COMPUTER-AIDED GRADING OF NEUROBLASTIC DIFFERENTIATION USING SUPPORT VECTOR MACHINE

A Dissertation Proposal

Submitted

By

Karamjeet Kaur

11309419

To

Department of Computer Science and Engineering

In partial fulfilment of the Requirement for the

Award of the Degree of

Masters of Technology in Computer Science and Engineering

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: Karamjeet Kaur	Registration No: 11309419
Batch: 2013-2015	Roll No. RK2305A04
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 grading of neuroblastic differentiation using support vector machine.
2. Grading of leukemia cancer.
3. Feature classification of neuroblastoma.

15745

Signature of Supervisor

PAC Remarks:

Topic '1' is approved. Research paper is also expected.

18/9/14

APPROVAL OF PAC CHAIRPERSON: **Signature:** **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

The purpose of this thesis is to perform computer-aided grading of neuroblastoma cancer by using support vector machine. The cancer is graded as poorly differentiating, undifferentiating and differentiated based on the level of mitiosis-karyorrhexis index, stage, DNA ploidy, mycn status, diagnostic category. This document defines about neuroblastoma cancer, its symptoms and causes. This also includes proper methodology to be used including feature-extraction and classification methods of image. It describes why support vector machine is selected for this grading. Different kernels of SVM are also compared in this proposed work. The attribute is also evaluated for each grade. The methods which are used for successful execution are defined in this document. Large database is used for properly differentiating the disease for its better treatment.

CERTIFICATE

This is to certify that **Karamjeet Kaur** has completed M.Tech dissertation proposal titled "**Computer-Aided Grading of Neuroblastic Differentiation 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 fulfilment of the conditions for the award of M.Tech in Computer Science and Engineering.

Date___________

Signature of Advisor

Name:

UID:

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 to the **Project Approval Committee Members** for their valuable comments and discussions. I would also like to thank to **Lovely Professional University** for the support on academic studies and letting me involve in this study.

Karamjeet Kaur

R. No. 11309419

DECLARATION

I hereby declare that the dissertation proposal entitled, **"Computer-Aided Grading of Neuroblastic Differentiation 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: ________________ Karamjeet Kaur

Reg. No. 11309419

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Cancer

The human body is made up of many cells. Normally cells of the body grow and then they divide to make new cells, and ends in a proper manner. When cancer cells begin, it reproduces to grow in uncontrollable way. It starts from one part of the body and spreads to other parts. It is disease in which cells grow out of control. DNA is present in all cells which performs various functions. Cells become cancerous when this DNA gets damaged. Infected cells grow and can harm various systems of the body. Damaged DNA may also be inherited. This is a life threatening disease.

The main cause of cancer is that cells which grow uncontrollably and not dies. Normal cells works accordingly and cancer cells continue to grow and divide. Cancers can be classified by cells:

- Which include internal and external body parts
- That is located in bone, muscles, fat, etc.
- Which begins from lymph node
- Which starts in nerves and many more

1.2 Cancers in adults and children

There are differences between the types of cancers that develop in children and that in elders. Adults generally tend to have cancers in lungs, pancreas, colon, breast, prostate, etc. Childhood cancer commonly includes leukemia, neuroblastoma, lymphoma, retinoblastoma, bone cancer and brain tumors. Childhood cancers spread more easily than cancer in adults. Cancers are rare in children as compared to adults. They are often caused due to DNA changes which happen in early phase of life. Various treatment like chemotherapy given to children results better and helpful as compared to adults. But such treatments can have side effects. Childhood cancers are not affected by environmental changes. From last many years treatment provided to many childhood cancers have improved.

1.3 Neuroblastoma

Neuroblastoma is a tumor of early childhood. It begins in the nervous system or nerve cells called neuroblasts. It generally starts in the adrenal glands (two small glands above the kidneys). It can also start from the neck, spinal cord and chest. It occurs commonly during early childhood period. It is a complex disease occurring in babies or infants or children less than 3 years of age. It starts from the cells of sympathetic nervous system which is the part of autonomic nervous system. It can be sometimes inherited from parent to the child but in most cases it is not inherited. It is more common in boys than in girls. Generally not every child having the symptoms of neuroblastoma develops that into cancer. Some may resist that and some having no symptoms or risk can still suffer from that.

It has a wide range of behavior, some neuroblastoma spread rapidly and other may grow slowly. Cancer cells may die on its own with no reason and may prevent the disease to grow. Every tumor may not be cancerous as described below:

- Ganglioneuroma is a non-cancerous tumor. It may be removed with the help of surgery.
- Ganglioneuroblastoma may have both cancerous and non cancerous parts. It contains nerve cells which do not grow normally as in the case of neuroblastoma. Its treatment is also similar to that of neuroblastoma.

1.3.1 Causes of Neuroblastoma

The exact cause of this cancer is unknown. It occurs when neuroblasts grow and divide uncontrollably which should instead develop into nerve cell. Sometimes genes of neuroblasts may be defected which may be one of the cause of this disease. These defected genes allow the nerves to grow uncontrollably. It generally not happens to the children of same family but it may happen rarely. Healthy cell grow out of control sometimes which can be the cause of this cancer. Neuroblasts found in newly born may grow or not. Many cases this disappears. Hence genetic mutation is the only one cause found yet which makes the cell grow in an un-normal manner. These are not caused due to any environmental changes. It can’t be passed to others, which means this cancer is not transferable. It is also not infectious.

1.3.2 Signs and Symptoms of Neuroblastoma

The signs and symptoms of neuroblastoma vary widely. It depends on which part of the body it has occurred. Starting symptoms are common so it is initially difficult to diagnose.

Specific Symptoms:

- Swollen stomach when it occurs in abdomen
- Problem in breathing and swallowing when it occurs in occurs in chest area
- Visible as lump when occurs around the neck
- Weakness in legs and problem in walking when happens in spinal cord

Common symptoms include:

- Swelling under the skin in neck or chest
- Back pain
- Pain in bones
- Trouble in breathing
- Dark circles around the eyes
- Weakness

Less common symptoms include:

- Fever
- Tiredness
- Infections and bleeding
- Increase in blood pressure
- Trouble in breathing

1.3.3 Risk factors for Neuroblastoma

Risk factors are the exposure of a person of getting a chance of disease like cancer. For all the cancers risk factors varies. Risk factors in adult cancer may include weight of body, diet plan, and other daily activities. But they do not play much role in childhood cancer like neuroblastoma. One risk factor for neuroblastoma can be age. It happens in very young children and rarely happens in children over age ten. Other factor may be heredity from parents. Inherited neuroblastoma often occurs in earlier age as compared to neuroblastoma that is not inherited. It may occur in children if their parents have gone through this disease in their early childhood. It is not necessary that everyone having this risk factor will develop this into cancer. And persons not having these factors can even have the risk of developing this cancer.

1.3.4 Staging of Neuroblastoma

Staging of a cancer describes about its size. It helps doctor to understand more about the cancer and this helps him in its proper treatment. The most common staging system includes the following stages:

- **Stage 1:** Cancer is present in one part of the body. It does not spread and can be removed by surgery.
- **Stage 2A:** Tumor is present only in one part that is it is localized but it cannot be removed completely by surgery.
- **Stage 2B:** In this stage also tumor is present in one body part only. It may or may not be removed by surgery. It spreads into nearby lymph nodes.
- **Stage 3:** It is not possible in this stage to remove tumor completely by surgery. Tumor can be large in this stage. Tumor is in one side of the body. It spreads to local areas but not in far apart parts of the body.
- **Stage 4:** It is advanced stage. Tumor in this stage spreads to far apart parts of body like lymph nodes, skin, bone marrow and the other organs.
- **Stage 4S:** It is also called as special neuroblastoma in this stage. It is found in children below one year old. This tumor spreads to liver, bone marrow and skin. All the tumors may be completely removed by surgery in this stage.

Newer staging system is also developed which stage cancer into **L1** (similar to stage 1, 2A and 2B described above), **L2** (corresponding to stage 3), **M** (corresponding to stage 4), and **MS** (corresponding to stage 4S).

1.3.5 Grading of Neuroblastoma

International Neuroblastoma Pathology Classification: This classifies tumor into unfavorable and favorable histology based on tumor subtype, patient age, and Mitiosis-karyorrhexis index (MKI). It is summarized in table below as described by (Shimada, H., et al., 2001).

Table 1: Mitiotic figures + karyorrhectic cells/ 5000 neoplastic cells

	Expressed as percent	Expressed as cell count
Low MKI	<2%	<100 / 5000 cells
Intermediate MKI	2-4%	100-200 / 5000 cells
High MKI	>4%	>200 / 5000 cells

Table 2: Favorable histology

Tumor Subtype	MKI	Patient Age
Poorly differentiated NB	Intermediate MKI	<1.5 yrs
Poorly differentiated NB	Low MKI	<1.5 yrs
Differentiating NB	Intermediate MKI	<1.5 yrs
Differentiating NB	Low MKI	<5 yrs

Table 3: Unfavorable histology

Tumor Subtype	MKI	Patient Age
Undifferentiated NB	Any MKI	Any
Poorly differentiated NB	High MKI	Any
Poorly differentiated NB	Intermediate MKI	>=1.5 yrs
Poorly differentiated NB	Low MKI	>=1.5 yrs
Differentiating NB	High MKI	Any
Differentiating NB	Intermediate MKI	>=1.5 yrs
Differentiating NB	Not Applicable	>=5 yrs

1.3.6 Treatment of Neuroblastoma

Plan of the treatment is based on factors like age of the child, cancer stage, cells which are the cause of cancer. The treatment of neuroblastoma cancer is based on the type of tumor risk which is based on these factors. Tumor can be of low, intermediate or high risk. Depending on the stage, chemotherapy can be usually given. In some cases it can be followed by radiotherapy. It helps in killing the remaining cancer cells. Other treatments may include stem cell transplant and immunotherapy. Treatment on risk basis is described below:

- Treatment options for low-risk neuroblastoma include surgery, chemotherapy with or without surgery or observation without biopsy.
- For intermediate-risk neuroblastoma usually chemotherapy is provided, surgery and observations is done in infants. Sometimes radiation therapy may be provided during emergency.
- Treatment for high-risk tumor includes all the above mentioned options. Its survival rates may be 40% - 50%.

1.4 Support Vector Machine

Support vector machine is used for classification and regression. It analyses data and recognize patterns. It is defined as systems which uses set of all hypotheses for its processing. It uses high collection of features which are used to characterize our data. This method is learned by an algorithm which helps in training. Theory of machine learning is used for classification of data with the help of this. It maximizes the accuracy of the data. It is a classifier which is derived from learning theory. This learning system thus includes:

- Set of hypotheses
- Collection of many features to classify data
- System is trained by an algorithm

When set of training examples are given, which may belong to different categories, then support vector machine helps in classifying new training examples into different groups. In this examples are represented as points. Set of examples are mapped in this space and then differentiated. New examples are then compared to that space and they are assumed to fall on particular side of the gap they belong to. Support vector machine thus finds a hyperplane as shown in figure below. In this example, the objects belong to either class brown or class purple. The separating line defines the hyperplane. As a result, maximum points of similar type results in same side of the class. Thus hyperplane maximizes the differentiation of two classes.

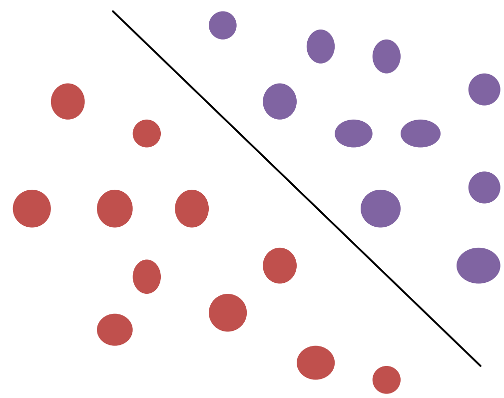


Figure 1.1: Hyperplane separating objects

1.4.1 Kernels in SVM

Kernel methods are used for analysis of patterns in machine learning. Kernel functions enable these methods to work in high dimensional feature space. Kernel can also be called a similarity function. It is the function that is provided by the user to the learning algorithm. For instance to classify images a single kernel function is defined. This kernel function along with the images is provided to the learning algorithm which classifies the image. Mathematically kernels are used as described below:

- Each kernel function can be presented as a dot product in a feature space.
- Machine learning algorithms can be further expressed as a dot product.

Reason for using kernels as compared to feature vectors is that it is easier to compute a kernel. For a particular kernel function the feature vector can grow much higher in size. Thus is easier to use kernel methods. Many algorithms are written to only use dot products which can be replaced by kernel methods. By this feature vector is not used at all. Thus user can work with efficiently performing kernels without the explicit computations of feature space coordinates. This technique is called the kernel trick. So the only restriction imposed is that only the methods which can be expressed as a dot product can use kernels. So if one feature vector is transformed into a higher feature space, then it is a new feature vector not a kernel. As kernels can do these transformations without raising the dimensions. The kernel functions represent the dot product of input feature space mapped into higher dimensional feature space by transformation Ø. It is represented as:

$$\mathbf{K\ (Xi.\ Xj) = \text{Ø}\ (Xi).\ \text{Ø}\ (Xj)}$$

Different kernels used in support vector machine models are:

1. Linear kernel: They are best to apply on linearly separable data. It does not perform any mapping. It is faster to train the classifier than with other kernels. So it performs well when there is a linear decision boundary. Mathematically it is represented as:

   $$\mathbf{K\ (Xi.\ Xj) = Xi.\ Xj}$$

2. Polynomial kernel: It is the non-stationary kernel. It represents the similarity of training samples over polynomials allowing learning of non-linear methods. In addition to looking at the given features, it also takes into account the combinations of input samples. For d degree polynomial the polynomial kernel is defined as (Men, H. et al, 2009):

   $$\mathbf{K\ (Xi.\ Xj) = (\gamma\ Xi.\ Xj + C)^{d}}$$

3. RBF kernel: The radial basis kernel function (Gaussian) is more popular function used in different learning algorithms. The RBF kernel on two sample Xi and Xj, represented as feature vectors, is defined as (Men, H. et al, 2009):

   $$\mathbf{K\ (Xi.\ Xj) = exp\ (-\gamma|\ Xi - Xj\ |^{2})}$$

4. Sigmoid kernel: It is also called Multilayer Perceptron (MLP) kernel. This function is originated from neural networks where bipolar sigmoidal function is used as activation function. When SVM use this kernel it is equivalent to a two layer Perceptron neural network. Slope γ and the intercept constant C are the two adjustable parameters in the sigmoid kernel. It is defined as (Men, H. et al, 2009):

   $$\mathbf{K\ (Xi.\ Xj) = tanh\ (\gamma\ Xi.\ Xj + C)}$$

1.4.2 Applications of SVM

SVMs can be used to solve various real world problems as described below:

- Text (and hypertext) categorization: the main task of this application is the grouping of original text (or hypertext) documents into a fixed number of previously defined categories. Examples include email filtering, web searching, and sorting documents by topic, etc.
- Image Classification: it can also be performed using SVM. Experimental results show that higher accuracy is achieved in this task as compared to other methods.
- Bioinformatics (Protein classification and Cancer classification): SVMs are helpful in medical sciences and proteins classification as described by (Yang, Z.R., 2004).
- Handwritten character recognition: handwritten content can be easily characterized or recognized using SVM.

1.5 Introduction to Image Processing

1.5.1 Image

An image refers to an array of square pixels that are arranged in rows and columns. It is defined as a two-dimensional array. It is denoted by f(x, y) where x and y are coordinates of plane. The intensity of an image at some particular coordinate is defined as the amplitude of that point. An image is called digital image when the values of x, y, amplitude are all discrete finite values. The elements of the image are known as picture elements or pixels.

1.5.2 Digital Image Processing

Digital image processing is a method of converting an image into digitized form. It also includes performing some operations on it so that some useful information can be obtained from that image. It is the sub-field of digital signal processing. It helps in making changes in the digital images with the help of computers. The output of this technique is either an image or features extracted from an image. Basically image processing consists of three steps: importing the image by digital photography, performing various operations for enhancement of image, and output obtained is the processed image or data based on image analysis. Image processing steps are performed for various purposes. It may include creating better image, obtaining the area of interest, differentiating the objects in an image, observing of invisible objects, and measuring of patterns.

1.5.3 Steps in Digital Image Processing

Fundamental steps of digital image processing are described in the diagram below:

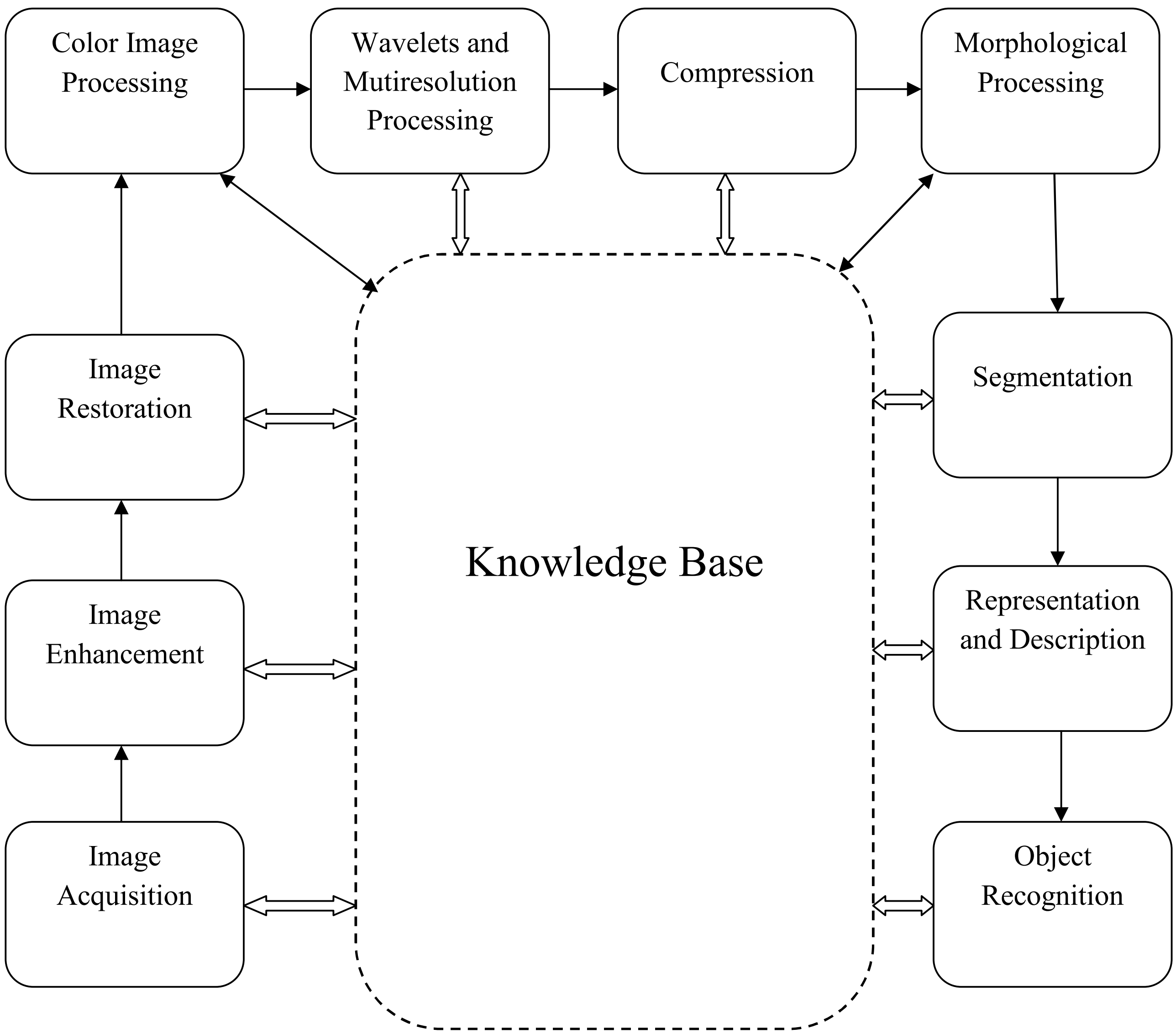


Figure 1.2: Fundamental Steps in Digital Image Processing

Image Acquisition: This is the first step of available fundamental steps of image processing. It is the process of acquiring an image. This step can be the simplest when the acquired image is already in the digital form. It may generally include preprocessing steps like scaling of an image.

Image Enhancement: It is the simplest area of digital image processing. This step is performed to improve the image so that further steps can be performed appropriately. The basic idea behind this step is to enhance the area of interest and obtain few characteristics of an image. Methods of image enhancement include:

- Spatial domain method which are procedures operating directly on pixels. This includes point processing which is the process of contrast enhancement. In this step image of higher contrast is achieved by darkening a particular level of image.
- Median filtering is the smoothing technique included in this step which removes the image blur. Max/min filtering is performed where the max or min values of the neighbourhood pixels replaces the candidate pixels.
- Image subtraction is the method where the difference between the corresponding pixel pairs is computed. It has area in background elimination.
- Histogram equalization can be performed which improves the appearance of the image.
- Image smoothing is the frequency domain method which reduces the camera noise effect and missing pixel values.
- Image sharpening method highlights the details in image and enhances the blurred part.

An example of enhanced image after performing deblurring operation using wiener filter is shown below:

Figure 1.3: Image Enhancement

Image Restoration: This step tries to restore the images which have been degraded. Image enhancement described above is a manipulative process which modifies an image while image restoration is an objective process, based on mathematical models. In this step first the degradation process is identified and an attempt is made to reverse it. Restoration techniques may include:

- Inverse filtering
- Linear filtering
- Non linear filtering

An example of image restoration is shown below:

Figure 1.4: Image Restoration

Color Image Processing: It is the area that uses digital images including color modeling and processing. Wavelets are used for representing images in various degrees. Primary colors are red, green, blue and secondary colors are magenta, cyan and yellow. The characteristics used to distinguish one color from another are hue, saturation and brightness. Hue describes the wavelength or dominant color as seen by the user. Saturation describes the relative purity and the amount of white light mixed with hue. Various available color models are: HIS (hue, saturation, intensity) model, HSV (hue, saturation, value) model, YIQ (luminance-inphase-quadrature) model, CMY model.

Full color image processing method includes processing component image individually and forming a processed image from individually processed component. Another method may be working with the color pixels directly.

Wavelets and Mutiresolution Processing: With the help of wavelets images can be represented in different resolution. Mutiresolution analysis is the representation of images in more than one scale. It is helpful in detecting features which were not detected in one resolution. Image compression, transmission, and analysis are the applications of this.

Wavelet transform decomposes the image to small waves of changing frequency and small time duration. Small objects can be seen at high resolutions using this. Steps in image processing by wavelet transform include: Decomposing the image into wavelet domain, then wavelet coefficients are changed according to the applications, then the image is reconstructed with the changed coefficients. Image pyramids is a collection of images at decreasing levels of resolution. The original image forms the base of the pyramid.

Compression: It handles the methods for decreasing the storage requirement for saving an image. Web page images and high resolution digital images can be compressed using this. Image compression plays an important role in many areas. Data compression is the method of reducing an amount of data required to represent the given amount of information. Images can have various redundancies. It can be coding redundancy which means image array contains more number of bits than that needed to represent the intensity. Other may be spatial and temporal redundancy and irrelevant information.

General image compression model is shown in the diagram below:

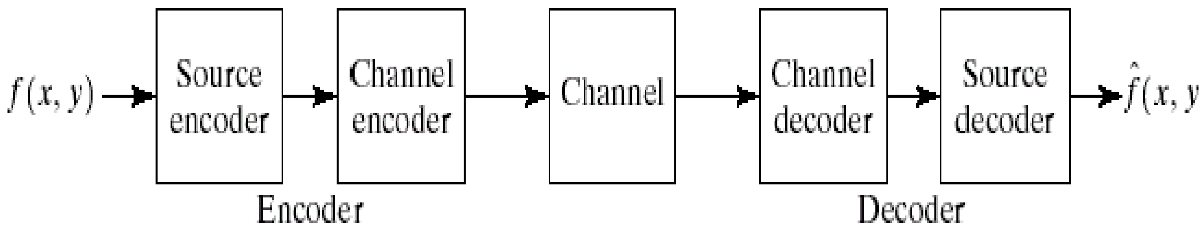


Figure 1.5: Image Compression Model

Compression can be lossy and loss-less compression. Lossy compression is based on compromising the accuracy of the reconstructed image. The distortion proposed by this compression is irreversible. It is basically a three step algorithm. The first step is to remove the redundancies. This is done by encoder. Then quantizer is applied which keeps irrelevant information out of compressed representations. This is an irreversible operation. This step is omitted when the compression is loss-less. Then as a final step fixed or variable length code is generated by system encoder. An example of images before and after compression is shown below:

Figure 1.6: Image before Compression

Figure 1.7: Image after Compression

Morphological Processing: It is the method of taking out the components of the image which are important. It is the description, analysis and identification of the smallest unit of image. Its basic operations are erosion and dilation. Using this operations opening and closing can be performed in an image. These operations are performed on binary images. Dilation process is responsible for the objects to grow in size and erosion process makes the objects smaller in size. The basic effect of opening is somewhat like erosion. It removes some of the bright pixels from edges of regions of foreground pixels. Closing is similar to dilation. It tends to enlarge the boundaries of foreground (bright) regions in an image. The exact operation is determined by the structuring element which consists of pattern specified as the coordinates of number of discrete points.

Segmentation: This step divides the image into parts. The output of this step is generally attributes of an image. Categories of segmentation process include:

- Threshold based segmentation: It is used for discrimination foreground from background. Histogram thresholding and slicing techniques are used to segment the image.
- Edge based segmentation: By this segmentation edges are detected and they are used to represent the object boundaries which are further used to identify these objects.

Edge based segmentation methods include Sobel, prewitt, log, canny, Roberts edge detectors.

- Region based segmentation: This is opposite to edge based segmentation which starts from the middle of an object and grows out until it meets the object boundaries.
- Clustering techniques: It is the process of identifying groups of similar image primitives which means organising the objects based on its attributes.

Figure 1.8: Image before segmentation

Figure 1.9: Image after segmentation

Representation and description: This step is followed by the segmentation step. After the image segmentation the boundary of a region is obtained. Transforming this data into suitable form is the solution for representing an image and making it suitable for computer processing.

Description of an image tells about important attributes information obtained from previous results. Its basic idea is for differentiating one type of objects from another.

Object recognition: This task includes identifying an object in an image. An image is given initially with some area of interest and few set of labels. The task of this step is to assign correct labels to the regions in that image. This step is closely similar to segmentation step.

Knowledge Base: It tells about the details of regions in an image which contains the area of interest. It may also be complex as it can have details of high resolution satellite images. List of details may also have possible defects.

1.5.3 Applications of Image Processing

Some major applications of digital image processing include:

- Image sharpening and restoration: In this operations on images can be performed like zooming, sharpening, blurring, conversion from one form to other, detection of edges. Some examples are shown below:

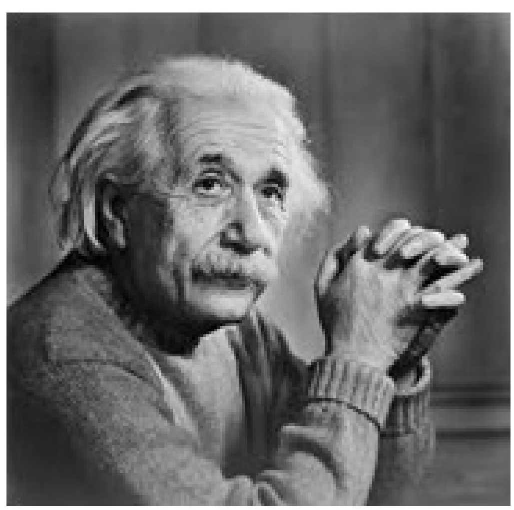


Figure 1.10: Original Image

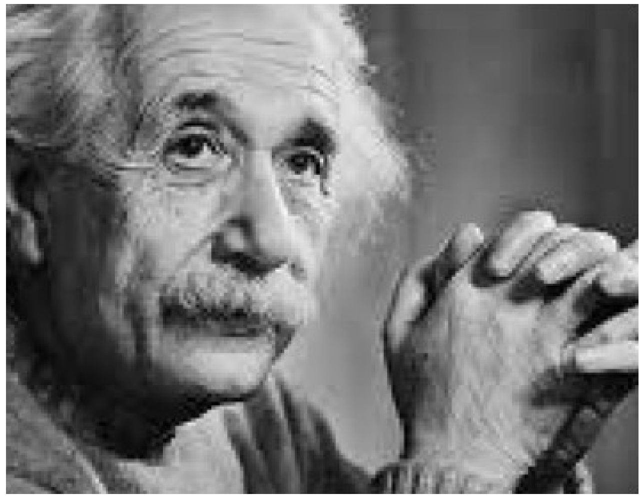


Figure 1.11: Zoomed image

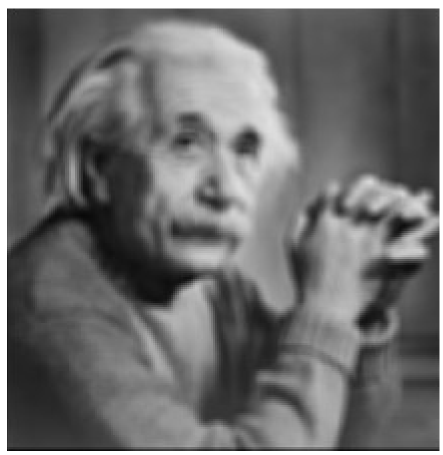


Figure 1.12: Blurred Image

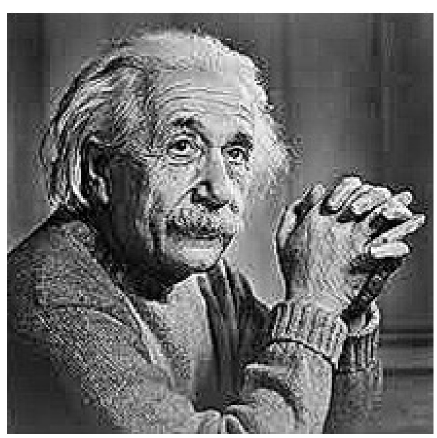


Figure 1.13: Sharpened Image

- Other applications are in medical field which includes:
    - Detection of cancer
    - UV imaging
- Remote Sensing
- Color Processing
- Pattern Recognition
- Video Processing
- Robot Vision

CHAPTER 2

REVIEW OF LITERATURE

Hsu, C.W., and Lin, C.J., (2002) presented a comparison between different methods of multiclass support vector machine. Support vector machine is an algorithm basically designed for binary classification. So in case of multi-classes several algorithms have been put forward. Solving multi-class problem all at once is cost effective, so large data is not considered in this work. Authors in this work implemented two all-together methods and compared the results with methods based on binary classification that are one-against-one, one-against-all and directed acyclic graph SVM (DAGSVM). But these experiments were limited to small data sets. As a result it is seen that two methods were efficient for practically implementing multiclass classification. These methods were one-against-one and DAG method. The future scope of this work is to test data having many classes.

Li, G.Z., et al., (2004) proposed feature selection for multi-class problems. This is done by support vector machines. Feature selection is a method covered under image processing and machine learning areas. Its role is to remove the unnecessary noisy features and improvement in performance of the system. Support vector machine is basically designed for binary classification. SVM as compared to other neural network methods like back propagation network has better generalization ability. In this work authors introduced a prediction risk based feature selection method. It uses multi-classification SVM. To evaluate the features measure of sum is calculated for equivalent features of individual binary SVM. Result is compared with optimal brain damaged based feature selection method. As a result it is seen that this introduced method shown better results as compared to earlier methods using SVM.

Yang, Z.R., (2004) presented the biological applications of support vector machines. Computer programs are now a day's becoming important for the classification process of biological data. Best predicted performer at present is support vector machine for binary classification. It increases the separation margin between two classes. In this work protein and DNA sequences are tested by SVM and its applications are discussed. Classification requires the set of features to train models and mapping is done from these features to class labels. For DNA and protein sequence analysis three stages are performed.

Liu, Y., and Zheng, Y.F., (2005) in their work presented multiclass SVM classification using reliability measures. Support vector machine is a learning algorithm basically designed for a binary classification. In case of multi classes this method is extended to one-against-all method which classifies M number of classes by taking a series of two class problems. One drawback of this method was that when all the m classifiers were collectively checked to get the final results, only decision function value is considered but the quality of classifiers were not kept into account. So to overcome this problem, in this work, reliability measures were introduced. The two measures used were static reliability measures (SRM) and dynamic reliability measure (DRM). SRM results in a constant value independent of test sample location. DRM is based on the test sample's location as reliability is checked by this method in the region surrounding the test sample. As a result of this work improved classification accuracy was achieved.

Samavi, S., et al., (2006) presented processing and compression of DNA microarray images. In this work pipeline architecture is presented. In this image processing steps are performed at rows one by one at each clock pulse. Different structures are used for compression techniques. This architecture proved highly scalable. DNA contains all the information required to make protein. A technique is proposed for microarray images. A functional algorithm is implemented for compression of images. Time gap is considered according to clock rate. The overall result proved to be scalable and versatile.

Gurcan, M.N., et al., (2006) in their work used the cell nuclei segmentation and detection algorithm. The method used morphological operations like closing and opening to distinguish cells. It also uses hysteresis thresholding operation. This method tested many cell nuclei, and compared the results with the original automated output. Morphological operations are used so that differences due to staining are reduced. It helps in properly segmenting the cell nuclei. In this work as a result it was observed that all cell nuclei were not detected and segmented. That is those which are faintly visible were unable to be processed. Segmented cell nuclei are used as a step for the segmentation of cytoplasm. This helps in understanding mitiosis and karyorrhexis signs. This also helped in differentiating different kinds of cell.

Kong, J., et al., (2007) presented in their work the analysis of image for grading of neuroblastic differentiation. Peripheral neuroblastic tumors (pNT's) are originated from developing neuronal cells of the sympathetic nervous system which is a part of autonomous nervous system. They are found mostly in young children. Neuroblastoma medication is linked to grading of neuroblastic differentiation. It includes three categories of grade: undifferentiated, poorly differentiated and differentiating. These grading features are used for automatic differentiation using image analysis method. In this work researchers represented an automated system that involves different steps for final classification. In this the misclassification rate was relatively low as compared to other methods used. Some misclassifications were due to cropped images which had small size. This does not contain all major elements of interest

Kong, J., et al., (2007) in their work presented neuroblastic differentiation in the presence of multi-resolution framework consisting of images of two or more size. This resulted in saving the mathematical cost of the system. This multi-resolution technique divides all input images into many different representations. Seven classifiers are together used in this work, and every system works without linking with other. These classifiers are then used together by a two-step integration strategy. Based on this the resultant decision is made. Three grades are defined in this system as undifferentiated (UD), poorly differentiated (PD), and differentiating (D). This system is tested on large number of images and much accuracy was obtained. This system also helped in acquiring better classification results. So as a final result the doctors where able to properly classify the disease.

Chapelle, O., and Keerthi, S., (2008) in their work considered multiclass problems where set of features are extracted for each class using support vector machine. Feature selection is an important method for text classifications and bioinformatics applications. In case of many noisy features this method is useful for reducing the load. In case of text classification when the average features are more in number then it takes more time for processing all features for individual class. So this work is motivated for simultaneous multiclass feature selection. Categories involved in different classification problems include a multi-class type in which to every example only one class is allotted and multi-labeled type in which more number of classes can be assigned. This methods performed better feature selection. For implementation of this work support vector machine is used as a particular model and new method is developed based on scaling factors.

Mathur, A., and Foody, G.M., (2008) presented multiclass and binary SVM classification. The main objective of classifying the data is to achieve better accuracy using very less number of training examples. Support vector machine is such classifier which is used in this work. It is primarily designed for binary classification. It can be extended to multiclass classification using a series of binary classifiers using different available strategies. But in this work another approach is used that is one-shot multiclass classification which shows better efficiency than the single and series of binary classification. Another advantage of this approach is that it showed much accuracy even when using less number of support vector as compared to other binary based approaches. SVM classification shows maximum accuracy as compared to other approaches like neural network, decision tree, etc. This is so because it requires small size of training samples to achieve maximum likelihood. Mainly focus is done on the confusion matrices achieved and training vectors number which indicates the cost of training data. It was also seen that complete confusion matrix was not achieved by one-against all strategy which can limit the result calculation.

Men, H., et al., (2009) in their work made classification tests based on support vector machine and neural networks individually. The different kernel functions used for classification using SVM were polynomial kernel and radial basis function. Cross-validation method is used to select the kernel function in order to achieve better accuracy in results. These results of SVM are further compared with the neural network method results. Linear SVM is based on creating a hyperplane as a solution. A penalty factor is also introduced in this work so that on the basis of that value generalization ability can be made better for the learning machine. On the basis of experiments separate results were achieved for both methods and compared. The results show that the SVM algorithm reaches much higher accuracy for classifying any data. Stability is also achieved by this method as compared to the other algorithm tested which is easily affected by some weighting factor.

Sertel, O., et al., (2009) in this study presented image analysis approach which operates on digitized neuroblastoma (NB) histological images. In this system the important role is played by the amount of cells undergoing either mitiosis or karyorrhexis. According to this work, the classification system, classified tumor as either favorable or unfavorable. One important method for analyzing behavior and the medical opinion of patient suffering from this disease is the understanding of surgical image samples. Critical step in this method is to identify the mitiosis-karyorrhexis index (MKI). MKI is the count of tumor cells present. Mitiosis is the process division of cell into two cells. Karyorrhexis is the division of the nucleus of a dying cell which is damageable in nature. A higher value of MKI shows a higher chance for the tumor to be unfavorable. MKI is calculated after a long microscopic study of hematoxylin and eosin (H&E) stained slides that too at high scales. The doctor study areas under slide and count the number of MK cells among many cells present.

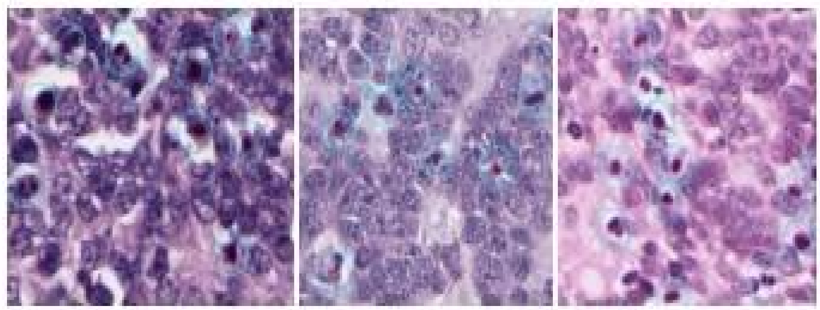


Figure 2.1: Sample images with MKI cells circled

Hong, S., et al., (2010) in their work presented the classification of chest DR image using support vector machine. The shortfall of poor ability of inference ability of earlier methods is improved. The two kind of images used are original and the defected image. This is done by improving the accuracy of two kinds of images. Support vector machine (SVM) method was used on the chest digital radiography images of children. The images were divided into normal and defected images. This system aims in building hyper-plane. It distinguishes different kind of data as shown below. This proves that the classification of these different types have large gap. This work resulted in efficiency in case of small samples. This result is used by doctors for early diagnosing of the disease. This results in good rate of accuracy.

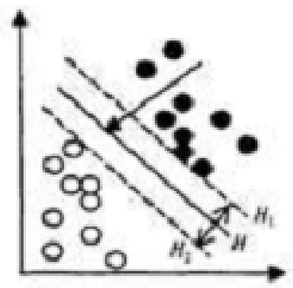


Figure 2.2: Optimal separating hyper-plane

Gu, H., et al., (2010) presented one-class support vector machine with relative comparisons. The hypersphere and the hyperplane version of a one-class SVM was proposed in this work. The feature space, in a hypersphere model, forms a minimum surrounding sphere in which many point are enclosed. And in the hyperplane model the given points are separated from the origin with more boundaries. In both the methods only positive samples are provided to map the data into the feature space. For improving the performance semi-supervised methods are employed which contains negative samples. In this work in one-class SVM certain constraints are considered which leads to convex quadratic programming (QP) problem. This as a result finds minimum sphere with most of the target points enclosed in it. This one-class SVM can be applied to different problems. In semi-supervised learning limited supervision is done. As a result this showed better accuracy as compared to other SVMs.

Stanescu, L., et al., (2011) presents study of methods for medical image segmentation. The images represent digestive system attributes. Segmentation of medical images is very vital in biomedical field. Three segmentation algorithms are studied in this work. They are successfully used in color images segmentation. For their comparison error measuring methods are used and different images are further compared. As a result one best method is obtained.

Saxena, P., and Singh, S.K., (2012) in their work done grading of Follicular Lymphoma and Neuroblastoma. Histopathological grading of FL is based on count of Centro Blasts (CB) per high power microscopic field (HPF) and categorization of neuroblastoma is done into favorable and unfavorable histology on the basis of International Neuroblastoma Pathology classification. In this work the main focus is to distinguish among nuclei, cytoplasm, extracellular material and red blood cells from H&E stained input image. In this Gabor function application technique is used for texture characterization which creates different kind of environments for visualizing image. Gaussian wavelet orientation transformation calculation of colored images before calculations of H&E stained histopathological images improved the classification performance.

Wang, X., et al., (2012) in their work used support vector machine (SVM) for image segmentation. It is the machine learning algorithm which can reduce the segmentation error caused by fast motion of the object. Image segmentation aims to classify an image to the foreground and the background. By extracting the features of foreground, the SVM classifier is trained and then it is used to classify the object image. This method reduces the segmentation error which is caused by fast motion of the object. In this paper firstly the frame difference combined with morphology of mathematics was applied to extract the object roughly. Then few characters of the image were computed for training SVM. Finally decomposed SVM binary decision tree was used for classification. They concluded that the selection of features and the character of the images have a great relationship. If the background and foreground are quite different, then the classification results were much better. Otherwise the SVM classifier may have misjudgment.

Saxena, P., et al., (2013) in their presented that Histopathological grading of Follicular Lymphoma (FL) is dependent on count of Centro Blasts (CB) per high power microscopic field (HPF) and categorization of neuroblastoma is done into favorable and unfavorable histology on the basis of International Neuroblastoma Pathology classification. In this work the main focus is to distinguish among nuclei, cytoplasm, extracellular material and red blood cells from H&E stained input image. In this Gabor function application technique is used for texture characterization by visualizing each object of image with all possible angles. By this method 12 different masks were obtained resulting in 12 different representations that are summarized in corresponding 12 images. Then these images are combined into one textured image which is represented as 3-dimensional representation of input image. Gaussian wavelet orientation transformation calculation of colored images before calculations of H&E stained histopathological images improved the classification performance.

Bhuiyan, A.H., et al., (2013) presented features extraction using image processing for skin cancer. Cancer is a disease in which normal cells grow in size and get affected. It is difficult to differentiate between normal and cancerous cells. Out of many forms of cancers skin cancers are most common. It is two types out of which the dangerous one that is the malignant melanoma is checked in this work. As the early diagnosing and treatment of cancer can reduce its risk, so this work is done to analyze the patients in their early stage of cancer. In this work three steps are performed to complete its task. Firstly supervised segmentation is done various parameters of skin shape types and textures. It is followed by feature extraction step and finally lesion recognition is done. Unsupervised methods are not used in this work because for extracting features of this particular cancer it is difficult task. Feature extraction is the important step for analyzing the images properly. In this work many digital images have been tested and analyzed. They were tested for asymmetry, border irregularity, color variation and diameter measurement. All these analyses are important for detection of this cancer. The risk of melanoma can be decreased by limiting the exposure to sunlight and protecting your skin.

Saxena, P., et al., (2013) in their work tested Neuroblastoma and Prostate cancer. An approach is proposed for determining the shape of every nucleus. This works on local intensity fitting techniques which defines the local fitting criterion function for image intensities in the neighborhood of each point. In this modified active contour model was used and an algorithm based on local intensity clustering. These models were able to segment overlapped nuclei and cover up entire area of image. Fewer computations and much less iterations were resulted as compared to previous techniques used.

Abu-Mahmoud, M.K., and Al-Jumaily, A., (2014) in their study proposed an automated system for discrimination between melanocytic nevi (birth mark or mole) and malignant melanoma (lesions or wound or tumor smaller than 6 mm in diameter that is merely a dot). The general approach for developing a computer aided diagnostic system for diagnosis of skin cancer is to find the location of a lesion and also to determine an estimate of the probability of a disease. Firstly pre-processing was done which included resizing, cropping, and hair removal to enhance its important features. This system used a number of features extracted from images of skin lesions through image processing techniques. Then before segmentation, methods to enhance the contrast of images are applied. Then extracted features are reduced by using sequential feature selection. Finally SVM classifier is used to diagnose tissues of patients suffering from melanoma. Higher accuracy is achieved in the proposed work.

Kumar, G., and Bhatia, P.K., (2014) review and applications of feature exaction in image processing. Features play important role in image processing. Before actually extracting features pre-processing techniques are applied on images. After that feature extraction steps are performed and the extracted features are helpful in identifying of images. The application discussed in this proposed work is for character recognition. Efficiency in classification and recognition is calculated. Feature extraction helps in evaluating the relevant shape. Its main goal is to obtain important information from the data. It is useful in various other applications also. Selecting of meaningful feature is also important additional features in work leads to worst performance. Two types of features that are local and global features are evaluated in this work. Usefulness of features is also explained in this work. Different feature extraction techniques have been developed and discussed in detail in this study. And as a result it is described which technique is better.

Chinnathambi, K., et al., (2014) feature extraction of cancer affected white blood cells. Cancer cells grow multiplicatively so it becomes difficult to count white blood cells (WBCs). In this work segmentation algorithm is proposed which helps in segmenting these cells so that area of interest can be easily found. This step is followed by feature extraction. If the data contain large amount of values, then by feature extraction redundant data can be removed. The aim of this work is to get more meaningful image by applying particular methods so that WBC count can be made easily and can be differentiated with cancer cells. Different segmentation methods are applied followed by the feature selection process. And finally this is compared with several images which are helpful in identifying tumors.

CHAPTER 3
PRESENT WORK

3.1 Problem Formulation

This thesis work deals with grading of neuroblastoma cancer using support vector machine. This is the cancer of early childhood which starts from sympathetic nervous system. It generally starts in the adrenal glands (two small glands above the kidneys). It can also start from the neck, spinal cord and chest. It occurs commonly during early childhood period. It is a complex disease occurring in babies or infants or children less than 3 years of age.

Specifically this study seeks to solve the problem of grading of neuroblastoma tumor and classification of attributes so that proper prognosis can be provided by doctors and disease can be properly cured. This work also evaluates the best kernel function as the attributes are tested with all them individually. This study intends to:

- Describe the performance of the attributes used for grading.
- Identify the best kernel function.
- Grade the disease as undifferentiated, poorly differentiated, and differentiating.

Further the methodology section describes the step by step evaluation of these factors. Firstly data is acquired and useful attributes are selected from that data. Support vector machine is applied with all possible kernel functions and results are obtained. Performance of all attributes is evaluated and grading is done. Firstly SVM is applied on image data. From that the best kernel function is evaluated. Then the same is applied on other data set consisting of clinical data. Out of this data useful attributes are selected which are important for grading of cancer. Then best kernel function is acquired and performance of grading is obtained.

3.2 Objectives

The main objective of this research topic is performing grading of neuroblastoma cancer for its proper treatment. If it is correctly diagnosed, then accordingly treatment will be provided and will give better results. Other objectives include increasing performance of the system and increase its efficiency so that doctors can use this to properly diagnose the cancer. For this work support vector machine classifier is used. This evaluates the performance of the attributes used for grading of the disease. Kernel functions are also evaluated for its performance. Finally the disease is classified for undifferentiated, poorly differentiated and differentiating cases. Matlab software is used for its implementation.

3.3 Methodology

Methodology used in this work is described as below:

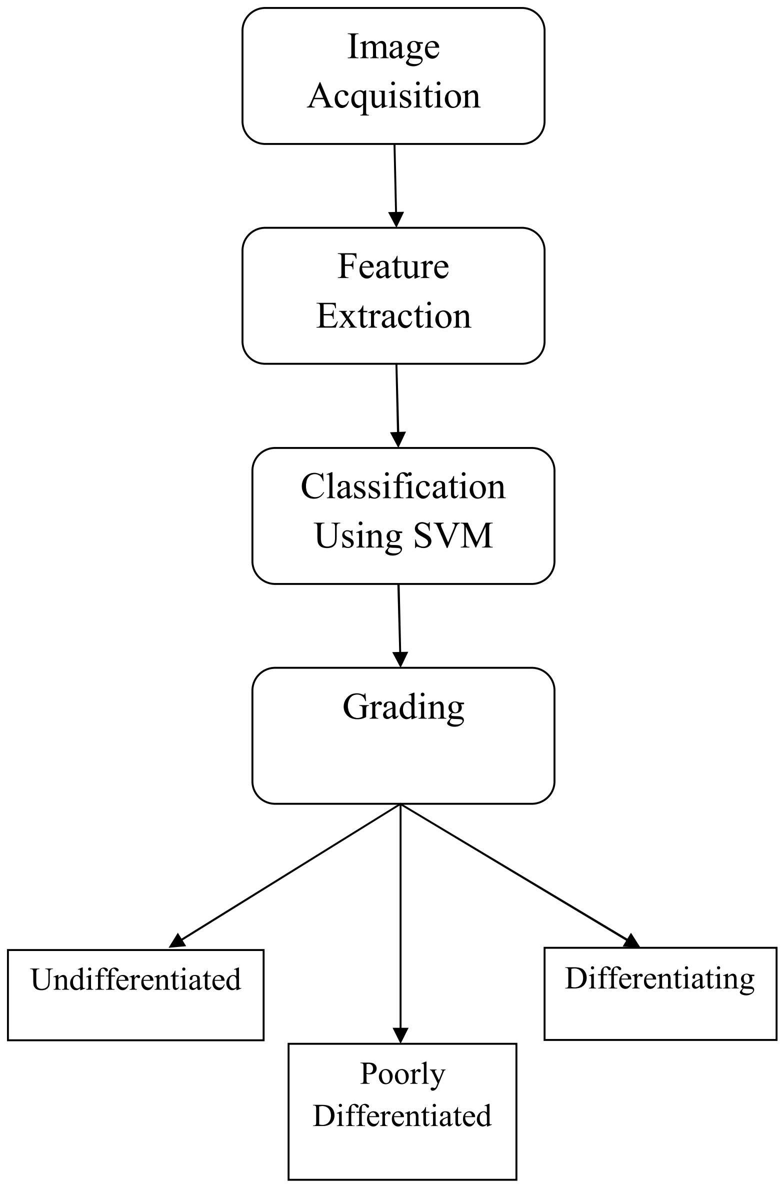


Figure 3.1: Research Methodology

Steps involved in that work include:

1. Image Acquisition: Images of neuroblastoma tumor are acquired from Pathology Education Informational Resource (PEIR) digital library and further steps were performed.

2. Feature extraction: Features extracted from these images were entropy, saturation, contrast, brightness. These are described as below:
    - **Entropy:** It is actually a physics concept which measures the amount of confusion or disorderliness in the system. A structure which is ordered has low entropy. As this system gets heated and changes to less ordered state, then entropy increases. Entropy can also be defined as number of states adopted by a system. A system with fewer states has low entropy and with more states has high entropy.

      In case of images gray levels of the images is responsible for change in entropy which corresponds to states. If the image is equalized using histogram, then the number of states is more and so is the entropy. While if image has two states, then entropy decreases. The entropy H is defined for an image as:

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

      M is the gray level of the input image and p_k is the probability of kth gray level.

      In this proposed work inbuilt function of matlab entropy is used to calculate its value for an image. This works on a grayscale image. So if image has more than two dimensions then this function treats that image as multidimensional grayscale image.
    - **Saturation:** It tells about the intensity of color in an image. Diagrammatically it is represented as steepness of the slopes of the curve. As saturation increases more pure color appears. It is also linked with contrast, brightness and sharpness of an image. With increase in saturation, these values also increase. Black and white images have no saturation.

      In this work saturation value is calculated for collected images and further tested.

- **Brightness:** It makes the image brighter or darker. It tells about the intensity of energy. It is the total value of energy. In this work brightness is calculated for all images in matlab and further used a feature for classification.
- **Contrast:** Contrast can be explained as the difference between the minimum and maximum intensity of pixel in an image. It may also be explained as amount of color or grayscale difference existing between different features of an image. Higher contrast image displays higher degree of color difference as compared to lower contrast image. This value is also calculated for dataset of images collected for grading and applying classifier.

3. Classification using SVM: Support vector machine is used for classification and regression. It analyses data and recognize patterns. It is defined as systems which uses set of all hypotheses for its processing. It uses high collection of features which are used to characterize our data. This method is learned by an algorithm which helps in training. Theory of machine learning is used for classification of data with the help of this. It maximizes the accuracy of the data. It is a classifier which is derived from learning theory. This learning system thus includes:
    - Set of hypotheses
    - Collection of many features to classify data
    - System is trained by an algorithm

    When set of training examples are given, which may belong to different categories, then support vector machine helps in classifying new training examples into different groups. In this examples are represented as points. Set of examples are mapped in this space and then differentiated. New examples are then compared to that space and they are assumed to fall on particular side of the gap they belong to. Support vector machine thus finds a hyperplane. As a result, maximum points of similar type results in same side of the class. Thus hyperplane maximizes the differentiation of two classes.

    In this work SVM is applied on features extracted from images. Entropy, saturation, contrast, and brightness are individually classified with support vector machine and further grading is performed.

4. Grading: Neuroblastoma tumor is classified as undifferentiated, poorly differentiated and differentiating. So after applying SVM on extracted features of acquired image, grading is performed and maximum performance is achieved. Each attribute is tested with every kernel function of SVM. The results obtained are discussed in chapter results and discussions.

In this work SVM is applied and tested on two different data sets. First dataset consist of images acquired from Pathology Education Informational Resource (PEIR) digital library. Firstly SVM is applied on images whose attributes extracted are entropy, saturation, contrast, brightness. This tells about the quality of the images. The results were obtained and compared. The other dataset consist of clinical data file of neuroblastoma cancer. This data was collected from National Cancer Institute's (NCI) website. NCI's Office of Cancer Genomics (OCG) provided three programs out of which data is collected from TARGET (Therapeutically Applicable Research to Generate Effective Treatments) data matrix. From this dataset important features were selected which are stage, histology, MKI (Mitiosis-Karyorrhexis index), diagnostic category, ploidy, MYCN status. This attributes were classified with SVM and grading is done as undifferentiated, poorly differentiated, and differentiating. Each attribute is evaluated with every kernel function and results and analysis were performed. These results are compared to previously obtained results of image database.

The attributes of clinical data are explained below:

- **Stage:** According to International Neuroblastoma Staging System (INSS), the dataset consists of following stages as described in section 1.3.4:
  - Stage 1
  - Stage 2a
  - Stage 2b
  - Stage 3
  - Stage 4
  - Stage 4s

- **Histology:** It has significant impact on prognosis and risk group assignment of neuroblastoma tumor. This dataset describes tumors by International neuroblastoma pathology committee (INPC) as:
    - Favorable
    - Unfavorable

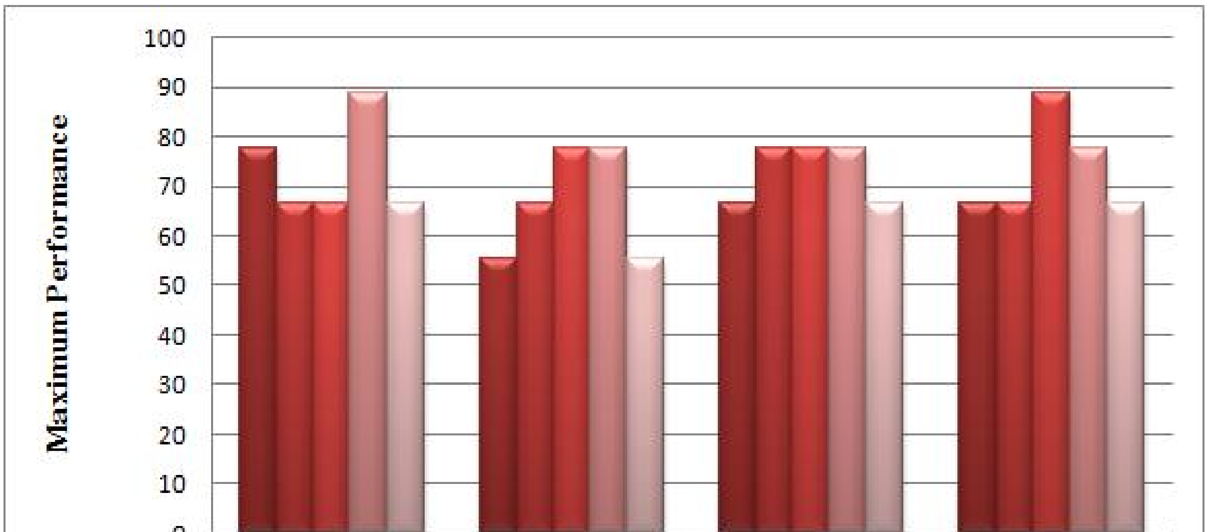
  Favorable and unfavorable are described on the basis of patient's age and MKI as described in section 1.3.5.
- **MKI:** MKI is the count of tumor cells present. Mitiosis is the process division of cell into two cells. Karyorrhexis is the division of the nucleus of a dying cell which is damageable in nature (Sertel, O., et al, 2009). A higher value of MKI shows a higher chance for the tumor to be unfavorable. INPC involves evaluation of tumor by MKI (Mitiosis-karyorrhexis index) of the neuroblastic cells. It has the following values:
    - Low
    - Intermediate
    - High
- **Diagnostic Category:** According to INPC, neuroblastoma is categorized as follows:
    - Neuroblastoma (Schwannian stroma-poor)
    - Ganglioneuroblastoma, intermixed (Schwannian stroma-rich)
    - Ganglioneuroma (Schwannian stroma-dominant), maturing subtype or Ganglioneuroblastoma, well differentiated (Schwannian stroma-rich)
    - Ganglioneuroblastoma, nodular (composite)
- **MYCN:** MYCN gene status is specified in dataset which is one of factor for classifying Neuroblastoma tumor. It is responsible for growth of a cell. Change in this factor causes cell to grow and divide faster. It is a protein which controls expressions of many target genes. Neuroblastoma with high amplification of this gene grows faster. In dataset this factor is specified by following values (Spitz, R., et al., 2006):
    - Not Amplified
    - Amplified
- **Ploidy:** It specifies the number of sets of chromosomes in the nucleus of a cell that is the amount of DNA in every cell. Cells with DNA index greater than 1 are called as hyperdiploid and with DNA index of one or less are diploid. In infants, hyperdiploid

cell indicates the earlier stage of neuroblastoma which can be treated easily. In dataset also this factor includes these values (Eckschlager, T., et al., 1999):

- Hyperdiploid (DNA Index >1)
- Diploid (DNA Index <=1)

CHAPTER 4
RESULTS AND DISCUSSIONS

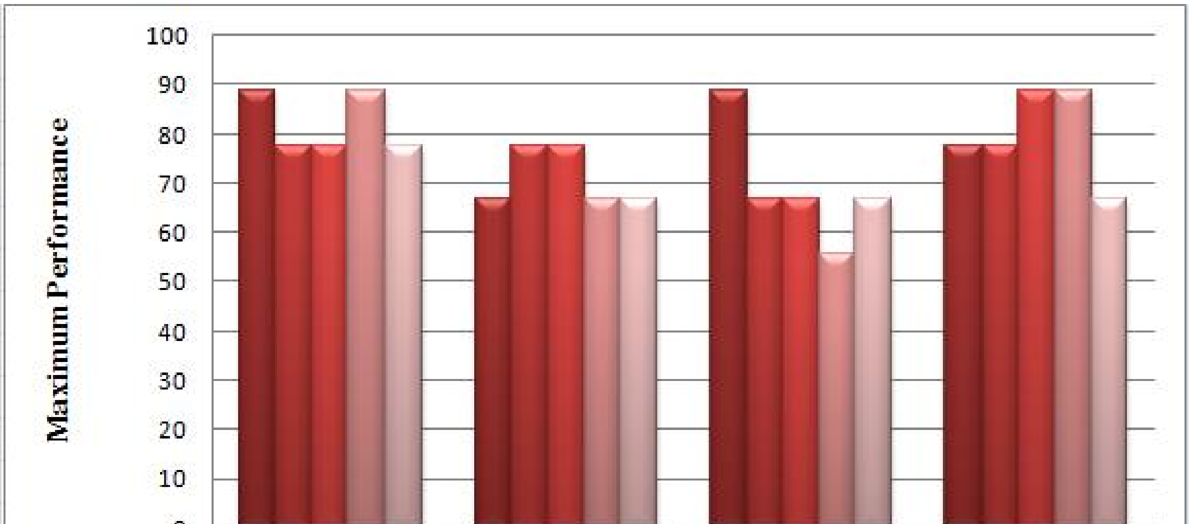
Support vector machine is applied on the data formed by features extracted from images that are entropy, brightness, contrast and saturation. For each feature, performance is evaluated by different kernel functions. Firstly data is trained and tested for undifferentiated cases. The results are shown in the figure 4.1 below. Maximum classification rate is seen by entropy and saturation. So these factors are important for this grade that is undifferentiated cases. Overall maximum performance of 88.89% is obtained. But almost equal performance can be seen so features cannot be evaluated properly. On taking average values it can be concluded that quadratic kernel shows maximum performance.



	Entropy	Brightness	Contrast	Saturation
Linear	77.78	55.56	66.67	66.67
Polynomial	66.67	66.67	77.78	66.67
Rbf	66.67	77.78	77.78	88.89
Quadratic	88.89	77.78	77.78	77.78
Mlp	66.67	55.56	66.67	66.67

Figure 4.1: Classification rate of undifferentiated cases for image data

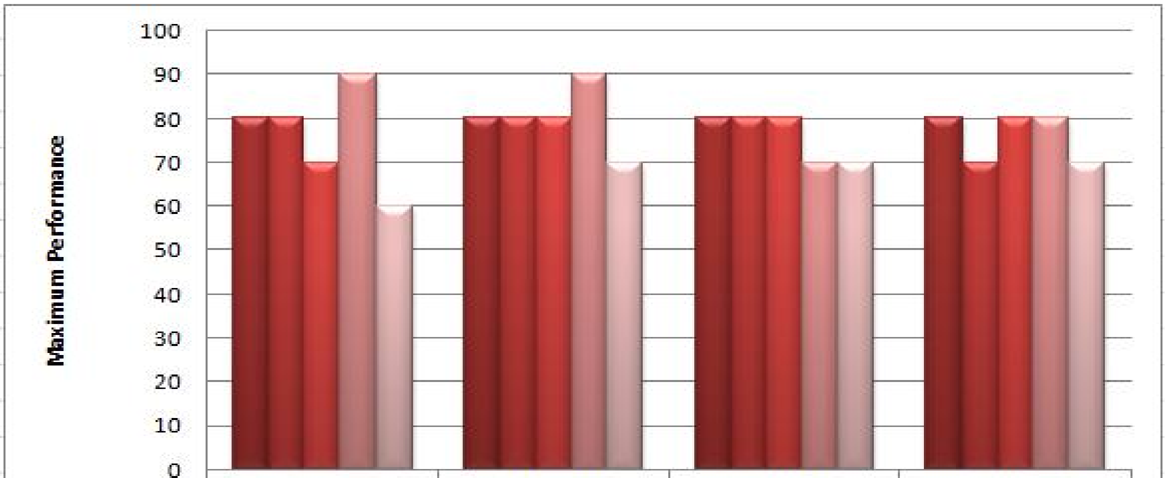
Next support vector machine is applied on poorly differentiated cases. Results obtained are summarized in figure 4.2 below. In this case all features obtained have equal maximum performance of 88.89% but with different kernel functions. Every attribute shows equal contribution so best cannot be evaluated by this data. By averaging the values, the best kernel evaluated is linear.



	Entropy	Brightness	Contrast	Saturation
Linear	88.89	66.67	88.89	77.78
Polynomial	77.78	77.78	66.67	77.78
Rbf	77.78	77.78	66.67	88.89
Quadratic	88.89	66.67	55.56	88.89
Mlp	77.78	66.67	66.67	66.67

Figure 4.2: Classification rate of poorly differentiated cases for image data

Finally the performance is evaluated for differentiating cases using SVM. The results are summarized in figure 4.3 below. Here the maximum performance can be seen by entropy and brightness with quadratic kernel. But the other features also show almost equal performance. The best performance obtained is 90%. By averaging the values, quadratic kernel can be evaluated best as in the case of undifferentiated cases.

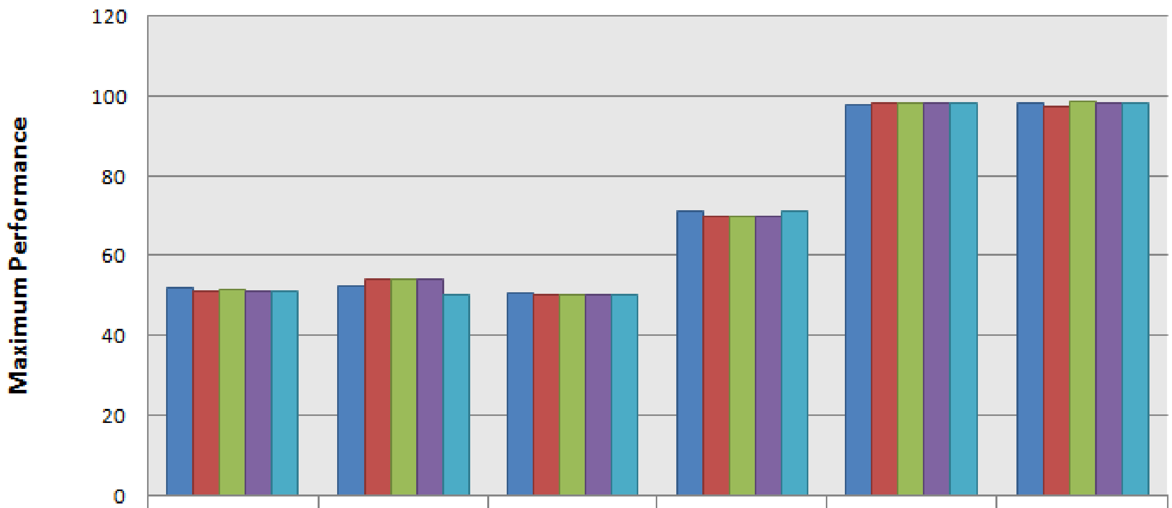


	Entropy	Brightness	Contrast	Saturation
Linear	80	80	80	80
Polynomial	80	80	80	70
Rbf	70	80	80	80
Quadratic	90	90	70	80
Mlp	60	70	70	70

Figure 4.3: Classification rate of differentiating cases for image data

Working on the data obtained from images does not properly classify the performance of each attributes. Almost equal performance is seen by them. Even the effect of using different kernel functions cannot be evaluated properly. In grade 1st and 3rd that is for undifferentiated and differentiating cases quadratic is evaluated as best and for poorly differentiated cases linear is best. So same methodology is applied but on different data set consisting of clinical data report and better results were obtained. Working on another attributes that are stage, MKI, diagnostic category, ploidy, histology, MYCN results and discussions are made and it can be concluded that they more efficiently evaluated the grading performance.

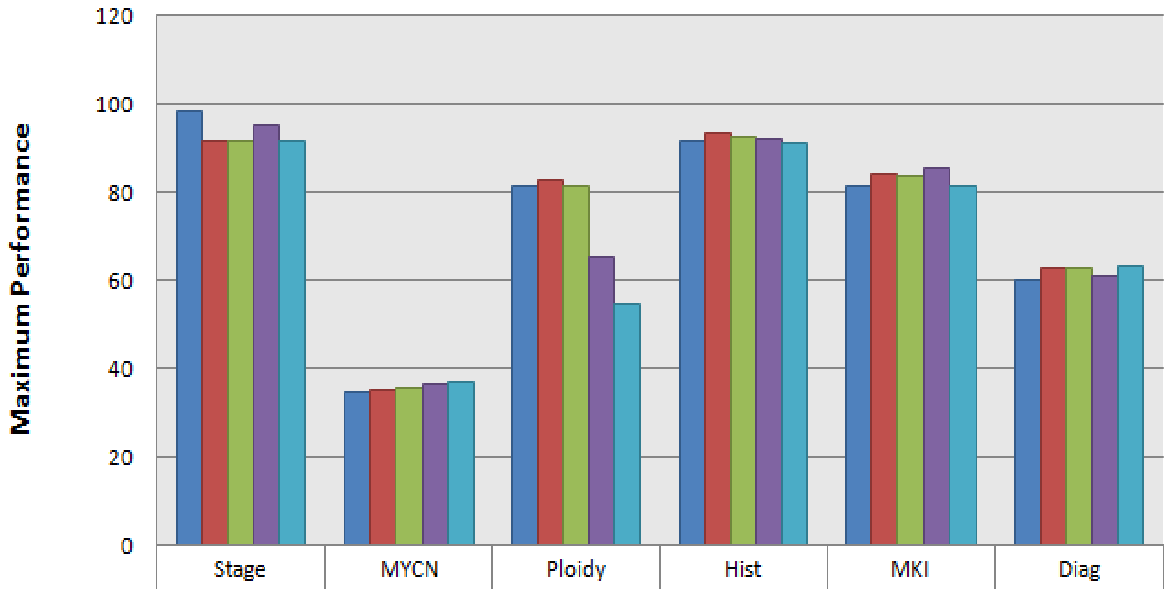
Working on clinical data, result shows that for undifferentiated or poorly differentiated cases, MKI and diagnostic category factor shows maximum performance in this case which is 98.12% and 98.44% respectively which implies that these attributes are most responsible for classification of this grade of neuroblastoma. MYCN amplification, stage and ploidy have not much contribution in this case. The approximate occurrence of stage, mycn and ploidy are 51.88%, 54.06% and 50.62% respectively for all kernel functions. Histology attribute has a maximum of 71.25% contribution in this classification using Multilayer Perceptron kernel. It is summarized in figure 4.4 below. On averaging the values of the results obtained it can be concluded that, best evaluated kernel in this case is rbf.



	Stage	MYCN	Ploidy	Hist	MKI	Diag
Linear	51.88	52.5	50.62	70.94	97.81	98.12
Polynomial	51.25	54.06	50.31	70	98.12	97.5
Rbf	51.56	54.06	50	70	98.12	98.44
Quadratic	51.25	54.06	50	69.69	98.12	98.12
Mlp	51.25	50.31	50	71.25	98.12	98.12

Figure 4.4: Classification rate of undifferentiated or Poorly Differentiating cases for clinical data

For differentiating cases maximum result is obtained by linear kernel with 98.12% by the stage attribute. It implies that if stage is correctly classified then differentiating cases of neuroblastoma can be classified. Histology attribute is almost equally important for this case with 93.42% performance. MYCN amplification factor shows least role in this case. MKI shows maximum of 85.27% contribution in differentiated cases. Ploidy attribute has a maximum of 82.76% contribution. Observed results are summarized in figure 4.5 below and on averaging the values it can be concluded that for this case rbf and polynomial kernel are evaluated as best.

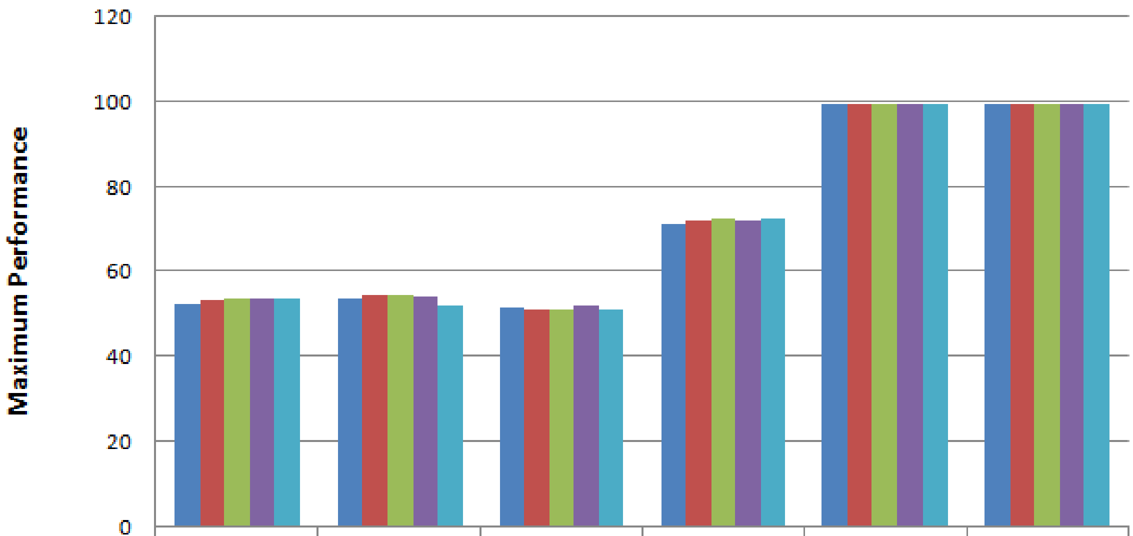


	Stage	MYCN	Ploidy	Hist	MKI	Diag
Linear	98.12	34.48	81.5	91.54	81.19	60.19
Polynomial	91.54	35.11	82.76	93.42	84.01	62.7
Rbf	91.54	35.42	81.5	92.48	83.7	62.7
Quadratic	94.98	36.36	65.2	92.16	85.27	60.82
Mlp	91.54	36.99	54.55	91.22	81.19	63.32

Figure 4.5: Classification rate of Differentiating cases for clinical data

For classifying unknown cases both MKI and diagnostic category attributes show maximum performance of 99.69%. Stage, MYCN and ploidy factors have less contribution in this case which are 53.92%, 54.55% and 52.04% respectively. Results are shown in figure 4.6 below. By this result also it is concluded that rbf is best kernel for this case.

So by this work grading performance is evaluated and best kernel is also achieved by applying support vector machine on two different data sets that are image data and clinical data. Results show that rbf is best kernel function and grading performance evaluates best attribute for a particular grade.

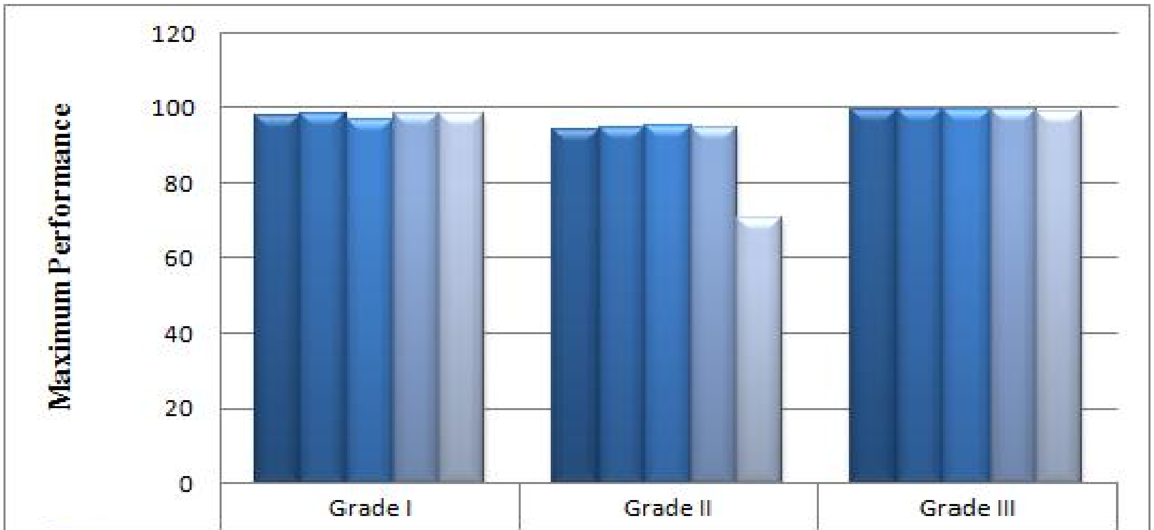


	Stage	MYCN	Ploidy	Hist	MKI	Diag
Linear	52.66	53.61	51.72	71.47	99.69	99.37
Polynomial	53.29	54.55	51.41	72.1	99.37	99.37
Rbf	53.61	54.55	51.41	72.41	99.37	99.69
Quadratic	53.92	54.23	52.04	72.1	99.37	99.37
Mlp	53.61	52.04	51.41	72.73	99.37	99.37

Figure 4.6: Classification rate of Unknown Cases for clinical data

Next SVM is applied on this data all taken together and classified and performance is evaluated for each grade. It means that test data consist of all six attributes taken together. Results are summarized in figure 4.7 below. It can be seen that grade I that is for undifferentiated and poorly differentiated cases maximum classification rate achieved is 98.75%. For grade II that is for differentiating cases maximum classification rate obtained is 95.3%. And finally for unknown cases that is for grade III maximum 99.69% classification rate is obtained.

In this data analysis on averaging the values obtained by different kernel functions rbf is the best evaluated kernel as in the above cases of clinical data. So it can be concluded that working on clinical data was more efficient for evaluating best kernel and for attribute performance for any grade.



	Grade I	Grade II	Grade III
Linear	97.81	94.36	99.37
Polynomial	98.44	94.67	99.69
Rbf	97.19	95.3	99.69
Quadratic	98.75	94.67	99.69
Mlp	98.44	70.53	99.06

Figure 4.7: Grading Performance

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

This work is done for grading of a cancer based on different features and attributes. Cancer used in this work is neuroblastoma which is nerve cell cancer. It happens mainly in new born and children less than 3 years of age. Classification is done by Support Vector Machine (SVM) classifier and grading based on International Neuroblastoma Pathology Classification (INPC). The result obtained shows the performance of each attribute taking part in classification. It grades the tumor as poorly differentiated, undifferentiated and differentiating based on the level of mitiosis-karyorrhexis index (MKI), diagnostic category, stage, histology, and ploidy and MYCN status. Each attribute is checked for different kernel functions present in SVM that are linear, quadratic, RBF, polynomial and MLP. From the results it can be concluded that MKI and diagnostic category are important features for undifferentiated or poorly differentiating cases. And stage and histology shows high performance while classifying differentiating cases.

Along with best attribute for a particular grade best kernel function is also evaluated. On working with images for all grades best kernel achieved was not same. On comparing the results with clinical data it can be seen that best kernel function is also evaluated. Rbf is best function for classification using SVM.

In this work each attribute is trained and tested individually by SVM classifier and this result is also compared with the results obtained by all attributes taken together. Future work may include multiple attributes taken together and tested for different grades. Performance may be evaluated by taking two or three or all features together. Limitation of this work is that the dataset collected may contain error values.

CHAPTER 6
REFERENCES

Abu-Mahmoud, M.K., and Al-Jumaily, A., (2014) “Novel feature extraction methodology based on histopathological images and subsequent classification by support vector machine”, *World Symposium on Computer Applications & Research (WSCAR)*, January, pp. 1-6

Bhuiyan, A.H., Azad, I., and Uddin, K., (2013) “Image Processing for skin cancer features extraction”, *International Journal of Scientific and Engineering Research (IJSER)*, Vol. 4, No. 2, February

Chapelle, O., and Keerthi, S., (2008) “Multi-class feature selection with support vector machines”, *Yahoo! Research*, February

Chinnathambi, K., Ramasamy, A., and Rajendran,P., (2014) “Modified segmentation algorithm and its feature extraction of cancer affected white blood cells ”, *International conference on Circuit, Power and Computing Technologies (ICCPCT),* March, pp. 1202-1210

Eckschlager, T., Pilat, D., Kodet, R., Dahbiova, R., Stankova, J., Jasinska, J., and Hrusak, O., (1996) “DNA ploidy in neuroblastoma”, *Neoplasma*, pp. 23-26

Gu, H., Zhao, G., and Qiu, J., (2010) “One-class support vector machine with relative comparisons”, *Tsinghua Science and Technology*, Vol. 15, No. 2, April, pp. 190-197

Gurcan, M.N., Pan, T., Shimada, H., and Saltz, J., (2006) “Image Analysis for Neuroblastoma Classification: Segmentation of Cell Nuclei” *IEEE 28th Annual International Conference of the Engineering in Medicine and Biology Society (EMBS)*, September, pp. 4844 – 4847

Hong, S., Tian-yu, N., Yan, K., and Hong, Z., (2010) “Chest DR Image Classification Based on Support Vector Machine”, *IEEE Second International Workshop on Education Technology and Computer Science (ETCS)*, Vol. 1, March, pp. 170 – 173

Hsu, C.W., and Lin, C.J., (2002) "A comparison of methods for multiclass support vector machine", *IEEE Transactions on Neural Network*, Vol. 13, No. 2, March, pp. 415-425

Kong, J., Shimada, H., Boyer, K., and Saltz, J., (2007) "Image Analysis for Automated Assessment of Grade of Neuroblastic Differentiation" *IEEE 4th International Symposium on Biomedical Imaging (ISBI)*, April, pp. 61-64

Kong, J., Sertel, O., Shimada, H., Boyer, K., Saltz, J., and Gurcan, M., (2007) "Computer-Aided Grading of Neuroblastic Differentiation: Multi-Resolution and Multi-Classifier Approach" *IEEE International Conference on Image Processing (ICIP)*, Vol. 5, October, pp. 525-528

Kumar, G., and Bhatia, P.K., (2014) "A detailed review of feature extraction in image processing system", *Fourth International Conference on Advanced Computing and Communication Technologies (ACCT)*, February, pp. 5-12

Li, G.Z., Yang, J., Liu, G.P., and Xue, L., (2004) "Feature selection for multi-class problems using support vector machines", *8th Pacific Rim International Conference on Artificial Intelligence (PRICAI),* Vol. 3157, August, pp. 292-300

Liu, Y., and Zheng, Y.F., (2005) "One-against-all multi-class SVM classification using reliability measures", *IEEE International Joint Conference on Neural Networks (IJCNN)*, Vol. 2, August, pp. 849-854

Mathur, A., and Foody, G.M., (2008) "Multiclass and Binary SVM Classification: Implications for Training and Classification Users" *IEEE Geoscience and Remote Sensing Letters*, Vol. 5, No. 2, April, pp. 241-245

Men, H., Gao, Y., Wu, Y., and Li, X., (2009) "Study on classification method based on support vector machine", *IEEE First International workshop on Education Technology and Computer Science (ETCS)*, March, Vol. 2, pp. 369-373

Samavi, S., Shirani, S., and Karimi, N., (2006) "Real-time processing and compression of DNA microarray images", *IEEE Transactions on Image Processing*, Vol. 15, No. 3, March, pp. 754-766

Saxena, P., and Singh, S.K., (2012) "Noble approach for texture classification of H&E stained histopathological image by Gaussian wavelet", *IEEE 12th International Conference on Intelligent Systems Design and Application (ISDA)*, November, pp. 375 – 379

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

Saxena, P., Singh, S.K., and Agrawal, P., (2013) "A heuristic approach for determining the shape of nuclei from H&E stained imagery*", IEEE Students Conference on Engineering and Systems (SCES)*, April, pp. 1 – 6

Sertel, O., Catalyurek, U.V., Shimada, H., and Guican, M.N., (2009) "Computer-aided Prognosis of Neuroblastoma: Detection of mitosis and karyorrhexis cells in digitized histological images" *IEEE Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*, September, pp. 1433 – 1436

Shimada, H., Umehara, S., Monobe, Y., Hachitanda, Y., Nakagawa, A., Goto, S., Gerbing, R.B., Stram, D.O., Lukens, J.N., and Matthay, K.K., (2001) "International Neuroblastoma Pathology Classification for Prognostic Evaluation of Patients with Peripheral Neuroblastic Tumors", *Cancer*, Vol. 92, No. 9, July, pp. 2451-2461

Spitz, R., Hero, B., Skowron, M., Ernestus, K., and Berthold, F., (2006) "MYCN-status in Neuroblastoma: characteristics of tumors showing amplification, gain, and non amplification", *European Journal of Cancer*, Vol. 40, No. 18, pp. 2753-2759

Stanescu, L., Burdescu, D.D., and Brezovan, M., (2011) "A comparative study of some methods for color medical image segmentation", *EURASIP Journal of Advances in Signal Processing*, December

Wang, X., Wang, S., Zhu, Y., and Meng, X., (2012) "Image segmentation based on Support Vector Machine", *IEEE 2nd International Conference on Computer Science and Network Technology (ICCSNT)*, December, pp. 202 – 206

Yang, Z.R., (2004) "Biological application of support vector machine", *Briefings in Bioinformatics*, Vol. 5, No. 4, September, pp. 328-338

Reference to WebPages:

http://www.ocg.cancer.gov

http://peir.path.uab.edu/library/

CHAPTER 7
APPENDIX

7.1 Abbreviations

- INPC: International Neuroblastoma Pathology Classification
- MKI: Mitiosis-Karyorrhexis Index
- NB: Neuroblastoma
- SVM: Support Vector Machine
- RBF: Radial Basis Function
- MLP: Multi-Layer Perceptron
- HIS: Hue-Saturation-Intensity
- HSV: Hue-Saturation-Value
- DAGSVM: Directed acyclic graph SVM
- SRM: Static Reliability Measures
- DRM: Dynamic Reliability Measures
- QP: Quadratic Programming
- FL: Follicular Lymphoma
- WBC: White Blood Cells
- PEIR: Pathology Education Informational Resource
- NCI: National Cancer Institute
- TARGET: Therapeutically Applicable Research to Generate Effective Treatments