CLASSIFICATION OF LIVER TISSUE BASED ON HISTOLOGICAL AND IMAGING FEATURES

Submitted in partial fulfillment of the requirement for the

degree of

Bachelor of Technology

In

Electronics and Communication Engineering

By

Shrestha Bansal **111043**

Gaurav Chhabra **111051**

B. Sarat Chandra **111053**

under the Supervision of

Dr. Jitendra Virmani

MAY 2015

JAYPEE UNIVERSITY OF INFORMATION AND TECHNOLOGY

CERTIFICATE

This is to certify that project report entitled "**Classification of Liver Tissue based on Histological and Imaging Features**", submitted by Shrestha Bansal (111043), Gaurav Chhabra (111051), B. Sarat Chandra (111053) in partial fulfilment for the award of degree of Bachelor of Technology in Electronics and Communication Engineering to Jaypee University of Information Technology, Waknaghat, Solan has been carried out under my supervision. This work has not been submitted partially or fully to any other University or Institute for the award of this or any other degree or diploma.

Date: 25/5/2015

Supervisor: Dr. Jitendra Virmani

Designation: Assistant Professor (Senior Grade)

Signature:

Shrestha Bansal (111043)

Gaurav Chhabra (111051)

B. Sarat Chandra (111053)

Acknowledgement

We owe a thanks to many people who have helped and supported us during this project. Our deepest thanks to **Dr. Jitendra Virmani**, our project guide for his exemplary guidance, monitoring and constant encouragement throughout the course of this project work. He has taken great pains to go through the project and make necessary corrections as and whenever needed. We are also grateful to **Mr. Mohan (ECE Project lab**) for his practical help and guidance. We would also like to thank all the faculty members of ECE department without whom the progress of this project would have been a distant reality. We also extend our heartfelt thanks to our family and well-wishers.

Date:

Name of the students:

Shrestha Bansal 111043

Gaurav Chhabra 111051

B. Sarat Chandra 111053

List of Abbreviations

ABNOR	Abnormal
AC	Autocorrelation
ANN	Artificial Neural Network
B-Mode	Brightness Mode
CAD	Computer-aided Diagnosis
CM	Confusion Matrix
E5	Edge Detector Filter of Length 5
E7	Edge Detector Filter of Length 7
E9	Edge Detector Filter of Length 9
FLL	Focal Liver Lesion
FS	Feature Set
FVL	Feature Vector Length
GA	Genetic Algorithm
GA–SVM	Genetic Algorithm–Support Vector Machine
GUI	Graphic User Interface
HCC	Hepatocellular Carcinoma
HEM	Hemangioma
HVS	Human Visual System
Hy-HCAD	Hybrid-Hierarchical Computer-aided Diagnostic System
ICA	Individual Class Accuracy
IROIs	Inside Regions of Interest
k-NN	k-Nearest Neighbor

Kurt	Kurtosis
L5	Level Detector Filter of Length 5
L7	Level Detector Filter of Length 7
L9	Level Detector Filter of Length 9
LHCC	Large Hepatocellular Carcinoma
LHCCI	Large Hepatocellular Carcinoma Image
MET	Metastatic Carcinoma
ML	Malignant Lesion
MPNN	Modified Probabilistic Neural Network
MRI	Magnetic Resonance Imaging
NN	Neural Network
NNE	Neural Network Ensemble
NOR	Normal
OCA	Overall Classification Accuracy
OL	Other Lesion
PC	Principal Component
PCA	Principal Component Analysis
PGIMER	Post Graduate Institute of Medical Education and Research
PML	Primary Malignant Lesion
PNN	Probabilistic Neural Network
R5	Ripple Detector Filter of Length 5
R7	Ripple Detector Filter of Length 7
R9	Ripple Detector Filter of Length 9
ROC	Region of Convergence

ROIs	Regions of Interest
S5	Spot Detection Filter of Length 5
S7	Spot Detection Filter of Length 7
S9	Spot Detection Filter of Length 9
SHCC	Small Hepatocellular Carcinoma
SHCCI	Small Hepatocellular Carcinoma Image
Skew	Skewness
S_p	Spread Parameter
SROIs	Surrounding Regions of Interest
Std	Standard Deviation
SVM	Support Vector Machine
SSVM	Smooth Support Vector Machine
TDs	Texture Descriptors
TEI	Texture Energy Image
TEM	Texture Energy Measure
US	Ultrasound
W5	Wave Detection Filter of Length 5
W9	Wave Detection Filter of Length 9

List of Figures

List of Tables

Abstract

Computer-aided diagnostic(CAD) system for characterization of different datasets for different diseases has the potential to assist the radiologists and clinicians in deducing whether the patient is suffering from the disease or not through histological processing by different machine learning algorithms which have to be applied on the given liver disorder datasets. In the absence of real time database, the current work of histological processing is based on a benchmark BUPA database created by University of California, Irvine. The current work aims to provide a means of economic assistance to the clinicians for the effective diagnosis of the liver diseases.

For this purpose, two types of classification techniques are used. First, Histological classification and second, classification through imaging features. In histological feature classification, five classifiers namely K Nearest Neighbor (KNN), Probabilistic Neural Network (PNN), Support Vector Machine (SVM), Neural Networks Toolbox (NN) and Smooth Support Vector Machines (SSVM) are used. In imaging classification, first step is feature extraction, then feature selection and ultimately, classification. Here, Ultrasound images are used to maintain the cost-effectiveness of the tool. Further, the work uses histological and imaging features collectively to further improve the overall efficiency of the designed tool.

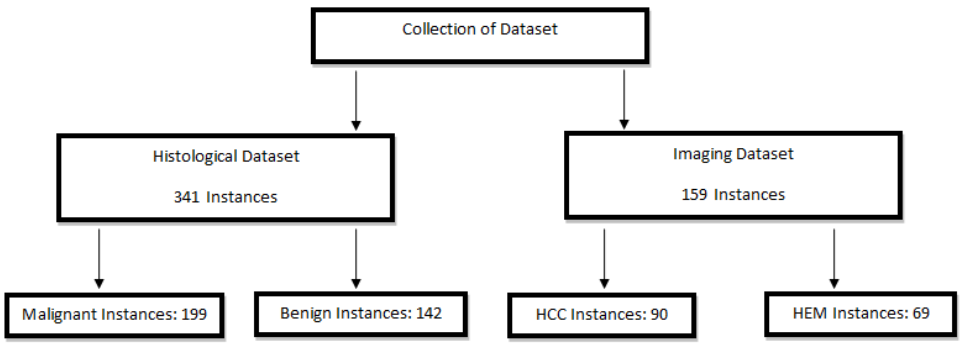

Fig. 1 Description of the database used in the current work.

Contents

Chapter 1

Introduction

1.1 Introduction

In this chapter, the importance of liver tissue is expressed and it is explained why the liver tissue was chosen as the area of research and what are the primary types of diseases affecting liver. Here, the focal liver diseases are explained and out of them Hemangioma and Hepatocellular Carcinoma are studied in detail. This chapter also includes the type of dataset used and the reason for using them. Further this chapter discusses about the need for a Computer Aided Diagnostic (CAD) System, objectives of the current research and the literature review.

1.2 Motivation

Liver is the most vital and largest organ of the human body. It performs many important functions like production and excretion of bile (a digestive fluid), synthesis of cholesterol, production of triglycerides (fats), metabolism of proteins, fats and carbohydrates, storage of vitamins and minerals, synthesis of plasma proteins, breakdown of insulin and other hormones, blood pressure management, blood detoxification, etc. Liver is a metabolically active organ necessary for survival. The working cells of the lever (called *hepatocytes*) have unique capability to reproduce whenever the liver is injured. Thus, liver regeneration can occur after surgical removal of a portion of the liver or after an injury that destroys a part of the liver. However, there is absolutely no way to compensate for long-term liver dysfunction, because of the diversity of functions it handles.

As liver is the largest solid organ of the human body, it becomes an easy target for many diseases. Liver diseases are widely recognized as an emerging public health crisis particularly in South Asian countries. In clinical diagnosis, liver diseases are always taken seriously as it is a vital organ, which performs very important functions required for sound operation of human body. Liver diseases are classified in two broad categories, i.e., *diffuse liver diseases* and *focal liver diseases*.

1.2.1 Focal Liver Diseases

In focal liver diseases, the abnormality is concentrated in a small localized region of the liver parenchyma which is often referred to as *focal liver lesion* (FLL). Liver Cysts, Hemangioma (HEM, i.e., a primary benign FLL), Hepatocellular carcinoma (HCC, i.e., a primary malignant FLL) and Metastatic carcinoma (MET, i.e., a secondary malignant FLL), are some of the commonly occurring focal liver diseases.

1.2.1.1 Hemangioma (*HEM*)

The hemangioma (HEM) is the most common primary benign FLL. It is a highly vascular benign FLL which is composed of tiny blood vessels. HEMs usually appear as a solitary lesion, but may also be multiple in 10 % of cases. In most of the cases, HEMs are small (< 3 cm) and are found incidentally. In very rare cases, these lesions are symptomatic; but it is sometimes difficult to diagnose these lesions as they can be indistinguishable from MET lesions. They appear in all age-groups but are more frequent in adult females. Once HEMs are detected in adult, they are stable in size, any further change in size and appearance is uncommon. HEMs found in children tend to be large and symptomatic. Many of these HEMs found in children regress with time, while others may have to be emoblized with coils under radiological guidance.

The sonographic appearance of HEMs varies considerably. In 70 % of cases, HEMs encountered in routine clinical practice are *typical HEMs*. These typical HEMs have a characteristic sonographic appearance; it appears as a round, homogeneous, hyperechoic, well defined lesion. These typical HEMs may sometimes exhibit posterior acoustic enhancement due to blood filled capillaries. *Atypical HEMs* are a great mimic and a definite diagnosis with conventional gray scale B-Mode US is difficult. Atypical HEMs can be isoechoic or even hypoechoic mimicking the sonographic appearance of certain atypical MET and HCC lesions. These atypical HEMs generally cause diagnostic problems as they may appear as hypoechoic lesions or as lesions with mixed echogenicity. Large HEMs (> 3 cm) are often heterogeneous and demonstrate spectrum of reflectivity based on the composition and central areas of degeneration. These large HEMs frequently exhibit slightly increased through-transmission with posterior acoustic enhancement. In case of atypical HEMs, where the diagnosis is not certain and a malignancy is suspected, administration of an

ultrasound contrast agent and further imaging like MRI scanning helps to characterize the lesion confidently.

1.2.1.2 Hepatocellular Carcinoma (HCC)

The *hepatocellular carcinoma* (HCC), also called as malignant hepatoma (liver cancer), is primary malignant FLL. HCC accounts for 80 to 90 % of all the malignant FLLs, amongst various primary FLLs. The US imaging modality is used world-wide for screening of HCCs. This occurrence of HCC is most common in adult population. It is the fifth most common cancer worldwide and the third leading cause of cancer related deaths. Worldwide, HCCs are detected with an estimated occurrence of 100000 - 300000 new cases per year. The occurrence of HCC is not uniform throughout, with highest occurrence rates in Sub-Saharan Africa and the Southeast Asia. The areas of low occurrence include North America and Northern Europe. Males have higher occurrence of HCC than females.

The risk factors which give rise to development of HCC are (*i*) cirrhosis, (*ii*) chronic infection with the hepatitis B and hepatitis C virus, and (*iii*) metabolic diseases. The symptoms of liver cancer vary among individuals. Many patients with primary liver cancer reveal no symptoms until the cancer develops to an advanced stage. In some cases, jaundice, general feeling of poor health, loss of appetite, weight loss, nausea, fever, fatigue, bloating, itching, swelling of legs, or weakness may be present. In certain cases abdominal pain or discomfort may also occur. It is worth mentioning that these symptoms can be vague and very similar to other diseases and conditions.

In 85 % cases, HCC occurs in patients with cirrhosis. The appearance of HCC on B-Mode US depends mostly on whether or not there is underlying cirrhosis. In fact, in radiology practice, cirrhosis is seen as precursor to development of HCC as the occurrence of HCCs on normal liver is very rare. Detecting *small HCCs* (SHCCs) developed on coarse and nodular cirrhotic liver parenchyma presents a daunting challenge for experienced radiologists. On the other hand, in rare cases when the HCC develops on normal liver parenchyma it can be easily diagnosed from its sonographic appearance, as it appears as a well differentiated HCC or as fibro lamellar HCC (which commonly appears with calcified areas).

The sonographic appearance of a *large HCC* (LHCC) is often inhomogeneous, whereas SHCCs can be hypoechoic and homogeneous. Experienced participating radiologists opined that *no sonographic appearance can be considered typical for HCC as there is a wide variability of sonographic appearances even within SHCCs and LHCCs*. The sonographic appearances of SHCC vary from hypoechoic to hyperechoic. LHCC appear frequently with mixed echogenicity.

1.2.2 Histological Processing

The microscopic level study of different organs of the body has helped in the evolution of medical science in the effective diagnosis of the diseases related to these organs. In the current study, these histological features of liver tissue are obtained by conducting several medical tests such as Mean Corpuscular Volume, Alkaline Phosphotase, Alamine Aminotransferase, Aspartate Aminotransferase, and Gamma-Glutamyl Transpeptidase. Based on these tests, a dataset is created which is further used in the work. In the current work, the histological database was obtained from the site of University of California, Irvine (UCI), which has uploaded this data for the purpose of research globally.

1.2.3 Ultrasound Imaging

The field of medical imaging and image analysis has evolved due to collective efforts from many disciplines like medicine, engineering and basic sciences. In current medical practice, imaging procedures are one of the major bases for diagnosis apart from other procedures like pathological examinations and biopsy. The overall objective of the medical imaging system is to acquire useful information about the physiological processes of the organs of the human body. The choice of the best imaging technique for any particular clinical application is based on several factors including resolution, speed, convenience, acceptability and safety. As an example, the US imaging modality is ideally suited for imaging the soft tissues, over other techniques accounting for all these factors. The other imaging modalities used for diagnosis of liver diseases include computed tomography (CT) and magnetic resonance imaging (MRI). The US, CT and MRI are all non-invasive imaging modalities. However, CT uses ionizing radiations, which are otherwise harmful for human body. On the other hand, US don't produce any known harmful effects on any of the tissues examined during clinical practice. The clinical relevance of the US

imaging modality is high worldwide due to its versatility, wide spread availability, portability and ease of operation in comparison to CT and MRI.

The US is particularly useful for differentiating between *cystic* and *solid FLLs*, whereas CT and MRI are particularly sensitive for differential diagnosis between solid FLLs. For differential diagnosis between solid FLLs, the radiologists don't rely on US examinations only, because of varying overlapping sonographic appearances between them. Therefore, for confirming their diagnosis the radiologists resort to administration of contrast agents, or additional imaging procedures (CT and MRI) which are costlier and time consuming, or invasive procedures such as biopsy. Furthermore, the diagnostic information extracted from the US examination is highly operator dependent; but this limitation can be overcome by proper training of the observer. In addition, obese patients can be difficult to scan with US and thus obtaining good quality diagnostic US images for these patients can be considerably difficult. Despite the disadvantages associated with US imaging modality, it is the most preferred option for screening of the liver, especially in the developing countries like India where most of the patients generally come from rural environment who cannot afford the financial burden of radiological procedures which are relatively costlier.

The aim of the present research work is to do value addition in the diagnostic performance obtained by most commonly available conventional gray scale B-Mode US imaging modality for diagnosis of liver diseases.

1.3 Sonographic Appearances of Different Liver Image Classes used in the Present Research Work

The brief details of the sonographic appearances of liver image classes used in the present research work are depicted below:

1.3.1 Sonographic Appearance of Normal Liver

The sonographic appearance of normal (NOR) liver is homogeneous with slightly increased echogenicity as compared to the right kidney. The NOR liver appears as a mid-gray organ with smooth outlining and homogeneous echotexture because of its uniform acoustic impedance on ultrasound. The sample of the Normal liver image from the image database is given in Fig. 1.1.

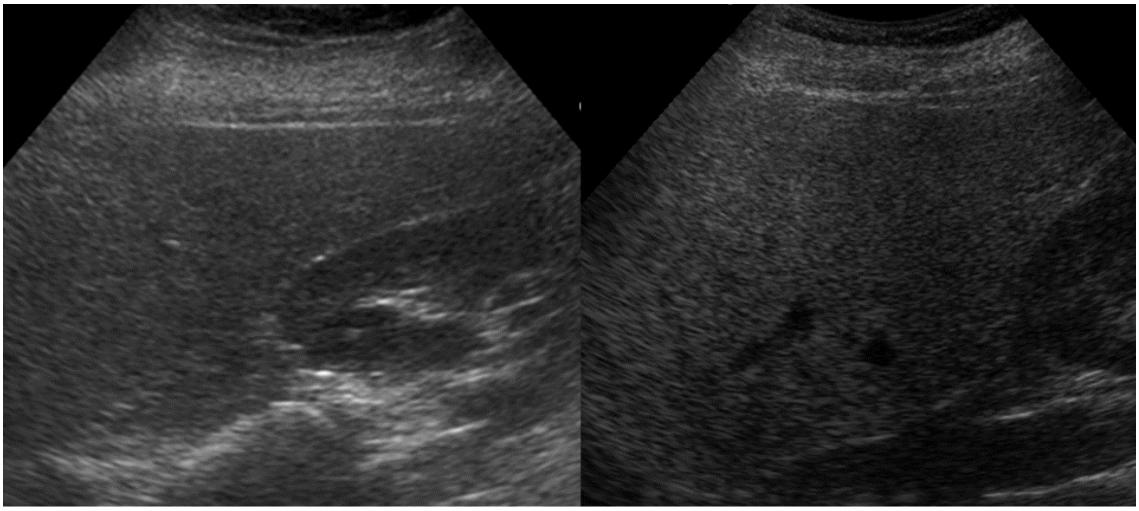

Fig. 1.1 Conventional gray scale ultrasound liver images with appearance of normal liver.
Note: Normal liver exhibits homogeneous echotexture with medium echogenicity.

The smooth liver parenchyma is interrupted by anechoic structures such as vessels (i.e., the hepatic veins, portal veins, hepatic arteries, etc. The capsule of the liver appears hyperechoic especially at its border with the diaphragm. The diaphragm appears as a curvilinear bright reflector. It is difficult to quantify the size of the liver as there are large variations in shape within normal subjects. The size of the liver is therefore assessed subjectively. All the NOR cases are considered as typical as there is no atypical appearance for normal liver tissue.

1.3.2 Sonographic Appearance of Small and Large HCCs

The sonographic appearances of Small HCC (SHCC) vary from hypoechoic to hyperechoic. Large HCC (LHCC) appears frequently with mixed echogenicity. Experienced participating radiologists opined that *no sonographic appearance can be considered typical for HCC* as there is wide variability of sonographic appearances even within small HCCs (SHCCs) and large HCCs (LHCCs).

The sample images of SHCC and LHCC cases from the image database are shown in Fig. 1.2.

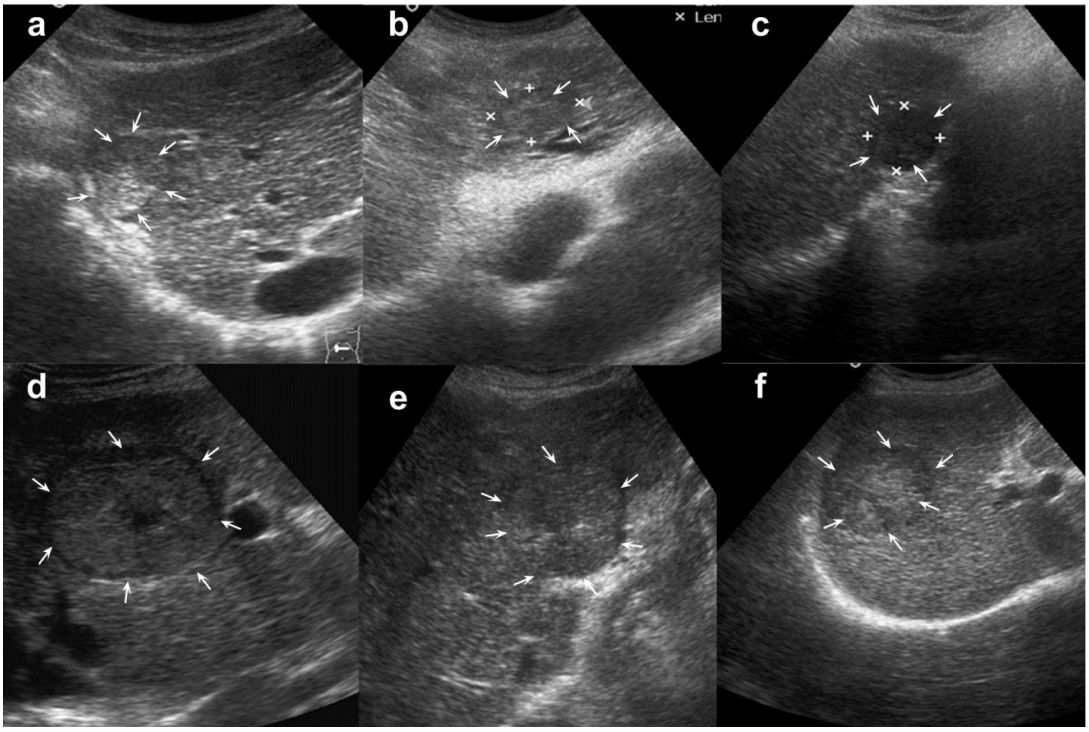

Fig. 1.2 Sample images of SHCC and LHCC variants from the image database: **(a)** Variant of SHCC with mixed echogenicity (coexistence of hyper-echoic and iso-echoic areas); **(b)** Isoechoic SHCC; **(c)** Hypoechoic SHCC; **(d-f)** Heterogeneous echotexture represents complex and chaotic structure exhibited by LHCC due to coexistence of areas of necrosis, fibrosis and active growth areas.
Note: Hypoechoic halo formation is visible in **(d)**, Necrotic area is visible at centre of LHHC in **(d)**.

1.4 Need for CAD Systems for Liver Diseases

The evolution of computer technology, medical image processing algorithms and artificial intelligence techniques has given ample opportunity to researchers to investigate the potential of computer-aided diagnostic systems for tissue characterization. Tissue characterization refers to quantitative analysis of tissue imaging features resulting in accurate distinction between normal and abnormal tissues. Thus, the result of tissue characterization is interpreted using numerical values. The overall aim of developing a computerized tissue characterization system is to provide additional diagnostic information about the underlying tissue which cannot be captured by visual inspection of B-Mode US images.

Ultrasonographic tissue characterization methods based on physical tissue models have been shown to be useful for improving the diagnostic accuracy of sonograms. Unfortunately, no physical model based diagnostic system have been developed for characterization of FLLs probably because these systems have been developed assuming single, homogeneous tissue model, whereas in case of FLLs the

variability in sonographic appearances within different lesions is quite large and quite often large HCCs and MET lesions are inhomogeneous.

For viable and useful sonographic characterization of FLLs, radiologists need to extract subtle sonographic information which may be difficult to extract visually, consistently and objectively. It is, therefore, expected that sophisticated computerized analysis of the texture patters of FLLs can yield objective characterization of lesions.

1.5 Objectives of the Present Study

The main objective of the research work presented in this thesis is to enhance the diagnostic potential of various forms of diagnosis of liver diseases by developing efficient CAD system designs using a comprehensive and representative histological and image database. To achieve this, various research objectives were formulated according to the needs of the radiologists, based on the practical difficulties faced by them in routine clinical practice. These research objectives are described below:

To develop a hybrid Computer Aided Diagnostic (CAD) system (CAD System-3) for detecting Liver disorders using histological features and imaging features.

Sub-Objectives:

1. To develop a CAD system based on histological features (CAD System-1).

2. To develop a CAD system based on imaging features (CAD System-2).

1.6 Literature Review

Visual criteria for diagnosing focal liver diseases from ultrasound images can be assisted by computerized tissue classification. Feature extraction algorithms are proposed in various studies to extract the tissue characterization parameters from liver images. The resulting parameter set is further processed to obtain the minimum number of parameters which represent the most discriminating pattern space for classification. This preprocessing step has been applied to distinct pathology-investigated cases to obtain the learning data for classification. The extracted features are divided into independent training and test sets, and are used to develop and compare both statistical and neural classifiers. The optimal criteria for these classifiers are set to have minimum classification error, ease of implementation and learning, and the flexibility for future modifications. Various algorithms of classification based on

statistical and neural network methods are presented and tested. The authors show that very good diagnostic rates can be obtained using unconventional classifiers trained on actual patient data.

Artificial Immune Recognition System (AIRS) classification algorithm, which has an important place among classification algorithms in the field of Artificial Immune Systems, has showed an effective and intriguing performance on the problems it was applied. This system, named as Fuzzy-AIRS was used as a classifier in the diagnosis of Liver Disorders, which are of great importance in medicine. The classifications of BUPA Liver Disorders datasets taken from University of California at Irvine (UCI) Machine Learning Repository were done using 10-fold cross-validation method. Reached classification accuracies were evaluated by comparing them with reported classifiers in UCI web site in addition to other systems that are applied to the related problems. Also, the obtained classification performances were compared with AIRS with regard to the classification accuracy, number of resources and classification time. Fuzzy-AIRS classified the Liver Disorders dataset with 83.36% accuracy. Fuzzy-AIRS obtained the highest classification accuracy according to the UCI web site. Beside of this success, Fuzzy-AIRS gained an important advantage over the AIRS by means of classification time. In the experiments, it was seen that the classification time in Fuzzy-AIRS was reduced about 70% of AIRS. By reducing classification time as well as obtaining high classification accuracies in the applied dataset, Fuzzy-AIRS classifier proved that it could be used as an effective classifier for medical problems.

In any of the research work reviewed, there has been no research on implementing a hybrid CAD system which includes both histological and imaging features.

1.7 Conclusion

In this chapter, the reason for selecting liver tissue was discussed and the various dataset used in the current work were explained. Based on the literature review done, the current work aims to test the performance of such a CAD System in classifying the liver tissues as hemangioma or hepatocellular carcinoma. In the next chapter, the methodology of the CAD System design for the current work is discussed.

Chapter 2

Methodology

2.1 Introduction

This chapter discusses the methodology that is used in this study. The first section describes the dataset representation. The next two sections describe the feature extraction and selection techniques. The feature classification section investigates the five classifiers that have been used in this research: K-nearest neighbor, Probabilistic neural network, Support vector machine, Smooth support vector machine, Neural network. This methodology plays an important role in implementing this research study accordingly. The details of the methodology are explained in detail in this chapter.

2.2 Collection and Pre-processing of dataset

A data set (or dataset) is a collection of data. Most commonly a data set corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. The data set may comprise data for one or more members, corresponding to the number of rows. The term data set may also be used more loosely, to refer to the data in a collection of closely related tables, corresponding to a particular experiment or event.

Data pre-processing is an important step in the data mining process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects. Data-gathering methods are often loosely controlled, resulting in out-of-range values (e.g., Income: −100), impossible data combinations (e.g., Sex: Male, Pregnant: Yes), missing values, etc. Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running an analysis.

If there is much irrelevant and redundant information present or noisy and unreliable data, then knowledge discovery during the training phase is more difficult. Data preparation and filtering steps can take considerable amount of processing time. Data pre-processing includes cleaning, normalization, transformation, feature extraction and selection, etc. The product of data pre-processing is the final training set.

2.3 Feature Extraction

In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be very redundant (e.g. the same measurement in both feet and meters, or the repetitiveness of images presented as pixels), then the input data will be transformed into a reduced representation set of features (also named feature vector). Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input.

Feature extraction involves reducing the amount of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which overfits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

2.4 Feature selection

In machine learning and statistics, feature selection, also known as variable selection, attribute selection or variable subset selection, is the process of selecting a subset of relevant features for use in model construction. The central assumption when using a feature selection technique is that the data contains many redundant or irrelevant features. Redundant features are those which provide no more information

than the currently selected features, and irrelevant features provide no useful information in any context.

Feature selection techniques are a subset of the more general field of feature extraction. Feature extraction creates new features from functions of the original features, whereas feature selection returns a subset of the features. Feature selection techniques are often used in domains where there are many features and comparatively few samples (or data points). The archetypal case is the use of feature selection in analyzing DNA microarrays, where there are many thousands of features, and a few tens to hundreds of samples. Feature selection techniques provide three main benefits when constructing predictive models:

- Improved model interpretability,
- Shorter training times,
- Enhanced generalization by reducing over fitting.

Feature selection is also useful as part of the data analysis process, as it shows which features are important for prediction, and how these features are related.

A feature selection algorithm can be seen as the combination of a search technique for proposing new feature subsets, along with an evaluation measure which scores the different feature subsets. The simplest algorithm is to test each possible subset of features finding the one which minimizes the error rate. This is an exhaustive search of the space, and is computationally intractable for all but the smallest of feature sets. The choice of evaluation metric heavily influences the algorithm, and it is these evaluation metrics which distinguish between the three main categories of feature selection algorithms: wrappers, filters and embedded methods.

Wrapper methods use a predictive model to score feature subsets. Each new subset is used to train a model, which is tested on a hold-out set. Counting the number of mistakes made on that hold-out set (the error rate of the model) gives the score for that subset. As wrapper methods train a new model for each subset, they are very computationally intensive, but usually provide the best performing feature set for that particular type of model.

2.5 Feature Classification

In machine learning and statistics, classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known.

In the terminology of machine learning, classification is considered an instance of supervised learning, i.e. learning where a training set of correctly identified observations is available. The corresponding unsupervised procedure is known as clustering, and involves grouping data into categories based on some measure of inherent similarity or distance.

Often, the individual observations are analyzed into a set of quantifiable properties, known variously explanatory variables, features, etc. These properties may variously be categorical ("A", "B" for blood type), ordinal ("large", "medium" or "small"), integer-valued (the number of occurrences of a part word in an email) or real-valued (a measurement of blood pressure). Other classifiers work by comparing observations to previous observations by means of a similarity or distance function.

An algorithm that implements classification, especially in a concrete implementation, is known as a classifier. The term "classifier" sometimes also refers to the mathematical function, implemented by a classification algorithm that maps input data to a category.

Terminology across fields is quite varied. In statistics, where classification is often done with logistic regression or a similar procedure, the properties of observations are termed explanatory variables (or independent variables, regresses, etc.), and the categories to be predicted are known as outcomes, which are considered to be possible values of the dependent variable. In machine learning, the observations are often known as instances, the explanatory variables are termed features (grouped into a feature vector), and the possible categories to be predicted are classes. There is also some argument over whether classification methods that do not involve a statistical model can be considered "statistical". Other fields may use different terminology: e.g. in community ecology, the term "classification" normally refers to cluster analysis, i.e. a type of unsupervised learning, rather than the supervised learning described in this article.

Following classifiers have been used in the current study:

2.5.1 Neural Network Classifier

In machine learning and related fields, ANN are computational models inspired by biological neural networks (the central nervous systems of animals, in particular the brain) and are used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown. Artificial neural networks are generally presented as systems of interconnected "neurons" which can compute values from inputs, and are capable of machine learning as well as pattern recognition thanks to their adaptive nature.

For example, a neural network for handwriting recognition is defined by a set of input neurons which may be activated by the pixels of an input image. After being weighted and transformed by a function (determined by the network's designer), the activations of these neurons are then passed on to other neurons. This process is repeated until finally, an output neuron is activated. This determines which character was read.

Like other machine learning methods - systems that learn from data - neural networks have been used to solve a wide variety of tasks that are hard to solve using ordinary rule-based programming, including computer vision and speech recognition.

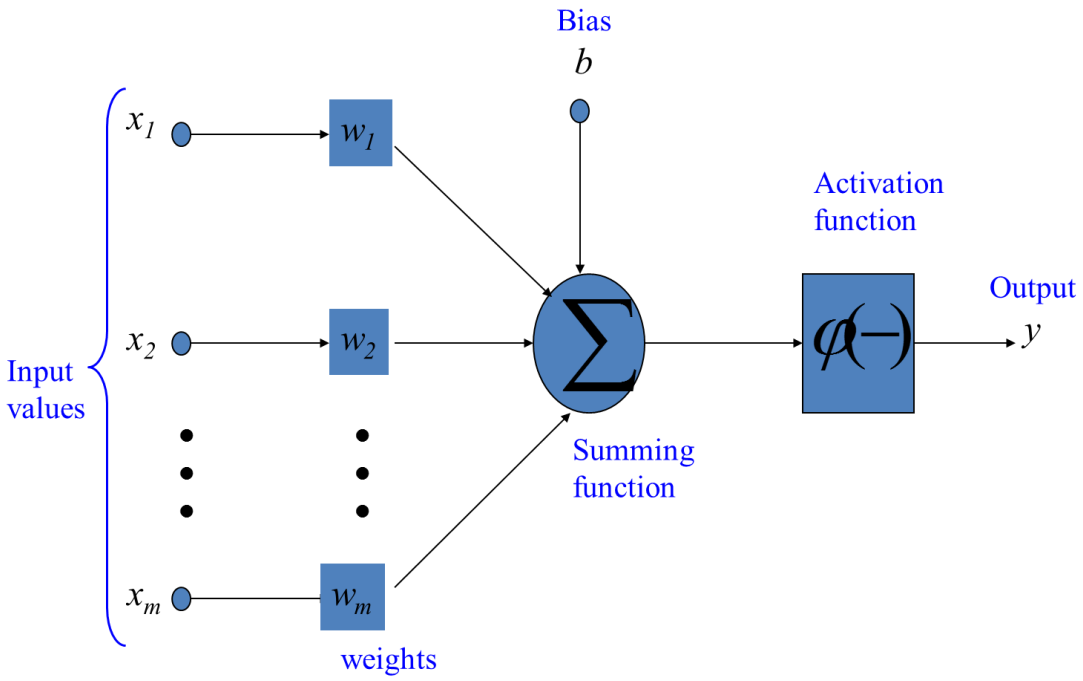

Fig. 2.1 Representation of an ANN

2.5.2 K-Nearest Neighbor Classifier:

In pattern recognition, the K-Nearest Neighbors algorithm (or KNN for short) is a non-parametric method used for classification and regression. In both cases, the input consists of the K closest training examples in the feature space. The output depends on whether KNN is used for classification or regression:

In KNN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its K nearest neighbors (K is a positive integer, typically small). If K = 1, then the object is simply assigned to the class of that single nearest neighbor.

In KNN regression, the output is the property value for the object. This value is the average of the values of its K nearest neighbors.

KNN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The KNN algorithm is among the simplest of all machine learning algorithms.

Both for classification and regression, it can be useful to weight the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. For example, a common weighting scheme consists in giving each neighbor a weight of 1/d, where d is the distance to the neighbor.

The neighbors are taken from a set of objects for which the class (for KNN classification) or the object property value (for KNN regression) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. A shortcoming of the KNN algorithm is that it is sensitive to the local structure of the data.

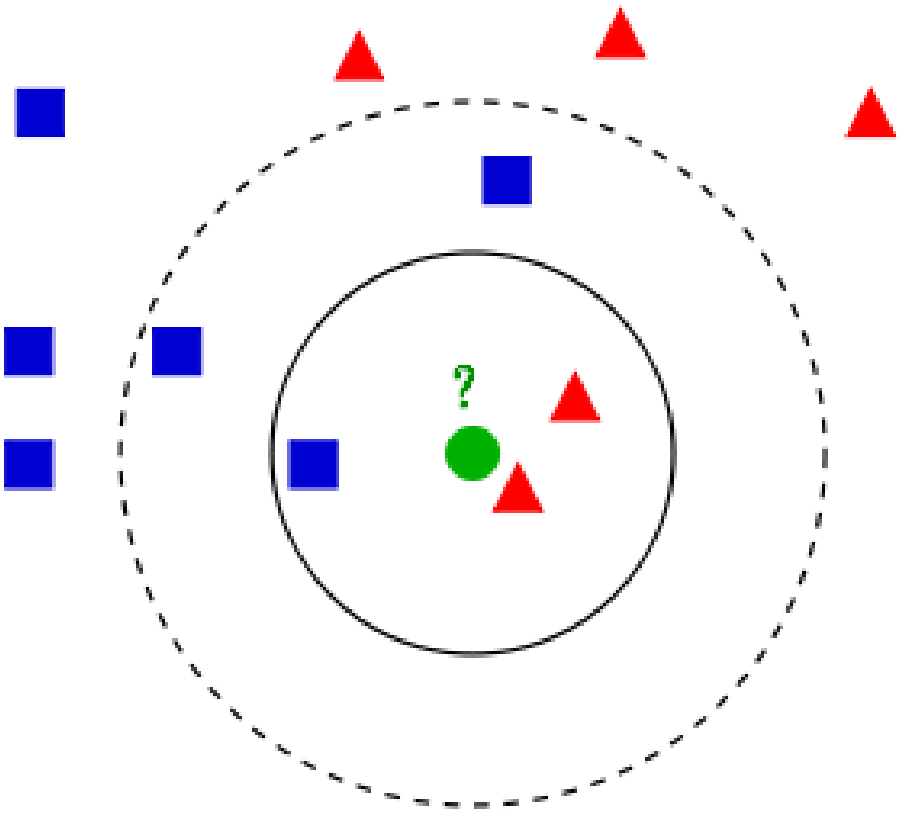

Fig. 2.2 Example of k-NN classification. The test sample (green circle) should be classified either to the first class of blue squares or to the second class of red triangles. If k = 3 (solid line circle) it is assigned to the second class because there are 2 triangles and only 1 square inside the inner circle. If k = 5 (dashed line circle) it is assigned to the first class (3 squares vs. 2 triangles inside the outer circle).

2.5.3 Probabilistic Neural Network:

A probabilistic neural network (PNN) is a feed-forward neural network, which was derived from the Bayesian network and a statistical algorithm called Kernel Fisher discriminant analysis. It was introduced by D.F. Specht in the early 1990s. In a PNN, the operations are organized into a multilayered feed forward network with four layers:

- **Input layer**

Each neuron in the input layer represents a predictor variable. In categorical variables, N-1 neurons are used when there are N numbers of categories. It standardizes the range of the values by subtracting the median and dividing by the interquartile range. Then the input neurons feed the values to each of the neurons in the hidden layer.

- **Pattern layer**

This layer contains one neuron for each case in the training data set. It stores the values of the predictor variables for the case along with the target value. A hidden neuron computes the Euclidean distance of the test case from the neuron's center point and then applies the RBF kernel function using the sigma values.

- **Summation layer**

For PNN networks there is one pattern neuron for each category of the target variable. The actual target category of each training case is stored with each hidden neuron; the weighted value coming out of a hidden neuron is fed only to the pattern neuron that corresponds to the hidden neuron's category. The pattern neurons add the values for the class they represent.

- **Output layer**

The output layer compares the weighted votes for each target category accumulated in the pattern layer and uses the largest vote to predict the target category.

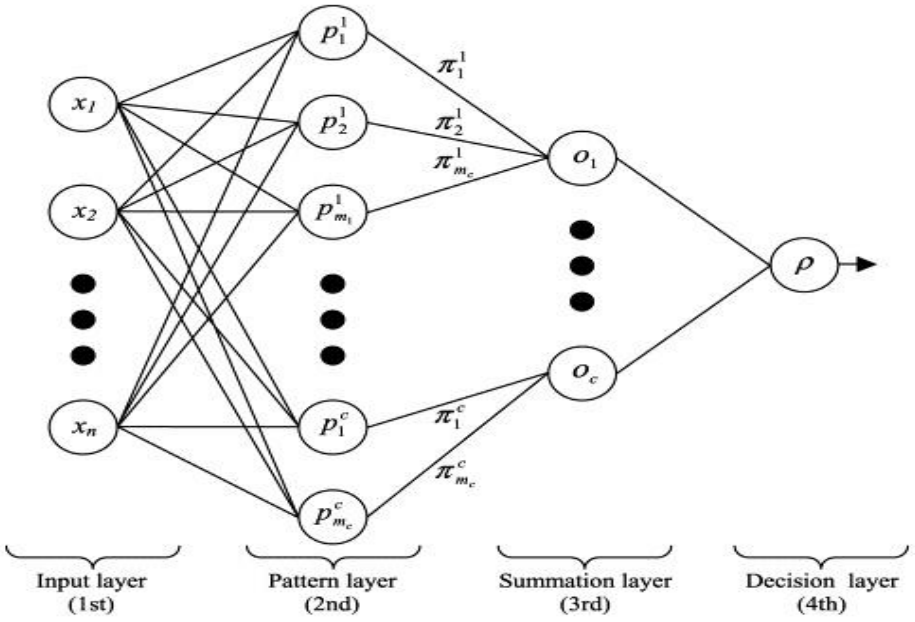

Fig. 2.3 A PNN structure that recognizes *c* classes

2.5.4 Support Vector Machine:

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that

analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

A support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

Whereas the original problem may be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space. To keep the computational load reasonable, the mappings used by SVM schemes are designed to ensure that dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function k(x,y) selected to suit the problem. The hyper planes in the higher-dimensional space are defined as the set of points whose dot product with a vector in that space is constant.

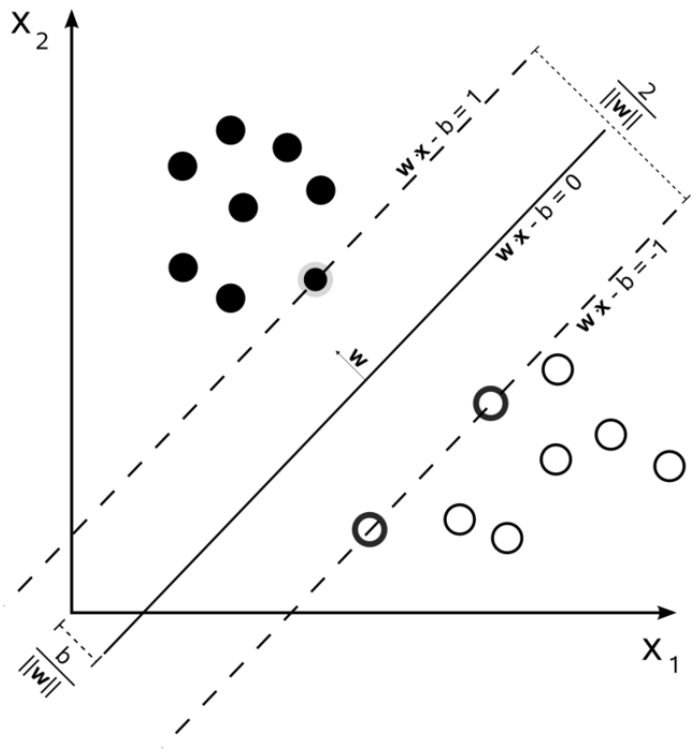

Fig. 2.4 Maximum-margin hyperplane and margins for an SVM trained with samples from two classes. Samples on the margin are called the support vectors.

2.5.5 Smooth Support Vector Machine (SSVM):

It is a new formulation of the support vector machine with linear and nonlinear kernel for pattern classification. The smooth support vector machine has important mathematical properties such as strong convexity and infinitely often differentiability. It uses Newton–Armijo algorithm to solve the SSVM and show that this algorithm globally and quadratically converges to the unique solution of the SSVM. SSVM gives higher tenfold cross validation correctness, thus making SSVM with a linear kernel very efficient for a large dataset.

2.6 Conclusion

After reading this chapter one can tell about the different feature reduction techniques that are used in this research. One can also have known all the classification methods that are being used to classify these features. Based on the features and that were discussed, the next two section i.e. Cad design using Histological features and have used the efficiencies of the classification methods mentioned in this chapter.

Chapter 3

Design of CAD System using Histological Features

3.1 Introduction

This chapter discusses the way in which the research has progressed from collection of dataset to the feature classifications [fig 3.1]. The first two sections have the details regarding the dataset that has been used. The next three sections describe the features and their extraction and selection techniques. The last section contains the efficiencies of the classifiers that have been used in this research: K-nearest neighbor, Probabilistic neural network, Support vector machine, Smooth support vector machine, Neural network.

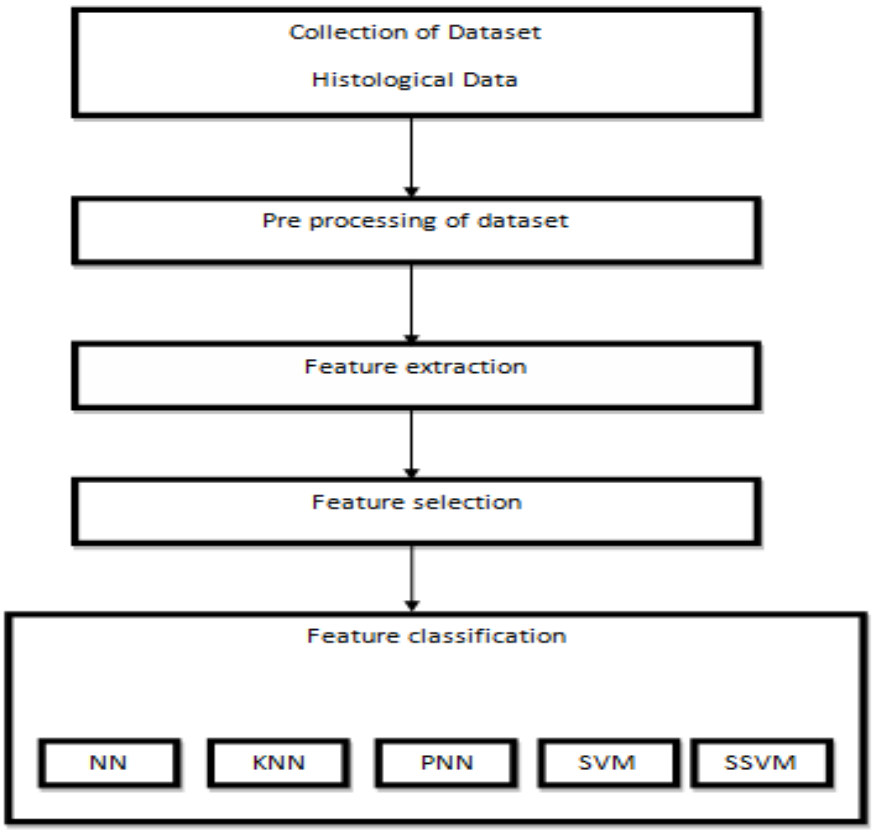

Fig. 3.1 Block Diagram of the CAD System 1

3.2 Collection of dataset

The database for liver disorders was taken from University of California, Irvine. This dataset is a standard dataset used globally by researchers working on classification of liver disorders.

The first 5 variables are all blood tests which are thought to be sensitive to Liver disorders that might arise from excessive alcohol consumption. Each line in the BUPA data file constitutes the record of a single male individual. It appears that if the number of drinks is more than five is some sort of a selector on this database.

3.3 Pre-processing of dataset

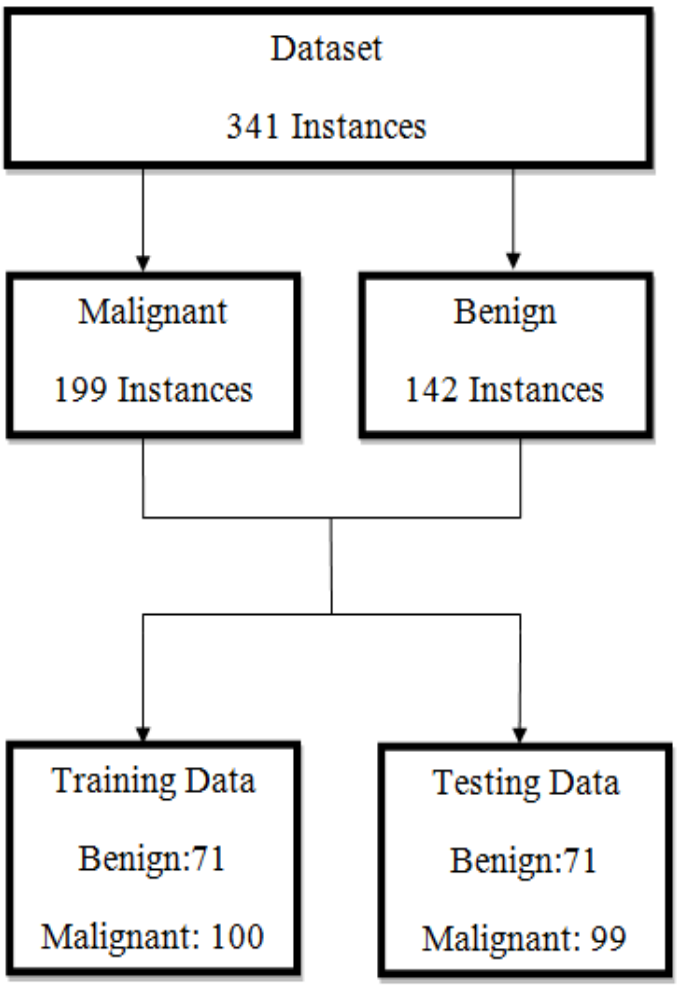

Fig. 3.2 Description of dataset

Table 3.1: Attribute information of Histological Features from BUPA UCI database for Liver disorders

Sr. No.	Attribute	Information
1	Mean Corpuscular Volume	A measure of the average volume of red blood cells. Normal values: MCV: 80 to 95 femtolitre.
2	Alkaline Phosphotase	An enzyme found in bloodstream. ALP helps break down proteins in the body and exists in different form. Normal range 20 to 140 IU/L.
3	Alamine Aminotransferase	The most commonly used indicators of liver damage enzymes normally found in liver cells. Normal Range: Female≤ 34 IU/L, Male≤ 45 IU/L
4	Aspartate Aminotransferase	Normally found in red blood cells, liver, heart, muscle tissue, pancreas, and kidneys. Normal Range: Female 6 - 34 IU/L, Male 8 - 40 IU/L
5	Gamma-Glutamyl Transpeptidase	Primarily present in kidney, liver, and pancreatic cells. GGT activity is elevated in any and all forms of liver disease. Normal range 7- 35 u/l
6	Drinks	Number of half-pint equivalents of alcoholic beverages drunk per day
7	Selector Field	Used to split data into two sets

3.4 Feature Extraction

The first CAD system design uses histological features, which are available in the form of benchmark database available for researchers globally. This database contains the data in form of features in numerical form, which can directly be used for classification. Thus, this system does not require any sort of feature extraction module.

3.5 Feature selection

Feature selection is used to select most efficient features out of a large bucket to reduce the computation time while maintaining the efficacy of the system. In this case, there are only six features and further selection may lead to reduction in efficiency, therefore, this module is skipped.

3.6 Feature Classification

The dataset was classified using the following classification techniques

- Neural Network
- k-Nearest Neighbour
- Probabilistic Neural Networks
- Support Vector Machines
- Smooth Support Vector Machine

The performance obtained by the above mentioned classification techniques on the BUPA UCI Liver disorder dataset is mentioned in the following table [Table 3.2].

Table 3.2: Performance of classification by NN, kNN, PNN, SVM and SSVM

Classifier	Confusion Matrix (CM)	Classification Accuracy (CA)	S_B	S_M
Neural Network (NN)	B: 30, 41 M: 12, 87	68.82% after training the network 4th time	42.25%	87.88%
K- Nearest Neighbors (kNN)	B: 39, 32 M: 15, 84	72.35% for k=9 with City-block distance	54.93%	84.85%
Probabilistic Neural Network (PNN)	B: 40, 31 M: 23, 76	68.24% for spread=6	56.34%	76.77%
Support Vector Machine (SVM)	B: 42, 29 M: 17, 82	72.94%	59.15%	82.83%
Smooth Support Vector Machine (SSVM)	B: 51, 20 M: 11, 88	81.76%	71.83%	88.89%

Note: S_B: Sensitivity to Benign; S_M: Sensitivity to Malignant; B: Benign; M: Malignant

3.7 Conclusion

After reading this chapter one can tell about the different features and their attribute information of these features. Also the best method of classification can also be recognized using the classification efficiency table of the 5 classifiers used in this section. The maximum efficiency obtained was 88.89 in the case of Smooth support vector machines.

Chapter 4

CAD Design using Imaging Features

4.1 Introduction

This chapter discusses the way in which the research on imaging has progressed from collection of dataset to the feature classifications [fig 4.1]. The first two sections have the details regarding the dataset that has been used. The next three sections describe the features and their extraction and selection techniques. The last section contains the efficiencies of the classifiers that have been used in this research: K-nearest neighbor, Probabilistic neural network, Support vector machine, Smooth support vector machine, Neural network.

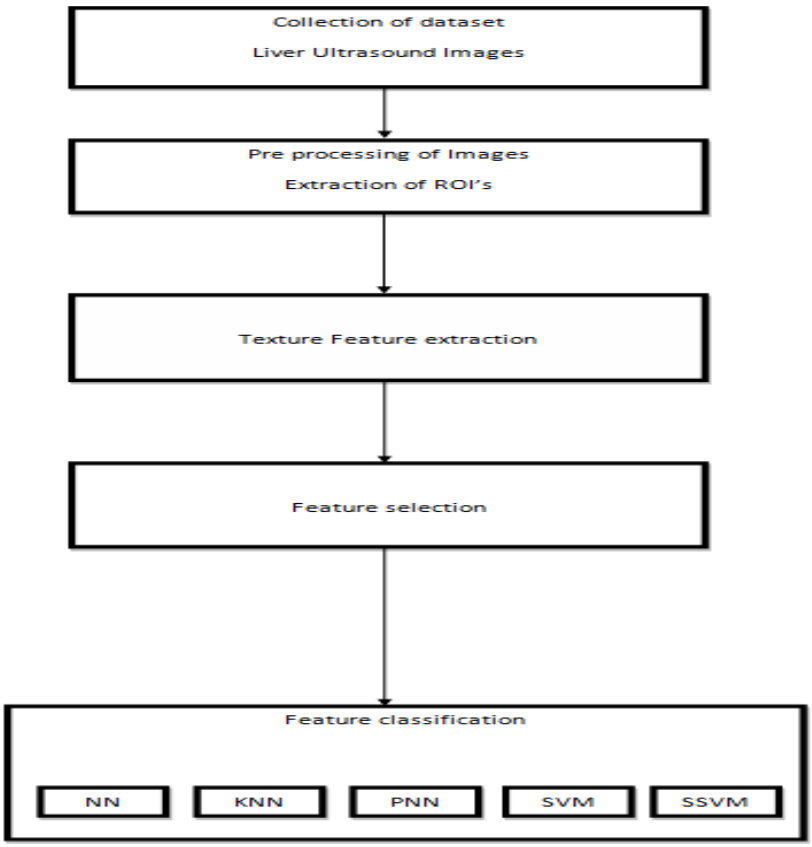

Fig. 4.1 Block Diagram of the CAD System 2

4.2 Collection of dataset

The dataset used in the current design was the Ultrasound images, which were acquired from Post Graduate Institute of Medical Education and Research, Chandigarh. Ultrasound images were used because they are much cheaper than MRI Images and do not cause any harm to the tissue as X-Ray Images do.

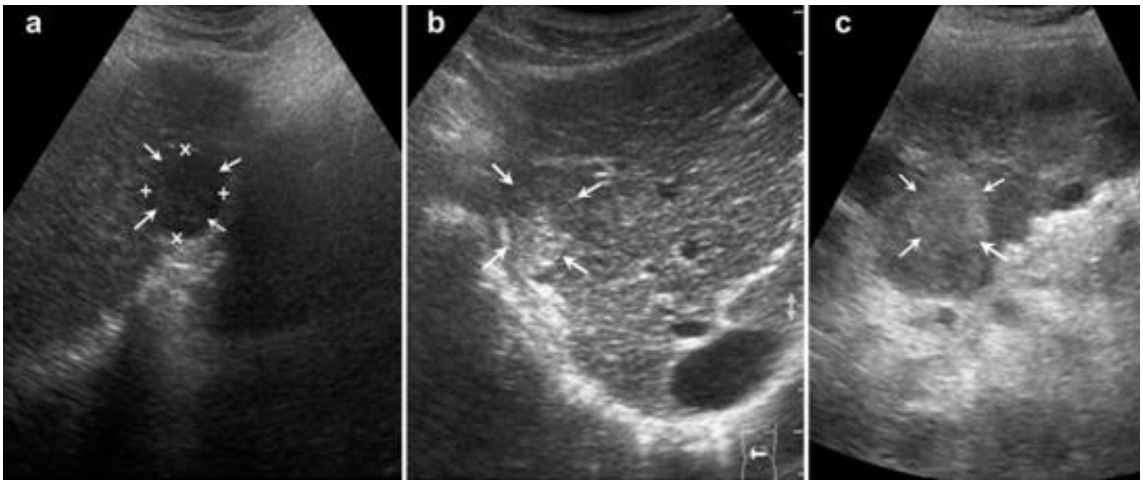

Fig. 4.2 Ultrasound Images of Liver Tissue showing (a) Hypo Echoic Echogenicity (b)Mixed Echogenicity (c) Hyper Echoic Echogenicity

4.3 Pre-processing of dataset

The region of interest (ROI) from the images were extracted using centroid based approach.

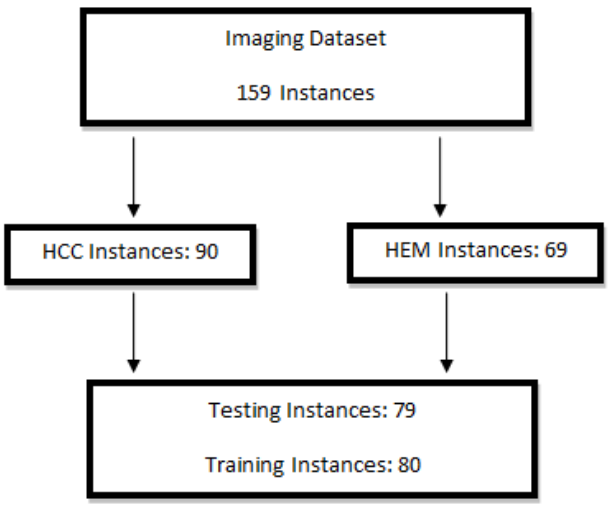

Fig. 4.3 Description of dataset

4.4 Feature Extraction

The Laws' Masks were then applied to extract the features necessary for classification.These ROIs were subject to various filters with specific masks to extract the five required fields- Level (L), Edge (E), Spot (S), Ripple (R) and Wave (W). These masks are created by pre-defined 1-Dimentional kernel vectors. In the current work, the masks of resolution three, five, seven and nine were used.

The vectors used were:

For resolution three:

$$L3 = [+1, \ +\ 2, \ +1]$$

$$E3 = [-1, \quad 0, \ +1]$$

$$S3 = [-1, \ +2, \ -1]$$

The 2-D masks formed from these vectors are

L3L3	E3L3	S3L3
L3E3	E3E3	S3E3
L3S3	E3S3	S3S3

For resolution five:

$$L5 = [+1,+4,+6,+4,+1]$$

$$E5 = [-1,-2,0,+2,+1]$$

$$S5 = [-1, 0, +2, 0, -1]$$

$$W5 = [-1,+2, 0,-2,+1]$$

$$R5 = [+1,-4,+6,-4,+1]$$

The 2-D masks formed from these vectors are

L5L5 E5L5 S5L5 W5L5 R5L5

L5E5 E5E5 S5E5 W5E5 R5E5

L5S5 E5S5 S5S5 W5S5 R5S5

L5W5 E5W5 S5W5 W5W5 R5W5

L5R5 E5R5 S5R5 W5R5 R5R5

For resolution seven:

$$\mathrm{L7} = [1,\ 6,\ 15,\ 20,\ 15,\ 6,\ 1]$$

$$\mathrm{E7} = [-1,\ -4,\ -5,\ 0,\ 5,\ 4,\ 1]$$

$$\mathrm{S7} = [-1,\ -2,\ 1,\ 4,\ 1,\ -2,\ -1]$$

The 2-D masks formed from these vectors are

L7L7	E7L7	S7L7
L7E7	E7E7	S7E7
L7S7	E7S7	S7S7

For resolution nine:

$$\mathrm{L9} = [1,\ 8,\ 28,\ 56,\ 70,\ 56,\ 28,\ 8,\ 1]$$

$$\mathrm{E9} = [1,\ 4,\ 4,\ -4,\ -10,\ -4,\ 4,\ 4,\ 1]$$

$$\mathrm{S9} = [1,\ 0,\ -4,\ 0,\ 6,\ 0,\ -4,\ 0,\ 1]$$

$$\mathrm{W9} = [1,\ -4,\ 4,\ -4,\ -10,\ 4,\ 4,\ -4,\ 1]$$

$$\mathrm{R9} = [1,\ -8,\ 28,\ -56,\ 70,\ -56,\ 28,\ -8,\ 1]$$

The 2-D masks formed from these vectors are

L9L9 E9L9 S9L9 W9L9 R9L9

L9E9 E9E9 S9E9 W9E9 R9E9

L9S9 E9S9 S9S9 W9S9 R9S9

L9W9 E9W9 S9W9 W9W9 R9W9

L9R9 E9R9 S9R9 W9R9 R9R9

Steps followed in Laws' mask analysis are:

- Convolving the image I(i,j) with each 2-D mask forming a texture image (TI) . e.g.

$$\mathrm{TI}_{\mathrm{E5E5}} = \mathrm{I}_{\mathrm{i,j}} \otimes \mathrm{E5E5}$$

- Normalizing the contrast of texture image.

$$\mathrm{Normalize}(\mathrm{TI}_{\mathrm{mask}}) = \frac{\mathrm{TI}_{\mathrm{mask}}}{\mathrm{TI}_{\mathrm{L5L5}}}$$

- The TIs are passed through Texture Energy Measurement (TEM) filters.

$$\mathrm{TEM}_{\mathrm{i,j}} = \sum_{\mathrm{u}=-7}^{7} \sum_{\mathrm{v}=-7}^{7} \left[\mathrm{Normalize}\left(\mathrm{TI}_{\mathrm{i+u,j+v}}\right)\right]$$

- By combining 25 TEM descriptors we obtain 15 rotationally invariant TEMs denoted as TR.

$$\mathrm{TR}_{\mathrm{E5L5}} = \frac{\mathrm{TEM}_{\mathrm{E5L5}} + \mathrm{TEM}_{\mathrm{L5E5}}}{2}$$

- From each TR five statistical parameters are obtained namely: Mean, Standard deviation (SD), Skewness, Kurtosis, Entropy

$$\mathrm{Mean} = \frac{\sum_{\mathrm{i}=0}^{\mathrm{M}} \sum_{\mathrm{j}=0}^{\mathrm{N}} [\mathrm{TR}_{\mathrm{i,j}}]}{\mathrm{M} \times \mathrm{N}}$$

$$\mathrm{SD} = \sqrt{\frac{\sum_{\mathrm{i}=0}^{\mathrm{M}} \sum_{\mathrm{j}=0}^{\mathrm{N}} (\mathrm{TR}_{\mathrm{i,j}} - \mathrm{Mean})^2}{\mathrm{M} \times \mathrm{N}}}$$

$$\mathrm{Skewness} = \frac{\sum_{\mathrm{i}=0}^{\mathrm{M}} \sum_{\mathrm{j}=0}^{\mathrm{N}} (\mathrm{TR}_{\mathrm{i,j}} - \mathrm{Mean})^3}{\mathrm{M} \times \mathrm{N} \times \mathrm{SD}^3}$$

$$\mathrm{Kurtosis} = \frac{\sum_{\mathrm{i}=0}^{\mathrm{M}} \sum_{\mathrm{j}=0}^{\mathrm{N}} (\mathrm{TR}_{\mathrm{i,j}} - \mathrm{Mean})^4}{\mathrm{M} \times \mathrm{N} \times \mathrm{SD}^4} - 3$$

$$\text{Entropy} = \frac{\sum_{i=0}^{M}\sum_{j=0}^{N}(TR_{i,j})^2}{M \times N}$$

4.5 Feature Classification

The dataset was classified using the classification techniques mentioned in chapter 2 and the performance is analyzed in the table [Table 4.1].

Table 4.1: Performance of classification by NN, kNN, PNN, SVM and SSVM

Classifier	Confusion Matrix (CM)	Classification Accuracy (CA)	S_B	S_M
Smooth Support Vector Machine (SSVM) using LAWS 3 mask	B: 42, 3 M: 17, 17	74.68%	93.33%	50.00%
Smooth Support Vector Machine (SSVM) using LAWS 5 mask	B: 43, 2 M: 13, 21	81.01%	95.56%	61.76%
Smooth Support Vector Machine (SSVM) using LAWS 7 mask	B: 44, 1 M: 29, 5	62.02%	97.78%	14.70%
Smooth Support Vector Machine (SSVM) using LAWS 9 mask	B: 41, 4 M: 25, 9	63.29%	91.11%	26.47%

Note: S_B: Sensitivity to Benign; S_M: Sensitivity to Malignant; B: Benign; M: Malignant

4.6 Conclusion

After reading this chapter one can tell about the different features and their attribute information of these features. Also the best method of classification can also be recognized using the classification efficiency table of the 5 classifiers used in this section. The maximum efficiency obtained was 97.78 in the case of Smooth support vector machine using LAWS 7 mask

Chapter 5

Hybrid CAD Design combining both Histological and Imaging Features

5.1 Introduction

This chapter discusses the way in which the research on both Histological and Imaging has progressed from collection of dataset to the feature classifications [fig 5.1]. The first sections has the details regarding the dataset that has been used. The next sections describe the features extraction technique. The last section contains the maximum efficiencies obtained by combining the Histological and Imaging results.

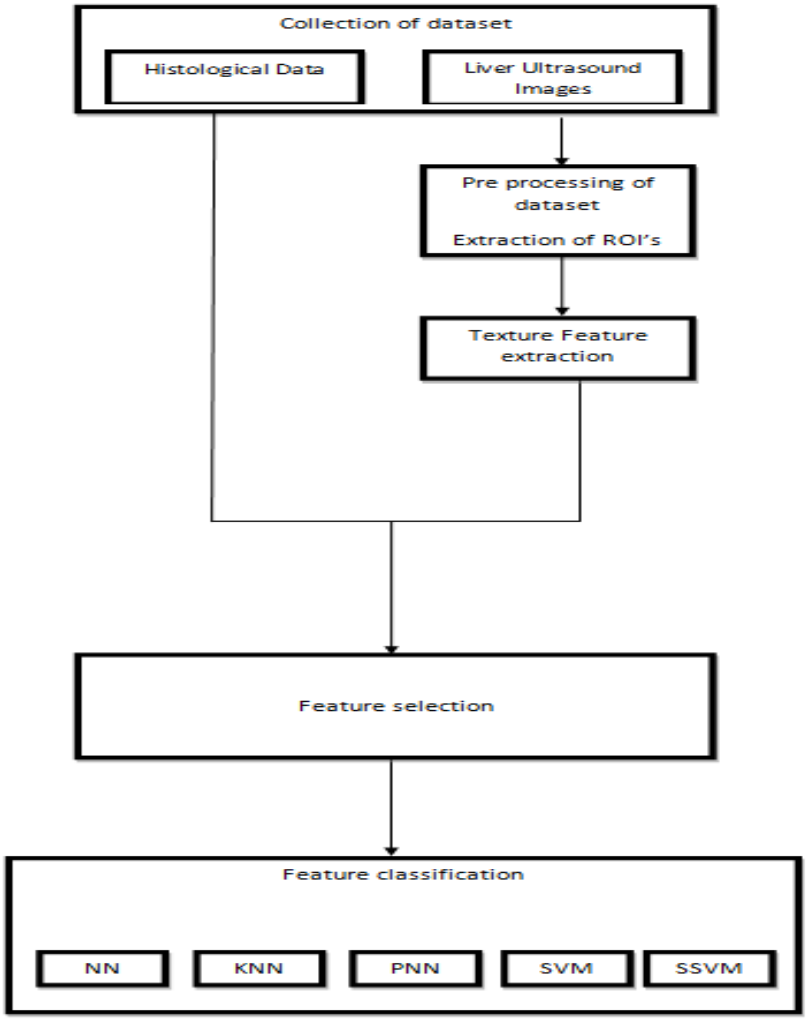

Fig. 5.1 Block Diagram of the Hybrid CAD System

5.2 Collection of dataset

The dataset used in the current design was the mix of histological features and the Imaging features extracted in chapters 3 and 4 respectively.

5.3 Feature Extraction

This CAD system design uses the hybrid features, which were created as part of CAD Design 1 and CAD design 2. This database contains the data in form of features in normalized numerical form, which can directly be used for classification.

5.4 Feature Classification

The dataset was classified using the classification techniques mentioned in chapter 2.

Table 4: The efficiencies obtained using various classification tools

Classifier	**Confusion Matrix (CM)**			**Classification Accuracy (CA)**	**S_B**	**S_M**
Smooth Support Vector Machine (SSVM) using LAWS 3 mask and histological features		B	M	79.74%	95.55%	58.82%
	B	43	2			
	M	14	20			
Smooth Support Vector Machine (SSVM) using LAWS 5 mask and histological features		B	M	86.07%	95.55%	73.52%
	B	43	2			
	M	9	25			

Note: S_B: Sensitivity to Benign; S_M: Sensitivity to Malignant; B: Benign; M: Malignant

5.5 Conclusion

After reading this chapter one can tell about the best results obtained from combing the histological and imaging features. Also the best efficiency obtained when Histological and Imaging features have been combined is 95.55 in 2 cases (i) Smooth Support Vector Machine (SSVM) using LAWS 3 mask and histological features (ii) Smooth Support Vector Machine (SSVM) using LAWS 5 mask and histological features.

Chapter 6

Challenges, Conclusion and Future Work

5.1 Challenges

The current research work involved a lot of complex tasks and some difficulties that required extensive hard work and extreme dedication. The following were the main challenges faced:

- Collection of data: The data of patients is classified and cannot be shared with anyone, even for research. So, we had to make do with the standard BUPA dataset available for researchers globally. While, the imaging database was collected from PGIMER. This created a lot of problems as the data was not of the same patient, thus, deterioration the results obtained.
- Application of NN: The nntool in MATLAB creates an ANN which cannot be saved. So, while applying the ANN, we had to save the data again and again. This data would change after every run as the network sometimes get stuck in the local minima and cannot achieve the global minima.
- Integrating SVM toolbox: The SVM classification technique required the integration of SVM toolbox, which required extensive research as no literature is available for this purpose. This procedure is very complex in itself as it requires several installations and may even lead to a hard disk failure.
- Application of SSVM: Not much research has been done on SSVM. However, to improve the efficiency of our design we had to apply this in any case. For this not much literature is available, so it involved extensive amounts of calculations and coding.

5.2 Conclusion and Future Work

The current work incorporating the features of both histological and imaging data improved upon the classification efficiency thus making the designed CAD system more efficient and accurate.

Due to the paucity of time, the current work could not include the Feature Extraction and Selection modules in many places. However, the number of features in the CAD

design using Imaging features is large, so it requires some sort of feature extraction methods such as:

- Principal Component Analysis: It is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. This transformation is defined in such a way that the first principal component has the largest possible variance, and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. The principal components are orthogonal because they are the eigenvectors of the covariance matrix, which is symmetric. PCA is sensitive to the relative scaling of the original variables.
- Kernel PCA: In the field of multivariate statistics, kernel principal component analysis is an extension of principal component analysis using techniques of kernel methods. Using a kernel, the originally linear operations of PCA are done in a reproducing kernel Hilbert space with a non-linear mapping.

Besides, various Feature Selection methods can also be used such as:

- Genetic Algorithm: In the field of artificial intelligence, GA is a search heuristic that mimics the process of natural selection. This heuristic is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover.

References

[1] T.S. Subhashini et. al., Automated assessment of liver tissue in digital legions, Computer Vision and Image Understanding 114 (2010) 33-43.

[2] P. Zhang et. al., Neural vs statistical classifier in conjunction with genetic algorithm feature selection in liver tissues, IEEE Congress on Evolutionary Computation 2, 1206-1213.

[3] www.archive.ics.uci.edu/datasets/liver+disorders

[4] www.csie.ntu.edu.tw/~cjlin/libsvm

[5] http://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

[6] www.cse-wiki.unl.edu/wiki/index.php/probabilistic_neural_network

[7] www.saedsayad.com/k_nearest_neighbors.htm

[8] http://en.wikipedia.org/wiki/Artificial_neural_network

[9] Lee, Y.J. and O.L. Mangasarian, 2001. A smooth support vector machine. J. Comput. Optimiz. Appli., 20: 5-22. DOI: 10.1023/A:1011215321374.

[10] http://www.ics.uci.edu/~mlearn/~MLRepository.html

[11] L. Kaufman, "Solving the quadratic programming problem arising in support vector classification," in Advances in Kernel Methods—Support Vector Learning, Bernhard Sch¨olkopf, Christopher J.C. Burges, and Alexander J. Smola (Eds.), MIT Press: Cambridge, MA, 1999, pp. 147–167.

[12] Kadah, Y.M "Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images", IEEE Transactions on Medical Imaging, (Volume:15 , Issue: 4) 466 – 478

[13] Kemal Polata, Seral Şahana, "Breast cancer and liver disorders classification using artificial immune recognition system (AIRS) with performance evaluation by fuzzy resource allocation mechanism", Expert Systems with Applications, Volume 32, Issue 1, January 2007, Pages 172–183

[14] K. J. Parker, "Ultrasonic attenuation and absorption in liver tissue", Ultrasound Med. Biol., vol. 9, pp.363 -369 1983

[15] D. Schlaps, et al., S. L. Bachrach, "Ultrasonic tissue characterization using a diagnostic expert system", Information Processing in Medical Imaging, 1986 :Martinus Nijhoff

[16] A. M. Youssef and A. A. Sharawi, "KSODATA clustering analysis for diffuse liver diseases", Proc. IEEE Symp. Ultrasound, 1990

[17] D. A. Christensen, Ultrasound Bioinstrumentation., 1988 :Wiley

[18] C. B. Burchardt, "Speckle in ultrasound B-mode scans", IEEE Trans. Sonics Ultrason., vol. SU-25, pp.1 -6 1978

[19] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, "Statistics of speckle in ultrasound B-scans", IEEE Trans. Sonics Ultrason., vol. SU-30, pp.156 -163 1983

[20] M. F. Insana, B. S. Garra, D. G. Brown, and T. S. Shawker, "Analysis of ultrasound images via generalized Rician statistics", Opt. Eng., vol. 25, pp.743 -748 1986

[21] E. Walach, A. Shmulewitz, Y. Itzchak, and Z. Heyman, "Local tissue attenuation images based on pulsed-echo ultrasound scans", IEEE Trans. Biomed. Eng., vol. 36, no. 2, pp.211 -220 1989

[22] S. Fields and F. Dunn, "Correlation of echographic visualizability of tissue with biological composition", J. Acoust. Soc. Amer. & mdash, Med. pt. I: Basic principles, special and biological state, vol. 54, pp.809 -812

[23] B. S. Garra, In Vivo Liver and Splenic Tissue Characterization by Scattering

[24] B. B. Gosnik, S. K. Lemon, W. Scheible, and G. R. Leupold, "Accuracy of ultrasonography in diagnosis of hepatocellular disease", AJR, vol. 133, pp.19 -23 1979

[25] A. Sharawi, Quantitative tissue characterization parameters for liver diseases, 1990 :Systems and Biomed. Eng. Dept, Faculty of Eng., Cairo Univ.

[26] Y. M. Kadah, A. A. Farag, A. M. Youssef, and A. S. Badawi, et al., "Statistical and neural classifiers for ultrasound tissue characterization", Proc. ANNIE-93, Artificial Neural Networks in Engineering, 1993