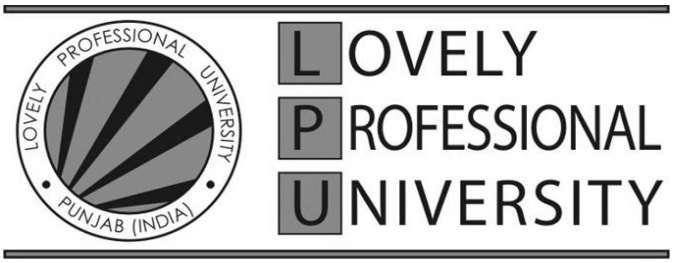

COMPUTER-AIDED DETECTION OF FOLLICULAR LYMPHOMA GRADING USING SUPPORT VECTOR MACHINE

A Dissertation Proposal

Submitted

By

Kirandeep Kaur

(11307315)

To

Department of Computer Science & Engineering

In partial fulfillment of the Requirement for the

Award of the Degree of

Master of Technology in Computer Science

Under the Guidance of

Mr. Sanjay Kumar Singh

(Asst. Professor, LPU)

May, 2015

PAC APPROVAL FORM

LOVELY PROFESSIONAL UNIVERSITY
Transforming Education Transforming India

School of: Computer Science and Engineering

DISSERTATION TOPIC APPROVAL PERFORMA

Name of the Student: Kirandeep Kaur	Registration No: 11307315
Batch: 2013-2015	Roll No. RK2305A15
Session: 2014-2015	Parent Section: K2305
Details of Supervisor:	Designation: Asst. Professor
Name: Sanjay Kumar Singh	Qualification: Mtech-CSE
U.ID : 15745	Research Experience: 3 yrs

SPECIALIZATION AREA: Intelligent System **(pick from list of provided specialization areas by DAA)**

PROPOSED TOPICS

1. Computer aided detection of centroblasts for follicular lymphoma grading using support vector machine.
2. Grading of histopathological image
3. Classification of lymphoma cells.

Signature of Supervisor 15/9/15

PAC Remarks:

Topic 1 is approved. Research Paper is also attached.

APPROVAL OF PAC CHAIRPERSON: Signature: 18/9/14 **Date:** 18/9/14

*Supervisor should finally encircle one topic out of three proposed topics and put up for approval before Project Approval Committee (PAC)

*Original copy of this format after PAC approval will be retained by the student and must be attached in the Project/Dissertation final report.

*One copy to be submitted to Supervisor.

ABSTRACT

Grading of follicular lymphoma is very important in order to treat this disease, as the grade specifies that how fast the lymphoma is spreading in the body. Depending on the grade whether it is high, moderate or low grade lymphoma, the treatment is given to the patient. This computer-aided system has been designed to help the doctors to diagnose/detect the grade of lymphoma in much lesser time and with better performance. This system performs grading of follicular lymphoma, that is, it detects the grades corresponding to each attribute feature of clinical data and also detects grades with all attributes combined. This system finds that which kernel function is best for classification taking different types of input data and data features. This system also shows that which clinical attribute is more important to predict a particular grade, that is, it tells the best attribute predictor for each grade.

ACKNOWLEDGEMENT

I would like to present my deepest gratitude to **Mr. Sanjay Kumar Singh** for his guidance, advice, understanding and supervision throughout the development of this dissertation study. I would like to thank the **Project Approval Committee members** for their valuable comments and discussions. I would also like to thank **Lovely Professional University** for the support on academic studies and letting me involve in this study.

Kirandeep Kaur

Reg. No.: 11307315

DECLARATION

I hereby declare that the dissertation proposal entitled, ***Computer-Aided Detection of Follicular Lymphoma Grading using Support Vector Machine*** submitted for the M. Tech Degree is entirely my original work and all ideas and references have been duly acknowledged. It does not contain any work for the award of any other degree or diploma.

Date:

Kirandeep Kaur

Reg. No: 11307315

CERTIFICATE

This is to certify that **Kirandeep Kaur** has completed M. Tech dissertation proposal titled ***Computer-Aided Detection of Follicular Lymphoma Grading using Support Vector Machine*** under my guidance and supervision. To the best of my knowledge, the present work is the result of her original investigation and study. No part of the dissertation proposal has ever been submitted for any other degree or diploma.

The dissertation proposal is fit for the submission and the partial fulfillment of the conditions for the award of M. Tech in Computer Science & Engineering.

Date:

Signature of Advisor

Name:

UID:

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1
INTRODUCTION

Lymphoma is a blood cancer of lymphocytes which are one of the types of white blood cells. Lymphocytes travel in body through lymph nodes, thymus, bone marrow and spleen that constitutes the lymphatic system. Lymphoma occurs when lymphocytes grow and multiply uncontrollably and form a solid mass which we call as tumor and it mainly affects the immune system of our body. The two main types of lymphoma are as follows:

- Hodgkin Lymphoma
- Non-Hodgkin Lymphoma

The Hodgkin lymphoma and non-Hodgkin lymphoma both originate in one of the types of white blood cell which we call as a lymphocyte which is an important component of our body's immune system. Both have similar symptoms but the conditions are different. The root of lymphoma is not known till now but there are few factors which have been identified. With help of biopsy it is determined whether the lymphoma is Hodgkin lymphoma or non-Hodgkin lymphoma. When biopsy result shows Reed-Sternberg cell in the tissue sample, that lymphoma is Hodgkin lymphoma and otherwise it is non-Hodgkin lymphoma. There are 2 types of lypmhocytes:

- B lymphocytes
- T lymphocytes

B lymphocytes are B cells which shelter our body against viruses and bacterias that is germs by producing antibodies. T lymphocytes are T cells, some of which directly destroy the cells which are infected with viruses, fungi, or bacteria.

1.1 Hodgkin Lymphoma (HL)

In Hodgkin's lymphoma, B-cells are affected. In this, Reed-Sternberg cells are found in the tissues which are irregular form of B-cells. The different kind of Hodgkin's lymphoma can be classified by differences seen under the microscope. Following are some risk factors:

- Infection with Epstein-Barr virus
- People in their 20s and people above 55 years

- A sibling having this lymphoma
- HIV

Types of Hodgkin lymphoma:

- Nodular sclerosing
- Mixed cellularity
- Lymphocyte rich, etc.

1.2 Non-Hodgkin Lymphoma (NHL)

In Non-Hodgkin's Lymphoma, B-cells are affected along with T-cells. These lymphocytes have their individual functions in providing immunity to the human body. Following are some risk factors:

- Mostly in people 60 years and above
- Some chemicals used in agriculture and nuclear radiation exposure to the body
- HIV
- A disease in which own cells of body are attacked by the immune system
- Certain viral and bacterial infections

Types of non-Hodgkin Lymhoma:

- B-cell lymphomas
    - Follicular Lymphoma
    - Mantle Cell Lymphoma, etc.
- T-cell lymphomas
    - Precursor T-lymphoblastic Lymphoma
    - Peripheral T-cell Lymphoma

1.3 Follicular Lymphoma

This is the most widespread category of Non-Hodgkin lymphoma. It is B-cell type of lymphoma in which B-lymphocytes develop into the lymphoma. In this, the B-lymphocytes develop in circular pattern and these lymphoma cells group together in clusters in a lymph node or other tissues to form a tumor. These enlarged irregular B-cells are called

centroblasts. This lymphoma is diagnosed by examining the shapes of centroblasts under microscope.

Follicular lymphoma is basically slow-growing type of lymphoma, takes several years to develop in the body. It typically occur in adults with age 60 years and older and is not commonly found in young people. Sometimes follicular lymphoma transforms into a fast-growing (aggressive) diffuse large B-cell lymphoma (DLBCL). It is slightly more commonly found in women than men and found very less in Asians and Blacks as compared to people of other ethnicities.

This disease is not inherited one. Children of the patient with this disease do not have any threat of developing it.

1.3.1 Follicular lymphoma symptoms

Follicular lymphoma grows slowly in the body in majority of cases and the mean time in which symptoms build up, the stage of lymphoma reaches 3rd or 4rth stage . There are various symptoms with which we can come to know that a person is having follicular lymphoma or not. People may have swollen lymph nodes, the size may increase or decrease over time. There can be blockage in normal flow in the digestion system , excretory system and blood vessels.

- Swelling of lymph nodes mainly in neck, groin areas and armpits
- High fevers that may come and go
- Heavy sweating at night
- Sudden weight loss with no other disease diagnosed
- Itching may occur and gets worse in some people after intake of alcohol
- Acute cough and difficulty in taking proper breaths
- Ache in abdomen
- Vomits after intake of alcohol
- No desire to eat

1.3.2 Follicular Lymphoma Diagnosis

This lymphoma can be diagnosed by taking whole or a piece of the inflated lymph node to study the cells. This procedure is called as biopsy. In this, the tissue sample is stained and visualized under the microscope to examine that grade of lymphoma and at which stage the

lymphoma has spread in the body. After diagnosis has been done some additional tests are performed to get more information about the extent to which lymphoma has spread in the body like Blood tests, Bone marrow biopsy, Computed tomography (CT) scan etc.

1.3.3 Follicular Lymphoma Staging

Staging tells that how much part of the lymphatic system has been affected by the lymphoma. It helps to decide whether treatment is required, what kind of treatment is needed. Following are the stages:

- Stage 1 – Only one lymph node region (neck, armpit, groin or chest) or only one lymph structure (lymph nodes, spleen, and thymus gland) is affected.
- Stage 2 – lymphoma has spread to two or more regions or structures of the lymph nodes on one side of diaphragm.
- Stage 3 – lymphoma has spread in structures or regions of lymph nodes on both of the diaphragm sides.
- Stage 4 – lymphoma has spread to many organs or parts of the body or tissues like bone marrow

While assigning a stage, one letter is also included, A or B and E, which denotes some symptoms fever, loss in weight or sweating at night etc. "A" tells that patient has such symptoms; "B" tells that patient does not have such symptoms. As if person is having stage 3A, tells that lymphoma has spread in lymph nodes on both of the diaphragm sides with no symptoms of fever, sweating etc. E signifies if tumor has extended outside the node.

1.3.4 Follicular Lymphoma Grading

Grading describes how the tumor looks under a microscope to know how it is progressing. It tells us at what speed the lymphoma is spreading in the body of patient. A low grade centroblast appears same as normal cell and grows at a slow rate. A high grade centroblast appears to be irregular and spreads faster. The centroblasts are looked under a microscope and then examined at what speed the tumor is growing and how aggressive it is.

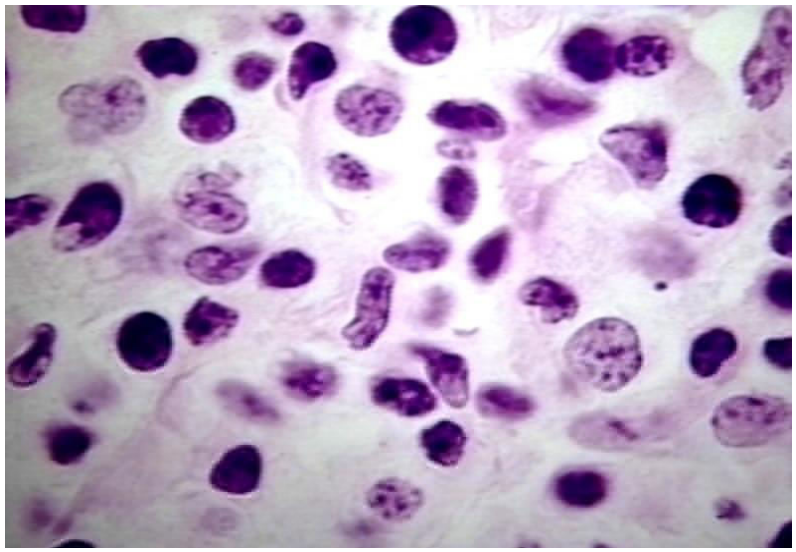

Figure 1.1: Follicular lymphoma Grade 2

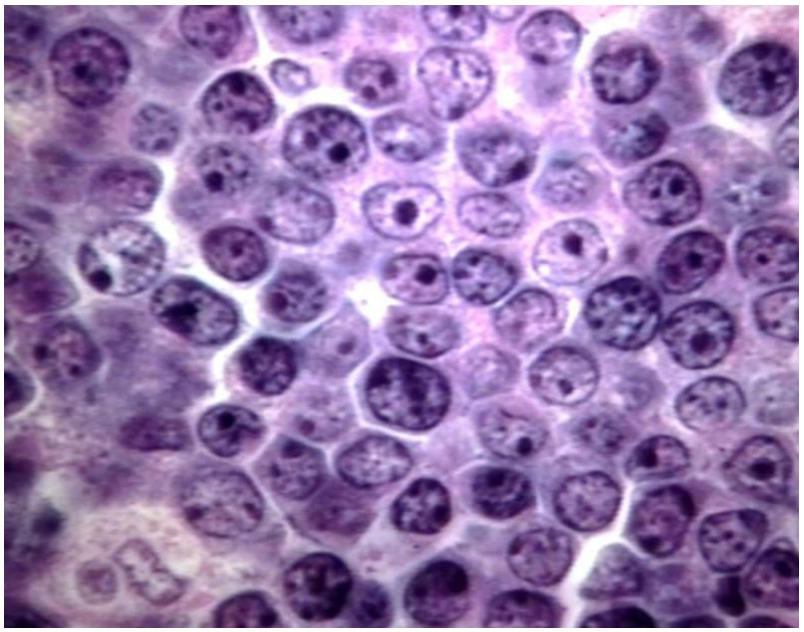

Figure1.2: Follicular Lymphoma Grade 3

There are 3 grades for rating follicular lymphoma:

- Grade 1 is low grade lymphoma which grows and spread at slow speed. The centroblasts appear same as normal body cells. There are mostly small cells.
- Grade 2 is moderate grade lymphoma which grows at little faster rate. The centroblasts appear different from normal body cells. There are a mixture of small cells and large cells.

- Grade 3 is high grade lymphoma which grows at very fast rate. The centroblasts appear very different irregular shape from normal body cells. There are mostly large cells.

Nowadays Grade 1 and 2 are considered low grade lymphoma and Grade 3A is High grade follicular lymphoma and Grade 3B is Diffuse Large B cell Lymphoma(DLBL).

The treatment for high grade and low grade lymphomas are different which also depends on the stages and symptoms of this disease.

Table 1.1: WHO Criteria Of Follicular Lymphoma Grades (Swerdlow S., etal., 2008)

Grade	Centroblasts
Grade 1	0 to 5 centroblasts / hpf
Grade 2	6 to 15 centroblasts / hpf
Grade 3	More than 15 centroblasts / hpf
Grade 3A	Centrocytes still present
Grade 3B	Solid sheets of centroblasts

Where hpf is high-power field which is the area to be observed specifying the magnification of lens used. Normally 400x magnification level is used.

Table 1.2: Architecture Patterns in Follicular Lymphoma (McNamara C., etal., 2011)

Follicular Lymphoma Pattern	Architecture
Follicular	More than75% follicular
Follicular and diffuse	25 to 75% follicular
Focally follicular	Less than 25% follicular
Diffuse	0% follicular

1.3.5 Follicular Lymphoma Staining

Staining is used to examine the cells of body so that there can be a clear distinction among the components of cells like nucleus, cytoplasm and other extra cellular part or highlighting red blood cells, white blood cells etc.. There are many techniques for staining like Silver staining uses silver to show proteins and DNA, Masson's trichrome, H&E staining etc. This system is using different stained follicular lymphoma images as its input.

Haematoxylin and eosin (H&E) staining is used mainly in histological images to examine minute parts of the cells in tissue. Haematoxylin makes nucleus of cell blue colored and eosin makes cytoplasm other extracellular substances pink colored. Red blood cells absorbs eosin stronly and so are colored red.

1.3.6 Follicular Lymphoma Treatment

Treatment is given to the patient after the grading and staging is done. Following are few ways of treatment:

- Chemotherapy: In this treatment the patient is given some kind of drugs and the cancer cells (centroblasts) are destroyed.
- Stem cell transplant: In this treatment, stem cells of the patient or from some donor are transplanted into patient's body as these blood cells are able to produce other kind of blood cells.
- Radiotherapy: In this treatment, high-energy X-rays destroy the centroblasts. It causes harm to the healthy cells

1.4 Image Processing

Image Processing is a method or technique in which the input image is converted into a form such that it becomes more efficient to visualize it properly. This is done so that the examiner can extract important information out of the image. Furthermore to perform various kind of operations on it so it can be used for some purpose.

1.4.1 Purpose of Image processing

Following tells us the purpose for doing image processing:

- For better visualization of the objects that are not visible with naked eye.
- To sharpen image for better view and restore the image that has been degraded.
- To retrieve the object of interest in the image this is called image retrieval.
- To distinguish among various objects in an image that is called image recognition.

1.4.2 Steps of Image Processing

Following are the steps that describe how an image is processed:

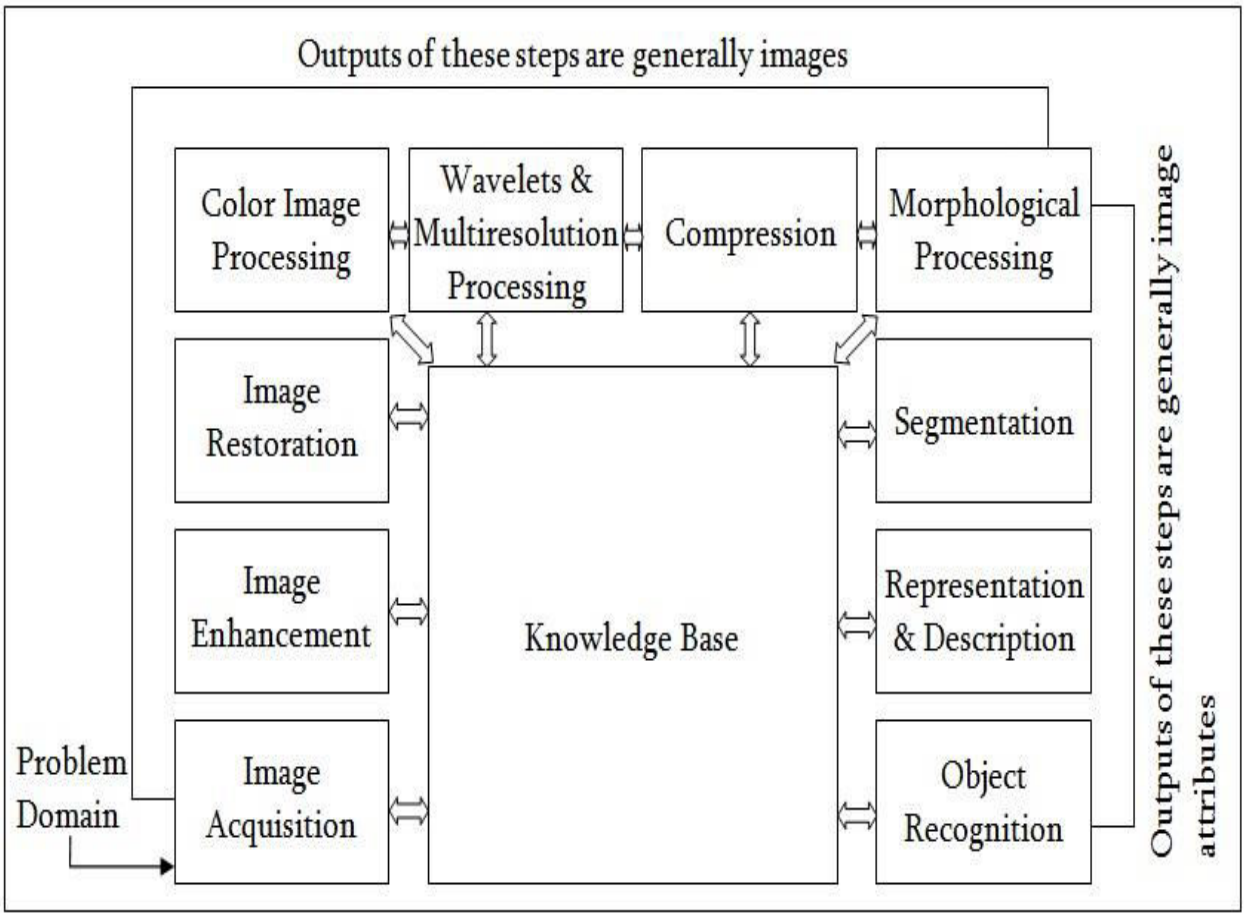

Figure 1.3: Steps of Image Processing

Image Acquisition:

This is the initial step of image processing in which the image is retrieved. The source can be any camera. The image acquired is purely unprocessed as it is the direct product of the hardware used to generate the image.

Image Enhancement:

In Image enhancement the image is modified into such manner so that it becomes better to visualize and for further analysis and experiments. We can remove the noise, make the image sharper, increase its brightness or contrast etc. All such kind of operations come under image enhancement. Following are the various approaches with which image can be enhanced:

- **Image sharpening:**

  Image can be sharpened so that it is less blurry in appearance and all the edges become clearly visible. This can be done by using unsharp filter.

Figure 1.4 (a): Input Image

Figure 1.4 (b): Image sharpened with unsharp filter

Image restoration:

This means to restore back the original image from its degraded image. The image can be degraded by any cause. It can be done by adding noise to the image.

- **Removal of noise:**

  There are various kinds of noise like salt and pepper, guassian noise, speckle noise, periodic noise etc. These noise can be removed by various filters like median filtering, average filter etc.

  In median filter, the output value of the pixel is the median of the neighborhood pixels and it removes the outliers, also is able to maintain the sharpness of the image.

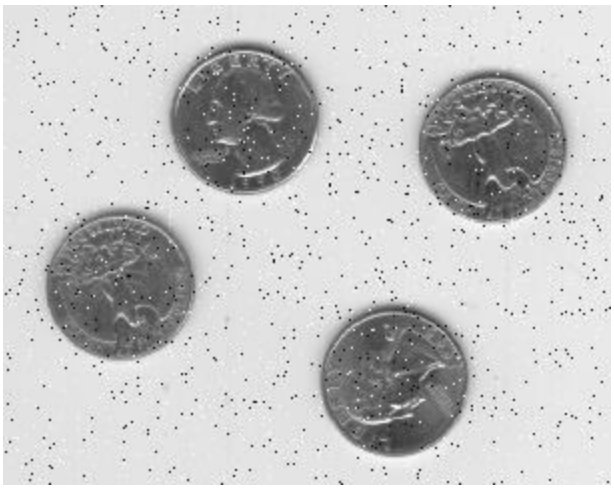

Figure 1.5 (a): Image with Salt and Pepper noise

Figure 1.5 (b): Salt and Pepper noise removed with Median Filter

Color Image processing:

This step is gaining more importance today as people today are using high definition digital cameras. So processing a colored image is of great concern today. For this there are few color models namely RGB, YCbCr, HSV, HIS, CMY. We can do conversions like RGB to HSV, RGB to CMY etc.

Wavelets and Multiresolution Processing:

Wavelets are small waves with varying frequency. These are used in context of multi-resolution images. These can be used in image processing for smoothing and blurring of an image.

The FWT(Fast Fourier Transform) decomposes the input image into four low resolution component images. There are some wavelet filters like haar, daubechies, symlets etc.

The inverse FWT combines the low resolution components back into original image.

Discrete Wavelet Transform can be used for image compression also.

Compression:

Image compression means to reduce the redundancy and to remove the data which is not important in the image which means to either totally remove the redundant data or to reduce the amount of redundancy and the data information which is not necessary for the user. In this way the size of the image gets reduced and it becomes easier to send the image data over the network.

Compression can also be seen in videos which means we can compress the video to reduce its size and becomes faster to send it over the network.

Compression can be of 2 types:

- **Lossless:** In this kind of compression the useful information is not lost. This type of compression is important in case of medical imaging, geological imaging, astronomical imaging etc.

Figure 1.6: Lossless Compression

- **Lossy:** In this kind of compression the useful information is lost. So it can be useful in cases where we don't care if some data gets removed like normal pictures of a person etc.

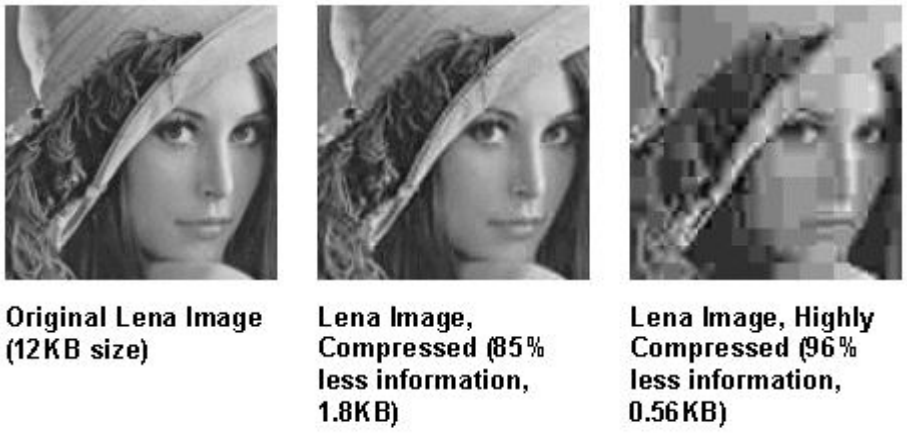

Figure 1.7: Lossy Compression

Morphological Processing:

Morphology deals with structural features of the image. Morphological operations are used to enhance the image to separate the elements that lie nearby to each other. These operations can be dilation, erosion, closing and opening operations.

In dilation, the object becomes thick or wide. In erosion, the object becomes thinner or get shrinked. These type of operations are performed to remove the small particles, to reduce the gap between the objects etc.

Figure 1.8: Image dilation

Segmentation:

In this the features of the image which have same characteristics are segmented that is it helps in clearly depicting the regions with same features. Segmentation can be done using various methods like Region Based and Edge Based.

In edge detection method, the boundary is identified to perform segmentation of the image. Edges are detected to identify the irregularities in the image. Edges are traced by identifying the pixel value and is then compared with the neighboring pixels. There are various edge detection methods like Gradient, log, sobel, laplacian, canny etc.

Canny Edge Detector

This detector detects the edges in the gray scale image, clearly showing the edges of the objects. Compared with other edge detection methods like sobel, log etc., this method gives the best outcome in terms of edge detection. To use this, first the image needs to be converted to gray scale then canny method is applied.

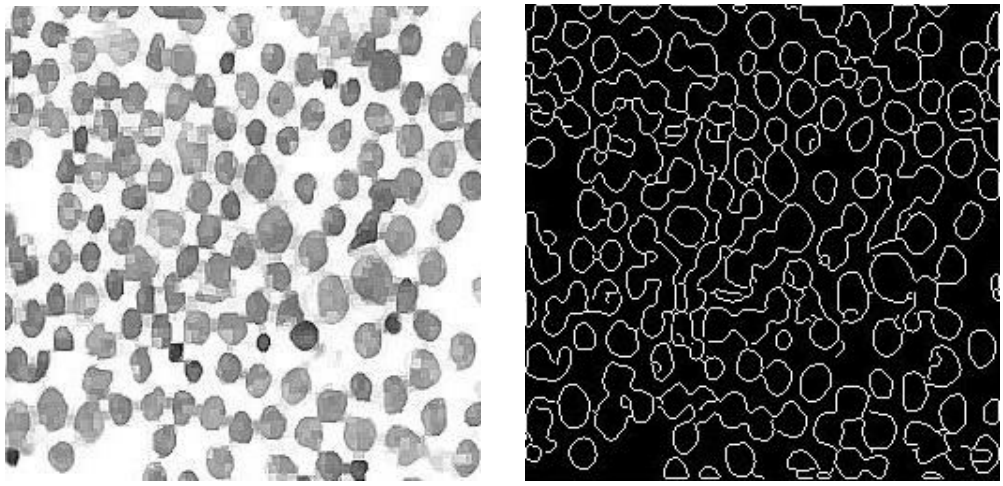

Figure 1.9: a) Gray-scale Image b) Segmented Image

Representation and Description:

After the segmentation of an image, the aggregated segmented pixels are represented and then described.

Representation can be seen in two terms. First in context of external characteristics that is the boundary and second in context of internal characteristics that are the pixels of the region. External Representation is used when our concern is of shape and internal representation is used when our concern is of color and texture. Boundary representation can be done using signatures, chain code, skeletons etc.

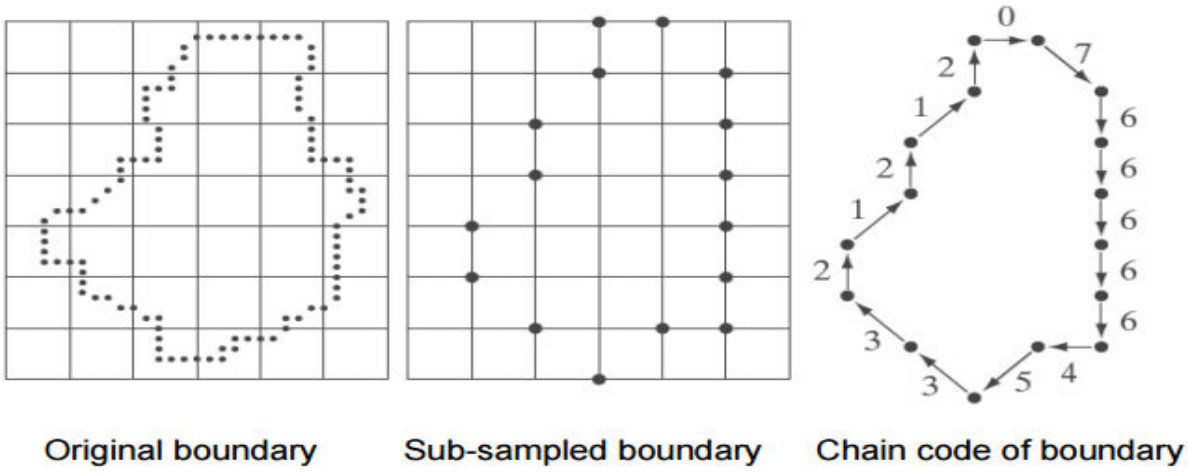

Figure 1.10: Boundary representation using chain code

The descriptor of a boundary can be its length, diameter of boundary, shape number etc.

Object recognition:

Recognising an object means to identify or find a particular object. This can be done in an image and even in a video. Like we want to find a criminal from a video recording etc. This is an example of object recognition. We can detect an object by detecting the edges or we can use features of the object to detect it etc.

1.5 Classification

Classification means to assign a class or category to the data input taken. This can be done by using various kind of classifiers such as Bayesian classifier, k-nearest neighbor method, support vector machine etc.

k-nearest neighbor method uses k nearest neighbors of the target whose class is to be determined. The class which majority of neighbors has is assigned to the target input as its class.

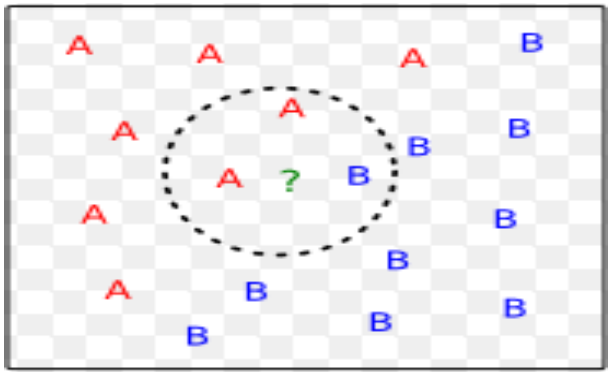

Figure 1.11: 3-nearest neighbor method

1.6 Support Vector Machine

Classification is done for classifying the objects/features in the dataset. There are various classification techniques available such as Artificial Neural Network, Decision Trees, Fuzzy Measure, Support Vector Machine.

SVM is based on supervised learning in which some training examples are presented, each having its own category, SVM allocate new input examples which we have not presented to machine before, into their respective categories. It builds a hyper plane such that the separation margin between the two kind of examples is maximized. It can handle large input data very efficiently. It can be used for both linearly separable patterns and non-linearly separable patterns also.

Support vectors are those examples which lie nearest to the optimal hyperplane. The examples lying on one side of hyperplane are positive examples and other are negative examples that are on the other side of the hyperplane.

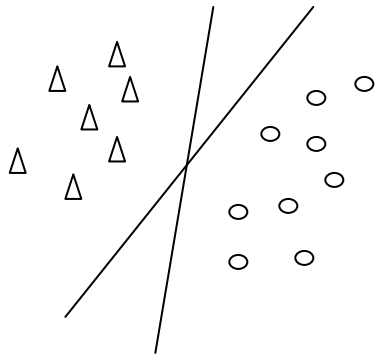

Figure 1.12: Determination of support vectors (in which plane they lie)

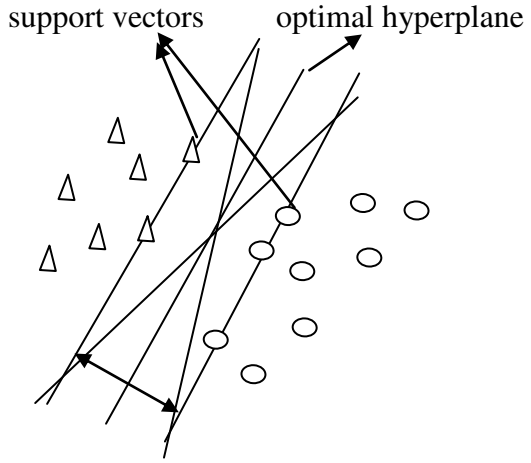

Figure 1.13: Idea of optimal hyperplane for linearly separable patterns

Figure 1.12 is showing a general layout of positive and negative examples where circles represent positive examples and triangles represent the negative examples. They are separated by hyperplane to separate the examples into 2 classes.

Figure 1.13 is showing the concept of linearly separable patterns in which both type of examples have been separated correctly. In this there are 5 support vectors 2 of negative examples and 3 of positive examples. The hyperplane is built in such a manner that the separation margin between the support vectors is maximum. It is correctly classifying the examples as negative which are to the left of hyperplane and as positive which are to the right of the hyperplane.

Support vector machine is special type of machine learning method in which supervised learning is used. In this the target is known to the machine, and given some data features it classifies the target. SVM is of great importance as its performance is better than other classification methods. To perform classification, firstly features are extracted from the image. Either the features can be extracted using feature extraction technique or can be extracted manually by experts. After that, the features are selected that on which features training will be performed. Then out of the total features we can take a few data for testing. When the dataset is trained for the target, then new test cases are employed to check the performance of correct classification rate. We can also use this training to know a better predictor for classification. It means the feature which gave maximum classification rate becomes the best predictor for classification.

Support vector machine has various types of kernels which are used to perform classification. All these kernels give different classification rates to the input sample. There are five different kernels namely, linear, quadratic, radial basis function (rbf), polynomial and multi-layer perceptron (mlp).

Using different kernel function K(xi,xj) can construct different types of learning machines with non-linear decision-making side on the input space, and lead to different support vector algorithms. (Hong S., etal., 2010)

Polynomial function is a d-order classifier. (Hong S., etal., 2010)
Polynomial kernel function:

$K(x_i \cdot x_j) = [(x_i \cdot x_j) + 1]^d$ (Men H., etal., 2009)

Radial basis function:

$K(x, x_i) = \exp(-\|x - x_i\|^2 / g^2)$ (Men H., etal., 2009)

Non-linear SVM is able to classify the set of examples which are not linearly separable by the hyperplane. For this, the SVM rearranges the original objects by the use of kernel functions which are mathematical functions. This process of rearranging the objects is called mapping and the mapped objects become linearly separable.

Chapter 2
REVIEW OF LITERATURE

Khouzani K. J. and Zadeh H. S., (2003) have done classification of cancer grades. The gleason grading is one in which the cancer is classified into five grades namely grade 1, grade 2, grade 3, grade 4 and grade 5. To perform this grading the researchers used k-nearest neighbor classifier. They took haematoxyline and eosin stained image of the tissue sample as their input for classification of grades. The features were extracted by wavelet packets, also by co-occurence matrices and multiwavelets. The features they used were the energy and entropy calculated from the images of the data sample taken for grading. Classification was best achieved using features extracted from multiwavelets with accuracy of 97%.

Sun B. Y., and Huang D. S., (2004) have developed to further improve the performance of classification using multi-class support vector machine. In this, a two-layered feed-forward neural network is cascaded to the outputs of support vector machine. Compared to those traditional methods of support vector machine, this method gives guaranteed better accuracy for classification.

Zorman M., et al., (2007) have made a system to identify the lymphoma by discovering the follicles present in the image of the sample taken. They carried out the work in two phases. In first phase they did image pre-processing steps to improve the quality o mage for better view for analysis and then did feature extraction and in second phase they used various set of approaches for classification. They used a new symbolic machine learning technique called rough sets. The results were compared with results derived from decision trees. Rough set theory is based on machine learning methods. The rough set theory gave better results for classification.

Swerdlow S., et al., (2008) have described WHO classification of grades. The World Health Organization (WHO) gave classification about neoplasms of the hematopoietic and lymphoid cells, represented a classification on the diagnosis of these tumors. WHO classification has defined some diseases like chronic lymphocytic leukemia (CLL) and plasma cell neoplasms. The issue of grading of follicular lymphoma was revisited in which grade 3 has been divided into grade 3A and grade 3B. The definitions of some categories of T-cell lymphomas have been refined. WHO also gave new definitions like Diffuse large

B-cell lymphomas. The major principle for classification they used is the recognition of various diseases using morphology, genetic, molecular, and clinical features. The diseases are calssified according to the cells appearances.

Sertel O., et al., (2008) have partitioned the images into its components in which they used unsupervised segmentation expressed in different colors. For segmentation they applied the lab color space, in which the difference among the 2 colors was uniform and Euclidean Distance was used. The K-means clustering algorithm has been used to represent the components of the image. Principal Component Analysis and Linear Discriminant Analysis as classifiers were used and also Bayesian classifier.

Li J., et al., (2008) have compared the performance of conventional support vector machine and the fuzzy support vector machine. The fuzzy support vector machine is able to classify the inputs which are not classified by the conventional support vector machine. For classifiable input patterns the SVM and fuzzy svm gave the similar results. But for those input patterns which were not classified by svm, in that case fuzzy SVM gave better performance. They took a database and made it into 8 categories and classified them.

Gurcan M. N., et al., (2009) have done detailed image analysis procedures for histopathology images. In this cytopathology and histopathology have been explained. Image preprocessing steps like color normalization , noise removal have been reviewed. Methods of histopathological images segmentation like the Linear Discriminant Analysis (LDA) for segmentation. Feature extraction methods like model based intermediate representation (MBIR), graphs etc. have been discussed. Feature selection methods have been discussed. For classification, support vector machine(SVM) and Adaboost have also been discussed.

Sertel O., et al., (2009) have established model-based intermediate representation (MBIR) for histological components which means to examine the divison of nucleus, cytoplasm etc. in the cells. They used color and texture feature extraction with MBIR to get the characteristics of tissue corresponding to every pixel. This technique gave good results specially in case of high grade lymphomas. They segmented the image using unsupervised clustering approach. Feature space dimensionality was reduced with Principal Component analysis, Linear Discriminant analysis.

Sami M. M., et al., (2009) have developed a system using image processing steps to differentiate oral borderline grades. Oral cancer is the cancer in mouth in which the cancer

cells are developed in the mouth. In this they took haemetoxylin and eosin image of the infected portion as input. Oral dysplasia is the first grade and second grade is carcinoma-in-situ. Epithelial segmentation was done followed by wide angle edgel method for the edgels selection. Then dilation was performed and then skeletonization thinning was done. And finally image was smoothened. The whole technique was based on comparison of drop-shaped similarity level between the neighbor rete ridges. It was concluded that the similarity level in dysplasia is greater than that of carcinoma-in-situ.

Huang P. W. and Lee C. H., (2009) have introduced a system to automatically detect the grade of prostate. Prostate carcinoma is most commonly found in men. In this they took stained images of the samples as inputs. The grades are from 1 to 5 in gleason grading system. The fractal analysis(Chaudhuri B. B, and Sarkar N., 1995) is used in image processing. In this the researchers extracted 5 features which are entropy, contrast, correlation, homogenity and energy. They used 3 classifiers Bayesian, k-nearest neighbor and support vector machine. Correct classification rates obtained were 91.2% with bayesian classifier, 93.7% with k-neaarest neighbor classifier and 93.7% with support vector machine classifier.

Men H., et al., (2009) have represented the technique of classification using support vector machine. In this a comparison between neural network algorithm and support vector machine has been shown. The results show that support vector machine gives better classification as compared to neural network algorithm. They have used two kernels of SVM which are polynomial kernel and radial basis function. Cross validation method was used to choose parameter for getting better accuracy for classification. The results showed an accuracy of 96% for classification.

Samsi S., et al., (2010) have used iterative watershed technique to split the overlapped objects. Initialy they performed segmentation of the follicles through clustering in which K-means clustering algorithm was used to cluster image in follicle and non-follicle regions. The texture features were extracted with median filter and gray level co-occurrence matrices. Fourier descriptors was used to smooth the follicle boundaries. This way they were able to detect follicles from non-follicular parts.

Boussaid K. B., et al., (2010) have used the principal component analysis (PCA) to extract texture color features. They used quadratic discriminate analysis (QDA) classifier to show

the centroblasts and non-centroblasts cells. The attributes of centroblasts were extracted by performing Otsu threshold, morphological operations like closing and opening, labeling, and area classification. They made an interface that was able to show centroblasts and non-centroblast cells accurately.

Sertel O., et al., (2010) have first converted the input image in 1-D single tone image with Principal Components Analysis. Then that image contrast was increased. Then unitone image was normalised. Then segmentation was done. The distinction of centroblasts and non-centroblasts cells was done with Gaussian mix model, expectation maximization (EM) algorithm. Posterior probability was used to construct a likelihood-based cell image.

Banu M. S., and Nallaperumal K., (2010) have used various color feature extraction approaches to analyze which one is better to have a better view on pathological images for diagnose. Techniques for extraction used were Pixel based Color Moments Descriptor , Color Histogram Moments Descriptor , Symmetrical Color Spatial Histogram and Binary Haar Color Descriptor and the color space used were HSVand CIE L*a*b* color spaces. The Binary Haar Color Descriptor gave the best result under HSV color space as compared to other approaches of color feature extraction used in this system.

Tosun A. B. and Demir C. G., (2011) have introduced an algorithm for the segmentation of histopathological tissue images. Segmentation of the tissue components has been done by graph construction and new texture features were defined by using Gray-level Run-length matrices. By calculating the "graph edge runs", this algorithm constructed "a graph run-length matrix". A graph-edge run is defined as a path starting from a node and contained those nodes that are reachable form set of edges of similar type. They used colon tissue images and proved that texture features obtained from "graph run-length matrices" lead to achieve more accuracy for segmentation accuracies.

Baoyong Z. and Yingjian Q., (2011) have used support vector machine to classify the input images into their corresponding classes. They used svm with ant colony optimization technique. As support vector machine has less rate of convergence, so a new approach has been used. In this the performance of support vector machine has been increased with ant colony technique. Ant colony algorithm has been used to select the kernel function. Classification performance of ant SVM is more than SVM.

Meng D., et al., (2012) have detected craters. Craters are formed on the surface of a planet when the meteorites collide with the surface of the planet. There are various parameters with the help of which we can determine the age of the crater that is from how long it is there. Those parameters can be dimension, pattern etc. of the crater. So to find the ages the scientists are interested in making their database. The researchers made an approach to automatically detect these craters. This system consists of two phases. The first one is selection of crater candidate region. The second phase is to detect the crater. The first phase is carried out by Kanade–Lucas–Tomasi (KLT) detector and the second phase uses Matrix-pattern-oriented least squares support vector machine as classifier to detect the craters.

Sivalingamaiah1 M. and Reddy B. D. V., (2012) have used multichannel Gabor filter to perform segmentation of color and grayscale images. In this a bank of gabor filters has been used in which after passing image through it, input image gets decomposed into number of filtered images as this filter is used for edge detection.

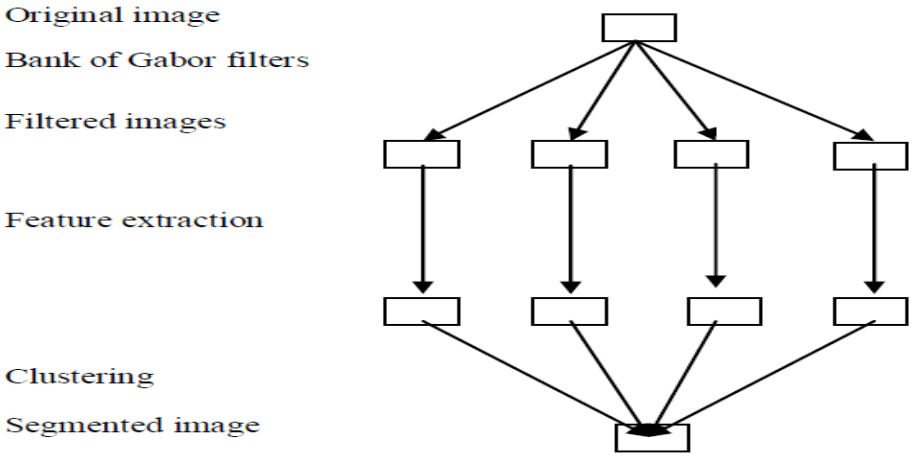

Figure 2.1: Texture Segmentation

Then feature extraction was done in which smoothing was obtained with low pass filtering. Then at last clustering of pixels was done and each pixel was labelled producing the segmented image.

Saxena P., et al., (2013) have proposed modified version of Active Contour Model and another algorithm based on local intensity clustering for image segmentation to determine the shape of nuclei. The criteria of clustering defined an energy in terms of a function. In this technique they have minimized the energy by steepest decent method for segmention. They

segmented the overlapped nuclei with less computation and much lesser iterations using this approach.

Saxena P., et al., (2013) have used image processing, the texture features of Follicular Lymphoma and Neuroblastoma tissues like nuclei, cytoplasm, extracellular material and red blood cells have been classified. In this, Gabor filters have been used for texture characterization of the H&E stained histopathological images by visualizing every object of image with many possible angles. In this, H&E stained images were convolved with 12 different 2D Gabor masks(filters) having different angles ranging from 30 degree to 360 degree. Then the output was 12 different convolved images and were combined into one textured image which represented as a 3-dimensional representation of input image.

Rezaeilouyeh H., et al., (2013) have classified the prostate cancer. This cancer is mostly common in men as compare to women. They have used image processing techniques to perform the grading of this cancer. In this they have used Gleason Grading as their system to perform grading. They have used shearlet transform to characterize the small parts of the cancerous cells. Every input image represented with histogram of shearlet coefficients that was used to classify the prostate cancer into benign and malignant tissues. The main benefit was that this system do not use segmentation and automatically performs the grading. They used support vector machine to perform gleason grading. They inputted the samples, used half for training and half for testing. They repeatedly trained the samples for 50 times and calculated the classification rates with an accuracy of 89%.

Amarreh I., et al., (2014) have done classification for epilepsy. Epilepsy is disorder of brain in which patient suffers from some mental problems, unconsciousness, abnormal functioning of brain etc. These are more commonly found in children. This abnormality is diagnosed by diffusion tensor imaging. In this the researchers took four attributes from diffusion tensor imaging which are fractional anisotropy, mean diffusivity, radial diffusivity and axial diffusivity. They used this with support vector machine for classification of epilepsy in children into two categories, first is children with epilepsy and second is children with remitted epilepsy and active epilepsy. The best attribute which can give best prediction with children with epilepsy was mean diffusity, with 90 to 100% sensitivity and with 96.6 to 100% specificity.

Zheng Y., et al., (2014) have done work on diagnosing breast cancer. For the diagnosis of breast cancer, content based image retrieval has been used in this system. In this the texture features and also pathological features have been extracted and also this system is used to discover the high- level semantics. This system used gabor features for represntaion of texture and also it parallely designed pathological features based in nucleus. This system used probabilistic latent semantic analysis(Hofmann T., 1999) system to find the high level semantics from the image data sample.

Chapter 3
PRESENT WORK

3.1 PROBLEM FORMULATION

Follicular lymphoma is a very dangerous cancer which leads to thousands of deaths a year. Detection of the Grade at an earlier stage is very important to save a person's life from death. Since the diagnosis of grade of lymphoma is being done manually by the pathologists. The pathologists detect the grade of follicular lymphoma by counting the centroblasts manually. This system will reduce the time taken for diagnosis and will detect the grade in lesser time.

This system will help pathologists to diagnose the grade through computer in an easier and faster way with better performance. This system will tell the best kernel function to be used to achieve better performance in detection of grades of follicular lymphoma. This system will be helpful for doctors to know that which attribute is more important to detect a particular grade. As per my knowledge, SVM with five different kernel functions linear, quadratic, rbf, polynomial and mlp have not been applied to color and texture features of image which are hue, saturation, value and entropy and also never been applied to clinical data of follicular lymphoma. So this system will use SVM with different kernel functions.

3.2 OBJECTIVE

This proposed system is designed to perform grading of follicular lymphoma cancer. This system is using support vector machine with 5 different kernel functions to achieve its objectives.

The main objectives of the research work are :

- to perform grading of follicular lymphoma to know that which grade lymphoma the patient is having in an easier and faster way.
- to have better performance of the grading system using SVM as it is able to handle a large number of input data very efficiently.
- to find the best kernel function which can be used for better classification.
- to find the best attribute as a predictor to detect/diagnose the grades.

3.3 METHODOLOGY

This system is using image processing techniques and support vector machine to perform grading of follicular lymphoma and to make a comparison between the kernel functions of support vector machine. This system is taking 2 types of inputs for classification of the data into corresponding grades. One is the histopathological data input that is 18 images and other is the clinical data of 110 follicular lymphoma patients having clinical report of the laboratory tests.

In case of histopathological data, color image processing is done. Then color and texture feature extraction is done to extract hue, saturation, value and entropy of the image. Then feature selection is done in which this system has taken all the four features extracted for training. This dataset has been talen online from Pathology Education Informational Resource (PEIR) Digital Library.

In case of clinical data input, there are a total of 25 attributes, so after feature selection, only five features were selected which are useful for diagnosing the grade of lymphoma. These are stage, stage group, tumor size, performance status and ldh ratio. The dataset has been taken online from Office of Cancer Genomics, National Cancer Institute under access of CGCI (Cancer Genome Characterization Initiative) Data Matrix Clinical.

Then training and testing is performed separately on both types of data corresponding to each feature. Testing is done on the complete trained data. This has been done by using support vector machine with five kernel functions that are linear, quadratic, rbf, polynomial and mlp. Using support vector machine on both types of input data, detection of each grade has been done. In case of histopathological data, detection of Grade 1, Grade 2 and Grade 3 has been done. In case of clinical data, the detection of Grade 1, Grade 2, Grade 3A and Grade 3B. The full methodology of this system has been shown in the following figure.

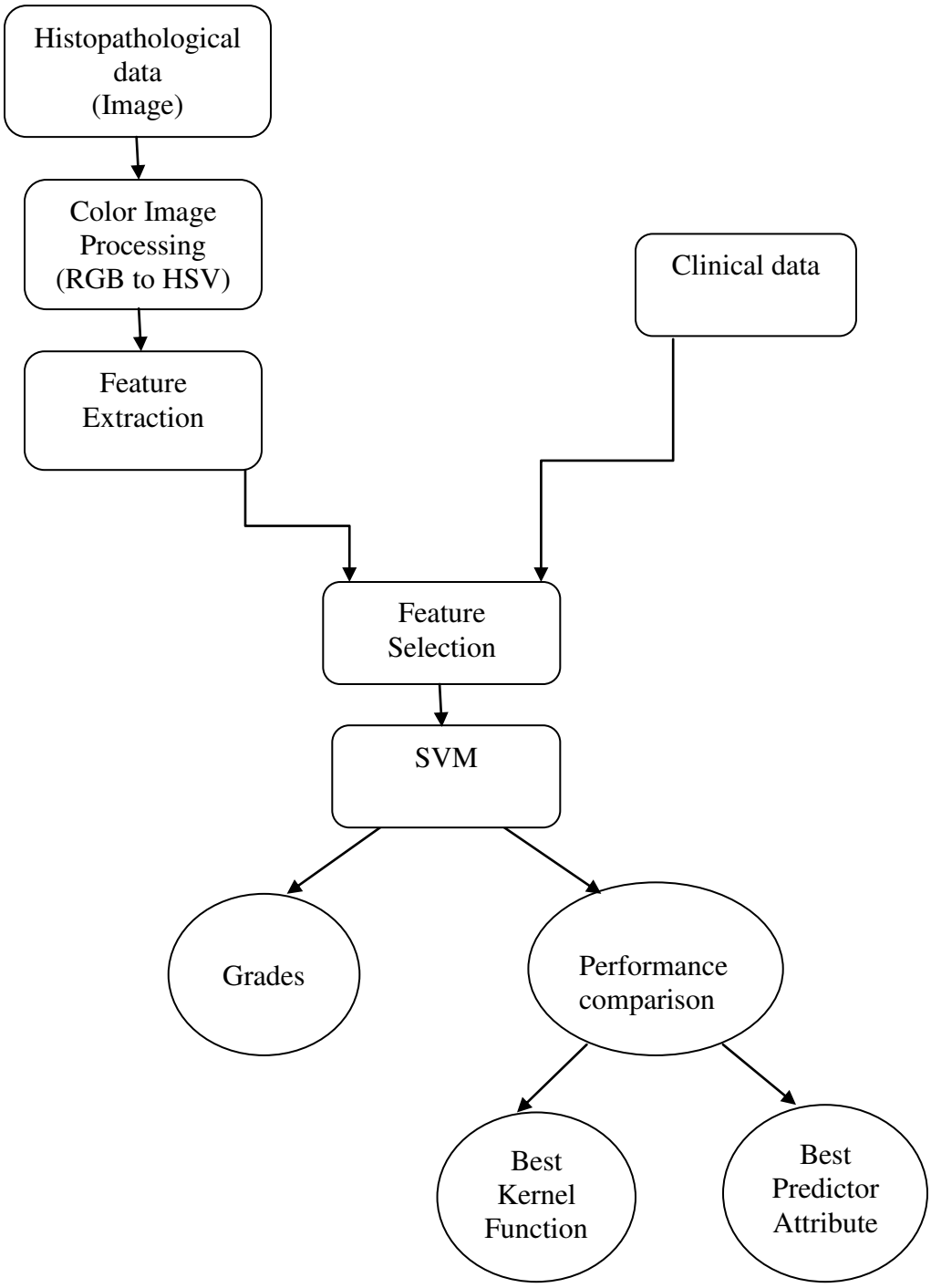

Figure 3.1: Flow Diagram of Methodology

3.3.1 Input

The input to this system is of two types:

- Histopathological data: This input data is in the form of images of biopsy samples of the patients having follicular lymphoma corresponding to different grades which are: Grade 1, Grade 2, Grade 3 as obtained from the source. There are a total of 18 images in the dataset.
- Clinical data: This input data is in the form of a report of 110 patients having follicular lymphoma. This contains the laboratory test results of all the patients.

3.3.2 Color Image Processing

Originally the images were in RGB color model. This system needs color features so the images were converted from RGB to HSV color model using color image processing technique.

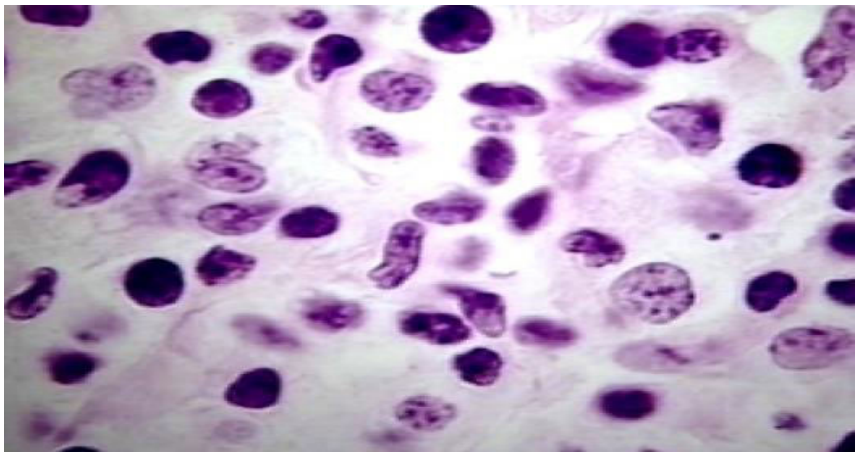

Figure 3.2: RGB Image

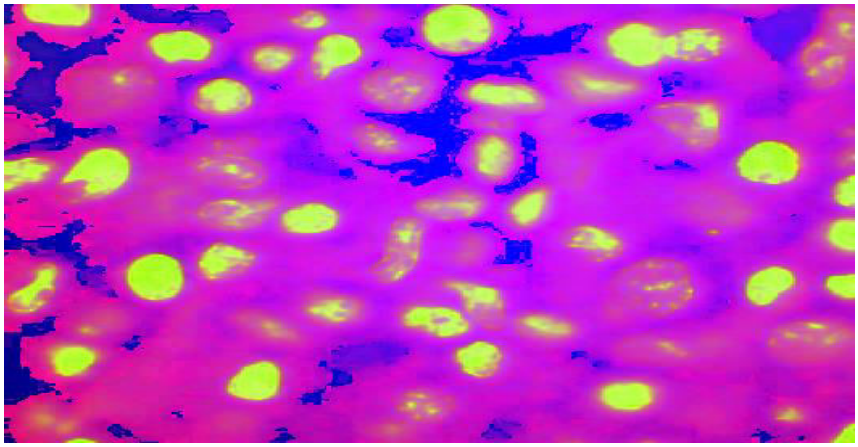

Figure 3.3: HSV Image

3.3.3 Feature Extraction

Features of an image consists of those features through which we can characterize the image and those features may be :

- Shape
- Color
- Texture

We can perform feature extraction using any of the three features. This system is using color features as Hue, Saturation and Value of the image. Also it has extracted texture feature as entropy of the image. This has been done to make a comparison between the kernel functions to know which is best. The features extracted are:

- Hue
- Saturation
- Value
- Entropy

For RGB to HSV :

R' = R/255

G' = G/255

B' = B/255

Cmax = max(R', G', B')

Cmin = min(R', G', B')

Δ = Cmax – Cmin (Samuel P. J., 2013)

The color features are:

Hue: Hue represents the shade of a color which means with which pure color it is similar. Hue is represented from 0 to 360 degree. Hue of red color is from 0, yellow from 60, green from 120, cyan from 180, blue at 240, magenta from 300.

$$H = \begin{cases} 60° \times \left(\frac{G' - B'}{\Delta} \bmod 6\right), C_{max} = R' \\ 60° \times \left(\frac{B' - R'}{\Delta} + 2\right), C_{max} = G' \\ 60° \times \left(\frac{R' - G'}{\Delta} + 4\right), C_{max} = B' \end{cases}$$ (Samuel P. J., 2013)

Saturation: Saturation of a color tells how much white color is there in the pure color. Like blood red is fully saturated while pink color is less saturated as there is a little white in pink color.

$$S = \begin{cases} 0, & \Delta = 0 \\ \dfrac{\Delta}{C_{max}}, & \Delta <> 0 \end{cases}$$

(Samuel P. J., 2013)

Value: Value of color represents the intensity or we can say brightness of that color.

$$V = C_{max}$$ (Samuel P. J., 2013)

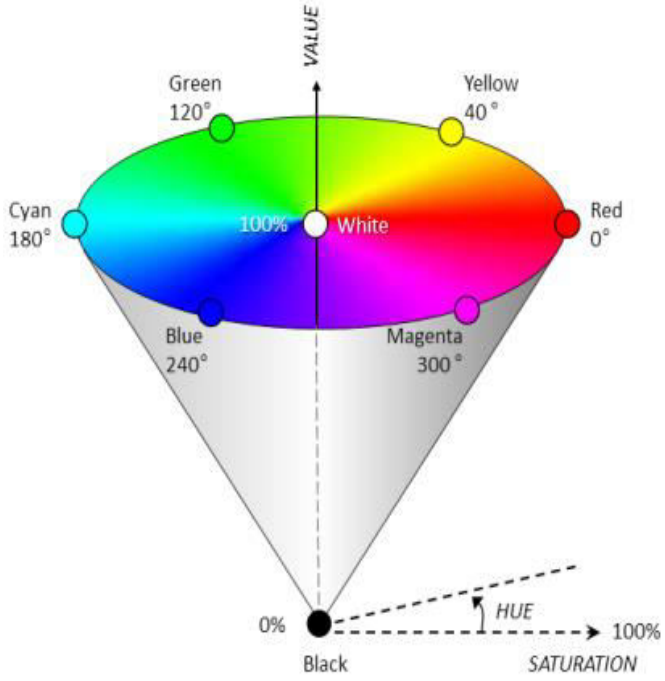

Figure 3.4: HSV color model

The texture feature extracted is:

Entropy: It is a texture feature which defines the gray level of the image pixels. If any image has pixels having same values, then its entropy is zero. The entropy of image is recognized by H:

$$H = -\sum_{k=0}^{M-1} p_k \log_2(p_k)$$

M denotes number of gray levels and p_k is the probability of gray level k.

3.3.4 Feature Selection

Feature selection is a technique in which only the relevant and important features are selected out of the total extracted features. With this, we select those features which are helpful in classification. In this system, feature selection has been done on both type of data sets:

In **histopathological data**, out of the four features extracted, all of the four features have been selected which have been used by support vector machine. They are :

- Hue
- Saturation
- Value
- Entropy

All these features have been discussed in section 2.3.3

In **clinical data**, out of the twenty five attributes, only five features have been selected according to their importance. These features are important to predict the grade of follicular lymphoma. So this system has selected the following features:

- Stage
- Stage Group
- Tumor Size
- Performance Status
- LDH Ratio

Stage:

Stage tells that to where and which body parts the lymphoma has spread. It specifies that which organs have been affected. There are 4 stages:

- Stage 1: Only one lymph node region (neck, armpit, groin or chest) or only one lymph structure (lymph nodes, spleen, and thymus gland) is affected.
- Stage 2: lymphoma has spread to two or more regions or structures of the lymph nodes on one side of diaphragm.
- Stage 3: lymphoma has spread in structures or regions of lymph nodes on both of the diaphragm sides.
- Stage 4: lymphoma has spread to many organs or parts of the body or tissues like bone marrow

Stage Group:

Stages have been categorized under two groups:

- Advanced: This group tells that patient has lymphoma on both sides of diaphragm and size of tumor is more than 10 cm. The patient has B symptoms (night sweats, etc.).
- Limited: This group tells that the patient has lymphoma on only one side of diaphragm and tumor size is small. The patient has no B symptoms.

Tumor Size:

Size of tumor is important parameter to classify the grades. This parameter tells how much mass of tumor has developed in the affected area. This size may vary depending upon the grade and stage of lymphoma. The tumor size here is taken in millimeters (mm).

Performance Status:

Performance status is useful in telling the physical condition of the patient, which can be used as a parameter to analyze the grade. WHO performance status is:

- 0:The person is completely energetic and can do all kind of work without restriction.
- 1: The person is limited to physically tiring activities but ambulatory and they can perform light activities in work.
- 2: Ambulatory and can take care of his/herself but cannot perform any activity at work.
- 3: The person is restrained in chair or bed more than 50% of his/her working hours.
- 4: Totally immobilized. Cannot take care of him / herself. Totally restrained to chair or bed.

LDH Ratio:

LDH ratio gives the amount of lactic dehydrogenase present in the blood. It is basically an enzyme present in the tissues of various parts of our body. It can be present in RBCs (red blood cells), kidney, brain etc. If the amount of LDH is greater than normal, it indicates that some of the tissues have been damaged. The cause of damage can be cancer or any other disease. So this also contributes in predicting grade of lymphoma.

3.3.5 Support Vector Machine

Support vector machine is special type of machine learning method in which supervised learning is used. In this the target is known to the machine, and given some data features it classifies the target. SVM is of great importance as its performance is better than other classification methods. To perform classification, firstly features are extracted from the image. Either the features can be extracted using feature extraction technique or can be extracted manually by experts. After that, the features are selected that on which features training will be performed. Then out of the total features we can take a few data for testing. When the dataset is trained for the target, then new test cases are employed to check the performance of correct classification rate. We can also use this training to know a better predictor for classification. It means the feature which gave maximum classification rate becomes the best predictor for classification.

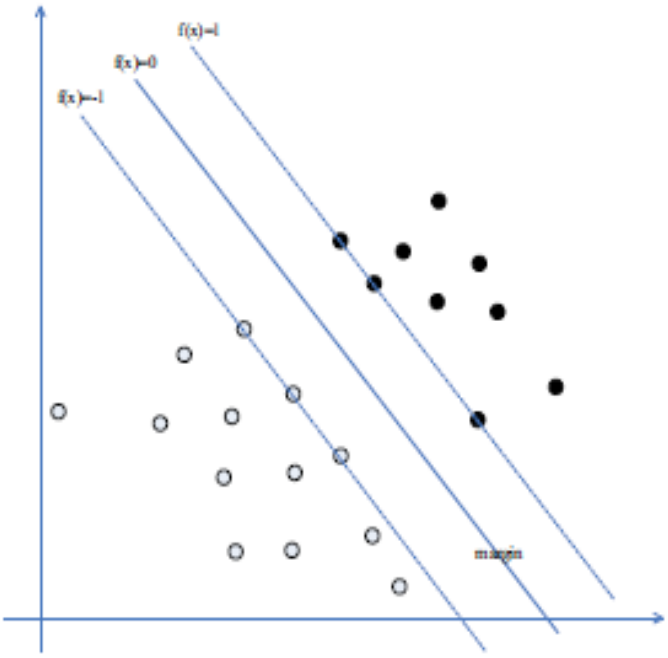

Figure 3.5: Support Vector Machine for linearly separable patterns(Zhu F., etal, 2011)

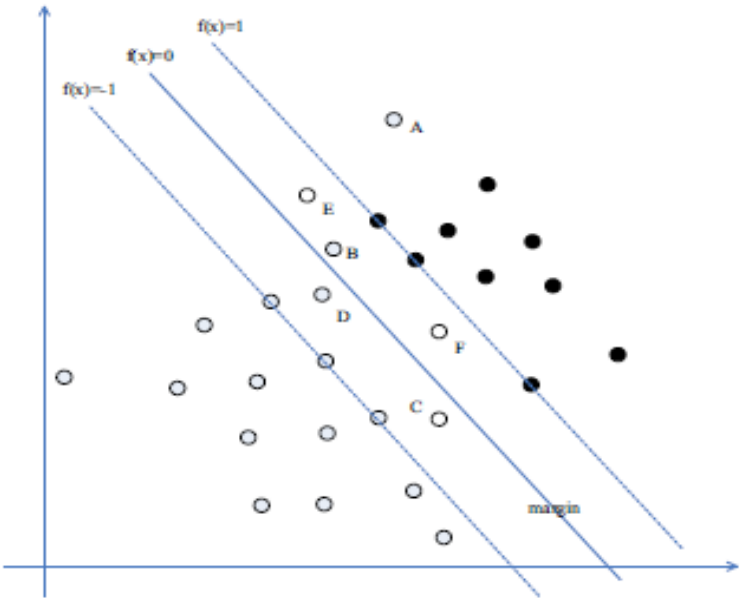

Figure 3.6: SVM for non-linear separable input patterns. A,B, E and F are misclassified samples(Zhu F., etal, 2011)

In figure 3.5, the support vector machine is classifying data set into two classes which are linearly separable.

In figure 3.6, the support vector machine is classifying non-linearly separable patterns. In this A, B, E and F are misclassified.

Non-linear SVM is able to classify the set of examples which are not linearly separable by the hyperplane. For this, the SVM rearranges the original objects by the use of kernel functions which are mathematical functions. This process of rearranging the objects is called mapping and the mapped objects become linearly separable.

This system has used support vector machine for detection of follicular lymphoma grades taking features extracted from both type of inputs. This system has compared the performance of each kernel function.

3.3.6 Outcomes

Outcomes are the grades of lymphoma, best kernel function and best attribute as a predictor for grading.

3.3.6.1 Grades

In case of histopathological image data, detection of grades is done into Grade 1, Grade 2, and Grade 3 according to the data set available.

In case of clinical data, detection of grades is done into Grade 1, Grade 2, Grade 3A and Grade 3B.

3.3.6.2 Performance Comparison

Two types of comparisons have been done. In one type of comparison, best kernel function has been identified. In other type of comparison, best predictor attribute has been identified for grades of follicular lymphoma.

- **Best Kernel Function:** After calculating the performance of all the kernels for each class taking all the features for detection of grades, the overall performance for each kernel is calculated which is the average of the observations. The kernel function with maximum overall performance is the best kernel function.
- **Best Predictor Attribute:** After calculating the performance for each attribute feature of clinical data corresponding to each grade, the attribute feature which contributes to the maximum overall performance is the best predictor for that grade. It tells that this attribute plays an important role in detecting/predicting a particular grade.

Chapter 4

RESULTS AND DISCUSSIONS

The results for both type of input data have been shown below separately:

4.1 Results with Histopathological image data:

The results show that detection of Grade 1, Grade 2, Grade 3 has been done depicting the performance of every kernel function corresponding to each feature and grade.

The overall performance for each kernel has been calculated by taking the average of all the performance values corresponding to each kernel.

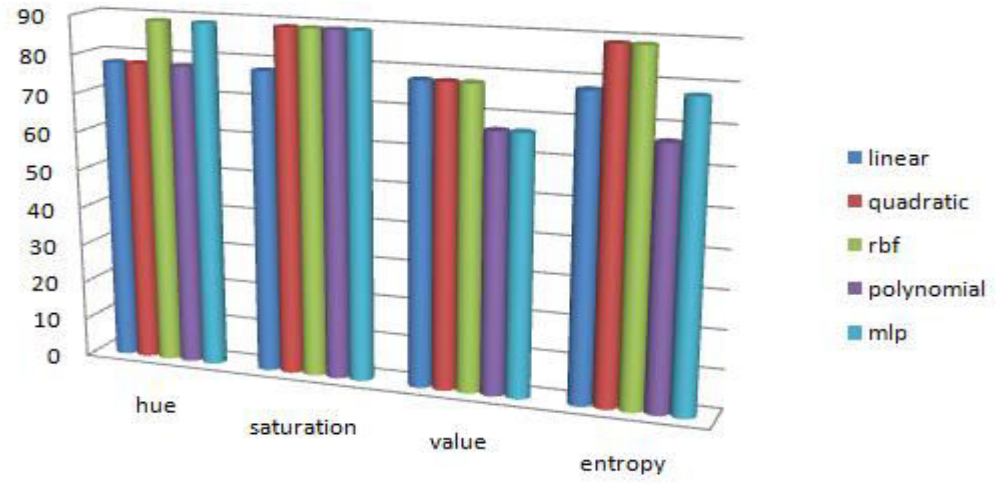

Figure 4.1: Performance Rate of Grade 1 from Histopathological Image data

Table 4.1: Overall Performance for Grade 1 from Histopathological Image data

	Hue	Saturation	Value	Entropy	Overall performance
Linear	77.78	77.78	77.78	77.78	77.78
Quadratic	77.78	88.89	77.78	88.89	83.335
Rbf	88.89	88.89	77.78	88.89	86.1125
Polynomial	77.78	88.89	66.67	66.67	75.0025
Mlp	88.89	88.89	66.67	77.78	80.5575

Table 4.1 depicts that the overall performance of rbf kernel is the maximum with 86.11% overall performance.

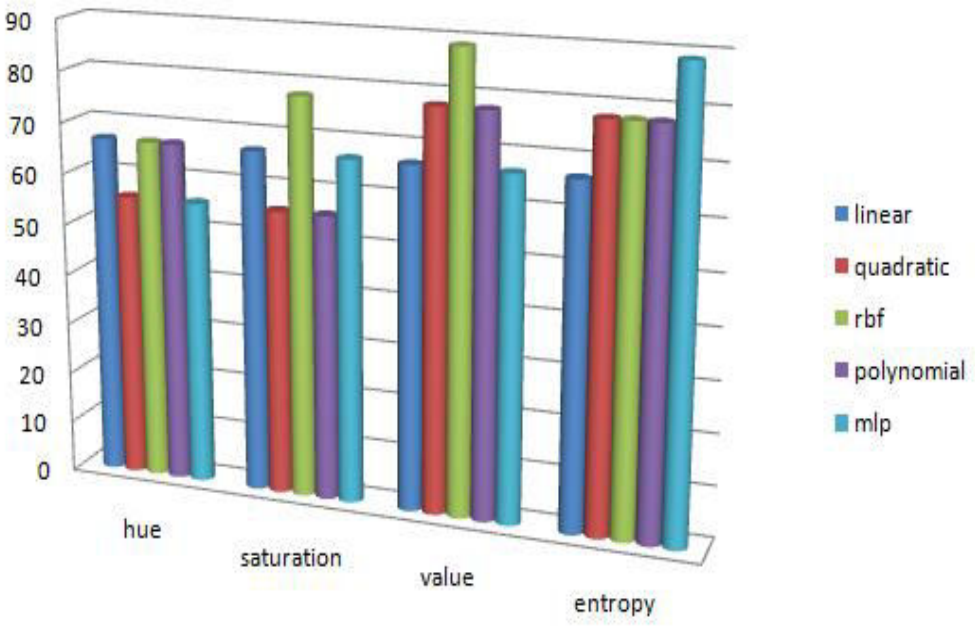

Figure 4.2: Performance Rate of Grade 2 from Histopathological Image data

Table 4.2: Overall Performance for Grade 2 from Histopathological Image data

	Hue	Saturation	Value	Entropy	Overall performance
Linear	66.67	66.67	66.67	66.67	66.67
Quadratic	55.56	55.56	77.78	77.78	66.67
Rbf	66.67	77.78	88.89	77.78	77.78
Polynomial	66.67	55.56	77.78	77.78	69.4475
Mlp	55.56	66.67	66.67	88.89	69.4475

Table 4.2 for grade 2 depicts that the rbf kernel function has the maximum performance of 77.78%.

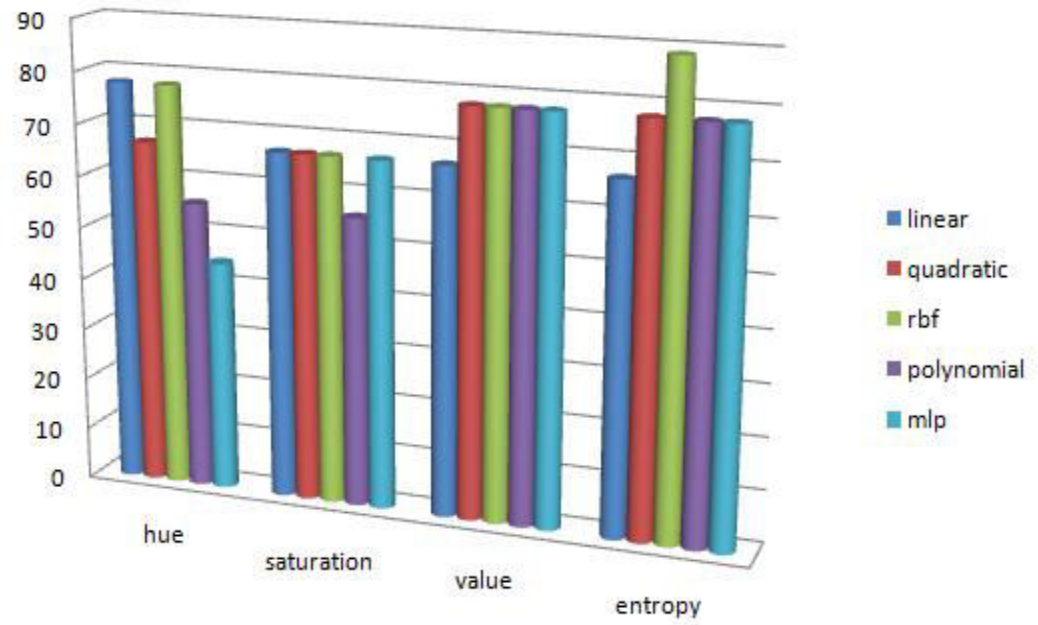

Figure 4.3: Performance Rate of Grade 3 from Histopathological Image data

Table 4.3: Overall Performance for Grade 3 from Histopathological Image data

	Hue	Saturation	Value	Entropy	Overall performance
Linear	77.78	66.67	66.67	66.67	69.4475
Quadratic	66.67	66.67	77.78	77.78	72.225
Rbf	77.78	66.67	77.78	88.89	77.78
Polynomial	55.56	55.56	77.78	77.78	66.67
Mlp	44.44	66.67	77.78	77.78	66.6675

Table 4.3 for grade 3 depicts that rbf kernel has the maximum overall performance of 77.78%.

Thus, we have concluded that rbf kernel is the best kernel function based upon this data, that is when we apply the SVM taking single column data.

4.2 Results with clinical data:

Grading with respect to each attribute feature to know which attribute feature is best predictor for grading.

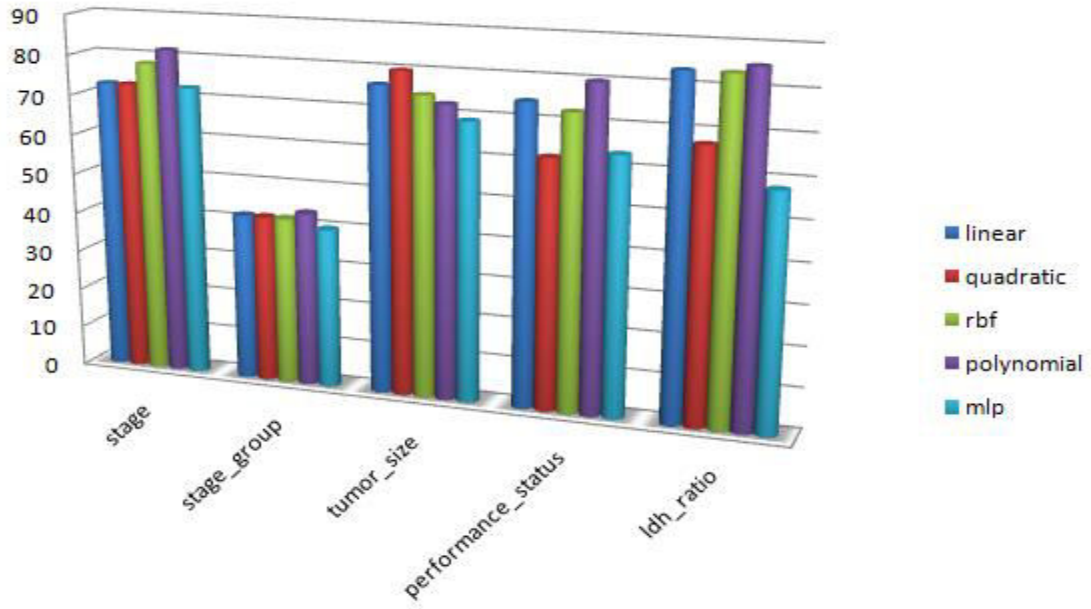

Figure 4.4: Performance Rate of Grade 1 from Clinical data

Table 4.4: Overall Performance for Grade 1 from Clinical data

	stage	Stage_group	Tumor_size	Performance_status	Ldh_ratio	Overall Performance
Linear	72.73	41.82	76.36	74.55	83.64	69.82
Quadratic	72.73	41.82	80	61.82	67.27	64.728
Rbf	78.18	41.82	74.55	72.73	83.64	70.184
Polynomial	81.82	43.64	72.73	80	85.45	72.728
Mlp	72.73	40	69.09	63.64	58.18	60.728
Average	75.638	41.82	74.56	70.54	75.636	

From table 4.4, it is clear that stage attribute is the best predictor for diagnosing grade 1 with overall average of 75.63%. Also we can see that ldh ratio is having equal average of 75.63% so it can also be a good predictor for grade 1. Here, polynomial kernel function gave maximum performance of 72.72%.

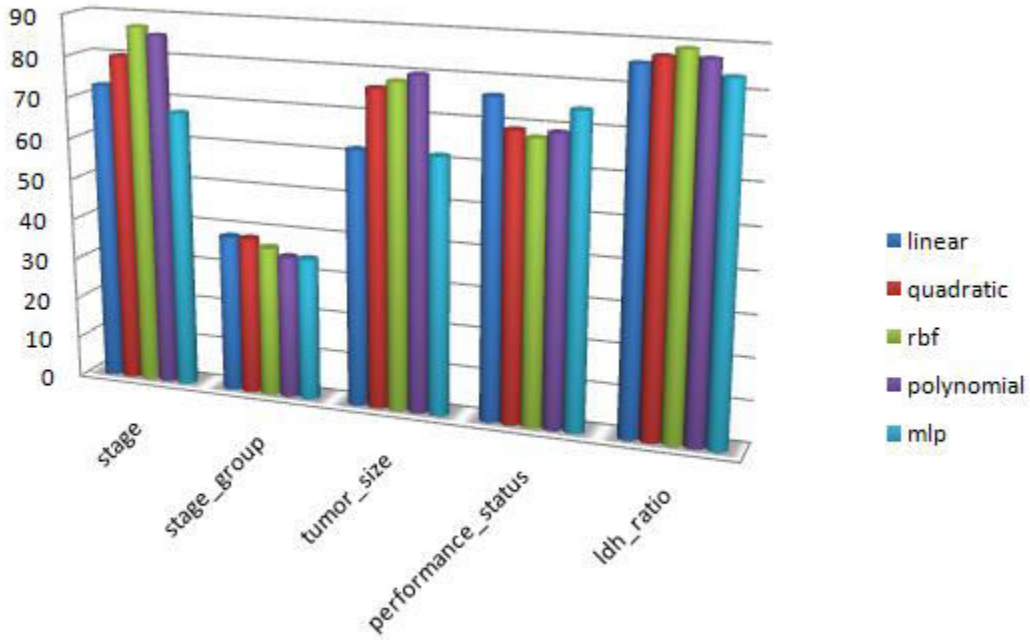

Figure 4.5: Performance Rate of Grade 2 from Clinical data

Table 4.5: Overall Performance for Grade 2 from Clinical data

	stage	Stage_group	Tumor_size	Performance_status	Ldh_ratio	Overall Performance
Linear	72.73	38.18	61.86	76.36	85.45	66.916
Quadratic	80	38.18	76.36	69.09	87.27	70.18
Rbf	87.27	36.36	78.18	67.62	89.09	71.704
Polynomial	85.45	34.55	80	69.09	87.27	71.272
Mlp	67.27	34.55	61.82	74.55	8364	64.366
Average	78.544	36.364	71.644	71.342	86.544	

From table 4.5, it is clear that ldh ratio is the best predictor attribute for detecting the Grade 2 with average of 86.54%. Also we can see that rbf kernel has maximum overall performance of 71.70%.

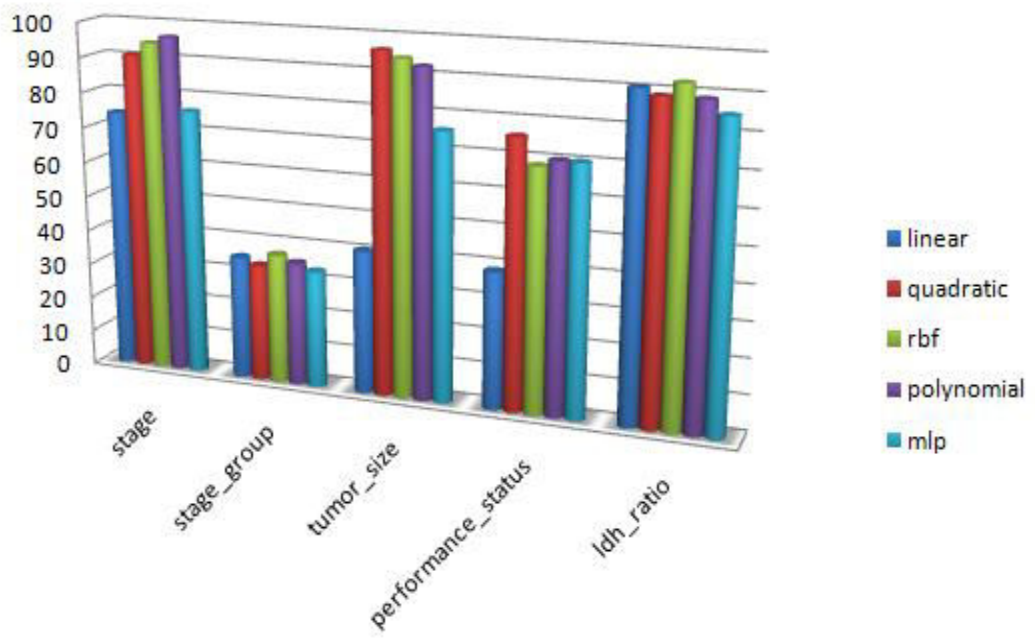

Figure 4.6: Performance Rate of Grade 3A from Clinical data

Table 4.6: Overall Performance for Grade 3A from Clinical data

	stage	Stage_group	Tumor_size	Performance_status	Ldh_ratio	Overall Performance
Linear	74.07	35.19	40.74	38.87	90.74	55.922
Quadratic	90.74	33.33	96.3	75.93	88.89	77.038
Rbf	94.44	37.04	94.44	68.52	92.59	77.406
Polynomial	96.3	35.19	92.59	70.37	88.89	76.668
Mlp	75.93	33.33	75.93	70.37	85.19	68.15
Average	86.296	34.816	80	64.812	89.26	

From table 4.6, it is clear that ldh ratio is the best predictor attribute for diagnosing the grade 3A with average of 89.26%. Also it clear that rbf kernel is the best kernel function with maximum performance of 77.40%.

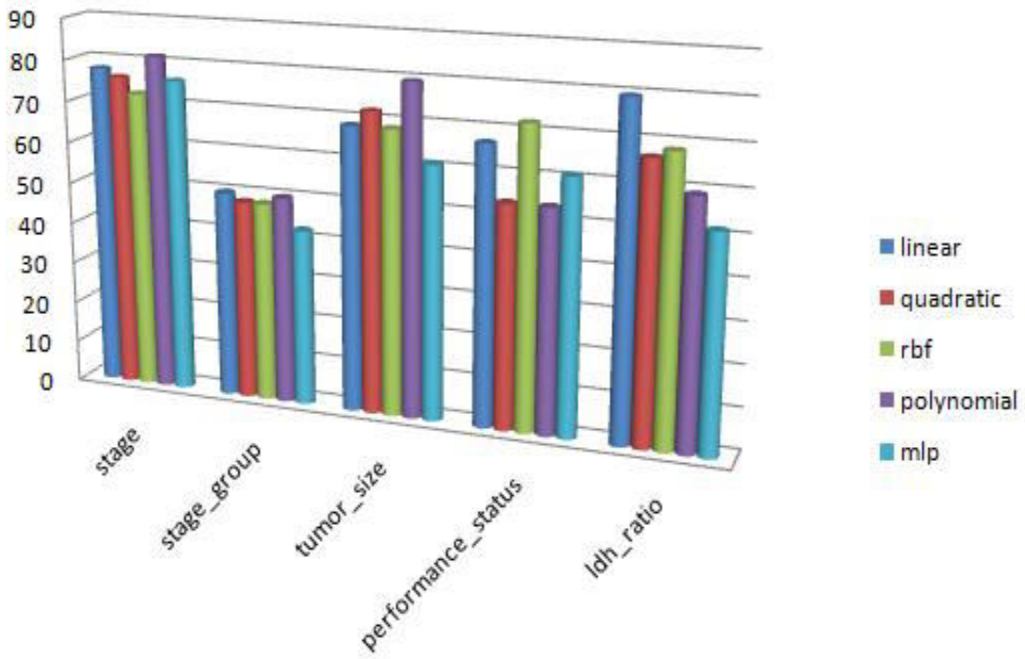

Figure 4.7: Performance Rate of Grade 3B from Clinical data

Table 4.7: Overall Performance for Grade 3B from Clinical data

	stage	Stage_group	Tumor_size	Performance_status	Ldh_ratio	Overall Performance
Linear	77.78	50	68.52	66.97	79.63	68.58
Quadratic	75.93	48.15	72.22	53.7	66.67	63.334
Rbf	72.22	48.15	68.52	72.22	68.52	65.926
Polynomial	81.48	50	79.63	53.7	59.26	64.814
Mlp	75.93	42.59	61.11	61.11	51.85	58.518
Average	76.668	47.778	70	61.54	65.186	

From table 4.7, it is clear that stage is the best predictor attribute for grade 3B with an average of 76.66%. Also we can see that rbf kernel is the best kernel function with overall performance of 65.92%.

4.3 Detection of Grades taking all the 5 attribute features of Clinical data

Using all the five attributes stage, stage group, tumor size, performance status and ldh ratio, detection of every grade has been done using all the five kernel functions.

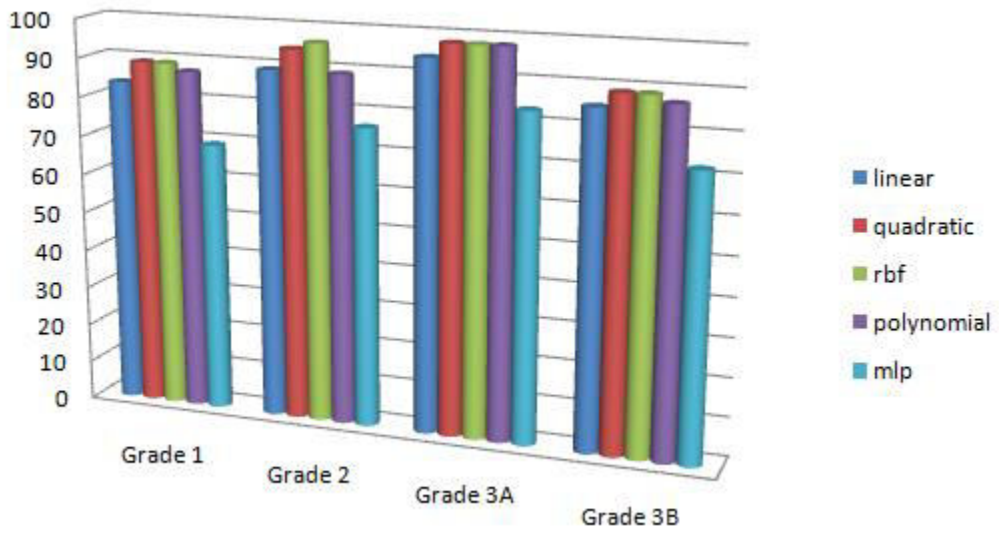

Figure 4.8: Performance Rate of all the Grades from Clinical data

Table 4.8: Overall Performance for all Grades from Clinical data

	Grade 1	Grade 2	Grade 3A	Grade 3B	Overall Performance
Linear	83.64	89.09	94.44	85.19	88.09
Quadratic	89.09	94.55	98.15	88.89	92.67
Rbf	89.09	96.36	98.15	88.89	93.122
Polynomial	87.27	89.09	98.15	87.04	90.38
Mlp	69.09	76.36	83.33	72.22	75.25

From table 4.8, we have seen the performance results of every kernel function for detecting each Grade. Here also we can see that rbf kernel is the best kernel function with overall performance of 93.12%. It is very clear from above table that Grade 3A has been detected with higher performance as compared to other grades.

Chapter 5
CONCLUSION AND FUTURE SCOPE

This system helps in performing grading of follicular lymphoma, that is, each grade has been detected corresponding to each attribute feature and also with combination of all the attribute features selected form the clinical data. With combination of attributes, the Grade 3A has been detected with higher performances as compared to other grades.

This system also helped in determining which attribute feature is best predictor for each grade, that is, tells that which attribute plays an important role in diagnosing a particular grade. The best predictor attribute for Grade 1 are stage and ldh ratio with 75.63% average performance, ldh ratio for Grade 2 with average performance 86.54%, ldh ratio for Grade 3 with average performance 89.26%, stage for Grade 4 with average performance 76.66%.

This system also helped in determining the best kernel function. Considering all the classification cases with different types of input and performing different kind of detection on different attributes, it is observed that in all cases rbf kernel has given maximum performance except for one. So, the best kernel function is rbf kernel function.

The Future of this work will be performing grading by counting the centroblasts in case of histopathological image data using some counting algorithm.

Chapter 6
REFERENCES

Amarreh I., Mary E. Meyerand, (2014) "*Individual classification of children with epilepsy using support vector machine with multiple indices of diffusion tensor imaging*", Elsevier, pp. 757–764.

Banu M. S., Nallaperumal K., (2010) "*Analysis of Color Feature Extraction Techniques for Pathology Image Retrieval System*", pp. 1-7.

Baoyong Z., Yingjian Q., (2011) *"Image Classification with Ant Colony Based Support Vector Machine"*, pp. 3260 – 3263.

Boucheron L., Can A., Madabhushi A., Rajpoot N., Yener B., (2009) "*Histopathological Image Analysis: A Review*" Department of Biomedical Informatics, the Ohio State University, Columbus, OH 43210 USA.

Boussaid K., Pennell M., Lozansk G., Shana'ah A., Gurcan M. N., ((2010) *"Effect of pathologist agreement on evaluating a computer-aided assisted system: recognizing centroblast cells in follicular lymphoma cases",* IEEE Conference: Biomedical Imaging, pp.1411-1414, April 2010.

Chaudhuri B. B, and Sarkar N., (1995) "*Texture segmentation using fractal dimension*", IEEE Trans. Pattern Anal. Mach. Intell., vol. 17, no. 1, pp. 72–77.

Hofmann T., (1999) "Probabilistic Latent Semantic Analysis," *The 22nd Annual ACM Conference on Research and Development in Information Retrieval*, San Francisco, pp. 289-296.

Hong S., Yan K., Hong, Z., 2010 "*Chest DR Image Classification based on SupportVector Machine",* Second International Workshop on Education Technology and Computer Science, vol. 1, pp. 170-173.

Huang P. W. and Lee C. H., (2009) "*Automatic Classification for Pathological Prostate Images Based on Fractal Analysis*", IEEE Transactions on Medical Imaging, Vol. 28, pp. 1037-1050.

Kourosh J. K., and Hamid S. Z., (2003) *"Multiwavelet Grading of Pathological Images of Prostate",* IEEE Transactions on Biomedical Engineering, Vol. 50, pp. 697 – 704

Li J., Huang S., He R., Qian K., (2008*) "Image Classification Based on Fuzzy Support Vector Machine"*, Vol. 1, pp. 68 - 71 .

Meng D., Yunfeng C., Qingxian W., (2013) "*Novel approach of crater detection by crater candidate region selection and matrix-pattern-oriented least squares support vector machine*", Chinese Journal of Aeronautics, pp. 385–393

McNamara C., Davies J., (2011) *"Guidelines on the investigation and management of follicular lymphoma",* 2011 Blackwell Publishing Ltd, British Journal of Haematology, 156, 446–467.

Men H., Gao Y., Wu Y., Li X., (2009) "*Study on Classification Method Based on Support Vector Machine*", Vol. 2, pp. 369 – 373.

Mustafa M. S., Saito M., Kikuchi H., Saku T., (2009) *"A computer-aided distinction of borderline grades of oral cancer",* IEEE International Conference on Image Processing, pp. 4205 – 4208.

Rezaeilouyeh H., La Rosa, Zhang J J., (2013) "*Prostate cancer detection and gleason grading of histological images using shearlet transform*", IEEE conference on signals, sytems and computers, pp. 268 – 272.

Samsi S., Lozanski G., Shana'ah A., Krishanmurthy K. A., Gurcan N. M., (2010), "*Detection of Follicles from IHC Stained Slides of Follicular Lymphoma Using Iterative Watershed*", IEEE Trans Biomed Eng. vol. 57, no. 10, pp. 2609–2612, 2010.

Samuel P. J., (2013) *"Face Image Retrieval with HSV Color Space using Clustering Techniques"*, SIJ Transactions on Computer Science Engineering & its Applications, Vol. 1, pp.

Saxena P., Singh S. K., Agrawal P., (2013) "*A Heuristic Approach for Determining the Shape of Nuclei from H&E Stained Imagery*", Students Conference on Engineering & Systems(SCES), IEEE, pp. 1-6.

Saxena P., Singh S. K., Agrawal P., (2013), *"Texture Classification of Biased Cytoplasmic Tissue Sample from Histopathological Imagery by Gabor Application*", Journal of Network and Innovative Computing ISSN 2160-2174 Volume 1 (2013) pp. 248-259.

Sertel O., Kong J., Catalyurek U. V., Lozanski G., Saltz J. H., Gurcan M. N., (2008), *"Histopathological Image Analysis Using Model-Based Intermediate Representations and Color Texture: Follicular Lymphoma Grading"*, Journal of Signal Processing Systems, vol. 55, Issue 1-3, pp. 169-183.

Sertel O., Kong J., Lozanski G., Saltz J., Gurcan M., (2008) "*Texture classification using non linear color quantization : application to histopathological image analysis*", IEEE, pp. 597-600.

Sertel O., Lozanski G., Shana'ah A., Gurcan M. N.,((2010) "*Computer-Aided Detection of Centroblasts for Follicular Lymphoma Grading Using Adaptive Likelihood-Based Cell Segmentation*", IEEE transactions on bio-medical engineering , Vol. 57, NO. 10, pp. 2613-2616.

Sivalingamaiah M., Reddy B. D., (2012) "*Texture Segmentation Using Multichannel Gabor Filtering*", IOSR Journal of Electronics and Communication Engineering Vol. 2, Issue 6 pp. 22-26.

Sun B. and Huang D., (2004) "*Texture Classification Based on Support Vector Machine and Wavelet Transform*", Vol. 2, pp. 1862 – 1864.

Swerdlow S., Campo E., Harris N., Jaffe E., Pileri S., Stein H., Thiele J., (2008) *"WHO classification of tumours of haematopoietic and lymphoid tissues*", vol. 2, World Health Organization, Lyon, France, fourth ed. ,2008.

Tosun A. B. and Demir C., (2011) *"Graph Run-Length Matrices for Histopathological Image Segmentation"*, IEEE Transactions on Medical Imaging, Vol. 30, No. 3, pp. 721-732, March 2011.

Zheng Y., Jiang Z., Shi J., Ma Y., (2014) *"Retrieval of Pathology Image For Breast Cancer Using PLSA Model Based On Texture and Pathological",* IEEE International Conference on Image Processing, pp. 2304 – 2308.

Zhu F., Ye N., Xu S., Gu X., (2011) "*Support Vectors Classification and Incremental Learning*", IEEE Joint International Information Technology and Artificial Intelligence Conference, Vol. 1, pp. 206-210.

Zorman M. and Kokol P., (2007) "*Symbol-Based Machine Learning Approach for Supervised Segmentation of Follicular Lymphoma Images"*, IEEE International Symposium on Computer-Based Medical Systems, pp. 115-120.

Book

WSW Shalaby - Manual of Gynecologic Oncology, 2011

Chapter 7
APPENDIX

APPENDIX A: ABBREVIATIONS

DLBL:	Diffuse Large B cell Lymphoma
HL:	Hodgkin Lymphoma
NHL:	Non-Hodgkin Lymphoma
HIV:	Human Immunodeficiency Virus
SVM:	Support Vector Machines
LDH:	Lactic Dehydrogenase
DNA:	Deoxyribonucleic Acid
RBF:	Radial Basis Function
MLP:	Multi Layer Perceptron