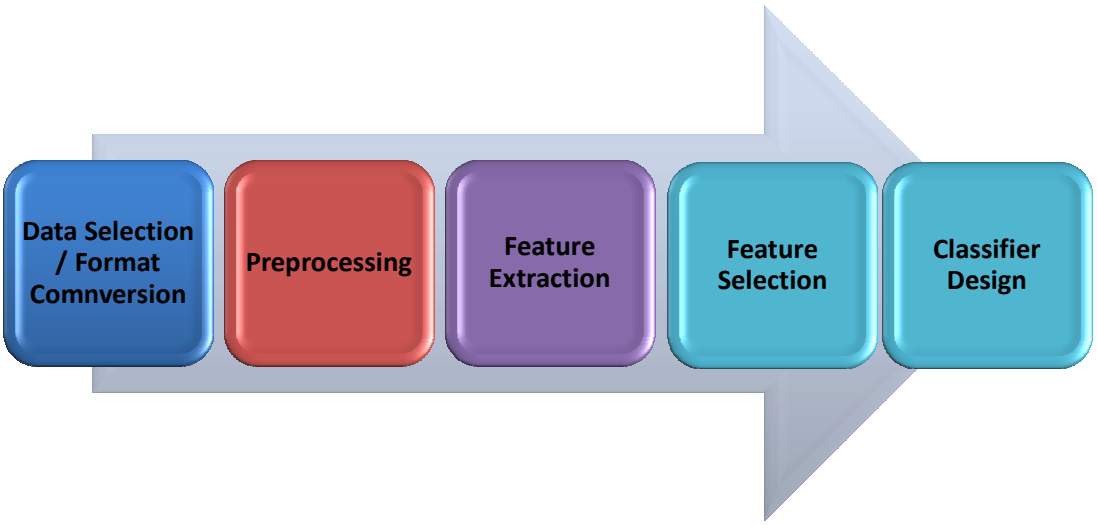


Figure 5: Texture Analysis Pipeline

(a) Preprocessing

Preprocessing stage is the first and the most important stage of the Texture Analysis Pipeline. The preprocessing stage reduces imaging artifact effects, such as noise and intensity non standardization. It is composed of three steps: noise filtering, Background segmentation and Intensity normalization [21]. In noise filtering the homogeneous areas of the images are smoothened while preserving sharp borders. Background segmentation separates the Region of Interest (ROI) from the surrounding area. Intensity Normalization methods helps in improving image compatibility. It reduces the variability introduced by different operators, different gain settings, and equipment variability and simplifies MR image comparability.

Figure 6: Data Preprocessing Steps

(b) Feature Extraction

The feature extraction stage calculates all the texture features used in tissue classification. The features can be extracted for a single voxel in the ROI as well as the mean of all the voxels in the entire ROI. It includes Statistical, Structural, Mathematical Model based and Transform based texture features. Statistical based methods are of Co-occurrence matrix type which includes 14 textural features or it can be Run length Matrix type which consists of 88 features for different angles. Similarly Structural based methods include features that are characterized by feature primitives and their spatial arrangements. Mathematical model based models consists of fractal models which generate empirical

models of each pixel in the image. Transform based features are wavelets totalizing 24 features

(c) Analysis

The first step in analysis is data scaling. This step improves the final accuracy in the classification process of the data analysis. Information gain is used as a measure in the feature selection for the model; it aims at finding the most relevant texture features for the image. After testing out several empirical tests, we chose only those texture features that have the highest information gain score as well as the highest percentage of correctly classified instances. Few classifiers that work very well for texture analysis feature selection are Support Vector Machines (SVM), Neural Networks (NN) and Ada Boost Classifiers. Parameters for each classifier are selected based on the problem being faced. Once the classifier is developed for problem it is tested on the testing data to justify the selection.

Figure 7: Data Analysis Steps

2.1.2 Texture Analysis Methodologies

The key categories of texture analysis methodology include [22]:

i) **Statistical Methods**, in which the texture is characterized by statistical distribution of intensity values Example of these methods, are Histogram, GLCM, and Run Length Matrix.

ii) **Structural Methods**, where the texture is characterized by feature primitives and their spatial arrangements

iii) **Mathematical model based Methods**, such as fractal models which usually generate an empirical model of all the pixels contained within that image considering the weighted average of the pixel intensities in its neighborhood.

iv) **Transform based Methods**, where the image is converted into new form using spatial frequency properties of the pixel intensity variations. Some examples of this method are Wavelet Transform, Fourier Transform and S transform.

Each of these methodologies has been briefly described as following:

(a) Statistical Methods

In statistical methods, texture is described by a collection of statistics of selected features. Statistical approach of texture analysis primarily describes texture of regions in an image using higher order moments of their grayscale histograms values [23]. Selecting various textural features from a Gray level co-occurrence matrix (GLCM) is apparently, the most commonly cited method for texture analysis [23]. In addition to the traditional statistical texture analysis methods, multivariate statistical techniques have also been considered for

extraction of textural features. If we consider an image as a matrix, the singular value decomposition (SVD) spectrum of the image texture is a summary vector represented by its singular values [23]. Alternatively, the run length matrix (RLM) includes higher-order statistics of the gray level histogram for an image. The RLM approach of texture analysis distinguishes fine textures of an image as having few pixels in a constant gray level run and coarse textures with many pixels in such a run [23].

(i) Histograms

In digital images, the allowed value for the grey level that can be given to a pixel is limited. The grey value is usually an integer ranging from 0 to 2b-1, where b denotes the number of bits of the image [22]. The histogram of an image is drawn by counting the number of pixels in the image that possess a given grey-level value. For example in a 12 bits image, the histogram may be represented by a graph, where the x-coordinates range from 0 to 4095 and y-coordinates represents the corresponding pixel count [22]. From the histogram many parameters may be derived, such as its mean, variance and percentiles.

(ii) Run Length Matrix

The run-length matrix is a technique where we search the image, always across a particular direction, for number of pixels that have the same grey-level value. Therefore, given a particular direction (for example, the vertical direction), the run-length matrix computes for each allowed grey-level value how many instances there are runs of, example, 2 consecutive pixels with the same grey-level value. Next it repeats the same for 3 consecutive pixels, then for 4, 5 and so on [22]. Thus using a single image, typically four matrices are generated, for the vertical, horizontal and two diagonal directions [22].

(iii) Haralick's co-occurrence matrix

The Haralick's co-occurrence matrix is a method that helps us to gather statistical information of an image or an image ROI based on distribution of pixels of that image. It is calculated by defining a direction and a distance i.e., the pairs of pixels separated by this distance. Once this has been done number of pairs of pixels is counted that contains a given distribution of grey-level values. Each entry in the matrix therefor corresponds to one similar grey-level distribution [22].

Co-occurrence matrix is a good way to describe shape by statistically sampling certain grey-levels in relation to other grey-levels. This matrix is square with dimension Ng, where Ng is the total number of gray levels in the image. The [i,j]th element of the matrix is produced by counting the total occasions a pixel with value i is adjacent to a pixel with value j and then subsequently dividing the whole matrix by the total number of such comparisons that are made. Each entry in the resulting matrix is considered as the probability that a pixel with value "i" is to be found that is adjacent to a pixel of value j.

$$\mathbf{G} = \begin{bmatrix} p(1,1) & p(1,2) & \cdots & p(1,N_g) \\ p(2,1) & p(2,2) & \cdots & p(2,N_g) \\ \vdots & \vdots & \ddots & \vdots \\ p(N_g,1) & p(N_g,2) & \cdots & p(N_g,N_g) \end{bmatrix}$$

Figure 8: General Format of Co-occurrence Matrix

(b) Structural Methods

This texture analysis technique characterizes a texture as the combination of well-defined texture elements such as regularly spaced parallel lines. The image texture is defined by the properties and placement rules of the texture elements. Different structural texture

analysis approaches have been recommended which ranges from utilizing different shapes of structuring elements to understanding real textures as distorted versions of ideal textures. However, as far as practical application of these methods is concerned, they are in limited use since they can only describe very regular textures [23].

(c) Mathematical Model Based Methods

In this approach of texture analysis a texture in an image is represented using sophisticated mathematical models (such as stochastic or fractal). The model parameters are estimated and used for the image analysis [22]. Mathematical model based texture analysis techniques generate an empirical model of each pixel in the image based on a weighted average of the pixel intensities in its neighborhood [23]. The disadvantage of these models is that the estimation of these parameters is computationally very complex [22]. The estimated parameters of the image models are used as textural feature descriptors. Examples of such model-based texture descriptors are autoregressive (AR) models, Markov random fields (MRF) and fractal models [23].

(i) Auto-Regressive Model

The auto-regressive model assumes a local interaction between image pixels in that the pixel grey level value is a weighted sum of the grey-level values of the neighboring pixels. The auto-regressive parameters are simply the set of weights used to establish these relations. It is expected that these relations are unique for a given type of object (or shape) in an image and, therefore, they may constitute a way of characterizing this object [22].

(ii) ***Fractal Model***

A fractal is a random geometric object that has an infinite nesting of structure at all scales. One can find fractal objects all around in nature, for example in coastlines, snowflakes, fern, mountains, trees, clouds and bacteria. Few of the most important properties of fractals are self-similarity, non-integer fractal dimension and chaos [24].

The Fractal Dimension (FD) has been used in detection of various biomedical applications such as brain tumor, breast tumor and lung tumor [25]. There are many fractal based algorithms that are being used in medical applications such as piecewise-threshold box-counting (PTBC), piecewise modified box-counting (PMBC), and piecewise-triangular prism surface-area (PTPSA) for detecting brain tumors in MR images based in 2D and 3D spaces [26].

The fractal dimension can be defined as the fraction of the number of self-similar pieces, N, to the magnification factor, 1/r, into which you split a figure. The equation for FD is as follows [27]:

$$FD = \frac{\ln\ (\text{number of self-similar pieces})}{\ln\ (\text{magnification factor})} = \frac{\ln N}{\ln(1/r)}$$

(d) Transform based Methods

Finally, the transform-based texture analysis method alters the image into a new form by using the spatial frequency properties of the pixel intensity variations. The success of these modern techniques is largely due to the type of transform they use to extract textural features from the image [23]. In this method the texture properties of an image

may be analyzed in the scale space or the frequency space. Transform based methods are based on the Gabor, Fourier, or Wavelet transforms [22].

(i) Wavelet Transform

The Wavelet transform is a spatial/frequency analytical tool which is being used extensively during past ten years and has been an area for research for many researchers. Wavelet transform is a traditional pyramid-type transform that decomposes signals to sub signals in low frequency channels [28]. However a drawback is that most significant information of a textured image often appears in the middle frequency channels therefore the conventional wavelet transform does not work properly in the texture context. To rectify this drawback, the transform is modified and an energy function is used to characterize the strength of a sub signal contained in a frequency channel requiring further decomposition. This idea leads formation of tree structured wavelet transform [28].

The methodology on which the wavelet transforms works is that it analyzes the frequency content in an image for different scales of that image. Therefore this analysis provides us with a set of wavelet coefficients corresponding to different scales and different frequency directions for that image. While calculating the wavelet transform of an image, each pixel is associated with a set of numbers known as the wavelet coefficients [29]. This wavelet coefficient represents the frequency content of the image at that point over a set of scales. These coefficients are used in computing the different texture parameters for the image. Figure 9 shows an example of a wavelet transform for the image shown in Figure 10. The top left corner of the image (Figure 9) depicts the low frequency and a

small-scale version of the original image. Whereas the other images in Figure 9 represents higher frequency versions of the original image but on different scales [22].

An example of a parameter derived from wavelet transform is the wavelet energy associated with a given scale and given direction. This parameter gives us the measure of the frequency content of the image on a given scale and in a given direction [22].

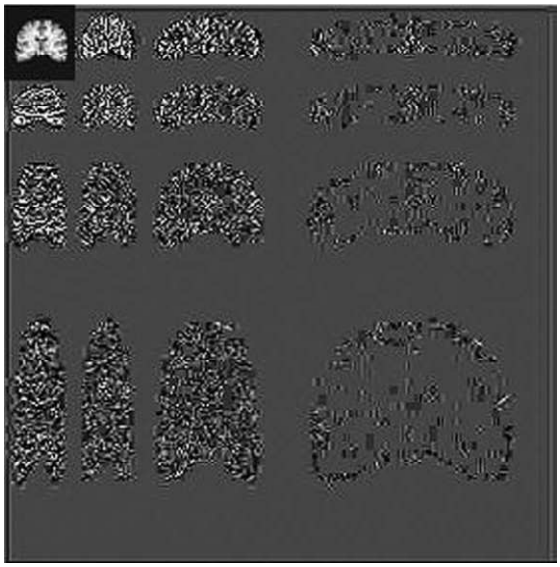


Figure 9: Wavelet transform of the image

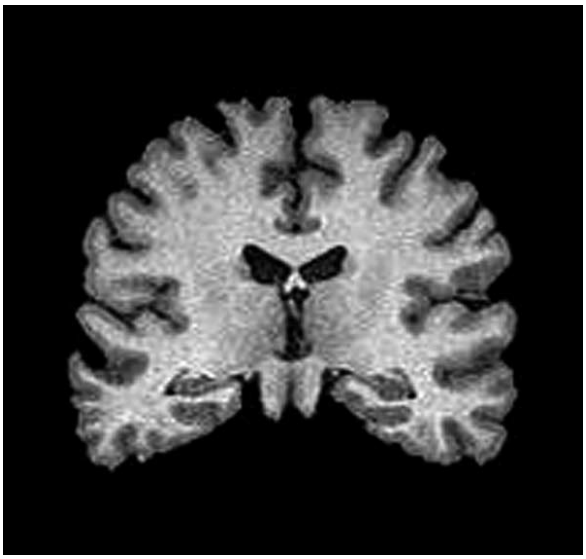


Figure 10: Coronal slice of T1-weighted cerebral MRI

Steps to construct two-dimensional Wavelet Packet Transform:

The wavelet transform provides us with an accurate and consolidated layout for the analysis and identification of a signal in an image at different scales. Wavelet Transform is often considered as a multiresolution analysis tool for the finite energy function. It can be implemented effectively alongside the wavelet packet transform and the pyramid-structured wavelet transform. Additional decomposition of a signal is performed by the pyramid-structured wavelet transform, in the low frequency regions. However, the decomposition of a signal in all low and high frequency regions is performed by wavelet packet transform [29].

The working of a wavelet transform starts with the decomposition of an image into sub images. The image is actually decomposed i.e., divided into four sub-bands and discrete wavelet transform is applied on and then it is critically sub-sampled as shown in Figure 11. These four sub images represent the frequency information of the original image in the frequency regions of LL, LH, HL, and HH respectively. Sub-bands labeled as LH1,

HL1 and HH1 represents the wavelet coefficients with finest scales amongst all the bands, whereas the sub-band LL1 corresponds to coarse level coefficients i.e., approximation image [20] .

The next step is decomposing the image further into the next coarse level of wavelet coefficients. The sub-band LL1 alone is used for further decomposition and critically sampled. This results in two-level wavelet decomposition. The process continues until some final scale is reached. The 2-D wavelet packet transform achieves a full decomposition by disintegrating all the frequency regions, as shown in Figure 12 [20].

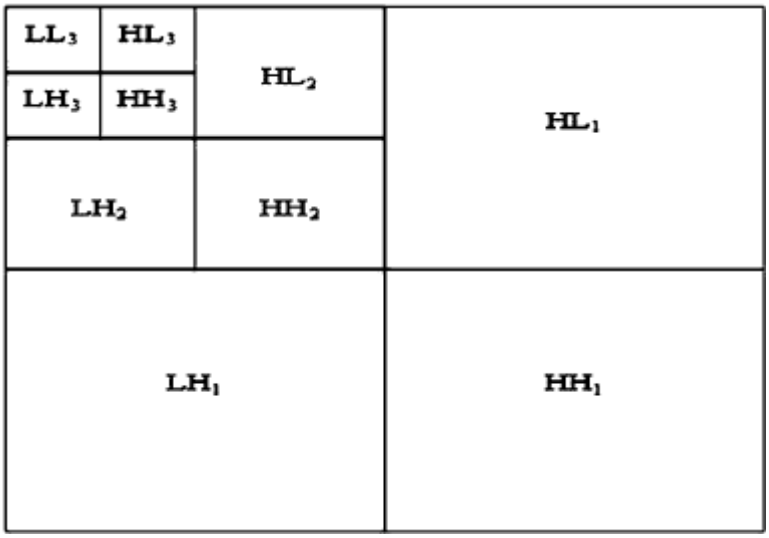


Figure 11: Three level 2-D PSWT decomposition of 128 x 128 image

LL_3	HL_3	LL_3	HL_3	LL_3	HL_3	LL_3	HL_3
LH_3	HH_3	LH_3	HH_3	LH_3	HH_3	LH_3	HH_3
LL_3	HL_3	LL_3	HL_3	LL_3	HL_3	LL_3	HL_3
LH_3	HH_3	LH_3	HH_3	LH_3	HH_3	LH_3	HH_3
LL_3	HL_3	LL_3	HL_3	LL_3	HL_3	LL_3	HL_3
LH_3	HH_3	LH_3	HH_3	LH_3	HH_3	LH_3	HH_3
LL_3	HL_3	LL_3	HL_3	LL_3	HL_3	LL_3	HL_3
LH_3	HH_3	LH_3	LH_3	HH_3	HH_3	LH_3	HH_3

Figure 12: Level 2-D wavelet packet decomposition of 128 x 128 image

Further after the decomposition the steps involved in the texture classification is shown in the Figures 13 and 14 [29].

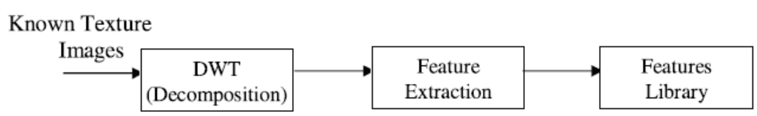


Figure 13: Texture training steps

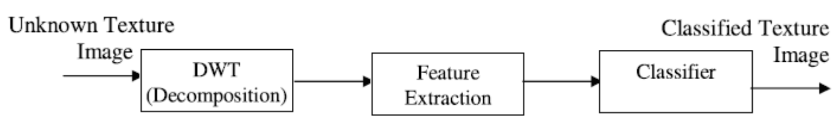


Figure 14: Texture classification steps

(ii) S-Transform

The S-transform (ST) relates closely to the continuous wavelet transform as it uses the complex Morlet mother wavelet and therefore it measures directly the local frequency composition in an image for each and every pixel. The S-transform has been successful in analyzing signals in various applications, such as ground vibrations, seismic recordings, gravitational waves, power system analysis and hydrology. The 1D S-transform has proved to be a useful tool for analyzing the medical signals, such as laser Doppler flowmetry, EEG and functional magnetic resonance imaging. The S-transform works satisfactorily for texture analysis of images in medical industry due to its optimum space-frequency resolution and close connection to the Fourier transform (FT).

The main obstacle of the S-transform algorithm in wider application of S-transform - based texture analysis for 2D images has been its redundant nature. In order to calculate and store the texture features of large medical images, extensive calculation time and a large memory space are required [18].

As a result, the S-transform of a 256×256 MR image takes almost one and half hours to calculate on one computer with memory requirements of almost 32 GB [18]. Therefore, previously the work that was done on 2D images had its limitation for analysis of only small ROIs and disintegrated to 1D spectrum. However because of the small ROIs, the resolution of the frequency spectra is reduced, and thereby reducing the sensitivity to complex texture changes. These shortcomings make the application of 2D-S-Transform to clinical medical applications not even difficult but also very impractical [18]. In modern era, the clinical texture analysis method requires an efficient algorithm that is

capable of providing complete information about all frequency components. However, even though there are few limitations, the 2D-S-Transform has demonstrated promising results in identification of differences in textures that is associated with neurological pathology [18].

(iii) Discrete Orthonormal S Transform (DOST)

The Discrete Orthonormal space-frequency transform (DOST) is a relatively new and effective approach for describing an image texture [18]. In order to obtain a rotationally consistent set of texture features, the DOST components can be combined together, which in turn accurately distinguishes between a series of texture patterns [18]. The DOST is highly efficient as it provides the multi-scale information and computational efficiency of wavelet transforms, when it provides the texture features as Fourier frequencies. It is better than other leading wavelet-based texture analysis techniques and is more efficient as compared to primitive Haralick's Co-occurrence Matrix [18].

One of the biggest advantages of DOST is that it speeds up the calculation of the S-Transform and eliminates the redundant nature of the space frequency domain. The DOST provides a spatial frequency illustration that is similar to the discrete wavelet transform. Along with all these advantages, the DOST has the additional benefits as well. It maintains the phase properties of the S- Transform and Fourier Transform and even maintains the ability to crash exactly back to the Fourier domain [18].

The computational accuracy of DOST is fast and straightforward. It allows us to analyze each and every pixel of an image within seconds. The DOST did very well in detecting the even a small change in contrast and spatial frequency when subjected to changes in

frequency domains at various levels of noise and evaluating against series of tailor-made images which have known frequency content and added noise. The DOST is robust to the presence of low or moderate noise levels. DOST is highly accurate in identifying single frequency components from the local spectra [18].

(iv) Fast Time Frequency Transform (FTFT)

FTFT is a method that is developed by Chun Hing Cheng and Ross Mitchell from Mayo Clinic. It is a fast and accurate way to generate a highly compressed form of the values of S Transform directly. It is used when N is so large that we cannot find and store the ST values first. It encodes the time frequency representation (TRF) information uniformly and so can then be used for analyzing the TRF correctly and processing the data efficiently and effectively. The compression that FTFT provides can help storage, transmission and visualization of S Transform. Using FTFT the values of S Transform can be calculated at individual points, called local spectra, instantaneously and accurately. This is useful for real-time monitoring, control, manipulation and filtering. This method is memory-efficient, robust and adaptive.

2.2 Machine Learning Algorithms used

For analyzing the data for the three HCC patients following machine learning algorithms were studies and models were developed using them:

(a) Decision Trees (J48 & Random Forest)

Decision tree is a simple yet widely used classification technique. They follow a nonparametric approach for classification models building. In other words, it does not require any previous assumptions regarding the type of probability distributions that the class and other attributes should satisfy. In a decision tree, every leaf node has an assigned class label. Attribute test conditions are used to separate records having different characteristics in the non-terminal nodes, which consist of the root node and other internal nodes. Decision trees, especially smaller-sized trees are relatively easier to interpret. They are quite robust to the presence of noise, especially when methods for avoiding over fitting. The accuracy of a decision trees is not adversely affected by the presence of redundant attributes.

(b) ADA Boost:

Ada Boost is an iterative technique that adaptively changes the distribution of training samples which helps the base classifiers to concentrate on examples that are difficult to classify. Ada boost algorithm assigns equal weights to all instances at the beginning in the training data. It then recalls the learning algorithm to develop a classifier for this data and then reweights each instance in according to

the classifier output. Therefore the weight of instances that were correctly classified is decreased and that of misclassified ones in increased.

(c) Bagging:

Bagging is also known as bootstrap aggregating. It is a technique that repeatedly samples from a dataset, with replacement, in accordance with uniform probability distribution. Every bootstrap sample has the same size as the original dataset. As we see that the sampling is done with replacement, therefore some of the instances might appear more than once in the same training set, while others might get eliminated from the training set. Bagging is a technique that improves on the generalization error by reduction in variance of the base classifiers. The stability of the base classifier decides the performance of bagging method. Bagging does not focus on any particular instance of the training data. This is due to the fact that every sample has an equal probability of getting selected. It is therefore less affected to over-fitting the model when applied to a noisy data.

(d) Support Vector Machines (SVM):

A classification technique that has received considerable attention is Support vector machine (SVM). This technique has originated from the statistical learning theory. SVM and has shown promising results in many practical applications. SVM works well with high-dimensional data and is not affected by the dimensionality problem. SVM performs capacity control by maximizing the margin of the decision boundary. Nevertheless, the user must still provide other parameters such as the type of kernel function to use and the cost function C for

introducing each slack variable. SVM works well for binary class categorical indicators.

(e) ***Artificial Neural Network (ANN):***

The study of artificial neural network (ANN) got its inspiration from simulation models on biological neural systems. Similar to human brain structure, an ANN comprises of an interconnected network of nodes and directed links. Multilayer neural networks with at least one hidden layer are universal approximators, i.e., they can be used to approximate any target functions. ANN can handle redundant features because the weights are automatically learned during training step. The disadvantage of ANN is that they are sensitive to the presence of noise in the training data and also they are a time consuming process, especially when the number of hidden nodes is large.

2.3 Feature Selection Techniques used

Feature selection is an important step in data analysis particularly when you are handling large number of features. Usually most of these features are irrelevant to the classification analysis and also dealing with large feature sets slows down the algorithms. In this experiment, significant features were selected for the dataset with two goals of attaining highest accuracy and selecting smallest set of features. Two feature selection techniques were used to determine the useful features in the dataset:

(i) Best First Technique

Best first is a selection technique that combines both forward selection and backward elimination rules. It is a method that does not just terminate when the performance starts to drop but keeps a list of all attribute subsets evaluated so far, sorted in order of the performance measure, so that it can revisit an earlier configuration instead. Given enough time it will explore the entire space, unless this is prevented by some kind of stopping criterion. It can search forward from the empty set of attributes, backward from the full set, or start at an intermediate point and searches in both the directions by considering all possible single-attribute additions and deletions.

(ii) Greedy Stepwise Technique

Greedy stepwise searches greedily through the space of attribute subsets. Like best first technique it may progress forward from the empty set or backward from the full set. Unlike best first technique, it does not backtrack but terminates as

soon as adding or deleting the best remaining attribute decreases the evaluation metric.

2.4 Resampling Techniques used

Resampling techniques were used since the two classes of the response were not balanced properly. There was a big imbalance between the two classes. There is a tendency that standard classifiers tend to be biased by the majority class and therefore they ignore the minority class. Particularly they tend to produce high predictive accuracy over majority class, but poor predictive accuracy over minority class. Resampling the classes generates a balanced dataset. For class balancing we used following resampling techniques:

- **Oversampling:** Oversampling is a method that balances the data-set by increasing the number of minority class instances .Oversampling generates new instances based on the values of known samples and thereby increasing the frequency of samples. This results in increase in number of instances for the minority class.

- **Under sampling:** Under sampling method extracts a smaller set of majority class while preserving all the minority classes.

- **SMOTE:** In SMOTE, we over sample the minority class by taking each minority class sample and inducing fabricated/fake examples along the line segments that joins any/all of the minority class k-nearest neighbors. Depending on the number of oversamples required, we randomly choose neighbors from the k-nearest neighbors [30]. An example for SMOTE can be seen in the Figure 15 below. Let X is the point selected in the minority class and let X_1 to X_4 is the selected nearest

neighbors to X. therefore R1 to R4 are the fabricated points that are created by random interpolation. We have used Euclidean distance to select the neighbors to the point X. in short what we just did in SMOTE is forming a new minority class examples by interpolating between several minority class examples that were in the neighbor to the selected point.

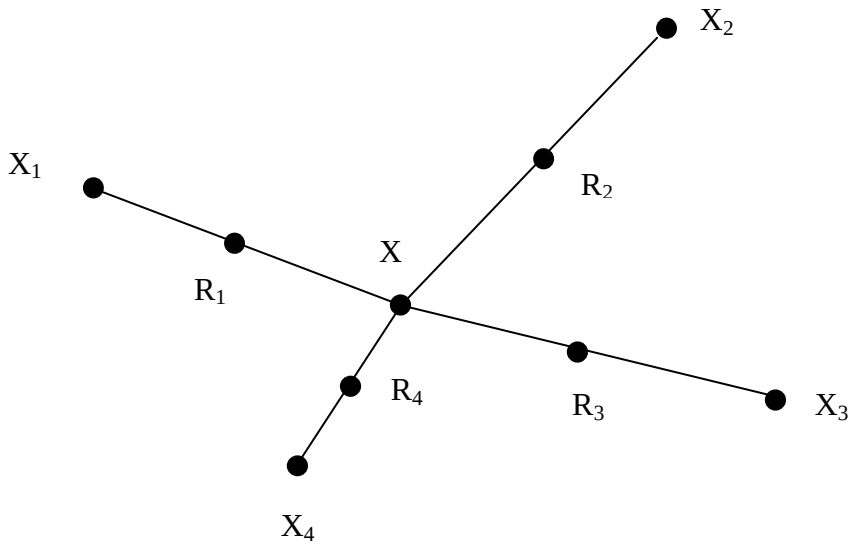


Figure 15: An illustration on how to create the fabricated data points in SMOTE

2.5 Data Preprocessing and Data Cleaning

One of the most important steps of model building is preprocessing of the dataset. For data preprocessing, a good understanding of dataset is very important. Data pre-processing consists of following steps: data cleaning, transformation, normalization, feature extraction and feature selection. The data preprocessing step is considered to be an important step as it can have a significant impact on how a supervised machine learning algorithm performs.

Data cleaning step involves handling the Missing values in the dataset, aligning the data into proper format, fixing the errors and outliers. Initial analysis of the dataset did not show presence of any missing values or outliers. Data alignment was done since the data collection activity was done on different dates and had to be aligned together in a common format.

2.6 Summary and Conclusion

After reviewing the existing techniques for Texture analysis we conclude:

1. Statistical Texture Analysis techniques are probably the most common and cited technique for Texture Analysis. Statistical methods vary from Grey level co-occurrence matrix (GLCM), developed in 1980's, and to more recent approaches based on multivariate statistical methods for textural feature extraction.
2. Structural texture analysis techniques characterize a texture as the combination of well-defined texture elements such as regularly spaced parallel lines. However, these methods appear to be limited in practical use since they can only describe very regular textures, making them limited in use and defined to certain class of textural patterns.
3. Mathematical model based methods for texture analysis attempts to represents texture in an image using sophisticated mathematical models. The disadvantage of this technique is that the computational complexity involved in the estimation of these parameters is large, which makes these models difficult to interpret.
4. Transform-based texture analysis method alters the image into a new form by using the spatial frequency properties of the pixel intensity variations. These

methods are based on the Fourier, Wavelet, Gabor and S-transforms. Their biggest advantage is that these methods can easily be adjusted to the problem in question making them most widely used texture analysis technique.

Based on the above observations we decided to work on Transform based Texture analysis technique FTFT-RIST which is based on S-Transform. As this is a new method that has been developed my Mayo Clinic and needs to be tested in terms of accuracy and computational speed.

CHAPTER 3

EXPERIMENTS AND ANALYSIS

3.1 Introduction

This chapter gives the in depth analysis and the experimentation for building the model, evaluating it and finally testing the prediction power and robustness. In this chapter firstly we would present the general workflow of the experiments that were conducted in this research. Data preprocessing and data cleaning were the next steps in analysis. Thereafter describing the datasets being used, what follows is the series of experiments that were conducted along with their conclusions.

3.2 General Workflow of the Experiments

MRE is known to be a promising imaging technique for HCC diagnosis. This research is to validate the applicability of MRE in conjunction with Texture Analysis pipeline. The study starts with drawing the ROIs, both tumor as well as Non tumor areas, on the MRE images with the help of the Radiologists. These images are tested on the Texture Analysis pipeline to present us with two types of dataset- pixel based and ROI based. These output datasets from the Texture Analysis pipeline are then subjected to data preprocessing where different sampling methods (undersampling, oversampling and SMOTE) are applied to get class balance in the dataset. Next various machine learning algorithms are applied on the data to develop a model, validate a model and final test the accuracy of the developed model. Major focus of this study has been concentrated on building the model, choosing the correct classifier to build a model, model validation and finally testing the build model. The following Figure 16 represents the general workflow of the study.

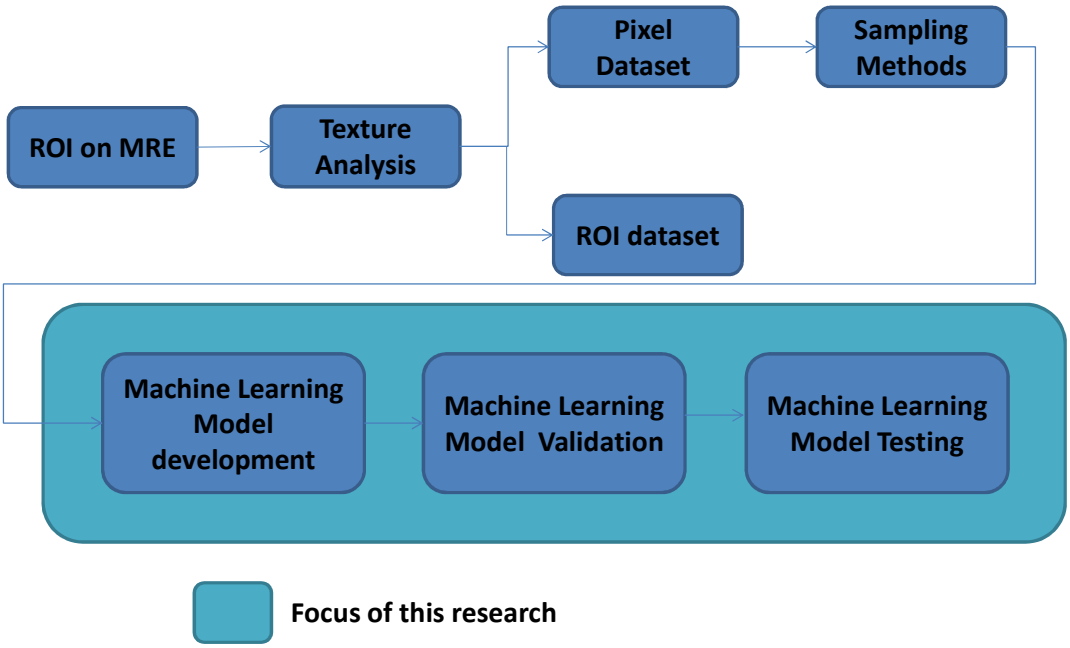
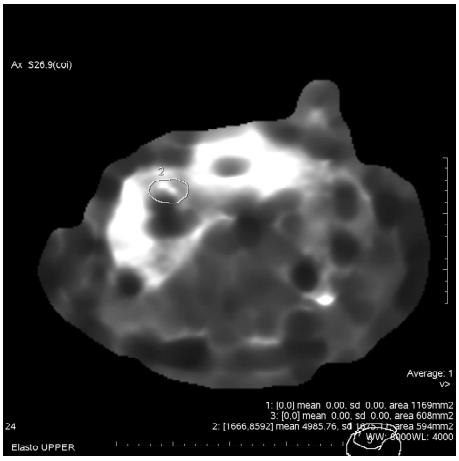


Figure 16: General Workflow of the Study

3.3 Dataset Description

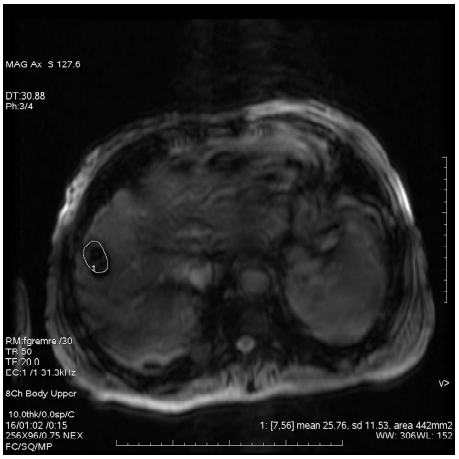
The dataset is derived from the MRE images for three patients suffering from Liver HCC. For Patients 1 & 3 we collected data from 1 ROI on tumor region and 2 ROIs on Normal regions for each patient. Patient 2 had two tumor regions on the Liver so we gathered the data from 2 ROIs on tumor region and 4 ROIs on Normal region.

Tumor ROIs

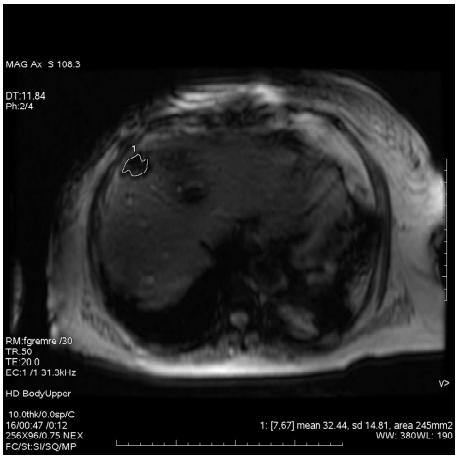


Patient 1

Patient 2 ROI 1



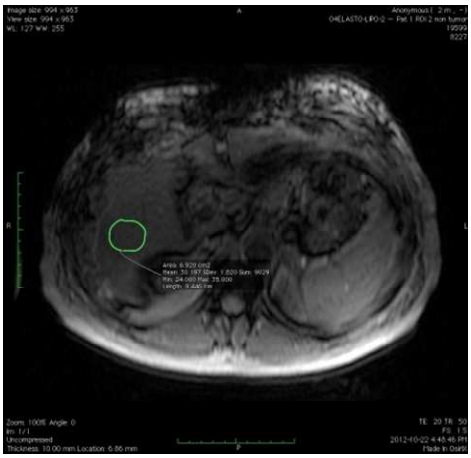
Patient 2 ROI 2



Patient 3

NON Tumor ROIs

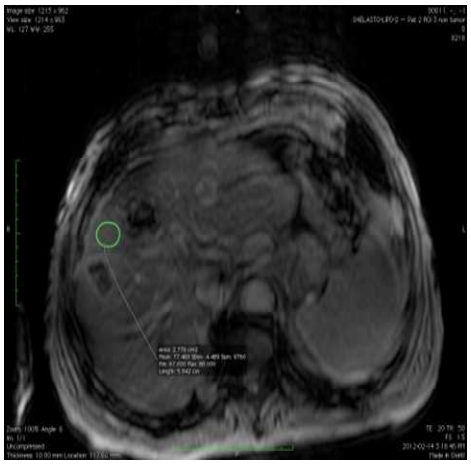
Patient 1 ROI 1



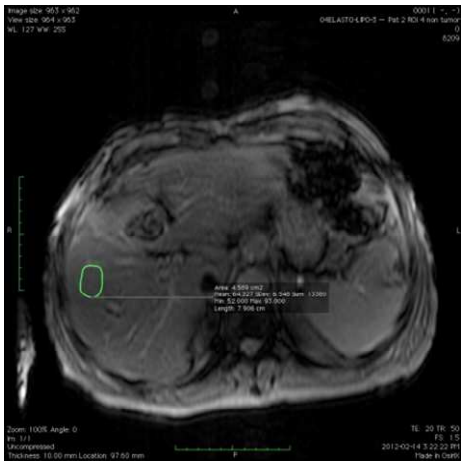
Patient 1 ROI 2

Patient 2 ROI 1

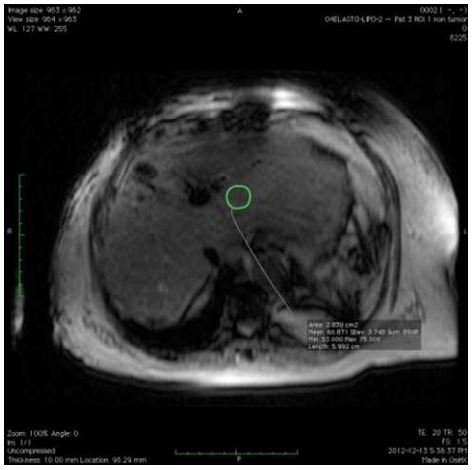
Patient 2 ROI 2



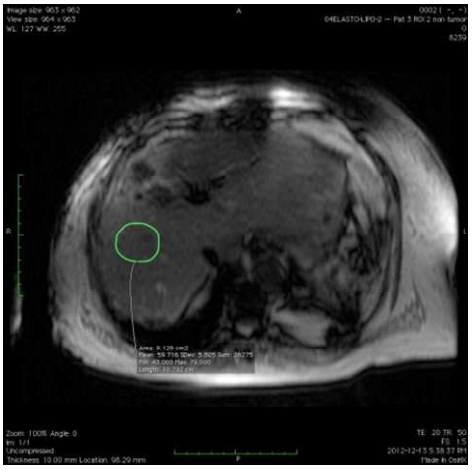
Patient 2 ROI 3



Patient 2 ROI 4



Patient 3 ROI 1



Patient 3 ROI 2

Following steps were followed to derive the data for the three patients liver cases:

Step 1: With the help of Radiologists drawing the Tumor as well as Non tumor ROIs on the Liver MRE images. This was done using Osirix, version "v4.1.2 32-bits". Save the images as Dicom file (.dcm).

Step 2: Convert the Images from Dicom files (.dcm) to XML files (.xml) that can be easily read by the software while converting them to Mask.

Step 3: Running the XML files of the images on "XML to Mask" Software to convert them to Mask before running them into FTFT-RIST algorithm.

Step 4: Final Step of deriving the data out of the MRE images is running the Mask images, obtained in the previous step, onto the FTFT-RIST algorithm of texture analysis.

We obtained two different datasets for a single image from FTFT-RIST algorithm. The first dataset is *ROI based dataset*, wherein the values of the features are the mean values of the all the pixels contained in the ROI that is selected on the image while running the image on the algorithm. Second dataset is the *Pixel based dataset* that provide us with feature values for every pixel contained within that ROI. Both the datasets were used while developing the classifier for the cases.

3.3.1 Pixel Based Dataset

The dataset consists of 19 variables and a binary Response variable. The response variable has values C0 and C1 representing two classes. C0 indicates class for Non tumor region whereas C1 indicates class with Tumor Region. Total 2718 instances are there in the dataset that cover both tumors as well as non-tumor regions for all the three patients. Table 2 lists down all the features in the Pixel Based dataset.

Table 2: Variables in Pixel based dataset

Variable Name	Variable Type	Description
X	Categorical	X coordinate of Pixel
Y	Categorical	Y coordinate of Pixel
Raw Signal	Categorical	Intensity value of the pixel
Band 1 Mean RIST	Numerical	It is the mean of the value in the RIST curve over the 1st frequency band
Band 2 Mean RIST	Numerical	It is the mean of the value in the RIST curve over the 2nd frequency band
Band 3 Mean RIST	Numerical	It is the mean of the value in the RIST curve over the 3rd frequency band
Band4 Mean RIST	Numerical	It is the mean of the value in the RIST curve over the 4th frequency band
Band5 Mean RIST	Numerical	It is the mean of the value in the RIST curve over the 5th frequency band
RIST Sum	Numerical	It is the sum of RIST values over all the pixels in the entire semicircular region in the RIST diagram
Mean of k in RIST Curve	Numerical	Measures the average frequency of texture at the pixel
SD of k in RIST Curve	Numerical	Measures the spread of frequency contents at the pixel
Semicircle RIST Homogeneity	Numerical	Measures the homogeneity as an average of squared wavelength in the texture.
Cell RIST Entropy	Numerical	Measures the randomness of the RIST values over the entire semicircular region in the RIST diagram
SD of RIST Sums by Sectors	Numerical	Measures the spread of RIST by sectors, i.e. by angles
SD of RIST Sums by Semicircles	Numerical	Measures the spread of RIST by semicircles, i.e. by radius.
SD of RIST Sums by Cells	Numerical	Measures the spread of RIST by cells, i.e. by angle and radius together.
RIST Sum in Major Sector	Numerical	The sum of RIST values over all the pixels in the major sector in the RIST diagram
Mean of k in Major RIST Curve	Numerical	Measures the average frequency of texture at the pixel in the direction of the major sector
SD of k in Major RIST Curve	Numerical	It measures the spread of frequency contents at the pixel in the direction of the major sector
Y (Response)	Categorical	Response that tells whether the tissue is tumor or non-tumor

3.3.2 ROI Based Dataset

The dataset consists of 68 variables and a binary Response variable. The response variable is same as in Pixel based dataset and has values C0 and C1 representing two classes. C0 indicating class for Non tumor region whereas C1 indicating class with

Tumor Region. Total 12 instances are there in the dataset that cover both tumor as well as non-tumor ROIs for all the three patients. Table 3 lists down all the features in the ROI Based dataset.

Table 3: Variables in ROI based dataset

Variable Name	Variable Type	Variable Name	Variable Type	Variable Name	Variable Type
Raw Signal Mean	Categorical	Band 5 Mean RIST Max	Numerical	SD of RIST sums by Sectors Min	Numerical
Raw Signal SD	Categorical	RIST Sum Mean	Numerical	SD of RIST sums by Sectors Max	Numerical
Raw Signal Min	Categorical	RIST Sum SD	Numerical	SD of RIST sums by Semicircles Mean	Numerical
Raw Signal Max	Categorical	RIST Sum Min	Numerical	SD of RIST sums by Semicircles SD	Numerical
Band 1 Mean RIST Mean	Numerical	RIST Sum Max	Numerical	SD of RIST sums by Semicircles Min	Numerical
Band 1 Mean RIST SD	Numerical	Mean of k in RIST curve Mean	Numerical	SD of RIST sums by Semicircles Max	Numerical
Band 1 Mean RIST Min	Numerical	Mean of k in RIST curve SD	Numerical	SD of RIST sums by cells Mean	Numerical
Band 1 Mean RIST Max	Numerical	Mean of k in RIST curve Min	Numerical	SD of RIST sums by cells SD	Numerical
Band 2 Mean RIST Mean	Numerical	Mean of k in RIST curve Max	Numerical	SD of RIST sums by cells Min	Numerical
Band 2 Mean RIST SD	Numerical	SD of k in RIST curve Mean	Numerical	SD of RIST sums by cells Max	Numerical
Band 2 Mean RIST Min	Numerical	SD of k in RIST curve SD	Numerical	RIST sum in major sector Mean	Numerical
Band 2 Mean RIST Max	Numerical	SD of k in RIST curve Min	Numerical	RIST sum in major sector SD	Numerical
Band 3 Mean RIST Mean	Numerical	SD of k in RIST curve Max	Numerical	RIST sum in major sector Min	Numerical
Band 3 Mean RIST SD	Numerical	Semicircle RIST homogeneity Mean	Numerical	RIST sum in major sector Max	Numerical
Band 3 Mean RIST Min	Numerical	Semicircle RIST homogeneity SD	Numerical	Mean of k in major RIST curve Mean	Numerical
Band 3 Mean RIST Max	Numerical	Semicircle RIST homogeneity Min	Numerical	Mean of k in major RIST curve SD	Numerical
Band 4 Mean RIST Mean	Numerical	Semicircle RIST homogeneity Max	Numerical	Mean of k in major RIST curve Min	Numerical
Band 4 Mean RIST SD	Numerical	Cell RIST entropy Mean	Numerical	Mean of k in major RIST curve Max	Numerical
Band 4 Mean RIST Min	Numerical	Cell RIST entropy SD	Numerical	SD of k in major RIST curve Mean	Numerical
Band 4 Mean RIST Max	Numerical	Cell RIST entropy Min	Numerical	SD of k in major RIST curve SD	Numerical
Band 5 Mean RIST Mean	Numerical	Cell RIST entropy Max	Numerical	SD of k in major RIST curve Min	Numerical
Band 5 Mean RIST SD	Numerical	SD of RIST sums by Sectors Mean	Numerical	SD of k in major RIST curve Max	Numerical
Band 5 Mean RIST Min	Numerical	SD of RIST sum by Sectors SD	Numerical	Y (Response)	Categorical

3.4 Experiments

We conducted three distinct experiments with the dataset for the patients.

Experiment 1: The objective of Experiment 1 is to quantitatively assess the diagnosis performance using Univariate approach on intensity information only. Specifically, the goal is to determine that intensity feature alone can distinguish comprehensively between the tumor region vs a non-tumor region for a HCC patient.

Experiment 2: The objective of Experiment 2 is to analyze the multiple features collected from Texture Analysis and quantitatively evaluate the diagnosis performance using ROI based data feature and pixel based data features. In addition to this another objective of this experiment would be to determine significant features and evaluate model performance on these features. This experiment contributes to the model building phase of the thesis study and concentrates on developing the classifier on WEKA and cross validating it using 10 folds cross validation.

Experiment 3: The objective of Experiment 3 is to evaluate the prediction power and the robustness of the developed model in HCC tissue characterization. This experiment contributes to the model testing phase of the thesis study.

3.4.1 Experiment 1: Intensity Based Analysis

In this experiment we have a bivariate dataset i.e., intensity values and response (C0 or C1), therefore we decided to analyze the data using Paired t-test and Scatter plot to determine that whether or not we draw concrete conclusions based on intensity value alone.

(a) Experiment 1.1: Paired t-test

The paired t-test is a common method of analysis of data that are paired and is based on the assumption that the differences between the paired observations are normally distributed. With paired observations, in which each value in one sample has a corresponding observation in the other sample, the task of comparing two samples can be simplified by making it a one-sample test, wherein the differences in each pair of observations constitute the newly formed sample. Thus, the mean value is the mean of the differences and the standard deviation represents the variability of the differences.

Let X and Y be the samples drawn from a population, then the test statistic (t-score) is defined by the equation:

$$t = \frac{[\overline{x} - \overline{y}]}{SE}$$

where SE, the standard error of the sampling distribution is computed as:

$$SE = \sqrt{[(\frac{s_x^2}{n}) + (\frac{s_y^2}{m})]}$$

x and y are the sample means, n and m are the sample sizes, and S_x and S_y are the sample standard deviations of X and Y respectively.

A paired t-test was conducted on the intensity values of both tumor and non-tumor region values for all 12 ROIs to check whether or not they are statistically different.

Hypothesis: H_0: Mean (Tumor) = Mean (Non tumor)

H_1: Means are different

Paired t-test results are shown in Table 4 as follows:

Table 4: Paired t-test results for intensity values

	N	Mean	SD	SE Mean
Tumor	679	47.984	20.058	0.77
Non Tumor	679	57.962	15.503	0.595
Difference	679	-9.978	24.026	0.922

At 95% for mean difference the CI was: (-11.788, -8.168)

Based on Paired t-test analysis we conclude that the intensity information from tumor and non-tumor tissues differs significantly, as CI does not include zero value in it. However,

1. The intensity information from tumor and non-tumor tissues differs significantly, as the CI does not include zero.
2. Paired t-test is population based test. If we are looking at each pixel, as observed from the scatter plot, for some regions we are able to tell the differences, for other regions we are not able to differentiate the pixels from tumor vs. non-tumor.
3. We need a method that can predict the diagnosis on the pixel based level for specific ROIs and whole liver.

(b) Experiment 1.2: Scatter Plot

A scatter plot also known as X-Y plot is one of the most useful and popular techniques for analyzing the data visually and exploring it. Analyzing a scatter plot we can identify the relationship between two attributes, cluster of points and outliers [31]. Such a relationship manifests themselves by any non-random structure in the plot.

The Figure17 represents the scatter plot of the intensity value vs tumor/non-tumor regions.

Red: Tumor Blue: Non tumor

Figure 17: Scatter plot of the Intensity values for tumor and non-tumor tissues

Based on the experiments we observe:

1. Scatter plot highlights some key areas of the intensity based study and therefore few good interpretations can be made by just analyzing the scatter plot of intensity values.

2. It is evident from the scatter plot that tumor is present in the liver for the intensity values less than 23 and greater than 100.

3. From intensity values between 43 and 48 presence of Tumor region is more likely than Non tumor region.

4. High number of Non tumor pixels lies between intensity values of 50 to 70.

5. At all the other intensity values it is equi-likely for a pixel to be tumor or Non tumor.

6. Although visually analyzing the scatter plots gave first impression that the data set is segregated in few regions as compared to other, a statistical test is needed in order to conclude that the intensity values are statistically different. A paired t-test was therefore done in order to confirm the results.

3.4.2 Experiment 2: Texture Analysis features based study

We have divided this experiment into two sections-*Pixel based study and ROI based study*. Each section has two experiments individually. First experiment of each section is developing a model using all the features of the dataset and second experiment is building the model with selected significant features. These experiments are done for both pixels based study as well as ROI based study.

(a) Experiment 2.1: Pixel Based Study-All features

We conducted pixel based classification analysis using all the 19 features. Ten fold cross-validation technique was used to validate the model. Cross-Validation is a statistical method of evaluating and comparing learning algorithms by dividing data into two segments: one used to learn or train a model and the other used for validating the model. In cross-validation technique, the training and validation set gets interchanged in successive rounds so that each data point gets a chance of being validated against each other. Following classifiers were used to build the model on the given dataset.

(a) Decision Trees (J48 & Random Forest)

(b) ADA Boost

(c) Bagging

(d) Support Vector Machines (SVM)

(e) Artificial Neural Network (ANN)

All the statistical classifier methods employed were able to differentiate between Tumor and Non Tumor regions for the patients with high accuracy. The table 5 summarizes the

% of correctly classified instances for different classifiers along with the resampling technique used on the dataset.

Table 5: Percentage of correctly classified instances for experiment 2.1

Classification Algorithm	Oversampling	Under sampling	SMOTE
J48	99.88	99.62	99.94
Random Forest	99.96	100.00	99.88
Ada Boost with J48	99.96	99.69	99.94
Ada Boost with Random Forest	99.96	99.85	99.91
Bagging with J48	99.84	99.77	99.97
Bagging with Random forest	99.96	99.92	99.94
Neural Network (ANN)	99.97	99.77	99.91
Support Vector Machines (SVM)	99.43	97.95	99.95

Worst performance Best performance

(b) Experiment 2.2: Pixel Based Study-with selected features

Two feature selection techniques were used to determine the useful features in the dataset:

- **Best First Technique**
- **Greedy Stepwise Technique**

Both the techniques selected same features to be significant for the dataset. The significant features selected are:

Table 6: Significant features selected for pixel based study

Significant Feature	Percentage Contribution
Band 1 Mean RIST	100
Band 4 Mean RIST	100
Band 5 Mean RIST	100
Mean of k in RIST Curve	100
SD of RIST Sums by Semicircles	80

All the statistical classifiers that we used in previous experiment were applied again on the selected features to determine the accuracy and computational speed. The results are summarized in the table 7 below:

Table 7: Percentage of correctly classified instances for experiment 2.2

Classification Algorithm	Oversampling	Under sampling	SMOTE
J48	99.88	99.61	99.71
Random Forest	99.92	99.77	99.88
Ada Boost with J48	99.92	99.77	99.88
Ada Boost with Random Forest	99.92	99.77	99.92
Bagging with J48	99.73	99.77	99.81
Bagging with Random forest	99.84	99.77	99.97
Neural Network (ANN)	95.28	94.60	98.91
Support Vector Machines (SVM)	93.72	93.84	99.70

Worst performance Best performance

(c) Experiment 2.3: ROI based study-All features

ROI based classification analysis was done using all the 68 features. Ten fold cross-validation technique was used to validate the model. Similar to experiment 2.1, following classifiers were used to build the model on the given dataset.

(a) Decision Trees (J48 & Random Forest)

(b) ADA Boost

(c) Bagging

(d) Support Vector Machines (SVM)

(e) Artificial Neural Network (ANN)

While conducting the data resampling, under-sampling technique was not used as the number of instances were low (12 rows of data) to perform under sampling. Therefore we used oversampling and SMOTE techniques for resampling the dataset for ROI based classification. The table 8 illustrates the % of correctly classified instances for different classifiers along with the resampling technique used on the dataset.

Table 8: Percentage of correctly classified instances for experiment 2.3

Classification Algorithm	Oversampling	SMOTE
J48	91.67	93.75
Random Forest	91.67	87.50
Ada Boost with J48	91.67	93.75
Ada Boost with Random Forest	83.33	87.50
Bagging with J48	91.67	87.50
Bagging with Random forest	75.00	87.50
Neural Network (ANN)	83.33	87.50
Support Vector Machines (SVM)	83.33	93.75

Worst performance Best performance

(d) <u>Experiment 2.4: ROI based study-with selected features</u>

Best First and Greedy Stepwise techniques were used determine the significant features and both the techniques yielded the same number of significant features enlisted below:

Table 9: Significant features selected for ROI based study

Significant Feature	Percentage Contribution
Raw Signal SD	100
Raw Signal Minimum	20
Raw Signal Maximum	90
Feature 0 Mean	30
Feature 3 Minimum	100
Feature 7 SD	100
Feature 11 SD	100
Feature 15 Mean	70

All the statistical classifiers that we used in previous experiments were applied again on the selected features to determine the accuracy and computational speed. The results are summarized in the table 10 below:

Table 10: Percentage of correctly classified instances for experiment 2.4

Classification Algorithm	Oversampling	SMOTE
J48	91.67	93.75
Random Forest	91.67	100.00
Ada Boost with J48	91.67	93.75
Ada Boost with Random Forest	91.67	100.00
Bagging with J48	91.67	93.75
Bagging with Random forest	91.67	100.00
Neural Network (ANN)	100.00	100.00
Support Vector Machines (SVM)	66.67	62.50

Worst performance Best performance

In Experiment 2 we studied and analyzed both pixel based and ROI based datasets. We also analyzed the data on selected features using feature selection techniques. From these experiments our conclusions are:

1. Pixel based analysis of the dataset seems to be moderately better than ROI based analysis. The accuracy is 99%. Our explanation is pixel based has more instances than the number of features and pixel based model may be more robust than ROI based model.
2. Since the models are having very good performance, that leaves less room for using features selection to improve performance. Yet, we do conclude comparable performance can be achieved with less # of features.
3. Amongst all the models studied, Trees (J48 & Random Forest) gives us a good result with high accuracy, easy to build and robust to outliers as well.
4. Trees classifiers are good at handling numerical as well as categorical predictors.

3.4.3 Experiment 3: Testing Prediction Accuracy of the build Classifier

Given the promising results from Experiment 2, we want to further assess the prediction power and the robustness of the developed models in tissue characterization. This experiment is therefore very crucial for the study, as it is performed to determine how accurate out model is and how confident we are in predicting using this model.

Similar to experiment 2, we have divided this experiment into two sections as well. The first section will be testing the model performance on ROI/pixel data, and determining the accuracy. This section will have two experiments with series of tests to check the robustness. Section 2 of this experiment will be testing the model performance on the whole liver-ROI/pixel data.

(a) Experiment 3.1: Testing the model on ROI/Pixel data

In this experiment we set aside data from four ROI regions (2 ROI regions of Tumor and 2 ROI regions of Non Tumor). This data was removed from the training dataset and saved as test data. The response column in the test data contains “actual” predictions C1 or C0. Similar approach for developing a classifier as used in Experiment 2 was used on the training data for Experiment 3. This experiment will help us determine the accuracy of the developed classifier.

The output will contain both the actual and predicted class. If in the class label for the test class we have '?' for each instance, the "actual" class label for each instance will not provide us with useful information, however the predicted class label will. The percentage of correctly classified instances of test dataset determines the accuracy of the developed classifier. Higher the accuracy, better the developed classifier can predict the

actual class. The following table 11 depicts the classifier developed using the training data and percentage of correctly classified instance using that classifier on the test data:

Table 11: Percentage of correctly classified instances for experiment 3.1

Classification Algorithm	Under sampling	Over sampling
J48	66.82	66.82
Random Forest	92.25	62.47
Ada Boost with J48	66.82	68.40
Ada Boost with Random Forest	64.77	76.75
Bagging with J48	56.05	56.77
Bagging with Random forest	66.70	90.92
Neural Network (ANN)	66.82	66.82
Support Vector Machines (SVM)	81.96	73.24

Worst performance Best performance

From the above experiment it seems that both SVM and trees (Random Forest) appears to be good approaches for predicting the response class for the pixel based study. Therefore to further test the robustness of these two selected classifiers we purposely and randomly remove some data from the existing dataset, using the remaining to build and validate the model, use the removed data to test the performance. The series of random experiments conducted and the prediction accuracy of SVM and Random forest have been summarized in the table below. Both SVM and Random forest provide consistent performance in prediction accuracy with high percentages.

Table 12: Prediction accuracy of SVM and RF for Patient # 3

Test Dataset	Train Dataset	SVM	RF
Tumor ROI from Patient 1	Remaining Dataset	92.5	90.1
Non-Tumor ROI from Patient 2	Remaining Dataset	100	85.3
Tumor ROI from Patient 3, Non Tumor ROI from Patient 1	Remaining Dataset	92.2	85.2
All Tumor ROIs	All Non Tumor ROIs	100	100
All Non Tumor ROIs	All Tumor ROIs	100	100

(b) Experiment 3.2: Prediction accuracy for Patient #3 data

In this experiment we specifically use Patient #1, 2 to develop the model and use Patient #3 to test the developed model. The objective of this test is primarily to determine that if we develop our model on a particular patient's data, whether or not it will be successful in predicting the response of some other patient, whose data has not been used in building the model.

Similar to the above experiment we have used pixel based dataset from Patient # 1, 2 to develop and train the classifier and used pixel based dataset from Patient #3 to test the developed classifier model. SVM and Random forest have been use as the classifier to develop the model as it is confirmed from Experiment 3.1 that these two classifiers outperforms the other classifiers as far as prediction accuracy is concerned. We have also used 10 fold cross validation techniques in this experiment.

The results of this experiment were over whelming, as it confirms the classifier selection. SVM prediction accuracy was 99.6% for predicting the response class of Patient #3 and Random forest was 82.92% accurate. This experiment therefore confirmed that the

selected classifier to develop the model was accurate enough to predict the response of any random patient without training the classifier for that patient's data.

(c) Experiment 3.3: Prediction accuracy for Whole Liver

This experiment was performed to predict the response class for all the pixels in the whole liver for patient# 3. Similar to experiment 3.2 this experiment will be testing the prediction power of the classifier. Along with this we will also be conducting sensitivity, specificity and accuracy tests to check the robustness of the developed model. The training data remains the same as the above experiment just the test data changes. In this experiment the test data used will be the whole liver for patient # 3. There are total 5955 instances in the whole liver test dataset which contains both tumor as well as non-tumor tissues. The Actual Response column in the dataset replaced by "?", therefore the actual class label for each instance will not contain useful information, but the predicted class label will. SVM will be used as the classifier to build and develop the model.

A small part of the output after the model was run on WEKA is shown in table 13 below.

Table 13: Prediction accuracy of SVM and RF for Patient #3, whole liver

Instance #	Actual Class	Predicted Class
1	?	C0
2	?	C0
3	?	C0
4	?	C1
5	?	C0
6	?	C1
7	?	C1
8	?	C0
9	?	C0

Since we do not have the actual class therefore the actual class is replaced by "?" and the predicted class label gives us the useful information about the response class for that pixel.

Further we performed sensitivity, specificity and accuracy tests to check the robustness of the developed model. Commonly used confusion matrix for binary classification problems is shown in Table 14 below.

Table 14: Confusion Matrix in general form

Outcome of the diagnostic test	Condition		
	Positive	Negative	Row Total
Positive	True Positive(TP)	False Positive(FP)	TP+FP
Negative	False Negative(FN)	True Negative(TN)	FN+TN
Column Total	TP+FN	FP+TN	N=TP+TN+FP+FN

In Table 14, TP refers to number of samples correctly identified as positive, FP refers to number of samples incorrectly identified as positive, TN refers to number of samples correctly identified as negative, and FN refers to number of samples incorrectly identified as negative.

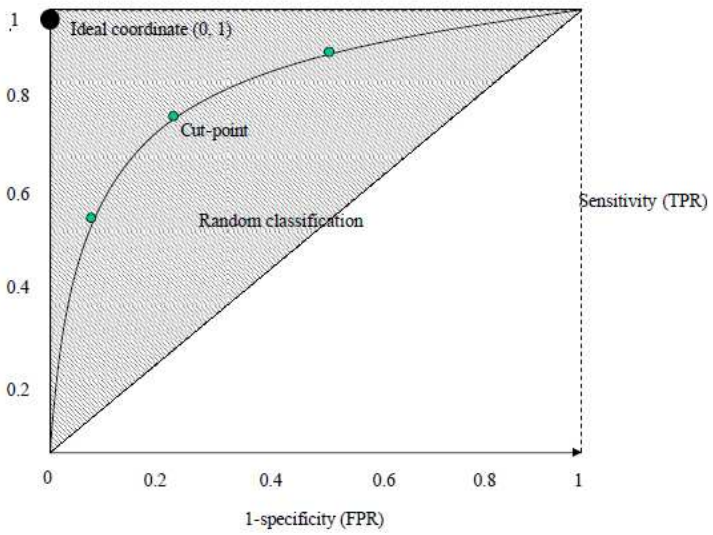


Figure 18: Receiver Operating Characteristics (ROC) space, the area under the ROC curve is known as Area under the Curve (AUC)

For a given diagnostic test, the true positive rate (TPR), false positive rate (FPR) and Accuracy can be measured using following formulas:

Sensitivity or TPR= TP / (TP+FN)

Sensitivity or TPR is the proportion of true positives that are correctly identified by a diagnostic test. It is a measure of how efficient the test is at detecting a disease.

FPR= FP / (FP+TN)

FPR is actually the proportion of positive tests among people without the disease or condition.

Specificity = TN / (FP+TN)

Specificity is the proportion of the true negatives correctly identified by a diagnostic test. It tells us how efficient our test is at detecting normal (negative) condition.

Accuracy = (TP+TN) / (TP+TN+FP+FN)

Accuracy is defined as the percentage of true results, be it true positive or true negative, in a population. It is a measure of the degree of precision for a diagnostic test on a condition.

The accuracy of a diagnostic test can also be measured by analyzing the area under ROC curve (AUC), shown in figure 16. Large area signifies an accurate the diagnostic test is.

The confusion matrix for Experiment 3.3 is:

Table 15: Confusion Matrix for Experiment 3.3

Outcome of the diagnostic test	Condition		
	Tumor	Non Tumor	Row Total
Tumor	TP=170	FP=820	990
Non Tumor	FN=0	TN=4965	4965
Column Total	170	5785	5955

Analysis of the dataset for Experiment 3.3 gave following results:

Sensitivity: 100.00%

Specificity: 85.82%

Accuracy: 86.23%

In Experiment 3 we studied the robustness and prediction power of the developed model. From these experiments our conclusions are:

1. SVM is a robust classifier in predicting the response class variable especially when the response has two categorical indicators, as in our case C1 & C0.

2. The classifier was able to predict the tumor tissues with high accuracy (sensitivity 100 %), but there is still room for improvement as far as predicting the correct non-tumor tissues as specificity and accuracy were around 85 % mark.

3. One disadvantage with SVM is its computational complexity, but WEKA takes good care of it and present s the output in nice and easily interpretable format.

4. Overall performance of SVM is superior compared to other classifiers tested

CHAPTER 4

CONCLUSION AND FUTURE WORK

Hepatocellular carcinoma (HCC) is a malignant tumor and it is considered as the seventh most frequent occurring cancer in human [1]. The incidence of HCC is mainly due to Hepatitis C, which also leads to chronic liver complications such as fibrosis, cirrhosis. Nonalcoholic Steatohepatitis (NASH) & Non-Alcoholic Fatty Liver Disease (NAFLD) are also major concerns that are growing really fast and it is expected that they will be surpassing Hepatitis C as the major cause in the near future. Most clinical research has focused on HCC early detection when the tumor might be curable by resection, liver transplantation, or ablation and a 5-year survival higher than 50% can be achieved [5].

Commonly used imaging techniques for HCC diagnosis include Ultrasound (US), Computed tomography (CT) and Magnetic resonance (MR). US has been largely replaced by CT and MRI due to its low sensitivity and specificity in HCC diagnosis [6]. As a result, liver disease diagnosis has mainly relied on CT and MRI imaging criteria. However, these two techniques are not without limitations. A retrospective analysis comparing the accuracy of radiologic staging with pathologic staging on liver disease patients found that imaging based diagnosis (both CT and MRI) resulted in a high number of false positives labeling common benign focal abnormalities in the liver as malignant tissue [7].

Another emerging imaging technique that has been under discussion over the last few years is Magnetic Resonance Elastography (MRE), developed by Mayo Clinic. It is a relatively new technique and under considerations for improvement, but MRE has proved to be a more accurate and reproducible technique compared to conventional imaging

features better diagnostic accuracy. In addition, most research published to date relies on subjective and variable assessment of imaging features. We believe Texture analysis is one potential technique to generate multiple, objective, reproducible, quantifiable features from medical images. Classification of healthy and diseased livers using modern imaging such as MRE in conjunction with Texture Analysis may address this clinical challenge [17]. Texture recognition is an important aspect of medical image analysis. Texture Analysis is of high importance in medical imaging analysis because, as the biological tissue becomes abnormal during a disease, its underlying texture could also change. Current imaging techniques, such as MR, are not capable enough of providing microscopic information of tissue that can be assessed visually. Texture Analysis can be successfully used for the separation of cirrhotic patients and healthy volunteers, and unknown patient data can be safely classified into the patient group. Different sets of Texture Analysis features can be used for a similar classification of patients. A combination of features significantly improves the ability of Texture Analysis to confirm the classification of the subjects.

In this research we evaluated the features from Texture Analysis, to effectively diagnose HCC for three patient's MRE images. The objective was to quantitatively assess and validate the applicability of MRE in conjunction with advanced imaging processing Texture Analysis pipeline and multi-variate analysis machine learning method for accurate HCC diagnosis. It was discovered that Texture Analysis pipeline is a useful tool to extract image signatures for these patients from their MRE images. During the analysis uni-variate approach for quantitatively assessing the diagnosis performance of MRE was also evaluated based on the intensity information of the tissues only. Three experiments

were performed, with each experiment having series of sub experiments to validate the applicability of MRE and Texture Analysis pipeline using multiple machine learning algorithms in WEKA. Major focus of this research has been concentrated on building the model, choosing the correct classifier to build a model, model validation and finally testing the build model. Based on our research and analysis we were able to draw concrete conclusions on the applicability of MRE along with Texture Analysis pipeline in HCC diagnosis, these conclusions are:

(a) Texture analysis pipeline is a useful tool to extract image signatures, both Texture Analysis and MRE proved to be promising imaging tools for HCC diagnosis.

(b) The model that was developed and validated had 99% accuracy in HCC tissue diagnosis.

(c) The model that was tested had up to 92% accuracy robustly in HCC tissue characterization.

(d) Our testing model can predict the HCC tissue for Patient# 3 with sensitivity of 100 %, specificity of 85.82 % and accuracy of 86.23 %

(e) Uni-variate study of the intensity information of tissues was able to give us a broad picture that the intensities of tumor and non-tumor regions differs significantly, but we were not able to precisely differentiate between the two region, therefore an advanced study was required.

(f) Multi-variate analysis is preferred in this study over uni-variate analysis, as it provides us with rich information to harvest the image at pixel level with high confidence.

(g) We observed that among all the image signatures, indeed, there may exist some redundancies, in this case, the system can benefit from using feature selection to narrow down to smaller number of significant features.

(h) We also observed pixel based model tends to moderately outperforms ROI based model. Yet, given the limited number of patient data available, this needs to be confirmed in the future research.

(i) We studied six machine learning methods; in general, SVM demonstrated outperformance in most experiments. This is due to the fact that SVM is known to be a good classifier for binary classification problem and robust in nature.

(j) We also conclude the model develop can robustly provide accurate predictions.

(k) The classifier was able to predict the tumor tissues with high accuracy (sensitivity 100%), but there is still room for improvement as far as predicting the correct non-tumor tissues as specificity and accuracy were around 85 % mark.

While we see that the results from the analysis are promising as far as applicability of MRE and Texture Analysis are concerned in HCC diagnosis, there is still plenty of room for improvement which can be implemented in future works:

(a) The model performance is far better than we initially expected, this maybe because the ROI was drawn precisely, which provides us a very good and representative dataset.

(b) This analysis was done on 2D, which proved its usefulness, yet with accuracy of 86.23 % with 2D. In addition, the specificity is only 85.82%. This concurs with conclusion from most existing investigations that current imaging techniques may result in a high number of false positives labeling common benign focal

abnormalities in the liver as malignant tissue [7]. We believe there is scope of improvement, for example exploring 3D MRE for better accuracy in terms of tissue characterization.

(c) This study involved only three patients due to limited number of available patients for HCC diagnosis study. Future studies must include more number of patients for further validation of the model.

(d) Phantom data may also be used for the validation study.

(e) More variability can be introduced in the model building, by allowing different radiologists to draw ROIs on the same MRE image for the patient. This will make sure that radiologist's variability is also included while building the model.

By these experiments we can improve on the current study and further higher accuracy can be achieved for characterizing the tissues better.

REFERENCES

[1] Massimo Colombo, "Hepatocellular carcinoma," Journal of Hepatology, Institute of Medicine, University of Milan, Milan, Italy, 1992.

[2] Asmaa Ibrahim Gomaa, Shahid A Khan, Mireille B Toledano, Imam Waked, Simon D Taylor-Robinson, "Hepatocellular carcinoma: Epidemiology, risk factors and pathogenesis," World Journal of Gastroenterology, 2008.

[3] Ramón Bataller and David A. Brenner, "Liver fibrosis," The Journal of Clinical Investigation.

[4] Gines, P., Cardenas, A., Arroyo, V., and Rodes, J., "Management of cirrhosis and ascites," N. Engl.J. Med., 2004.

[5] Sameer Parikh, MD, David Hyman, MD, MPH, "Hepatocellular Cancer: A Guide for the Internist," The American Journal of Medicine , Department of Medicine, Baylor College of Medicine, Houston, Tex., 2007.

[6] Byung Ihn Choi, "The Current Status of Imaging Diagnosis of Hepatocellular Carcinoma," American Association for the Study of Liver Diseases.

[7] Richard B. Freeman, Abigail Mithoefer, Robin Ruthazer, Khanh Nguyen, Anthony Schore, Ann Harper and Erick Edwards, "Optimizing Staging for Hepatocellular Carcinoma Before Liver Transplantation: A Retrospective Analysis of the UNOS/OPTN Database," American Association for the Study of Liver Diseases, LIVER TRANSPLANTATION, 2006.

[8] Jonathon M. Willatt, Hero K. Hussain, Saroja Adusumilli, Jorge A. Marrero, MD, "MR Imaging of Hepatocellular Carcinoma in the Cirrhotic Liver: Challenges and Controversies," Radiology: Volume 247: Number 2—May, 2008.

[9] Jorgen Arendt Jensen, "Medical ultrasound imaging," Elsevier Ltd., 2006.

[10] Fattovich G, Stroffolini T, Zagni I, Donato F., "Hepatocellular carcinoma in cirrhosis: incidence and risk factors," Gastroenterology, 2004.

[11] Antonio Di Ieva, Fabio Grizzi, Elisa Rognone, Zion Tsz Ho Tse, Tassanai Parittotokkaporn, Ferdinando Rodriguez y Baena, Manfred Tschabitscher, Christian Matula, Siegfrid Trattnig & Riccardo Rodriguez y Baena, "Magnetic resonance elastography: a general overview of its current and future applications in brain imaging," Neurosurg Rev, 2010.

[12] Stephen Hughes, "Medical ultrasound imaging," IOP Publishing Ltd, 2001.

[13] K.Klingenbeck and H.E.Reinfelder, "Medical Imaging Techniques," ISPRS Journal of Photogrammetry and Remote Sensing, Elsevier Science Publishers B.V., Amsterdam, 1990.

[14] Carmen Ayuso, Jordi Rimola, A´ngeles Garcı´a-Criado, "Imaging of HCC," Springer Science+ Business Media, LLC, 2011.

[15] Byung Ihn Choi, Jeong Min Lee, "Advancement in HCC imaging: diagnosis, staging and treatment efficacy assessments," Japanese Society of Hepato-Biliary-Pancreatic Surgery and Springer, 2009.

[16] Carlos Rodriguez de Lope, Silvia Tremosini, Alejandro Forner, Maria Reig, Jordi Bruix, "Management of HCC," Journal of Hepatology, 2012.

[17] Daniel Jira´k, Eng, Monika Dezortova´ , PhD, Pavel Taimr, MD, and Milan Ha´ jek, DSc, "Texture Analysis of Human Liver," JOURNAL OF MAGNETIC RESONANCE IMAGING.

[18] Sylvia Drabycz, Robert G. Stockwell and J. Ross Mitchell, "Image Texture Characterization Using the Discrete Orthonormal S-Transform," Journal of Digital Imaging, Vol 22, No 6 (December), 2009.

[19] Sylvia Drabycz, Gloria Roldán, Paula de Robles, Daniel Adler, John B. McIntyre, Anthony Magliocco, J. Gregory Cairncross, J. Ross Mitchell, "An analysis of image texture, tumor location, and MGMT promoter methylation in glioblastoma using magnetic resonance imaging," NeuroImage , 2010.

[20] Zhi-Zhong Wang and Jun-Hai Yong, "Texture Analysis and Classification with Linear Regression Model Based on Wavelet Transform," IEEE Transactions on Image Processing, Vol 17, No. 8, AUGUST, 2008.

[21] Maryana de Carvalho Alegro, Alexandre Valotta Silva, Silvia Yumi Bando, Roseli de Deus Lopes, Luiz Henrique Martins de Castro, Wen HungTsu, Carlos Alberto Moreira-Filho, and Edson Amaro, Jr, "Texture Analysis of High Resolution MRI Allows Discrimination Between Febrile and Afebrile Initial Precipitating Injury in Mesial Temporal Sclerosis," Magnetic Resonance in Medicine 68:1647–1653, 2012.

[22] G. Castellano, L. Bonilha, L.M. Li, F. Cendes, "Texture analysis of medical images," Clinical Radiology, 2004.

[23] Manish H. Bharati, J. Jay Liu, John F. MacGregor, "Image texture analysis: methods

and comparisons," Chemometrics and Intelligent Laboratory Systems, 2004.

[24] Khan M. Iftekharuddin, Wei Jia, Ronald Marsh, "Fractal analysis of tumor in brain MR images," Machine Vision and Applications, 2003.

[25] Justin M. Zook, Khan M. Iftekharuddin, "Statistical analysis of fractal-based brain tumor detection algorithms," Magnetic Resonance Imaging, 2005.

[26] K. M. Iftekharuddin, J. Zheng, M. A. Islam, R. J. Ogg, F. Lanningham, "Brain Tumor Detection in MRI: Technique and Statistical Validation".

[27] G.M. Behery, "SIMULATION OF FRACTAL DIMENSION EVALUATIONS," International Journal of Modelling and Simulation, Vol. 26, No. 2, 2006.

[28] M. SASIKALA and N. KUMARAVEL, "A wavelet-based optimal texture feature set for classification of," Journal of Medical Engineering & Technology, 2008.

[29] S. Arivazhagan , L. Ganesan, "Texture classification using wavelet transform," Elsevier Science, 2002.

[30] Julia´n Luengo, Alberto Ferna´ndez, Salvador Garcı´a, Francisco Herrera, "Addressing data complexity for imbalanced data sets: analysis of SMOTE-based oversampling and evolutionary undersampling," Springer-Verlag, 2010.

[31] Daniel A. Keim, Ming C. Hao, Umeshwar Dayal, Halldor Janetzko and Peter Bak, "Generalized scatter plots," Macmillan Publishers Ltd. 1473-8716, 2010.

[32] Jorg Schrader, Timothy T. Gordon-Walker, Rebecca L. Aucott, Marielle van Deemter, Alexander Quaas, Shaun Walsh, Daniel Benten, Stuart J. Forbes, Rebecca G. Wells and John P. Iredale, "Chemotherapeutic Response, and Dormancy in Hepatocellular Carcinoma Cells," HEPATOLOGY, Vol. 53, No. 4,, 2011.

[33] S. Arivazhagan, L. Ganesan, "Texture classification using wavelet transform," Pattern Recognition Letters, 2003.