Computerized Cancer Malignancy Grading

of Fine Needle Aspirates

Lukasz Jeleń

A Thesis

in

The Department

of

Computer Science and Software Engineering

Presented in Partial Fullfilment of the Requirements
for the Degree of Doctor of Philosophy (Computer Science) at
Concordia University
Montreal, Quebec, Canada

April, 2009

Canada

ABSTRACT

Computerized Cancer Malignancy Grading
of Fine Needle Aspirates

Lukasz Jeleń, Ph.D.

Concordia University, 2009

According to the World Health Organization, breast cancer is a leading cause of death among middle–aged women. Precise diagnosis and correct treatment significantly reduces the high number of deaths caused by breast cancer. Being successful in the treatment strictly relies on the diagnosis. Specifically, the accuracy of the diagnosis and the stage at which a cancer was diagnosed. Precise and early diagnosis has a major impact on the survival rate, which indicates how many patients will live after the treatment.

For many years researchers in medical and computer science fields have been working together to find the approach for precise diagnosis. For this thesis, precise diagnosis means finding a cancer at as early a stage as possible by developing new computer aided diagnostic tools. These tools differ depending on the type of cancer and the type of the examination that is used for diagnosis. This work concentrates on cytological images of breast cancer that are produced during fine needle aspiration biopsy examination. This kind of examination allows pathologists to estimate the malignancy of the cancer with very high accuracy. Malignancy estimation is very important when assessing a patient's survival rate and the type of treatment.

To achieve precise malignancy estimation, a classification framework is presented. This framework is able to classify breast cancer malignancy into two malignancy classes and is based on features calculated according to the Bloom–Richardson grading scheme. This scheme is commonly used by pathologists when grading breast cancer tissue. In Bloom–Richardson scheme two types of features are assessed depending on the magnification. Low

magnification images are used for examining the dispersion of the cells in the image while the high magnification images are used for precise analysis of the cells' nuclear features. In this thesis, different types of segmentation algorithms were compared to estimate the algorithm that allows for relatively fast and accurate nuclear segmentation. Based on that segmentation a set of 34 features was extracted for further malignancy classification. For classification purposes 6 different classifiers were compared. From all of the tests a set of the best preforming features were chosen.

The presented system is able to classify images of fine needle aspiration biopsy slides with high accuracy.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cancer treatment and diagnosis is an active field among researchers in various domain around the world. Detecting a cancer at an early stage greatly increases the probability of successful treatment. Breast cancer is the most often diagnosed cancer among middle–aged women. According to the World Health Organization there are 7.6 million deaths due to cancer each year out of which 502,000 are caused by breast cancer alone. According to the American Cancer Society 1.3 million women are diagnosed with breast cancer each year worldwide and about 178 thousand new cases of invasive breast cancer will be diagnosed in the United States itself [180]. In 2007 there were approximately 40,460 deaths caused by this cancer in the United States. Although the number of deaths caused by this cancer has reduced during last few years, it still remains the second most deadly cancer [8, 180]. This is illustrated by figure 1.1.

To reduce the high death rate, accurate diagnosis is essential. For many years doctors have been trying to find the best ways to treat breast cancer. Being successful in the treatment is a key to reducing the high death rate. To successfully cure a patient with breast cancer we need to diagnose it as early as possible. Cancers in their early stages are vulnerable to treatment while cancers in their most advanced stages are usually almost impossible to treat.

The most common diagnostic tools are a mammography, a histological examination and a fine needle aspiration biopsy (FNA). Mammography, which is a non-invasive method, is most often used for screening purposes rather than for precise diagnosis. It allows a

physician to find possible locations of microcalcifications and other indicators of breast cancer in the breast tissue. When a suspicious region is found, the patient is sent to a pathologist for a more precise diagnosis. This is when the FNA is taken. A fine needle aspiration biopsy is a minimally invasive method that allows the pathologist to describe the type of the cancer in detail. Using this method doctors can very adequately describe not only the type of the cancer but also its genealogy and malignancy by observing some of the preserved histological structures. They can also foresee the course of cancer development by attributing to it a predictive factor. Based on that description, a patient can be assigned for surgery to remove the suspicious region. Apart from FNA and mammography examinations, a histological examination can be performed that also allows pathologists to describe the course of the cancer and based on that description they can apply an appropriate treatment. During this examination a sample tissue of the suspicious region, found by FNA and mammography, is taken from the patient during operation and later processed in a histological laboratory to assign a prognostic factor, that describes patients' survival after undertaking treatment.

The stage of the cancer depends on the malignancy factor that is assigned during an FNA examination. The determination of the malignancy is essential when predicting the progression of cancer. There are certain features in the cytological slides that are taken into account. These features are used to assign a malignancy grade to the diagnosed tissue. One of the most popular and widely used schemes for grading cytological tissue is the Bloom-Richardson grading scheme [19].

1.1   The problem

The work presented and described in this thesis concentrates mainly on the malignancy grading of fine needle aspiration biopsy slides that are used by pathologists for diagnosis. Cancer malignancy grading is a very subjective procedure that depends on many different factors starting with experience of a pathologist through the number of similar cases seen
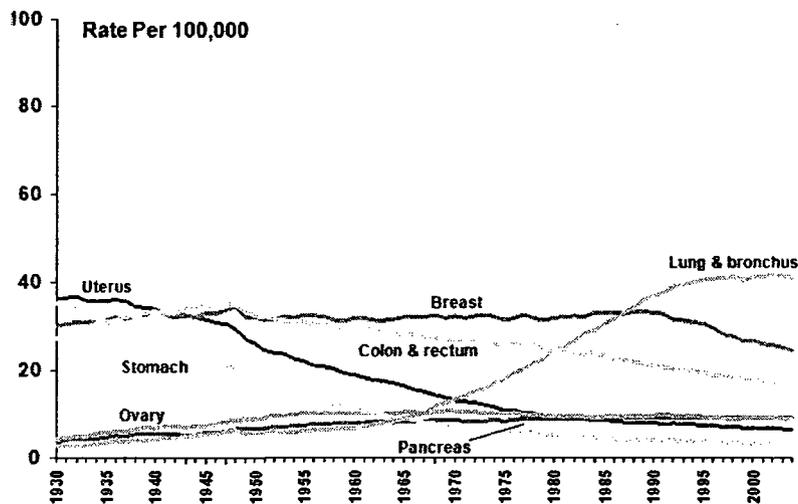
Figure 1.1: Cancer death rates among women in the US, 1930–2004. Taken from [180]

during the day and ending with his or her fatigue. There are many other factors, including the technical strength of the slide, staining and fixation method and quality, that can have a tremendous impact on the diagnosis.

The main idea of this work was to help pathologists to make a more objective diagnosis especially when similar cases are seen during diagnosis. Systematic and repetitive diagnosis is very important and allows pathologists to suggest a proper treatment, especially when a malignancy grade is hard to assess. Misclassification can lead to the proposition of an inappropriate treatment. In this work, the problem of malignancy classification is addressed and the proposed approach is briefly described in the following section (§1.2).

1.2 The approach

Breast cancer malignancy grading is based on the grading scheme described by Bloom and Richardson in [19], which is a widely used grading scheme among pathologists around the world. In general, the malignancy is assigned one of the three grades: low, intermediate, or high. That assignment is based on three factors that take into consideration different nuclear features of the cancerous cell. Each of these factors is assessed on a point based scale and the final grade depends on the summation of all the awarded points.

The Bloom–Richardson scheme can be generalized to a classification problem with three

3

classes. Each class will represent one of the three malignancy grades. The approach presented here is purely based on this generalization and malignancy grading is treated as a classification problem with some constraints that will be addressed in the subsequent chapters.

In this thesis, a fully functional grading system is developed. The system takes a FNA slide as input and returns a malignancy grade to the pathologist as output. This grading framework uses some well–known image processing algorithms that are used with more advanced segmentation techniques which allowed for nuclei segmentation from input images and for further feature extraction. In this thesis four segmentation techniques are compared. These techniques include elliptical Hough transform, level sets, fuzzy c–means and gray–level quantization. Their ability to represent the nuclear boundary is tested. Based on the segmentation results 34 features are calculated. The extracted features are then used with some the well–known classification algorithms. Here, multilayer perceptron, self–organizing maps, principal component based neural networks and support vector machines are compared. To show the strength of the proposed features, their discriminatory powers are presented. The algorithms taken into consideration are adequate for the characteristics of cytological images, i.e., nuclei extraction and features of the cancerous cell.

The research shows promising results, with classification errors lower that 5% in some cases. The presented validation of the features used for malignancy classification show that the features chosen can discriminate between malignancy classes with high accuracy. It can also be noticed that the classification error depends on the segmentation technique used for nuclei separation. Here, four well–known segmentation algorithms were used to test which technique is able to separate the nuclei with the lowest loss of classification accuracy. It was shown that the combination of multilayer perceptron and level set segmentation provided the best classification performance. From the segmentation results, it can also be noticed that level sets provide the best nuclear boundary representation out of the four compared techniques.

4

Results presented in this thesis show that such a system can not only be used during breast cancer diagnosis to help pathologists with their decisions, but also for analysis of nuclear shapes when a segmentation and feature extraction parts of the system are used. According to the expert pathologist, an introduction of such an automated grading system is essential to aid the breast cancer diagnosis process. The system proposed here is tested on the database that consists of the more difficult cases than these that are assessed every-day in the pathological laboratory. Taking that into consideration, we can assume that the good performance of the described system on the collected database suggests a good performance on the real data, that shall be less difficult to grade than this described here.

1.3   Main contributions

In this thesis a scheme for breast cancer malignancy grading is proposed. This scheme involves three main stages: segmentation, feature extraction and classification. In this work, the following contributions can be outlined:

- The introduction of three new features based on the cells' ability to form groups [86]. This features were extracted from low magnification images and provided a significant information during classification [88]. The introduction of the low magnification features allowed for the reduction of the classification error rate.

- The proposition of a set of 31 features extracted from high magnification images to choose a set of the features with the highest discrimination power and ability to separate malignancy grades.

- The reduction of the feature set based on correlation to propose a set of the most significant features for classification.

- The proposition of a set of 5 polymorphic features to represent nuclear structures in FNA sides [84] in accordance to the Bloom–Richardson grading scheme.

5

- The comparison of four segmentation algorithms to propose the most efficient segmentation method for the extraction nuclei from FNA slides [85, 87]

- The comparison of 6 classifiers to test their classification ability to distinguish between malignancy cases and the proposition of the classifier for the automated cancer malignancy grading.

All of the above contributions are described in the subsequent chapters of this thesis as mentioned in section 1.4

1.4 Structure of the thesis

This thesis consists of seven chapters preceded by a glossary of acronyms and variable names used within the text.

Chapter 2 presents an overview of breast cancer from the medical point of view with some elementary biological information about cancer. Section 2.3 presents the medical approach to breast cancer diagnosis with a focus on cytological examination, which is of a concern of this thesis. This section also describes a literature review of computer aided cancer diagnosis and malignancy grading. The remaining sections of this chapter include a review of image processing and segmentation techniques used in this work. It also describes some fundamentals of microscopy. In chapter 3 a background information about digital images and a description of the basic image processing algorithms can be found.

Chapter 4 is devoted to segmentation algorithms. It presents an overview of the algorithms used in this research and comparative results of their ability to represent nuclear boundaries. In Chapter 5 we can find description of extracted features as well as their definition and validation of their discriminatory power. Following features extraction and validation, a classification scheme is described in Chapter 6 with a presentation of the achieved error rates for each tested classifier. Last chapter of this thesis consists of conclusions and references to possible future work.

Chapter 2

Breast Cancer and Its Diagnosis

2.1 Introduction

In this chapter we will describe a process in which a healthy cell undergoes mutation and becomes abnormal and therefore cancerous. We will also show how healthy and cancerous cells differ and how it is possible to discriminate between the two types of cells. In section 2.3 medical and computer–aided approaches are described to show how computerized approaches can help during the diagnostic process. This section also includes a literature review on breast cancer diagnosis and malignancy grading. As we will see in this chapter, breast cancer diagnosis is a very active field of research while malignancy grading is less active. Assigning the malignancy grade is a crucial stage during cytological diagnosis because depending on that grade an appropriate treatment is suggested. Grading malignancy is more complicated and a more subjective procedure than pure diagnosis between benign and malignant cases. When grading malignancy, differences between malignancy grades are not as distinct as they are between benign and malignant cases. This is why grading is such a subjective procedure that depends on pathologist experience and a number of other factors. This makes the problem of malignancy grading a very interesting classification problem.

2.2 Cancer biology

Cells are fundamental components of all living organisms. They are responsible for all functions necessary for an organism to live. Multicellular organisms consist of groups of specialized cells such as tissues and organs [23]. They are responsible for specific functions of the organism. While the structure of a cell differs depending on its function, all cells have the same basic organelles such as nucleus, Golgi apparatus, mitochondria, ribosomes and rough and smooth endoplasmic reticulum (Fig. 2.1 taken from [23]).
In particular, for the purpose of our research we will concentrate on the most conspicuous
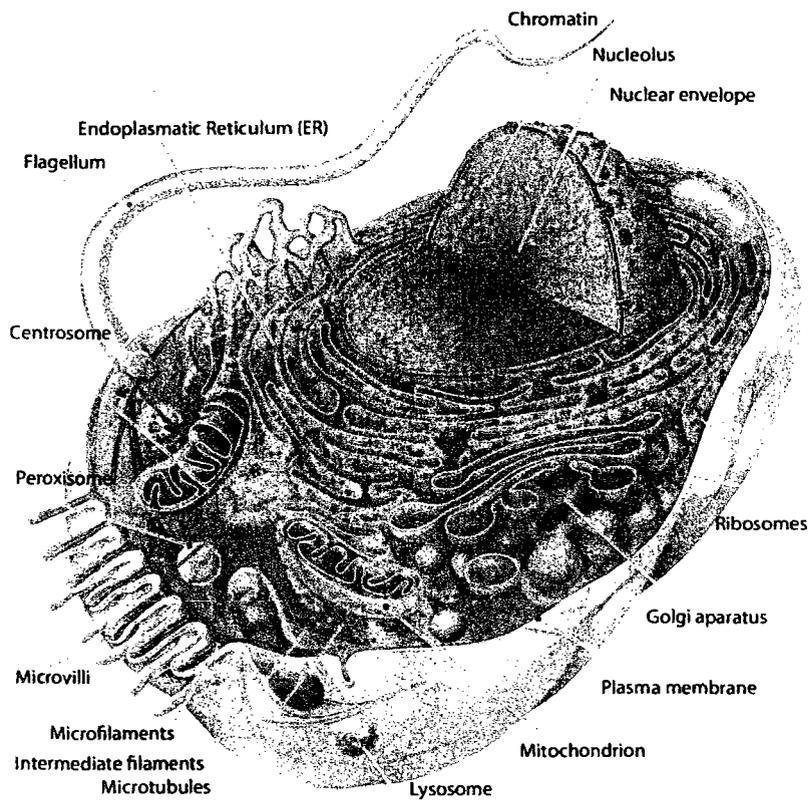
Figure 2.1: Overview of an animal cell. Taken from [23].

cell organelle, which is the nucleus. The cell nucleus is a characteristic of all organisms except for bacteria and archaea. Also, nuclear structures are the features that are taken into consideration when grading cancerous cells, as described in chapter 2.3.2.

Fig. 2.2a shows a nucleus and its internal structures. The nucleus contains the most of the cell's DNA [23] and is essential for controlling the main processes of the cell. DNA contains information about the order of amino acids in all proteins that can be synthesized by the particular cell. RNA (ribonucleic acid), which is also a component of a nucleus, is responsible for the actual protein synthesis.

We can distinguish the following nuclear structures:

- **Nuclear envelope** – consists of two separate membranes that are joined together at the pores.

- **Pores** – regulate the entry and exit of certain macromolecules and particles [23]

- **Nucleolus** – here a ribosomal ribonucleic acid (rRNA) is synthesized and stored.

- **Chromatin** – fibrous material created by fusing DNA double helix with histone molecules to form so called 'beads' as shown in Fig. 2.2b.
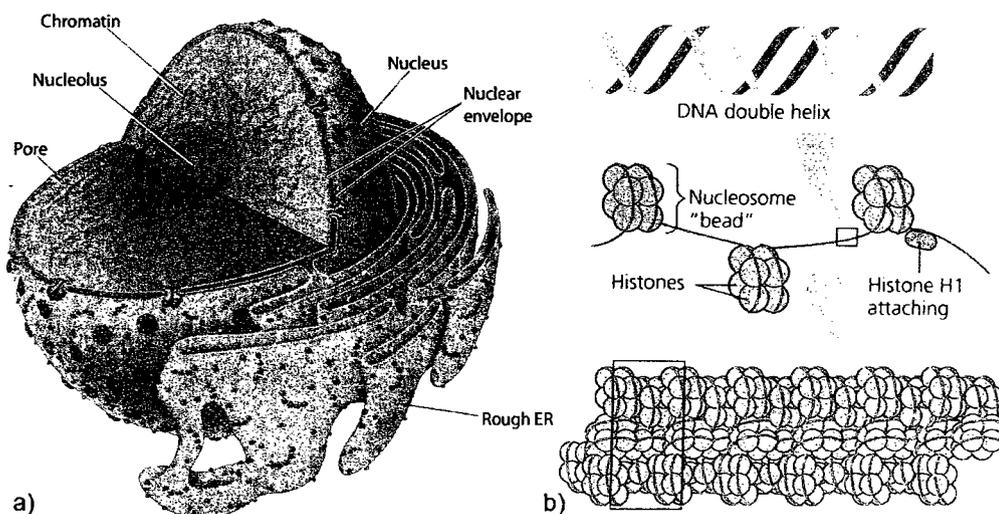
Figure 2.2: Overview of a chromatin and nucleus. Taken from [23]. a) Nucleus, b) Chromatin.

The life of all organisms is based on certain stages that all cells have to go through. These stages are called the cell's life cycle. One of the most important stages of a cell cycle is cell division, which is closely regulated by different factors. During this process new cells are generated to replace dead or damaged ones as well as to guarantee the growth of the organism [23]. There are two types of cell division, one called mitosis and another called meiosis. Meiosis occurs only during the formation of gametes and reduces the number of chromosomes by half. This number is doubled during fertilization process. Mitosis is responsible for the growth of an organism. During a mitotic division two identical cells are formed. A cell that undergoes division is called a mitotic figure.

Sometimes a healthy cell can undergo a genetic transformation that changes it to a cancerous cell. In most cases, the body's immune system destroys such cells. If, for some reason, the cell resists destruction, it may start to divide in an uncontrollable manner forming a tumor, also called a cancer. Cancer is a clump of mutant cells within normal tissue. Fig. 2.3 illustrates such a process, where a single healthy cell grows to form a malignant cancer. When a cancer is formed it can stay in this stage for many years until one of its cells undergoes another mutation that modifies its properties allowing changes in shape, orientation and behavior, Fig. 2.3 b and c. Fig. 2.3 b shows an example of hyperplasia which is a focal growth of a breast tissue in the breast duct. Fig. 2.3 c illustrates the cellular abnormality that is contained within a breast tissue. This type of a cancer is also referred to as dysplasia. When no other mutations are present, the cancer is said to be in situ, which means that it is still contained within the same tissue. When other genetic mutations occur, Fig. 2.3 d, that let cancerous cells to invade neighboring tissues, the cancer becomes malignant. The grade of a cancer's malignancy describes its ability to spread within an organism. Such a transmission of cancerous cells from a primary growth elsewhere in the organism is called metastasis. The mutations described above lead to changes in the shape of the original cell. These changes are taken into account during diagnosis and malignancy grading. Fig. 2.4 illustrates the significant variation between healthy and cancerous cell. The most important differences are:
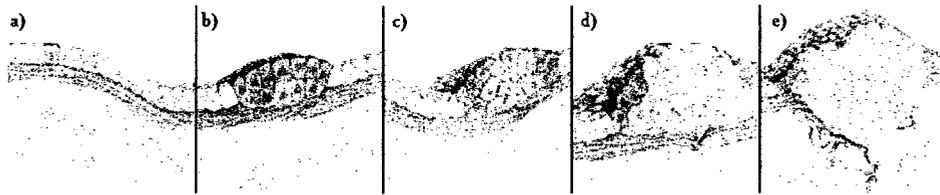
Figure 2.3: Upgrowth of a Malignant Cancer within a tissue. Taken from [23]. a) Cell modified genetically, b) Hyperplasia, c) Dysplasia, d) Cancer in situ, e) Malignant Cancer.

- Shape variations – the shape of a healthy cell is circular while a cancerous cell can assume arbitrary shapes

- Nucleus size – cancerous cells also tend to have larger nucleus in comparison to healthy ones

- Nucleolus – the nucleus of a mutated cell doesn't have to contain nucleolus but it also can have multiple nucleoli while regular cells nucleus has only one

- Nucleolus size – if there are any nucleoli they are usually of an irregular shape and dominant in the nucleolus

- Chromatin – in healthy cell it is equally distributed in the nucleus. In the genetically changed nucleus, chromatin is placed irregularly and sometimes in shape of irregular ring placed underneath a nuclear membrane

These are the most important differences that are taken into account during a cancerous tissue examination.
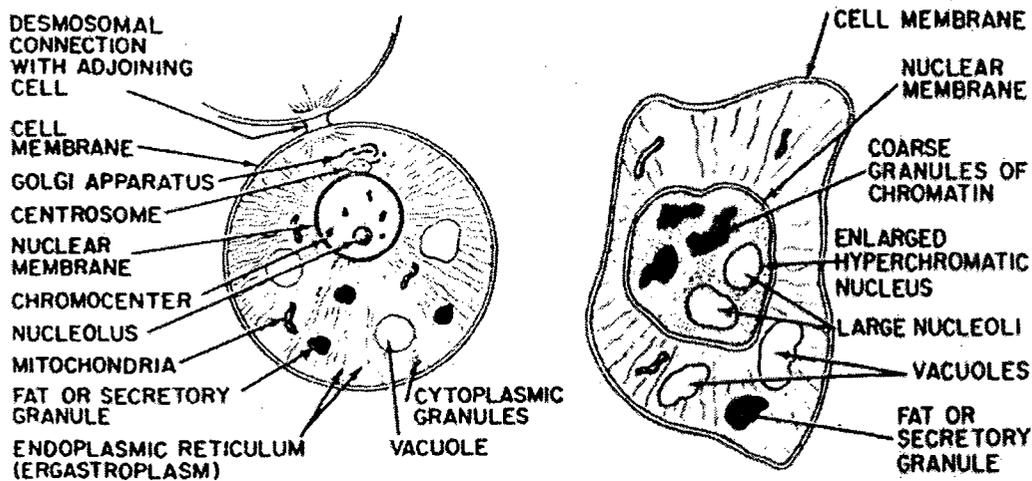
Figure 2.4: Comparison of a healthy (left) and cancerous (right) cells. Taken from [23].

2.3 Breast cancer diagnosis

Computers have been influencing our lives for many decades. Their applications are more and more advanced to solve every-day problems. Since their introduction to medicine, computers have been very helpful in assisting doctors with their diagnoses. As mentioned in chapters 1 and 2.2, cancers are very dangerous diseases and their early diagnoses can reduce the number of deaths they cause. For this purpose many computerized systems have been developed to assist doctors to make diagnoses that are more precise and more objective. In the literature we can find wide range of applications for variety of more or less deadly cancers [106, 145, 10, 74, 176]. Below, a literature review on computer–aided breast cancer diagnosis and grading is presented.

2.3.1 Medical diagnosis of breast cancer

The anatomy of a healthy woman breast contains lobules that are connected with a nipple by ducts. These structures are supported by fat tissue. Breast cancer is an abnormal growth of cells originating from ducts and lobules.

Out of all cancers, breast cancer is not only one of the most often diagnosed cancer but also one of the most deadly cancer among middle-aged women. To reduce the high number of deaths it is crucial to perform screening examinations. Regular screening allows for

12

reducing the death rate significantly. Early detection and effective treatment can reduce the death rate by up to 30%.

A screening examination consists of mammography, ultrasound and palpatic examination. The last one can be performed at home by the patient herself or by a doctor. During mammographic examination, doctors can detect lesions that are very small and in a stage that cannot be distinguished during self-examination. This examination is usually performed for women with symptoms of breast cancer. During an ultrasound examination we can detect the same lesions as in mammography without the risk of excessive radiation. The ultrasonography does not use the x-ray radiation and therefore is safer for the patient, although it can not be used for regular screening due to the fact that the microcalcifications are not as clearly visible as in the mammography and can lead to misclassification of lesions [99].

Both methods described above are said to have about 25% of false-positive diagnosis. Also, their interpretation can vary depending on the radiologist [13].

To establish a precise diagnosis, a biopsy examination is required. There are different types of biopsies but for the purpose of this study we will concentrate only on Fine Needle Aspiration Biopsy (FNA). During cancer examination a part of an abnormal tissue is collected. The collection is performed by syringe with a needle, with an outer diameter that is smaller that 1mm (typically 0.4 to 0.7mm). To locate a cancer within the breast tissue, ultrasound or mammography is used. The specimen obtained is then put on the glass slide and stained. The type of staining depends on the type of cell structures required to be visualized. Fig. 2.5 shows a stained breast tissue specimen.

When the specimen is stained, the microscopic examination is performed, during which the type of cancer is recognized as well as its malignancy grade and prognostic and predictive factors [2, 3]. Prognostic factors allow pathologists to foresee the overall survival (OS) and disease–free survival (DFS) rates while predictive factors allow them to foresee the reaction to undertaken treatment.

Besides the diagnostic examinations described above, Intra-operative and Histopatholog-ical examinations are also performed. Intra-operative examination allows quick diagnosis from a frozen sample during operation. Histopathological examination is the most accurate among all examination methods described here. It uses a paraffin preserved tissue sample. Unfortunately this method is the most time consuming due to the preservation process followed by a suitable staining.
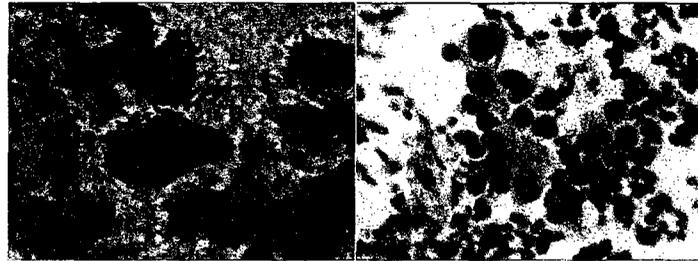
Figure 2.5: Fine Needle Aspirate of a breast recorded with different magnifications. a) 100x, b) 400x.

2.3.2   Bloom-Richardson Grading Scheme

The Bloom-Richardson (BR) grading system is the most common malignancy grading scale used by pathologists. This system was originally introduced by Bloom and Richardson in 1957 for grading histological pictures [19]. The same scale can easily be used to assess malignancy for cytological smears. According to this system there are three factors that are taken into account while grading cancerous tissue. Each of the three factors is evaluated on a three-point scale according to the following description:

1. **Degree of structural differentiation (SD)** - In histopathological slides this is also described as tubule formation. Since in cytological smears tubules are not preserved, the scoring, given below, for this factor is based on the classification of cell groupings within a smear, Fig. 2.6.

   - One point - cells in the image are grouped regularly.

   - Two points - both grouped and single spread cells found within the image.

14

○ Three points - cells spread irregularly.
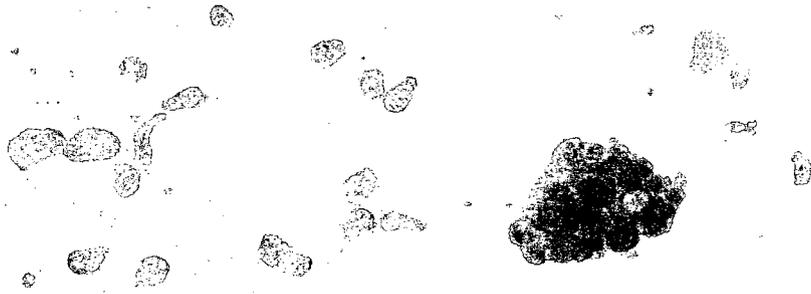
Figure 2.6: Cell groupings. a) Spread cells. b) Grouped cells.

2. **Pleomorphism (P)** - This factor takes into consideration differences in size, shape and staining of the nuclei. This scoring is fairly straightforward because with the growth of irregularity of the nuclei the prognosis becomes worse.

- One point - nuclei with uniform size, shape and staining.

- Two points - moderate variation is found.

- Three points - very significant variations.
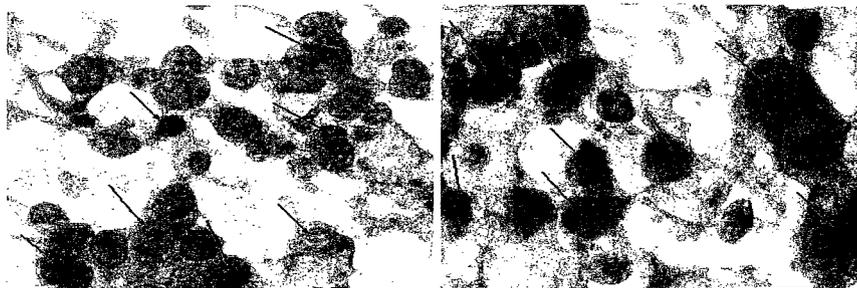
Figure 2.7: Size of nuclei,shape and staining variations. a) G2 Ductal Carcinoma, b) G3 Ductal Carcinoma

In Fig. 2.7 one can notice that G2 Ductal Carcinoma has more uniform nuclei than G3 Carcinoma. It can also be seen that a high malignancy case shows more staining variations. The darker shade in the image represents chromatin

concentrations. In healthy cells, chromatin is uniformly distributed within a nucleus while in cancerous cells chromatin is clumped and is visible as dark spots inside the nucleus. Arrows in Fig. 2.7 indicate the variations between the two grades.

3. **Frequency of hyperchromatic and mitotic figures (HMF)** - Fig. 2.8 shows an example of the mitosis. In the center part one can notice a thin border between two nuclei. A darker shade of staining can also be seen. The main objective of this factor is to assess the number of mitotic figures in the image. The more cases of mitosis one finds, the worse the prognosis is.

   - One point - occasional figures per field are found.

   - Two points - smears with two or three figures in most fields.

   - Three points - more than three figures per field are found

Figure 2.8: Example of a mitosis.

All three factors are initialized to zero. According to BR scheme, malignancy of the tumor is assigned a grade that depends on the quantitative values of the above factors and is determined by the following equation:

$$G = SD + P + HMF, \tag{2.1}$$

As one can see, the final grade is obtained by the summation of all the awarded points for each described factor. Depending on the value of G, the tumor is assigned one of three grades:

- Grade I - Low malignancy

- Grade II - Intermediate malignancy

- Grade III - High malignancy

These grades are determined according to the Fig. 2.9 (taken from [19]):

$$\overbrace{\underbrace{3\ 4\ 5}_{Grade I}\quad \underbrace{6\ 7}_{Grade II}\quad \underbrace{8\ 9}_{Grade III}}^{Points}$$
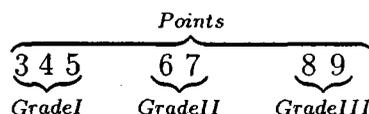
Figure 2.9: Grade distribution.

There are many different grading schemes used for cancer malignancy grading depending on the type of a cancer. For example, the Gleason grading scheme is used for grading the malignancy of a prostatic cancer [83]. It uses 5 features that are summed for the estimation of a final grade. For grading cervical cancer malignancy World Health Organization adopted the grading scheme proposed by Reagan in 1979 [43, 152]. In the case of gastric cancer the Goseki grading scheme is used as described in [53].

The basis of all malignancy grading methods is similar to that of the Bloom–Richardson scheme. They are all based on assessing the cells pleomorphism, tubules and mitosis. The variations consist of additional features that are taken into consideration. The most common variation of Bloom–Richardson scheme is the Nottingham–Bloom–Richardson scheme that additionally takes lymph node metastases into consideration [131]. Although, there are numerous variations of the grading scheme, in case of ductal breast carcinoma FNA, the Bloom–Richardson grading scheme in its original form is used.

The evaluation of the malignancy of the tumor indicates the likelihood that the case can undergo metastasis at the time or after the treatment. It also has an impact on the patient's type of treatment. Therefore it not only has a prognostic but also a predictive

value.

2.3.3 Computer–aided breast cancer diagnosis

Breast cancer diagnosis is a very wide field of research studying only medical issues but also computer science issues. As discussed in chapter 2.3.1, breast cancer diagnosis is a multi-stage process that involves different diagnostic examinations.

Pattern classification is a well–known problem in the field of Artificial Intelligence concerned with the discrimination between classes of different objects [55]. We can use the same techniques in cancer diagnosis to assist doctors with their decisions. Cheng et al. [34] provided an extensive survey on automated approaches in mammograms classification and importance of computer assisted diagnosis. Since mammography is one of the preliminary tests performed to locate abnormalities in the breast tissue, it is used for screening purposes and has raised a lot of interest within the scientific community [21, 32, 34, 35, 49, 70, 208]. This study is concerned with another important diagnostic tool, which is biopsy examination. As already mentioned, a biopsy can provide a detailed description of a disease that is crucial for determination of its treatment. This field of breast cancer examination is also an interest to many scientists. I will concentrate on some of the techniques used for classification and detection of the cancerous nuclei since it is very closely related to the research presented in this thesis.

To the best of our knowledge, the computerized breast cytology classification problem was first investigated by Wolberg et al. in 1990 [205]. The authors described an application of a multi-surface pattern separation method to cancer diagnosis. The proposed algorithm was able to distinguish between a 169 malignant and 201 benign cases with 6.5% and 4.1% error rates, respectively depending on the size of the training set. When 50% of samples were used for training, the method returned a larger error. Using 67% of sample images reduced the error to 4.1%. The same authors introduced a widely used data-base of pre-extracted features of breast cancer nuclei obtained from fine needle aspiration biopsy images [118]. Later, in 1993, Street et al. [184] used an active contour algorithm, called 'snake' for precise

18

nuclei shape representation. The authors also described 10 features of a nucleus used for classification. They achieved a 97.3% classification rate using multi-surface method for classification.

The features described by the authors are mainly geometrical features of the nucleus. These features are:

- Radius – defined as an average of the radial line segments lengths from the centroid of the nuclei to the snake points on the boundary.

- Perimeter – is the length of the boundary of a polygon connecting snake points.

- Area – is a number of pixels inside the closed snake curve.

- Compactness $= \frac{perimeter^2}{area}$

- Smoothness of a nuclei contour – defined as an average difference between the length of a radial line and the mean length of the lines surrounding it as shown in Fig. 2.10.
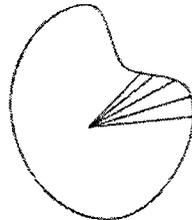
Figure 2.10: Radial lines used for smoothness computation. Taken from [184].

- Concavity – a measure of nucleus concavity. This is performed by drawing chords between non-adjacent snake points and measuring the extent to which the boundary of the nucleus lies on the inside of each chord, as seen in Fig. 2.11. The length of the chord that is outside of the nuclei is considered as a measure of the concavity. The concavity is larger when the length of the exterior chord increases.
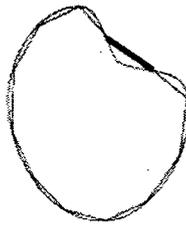
Figure 2.11: Chords used for concavity calculations. Taken from [184].

- Concave points – measures number of concavities and not their magnitude.

- Symmetry – Here, the major axis (longest chord through the center) is found. Next, length difference between lines perpendicular to major axis to nucleus boundary in both directions are measured.
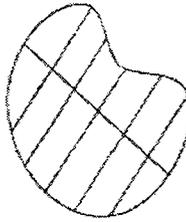
Figure 2.12: Segments used for the calculation of symmetry. Taken from [184].

- Fractal Dimension of a cell – this is approximated using a 'coastline approximation' method. Authors measure the perimeter of the nucleus using increasingly larger segments, see Fig. 2.13. Next, they plot the obtained values on a log scale and calculate the downward slope which gives an approximation to the fractal dimension. Higher values of the feature provide higher probability of malignancy.
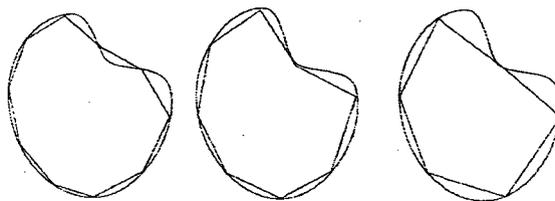
Figure 2.13: Sequence of measurements of fractal dimension. Taken from [184].

- Texture – authors define texture as an average gray scale intensity of the nucleus.

Based on the above features, Street [182], in his PhD Thesis, introduced a system called XCyt, that was later improved and described in 2000 [183]. In 1999, Lee and Street [104] described an iterative approach for automated nuclei segmentation as an addition to the previously described framework. In 2003, they introduced flexible templates to their iterative Generalized Hough Transform approach for segmentation. They created a set of predefined templates of a nuclei and each iteration shuffles the templates in such a way that those that were used the most often during the previous iteration are visited first to save time. The authors were able to segment nuclei with 78.19% accuracy [105]. They also introduced a neural network approach for classification stage, achieving 96% accuracy. Classification was based on the features previously described by Street et al. [184].

All work presented above was based on the Wisconsin Breast Cancer Database (WBCD) introduced by Mangasarian et al. [118]. This data-base consists of pre-extracted nuclear features and is widely used among researchers. Features included in the data-base are the features proposed by Street et al. [184]. WBCD [118] and its variations [203, 204] are the only data sets publicly available. Therefore, the majority of work in this field is performed on this data-base and involves research on different classification algorithms.

In 1998, Walker et al. [197, 198] introduced Evolved Neural Networks for breast cancer classification and tested their algorithm on WBCD data-base achieving 96% correctness. Nezafat et al. [135] used WBCD to compare several classification algorithms such as k-nearest neighbor classifier, radial-basis function, neural networks, multilayer perceptron and probabilistic neural networks. The authors showed that among these classifiers, multilayer perceptron with one hidden layer performed the most efficiently giving 2.1% error rate. Additionally they also compared and reported which of the features extracted by Wolberg et al. [205] were most significant for classification.

In 2002, Estevez et al. [57] introduced a different approach for classification based on the Fuzzy Finite State Machine, but their system performed rather poorly giving 19.4% error for the testing set of images. To extract features, the authors first manually segment nuclei

from the image and then apply a low-pass filter and in the following step topological map of a nuclei is created. The extracted features are texture based. Motivation for them was that benign cell textures have bigger homogenous gray areas and more concentric contours than malignant cell textures.

Bagui et al. [7] recently introduced a classification algorithm applied to WBCD. The authors described a generalization of the rank nearest neighbor rule and obtained results that show a 97% recognition rate, which, according the authors, is better than that previously reported in the literature. From the above discussion we can deduct that majority of work in the field of breast cancer detection and classification was performed by Street et al. and Wolberg et al. We can find other approaches such as wavelet based approach of Weyn et al. [200]. Here the authors introduce a textural approach for chromatin description and claim that it has a 100% recognition rate.

Another approach is one introduced by Schnorrenberg et al. [168] that uses receptive fields for nuclei localization as an integral part of a bigger system, called 'BASS. ' In 1996, they introduced a content–based approach [169] and provided an extensive survey on existing histopathological systems [170]. The authors presented two types of color-based features, luminance-based local features and global features. Luminance features were obtained from image RGB values. Global features are the variance and average of luminance in the image. They also introduce one texture measure that is calculated according to the luminance variance and current nucleus luminance. Approaches presented by Schnorrenberg et al. are mostly based on histological samples rather than cytological. In 2000, they presented a description of features used in their research [171] on classification of cryostat samples during intra-operative examination based on feed-forward neural networks achieving the highest accuracy of 76% on their own database.

In the literature we can also find some other approaches that involve segmentation of a breast cancer nuclei rather than classification. In 1996, Belhomme et al. [14] proposed a watershed based algorithm for segmentation of breast cancer cytological and histological images.Their algorithm is a more general version of the method described by Adams and

Bischof [1]. The generalization involves the usage of numerous merging criteria. Authors use the segmentation principles described by Beucher in his PhD thesis [15]. This involves the decomposition of the segmentation procedure into two steps. In the first step, the image is simplified based on a set of markers. The second stage involves region decomposition by the construction of the watershed lines [14]. The algorithm proposed by Belhomme *et al.* is the extension of the Beucher and Meyer [16] method by introduction of a general segmentation operator.

In 1998, Olivier *et al.* [109] introduced another extension to the watershed algorithm in addition to that of Belhomme *et al.* Their extension incorporates the color information in the image regardless of the color space. The authors compared their segmentation results against the segmentation performed by three experts and they reported the correctness of their method to be between 89.2% and 98.3% for the nuclei.

Another approach to nuclear segmentation is based on fuzzy c–means clustering and multiple active contours models described by Schüpp *et al.* [172]. The authors describe a level set active contours method, where the initial level set is obtained by the fuzzy c–means algorithm.

2.3.4 Computer–aided breast cancer malignancy grading

In the previous section we described different approaches for breast cancer diagnosis. Most of those systems discriminate only benign and malignant cases. For good diagnosis it is crucial to evaluate the malignancy grade as already described in section 2.3.1. In [119] we can see attempts at prognostication along with nuclear classification. For their grading approach, the authors used only nuclear features of a cell, which correspond to the second factor in Bloom-Richardson grading scheme (see section 2.3.2). They were estimating the prognosis of the breast cancer according to these features. Further attempts for malignancy grading include VLSI approach introduced by Cheng et al. [33] in 1991 and applied in 1998 to breast cancer diagnosis [36]. In this method, the authors propose a parallel approach to tubule grading for histological slides. The authors divided their algorithm into four stages.

The first stage consists of image enhancement for which purpose they use median filtering to remove artifacts. In stage two, the authors locate possible tubule formations by image thresholding with a threshold level known *a priori*. The next stage is a classification stage, where regions are classified as tubular formations. The features used in this study consists of brightness, bright homogeneity, circularity, size, and boundary colors. In the fourth stage, the authors count the number of tubular formations. The work presented by the authors not only deals with histology but also only mentions grading using only one factor on the Bloom-Richardson scale. The authors showed time improvement of the parallel algorithm that grades tubules to $O(n)$ time while previously reported run time complexities were $O(n^2)$, where n is the size of the input data. In 1991, MacAulay *et al.* [115] introduced a graphics package for Bloom-Richardson grading of histological tissue. Their application acts as a typical graphics program that allows user to pick the nuclei from the image and perform some basic calculations. This process is almost completely user dependent. The authors provide an extensive description of the interface of the package but no further information on computation grading was found. Another approach found in literature is an algorithm based on wavelet texture description of chromatin [200]. This work was also performed on histological slides. The features calculated by the authors are calculated according to wavelet parameters and are divided into three groups. The first group are co-occurrence parameters that describe the color intensity in the image. The second set of parameters are densitometric parameters that are based on intensity values of the nucleus. The third group consists morphometric parameters that describe the geometry of the nucleus. Authors performed tests on their data-base of 83 histological slides and claim to have 100% classification rate. Such a high rate suggests a good separation between the classes.

In 2004, Gurevich and Murashov [73] proposed a method for chromatin structure analysis based on scale–space approach of Florack and Kuijper [58]. The authors claim that chromatin distribution corresponds to the grade of malignancy. This statement is supported by additional studies of Rodenacker [156, 157, 158] and Weyn *et al.* [199]. The authors

also mention another approach to chromatin description. This method uses heterogeneity, clumpiness, margination and radius of particles and was introduced by Young *et al.* [207]. The algorithm of Guverich and Murashov uses topological properties of iso–intensity manifolds in the spatial extrema neighborhoods [73]. Their algorithm is able to measure the number of chromatin particles in the input image. For testing purposes the authors trained several classifiers achieving a classification rate between 72% and 85.4%. In 2006, Gurevich *et al.* [72] described a system for automatic analysis of cytological slides for the lymphatic system tumors. The authors used a Gaussian filter for segmentation of a nuclei from the previously extracted blue channel of the image. The feature extraction part of the proposed system is the same as in [73] plus an additional 47 features described by Churakova *et al.* [39]. These features include a well known and widely used morphological features such as the area of a nuclei, histogram features and features based on a Fourier spectrum of a nucleus [72]. In this paper, the same choice of classifiers was used as in [73] but the accuracy increased and is claimed by the authors to be above 90%. The authors did not provide an accurate error rate of their experiments and therefore it is difficult to assess the accuracy of the proposed system.

To the best of our knowledge, currently there is no publicly available database and most of the approaches presented in the literature are tested on the databases created by the authors, which makes the comparison of the obtained classification results difficult. The only commonly used database that we came across during this study is the Wisconsin Breast Cancer Database, which was described earlier in this thesis. This database is freely available from the authors web page [205]. In this study, some of the proposed features are the same as in WBCD but the testing of the presented system on that database would be limited only to the classification stage due to the fact that WBCD is the database of pre-extracted features.

The most recent development in this area is commercial system released in 2005 by QinetiQ for automated histopathological tissue grading [147]. According to the specifications and discussion with a pathologist, the results obtained by this system seem to be difficult

to confirm. According to the authors, their system showed performance similar to the pathologists during clinical evaluation that was performed on 100 patients.

2.4   Contributions

This thesis is an attempt of making the breast cancer diagnosis more repeatable and objective task to assist a pathologist during the decision making process. Malignancy grading is a very difficult task and as can be noticed from the literature review there weren't many computerized approaches to breast cancer malignancy grading. A part of this thesis was also devoted to the collection of the database of fine needle aspiration biopsy images (see Appendix A). Unlike Cheng et al. [33] our approach uses cytological data rather than histopathological. In [73, 157, 200, 207] authors extract features that are based on chromatin description. In this thesis, we additionally extracted features based on nuclear polymorphy. We have also proposed a new set of 3 features that described the cell' ability to form groups. These features are calculated from the low magnification images. The use of this kind of images makes the presented system a bimodal system. Unlike any other presented and previously described system in this thesis we introduce a multimodal malignancy grading system. Also, in this study we have extracted several features that allowed for the discrimination between two malignancy classes. Most of the features are similar to these previously described in the literature [182, 183]. Here, we proposed two techniques that can be used for the evaluation of these features to choose the most representative set of features that can be used for classification purposes. In this thesis, unlikely to other approaches for malignancy grading, we have compared 6 classifiers to propose one that is able to classify breast cancer malignancy most precisely.

2.5 Chapter Summary

This chapter was mainly devoted to defining breast cancer, starting from basic medical view on healthy cell trough the mutation process and cancerous cell formation to diagnosis procedures and problems related to that diagnosis.

From the discussion in previous sections, one can notice that cancer is a very serious disease with a high death rate. In this thesis we focus exclusively on breast cancer which is the one of the most deadly cancer affecting middle–aged women. The key to reduce the high death rate is to find a very efficient diagnostic method that will allow for a diagnosis of a cancer in as early a stage as possible. This thesis is an attempt to solve this problem by making malignancy grading more objective and, what is probably more important, repeatable. Nowadays, pathologists struggle with the problem where the malignancy grade can depend on the pathologist that performs the grading. The computerized approach allows for the repeatable decision making diagnosis with the application of the classifier that makes the same decisions every time for the same features. As we will see in subsequent chapters, system presented in this thesis allows repeatable grading of breast cancer malignancy according to Bloom and Richardson Scheme (see section 2.3.2). The presented results support the proposition that the system presented here can be very useful in the pathological laboratory and can help pathologist with his/her diagnosis.

Chapter 3

Microscopy, Image Representation and Image Analysis

3.1 Introduction

Analysis of the content of the image is a very important task in any image processing problem. There are many different algorithms that can be used for this kind of analysis. In this chapter we will present some fundamentals of light microscopy along with a description of some other imaging techniques. We will also review the basic image processing and analysis techniques such as image histograms, morphological operations, edge detection and connected components. The reader can also find a description of shape descriptors that can be used to represent and analyze the shapes in the input image.

3.2 Fundamentals of Microscopy

Cells are the fundamental structures of each organism. Being able to visualize and observe the cell is very crucial in most biological studies [153]. Microscopy allows for an observation of different biological strictures on different levels. Numerous types of microscopy and imaging techniques allow for visualization of these levels, from tissues to cells or even nuclei [122, 143]. Here, four types of microscopic techniques are presented to show the reader the differences between them.

3.2.1 Light Microscopy

Light microscopy is the most popular and the easiest method for observation of small objects. It is the most used research tool in biology and medicine. It allows for noninvasive observation of cell shapes as well as their movement and some of their functions.

Nowadays, there are many different types of microscopes from bright and dark field microscopes [27], through phase and contrast microscopes up to fluorescence and confocal microscopes, which are specialized for advanced research, clinical and industrial applications.

The simplest and most well–used microscope is a bright field microscope. Elements of this kind of microscope are a base of construction for all other types of light–based microscopes. The main mechanical parts (Fig. 3.1a) of the microscope are:

- frame – lies on the base of the microscope and constitutes a support for other mechanical and optical elements,

- stage – is a small table on which the specimen is placed. It allows for movement of the specimen for better examination,

- nosepiece objective turret – allows for changing objectives that are mounted on the bottom. It also allows for magnification changes of the observed specimen,

- coarse and fine focus knobs – allow for coarse and fine focusing of the viewed image.

Optical elements (Fig. 3.1b) are engaged to change the light beam to light a specimen on the stage and then pass through it to make a magnified image in the eyepiece [133]. Main and necessary optical elements are:

- light source – its main task is to supply light to the microscope. The simplest light source can be a mirror that is adjusted to reflect a sunlight to the objective in the turret. More advanced light sources are built with a light bulb or even with LED diode,

- condenser – condenses a light from a light source on the specimen,

- shutter – usually placed under the condenser to decrease or increase amount of light that falls on the specimen,

- objective – mounted on the bottom of the turret, produces magnified, real and obverse image of the specimen,

- eyepiece – mounted on top of the microscope and allows for observation of the image produced by objective. Objectives in the eyepiece also produce magnified and obverse image of the specimen but it is a virtual image.

We can distinguish different kinds of objectives depending on the magnification. Small magnification objectives are called dry objectives and large magnification objectives are called wet because they require a use of immersion oil between the specimen and the objective. The oil has the same refraction coefficient as glass and therefore improves the quality of observed image. Between the objective and eyepiece one can also find additional optical elements such as different filters and mirrors.
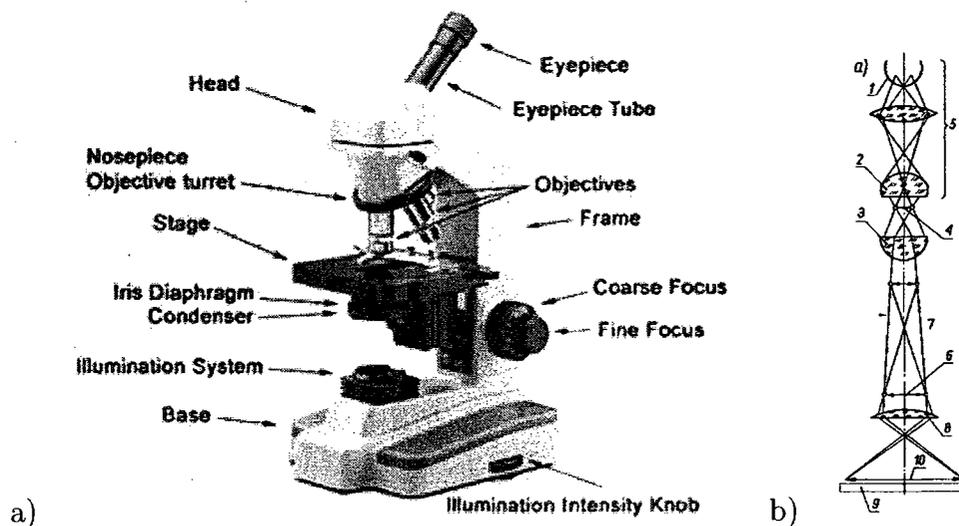
Figure 3.1: Elements of a light microscope a) Mechanical parts, b) Optical parts.

3.2.2 Confocal Microscopy

Confocal microscopy is usually used for visualization of 3D cellular structures and their dynamic processes. The scanning confocal microscope was invented in 1955 by Minsky [126] and nowadays is widely used around the world.

In comparison with a light microscope, the confocal microscope consists of the confocal pinholes, that are responsible for the passage of only the light from the focus plane [153]. Only the light that passes through the pinhole can reach the detector. This setup is visualized in Fig. 3.2 a.

Images taken by the confocal microscope are high resolution images with a high clarity. They allow for a construction of the 3D representation of an observed sample or a cell in our case. This representation is constructed form a collection of slices recorded on a different levels of focus through the thickness of the object (see fig. 3.2 b) [153].
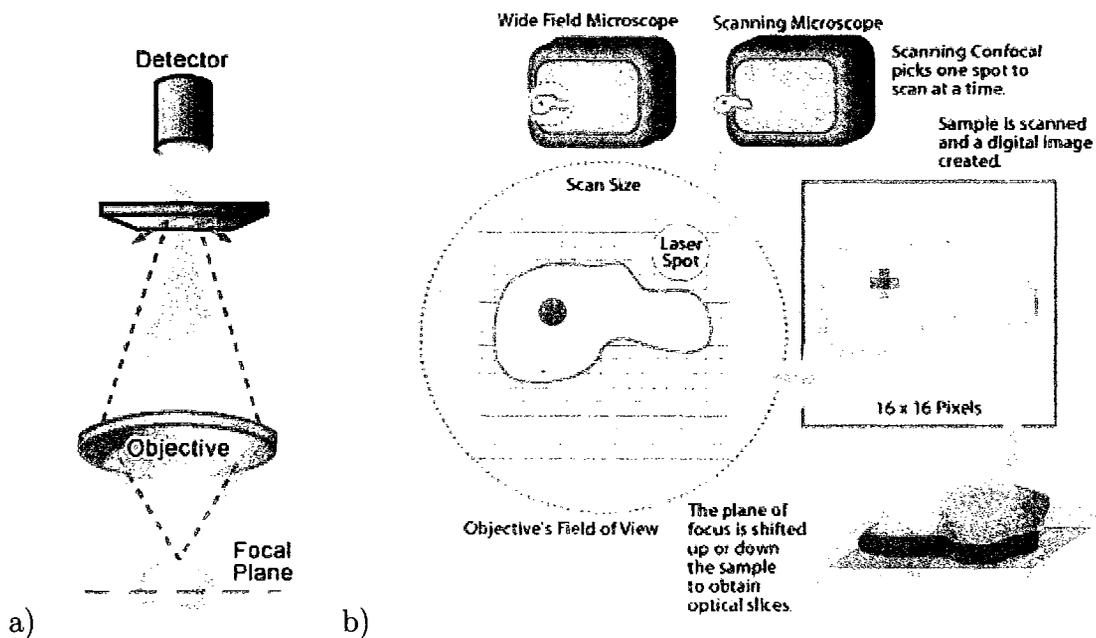
Figure 3.2: Confocal microscope. Taken from [153]. a) An example of a confocal pinhole, b) Scanning of a sample.

Although the samples used for confocal microscopy are usually thicker than in the conventional light microscopy, they both need special preparation to be able to visualize the contrasts between the areas of interest. The staining used with confocal microscopy is different than that for light microscopy, although it can be easily adapted. The limiting factor that needs to be taken into consideration is the thickness of the tissue because the diffusion of the stain is different for thicker and thinner samples. This is considered as a major limiting factor. Staining used with confocal microscopy requires stating that reflects, absorbs the light or is fluorescent [153]. Fig. 3.3 shows an example of the confocal microscopy images.
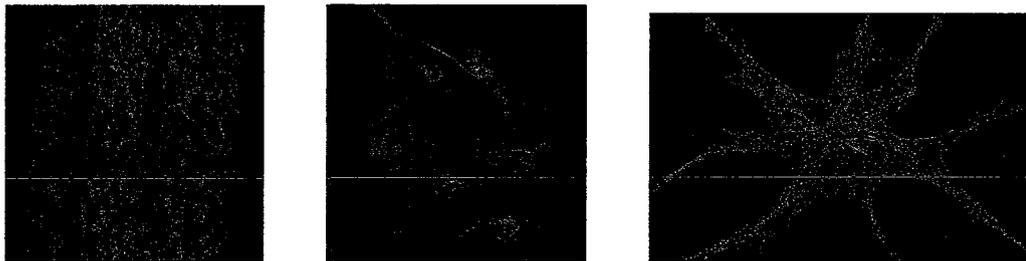
Figure 3.3: Confocal microscopy images. Taken from [153, 190, 108]

3.2.3 Fluorescence Microscopy

Fluorescence microscopy in very similar to the conventional light microscopy. They both use a light source that illuminates the samples to produce the magnified image of the specimen. The difference between these methods is that fluorescence microscopes use higher intensity light that causes fluorescent excitation in the sample. This phenomenon is used to produce the magnified image based on the emitted light by the sample. The emitted

light has a longer wavelength than the original light source that was used to illuminate the specimen (see fig. 3.4)[139].
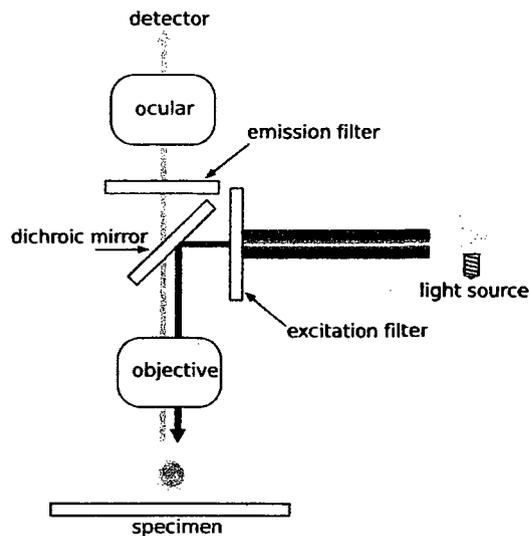
Figure 3.4: Schematic of a fluorescence microscope. Taken from [201]

In this kind of microscopy, a special type of staining procedure is used. In this procedure, cells are stained with a fluorescent dye that allows for better light emission and therefore easier measurement of the fluorescence. This allows for visualization of single cells as well as certain parts of the cell [4].

Because the objective of the fluorescence microscope uses the same optics both for excitation and emission, a dichroic mirror is used for the separation of these paths. Here, the excitation light is reflected from the mirror towards the objective while the emission is passed through to the detector for visualization of the specimen [139].

Fluorescence microscopy is widely used for visualization of biological structures [12, 46, 47, 48, 63, 107, 128]. Most of these approaches involve the visualization of nucleic acids, such as DNA [12, 63, 128] or RNA [48]. There are also some other applications of this technique, such as cell population visualization [107]. All of these approaches prove that this is a powerful tool that allows for very reliable representation of cellular structures. Examples of the fluorescence images are presented on the fig. 3.5.
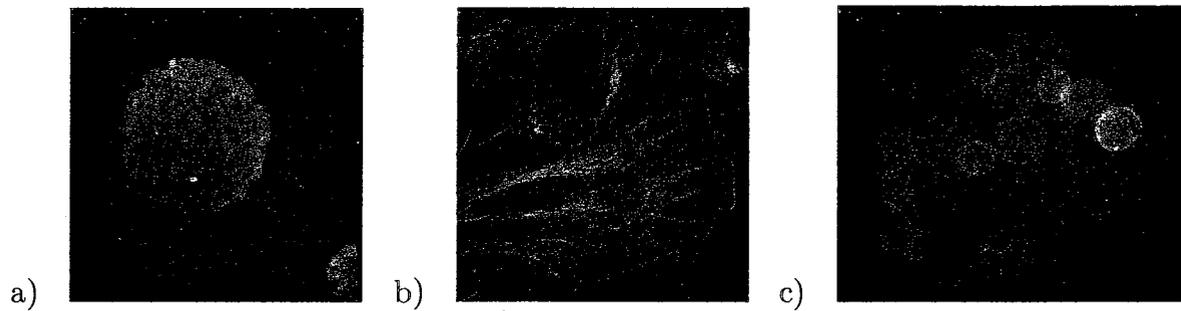
Figure 3.5: Examples of the fluorescence images. Taken from [201] a) Human lymphocyte nucleus, b) endothelial cells, c)Yeast cell membrane.

3.2.4 Phase Contrast Microscopy

The phase contrast microscope was invented in 1935 by Fritz Zernike. It uses a phenomena of phase shifting on the light that goes through the sample. These shifts are then converted into the contrast changes in the image [65, 202].

In general, this type of microscopy generates the coherent monochromatic light rays that passes through the condenser assuming a ringlike shape. Next, it is focused on the object plane in which some of the light is diffracted and some goes through the object. Both of the beams are then separated and directed toward the phase plate. This allows for changing the phase of one of the rays and then in the image plane the beams interfere to create an image [178]. Fig. 3.7 illustrates the principals of the phase contrast microscopy.

This type of microscopy is widely used for cytological examination in gynecology [209, 124, 125] for visualization of the cellular structures [163]. Apart form these, the phase contrast microscopy is found to be useful in urology and nephrology for examination of epithelial and intracellular structures [62, 130, 50, 159, 136, 149].
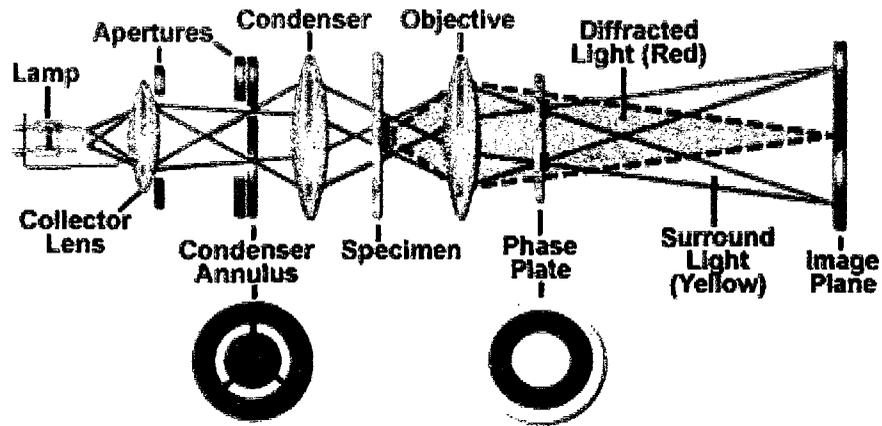
Figure 3.6: Schematic of a phase contrast microscope. Taken from [129]
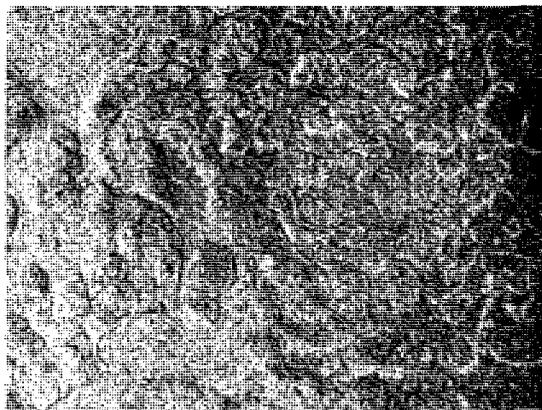
Figure 3.7: Image from the phase contrast microscope. Taken from [123]

3.2.5　Architecture of the Proposed System

In this study we make use of the bright field light microscope with an additional mirror mounted behind the objective. The role of the mirror is to split the image into to visible images. First image is visible through the eyepiece and the second image is projected to the camera and later recorded by the CCD camera and sent to the computer with MultiScan acquisition software. An example of a similar system is shown on the Fig. 3.8.
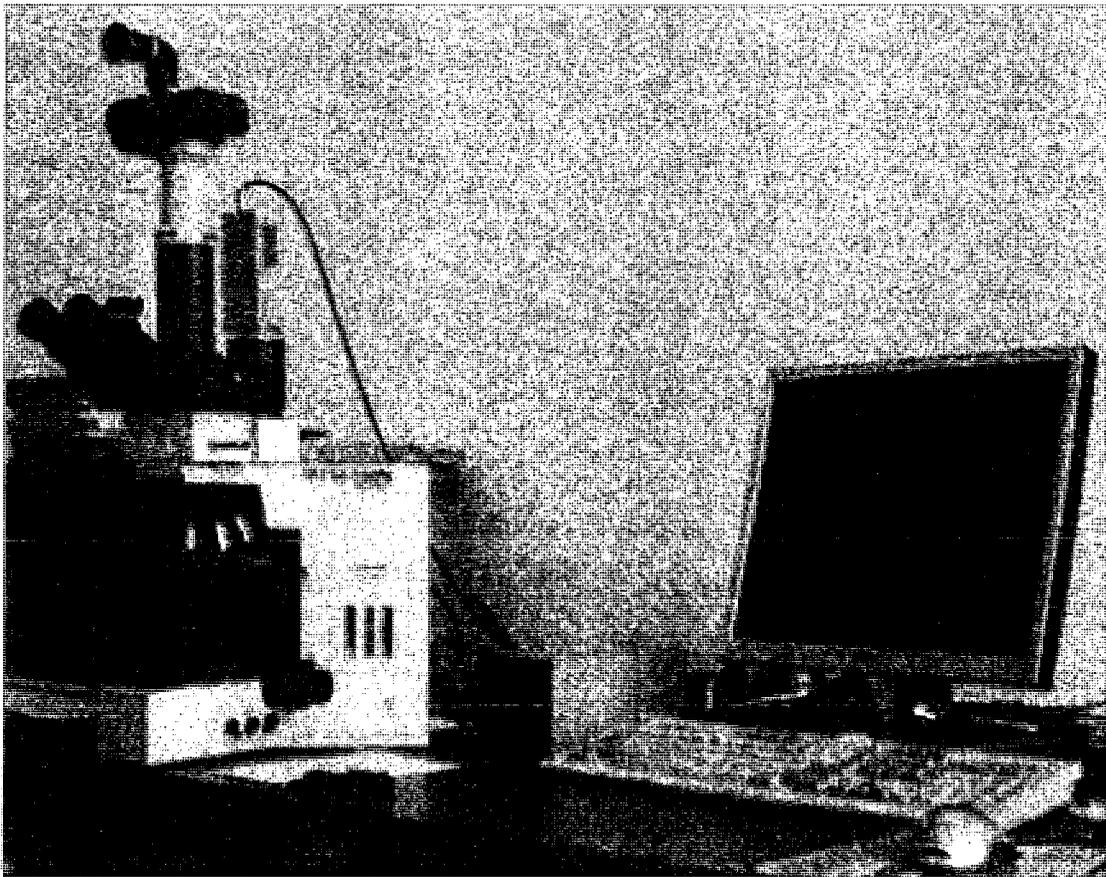
Figure 3.8: Digital microscopy acquisition system.

Fig. 3.9 shows a diagram of the system proposed in this thesis. The following stages of the system can be distinguished:

- Image acquisition – Here, the slide placed under the microscope is captured by a CCD camera and saved on the computer hard drive for further processing. At this stage, images with low (100x) and high (400x) magnification of the same tissue are taken.

- Preprocessing – In this stage, the image recorded during the acquisition step is converted from the RGB color base to gray level. This is performed by extraction of the image red channel.

- Segmentation and features extraction – This is the most important stage of the system, since good segmentation and feature extraction have major impact on the classification results. Here, the images are divided depending on their magnification. Images with higher magnification require more sophisticated methods for segmentation than those recorded in lower magnification. When nuclei and nuclear groups are extracted, the features are extracted and passed to the next stage of the system.

- Classification – During the classification stage, the features are classified based on the description of the Bloom–Richardson scheme. The classifier shall be able to distinguish between low (G1), intermediate (G2) and high (G3) malignancy grades and outputs the assigned grade. Because the low malignancy cases are very rare, the training and testing sets were reduced only to two classes.
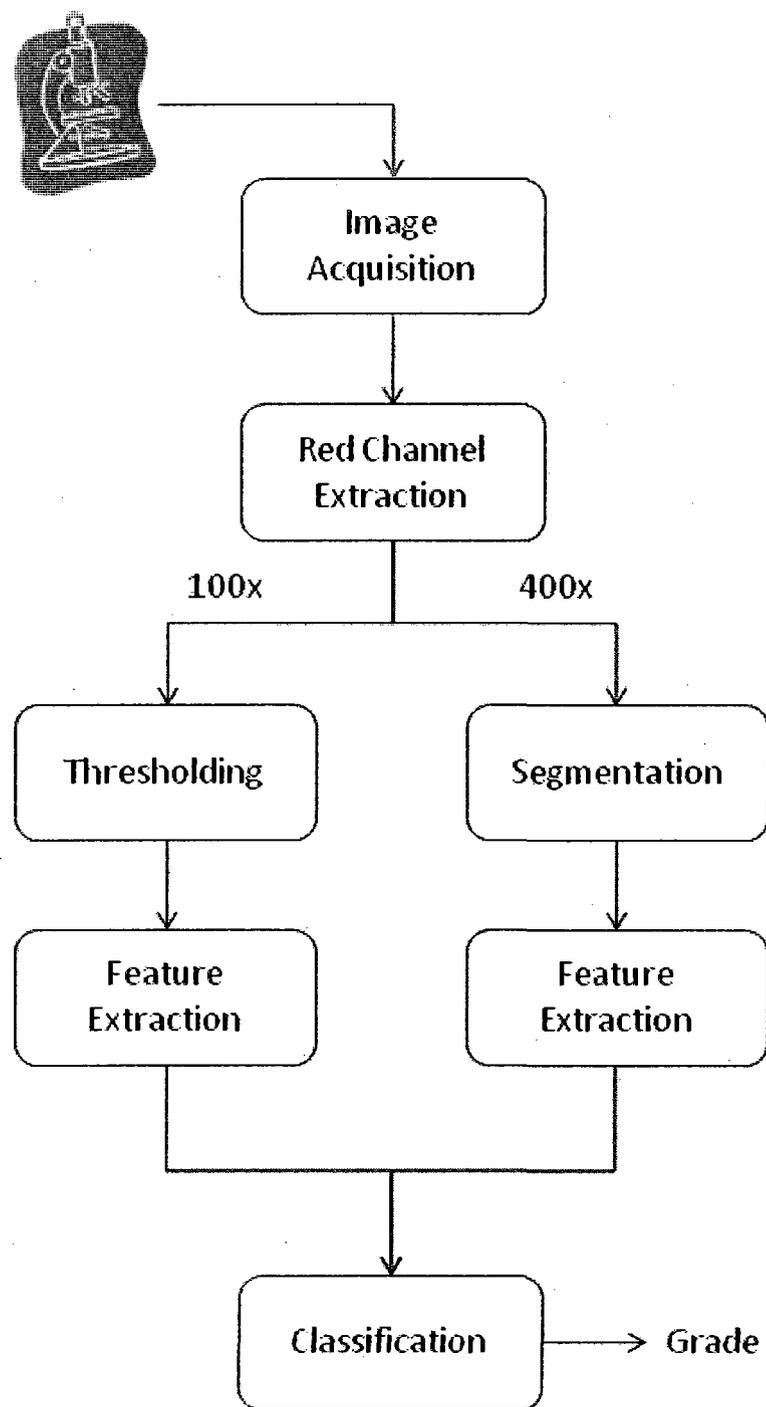
Figure 3.9: Breast cancer grading – system diagram.

3.3 Image Representation

To be able to further investigate images obtained from microscopic examination, we first should define what the image representation is. This information allows us to define some standards about digital images. Here we take into consideration gray level, color and binary images.

Definition 1. *An image I of a size M by N pixels is a function that assigns each pixel (x,y) a value between 0 and 255 (Eq. 3.1).*

$$I : [1, M] \times [1, N] \to [0, 255]$$
$$(x, y) \mapsto I(x, y).$$

$$(3.1)$$

Function $I(x, y)$ is a gray level image where 0 represents black color and 255 denotes white [178]. Color images consist of three components which represent one of the three primary colors: red (I_r), green (I_g) and blue (I_b). This is the most often used color image representation, called RGB representation. Here the color of the pixel is determined as an combination of the three components: I_r, I_g and I_b.

Binary image representation is a little different from those previously presented. This kind of image can only assume two pixel values, 0 for black (false) and 1 for white (true) [178, 195]. In this thesis we assume that true pixels represent an object and false (black) pixels represent a background. If we denote the number of white pixels by n and $\mathcal{I}$ as:

$$\mathcal{I} = \{\mathbf{p_i} : i \in \{1, ..., n\}\}, \quad \mathbf{p_i} = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \tag{3.2}$$

then we can define n as:

$$n = |\mathcal{I}|. \tag{3.3}$$

When analyzing an image it is often necessary to process the information not only from a particular pixel, but also the values of the neighborhood of that pixel. There are two most common representations of the pixels neighborhood. Their definition depends on the number of neighboring pixels taken into consideration, and these are the four–connected

	N	
W	(x,y)	E
	S	

NW	N	NE
W	(x,y)	E
SW	S	SE

a) $N_4(x,y)$ b) $N_8(x,y)$
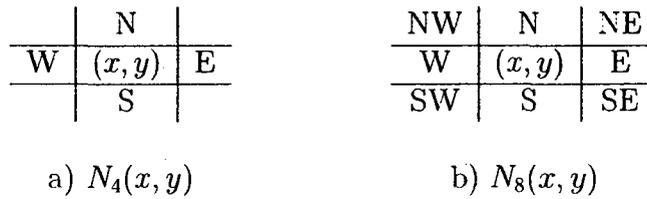
Figure 3.10: Neighborhoods of the pixel

neighborhood $N_4(x,y)$ and the eight–connected neighborhood $N_8(x,y)$ [178].

Fig. 3.1 illustrates both of the neighborhoods. Fig. 3.1a represents the 4–connected neighborhood, where neighboring pixels are refereed to as *north, south, west* and *east*. The 8–connected neighborhood in addition to N_4 pixels also includes diagonal pixels that are referred to as *northwest, northeast, southwest* and *southeast*. All of these directions describe the following pixels of the neighborhood of the pixel (x,y):

$$WN = (x-1,y-1), \quad N = (x,y-1), \quad SN = (x+1,y-1),$$

$$W = (x-1,y), \qquad\qquad E = (x+1,y),$$

$$SW = (x-1,y+1), \quad S = (x,y+1), \quad SE = (x+1,y+1).$$

3.4   Image Analysis Techniques

3.4.1   Image Histograms

The histogram of an image describes the occurrence frequency of intensity values in that image [196]. In the histogram the lowest value is 0 and its maximal value is $2^n - 1$, where color is represented by n bits. In case of intensity images the maximal value of the histogram is 255 (8 bits). The height of a peak is equal to the number of corresponding intensity values on the x-axis.

Histogram analysis is very helpful in image processing. It provides information about intensity ranges in the image. In reality, a histogram, in statistical terms, is a probability density function $h(x)$. The area of the histogram, called a cumulative histogram, can be calculated from the equation 3.4.

$$\int_0^X h(x)dx \tag{3.4}$$

where $h(x)$ is a histogram function, $h(x)dx$ is the area of a histogram column [195].

For digital images, we calculate the histogram as the normalized sum of all pixels with a certain intensity value:

$$h(x) = \frac{\sum(x)}{N} \tag{3.5}$$

where $\sum(x)$ is a number of all pixels with intensity value equal to x and N is a number of pixels in the image.

Based on a calculated histogram we can apply basic enhancements based on its intensity. If the image is too dark and its intensity distribution is very narrow, then we can stretch the histogram to obtain a distribution covering all the values in the intensity range of the image. This process is called histogram stretching. Another enhancement is based on contrast and can be achieved by shifting the histogram by a certain value to the left or right [178].

Histograms also allow us to calculate additional parameters that describe different properties of the image and histogram itself. If a calculated histogram is $P(g)$, where g is a pixel value, then we can defined the following parameters:

1. Mean value

$$\overline{g} = \sum_{g=0}^{L-1} gP(g) = \sum_{row} \sum_{column} \frac{I(x,y)}{M} \tag{3.6}$$

where L is a number of possible intensity levels and M is the total number of pixels.

2. Standard deviation

$$\sigma = \sqrt{\sum_{g=0}^{L-1}(g-\overline{g})^2 P(g)} \tag{3.7}$$

3. Histogram asymmetry

$$SKEW = \frac{1}{\sigma^3} \sum_{g=0}^{L-1}(g-\overline{g})^3 P(g) \tag{3.8}$$

or:

$$SKEW = \frac{\bar{g} - max(g)}{\sigma} \qquad (3.9)$$

4. Histogram width

$$HW = \frac{1}{4} \sum_{g=0}^{L-1} (g - \bar{g})^4 P(g) - 3 \qquad (3.10)$$

5. Energy

$$E = \sum_{g=0}^{L-1} [P(g)]^2 \qquad (3.11)$$

6. Entropy

$$H = -\sum_{g=0}^{L-1} P(g) \log_2 P(g) \qquad (3.12)$$

The parameters defined above are characteristic features of an image histogram and therefore can be then used to compare two or more images.

3.4.2 Texture Analysis

Texture is a very valuable source of information when analyzing structures in an image. Based on textural analysis both image segmentation [75] and classification [77] can be performed. Typically texture analysis is performed on gray level images [28, 29, 37, 38, 54, 113, 197] but it can also be used for color texture analysis [93, 100, 181, 188, 206]. Gray level textures are usually chosen due to the fact that they involve less computation complexity and providing accurate results at the same time. There are many different algorithms used for texture analysis like simple histogram calculations (see section 3.4.1). More advanced and more accurate algorithms use Markov Random Field theory [29, 54, 113, 198] and the Gray Level Co-occurrence Matrix [38, 40, 77, 197]. In [40] the authors proposed a new method for the determination of the Gray Level Co-occurrence Matrix. Their method is faster than previously known algorithms.

There are other methods for texture analysis that can be found in the literature. These methods include variograms [26, 116, 177] and segmentation based on region growing [31, 189, 211] that includes the neighboring points to the area of interest by searching for textural similarities [30, 6].

Textural segmentation is a powerful tool in image analysis and was shown to be useful for segmentation and classification of the cellular structures in microscopic images [5, 121, 178, 197, 198]. The textural segmentation of nuclei in fine needle aspiration biopsy images is discussed in more detail in section 4.2.5.

3.4.3 Morphological Operations

Morphological operations are a very useful tool in image processing allowing us to analyze shapes in an image. Some of the morphological techniques include erosion, dilatation, opening, closing, thinning and skeletonisation.

- **Erosion** – We define the erosion of a figure A with structuring element B as a set of central points of all structuring elements, which are completely contained inside figure A. Fig. 3.11b illustrates this procedure. We can also write an erosion as:

$$E(A, B) = A \ominus (-B) = \cap_{\beta \in B}(A - \beta) \qquad (3.13)$$

- **Dilation** – The definition of dilation is formulated analogically to the definition of erosion. Therefore, dilation of figure A with structuring element B is a set of central points of all structuring elements for which at least one point overlays with the figure A. We can find an example of this operation in Fig. 3.11. Mathematical expression of dilation is described by equation 3.14

$$D(A, B) = A \ominus B = \cup_{\beta \in B}(A + \beta) \qquad (3.14)$$
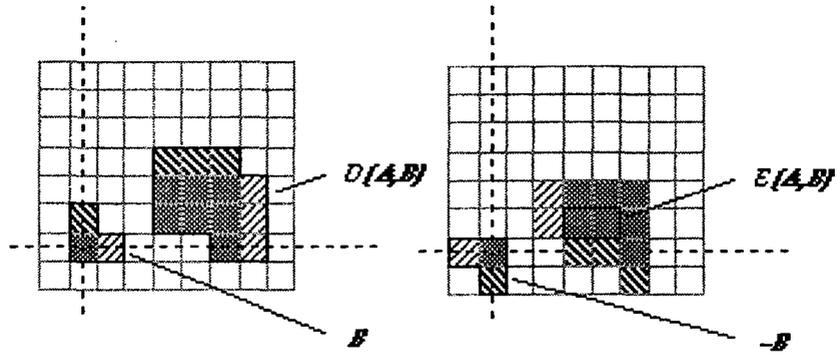
Figure 3.11: Dilation (left) and Erosion (right)

The application of erosion and dilation results in edge smoothing of the figure and therefore they are often used as simple morphological filters. The problem with these methods is that they interfere with the area of a figure. Erosion decreases the area while dilation enlarges it. Considering that fact we can notice that the order in which these two operations are used is significant. Opening (o) and closing (•) are two operations that combine erosion and dilation.

- **Opening** – An application of erosion followed by dilation and can be written as:

$$O(A, B) = A \circ B = D(E(A, B), B) \tag{3.15}$$

- **Closing** – this operation first applies dilation which is then followed by erosion. Its mathematical representation is as follows:

$$C(A, B) = A \bullet B = E(D(A, -B), -B) \tag{3.16}$$

The above operations are very useful in image processing. Opening allows for removing small and unnecessary objects and can also be used for separation of some combined shapes. Closing on the other hand fills small cavities and holes in the figure. What is most important that both of these operations don't change the shape of big objects with smooth edges.

| thresholding | dilation | erosion | opening | closing |
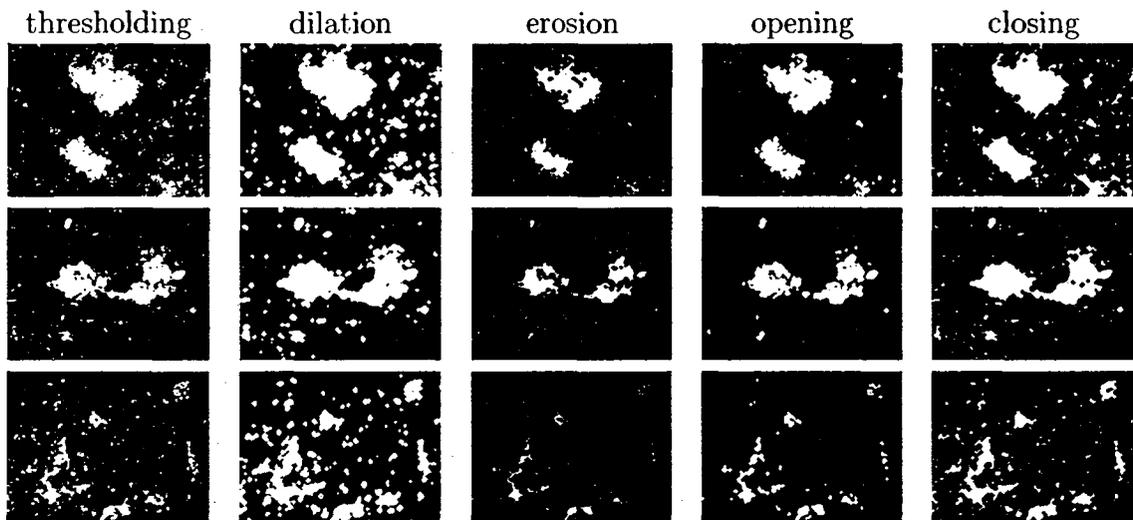
Figure 3.12: Example of morphological operations.

3.4.4 Connected Components Labelling

Estimating connected components within an image allows for more precise analysis of the information provided in that image. It can be seen as the extraction and representation of a single nuclei from an input image. If we consider a binary image $\mathcal{I}(x, y) = v$, where $v = 1$ or $v = 0$, then we can say that (x, y) is a connected component to (x', y') with respect to v if $(x, y) = (x_0, y_0), (x_1, y_1), ..., (x_n, y_n) = (x', y')$ and $\mathcal{I}(x_i, y_i) = v$ for $i = 0, 1, ..., n$, where $(x_i, y_i) \in N_8(x_{i-1}, y_{i-1})$, $i > 0$. A connected component of value v is a set of pixels in which every pair of pixels is connected and assumes a value v. Fig. 3.13 shows an example of the binary image and its labelled connected components.

Definition 2. *A connected components labelling of a binary image $\mathcal{I}$ is a labelled image $\mathcal{LI}$ in which the value of each pixel is the label of its connected component.*

If we consider a *label* to be a unique description of a component, the *connected component labelling* can be defined according to the definition 2, taken from [174].

Using this definition, the algorithm for computation of the labelled components can be described. Shapiro and Stockman [174] described a recursive algorithm for finding the component from a binary image $\mathcal{I}$ of a size (MaxRow +1) × (MaxCol +1). The algorithm first negates all the 1–pixels to assign them a value of -1 in order to distinguish unprocessed
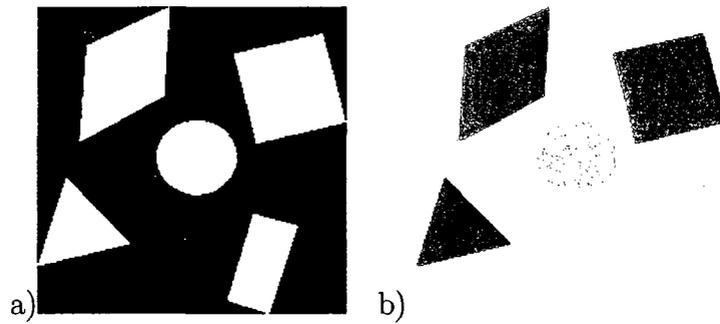
Figure 3.13: Illustration of a connected components. a) Binary image, b) Labelled image.

pixels (-1) from those that belong to the label 1. The algorithm finds pixels with a value -1 and assigns them with a new label. In the next step it calls a recursive searching procedure to find all their -1 neighbors. Another procedure used in this algorithm searches and returns a set of pixels that belong to the 4 or 8-connected neighborhood of the pixel (L, P).

Fig. 3.14 shows an example of image labelling applied to the fine needle aspiration biopsy images.

3.4.5 Edge Detection

Edge detection allows us to identify localization of the objects in an image and therefore plays a key role in image processing. In general edge detection is a combination of three types of image operations: smoothing, convolution and thresholding [38, 45, 76, 120, 132, 193]. An edge is defined as a change in the intensity. We can distinguish between the three types of edges. These types are the most often seen in images. The first type of edges (Fig. 3.15a) is the easiest to detect and is characterized by a sudden intensity change. The second type is a local intensity change (Fig. 3.15b) and the third is based on a texture change (Fig. 3.15c). Each edge type requires a different detection approach. One of the
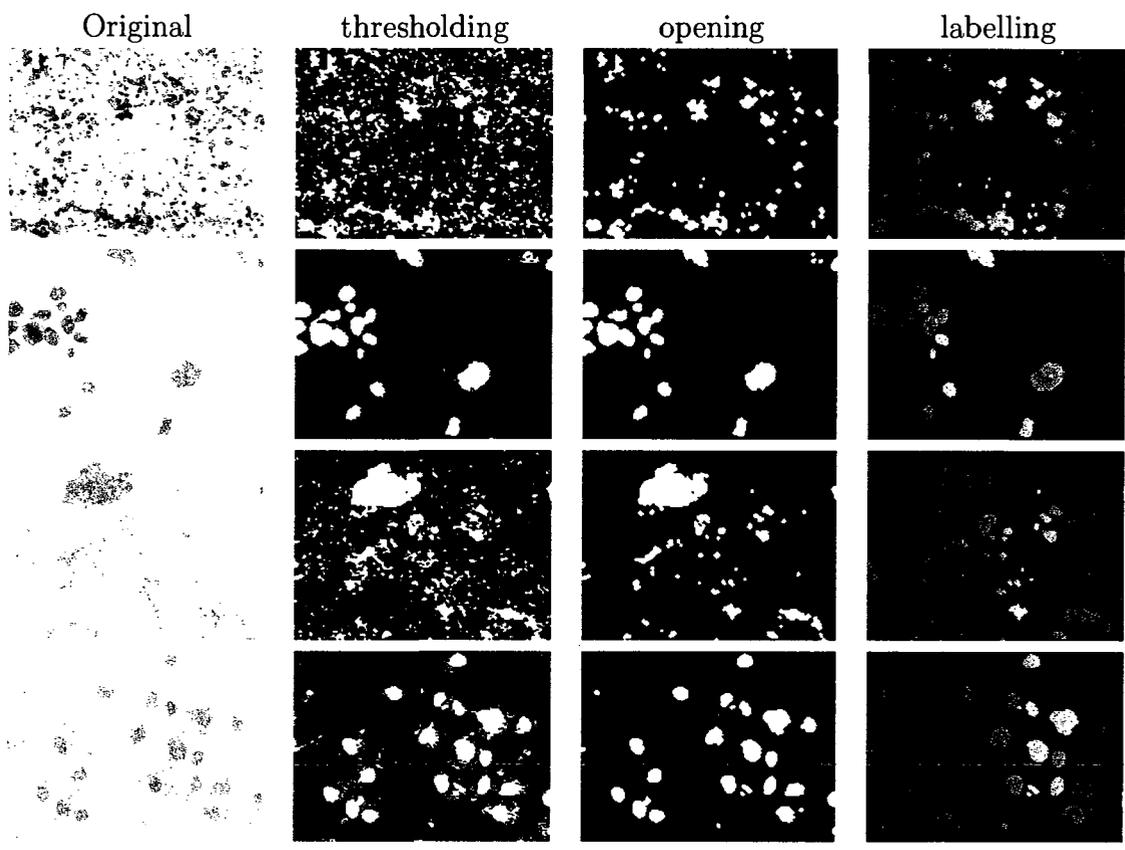
Figure 3.14: Example of nuclei labelling.
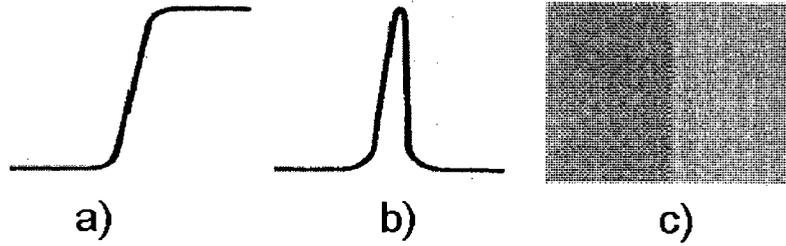
Figure 3.15: Different types of edges.

most popular and simplest approaches is based on the first and second derivatives. In reality, it is difficult to calculate the derivative from the image but a very good approximation of the derivative can be used. That is a local gradient that for an image $I(x, y)$ is defined as:

$$G_x = \frac{\partial I(x, y)}{\partial x} \qquad (3.17)$$

$$G_y = \frac{\partial I(x, y)}{\partial y} \qquad (3.18)$$

In digital images, gradient can be calculated as a difference of two neighboring pixels intensities using the so–called Robert's operators in four different directions. Robert's operators are defined as:

+1	0
0	-1

G_x

0	+1
-1	0

G_y

Using values calculated according to the above description we can calculate the gradient's magnitude (Eq. 3.19) and direction (Eq. 3.20).

$$\|\nabla f\| = \sqrt{G_x^2 + G_y^2} \qquad (3.19)$$

$$\phi = \arctan \frac{G_y}{G_x} \qquad (3.20)$$

A Robert's operator is very sensitive to noise and therefore performs poorly on noisy

images. Better results can be obtained using Prewitt's and Sobel's operators that are defined as:

1	1	1
0	0	0
-1	-1	-1

0^0

-1	0	1
-1	0	1
-1	0	1

90^0

Prewitt

1	2	1
0	0	0
-1	-2	-1

0^0

-1	0	1
-2	0	2
-1	0	1

90^0

Sobel

Another edge detection method is Laplacian. This method uses second derivatives and for image $I(x,y)$ is defined as:

$$\nabla^2(x,y) = \frac{\partial^2 I(x,y)}{\partial^2 x} + \frac{\partial^2 I(x,y)}{\partial^2 y} \qquad (3.21)$$

or in discrete from:

0	1	0
1	-4	1
0	1	0

1	1	1
1	-8	1
1	1	1

Directions:
Two & Four

This method is very sensitive to noise but when applied provides us with very thin edges. The usage of edge detection algorithms is a tradeoff between detection precision and noise reduction.

The most accurate edge detection method was introduced in 1986 by Canny [25]. This method seeks for the zero–crossings of the second derivative of the image convolved with a filter in the gradient direction for which the magnitude is larger than a predefined threshold. The threshold value is estimated according to the image statistics. The zero–crossings relate to the derivative's maxima and minima in this direction, which are considered as

edges in the image. This can be described with equation 3.22.

$$\frac{\partial^2(G * I)}{\partial n^2} = \frac{\partial([\frac{\partial G}{\partial n}] * I)}{\partial n} \qquad (3.22)$$

where n is the direction of the gradient in the smoothed image [24, 25] and $*$ is convolution. In this method the Gaussian filter is used for smoothing the image and the static threshold value is exchanged for the hysteresis threshold which can be adopted depending on the content of the image. Canny edge detection uses double thresholding in which pixels above the high threshold value are considered as edges and these below that value can be classified as edges if they are greater than the low threshold value. The performance of the Canny method depends on the Gaussian filter and the chosen threshold values. The larger values of the standard deviation σ for the filter allow for better edge detection in the noisy images and the detection of larger edges. Small values can introduce more faulty edges, although they allow for the detection of finer edges due to the smaller blurring of the original image. The quality of the Canny edge detection purely depends on the choice of these parameters which should be determined depending on the desired results and the amount of noise in the image.

The application of an edge detection algorithm is usually considered as one of the first stages of image analysis systems. Edges extracted by these methods are usually thicker than one pixel and therefore there is a need for additional processing of such edges. The process of reducing the width of the edge in the image is called thinning. This procedure was first described by Sherman in [175] and found itself useful in numerous application nowadays [56, 161, 166, 179, 187]. We can find applications of thinning in such areas as examination of electronic circuits, chromosome shape analysis, fingerprint classification and letter recognition [137]. In this study the thinning algorithm was used for the detection of nuclear boundaries during initial level set extraction from original images. Fig. 3.16 shows an example of the described edge detection algorithms applied to the fine needle aspiration biopsy images.

| Original | Roberts | Sobel | Laplacian | Canny |
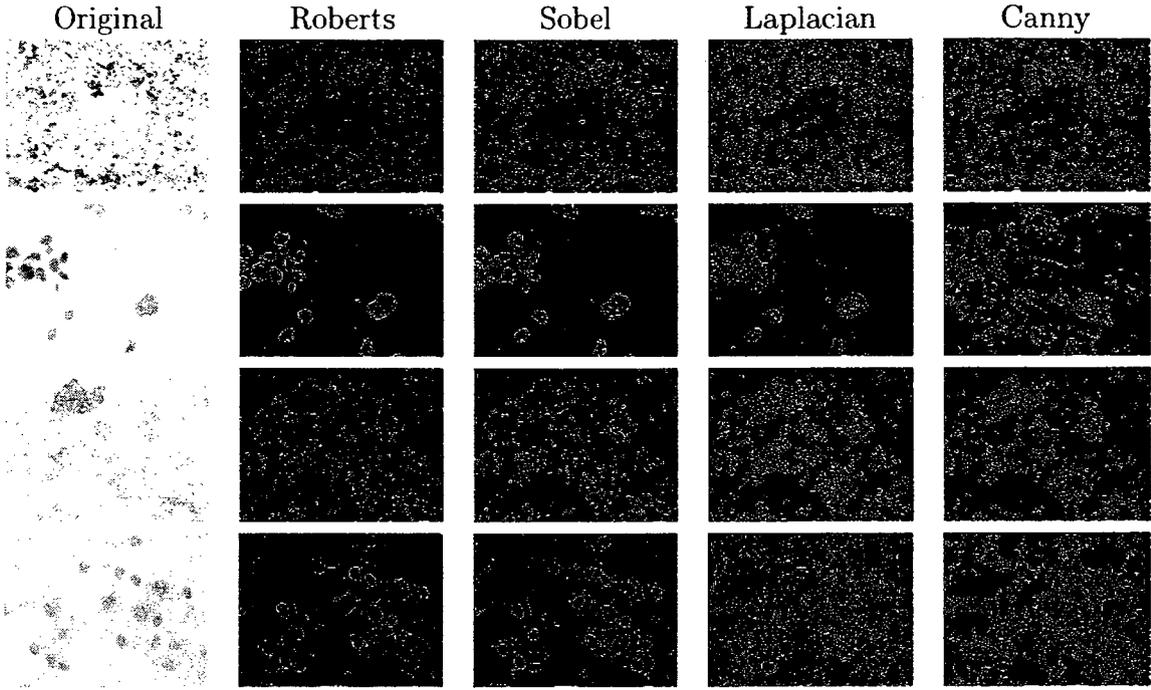
Figure 3.16: Comparison of edge detection algorithms for nuclear boundary extraction.

3.4.6 Moments

A moment is defined by the following equation:

$$m_{pq} = \sum_x \sum_y x^p y^q \mathcal{I}(x, y), \tag{3.23}$$

where $(p + q)$ is the order of the moment [195]. Now, if we redefine $(\overline{x}, \overline{y})$ as:

$$\overline{x} = \frac{m_{10}}{m_{00}} \quad and \quad \overline{y} = \frac{m_{01}}{m_{00}} \tag{3.24}$$

then the central moments, that are translation invariant, can be defined by:

$$\mu_{pq} = \sum_x \sum_y (x - \overline{x})^p (y - \overline{y})^q \mathcal{I}(x, y). \tag{3.25}$$

Normalization of equation 3.25 will allow for the definition of momentum–based, RST–invariant features. The normalization of the central moment is performed with the following equation:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \tag{3.26}$$

where

$$\gamma = \frac{p + q}{2} + 1 \quad \text{for } (p + q) = 2, 3, 4, \ldots$$

Apart from central invariant moments, we can also distinguish moments that are translation, rotation and scaling invariant. These moments are called affine moment invariants (AMI) [81]. In [185] the authors described a graph method for generating affine moment invariants and Rosin [162] showed that AMIs can be used for the identification of ellipses. If we use I_1 defined by equation 3.27 we can determine if the figure in question is an ellipse.

$$I_1 = \frac{\mu_{20}\mu_{02} - \mu_{11}^2}{\mu_{00}^4} \tag{3.27}$$

For an ellipse, the result of the equation 3.27 is $I_1 = \frac{1}{16\pi^2}$.

3.4.7 Shape Descriptors

During the preprocessing and segmentation stages, an FNA image underwent a number of operations, including thresholding and extraction of contours. It can now be transformed to a binary representation for shape analysis. At this stage shape descriptors are calculated to obtain numerous representations of the shape in the image. It is not possible to recreate the shape in the image from this set of numbers, but it is possible to analyze the shape variation between given examples. Shape descriptors should efficiently differ for different shapes to allow good discrimination between them.

In the literature we can find numerous examples of shape descriptors applications for better shape classification [41, 66, 67, 91, 102, 142]. There are many different shape descriptors. Some of them only depict well known parameters such as area, perimeter, circularity, etc. Other will describe more complex features such as circular or elliptical variance, convexity, or eccentricity. Here, we present some of the descriptors used in literature for different types of application, not only for nuclei representation.

- Area (A) – the number of pixels inside the shape:

$$A = \sum_{(x,y) \in S} S(x,y). \tag{3.28}$$

- Perimeter (P) – the number of pixels on the contour (c) of the shape:

$$P = \sum_{(x,y) \in C} c(x,y). \tag{3.29}$$

- Circularity (Cir) – describes how much the shape differs from the circle [195]:

$$Cir = \frac{P^2}{A}. \qquad (3.30)$$

Circularity of the circle will yield a value of 4π while all other shapes will assume the value of circularity larger than 4π.

- Eccentricity (Ecc) – in general, eccentricity is defined as a ratio of the length of the longest chord of the shape to the longest chord perpendicular to it. In this study we use a modified definition for eccentricity. According to this definition, the eccentricity is defined as the ratio of the distance between the foci of an ellipse that has the same second–moments as the shape, and its major axis length. Values of this feature vary between 0 for a circle and 1 for a line segment.

- Elongation (El) – is defined as a ratio of the height (h) and width (w) of the smallest bounding box which encloses the shape. The bounding box shall be rotated in the same direction as a nucleus.

$$El = \frac{h_{Bb(S)}}{w_{Bb(S)}} \qquad (3.31)$$

where $Bb(s)$ is the smallest bounding box which encloses the extracted shape.

- Rectangularity (Rec) – describes how rectangular the shape is:

$$Rec = \frac{A}{A(Bb(S))}. \qquad (3.32)$$

Values of this feature vary between 0 for a line and 1 for a rectangle.

- Orientation – describes the direction of the shape in the image. A more detailed description of this feature can be found in section 5.2.2.

Peura and Ilvarinen in [142] described a following group of shape descriptors called simple shape descriptors (SSD):

- Convexity – ratio of the shape's convex hull to the length of its contour:

$$C = \frac{CH(S)}{|S_C|} \qquad (3.33)$$

where $CH(S)$ is a convex hull of the shape S and $|S_C|$ is a contour length of S. In this study we use a modified definition of this feature, which is described in section 5.2.2 in more detail.

- Ratio of the principal axes:

$$PRAX = \frac{c_{yy} + c_{xx} - \sqrt{(c_{yy} + c_{xx})^2 - 4(c_{xx}c_{yy} - c_{xy}^2)}}{c_{yy} + c_{xx} + \sqrt{(c_{yy} + c_{xx})^2 - 4(c_{xx}c_{yy} - c_{xy}^2)}}, \tag{3.34}$$

where the image covariance matrix is

$$Cov = \begin{bmatrix} c_{xx} & c_{xy} \\ c_{yx} & c_{yy} \end{bmatrix} = \frac{1}{|\mathcal{I}|} \sum_{p \in \mathcal{I}} (p - \mu)(p - \mu)^T \tag{3.35}$$

and the image center of mass is

$$\mu = \frac{1}{|\mathcal{I}|} \sum_{p \in \mathcal{I}} p. \tag{3.36}$$

- Compactness – ratio of the circle with the same area as the area of the shape to the shape contour length:

$$COMP = \frac{2\sqrt{\mathcal{I}\pi}}{|\mathcal{I}_E|}. \tag{3.37}$$

- Circular variance – mean squared fitting error of the shape to the circle:

$$CV = \frac{1}{\mathcal{I}\mu_r^2} \sum_{p \in \mathcal{I}} (\|p - \mu\| - \mu_r)^2, \tag{3.38}$$

where an average radius length is

$$\mu_r = \frac{1}{\mathcal{I}} \sum_{p \in \mathcal{I}} \|p - \mu\|. \tag{3.39}$$

- Elliptical variance – mean squared fitting error of the shape to the ellipse:

$$EV = \frac{1}{\mathcal{I}\mu_c^2} \sum_{p \in \mathcal{I}} (\sqrt{(p - \mu)^T Cov^{-1}(p_i - \mu)} - \mu_c)^2, \tag{3.40}$$

where an average radius length is

$$\mu_c = \frac{1}{\mathcal{I}} \sum_{p \in \mathcal{I}} \sqrt{(p - \mu)^T Cov^{-1}(p_i - \mu)}. \tag{3.41}$$

54

Depicted descriptors are able to efficiently describe the shape in the image only when the contour image is available, therefore it should be preceded by the edge detection algorithm. The strength in representing the shapes is dependent on the precision with which the background information was removed before conversion of the image into its binary representation. In section 5.2.2 of this thesis we propose the set of features for shape analysis of the breast cancer nuclei. The proposed descriptors are then used as features in the classification stage of the malignancy grading system.

3.5 Conclusions

In this chapter we have reviewed methods that can be used for the representation of the elements in the image. Here, we can also find a description of fundamental knowledge about the microscopy with the review of different areas where microscopes found to be very useful. The examples presented show the behavior of the depicted methods when applied to the images used in this thesis. From fig. 3.16 it can be noticed that the simple edge detection algorithms do not lead to a good nuclear boundary representation and it can be assumed that a more sophisticated method is needed for such a representation. The reviewed edge detection algorithms provide too much noisy information in form of additional boundaries that can be later lead to mistaken results. So, in chapter 4 more advanced methods are studied for precise nuclear boundary extraction and representation. The algorithms described in this chapter are very popular and are widely used in the scientific community. With the use of such techniques we are able to extract nuclei in the fine needle aspiration biopsy slides and then we can extract some valuable features not only based on the shape descriptors but also with the application of histograms and moments. In chapter 5 we will describe the features that were extracted from the FNA images to represent the breast cancer nuclei. This representation is then used in chapter 6 for classification of cancer malignancy according to the Bloom–Richardson grading scheme.

Chapter 4

Nuclear Segmentation

4.1 Introduction

Image segmentation is a process in which similar regions of an image are identified. Segmentation is a crucial stage in any classification system because it improves feature extraction. Good features are the most important part of the classification process. In this section we will look into some segmentation techniques that can be used when estimating a cancer's malignancy.

To properly segment images from the database (see Appendix A), the task needs to be divided into two subproblems. The first problem deals with segmentation of images recorded at low magnification and the second problem is segmentation of high magnification images. There are two separate techniques used for segmentation in each problem domain.

In this chapter the reader can find the review of the segmentation algorithms used for nuclear boundary representation followed by the segmentation results on FNA images. Results show a comparison of the boundary representation to illustrate the differences between all methods.

4.2 Image Segmentation Techniques

4.2.1 Thresholding

Thresholding is one of the fundamental segmentation methods. For this algorithm we need to specify a threshold value T (in the same color convention as the input image). Next, pixels are divided into two classes according to the equation 4.1.

$$g(x,y) = \begin{cases} 1 & \text{if } f(x,y) \geq T \\ 0 & \text{if } f(x,y) < T \end{cases} \tag{4.1}$$

where $g(x,y)$ is the output image and $f(x,y)$ is the input image.

We can also introduce multiple threshold values ($T = [T_1, T_2, ..., T_N]$) which will lead to segmentation into $N+1$ regions. Eq. 4.2 describes thresholding into three different regions using two threshold levels (T_1 and T_2).

$$g(x,y) = \begin{cases} A_1 & \text{if } f(x,y) \leq T_1 \\ A_2 & \text{if } T_1 < f(x,y) \leq T_2 \\ A_3 & \text{if } f(x,y) > T_2 \end{cases} \tag{4.2}$$

where A_1, A_2 and A_3 represent segmented regions.

4.2.2 Hough Transform

Hough Transform (HT) is a method that allows us to detect shapes in images. It performs well for noisy images and also when object edges are not fully preserved. HT was initially introduced to locate lines and analytical curves within images [80, 160]. In general, Hough Transform maps image points into Hough Space also called an accumulator. Mapping is based on an analytical description of the shape of interest.

The concept of curve detection can be easily explained using a line as an example. If L passes through some point (x,y), Fig. 4.2.2a. We can then represent that line analytically with the equation 4.3:

$$L : y = ax + b, \tag{4.3}$$

where a is a slope and b is the intersection point on the y–axis.

In such a representation x and y are free parameters. If we change the above representation in such a way that a and b become free parameters and x and y are fixed, then line L will be described by the following equation:

$$L_{ab} : b = y - ax, \qquad (4.4)$$

where x is a slope and y is the intercept on the b–axis.

The linear Hough Transform is based on the above dual representation of a line and the fact that all points on the line L will generate a set of lines in ab space that intersect at point (a, b).

With the above representation a problem arises for vertical lines for which $a = b = \infty$. To overcome this problem a parametric representation was introduced:

$$x \cos \alpha + y \sin \alpha = r, \qquad (4.5)$$

where r is the length of a normal from the line to the origin and α is the orientation of r with respect to the x-axis, see Fig. 4.2.2.

The process of detecting lines reduces to finding intersections in the Hough space. Typ-

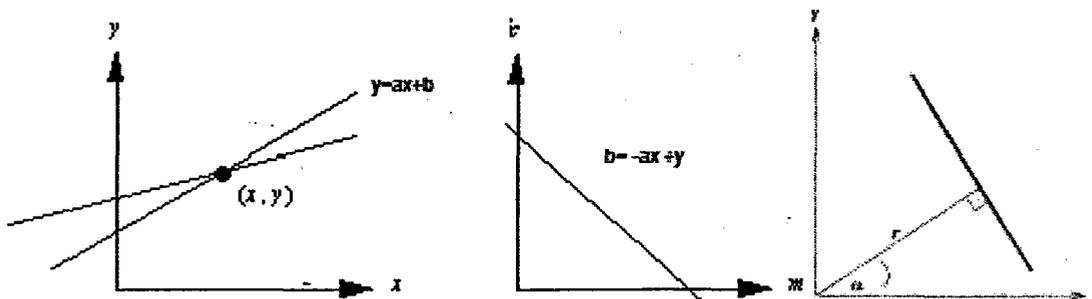Figure 4.1: Illustration of a Hough Transform. a) Line L, b) Dual representation of L, c) Parametric representation

ically it is achieved by creating an accumulator array Acc, which size depends on the maximum values for a and b. In the next step, for each image point (x, y) the position in Acc, that satisfies equation 4.4, is incremented. Finding maxima in the accumulator array

58

is equivalent to locating that points in the ab space are on a line corresponding to a and b which index the maximum cell.

We can now generalize the above discussion to other analytical curves by substituting equation 4.3 or 4.5 with an equation for the curve of interest. For example, for ellipse we would substitute the above equations with equation 4.6 or 4.7.

$$\frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} = 1, \tag{4.6}$$

$$x_0 = x - a\cos\alpha, \quad y_0 = y - b\sin\alpha \tag{4.7}$$

The main drawback of this procedure is that the more free parameters needed to describe the curve the larger the accumulator space becomes. To resolve this problem, Ballard [9] introduced the Generalized Hough Transform (GHT) to locate arbitrary shapes. This process is similar to the previous idea of mapping. Before the actual mapping to the Hough space, a representation of the shape to be detected is needed. This is achieved by the creation of a look-up table, called R-Table, using the directional information at the boundary points. This procedure allows us to eliminate the need of an analytical description of the shape, as was the case for lines, circles and ellipses.

To construct an R-Table we choose a reference point $X_R = (x_r, y_r)$ inside the shape of interest. Next, for each boundary point $X_B = (x_b, y_b)$ we calculate the following:

$$r = |X_B - X_R|. \tag{4.8}$$

We now store r in the R-Table as a function of ϕ, $r(\phi)$. The second stage consists of mapping of an unknown shapes into a Hough space. For each boundary point X_B, we calculate a gradient direction according to equation 3.20. We then look for r in the R-Table and increment all the positions in accumulator array that correspond to $X_B + r$.

To be able to locate arbitrary shapes using the generalized version of HT, we need to include scaling and rotation factors in our calculations. This will guarantee the performance is resistant to size and orientation changes. Although the accuracy is improved, the size of an accumulator has to be increased. Including scaling and rotation will require an accumulator

of the form $Acc(X_B, s, \phi)$, where $X_B = (x_b, y_b)$. As we can see, a 4-dimensional accumulator is needed.

4.2.3 Level sets

Active contours, deformable models, or snakes, allow us to precisely reflect boundaries of objects in the input image. The algorithm takes as an input an initial guess for the contours of the boundaries that is then fitted to the actual boundaries of desired objects in the image. With active contours two types of energies are considered during the fitting process. The first type is internal energy (E_I) that is defined within the curve. It is responsible for maintaining the smoothness of the curve during deformation, or in other words it measures the degree of bending along the curve. External energy (E_E), which is the second type of energy, is calculated from the image. Its main task is to measure the edginess of the region through which the initial contour passes. The boundary fitting is based on minimization of the energy which is a sum of the two associated energies (Eq. 4.9).

$$E = E_I + E_E \tag{4.9}$$

Internal energy is defined by equation 4.10 and external energy is defined by equation 4.11

$$E_I = \sum \alpha \|X_i - X_j\| + \beta \|X_{i-1} - 2X_i + X_{i+1}\|, \tag{4.10}$$

$$E_E = \sum \exp(-\|\nabla f(X_i)\|), \tag{4.11}$$

where $X_i = [x_i y_i]^T$ is a snake point. Minimization of the first term returns snake points that are close together. If this is negative it means that snake shrunk to the single point and therefore minimized. If we minimize the second term we obtain a small bending of the curve.

From the above we can see that snakes are described parametrically. The other type of representation are level set functions that describe the boundary curve implicitly. According to Suri et al. [186] level sets are designed to handle topological changes but are not robust to discontinuous boundaries.

For the purpose of this study the level set method proposed by Li *et al.* [110] was implemented for nuclear segmentation. Level sets were first described in 1988 by Osher and Sethian [140] as a method for capturing moving fronts. In the level set formulation, the segmentation problem is equivalent to the computation of a surface $\Gamma(t)$ that propagates in time along its normal direction. The Γ surface is also called a propagating front, which according to Osher and Sethian [140] is embedded as a zero level of a time–varying higher dimensional function $\phi(x,t)$:

$$\Gamma(t) = \{x \in \mathcal{R}^3 | \phi(x,t) = 0\} \tag{4.12}$$

An evolution equation for an interface Γ, where Γ is a closed curve in R^2, can be written in a general form [173] as:

$$\frac{\partial \phi}{\partial t} + F|\nabla \phi| = 0, \tag{4.13}$$

The function ϕ describes a surface defined by $\phi(x,t) = d$, where d is a signed distance between x and the surface Γ. If x is inside (outside) of Γ then d is negative (positive). The function F is a scalar speed function that depends on the image data and the function ϕ. The main drawback of this procedure is that during the evolution, ϕ can assume sharp or flat shapes. To overcome this problem, ϕ is initialized as a signed distance function before evolution. Later, during evolution, it is periodically reshaped to be a signed distance function [110].

In our framework, we make use of variational level sets, introduced by Li *et al.* [110], that are more robust than those originally proposed by Osher and Sethian because they incorporate shape and region information into the level set energy functions.

In 2005, Li *et al.* [110] proposed a modification of traditional variational level sets to overcome the problem of re–initialization of function ϕ to a sign distance function within the evolution cycle. They proposed an evolution equation of the form:

$$\frac{\partial \phi}{\partial t} = -\frac{\partial \mathcal{E}}{\partial \phi} \tag{4.14}$$

where $\frac{\partial \mathcal{E}}{\partial \phi}$ is a Gateaux derivative of the energy function $\mathcal{E}$ and is represented by the equation 4.15:

$$\frac{\partial \mathcal{E}}{\partial \phi} = -\mu \left[\Delta \phi - \mathrm{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) \right] - \lambda \delta(\phi) \, \mathrm{div} \left(g \frac{\nabla \phi}{|\nabla \phi|} \right) - \nu g \delta(\phi), \qquad (4.15)$$

where Δ is the Laplacian operator, *div* is the divergence operator and $\mu > 0$ is a parameter controlling the effect of penalizing the deviation of ϕ from a signed distance function.

All level set methods start with an initial level set function. The closer the initial level set function is to the final segmentation, typically, the more likely the level set method will quickly converge to the segmentation. Therefore, to automate the segmentation process and start with a good initial level set function, we make use of the iterative clustering approach for automatic image thresholding described in section 4.2.6.

4.2.4 Fuzzy c–means Segmentation

Another segmentation method taken into consideration is a fuzzy approach of Klir and Yuan [94] that can be used to partition the image information to extract nuclei. In general, a set of data $X = \{x_1, x_2, ..., x_n\}$ is supposed to be divided into c clusters with assumption that $P = \{A_1, A_2, ..., A_c\}$ is a known pseudo–partition and A_i is a vector of all memberships of x_k to cluster i. Now, using equation 4.16 the centers of the c clusters can be calculated [191].

$$v_i = \frac{\sum_{k=1}^{n} [A_i(x_k)]^m x_k}{\sum_{k=1}^{n} [A_i(x_k)]^m}, \quad i = 1, 2, ..., c \qquad (4.16)$$

where $m > 1$ is a weight that controls the fuzzy membership. The memberships are defined by equation 4.17 if $\|x_k - v_i\|^2 > 0$ for all $i \in \{1, 2, ..., c\}$ and if $\|x_k - v_i\|^2 = 0$ for some $i \in I \subseteq \{1, 2, ..., c\}$ the memberships are defined as a nonnegative real number satisfying equation 4.18 for $i \in I$.

$$A_i(x_k) = \left[\sum_{j=1}^{c} \left(\frac{\|x_k - v_i\|^2}{\|x_k - v_j\|^2} \right)^{\frac{1}{m-1}} \right]^{-1} \qquad (4.17)$$

$$\sum_{i \in I} A_i(x_k) = 1 \qquad (4.18)$$

The clustering algorithm seeks a set P that minimizes the performance index $J_m(P)$ which is defined by equation 4.19 and the optimal solution to this problem can be found in [17].

$$J_m(P) = \sum_{k=1}^{n} \sum_{i=1}^{c} [A_i(x_k)]^m \|x_k - v_i\|^2. \tag{4.19}$$

Compared to all the other segmentation techniques, fuzzy c–means algorithm (FCM) doesn't need any additional processing and was applied to segment the color information in the image. In [172], Schüpp *et al.* presented an approach which is similar to the one used in this thesis, where the initial level set is obtained by the fuzzy c–means algorithm. Here, we use automatic thresholding for initial level set creation and a fuzzy c–means method is used as a segmentation method itself. From the segmentation results we can see that fuzzy c–means is a sufficient segmentation method when used by itself and the level set would only smooth the nuclear boundary.

4.2.5 Textural Segmentation

A third segmentation technique image textural description based on second order statistics was used to generate grey level co–occurrence texture features [77]. Here, for a spatial window inside the image the conditional joint probabilities C_{ij} are calculated according to equation 4.20 for all pairwise combinations of grey levels assuming that the distance between the pixels is known.

$$C_{ij} = \frac{P_{ij}}{\sum_{i,j=0}^{G-1} P_{ij}} \tag{4.20}$$

where P_{ij} is the frequency of occurrence of two grey levels i and j and G is the number of quantized grey levels. The probabilities are stored as a gray level co–occurrence matrix. where the (i,j) element of the matrix represents the probability C_{ij}. To identify textures within an image we derive four features from the dependency matrix. The four features consist of entropy, correlation, inertia (correlation) and energy. These properties are described by the following equations:

$$Entropy = -\sum_{i,j=0}^{G-1} C_{ij} \ln C_{ij}, \tag{4.21}$$

$$Contrast = \sum_{i,j=0}^{G-1} C_{ij}(i-j)^2, \tag{4.22}$$

$$Inertia = \sum_{i,j=0}^{G-1} \frac{(i-\mu_x)(j-\mu_y)C_{ij}}{\sigma_x \sigma_y}, \tag{4.23}$$

$$Energy = \sum_{i,j=0}^{G-1} C_{ij}^2, \tag{4.24}$$

where σ is the standard deviation and μ is the mean.

To be able to efficiently segment an image using this method, an image needs to be converted to the gray level representation.

4.2.6   Segmentation of Cells Groupings

When segmenting groups from 100x magnification images, there is no need for a sophisticated segmentation algorithm such as the Hough transform or level sets. Here we are concerned only in a rough estimation of the size of the groups present in the image rather than in their precise description. In this study, areas of the image where cells grouping are visible have been segmented.

For this purpose we make use of an iterative clustering approach for automatic image thresholding. This method was first proposed in 1978 by Riddler and Calvard [154]. In principle, their method seeks a threshold T, represented by a curve, within an image, that is restricted to have a bimodal histogram and the final threshold level is calculated according to the following equation:

$$T = \frac{\mu_1 + \mu_2}{2}, \tag{4.25}$$

where μ_1 and μ_2 are the means of the components separated by T.

It is necessary to mention that the image red channel provides best information. From Fig. 4.2 it can easily be noticed that red channel doesn't carry as much background information as the remaining green and blue channels. Due to this reasoning the image red channel is used for calculation of the threshold level and further segmentation of the image.
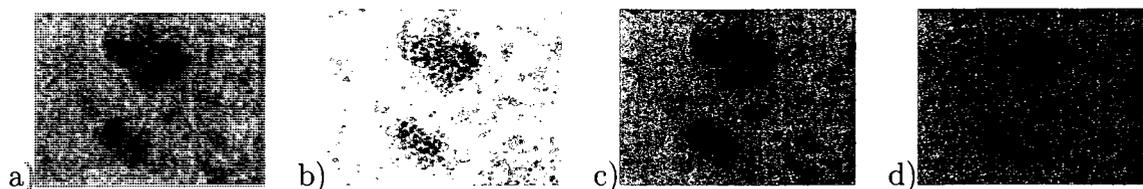
Figure 4.2: Comparison of image RGB channels – 100x magnification. a) Original image (RGB), b) Red channel (R), c) Green channel (G), d) Blue channel (B).

4.2.7 Nuclei Segmentation

Contrariwise to low magnification images, high magnification smears need a sophisticated and precise method of segmentation. These kinds of images are used for extraction of nuclear features and therefore as precise a nuclei representation as possible is required. As was previously mentioned there were numerous attempts for nuclear segmentation in literature. Out of these, level sets seem to be the best choice since they provide good segmentation results in a reasonable amount of time.

In this work four well known segmentation algorithms were compared, to test their applicability to breast cancer nuclei segmentation. The compared methods include the Hough transform, level sets, fuzzy c–means and texture based segmentation. Comparative results are presented in section 4.3.

For high magnification images we also use the red channel because it provides best information about nuclear structures. During the staining process nuclei stain with shades of purple and when the red channel is extracted all the nuclear features are preserved while the background information is removed (see Fig. 4.3). This observation leads us to extract and threshold the image red channel and then to use it for nuclear segmentation.
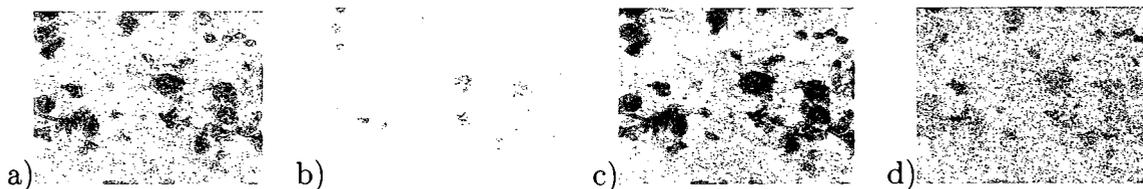

Figure 4.3: Comparison of image RGB channels – 400x magnification. a) Original image (RGB), b) Red channel (R), c) Green channel (G), d) Blue channel (B).

4.3 Segmentation Results

In this section segmentation results on FNA biopsy images at 100x and 400x magnification are presented. Due to the nature of the images taken into consideration, the segmentation task was divided into two subproblems. Images with low magnification that are used for group–based feature extraction do not require an implementation of complicated segmentation algorithms such as Hough transform or level sets. Simple thresholding is good enough for a rough estimation of the groups in the image [86]. Figs. 4.4 and 4.5 show an example of the segmentation of 100x magnification images. From the figures we can see that applying simple thresholding with automatic threshold level estimation is a good method for segmenting these kind of images. We can see that groups are well separated and the background information is removed. Application of morphological filters with small oval structuring element can eliminate small artifacts as can be seen on Fig. 4.6.
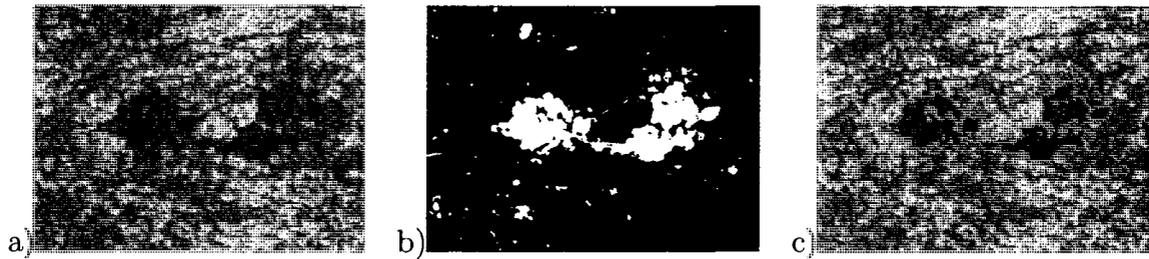

Figure 4.4: Example of segmentation results for 100x magnification image with intermediate malignancy grade. a)Original image, b) Thresholded image, c) Contour
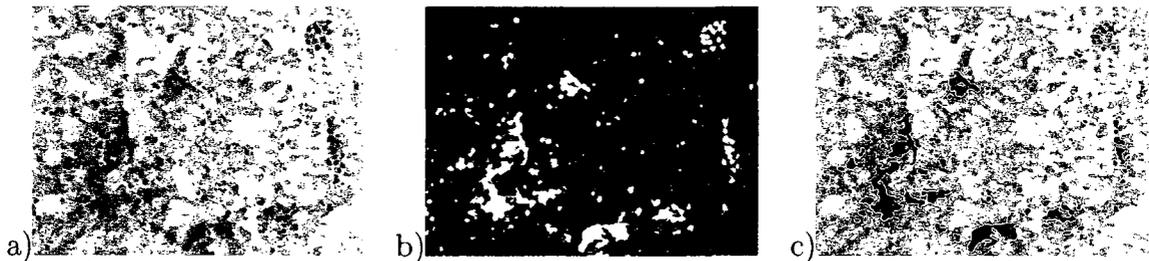

Figure 4.5: Example of segmentation results for 100x magnification image with high malignancy grade. a)Original image, b) Thresholded image, c) Contour
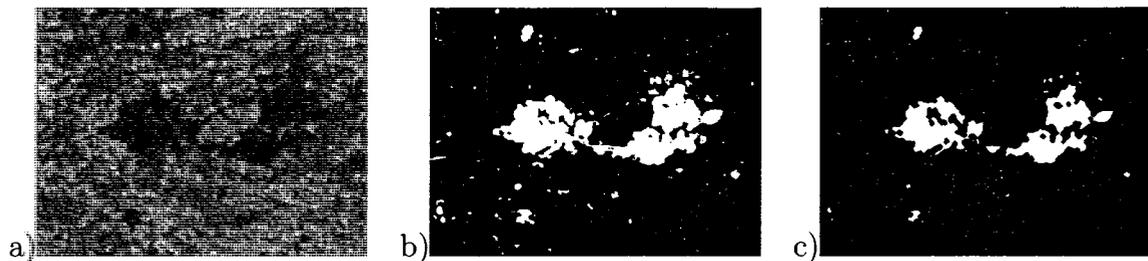
66

Figure 4.6: Elimination of small artifacts by morphological filtering. a) Original image, b) Thresholded image, c) Eroded image

The second type of images in the database (see Appendix A) consist of FNA smears recorded at 400x magnification. These images require more sophisticated methods of segmentation than 100x magnification images because they are used for extracting features related to cells nuclei. These features need a good estimation of nuclear boundaries for the calculation of shape–based features and the determination of the staining intensity of possibly all the nuclei in the image. For this purpose we compared different segmentation techniques to show which one performs best and works in a reasonable time. Here we compared well known methods such as the Hough transform and level sets. From these algorithms, the Hough transform (HT) is a well know technique for elliptical objects segmentation and at first seems to be a very good choice for nuclei segmentation [87]. From the literature review we know that this is the most popular technique used for this kind of segmentation. The result of the application of HT to images in our database is presented on Fig. 4.7b. For comparison, Fig. 4.7c presents level set segmentation for the same image. From the figure we can see that the HT does not reflect the nuclear boundary as precisely as the level set. It is necessary to note that to obtain a final contour with HT an additional active contour algorithm for precise boundary estimation shall be applied, while with level set being an active contour algorithm itself, there is no need for additional methods. This observation allowed us not to use the HT with an additional technique. Introduction of an additional active contour algorithm would have an impact on the running time of the segmentation which is essential during medical diagnosis. Long running times will make such a system less interesting for pathologists since it will be faster for them to look and

diagnose the slide themselves.

This observation led to the introduction of two additional segmentation algorithms for comparison purposes. Both methods proved to be powerful in medical imaging as mentioned in previous chapters. They are also faster than the Hough transform. Here, fuzzy c–means (FCM) and gray level quantization (GLQ) methods were used for nuclei segmentation. Fig. 4.8 shows segmentations obtained with these two methods for single nuclei. From the figure we can see that FCM provides us with a better separation of nuclei than GLQ.

It can be easily noticed that during the segmentation process sometimes it might be difficult to set a border between two nuclei which are close to each other. This can happen when the two nuclei overlap either completely or partially. In pathology, this problem is solved by taking into consideration only the visible parts of each nuclei. In this research, we also disregard these parts of nuclei that are hidden underneath other nuclei. The features calculated in chapter 5 are averaged for each image, which allows us to minimize the impact of the overlapping cells or nuclei.
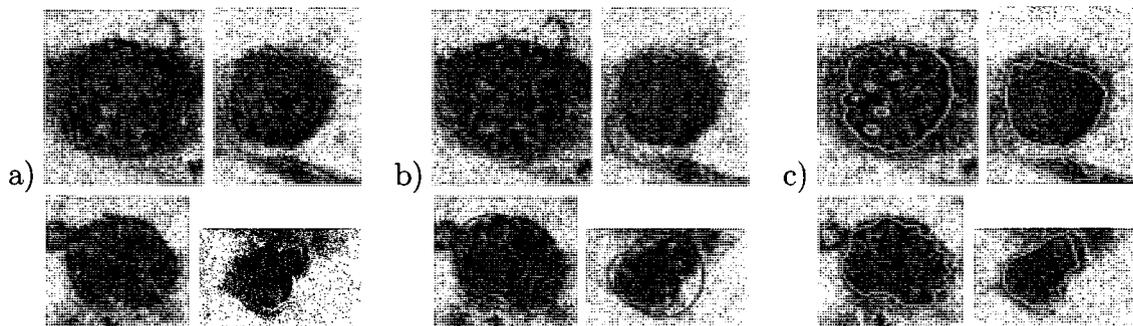

Figure 4.7: Segmentation results for 400x magnification images. a) Original images, b)Hough transform segmentation, c) Level set segmentation.
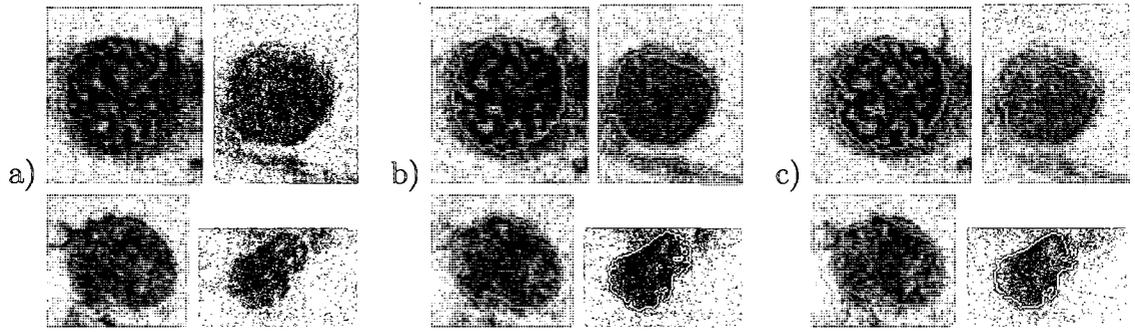
Figure 4.8: Segmentation results for 400x magnification images. a) Original images, b)Fuzzy c–means, c) Gray level quantization.

Segmentation	Case	
algorithm	1	2
HT	3.02	4.25
FCM	0.01	0.01
GLQ	0.01	0.01
LS	0.25	0.26

Table 4.1: Segmentation running times for images from Fig. 4.9 in minutes.

4.4 Conclusions

This chapter of the thesis was concerned with the segmentation of nuclei. The presented results show the behavior of four well known and widely used segmentation algorithms. From these results the following conclusions can drawn:

1. Hough Transform

   - Precise method for nuclei localization. Out of the four compared methods, HT provides the best localization of nuclei in the image [87].

   - The boundaries of nuclei are not well represented (see Fig. 4.7b).

   - To obtain a precise boundary representation, HT requires an additional active contour algorithm.

   - Introduction of an additional algorithm increases the running time of the segmentation and therefore makes it less attractive for the pathologists.
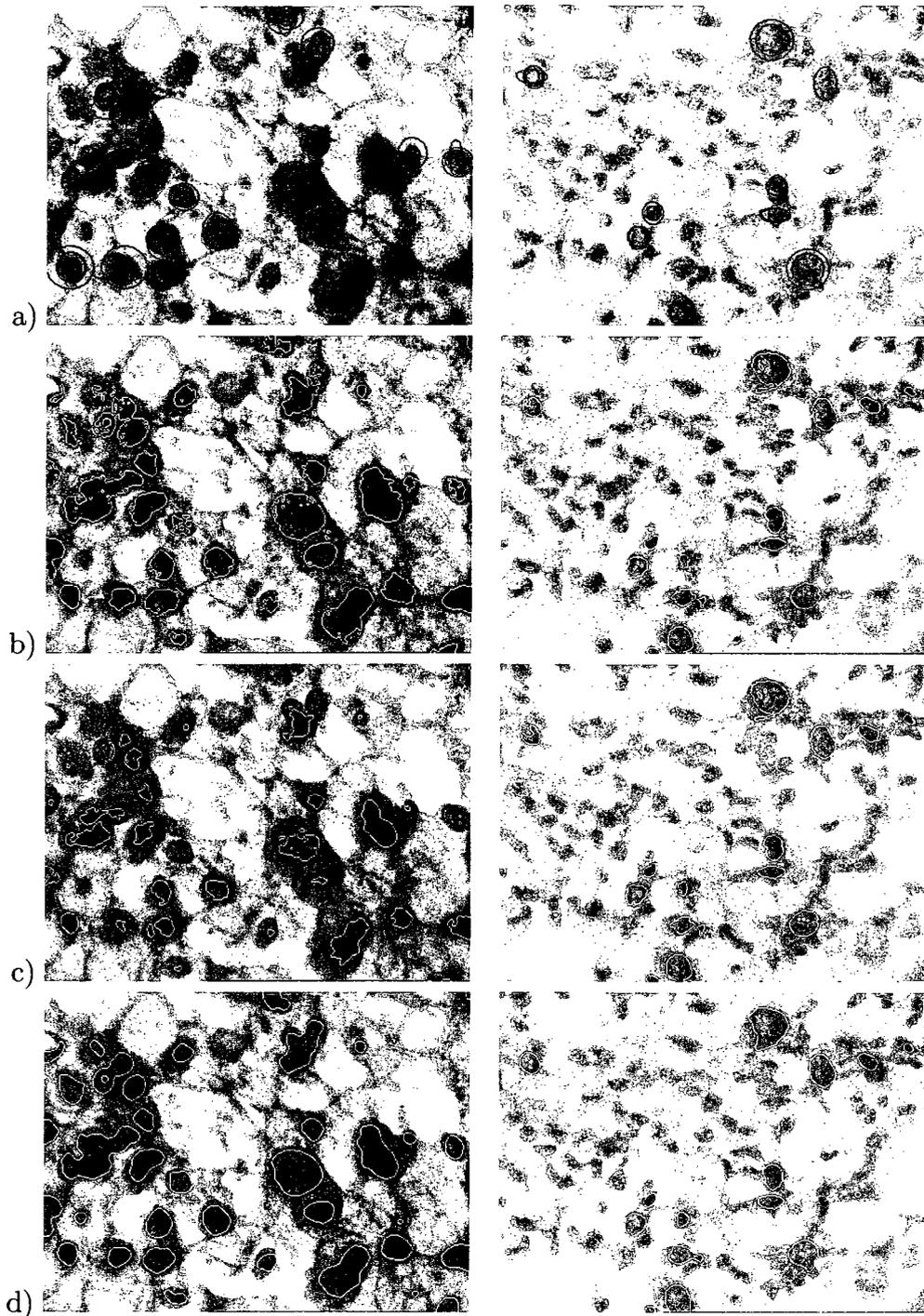
Figure 4.9: Comparison of segmentation results for 400x magnification images. a) Hough transform, b)Fuzzy c–means, c) Gray level quantization, d) Level set.

- There is a possibility of multiple and confusing nuclei localization (see Fig. 4.9a).

2. Level Sets

  - Precise method for boundary representation.

  - This method requires an initial level set – initial boundary.

  - Comparing to HT, the running time of level sets is much faster (see Table 4.1). The initial boundary can be obtained with thresholding which reduces the computation load.

  - Provides a very good trade–off between the running time and boundary representation precision.

3. Fuzzy C–means

  - This method is based on color classification of the neighboring pixels.

  - Provides good representation of the boundaries.

  - The boundary representation is better than that obtained with HT and Gray level quantization but not as good as with level sets.

  - Fuzzy c–means does not require any initial boundary calculations.

4. Gray Level Quantization

  - Segmentation is based on a textural description of the nuclei.

  - Similar to fuzzy c–means, gray level quantization does not require any initial contour calculations.

  - The boundary representation is better than with HT but not as precise as with either level sets or fuzzy c–means.

In general, it can be noticed that level sets, fuzzy c–means and gray level quantization are a better choice for nuclei segmentation when compared to Hough transform alone. In the remaining parts of this thesis we will concentrate only on these three algorithms.

Fig. 4.9 shows a comparison of the segmentation methods and Fig. 4.9a–c illustrates the techniques chosen for breast cancer nuclei segmentation. From this figure it can be noticed that although FCM provides better nuclei segmentation then GLQ, the level sets represent the nuclear boundary the most precisely. It can easily be noticed that GLQ algorithm loses a lot of nuclear information during the segmentation.

The presented results show that the proposed choice of segmentation techniques is adequate and should lead to a good feature extraction and class separation [85]. In the following section the proposed set of features is presented.

Chapter 5

Feature Extraction

5.1   Introduction

This chapter describes different techniques used for the extraction of features from fine needle aspiration biopsy slides. In section 5.2 a reader can find a detailed definition of extracted features with discrimination between the BR factors followed by a description of the discriminatory power calculations and feature correlation.

5.2   Nuclear Feature Extraction

The features presented here were chosen to correspond to the indicators used for the Bloom–Richardson scheme. In this thesis two kinds of images were taken into consideration when extracting features for breast cancer malignancy grading. The magnifications of images correspond to the magnifications used during a cytological examination of a breast tissue. For the calculation of the low magnification features, images recorded with 100x magnification are used and for the calculation of high magnification features, images obtained with 400x magnification are used.

5.2.1   Low Magnification Features

In this study we calculated three features based on the cells' ability to form groups. These features are defined based on the number of groups and their area. Single cells, that are present in the images with high malignancy, are also represented as a group that consists of only one cell. If we take into consideration the description of Bloom and Richardson [19]

73

of the dispersion measure and the nature of the images taken during the FNA examination (see Fig. 5.1) we can see that groups with larger area are less malignant that those with smaller areas. Analogically, we can say that images with a larger number of groups are more malignant that those that contain only one or few groups (see Fig. 5.1c and d). Taking that into consideration we propose the following three features as a measure of
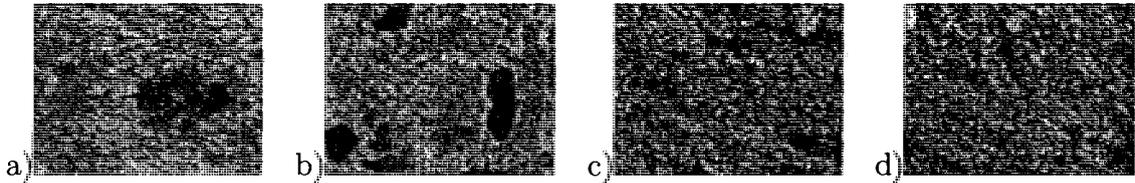

Figure 5.1: Illustration of cells groupings in low magnification images. a) and b) Intermediate malignancy, c) and d) High malignancy.

cancer malignancy:

- **Average area** (A_{100}) – is calculated as the average number of nuclei pixels. This feature represents the tendency of cells to form groups. If A_{100} is large then there is one or a few large groups in the image.

- **Number of groups** (NG) – To measure this feature we calculate the number of groups in the image that weren't removed during the segmentation process. Segmentation allows us to remove unwanted information in the image. In this case we remove the background information.

  If NG is large then there are numerous groups in the image, which suggests a high malignancy case.

- **Dispersion** (D) – We define the dispersion as a variation of cluster areas (A_c) which is determined by the following equation:

$$\frac{1}{D} = \frac{1}{n-1} \sum_{i=1}^{n} (A_c - A_{100})^2, \qquad (5.1)$$

where n is a number of cell clusters in the image and A_{100} is the average area.

Large values of this feature represent less dispersive cells and therefore a lower malignancy of the cancer.

According to the above description, we calculated features for all of the 34 low magnification images in the database. Calculated features were then used in conjunction with features extracted from images recorded with high magnification. In Table 5.1 some of the calculated low magnification features are collected. These features are illustrated on Fig. 5.2 and represent the most distinct cases from both malignancy classes. From the ta-

Feature	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Av. Area	173.9	122.5	266.1	149	4038.9	2216
No. of Groups	30	10	48	1	6	52
Dispersion	55.56	62.5	7.87	0	0.009	0.031
BR Grade	G3	G3	G3	G2	G2	G2

Table 5.1: Some of the calculated low magnification features along with pathologist grading.

ble we can see that the average areas of the groups for high malignancy cases have smaller values than for those with intermediate malignancy. We can notice the same pattern for number of groups which typically have larger values for more malignant images. Taking only number of groups into consideration would introduce discrepancy into our classification task because we can have more groups with large areas (i.e., Case 6) in case of intermediate malignancy. For this reason area and number of groups should be combined and used as a feature vector of two features. Here, we also introduced a third feature that should also be added to a feature vector thus improving our classification (see Table 6.2). From Table 5.1 we can see that dispersion values grow with increasing cancer malignancy.
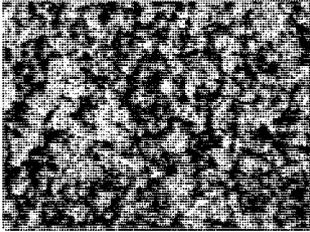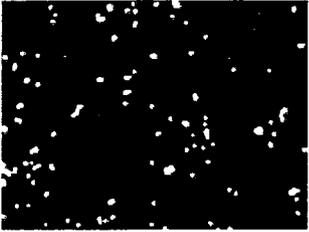
	Original	Groups	Grade
Case 1			G3
Case 2			G3
Case 3			G3
Case 4			G2
Case 5			G2
Case 6			G2
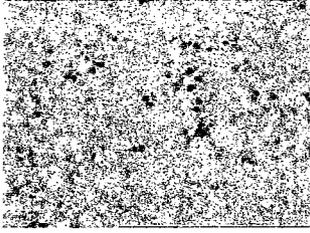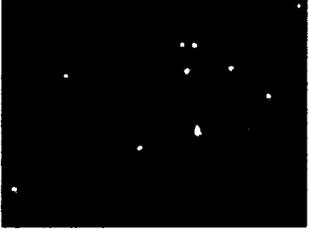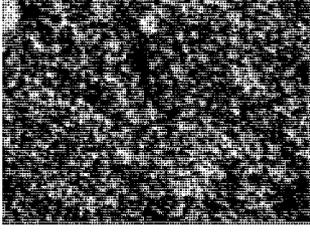
Figure 5.2: Cell groupings.

5.2.2 High Magnification Features

To precisely estimate the necessary features, only the nuclear features, rather than cellular features, are taken into consideration because only these features are the most discriminant and these are the features graded by pathologists. There are numerous features that can be extracted for this purpose. In this section, a set of features is described and their discriminatory powers are tested. The described features can be divided into five categories [195, 196].
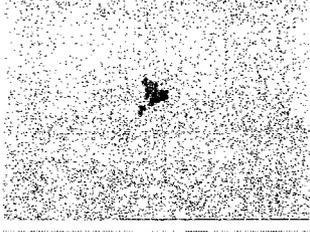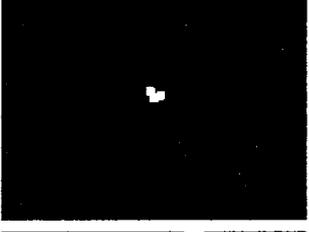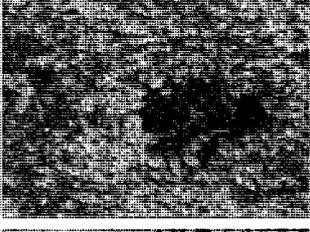
1. **Binary features** – features calculated based on the binary image $(\mathcal{I})$. A set of nuclei in the image, $N = \{N_1, N_2, ..., N_n\}$, can be defined as a collection of all connected components and the nucleus N_i is a set of pixels that are contained in the extracted nuclei. Using this definition, the following features can be extracted from $\mathcal{I}$:

   • **Area** (A_i) – calculated as the sum of all pixels (x, y) of the nucleus(N_i) [117, 183]:

   $$A_i = \sum_{(x,y) \in \mathcal{I}} \mathbb{I}_{N_i}(x, y). \tag{5.2}$$

   where $\mathbb{I}_{N_i}$ is an indicator function of set N_i.

   • **Perimeter** (p_i) – the length of the nuclear envelope. Calculated as length of the polygonal approximation of the boundary (B_i)[182, 183]:

   $$P_i = \sum_{(x,y) \in B_i} B_i(x, y). \tag{5.3}$$

   • **Convexity** (C_i) – calculated as the ratio of nucleus area and its convex hull [212], which is the area of the minimal convex polygon that contains the

nucleus(see Eq. 5.4).

$$C_i = \frac{A_i}{Area(CH(N_i))}.$$

(5.4)

where $CH(N_i)$ is a convex hull of the ith nucleus.

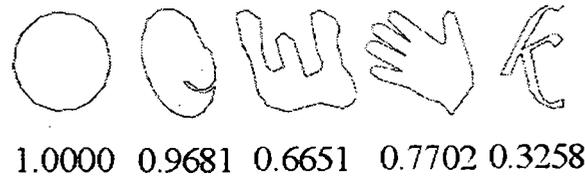Convex shapes will yield a value of 1, while concave shapes will have a value less than 1 (see Fig. 5.3).

1.0000 0.9681 0.6651 0.7702 0.3258

Figure 5.3: Graphic representation of convexity, taken from [212]

- **Eccentricity** (Ecc_i) – allows us to track how much a segmented nucleus differs from a healthy nucleus [117]. Healthy nuclei will assume circular shapes while cancerous nuclei can assume arbitrary shapes. We calculate eccentricity as the ratio of the distance between the foci of an ellipse fitted to the nucleus that has the same second–moments as the extracted nuclei, and its major axis length. Values of this feature vary between 0 and 1. These are degenerate cases because a shape whose eccentricity is 0 is actually a circle, while a shape whose eccentricity is 1 is a line segment (see Fig. 5.4).
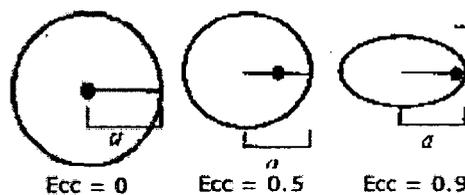

Figure 5.4: Graphic representation of eccentricity.

- **Centroid** (Ctr_i) – finds a center point of a nucleus along each row (X) and column (Y). It is also often called a center of mass of the object. The

centroid $(\overline{x}_i, \overline{y}_i)$ can be calculated as [195]:

$$\overline{x}_i = \frac{1}{A_i} \sum_{j=0}^{X-1} \sum_{k=0}^{Y-1} jN_i(j,k); \qquad \overline{y}_i = \frac{1}{A_i} \sum_{j=0}^{X-1} \sum_{k=0}^{Y-1} kN_i(j,k). \qquad (5.5)$$

- **Orientation** (Or_i) – this feature is also called an axis of least second moment and provides us with the information about the orientation of the nuclei. It is considered as an axis of least inertia. When the coordinate system is placed at $(\overline{x}_i, \overline{y}_i)$ then the orientation Or_i can be defined as [195]:

$$Or_i = tan(2\theta_i)$$

$$ (5.6) $$

$$tan(2\theta_i) = 2\frac{\sum_{j=0}^{X-1} \sum_{k=0}^{Y-1} jkN_i(j,k)}{\sum_{j=0}^{X-1} \sum_{k=0}^{Y-1} j^2 N_i(j,k) - \sum_{j=0}^{X-1} \sum_{k=0}^{Y-1} k^2 N_i(j,k)}.$$

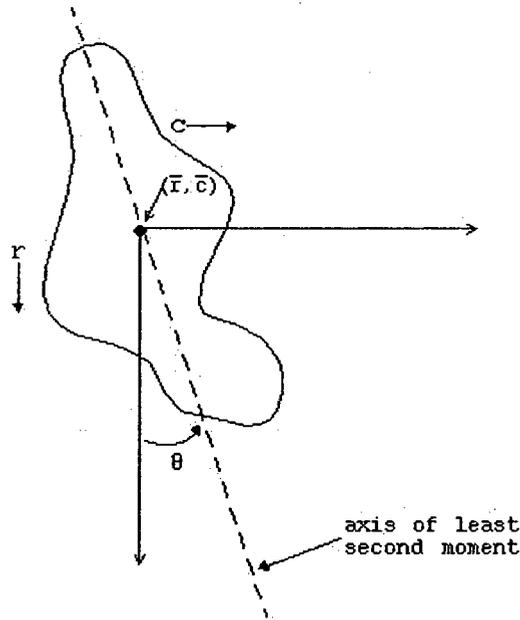The angle θ_i is measured counterclockwise from the x–axis (see Fig. 5.5).

Figure 5.5: Graphic representation of the orientation feature.

- **Projection** (Prj_i) – this feature is calculated as a sum of all pixels along rows and columns of the nucleus image [117, 195]. Summation of all the rows provides us with a horizontal projection $h_i(x)$ and summation of all the

79

columns determines the vertical projection $v_i(y)$ (see Fig. 5.6):

$$h_i(x) = \sum_{k=0}^{Y-1} N_i(x,k); \quad v_i(y) = \sum_{j=0}^{X-1} N_i(j,y) \tag{5.7}$$

and

$$Prj_i = (h_i, v_i). \tag{5.8}$$

									h(x)
0	0	0	0	1	0	0	0	0	1
0	0	0	0	1	1	0	0	0	2
0	0	1	1	1	1	1	0	0	5
0	1	1	1	1	1	1	0	0	6
0	0	0	0	0	1	1	0	0	2
0	0	0	0	0	0	0	0	0	0

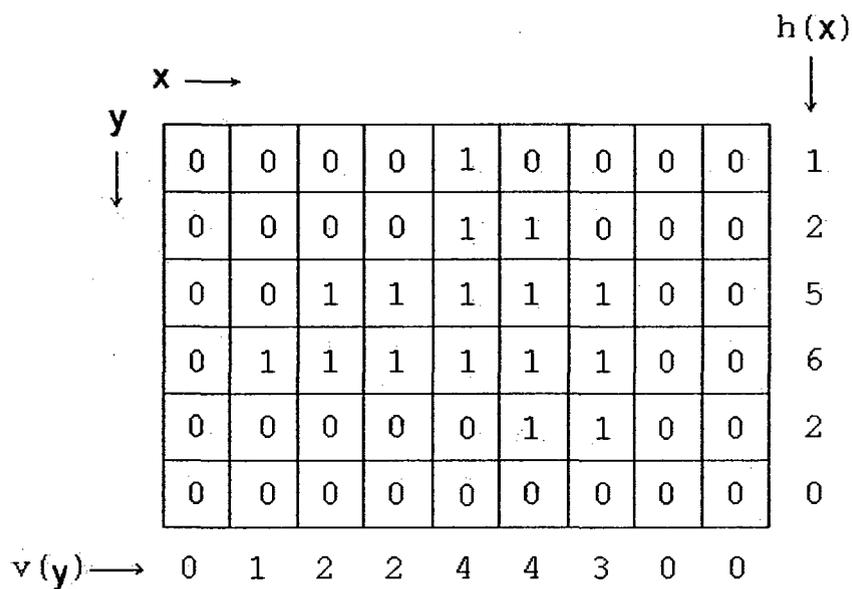x →   y ↓   h(x) ↓   v(y) →   0 1 2 2 4 4 3 0 0

Figure 5.6: Illustration of the projection feature.

2. **Momentum based features** – moments are generally used for the extraction of features that are rotation, scaling and translation (RST) invariant. Based on the normalized central moments, η_{ij}, the following seven momentum–based features

are calculated [196]:

$$\varphi_1 = \eta_{20} + \eta_{02} \tag{5.9}$$

$$\varphi_2 = (\eta_{20} - \eta_{02}^2) + 4\eta_{11}^2 \tag{5.10}$$

$$\varphi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{5.11}$$

$$\varphi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{5.12}$$

$$\varphi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2]$$
$$+ (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \tag{5.13}$$

$$\varphi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$
$$+ 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \tag{5.14}$$

$$\varphi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2]$$
$$- (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \tag{5.15}$$

3. **Histogram based features** – the image histogram describes the occurrence frequency of intensity values in the image (see §3.4.1). Features based on the histogram are considered to be statistical features and the histogram is considered to be a probability distribution function of grey level values in the image [195]. These features allow for the determination of the characteristics of that distribution for the nuclei image taken into consideration. The histogram $P(g)$ can be defined as a probability function of the form:

$$P(g) = \frac{N(g)}{M}, \tag{5.16}$$

where $M = X * Y$ is the number of pixels in the image and $N(g)$ is the number of pixels at gray level g.

Based on this definition five histogram based features are calculated. These features reflect the parameters of the histogram described in section 3.4.1. The five features are the mean, standard deviation, skew, energy and entropy. Fig. 5.7 illustrates the behavior of these features.

81
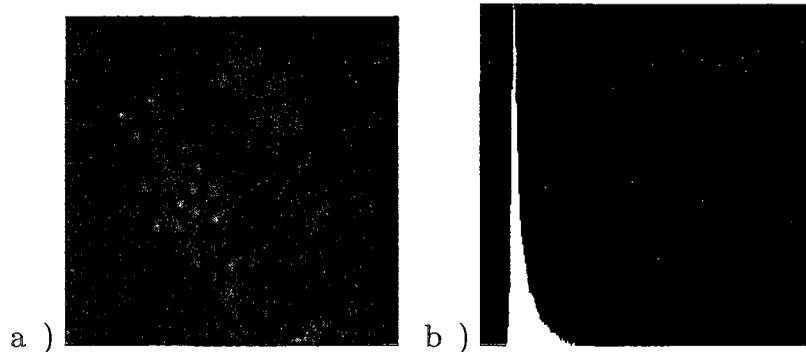
Mean	Std. Dev.	Skew	Energy	Entropy
37	35	5.3	0.050	4.94
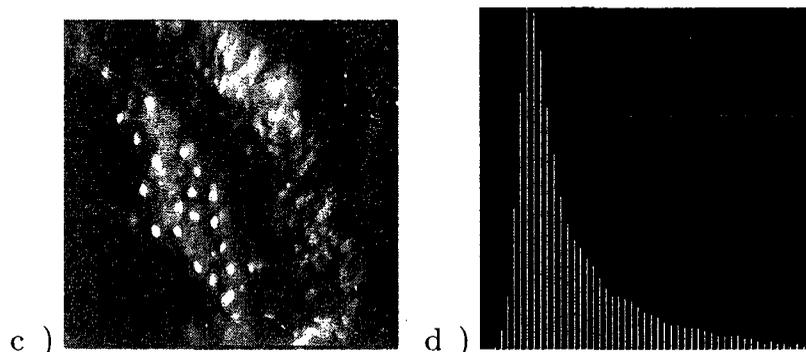
Mean	Std. Dev.	Skew	Energy	Entropy
75	56	1.7	0.051	4.76

Figure 5.7: Illustration of histogram based features, taken from [195].

4. **Textural features** – textural features are used to measure the texture information of the image [75, 188, 195]. Here, the texture of the nucleus is taken into consideration. To extract textural features, a co-occurrence matrix is calculated (see section 4.2.5), which provides us with information about the relation of pairs of pixels and their corresponding grey levels. If we assume that the distance between the pixels and the directions are given, then we can extract five GLCM based features. In section 4.2.5, energy (eq. 4.24), entropy (eq. 4.21) and inertia (eq. 4.23) are defined. Additionally to the textural features described in §4.2.5

we can extract the following two features:

$$Inverse\ Difference\ =\ \sum_i \sum_j \frac{c_{ij}}{|i-j|};\ \ for\ i \neq j, \qquad (5.17)$$

$$Correlation\ =\ \frac{1}{\sigma_x \sigma_y} \sum_i \sum_j (i - \mu_x)(j - \mu_y)c_{ij} \qquad (5.18)$$

where

$$\mu_x = \sum_i i \sum_j c_{ij},$$

$$\mu_y = \sum_j j \sum_i c_{ij},$$

$$\sigma_x = \sqrt{\sum_i (1 - \mu_x)^2 \sum_j c_{ij}},$$

$$\sigma_y = \sqrt{\sum_j (1 - \mu_y)^2 \sum_i c_{ij}}.$$

A sixth textural feature that can be introduced is calculated as an average grey level (eq. 5.19), as proposed by Street $et\ al.$ [184].

$$r_i = \frac{1}{k} \sum_{1=0}^{k} \overline{r_i}, \qquad (5.19)$$

where k is the number of nuclei in the image and r_i is define by eq. 5.20

$$\overline{r_i} = \frac{1}{n} \sum_{x,y \in N_i} I_{r_i}(x, y) \qquad (5.20)$$

5. **Color based features** – Color images consist of three components each representing a primary color. Each of these components can be treated as a separate intensity image. To calculate color features we can use any of the previously defined features and apply it to each color band. This requires three times as many calculations as for the grey level image. To overcome this problem, the spherical coordinate transform (SCT) can be applied to the RGB image which takes into consideration the relation between the RGB channels. More precisely, it transforms the image information of RGB channels into a 2–D color

space with brightness L (see Fig. 5.8). The transformation from RGB values into the SCT components is given by the following set of equations [195]:

$$L = \sqrt{R^2 + G^2 + B^2},$$
$$\angle A = \cos^{-1}\frac{B}{L},$$
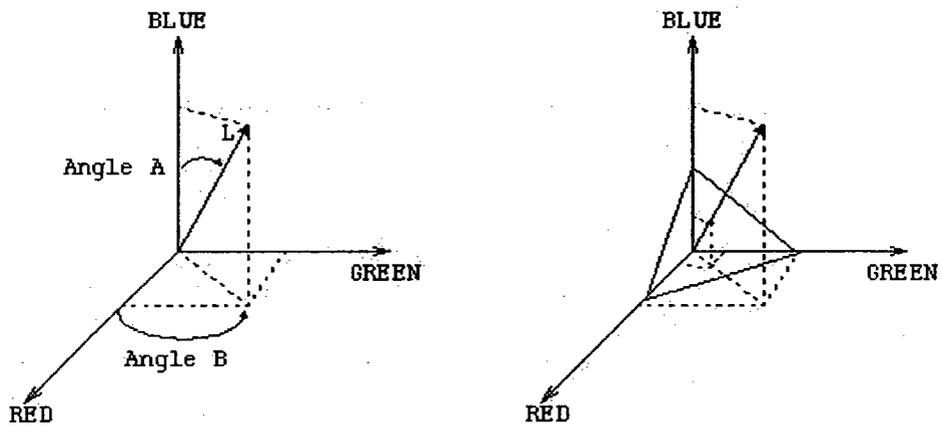$$\angle B = \cos^{-1}\frac{R}{L\sin\angle A}. \tag{5.21}$$

Figure 5.8: Illustration of the spherical coordinate transform, taken from [195].

The features defined in this section can also be divided into two categories depending on the type of the polymorphy that is being described. The first category describes nuclei changes in shape and size (see Fig. 5.9 a) and b)) and the second describes the changes in color or textures, which is equivalent to staining intensity changes (see Fig. 5.9 c)). To describe size changes the binary, momentum and spectral features can be used while for the description of staining changes the histogram, textural and color–based features are more suitable.
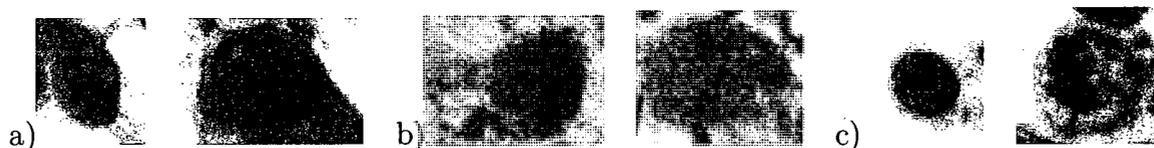

Figure 5.9: Illustration of cells polymorphy. a) and b) Intermediate malignancy, c) and d) High malignancy.

Tables 5.2 and 5.1 present some of the calculated features for both types of images and their pre–assigned grade given by an expert pathologist. From the tables we can notice that the features confirm the biological cells features that are characteristic for more and less cancerous cases. This can also be noticed from Fig. 5.10 which shows the cases presented in Table 5.2.

Based on the literature review, in this study, we extracted 34 features describing the nuclei in the image. If we compare the results from Table 5.2 we can notice the correlation of some of these features with the malignancy of the cancer. Analogously to the 100x features, the tendency of some of feature values reflects the malignancy stages of the cancer. It can be easily noticed that the areas of nuclei in the images are larger for high malignancy that for low malignancy cases. Since perimeter is closely related to area, and for large areas perimeter will also be large, the same reasoning can be applied. Beside the shape–related features we also extracted features that represent nuclear circularity and we calculate textural features for gray levels statistics and for the image red channel alone. Cancerous nuclei assume shapes that are irregular and not necessarily circular.

Feature	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Area	198	429	414	429	1177	256
Perimeter	51	70	74	94	137	60
Eccentricity	0.6510	0.7144	0.6746	0.7283	0.5884	0.6650
Convexity	0.9534	0.9499	0.9400	0.8643	0.9104	0.9377
x–centroid	170.4	197.3	175.4	161.6	181.2	190.4
y–centroid	133.7	119.4	156.7	154.9	121.3	130.7
Orientation	8.77	5.42	0.62	-13.92	1.55	-5.93
ver. projection	10.29	15.17	15.21	12.85	22.34	11.41
hor. projection	11.44	15.87	15.21	13.74	26.57	11.05
φ_1	0.7459	1.0655	1.5626	1.3458	1.1824	0.9293
φ_2	0.3363	0.8824	1.2264	1.0181	0.9855	0.6444
φ_3	0.1862	1.0395	0.7429	0.7311	1.0234	0.6033
φ_4	0.4785	1.5412	3.4390	2.6076	1.8912	1.0022
φ_5	0.0726	0.6433	-0.2439	0.2152	2.0803	0.1548
φ_6	0.2774	1.4460	3.8077	2.6270	1.8762	0.8042
φ_7	0.0349	2.4644	4.1495	2.1059	-0.0126	0.9827
hist. mean	5.54	3.41	10.35	5.61	15.90	16.05
hist. Std.Dev.	21.67	20.42	26.92	24.77	38.59	38.53
hist. skew	3.79	5.90	2.39	4.38	2.12	2.07
hist. energy	0.8773	0.9458	0.7462	0.9006	0.7213	0.7187
hist. width	881783	1577365	946988	1971352	3200727	3029530
text. energy	0.8566	0.9395	0.7298	0.8854	0.7163	0.6931
text. inertia	1608	438	2309	1124	1703	3355
text. correlation	0.7926	0.8747	0.8495	0.8230	0.8986	0.7999
text. homogeneity	0.9753	0.9933	0.9646	0.9828	0.9739	0.9486
L hist. mean	0.0311	0.0107	0.0557	0.0226	0.0505	0.0627
L hist. Std.Dev.	2.19	1.46	2.77	1.87	2.85	3.32
L hist. skew	73.4	138.3	52.8	86.8	58.6	54.1
L hist. energy	0.8773	0.9458	0.7459	0.9006	0.7210	0.7185
L hist. width	32252	21847	42563	23687	58257	90321
Avg. gray level	87.4	124.2	75.9	110.0	105.4	105.3
BR Grade	G3	G3	G3	G2	G2	G2

Table 5.2: Some of the calculated high magnification features along with pathologist grading for selected cases.

More malignant cases usually assume cylindrical shapes while less malignant will be more circular. When comparing the obtained results in Table 5.2 it can be noticed that these tendencies are preserved.

5.3 Feature Validation

5.3.1 Kolmogorov–Smirnov statistics

The features described in the previous section (§5.2) were used for training and testing classifiers (see §6). To estimate the classification performance and the power of extracted features their discriminatory powers were calculated according to the procedure shown in this section. For the purpose of this study two discriminatory power measures based on Kolmogorov–Smirnov statistics were used.

Kolmogorov–Smirnov (KS) measure is a statistical measure where features are presented as a distribution related to a data [42]. Originally, Kolmogorov and Smirnov introduced a statistical procedure which uses the maximum vertical distance between two distributions to check how well the distributions match each other [42].

The idea of the Kolmogorov measure is to test the goodness of the fit of a random sample from an unknown distribution to check the hypothesis that this distribution is actually a specified function. Fig. 5.11 illustrates this algorithm. From the figure we can see that if the Kolmogorov test value is small then our data distribution can be assumed to represent accurately a specified distribution. This description of the Kolmogorov statistics, also called the one–sample Kolmogorov–Smirnov test, is a valuable test but it is assumed that the hypothesized distribution is fully specified. The original description of Kolmogorov goodness of fit has numerous modifications. These modifications allow the use of this test when distribution parameters are estimated from experimental data.

Here, the Smirnov modification of the Kolmogorov measure is described. It is also called a two–sample Kolmogorov–Smirnov (KS) test due to the fact that it tests two samples from different data distributions [42]. The idea of the test is the same as the one–sample test. If we consider two independent samples, first of a size n, $X_1, X_2, ..., X_n$, and the

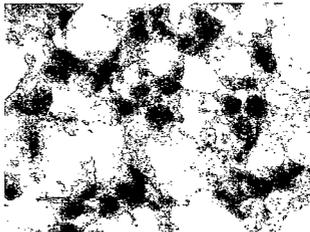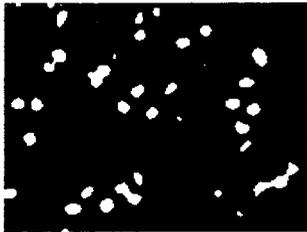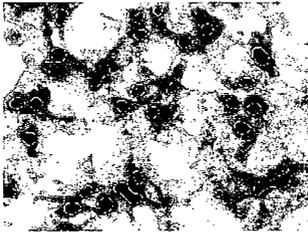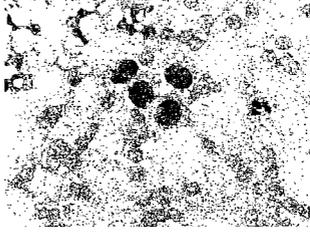|	Original	Segmented nuclei	Nuclear contours	Grade
Case 1				G3
Case 2				G3
Case 3				G3
Case 4				G2
Case 5				G2
Case 6				G2
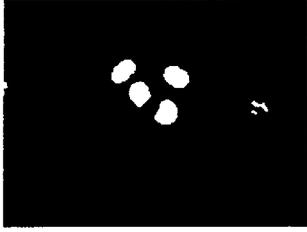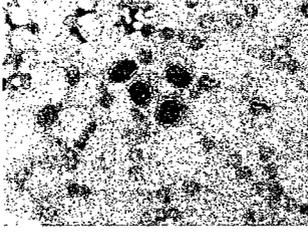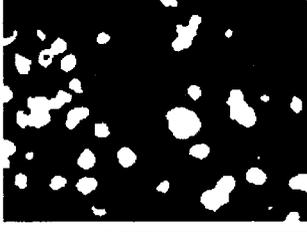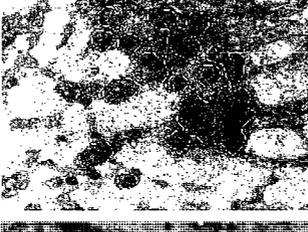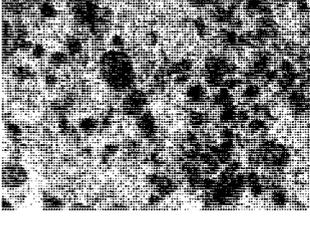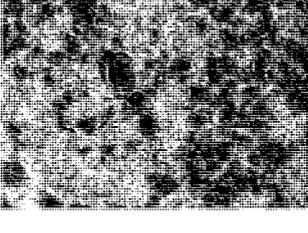
Figure 5.10: Segmented nuclei and nuclear contours.

second of a size m, $Y_1, Y_2, ..., Y_n$ then we denote $F(x)$ and $G(x)$ respectively to be their unknown distributions. To perform the KS test we take into consideration the following
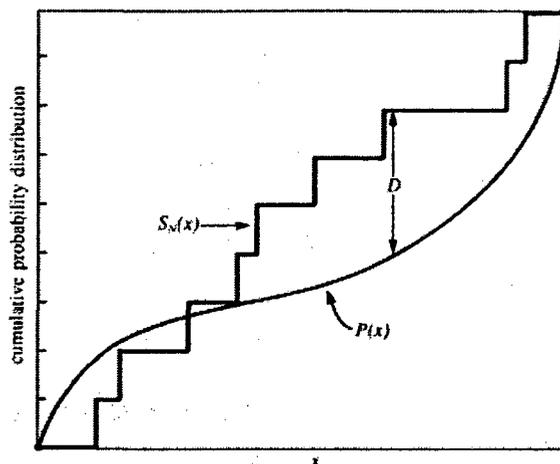
Figure 5.11: Graphical representation of the Kolmogorov statistics, taken from [42].

assumptions:

1. Data consists of random samples.

2. Two samples are mutually independent.

3. The measurement scale is at least ordinal.

4. Variables are assumed to be continuous.

and the following hypotheses are tested:

$$H_0 : F(x) = G(x) \qquad for\ all\ x\ from\ -\infty\ to\ +\infty$$

$$H_1 : F(x) \neq G(x) \qquad for\ at\ least\ one\ value\ of\ x$$

If we now denote $S_1(x)$ and $S_2(x)$ as empirical distributions based on the X and Y samples respectively then we can define the test statistics T as the greatest distance between $S_1(x)$

and $S_2(x)$, see eq. 5.22. We reject a hypothesis H_0 at the level of significance α if $T > 1-\alpha$. If the hypothesis is rejected, the function returns 0 and 1 otherwise.

$$T = \sup_x |S_1(x) - S_2(x)| \qquad (5.22)$$

Fig. 5.12 shows a graphical representation of the KS test. From the figure we can see that if the distance between two empirical distributions is large then we can assume that the data point from these distributions are drawn from two different populations.
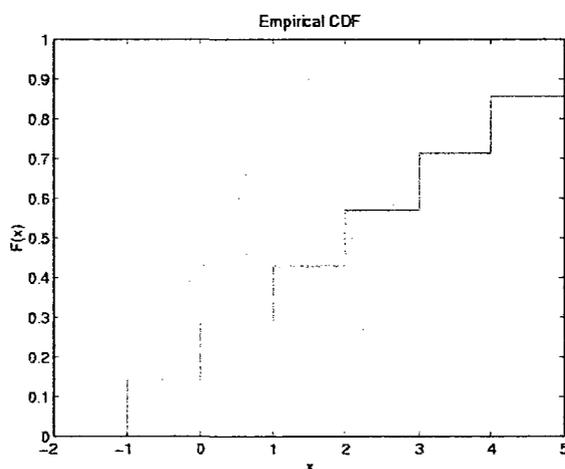
Figure 5.12: Graphical representation of the Kolmogorov–Smirnov test.

5.3.2 Feature discriminatory power

For all of the features, the discriminatory powers were calculated to show which of the presented features are the most distinctive in terms of malignancy classification and the results are presented in Table 5.3. As mentioned in section 5.3.1 KS measure tests if the data comes from two different distributions. For this test we present our features as distributions. To check the discriminatory power of our features we calculate the maximum distance between distributions of high and intermediate malignancy classes. Fig. 5.13

presents plots of such distributions from which feature discriminatory powers were calculated. for comparison purposes three different variations of the Kolmogorov–Smirnov tests were performed. The first test shows if the samples are drawn from the same distribution. If the samples come from two different distributions then the KS Test returns a value 0. The remaining two variations show us how different the samples are. The assumption is that the farther the samples are from each other, the higher is the probability of coming from two different distributions is. This tendency is represented in the Table 5.3 by AKS and SKS Tests.

From Fig. 5.13 and Table 5.3 we can see that 100x features offer better separation between the malignancy classes than 400x features. Although structural features present better separation than shape features, the pleomorphic features provide better classification than 100x features when only that subset of features is taken into consideration. This can be seen from the classification results presented in the next section.

5.3.3 Feature Correlation

With such a large number of features, questions may arise about the performance of the classifier. It can be assumed that some of the features are related to one another which causes the introduction of erroneous information to the training set. To minimize such a correlation, the degree of statistical dependence between features was calculated. Such a dependence is determined by calculation of the features' covariance. Covariance σ_{xy} is a statistical measure of the deviation between variables x and y and it can be defined as:

$$\sigma_{xy} = \varepsilon[(x - \overline{x})(y - \overline{y})] = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} (x - \overline{x})(y - \overline{y}) P(x, y), \qquad (5.23)$$

where $P(x, y)$ is a joint probability mass function.

Analogically, we can define a covariance matrix Σ which contains covariances σ_{ij} at ij^{th} element of x_i and y_j according to the following formula:

$$\sigma_{ij} = \sigma_{ji} = \varepsilon[(x_i - \overline{x}_i)(y_j - \overline{y}_j)] \quad i, j = 1, 2, ..., n. \qquad (5.24)$$
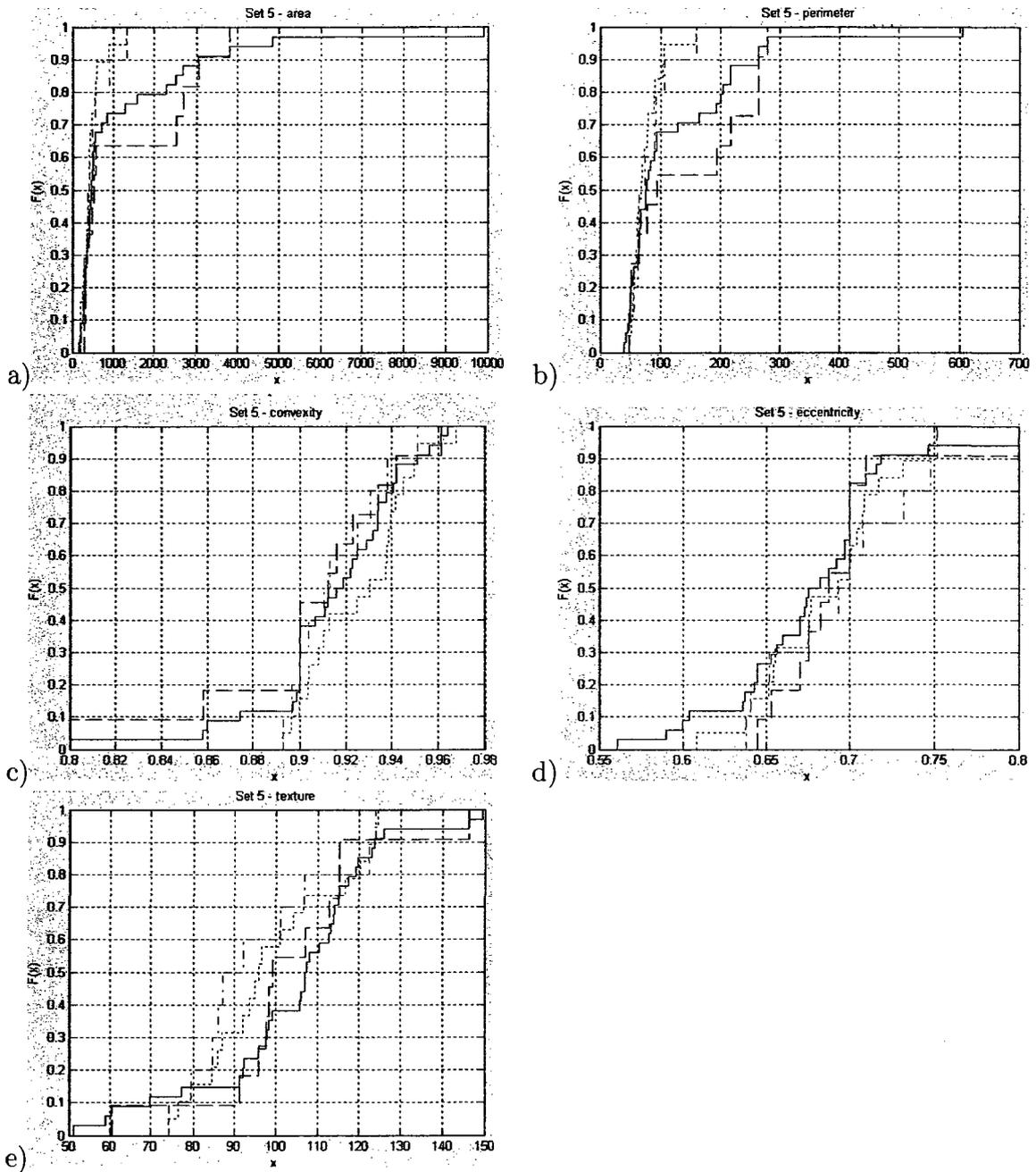
Figure 5.13: Empirical distributions of the features.

Features	KS Test			AKS Test			SKS Test		
	FCM	GLQ	LS	FCM	GLQ	LS	FCM	GLQ	LS
$Area_{100}$	0	0	0	20.356	20.356	20.356	0.429	0.429	0.429
No. of Groups	1	1	1	8.432	8.432	8.432	0.084	0.084	0.084
Dispersion	0	0	0	22.409	22.409	22.409	0.494	0.494	0.494
Area	1	1	1	4.324	8.671	9.032	0.009	0.076	0.088
Perimeter	1	1	1	7.326	7.074	7.841	0.069	0.063	0.076
Eccentricity	0	1	1	9.891	3.618	5.065	0.114	0.016	0.016
Convexity	0	1	1	7.321	6.003	4.915	0.078	0.049	0.032
x-centroid	1	1	1	6.635	3.962	3.376	0.055	0.022	0.011
y-centroid	1	1	1	8.426	6.312	2.679	0.091	0.044	0.011
Orientation	1	1	1	8.212	7.118	6.376	0.086	0.063	0.038
ver. projection	0	1	1	12.468	9.291	3.206	0.151	0.089	0.009
hor. projection	0	1	1	10.959	8.476	4.062	0.129	0.077	0.021
φ_1	1	1	1	6.556	4.394	7.053	0.046	0.017	0.057
φ_2	1	1	1	5.915	4.471	7.465	0.040	0.018	0.060
φ_3	1	1	1	5.438	3.965	7.462	0.030	0.013	0.066
φ_4	1	1	1	4.247	3.926	6.241	0.030	0.010	0.046
φ_5	1	1	1	5.485	4.071	8.338	0.035	0.005	0.074
φ_6	1	1	1	5.359	3.768	6.321	0.032	0.010	0.043
φ_7	1	1	1	1.935	2.974	3.606	0.004	0.005	0.010
hist. mean	1	1	0	5.274	5.365	11.912	0.031	0.039	0.152
hist. Std.Dev.	1	1	0	4.853	5.147	11.276	0.032	0.034	0.137
hist. skew	1	1	0	4.065	4.403	10.659	0.021	0.028	0.122
hist. energy	1	1	1	4.315	4.426	10.006	0.018	0.028	0.098
hist. width	1	1	1	4.200	6.153	8.424	0.023	0.046	0.076
text. energy	1	1	1	4.009	4.859	9.529	0.019	0.039	0.080
text. inertia	1	0	1	4.788	8.291	6.750	0.029	0.095	0.054
text. correlation	1	0	1	3.138	9.165	7.024	0.009	0.096	0.058
text. homogeneity	1	0	1	4.788	8.406	6.750	0.029	0.097	0.054
L hist. mean	1	1	1	5.732	5.585	8.418	0.044	0.033	0.079
L hist. Std. Dev.	1	1	1	5.421	6.053	9.688	0.040	0.042	0.095
L hist. skew	1	1	1	5.532	4.406	7.544	0.035	0.020	0.060
L hist. energy	1	1	1	4.315	4.426	10.006	0.018	0.028	0.098
L hist. width	1	1	0	4.597	4.635	9.041	0.038	0.023	0.099
Avg. Gray level	1	1	1	2.715	4.903	5.971	0.008	0.027	0.034

Table 5.3: Features discriminatory powers for $\alpha = 0.05$.

The covariance matrix is a square, symmetric matrix where diagonal elements represent the variance of x and off–diagonal elements are the covariance values. The following equation illustrates the Σ matrix:

$$\Sigma = \begin{bmatrix} \varepsilon[(x_1 - \overline{x}_1)(x_1 - \overline{x}_1)] & \varepsilon[(x_1 - \overline{x}_1)(x_2 - \overline{x}_2)] & \dots & \varepsilon[(x_1 - \overline{x}_1)(x_n - \overline{x}_n)] \\ \varepsilon[(x_2 - \overline{x}_2)(x_1 - \overline{x}_1)] & \varepsilon[(x_2 - \overline{x}_2)(x_2 - \overline{x}_2)] & \dots & \varepsilon[(x_2 - \overline{x}_2)(x_n - \overline{x}_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon[(x_n - \overline{x}_n)(x_1 - \overline{x}_1)] & \varepsilon[(x_n - \overline{x}_n)(x_2 - \overline{x}_2)] & \dots & \varepsilon[(x_n - \overline{x}_n)(x_n - \overline{x}_n)] \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \dots \\ \sigma_{d1} & \sigma_{d2} & \dots & \sigma_{dd} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1d} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \dots \\ \sigma_{d1} & \sigma_{d2} & \dots & \sigma_d^2 \end{bmatrix} \qquad (5.25)$$

The values obtained by the determination of the covariance tell us if x and y are closely related or not. Covariance will assume large values if they are similar and small for those not closely related.

Here, the correlation coefficients where calculated to visualize dependencies between extracted features. Correlation coefficient is a normalized covariance defined as:

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}. \qquad (5.26)$$

The coefficient values must be between -1 and $+1$ meaning maximal negative and positive correlation respectively.

Fig. 5.14 visualizes the correlation between features. In the figure white represents high correlation with a coefficient value equal to 1. Correlation coefficient assumes a value of -1 for not correlated features which are represented as black in Fig. 5.14.

In the figure, two areas of the least correlation can be noticed. These areas include textural and luminescence features, which are the least correlated amongst all features described here. The shape based features show some correlation to other features. In the figure we can also notice that the highest correlation was recorded for the same type of features such as momentum features which have the highest correlation of all. Similar behavior can be noticed for shape based features. If we take the 100x magnification features into

94

consideration, we can see that these features show moderate correlation both to the 400x features and between themselves.

5.4 Conclusions

In this chapter a set of extracted features were presented. There are 34 features calculated based on the segmentation results from chapter 3. From this chapter we can notice that there are two types of features that have to be calculated. The type of the feature depends in the image magnification. It can easily be noticed that low magnification and high magnification features will be able to discriminate between malignancy cases by themselves but the combination of both should provide best results. When compared against classification results by a pathologist, one can see that this is true.

The results presented in this chapter also show that the collection of features proposed for this study have good discriminatory powers. The correlation map shows interactions between features and allows for elimination of those features that closely correlated with other features. We can then replace these features with only one. This procedure allows for reduction of a size of the feature vector that should reduce the complexity of the classification algorithms. This reasoning was taken into consideration in section 6.11 were the results on the smaller data sets are presented.

a) Correlation Map, Variables in Original Order

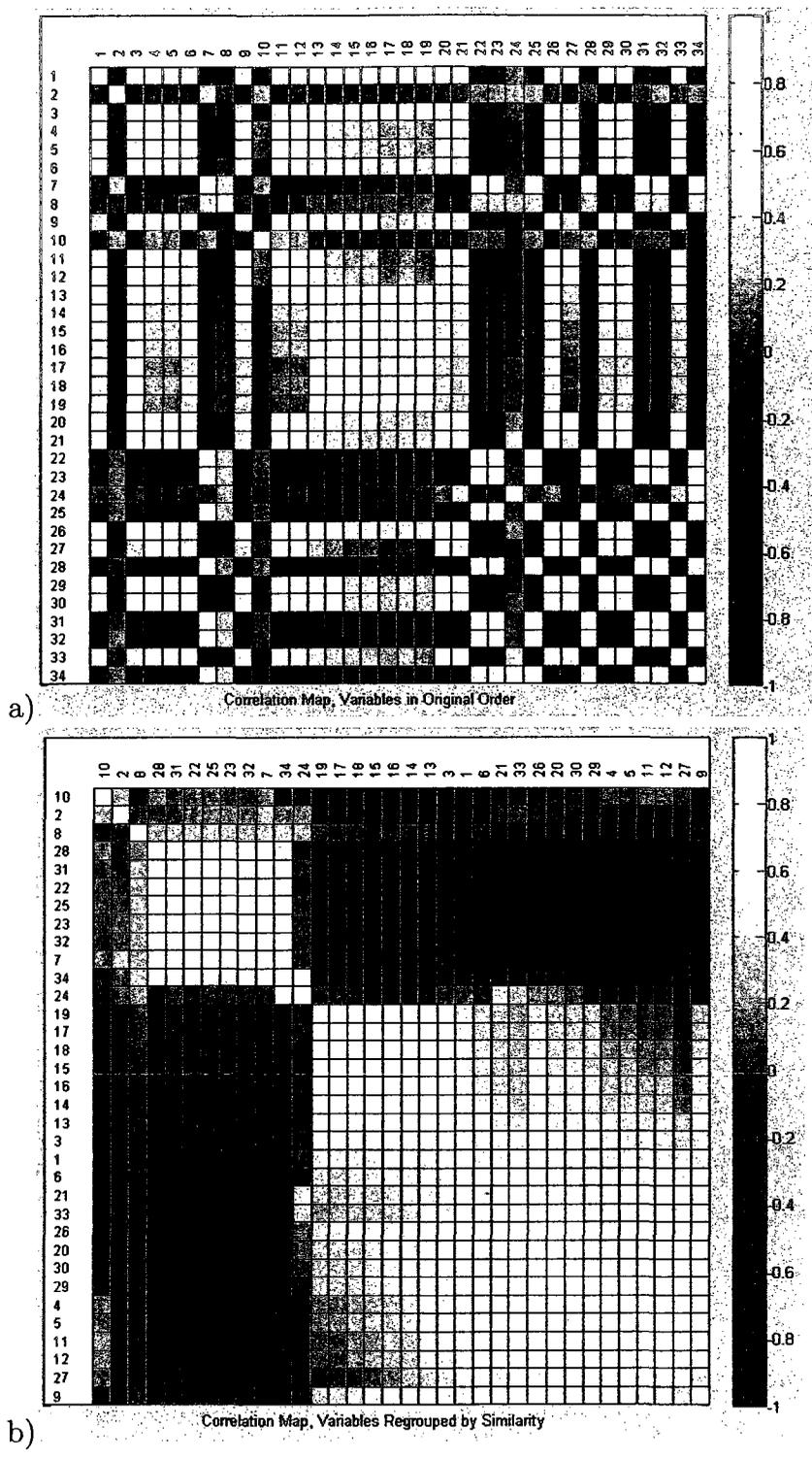b) Correlation Map, Variables Regrouped by Similarity

Figure 5.14: Graphical representation of feature correlation coefficients. a) Ordered by features, b) Grouped according to the correlation.

Chapter 6

Malignancy Classification

6.1 Introduction

This chapter describes the malignancy classification process and gives a description of the classifiers tested during this research. Pattern classification deals with the discrimination between classes of patterns. The process of discrimination is not guaranteed to work for all patterns. Each classification method is expected to misclassify some of the input patterns. This property leads to for the definition of an error rate that is a measure of each known classification algorithm. The lower the error the better classification performance of our system. Analogically we can define a classification rate that will tell us how well the system classifies the unknown patterns. Ideally this rate should be 100%, therefore the closer we get to that ideal rate the better that classification performed by the system. To classify an FNA cytological tissue we use classifiers that take a feature vector as an input and respond with a two element output vector $(1,0)^T$ for intermediate malignancy and $(0,1)^T$ for high malignancy. In this chapter, the classification methods will be presented that will be studied for their ability to classify malignancies.

In this chapter, the reader can find detailed information about classifiers used for breast cancer malignancy grading. A description of training and testing sets used in this study is also provided. Section 6.5 introduces neural networks in general and is followed by a detailed presentation of multilayer perceptron and principal component based neural networks. The description provides a reader with additional information on the limitations

and initial conditions of the classifiers used. The remaining sections of this chapter present the other classifiers used in this study for comparison purposes. We tested support vector machines, self–organizing maps, k–nearest neighbor rule and tree classifiers. In this chapter, a reader can also find a description of the cross–validation technique that was applied in this thesis as well as a description of the collected database. At the end of this chapter, a summary of the presented results, and some conclusive insights about their impact and usefulness for breast cancer malignancy grading is presented.

6.2 Naive Bayes Classifier

Pattern classification deals with the problem of assigning a certain class to the pattern in question. There are numerous techniques that can be applied for data classification. The most popular and fundamental algorithm is a classifier based on the Bayesian decision theory [55, 82, 141]. This is a statistical approach that takes into consideration a probability and costs that are associated with the classification of the data to the certain class. This rule can be applied under the assumption that all of the relevant probabilities have to be known and attributes of a feature vector $x = [x_1, x_2, .., x_{D_f}]^T$ are statistically independent [55, 90] and therefore the *posterior* probability for consecutive attributes can be described with the following equation [178]:

$$p(X = x|\omega = i) = \prod_{j=1}^{D_f} p(X_j = x_j|\omega = i),$$
(6.1)

where X and ω are the random variables.

The classification problem for c number of classes is based on defining a set of discrimination boundaries $g_i(x)$, $i = 1, ..., c$ and classifying x to class ω_i if

$$g_i(x) > g_j(x) \qquad \text{for all } j \neq i$$
(6.2)

It can be noticed that Bayes classifier needs to calculate c discriminant functions and the selected pattern corresponds to the largest discriminant [52, 55, 78, 155, 192].

If we take into consideration a continuous univariate normal or Gaussian distribution

(Eq. 6.3), then we can have a family of distributions for all attributes and classes.

$$p(X_j = x_j | I = i) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp\left(-\frac{(x_j - \mu_{ij})^2}{2\sigma_{ij}}\right), \qquad (6.3)$$

where i is the class number, j is the feature vector attribute, the expected value (mean) of x can be described as:

$$\mu_{ij} = \frac{1}{N_i} \sum_{k \in S_i} x_j^{(k)}, \qquad (6.4)$$

and the variance (expected squared deviation) is:

$$\sigma_{ij}^2 = \frac{1}{N_i} \sum_{k \in S_i} (x_j^{(k)} - \mu_{ij})^2. \qquad (6.5)$$

It can be noticed that the Gaussian normal density is fully specified by the mean μ_{ij} and variance σ_{ij}. This fact can be then used for classifier training purposes which relies on the estimation of the unknown mean and variance according to the Eqs. 6.4 and 6.5 with the constraint that for $i - th$ class, the summation is performed only on the training set elements that belong to the decision region of the class. Now, it is possible to rewrite the $i - th$ decision boundary $g_i(x)$, which will yield Eq. 6.6 after elimination of the constant parameters.

$$g_i(x) = N_i \prod_{j=1}^{D_f} \frac{1}{\sigma_{ij}} \exp\left(-\frac{(x_j - \mu_{ij})^2}{2\sigma_{ij}}\right) \qquad (6.6)$$

The Bayes rule for classification is a popular method used in different aspects of pattern recognition. We can find some approaches to image segmentation based on Bayes classifier [210] as well as in medical applications [151].

6.3 K–Nearest Neighbor Rule

K–nearest neighbor (KNN) is one of the simplest classification algorithms. It is based on the distance calculation between the pattern in question and its k neighbors. The decision is made based on the closest association between the pattern and the neighbors. The pattern is classified to the closest class in terms of a distance amongst its k neighbors.

The training procedure is very simple and is based on recording the entire training set. Testing usually uses a Euclidean distance for calculation of distances between training samples and the tested sample. The class assigned to the sample is the one for which the distance is smallest. To be able to calculate the Euclidean norm it is usually necessary to normalize the data to avoid any data inconsistency.

The literature review showed that the KNN rule is a popular method for classification [69, 71, 111, 178]. We can find different application areas such as text and character recognition [11, 71] and numerous medical applications [20, 69, 111, 114].

6.4 Decision Trees

Decision trees belong to the group of more complicated classification methods than the rest of the algorithms presented in this study [22, 60, 95]. Most of the traditional pattern recognition algorithms are based on the feature vectors that are real–valued and some kind of metric can be applied to them [55]. Tree classifiers on the other hand are able to solve classification problems that involve nominal data such as a list of attributes like fruit colors and sizes.

Decision trees are constructed in a way where the classes are held in the leaves of the tree and the decision rules are kept in the interior nodes including the root [18, 165, 178]. Fig. 6.1 shows an example of a typical decision tree.
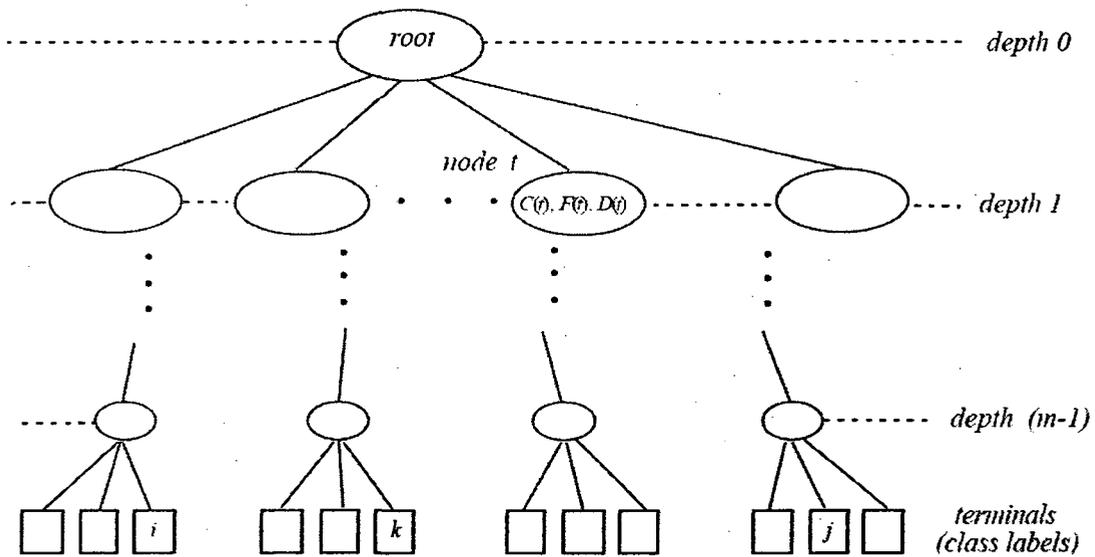
Figure 6.1: Example of a general decision tree, where C(t) is a subset of classes accessible from node t. F(t) is a feature subset used at node t and D(t) is a decision rule used at node t. Taken from [165].

The path from the root to the correct leaf is called a decision path. The construction of the tree is a complicated process that will not only reflect the training set but it also should generalize the knowledge about the problem [59, 68, 146, 148]. In this study we make use of the CART (Classification And Regression Trees) method described by Breiman *et al.* [22] that provides a general framework for decision tree construction. In general, the tree–growing process declares the node to be a leaf or finds another property that can be used to split the data represented at the node into subsets creating new nodes. This is process is run recursively until all the data is represented by the constructed tree.

Tree classifiers are found in many different applications, such as networking [112] or medicine [79, 144] for example. In [79] Hothorn and Lausen showed good performance of tree classifiers for glaucoma classification and in [144] Polat and Günes show that decision trees were able to classify dermatological lymphographic data with 96.71% and 87.95% accuracy respectively. The good behavior of the decision trees presented in literature were a motivation to test their performance for breast cancer malignancy classification.

6.5 Neural Networks

The idea of neural networks is based on the real interactions of human nerve system. The basic element of the neural network is the neuron, or sometimes also called a perceptron. It is a mathematical model of a biological neuron [150, 194]. In Fig. 6.2 we can see the basic schematic of a neuron. Combining a few neurons together in such a way that the neurons can interact with one another make a neural network that is able to process input data and provide us with a certain decision.

We can distinguish three types of networks depending on their architecture. If there is only one layer of neurons then it is called a single-layer neural network. A multilayer network has at least two separate layers and the output signal depends on information received from the previous layer, Fig. 6.3. The third kind of a neural network is called a recurrence
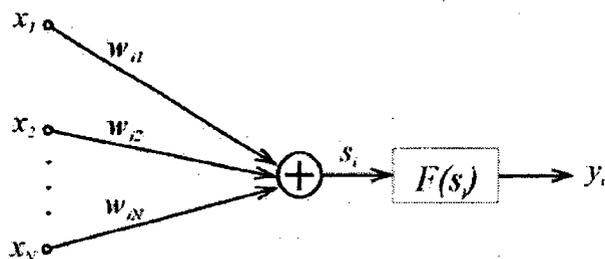
Figure 6.2: Example of a neuron in a neural network

neural network because the output signal depends not only on the inputs but also on the output through the loop-back.

Each neuron accepts an input signal of the form $X = [x_1, x_2, ..., x_n]$ and each of the sub-signals are assigned a weight. $F(s_i)$ is called an activation function of the neuron and depending on the type of the neuron activates its output. For a perceptron, which is the simplest neuron model, the activation function is of the form:

$$y_i(s_i) = \begin{cases} 1 & \text{if } s_i > 0 \\ 0 & \text{if } s_i \leq 0 \end{cases} \tag{6.7}$$
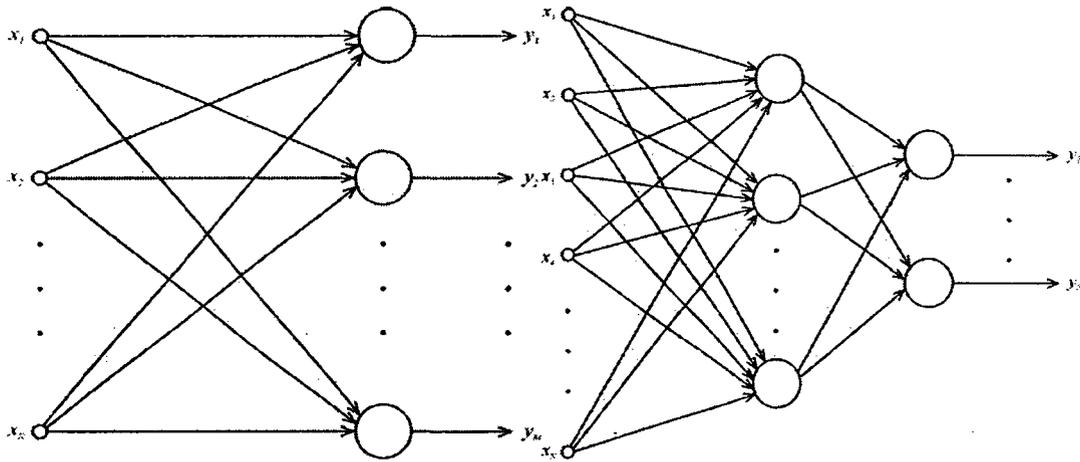
Figure 6.3: Neural Networks. a) single layer, b) multilayer

where s_i is an output signal calculated from equation 6.8,

$$s_i = \sum_{j=0}^{N} w_{ij}x_j \tag{6.8}$$

Another type of neuron is called a sigmoidal and it is characterized by an activation function of the form:

$$f(x) = \frac{1}{1 + e^{\beta x}} \tag{6.9}$$

Before we can use our neural network it is necessary to train it such that it will be able to recognize patterns. Training of a neural network is based on adjustments of weights depending on the output value. Training is performed on known patterns for which the output is known. Such a set of known patterns is called a training set. Analogously a set of unknown pattern is called a testing set.

6.5.1   Multilayer Perceptron (MLP)

MLPs are simple and one of the most widely used neural networks trained with a back-propagation method in a supervised manner [55]. These networks are powerful and are able to approximate arbitrary functions [101]. The backpropagation learning allows for error propagation through the network and adaptation of the weights of the hidden neurons. This error correction training assumes that the desired network response is known a

priori, which is usually the case in pattern recognition.

Here, an improved backpropagation training algorithm is used, called momentum learning. It was introduced by Rumelhart et al. [164] to make the training procedure faster and for better convergence. This improved learning technique uses equation 6.10 for updating the weights in hidden layers in which the step size is denoted by η and the local error by δ_i which is determined with equation 6.11 for the output layer and with equation 6.12 for the hidden layer [78].

$$w_{ij}(n+1) = w_{ij}(n) + \eta\delta_i(n)x_j(n) + \alpha(w_{ij}(n) - w_{ij}(n-1)) \tag{6.10}$$

where $\alpha \in [0.1, 0.9]$ is the momentum term and n is the iteration.

$$\delta_i(n) = e_i(n)\varphi_i'(v_i(n)) \tag{6.11}$$

where $e_i(n)$ is the error signal for that neuron and $\varphi_i'(v_i(n))$ is a derivative of the activation function.

$$\delta_i(n) = \varphi_i'(v_i(n)) \sum_k \delta_k(n)w_{ki}(n) \tag{6.12}$$

where $\delta_k(n)$ is a local gradient at neurons in the kth layer that are connected to the hidden neuron j.

To follow the progression of learning, the evolution of the mean squared error is checked and represented as a learning curve. From the curve we are able to control the learning parameters. The difficulty of the classification task is also visible. Analyzing the smoothness of the curve, the step size can be estimated. When the curve oscillates, the step should be decreased and if the curve is smooth then we can increase the step size to make the training faster. Further analysis of the curve can lead to decisions on the network topology. For example, the curve stabilization after a few iterations with unsatisfactory error rate suggests additional hidden neurons or layers, or even a completely different architecture.

6.5.2 Principal Component Analysis (PCA) Neural Networks

PCA networks are a combination of supervised and unsupervised trained neural networks. PCA analysis finds a solution in an unsupervised manner from input data and then a

supervised MLP is used for classification of the components [138].

Typically, PCA analytically finds the projections into an orthogonal set of directions within the input space, called eigenvectors. The projection corresponds to the eigenvalues. According to Sanger [167] and Oja [138] this approach can be accomplished with a single layer neural network that uses the Hebbian rule for training. Fig. 6.4 shows a typical neural network that is used to solve the PCA problem. This network takes p components as input and yields m outputs as described by equation 6.13. The number of outputs has to smaller than the number of input components, and weights are adjusted according to the Hebbian rule described by equation 6.14.
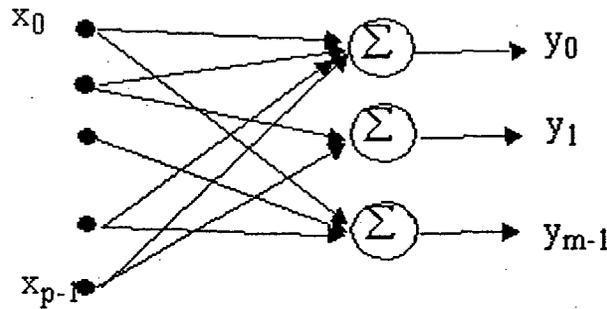
Figure 6.4: PCA neural network, taken from [134]

$$y_i(n) = \sum_{i=0}^{p-1} w_{ij}(n)x_i(n) \qquad for\, j = 0, 1, ..., m - 1 \qquad (6.13)$$

$$\Delta w_{ij}(n) = \eta[y_j(n) - y_j(n)\sum_{k=0}^{j} w_{ki}(n)y_k(n)] \qquad \begin{matrix} i = 0, 1, ..., p - 1 \\ j = 0, 1, ..., m - 1 \end{matrix} \qquad (6.14)$$

As previously mentioned, PCA networks find eigenvectors without calculation of the correlation function and the weights are related to the principal components. This leads to finding corresponding eigenvalues on the output of the network and these can be further used as an input to another neural network.

Here, a feature vector is presented as the input of the PCA network and the estimated principal components are then used as the input to a MLP neural network for malignancy classification. The size of the eigenvector is equal to the size of the feature vector. The

advantage of this method is that it reduces the number of inputs for the MLP, which positively affects a required number of training pattern of our network. This results in shorter training times.

The main disadvantage of this method is poor classification when data points are poorly separated. In this case the estimation of eigenvectors is distorted resulting in poor estimation of the principal components.

6.6 Support Vector Machines (SVM)

Support Vector Machines are used to separate two or more classes of patterns or data points by constructing a boundary between them, see Fig. 6.5. An unknown point will be classified according to its orientation with respect to the boundary. To estimate the boundary

Figure 6.5: Illustration of class' separation. a) straight line, b) curve

between classes we use boundary points from each class. These points are called support vectors. This procedure is an iterative approach that minimizes some error function, e.g.,

$$\frac{1}{2}w^T w + C \sum_{i=1}^{N} \varepsilon_i \tag{6.15}$$

with the following restrictions:

$$y_i(w^T \phi(x_i) + b) \geq 1 - \varepsilon_i \quad and \quad \varepsilon_i \geq 0, \quad i = 1, ..., N \tag{6.16}$$

where C and b are constants, w is the weight vector, ε_i is a bias value that deals with overlapping cases and ϕ is a kernel function that transforms input data into the feature

space. The constant C has a major influence on the error rate and has to be carefully estimated during the training process.

Depending on the error function we can distinguish between different SVMs and also there are different kernels used out of which the most common is the Radial Base Function (RBF) kernel. Kernels can be represented as follows:

$$\phi = \begin{cases} x_i * y_i & \text{linear} \\ (\gamma x_i x_j + const.)^{pol} & \text{polynomial} \\ \exp(-\gamma |x_i - x_j|^2) & \text{RBF} \\ \tanh(\gamma x_i x_j + const.) & \text{sigmoidal} \end{cases} \tag{6.17}$$

The SVMs used in this study use the idea of large margin classifiers for training, that provide a good generalization of the problem. Large margin classifiers use a kernel–based method for data separation.

The learning process uses the Adatron algorithm [61] which guarantees the convergence to the solution assuming that the solution exists. According to the authors this method is able to learn nonlinear decision boundaries. This algorithm can be extended by substitution of the inner product of patterns in the input space by the kernel function (see eq. 6.18) which allows for maximization of the following function

$$J(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j G_{2\sigma^2}(x_i - x_j) \tag{6.18}$$

with the following constraints:

$$\sum_{i=1}^{N} d_i \alpha_i = 0, \quad \alpha_i \geq 0, \quad \forall i \in 1, ...N.$$

where x_i, x_j are feature vectors, $d_i, d_j \in \{2, 3\}$ are malignancy grades, α_i are multipliers and G is a Gaussian kernel with variance σ.

$$g(x) = d_i (\sum_{j=1}^{N} d_j \alpha_j G_{2\sigma^2}(x - x_j) + b) \tag{6.19}$$

where g is a decision boundary.

$$\begin{cases} \alpha_i(n+1) = \alpha_i(n) + \Delta\alpha_i(n), & if \quad \alpha_i(n) + \Delta\alpha_i > 0 \\ \alpha_i(n+1) = \alpha_i(n) & if \quad \alpha_i(n) + \Delta\alpha_i \leq 0 \end{cases} \tag{6.20}$$

$$M = \min_{i} g(x) \tag{6.21}$$

where α_j is nonzero if and only if x_j is a support vector.

Training starts with a starting multiplier $\alpha_i = 0.1$ and a terminating threshold $t = 0.01$. We calculate $\Delta\alpha_i = \eta[1 - M]$ and perform an update according to equation 6.20 as long as $M > t$, η is a predefined learning rate.

6.7 Self–organizing Maps (SOM)

In addition to the methods that use a supervised techniques for training we also tested Self–organizing maps as a representative of the unsupervised methods. SOMs networks reduce the input space into representative features according to self–organizing process and are trained in an unsupervised manner [98]. These networks consist of only one layer with a linear transfer function for its neurons. It uses a comprehensive learning algorithm for weights estimation. This procedure updates the weight of only one, winning neuron for each input pattern. According to Kohonen [98] the introduction of an additional weight change of the neighboring neurons with smaller step size results in better correspondence to the features of the input data.

To train the SOM network we start with initialization of weights with small random values and for each input data a winning neuron is found according to equation 6.22. Neighboring weights are then calculated according to equation 6.23.

$$\overrightarrow{i}(x) = \arg\min_{j} \|\overrightarrow{x}(n) - w_j\| \tag{6.22}$$

$$w_j(n + 1) = w_j(n) + \eta(n)[x(n) - w_j(n)] \tag{6.23}$$

According to Kohonen, the neighborhood taken into consideration should be Gaussian and he suggests the neighborhood description according to equation 6.24.

$$\Lambda_{j,j^o}(n) = \exp\left(-\frac{|r_j - r_{j^o}|^2}{2\sigma^2(n)}\right) \tag{6.24}$$

where j^o is the winning neuron and $|r_j - r_{j^o}|$ is a distance between the winning node and the $j - th$ node. From this we can notice that this is an adaptive procedure, because

the neighborhood and learning rate depend on the current iteration. Due to this fact our neighborhood should be as large as the output space at the start and should be decreasing during the iterations according to the equation 6.25. We can take the same reasoning for the step size, which should be big at the beginning and progressively decrease according to equation 6.26 until it reaches zero.

$$\sigma(n) = \frac{1}{c_\sigma + d_\sigma n} \tag{6.25}$$

$$\eta(n) = \frac{1}{a_\eta + b_\eta n} \tag{6.26}$$

where a_η, b_η, c_σ and d_σ are constants.

When stabilization of the SOM is reached, the output of the map is presented as an input to another classifier for further classification. Here, we use the MLP architecture for neighborhoods classification to assign them one of the malignancy grades. This procedure make MLP learning easier and faster by reduction of the dimensionality of the input space.

6.8 Set Partitioning – Cross–validation

Cross–validation is a statistical method of partitioning of the data into subsets for further analysis. This method is also called a rotation estimation [51, 96]. In principal, one subset is used for testing while the remaining subsets are used for training the classifiers. Fig. 6.6 illustrates the partitioning of the data into four subsets. In general, this is called a K–fold cross–validation for $K = 4$. The idea behind K–fold cross–validation is that the data is partitioned into K subsets from which one is retained for testing and the remaining K–1 subsets are used for training. This process is repeated K times and each time a different fold is used for testing.

Leave–one–out cross–validation involves a degenerate case of K–fold cross–validation, where K is equal to the number of all samples in the data set. Here only one sample is used for validation purposes and requires K–1 repetitions (see Fig. 6.7). For large data sets this is a very time consuming method. The most commonly used K for cross-validation is $K = 10$ which provides good estimation of the error rate. The overall error is usually calculated as
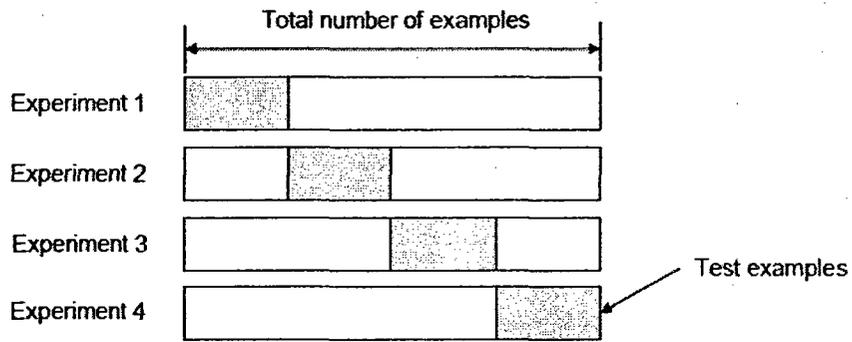
Figure 6.6: 4-fold cross–validation, taken from [44].

an average error of all the runs and can be calculated as:

$$Err = \frac{1}{K} \sum_{i=1}^{K} Err_i \qquad (6.27)$$
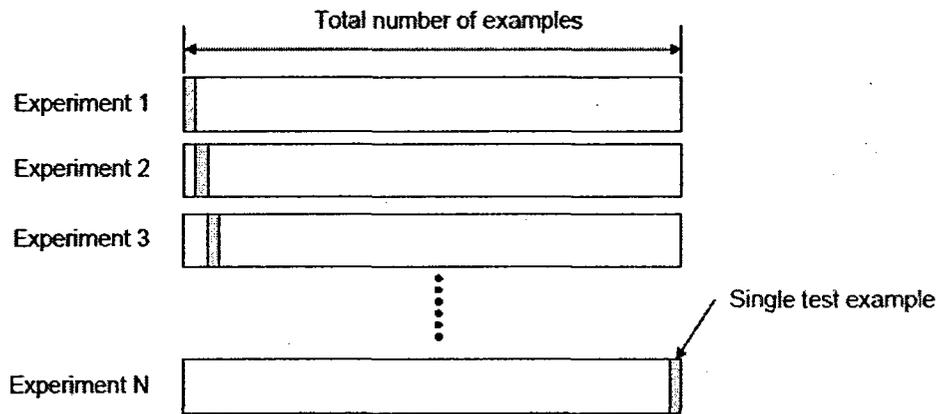
Figure 6.7: Leave–one–out cross–validation, taken from [44].

6.9 Classification Algorithm Evaluation

To be able to say how the proposed classifiers behave and to be able to evaluate their performance there is a need for introduction of quantitative criteria. The most popular and reliable evaluation method is based on the confusion matrix that contains information about the actual and predicted classifications [97]. The fields of the matrix are filled depending on the classification result of the tested samples. Fig. 6.8 shows an example of a general confusion matrix for the two class problem.

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Figure 6.8: Two class problem confusion matrix.

The TP, TN, FP and TN entries are assigned during classification and they have the following meaning:

- TP – True Positives – the number of Positive classifications correctly classified as Positive.

- TN – True Negatives – the number of Negative classifications correctly classified as Negative.

- FP – False Positives – the number of Negative classifications incorrectly classified as Positive.

- FN – False Negatives – the number of Positive classifications incorrectly classified as Negative.

Based on these values we can now define *sensitivity* and *specificity* of our system. The system sensitivity is a measure of the system's ability to make true positive classifications

and can be defined as:

$$Sens = \frac{TP}{TP + FN}.$$ (6.28)

The specificity is an ability of true negative classifications and its definition is:

$$Spec = \frac{TN}{TN + FP}.$$ (6.29)

Sensitivity and specificity define the behavior of the proposed classification system and the closer *Sens* and *Spec* values are to 1 the better is the performance of such a system. In oncological related classification it is very important to keep both of these values as high as possible. Here we can introduce another measure to evaluate the overall quality of the recognition as:

$$Q = \frac{Sens + Spec}{2}.$$ (6.30)

6.10   Database of Fine Needle Aspirates

The database used in this study is a collection of images recorded during cytological examinations of the breast. The images were collected at the Department of Pathology, Medical University of Wrocław, Poland by prof. Michał Jeleń. The database consists of 101 FNA biopsy images (see Appendix A)that were graded by an expert pathologist. During this study we assumed the pathological gradings to be our 'gold standard.'

All the collected biopsies were recorded with two different magnifications using the setup described in section 3.2.1 and stored as bitmap files of the size 764×572 pixels. The first subset of 47 FNA aspirates consists of images with 100x magnification. Images from this subset are used to calculate features related to the first factor of the Bloom–Richardson grading scheme. They allow us to extract features based on the cells' tendency to form groups. Healthy and low malignant cases (see Fig. 6.9a) tend to form one or two large groups in the image, while those cases with large malignancy are loosely spread and groups usually consist of only a few cells. Based on that information three features are proposed: average area of groups, number of groups and cells dispersion measure. These features were described in more details in section 5.2 and their discriminatory powers were presented in

section 5.3.

The second subset of images in the database consists of 54 FNA aspirates recorded with 400x magnification. These images were used for calculation of all the features describing nucleus pleomorphism and they include binary histogram, momentum, textural and color based features. These shape–based features provide us with valuable information about cells nuclei. Here we make use of the differences between healthy and cancerous cells as described in section 2.2. Essentially, low malignancy cases have uniform size and staining while in more malignant cases this tendency is disturbed and nuclei in the image will assume nonuniform sizes and will have stronger staining variations (see Fig. 6.9). A detailed description of these features can be found in section 5.2. It can be noticed that there are 6 instances where 2 images with 400x magnification were taken for one 100x magnification image and one instance were 3 high magnification images were taken for one low magnification image. This is due to the fact that there were two or more suspicious regions in the 100x image. These cases are treated as separate cases and no prior knowledge about that fact is taken into consideration during testing. All of the images in the database were
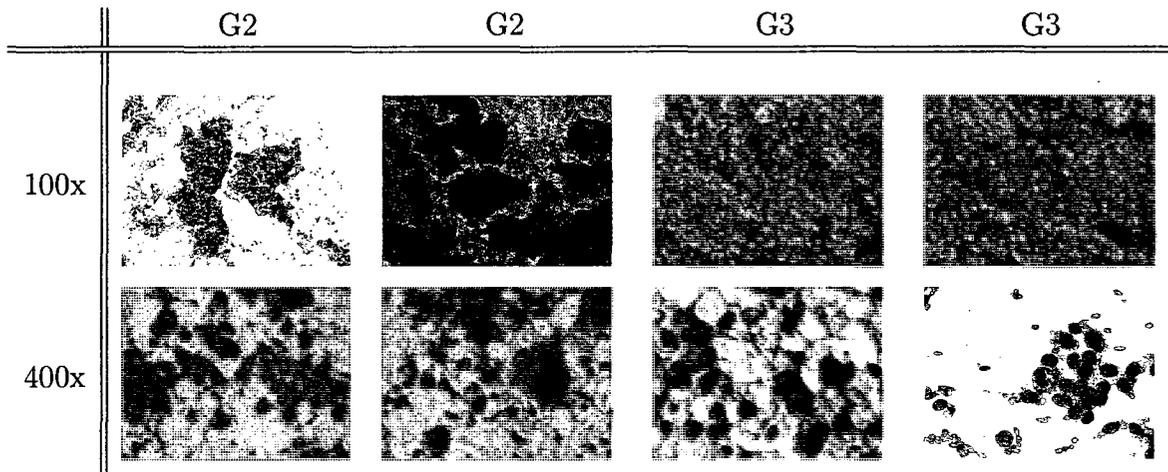
Figure 6.9: Example of images from the database.

stained with the Haematoxylin and Eosin technique (HE) which yielded purple and black

stain for nuclei, shades of pink for cytoplasm and orange/red for red blood cells. All the images were obtained by an Olympus BX 50 microscope with mounted CCD–IRIS camera connected to a PC computer with MultiScan Base 08.98 software.

6.11 Classification Results

To check the behavior of the presented features, several tests have been performed. We tested the performances of different classifiers on different testing sets. The first set training and testing sets is presented in Table 6.1. These sets were chosen randomly and they consist of different number of samples used for training the classifiers [84, 88]. Tests on a different sets were also performed with cross validation method and will be described later in this section. Additionally, tests on different features were also performed.

To be able to fully test the classifiers performance there is a need to calculate the error rates of the training process. This will allow for the evaluation of the training procedure and further enhancement of the feature set to be more distinctive. Here, low and high

	G2		G3	
	Training [%]	Testing [%]	Training [%]	Testing [%]
set #1	50	50	30	70
set #2	50	50	50	50
set #3	50	50	70	30
set #4	50	50	90	10
set #5	30	70	50	50
set #6	70	30	50	50

Table 6.1: Training and testing sets used for classification.

magnification features were tested separately to check for their ability to classify breast cancer malignancy when only structural or pleomorphic features are taken into consideration [88]. Table 6.2 presents classification results obtained for 100x magnification images. From that table we can notice that KNN for $k = 9$ has the highest average error rate of 48.3% for all performed tests, and its overall performance for all k's taken into consideration is worse than for any other classifier. The tree classifiers performed better that KNN

114

achieving 40.9% error which is still significantly higher than other compared algorithms. The best performance for these type of images was recorded for the SOMs network with an average error rate of 10.56% and best for test set # 6, where the most G3 cases were taken for testing. The best recorded error rate in this case was 7.8%. Such a large difference of the classification rate can be caused by the fact that KNN and tree classifiers are a non-parametric methods and they require a large data set to achieve good classification performance. From this discussion we can conclude that although structural features provide good separation between malignancy classes, as proved by the KS measure, they do not provide good classification and therefore are not very good for malignancy classification when used by themselves. Hence we can conclude that additional features are needed to support the classification.

To test the performance of high magnification images, we used the same set structure as

Classifier		Set #1	Set #2	Set #3	Set #4	Set #5	Set #6	Avg. Performance
SOM		9.33%	16.21%	11.42%	9.96%	8.67%	7.8%	10.56%
MLP		11.99%	10.9%	12.36%	10.21%	9.79%	8.75%	10.67%
SVM		11.96%	15.4%	14.29%	13.23%	17.16%	8.67%	13.44%
PCA		16.61%	10.65%	11.51%	12.28%	8.96%	7.79%	11.3%
TREE		20%	80%	20%	20%	80%	20%	40.9%
KNN	1	30%	70%	20%	30%	70%	30%	42.9%
	2	30%	70%	10%	20%	80%	30%	42.9%
	3	30%	70%	20%	30%	80%	30%	43.8%
	4	30%	70%	20%	20%	80%	30%	43.1%
	5	30%	70%	30%	40%	80%	30%	46.6%
	6	30%	80%	20%	30%	70%	30%	44.8%
	7	40%	70%	20%	30%	80%	30%	46.5%
	8	30%	80%	10%	20%	80%	30%	43.2%
	9	30%	80%	20%	30%	80%	30%	48.3%
	10	30%	80%	20%	30%	80%	30%	46.1%
	Avg.	31%	74%	19%	28%	78%	30%	44.82%

Table 6.2: Error rates of tested classifiers for 100x magnification images.

for 100x magnification images (see Table 6.1) and the summary of classification rates is presented in Table 6.3. Here, it can be can noticed that SVMs have the lowest error rate of all the classifiers for set #5, which consists of only 30% of G2 cases. Such a low error rate tells us that the cases used for in this set were the most distinct for both classes and

that G2 cases in this set are the most characteristic and the highest class separation was found. This can also be proved by the distribution plot on Fig. 5.13 for this set. Other test sets didn't provide such a good classification giving average error rates even higher than for 100x images. Out of all the tested classifiers MLP performed best achieving an average error rate of 17.3% with the lowest error for set #5, 6.73%. The worst average performers was again recorded for KNN classifier with $k = 1$ achieving 52.4% error. Here, the same observation as for 100x images can made that the KNN performance is significantly worse than for all of the remaining algorithms. We can also notice that the tree classifier achieved better classification accuracy than KNN with an error of 39.2%, which is still rather high in comparison to the remaining classifiers.

Having in mind the classification rates for 100x and 400x magnifications, additional tests

Classifier		Set #1	Set #2	Set #3	Set #4	Set #5	Set #6	Avg. Performance
SOM		23.28%	21.91%	19.49%	22.43%	6.32%	23.75%	19.53%
MLP		23.18%	16.47%	22.76%	12.44%	6.73%	24.71%	17.30%
SVM		21.55%	20.25%	20.91%	23.15%	5.76%	24.71%	19.39%
PCA		22.84%	16.32%	20.88%	22.01%	7.22%	23.66%	18.82%
TREE		40%	40%	30%	50%	40%	30%	39.20%
KNN	1	50%	50%	60%	60%	50%	40%	52.40%
	2	40%	40%	50%	50%	40%	40%	42.50%
	3	40%	50%	60%	60%	50%	40%	51.70%
	4	40%	30%	30%	60%	40%	40%	42.00%
	5	40%	40%	50%	70%	50%	40%	48.90%
	6	40%	40%	30%	60%	30%	40%	41.60%
	7	40%	40%	40%	70%	60%	40%	48.60%
	8	40%	30%	30%	40%	50%	40%	41.20%
	9	40%	30%	30%	70%	50%	40%	45.60%
	10	40%	30%	30%	40%	50%	40%	40.20%
	Avg.	41%	38%	41%	58%	47%	40%	45.47%

Table 6.3: Error rates of tested classifiers for 400x magnification images.

on a subset of features were performed. For this purpose, classifiers were fed a feature vector that consists of all the 100x features and five 400x features. For this test areas (A_{100} and A_i), number of groups, dispersion, perimeter, eccentricity, convexity and average gray level were used as a feature vector. As can be seen in Table 6.3 test set #5 provided the best learning with the most distinct cases for 400x images. If we look at the performance

Classifier		Set #5	Best	Worst	Average
SOM		5.22%	5.22%	17.27%	11.53%
MLP		3.30%	3.30%	15.77%	9.81%
SVM		6.76%	6.76%	16.73%	13.30%
PCA		8.16%	7.65%	10.01%	10.92%
TREE		40%	10%	40%	18.1%
KNN	1	20%	20%	30%	25.2%
	2	20%	20%	40%	28.8%
	3	20%	20%	40%	30.0%
	4	20%	20%	40%	29.4%
	5	20%	20%	40%	32.8%
	6	20%	20%	40%	32.0%
	7	20%	20%	40%	32.9%
	8	20%	20%	40%	32.7%
	9	30%	30%	50%	35.7%
	10	20%	20%	40%	32.7%
	Avg.	21%	21%	40%	31.22%

Table 6.4: Error rates of tested classifiers for 100x and 400x features.

of that set for the 100x images we can see that the recorded error rates were reasonable, except for SVMs. Although it was the best performing classifier for high magnification images we can still say that set #5 would be the best choice for training our classifier with all the extracted features. This reasoning can be proved with the results obtained from the tests and summarized in Table 6.4. Here, the tests were performed only on set #5 since it was the best performing and as we can see from the table the results confirm our reasoning. Taking this into consideration we can say that the use of the most distinct cases for training purposes is essential. This reasoning could be applied not only the problem of breast cancer grading but also for other cancerous data.

As we can see the classifiers performed best when both 100x and 400x features together were used. In addition to these results, a comparison of the classification behavior on differently segmented nuclei was also performed. Additional 400x features were added to the feature vector and their influence on the classification results was studied. To perform these tests a cross–validation method was used for data partitioning into training and testing sets. For the purpose of this study, the most common and well known variation of k–fold cross–validation was used, for $k = 10$. All four classifiers were trained according to the principals of cross–validation and their errors were recorded. Table 6.5 summarizes all of the recorded error rates for all used classifiers and segmentation algorithms.

Classifier		Cross Validation Set										Avg.
		1	2	3	4	5	6	7	8	9	10	
SOM	LS	14.64%	33.57%	22.95%	12.58%	12.89%	19.92%	2.46%	4.88%	38.34%	26.09%	18.83%
	FCM	15.03%	38.15%	31.51%	25.91%	9.34%	10.52%	2.89%	4.66%	15.42%	31.23%	18.47%
	GLQ	14.83%	22.45%	26.04%	14.63%	21.29%	18.69%	6.31%	15.77%	13.18%	35.97%	18.92%
MLP	LS	2.78%	11.13%	23.23%	10.50%	11.14%	14.48%	2.36%	2.62%	28.01%	16.89%	12.31%
	FCM	20.66%	37.47%	7.21%	11.65%	10.13%	6.23%	1.08%	4.16%	19.77%	17.51%	13.59%
	GLQ	5.58%	33.67%	8.65%	5.87%	5.07%	29.91%	12.73%	1.80%	7.90%	15.75%	12.69%
SVM	LS	14.59%	23.03%	31.78%	18.92%	16.93%	23.94%	9.41%	5.25%	30.47%	26.20%	20.05%
	FCM	25.30%	33.62%	25.46%	19.16%	10.91%	14.22%	6.08%	13.60%	23.94%	28.00%	20.03%
	GLQ	13.30%	29.67%	22.81%	17.85%	12.94%	26.84%	16.63%	19.78%	11.35%	25.45%	19.66%
PCA	LS	13.19%	18.21%	30.53%	10.61%	9.55%	15.79%	3.01%	2.06%	32.84%	26.63%	16.24%
	FCM	1.65%	46.01%	29.67%	30.67%	3.23%	21.69%	17.07%	19.85%	24.42%	27.86%	22.21%
	GLQ	11.26%	40.29%	11.17%	13.23%	13.88%	24.25%	6.14%	14.28%	11.21%	13.10%	15.88%
TREE	LS	0.0%	60.0%	40.0%	20.0%	40.0%	20.0%	0.0%	0.0%	60.0%	44.6%	28.46%
	FCM	40.0%	60.0%	20.0%	20.0%	0.0%	20.0%	0.0%	0.0%	60.0%	55.6%	27.56%
	GLQ	0.0%	40.0%	0.0%	20.0%	20.0%	20.0%	0.0%	20.0%	16.7%	33.3%	15.00%
KNN 1	LS	20.0%	40.0%	80.0%	60.0%	20.0%	80.0%	20.0%	20.0%	40.0%	66.7%	44.7%
	FCM	40.0%	40.0%	60.0%	40.0%	60.0%	40.0%	100.0%	0.0%	80.0%	88.9%	54.9%
	GLQ	60.0%	80.0%	80.0%	20.0%	20.0%	20.0%	80.0%	60.0%	80.0%	88.9%	58.9%
KNN 2	LS	0.0%	80.0%	40.0%	60.0%	20.0%	40.0%	20.0%	20.0%	20.0%	88.9%	38.9%
	FCM	20.0%	80.0%	40.0%	40.0%	60.0%	0.0%	80.0%	0.0%	20.0%	88.9%	42.9%
	GLQ	0.0%	80.0%	40.0%	40.0%	20.0%	0.0%	60.0%	60.0%	20.0%	88.9%	40.9%
KNN 3	LS	0.0%	20.0%	40.0%	60.0%	20.0%	80.0%	40.0%	20.0%	20.0%	66.7%	36.7%
	FCM	60.0%	40.0%	80.0%	40.0%	20.0%	60.0%	80.0%	20.0%	60.0%	88.9%	54.9%
	GLQ	40.0%	80.0%	80.0%	50.0%	0.0%	20.0%	60.0%	60.0%	40.0%	55.6%	48.6%
KNN 4	LS	0.0%	20.0%	40.0%	40.0%	20.0%	20.0%	0.0%	20.0%	0.0%	88.9%	30.9%
	FCM	20.0%	40.0%	60.0%	40.0%	40.0%	60.0%	20.0%	60.0%	40.0%	88.9%	40.9%
	GLQ	0.0%	80.0%	40.0%	40.0%	60.0%	0.0%	40.0%	20.0%	40.0%	88.9%	44.9%
KNN 5	LS	0.0%	20.0%	40.0%	40.0%	60.0%	80.0%	20.0%	20.0%	0.0%	77.8%	35.8%
	FCM	40.0%	40.0%	80.0%	40.0%	60.0%	60.0%	40.0%	60.0%	40.0%	88.9%	50.9%
	GLQ	0.0%	80.0%	60.0%	20.0%	60.0%	0.0%	60.0%	20.0%	40.0%	88.9%	48.9%
KNN 6	LS	0.0%	20.0%	40.0%	40.0%	60.0%	80.0%	20.0%	0.0%	20.0%	88.9%	36.9%
	FCM	20.0%	80.0%	60.0%	0.0%	60.0%	40.0%	0.0%	20.0%	20.0%	88.9%	42.9%
	GLQ	0.0%	80.0%	40.0%	0.0%	60.0%	0.0%	60.0%	0.0%	40.0%	77.8%	42.9%
KNN 7	LS	20.0%	20.0%	60.0%	0.0%	60.0%	80.0%	40.0%	20.0%	40.0%	77.8%	43.8%
	FCM	20.0%	80.0%	60.0%	20.0%	80.0%	40.0%	0.0%	40.0%	20.0%	88.9%	42.9%
	GLQ	20.0%	80.0%	40.0%	40.0%	60.0%	0.0%	60.0%	40.0%	40.0%	88.9%	46.9%
KNN 8	LS	0.0%	60.0%	60.0%	20.0%	20.0%	80.0%	20.0%	20.0%	40.0%	77.8%	39.8%
	FCM	20.0%	80.0%	40.0%	40.0%	60.0%	20.0%	60.0%	20.0%	20.0%	88.9%	40.9%
	GLQ	0.0%	80.0%	40.0%	0.0%	60.0%	0.0%	20.0%	0.0%	40.0%	88.9%	38.9%
KNN 9	LS	0.0%	20.0%	60.0%	0.0%	60.0%	80.0%	20.0%	40.0%	40.0%	77.8%	39.8%
	FCM	20.0%	80.0%	60.0%	60.0%	60.0%	40.0%	20.0%	20.0%	20.0%	88.9%	46.9%
	GLQ	0.0%	80.0%	60.0%	40.0%	20.0%	0.0%	60.0%	40.0%	20.0%	88.9%	44.9%
KNN 10	LS	0.0%	40.0%	60.0%	40.0%	60.0%	80.0%	0.0%	20.0%	20.0%	88.9%	32.9%
	FCM	0.0%	80.0%	60.0%	40.0%	60.0%	0.0%	40.0%	20.0%	20.0%	88.9%	38.9%
	GLQ	0.0%	80.0%	60.0%	40.0%	60.0%	0.0%	40.0%	0.0%	20.0%	88.9%	38.9%

Table 6.5: Error rates obtained with cross–validation sets.

From the table we can notice that the best average performance was recorded for multilayer perceptron when the level set method was used for nuclei segmentation. MLP achieved an average error of 12.31% while for the remaining two segmentation algorithms their recognition error increased and were 12.69% and 13.59% for textural and fuzzy c–means, respectively. The rest of the tested classifiers performed significantly worst than MLP achieving the highest error of 20% for SVMs with level set and fuzzy c–means segmentations. These results are visualized in Fig. 6.10. From the figure we can easily notice the better MLP recognition rate, which was as good as 98.92% for set #7 using fuzzy c–means segmentation with 77.2% sensitivity and 60.04% classification quality (see Table 6.6). Table 6.6 shows the calculated sensitivity, specificity and quality values for all of the tested classifiers. From this table one can see that the MLP has the best quality measures for all of the compared segmentation algorithms. Only SOMs seem to have higher quality for level set segmentation. The worst classification quality can be noticed for KNN($k = 1$) with textural segmentation.

In chapter 5 feature validation results were presented. Based on the results of the KS measure and features correlation we can choose smaller sets of features that will decrease the size of the feature vector while keeping approximately the same error rate. According to this reasoning, different features sets were chosen and their performance was tested. Two tests were performed depending on the validation technique. First, the correlation results were used to reduce the number of features by taking into consideration only one or two features from the most correlated features. After this reduction, the classifiers were trained and their performance was tested with a 15-element feature vector. The features that were taken into use are: $area_{100}$, number of groups, dispersion, perimeter, convexity, x-centroid, orientation, vertical projection, ϕ_3, histogram mean, histogram energy, textural homogeneity, red channel histogram mean, red channel histogram skew and red channel histogram width. The classification results are collected in Table 6.7 and the sensitivity and quality measures are summarized in Table 6.8. From these tables it can be noticed that feature vector reduction did not significantly increase the classification error rate.
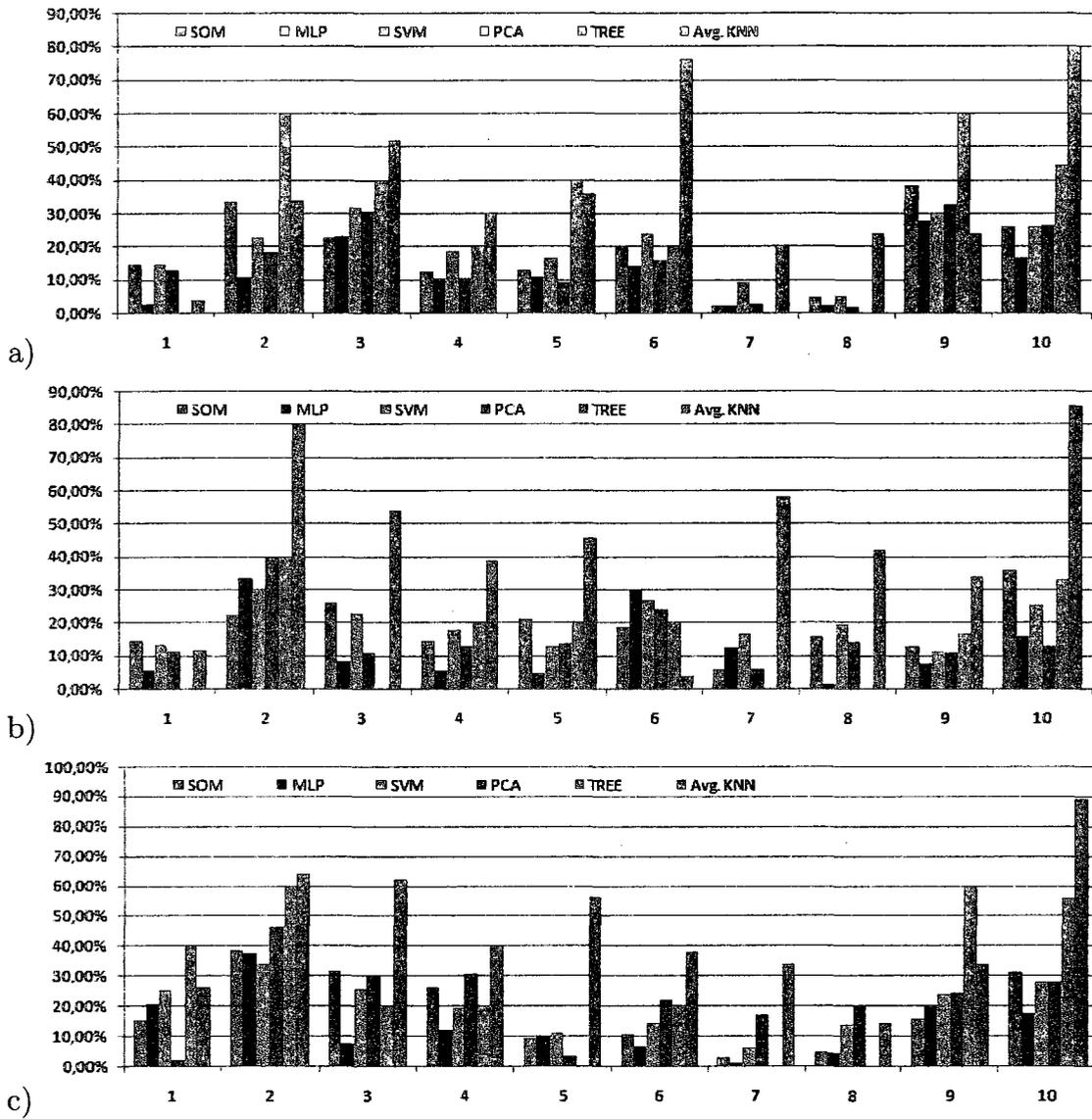
Figure 6.10: Error rates bar charts. a)Level set, b) Gray level quantization, c) Fuzzy c–means.

Classifier			Sensitivity	Specificity	Quality
SOM		LS	84.20%	45.80%	65.00%
		FCM	76.00%	15.40%	45.71%
		GLQ	72.00%	23.30%	47.67%
MLP		LS	77.30%	48.90%	63.11%
		FCM	77.20%	42.90%	60.04%
		GLQ	73.30%	25.80%	49.58%
SVM		LS	84.00%	36.70%	60.33%
		FCM	65.00%	40.00%	52.50%
		GLQ	84.70%	25.00%	54.83%
PCA		LS	60.00%	36.30%	48.13%
		FCM	61.30%	27.50%	44.42%
		GLQ	83.30%	31.30%	57.29%
TREE		LS	82.20%	28.30%	55.25%
		FCM	83.20%	41.30%	62.21%
		GLQ	84.70%	39.20%	61.92%
KNN	1	LS	70.20%	19.20%	44.67%
		FCM	62.80%	13.30%	38.08%
		GLQ	45.50%	22.50%	34.00%
	2	LS	85.30%	11.70%	48.50%
		FCM	90.00%	0.00%	45.00%
		GLQ	88.00%	6.70%	47.33%
	3	LS	76.80%	31.70%	54.25%
		FCM	66.30%	11.70%	39.00%
		GLQ	69.50%	13.80%	41.63%
	4	LS	86.70%	29.20%	57.92%
		FCM	84.20%	8.30%	46.25%
		GLQ	87.50%	0.00%	43.75%
	5	LS	74.70%	30.40%	52.54%
		FCM	66.50%	13.30%	39.92%
		GLQ	82.20%	0.00%	41.08%
	6	LS	78.00%	19.20%	48.58%
		FCM	88.70%	0.00%	44.33%
		GLQ	90.00%	0.00%	45.00%
	7	LS	66.20%	25.40%	45.79%
		FCM	83.70%	5.00%	44.33%
		GLQ	85.50%	0.00%	42.75%
	8	LS	82.20%	15.40%	48.79%
		FCM	90.70%	0.00%	45.33%
		GLQ	94.00%	0.00%	47.00%
	9	LS	70.20%	25.40%	47.79%
		FCM	83.30%	0.00%	41.67%
		GLQ	86.70%	0.00%	43.33%
	10	LS	86.70%	21.70%	54.17%
		FCM	92.70%	0.00%	46.33%
		GLQ	92.70%	0.00%	46.33%

Table 6.6: Sensitivity, specificity and quality measures, as defined in Eqs. (6.28–6.30)

When we take MLP into discussion, which performed best on all features, we can notice that only for level set segmentation the error rate increased significantly from 12.31% to 16.76% while for the remaining two segmentation methods the error change was at most 0.5%. From Table 6.8 it can be noticed that the feature vector reduction allowed for the better classification quality, achieving 66.33% from 60.04% with MLP using fuzzy c-means segmentation.

The second test performed on the reduced feature vector involved the use of the KS measure. Table A.1 summarizes the recorded error rates for all the segmentation methods. There were different sets chosen for each segmentation algorithm depending on the KS measure. The choice of features is presented in Table 6.10. Only those features were chosen that allowed for clear separation between malignancy classes. For this test only the MLP classifier was trained due to its previous best performance. From the table we can see that the best performance was recorded for fuzzy c-means segmentation with an error rate of 10.59%. In this test, level set segmentation gave the worst classification results with 19.85% error.
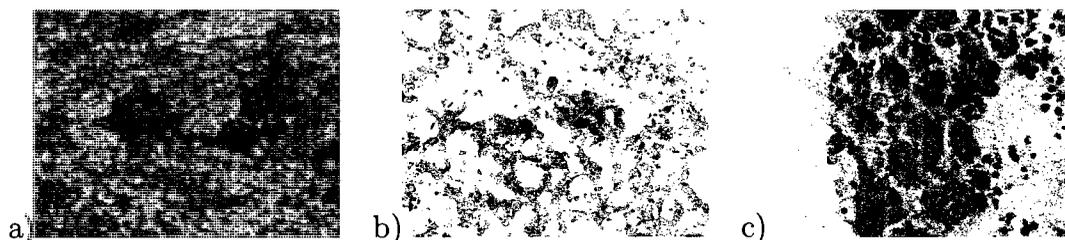
Figure 6.11: Classification results for 100x images.a) Correctly classified image, b) Intermediate malignancy misclassified as high, c) High malignancy correctly classified.

To visualize the classification task, some of the correctly and incorrectly classified images are presented on Fig. 6.11, 6.12 and 6.13. Fig. 6.11a presents a correctly classified image while Fig. 6.11b and Fig. 6.11c show misclassified cases of intermediate malignancy classified as high and high malignancy misclassified as intermediate. Respectively for 400x

Classifier		1	2	3	4	5	6	7	8	9	10	Avg.
SOM	LS	1.57%	17.09%	21.54%	6.06%	21.24%	22.39%	12.34%	2.56%	35.47%	15.37%	15.56%
	FCM	35.67%	32.30%	25.29%	14.73%	4.79%	4.18%	4.12%	3.20%	29.00%	19.97%	17.33%
	GLQ	4.65%	21.86%	20.01%	19.06%	11.23%	34.52%	17.27%	23.66%	20.68%	27.56%	20.05%
MLP	LS	2.04%	11.73%	21.46%	15.61%	12.63%	25.06%	11.02%	8.66%	38.39%	20.95%	16.76%
	FCM	17.01%	41.65%	4.78%	13.11%	6.32%	5.84%	2.44%	3.33%	20.37%	19.97%	13.48%
	GLQ	4.72%	30.43%	13.07%	2.51%	9.19%	25.95%	12.33%	1.69%	8.62%	23.94%	13.25%
SVM	LS	14.59%	23.03%	31.78%	18.92%	16.93%	23.94%	9.41%	5.25%	30.46%	26.19%	20.05%
	FCM	25.30%	33.62%	25.46%	19.17%	10.91%	14.22%	6.08%	13.61%	23.94%	28.00%	20.03%
	GLQ	8.36%	28.20%	18.59%	8.48%	15.24%	27.87%	14.90%	17.72%	14.69%	23.87%	17.79%
PCA	LS	3.86%	18.81%	20.11%	19.07%	24.75%	20.21%	11.81%	13.99%	39.91%	12.24%	18.48%
	FCM	18.55%	44.42%	28.56%	17.89%	4.28%	2.00%	2.45%	13.14%	31.98%	27.36%	19.06%
	GLQ	21.52%	40.97%	20.57%	1.76%	8.24%	35.39%	11.81%	9.91%	10.39%	21.51%	18.21%
TREE	LS	0.0%	40.0%	40.0%	60.0%	20.0%	20.0%	0.0%	0.0%	60.0%	44.4%	28.44%
	FCM	40.0%	60.0%	20.0%	20.0%	40.0%	20.0%	0.0%	0.0%	20.0%	22.2%	24.22%
	GLQ	0.0%	40.0%	20.0%	20.0%	60.0%	20.0%	0.0%	0.0%	20.0%	55.6%	23.56%
KNN 1	LS	20.0%	40.0%	40.0%	0.0%	80.0%	40.0%	60.0%	60.0%	60.0%	44.4%	44.4%
	FCM	20.0%	60.0%	20.0%	80.0%	80.0%	40.0%	60.0%	40.0%	60.0%	66.7%	52.67%
	GLQ	20.0%	20.0%	40.0%	20.0%	60.0%	60.0%	20.0%	40.0%	20.0%	44.4%	32.44%
KNN 2	LS	0.0%	40.0%	40.0%	0.0%	80.0%	20.0%	0.0%	0.0%	40.0%	66.7%	28.7%
	FCM	20.0%	80.0%	40.0%	40.0%	60.0%	0.0%	40.0%	20.0%	20.0%	88.9%	40.9%
	GLQ	20.0%	40.0%	40.0%	20.0%	40.0%	40.0%	0.0%	0.0%	60.0%	55.6%	27.6%
KNN 3	LS	20.0%	40.0%	80.0%	20.0%	80.0%	40.0%	20.0%	0.0%	60.0%	55.6%	41.6%
	FCM	20.0%	80.0%	60.0%	40.0%	60.0%	40.0%	60.0%	20.0%	40.0%	77.8%	49.8%
	GLQ	0.0%	20.0%	40.0%	20.0%	60.0%	40.0%	20.0%	60.0%	60.0%	44.4%	38.4%
KNN 4	LS	0.0%	80.0%	40.0%	40.0%	80.0%	40.0%	0.0%	20.0%	20.0%	55.6%	37.6%
	FCM	20.0%	60.0%	60.0%	60.0%	40.0%	0.0%	20.0%	0.0%	60.0%	88.9%	40.9%
	GLQ	20.0%	60.0%	60.0%	20.0%	60.0%	40.0%	20.0%	0.0%	60.0%	44.4%	36.4%
KNN 5	LS	0.0%	60.0%	40.0%	40.0%	60.0%	60.0%	20.0%	0.0%	20.0%	44.4%	34.4%
	FCM	20.0%	60.0%	80.0%	40.0%	60.0%	20.0%	40.0%	0.0%	20.0%	88.9%	44.9%
	GLQ	20.0%	60.0%	60.0%	20.0%	40.0%	40.0%	40.0%	20.0%	60.0%	44.4%	38.4%
KNN 6	LS	0.0%	80.0%	40.0%	40.0%	80.0%	60.0%	0.0%	20.0%	20.0%	55.6%	37.6%
	FCM	20.0%	80.0%	80.0%	40.0%	60.0%	20.0%	0.0%	20.0%	20.0%	88.9%	42.9%
	GLQ	0.0%	80.0%	60.0%	20.0%	60.0%	40.0%	0.0%	0.0%	0.0%	55.6%	31.6%
KNN 7	LS	0.0%	40.0%	40.0%	20.0%	60.0%	60.0%	20.0%	0.0%	0.0%	55.6%	29.6%
	FCM	20.0%	60.0%	80.0%	40.0%	60.0%	20.0%	20.0%	20.0%	0.0%	88.9%	40.9%
	GLQ	20.0%	100.0%	60.0%	20.0%	60.0%	60.0%	40.0%	0.0%	20.0%	55.6%	43.6%
KNN 8	LS	0.0%	40.0%	40.0%	20.0%	80.0%	60.0%	0.0%	0.0%	20.0%	66.7%	32.7%
	FCM	20.0%	60.0%	80.0%	40.0%	60.0%	20.0%	0.0%	0.0%	20.0%	88.9%	40.9%
	GLQ	20.0%	100.0%	60.0%	20.0%	60.0%	20.0%	20.0%	0.0%	0.0%	88.9%	36.9%
KNN 9	LS	20.0%	40.0%	80.0%	40.0%	80.0%	40.0%	20.0%	20.0%	0.0%	55.6%	35.6%
	FCM	40.0%	60.0%	60.0%	20.0%	60.0%	40.0%	20.0%	20.0%	0.0%	88.9%	44.9%
	GLQ	40.0%	60.0%	60.0%	20.0%	60.0%	40.0%	20.0%	0.0%	0.0%	88.9%	38.9%
KNN 10	LS	20.0%	40.0%	80.0%	40.0%	80.0%	40.0%	20.0%	0.0%	0.0%	66.7%	34.7%
	FCM	20.0%	80.0%	80.0%	40.0%	60.0%	40.0%	0.0%	0.0%	20.0%	88.9%	42.9%
	GLQ	20.0%	80.0%	80.0%	20.0%	60.0%	40.0%	20.0%	0.0%	20.0%	88.9%	34.9%

Table 6.7: Error rates obtained with the reduced feature vector.

123

Classifier			Sensitivity	Specificity	Quality
SOM		LS	83.20%	30.80%	57.00%
		FCM	81.20%	38.80%	59.96%
		GLQ	81.20%	38.80%	59.96%
MLP		LS	63.30%	42.90%	53.13%
		FCM	87.70%	45.00%	66.33%
		GLQ	92.00%	39.20%	65.58%
SVM		LS	80.70%	26.70%	53.67%
		FCM	65.00%	40.00%	52.50%
		GLQ	92.00%	21.70%	56.83%
PCA		LS	64.70%	26.30%	45.46%
		FCM	59.80%	37.50%	48.67%
		GLQ	75.70%	30.00%	52.83%
TREE		LS	72.20%	32.50%	52.33%
		FCM	88.20%	38.30%	63.25%
		GLQ	94.70%	23.80%	59.21%
KNN	1	LS	64.50%	35.00%	49.75%
		FCM	59.80%	23.30%	41.58%
		GLQ	77.70%	29.20%	53.42%
	2	LS	86.30%	27.50%	56.92%
		FCM	92.00%	0.00%	46.00%
		GLQ	94.00%	17.10%	55.54%
	3	LS	67.80%	30.00%	48.92%
		FCM	71.80%	9.60%	40.71%
		GLQ	72.70%	25.80%	49.25%
	4	LS	81.50%	18.80%	50.13%
		FCM	84.30%	3.30%	43.83%
		GLQ	83.70%	15.80%	49.75%
	5	LS	79.50%	27.50%	53.50%
		FCM	75.00%	10.80%	42.92%
		GLQ	81.70%	15.80%	48.75%
	6	LS	84.00%	8.80%	46.38%
		FCM	79.00%	8.30%	43.67%
		GLQ	92.70%	18.80%	55.71%
	7	LS	82.00%	33.80%	57.88%
		FCM	77.00%	20.80%	48.92%
		GLQ	65.50%	28.80%	47.13%
	8	LS	87.30%	12.50%	49.92%
		FCM	84.00%	7.50%	45.75%
		GLQ	76.00%	25.00%	50.50%
	9	LS	80.00%	23.80%	51.88%
		FCM	80.00%	7.50%	43.75%
		GLQ	70.00%	30.00%	50.00%
	10	LS	82.00%	22.50%	52.25%
		FCM	84.00%	5.00%	44.50%
		GLQ	72.00%	30.00%	51.00%

Table 6.8: Sensitivity, specificity and quality measures for reduced feature vector.

sets	LS	GLQ	FCM
1	3.65	2.73	10.41
2	32.47	17.43	12.11
3	30.32	4.45	13.42
4	11.86	11.46	12.73
5	19.19	22.95	11.15
6	37.53	25.97	11.27
7	2.53	2.6	8.25
8	3.08	2.7	2.09
9	29.24	5.33	13.94
10	28.63	16.69	10.5
Avg.	19.85	11.23	10.59

Table 6.9: Error rates recorded for various feature vectors that provided best malignancy separation using MLP.

LS	GLQ	FCM
A_{100}	A_{100}	A_{100}
dispersion	dispersion	dispersion
histogram mean	eccentricity	textural inertia
histogram skew	convexity	textural correlation
red histogram width	vertical projection	textural homogeneity
	horizontal projection	

Table 6.10: Features chosen for classification according to the KS Test.

magnification images Fig. 6.12a shows a correctly classified case of 400x magnification image, while Fig. 6.12b and Fig. 6.12c present misclassified cases. Similar results for all of the images are summarized in Fig. 6.13, where Fig. 6.13a presents a correct classification and Fig. 6.13b and Fig. 6.13c incorrect classifications. These images can help to visualize the task of malignancy classification showing the most difficult cases that were typically misclassified. These cases are difficult to asses because the high polymorphy in the image suggests the high malignancy case but the high concentration of the nuclei shows intermediate malignancy. This cases could also be misclassified by a pathologist [89] and they most probably are the border cases.

6.12 Conclusions

In this chapter, the classification results obtained with various classifiers were presented. From the results it can be seen that the task of breast cancer malignancy classification is

Figure 6.12: Classification results for 400x images. a) Correctly classified image, b) Intermediate malignancy misclassified as high, c) High malignancy correctly classified.

Figure 6.13: Classification results for all images. a) Correctly classified image, b) Intermediate malignancy misclassified as high, c) High malignancy correctly classified.

a complex and difficult problem. Feature extraction is a very important stage and the collection of relevant features is crucial. The features presented in this study show relevance to the malignancy stages of the cancer, which is supported by the obtained classification results. Some of the classification rates could be improved if more data would be available. As already mentioned, it is very difficult to collect medical data due to legal regulations. The results presented in this section also show which segmentation algorithm is most suitable for nuclei extraction. It also important to notice that introduction of structural features extracted from low magnification images improves the classification rate. These novel features were introduced to represent the most important factors taken into consideration by a pathologist during a FNA examination. From the results obtained from the KS measure we can see that these features provide very good separation between malignancy classes. This can also be proved by the classification rates, that for set #5 are above 90% which is rather high, especially for tasks where the separation between classes isn't very obvious. When we combine the two types of features we can see that the recognition rate is high and is above 96%, which is a very good result for this type of the problem [88, 84, 86]. As a comparison, the accuracy of the pathological diagnosis for breast cancer is between

80–98% [89, 64].

It can be noticed that FNA images are difficult to analyze and some of the extracted features need sophisticated segmentation methods such as level sets or the Hough transform. In this thesis different segmentation approaches as well as classification methods were compared to show their usefulness in automated malignancy grading. Also, numerous features were extracted for that purpose. They were compared and their discriminatory powers were calculated. Feature discriminatory power tells us about their ability to represent a particular malignancy class. The process of grading the malignancy is fully automated and no user interaction is necessary, except for choosing an input image to be classified.

From the results presented here, we can estimate which of the extracted features are the most powerful and also, which classifier is the most suitable for computerized breast cancer malignancy classification. It is necessary to note that FNA images are relatively unique, and some of the techniques presented in this thesis may not be suitable when different staining standards are used. Gathering medical data for research studies is extremely difficult nowadays and no public database of images exists for comparison. To the best of our knowledge, there is only one public database described by Mangasarian *et al.* in [118] which consists of only pre–extracted features. In the absence of a proper accessible database of images, we had to construct our own database of FNA images, leading to a relatively small database since this kind of images are difficult to obtain. If we compare the results obtained by Street *et al.* and Mangasarian *et al.* [183, 118] with these presented in this thesis we can see that the obtained classification rates are promising and the malignancy grading system performs well on the FNA data.

Chapter 7

Conclusions

7.1 Open Problems and Future Research

The research presented in this thesis is concerned with the computerized breast cancer malignancy grading. To be able to efficiently classify between the malignancy grades several image processing and machine learning algorithms were used. It was shown that the choice of the technique is sufficient but can be further exploited to make the classification task better.

Up to date, there are a few open point that can be outlined here:

- Data set enlargement – for better training there is a need for a larger number of images and cases in the database.

- Feature set enlargement – it would be worth while to extract more features to test their discriminatory powers.

- Clinical tests – to be able to precisely correlate the results obtained with cross validation it is necessary to compare the performance of the proposed system with the pathological grading. This can be performed during the clinical tests.

- Single cell classification – using single cells for classification shall provide better classification.

- Occlusion handling – separation of the occluded cells shall allow for more reliable classification.

- Introduction of the low malignancy grade – to be able to classify the malignancy more precisely there is a need of low malignancy data collection to train the classifier and allow it to choose a malignancy grade other than intermediate or high.

- Implementation of the convolutional neural networks for automatic feature extraction.

- Tests on a different set of cancer data – to test if the system is applicable to other cancers it would be interesting to test on different cancers where biopsy can be taken (e.g., cervical, prostate, or kidney).

- Combination with other modalities – the system presented here depends on the staining technique and in the future it test the feature extraction on different staining techniques.

The open problems listed above can be an entry point for further research in this area. The clinical tests shall be provided to validate the system accuracy in comparison to the pathologist. There can also be additional research performed on all stages of the described system. Additional algorithms on nuclei segmentation level would need to reviewed to allow for the separation of occluded nuclei. The occlusion handling will allow for the classification of single nuclei to make the decision making process even more objective. After the segmentation, the feature extraction can be performed automatically, for example with the use of convolutional neural networks [103, 92, 127].

7.2   Summary

Computerized malignancy grading, presented here, is performed according to the Bloom–Richardson grading scheme, which is the most popular grading scheme used by pathologists around the world for this type of cancer. According to this scheme, malignancy is classified into three classes of low, intermediate and high malignancy. The discrimination between the classes is done by estimation of certain features of the fine needle aspiration biopsy

slides. The variations between these features allow for malignancy discrimination. There are two types of features taken into consideration. The first group of features describes cells ability to form groups or to spread around the image. Lower malignancy cases will have one or a few large groups of cells while the high malignancy will have cells loosely spread around the image. In this thesis a set of three features was proposed. These features take into consideration area of the segmented groups, their number and the dispersion ratio of cells. Classification and validation results show that these features provide very good separation between malignancy classes and when combined with shape–based features the classification rate increased significantly from 82.7% for 400x features to 90.19% for all features taken into consideration [84, 86, 88]. Classification results based only on the proposed features also show good performance achieving an error of 10.67%, which isn't significantly larger then that obtained for all of the features. Additional performed tests show the comparison of numerous shape–based features. These features describe shape and staining variations of the nuclei in the image. We also compared four different segmentation methods such as the Hough transform, level sets, fuzzy c–means and textural (gray level quantization) segmentation. From the results it can easily be noticed that the most obvious choice for segmentation of elliptical shapes, which is the Hough transform, didn't provide good segmentations when used without any additional supporting active contour algorithm. Out of the four compared methods, level sets provided the best nuclear boundary representations which led to better classification performance when compared to FCM and GLQ. From the presented results it can also be noticed that feature extraction also plays a very important role during malignancy classification. The proper estimation of features can have a very big impact on classification, which was shown by the results obtained with features chosen according to the KS measure. From these results we can see that the classification error for features extracted from level sets segmentations are largest out of the three segmentation methods.

From the medical point of view, the application of a system for computerized cancer malignancy grading is very important and would assist doctors with their diagnosis. Computerized malignancy grading will allow for repeatability in the decision making process, which is of a big concern among the pathological community [89]. The computerized scheme described in this thesis complies with this requirement and the obtained classification results are very good.

Bibliography

[1] R. Adams and L. Bischof. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16:641–647, 1994.

[2] A.K Agrawal, M. Jeleń, Z. Grzebieniak, P. Żukrowski, J. Rudnicki, and E. Nienartowicz. Androgen Receptors as a Prognostic and Predictive Factor in Breast Cancer. *Folia Histochemica et Cytobiologica*, 46(3):269–276, 2008.

[3] A.K Agrawal, M. Jeleń, J. Rudnicki, Z. Grzebieniak, P. Żukrowski, and E. Nienartowicz. Molecular Markers (c-erbB-2, p53) in Breast Cancer. *Folia Histochemica et Cytobiologica*, 46(4):449–455, 2008.

[4] Fluorescence microscopy, 2009. http://albertderoos.nl/?p=61.

[5] F. Albregsten, Y. Kanagasingam, G. Farrants, and H.E. Danielsen. Texture discrimination of normal and malignant mouse liver cell nuclei. *Selected papers from the 7th Scandinavian Conference on Image Analysis*, pages 324–335, 1992.

[6] H.M. Ararat and A.K.C. Wont. Textured–based image segmentation. In *Proceedings of Computer Visor and Pattern Recognition*, pages 1–7, 1986.

[7] S.C. Bagui, S. Bagui, K. Pal, and N.R. Pal. Breast Cancer Detection using Rank Nearest Neighbor Classification Rules. *Pattern Recognition*, 36(1):25–34, 2003.

[8] S.G. Baker, B.S. Krameria, and P.C. Prorok. Comparing breast cancer mortality rates before–and–after a change in availability of screening in different regions: Extension of the paired availability design. *BMC Medical Research Methodology*, 4:12, 2004.

[9] H.D. Ballard. Generalizing the Hough Transform to Detect Arbitrary Shapes. *Pattern Recognition*, 13(2):111–122, 1981.

[10] P. Bamford and B. Lovell. A Water Immersion Algorithm for Cytological Image Segmentation. In *Proceedings of the APRS Image Segmentation Workshop*, pages 75–79, University of Technology Sydney, Sydney, 1996.

[11] L. Baoli, Y. Shiwen, and L. Qin. An improved k-nearest neighbor algorithm for text categorization. In *Proceedings of the 20th International Conference on Computer Processing of Oriental Languages, ICCPOL, ShenYang, China*, pages 469–475, 2003.

[12] E. Barragn, B. Gordillo, G. Vargas, M.T. Cortez, B.E Jaramillo, S. Villa-Trevio, S. Fattel-Fazenda, J. Ortega, and L. Velazcoe. Dna replication inhibition and fluorescence microscopy of non-cationic porphyrins in malignant cells. *Archive for Organic Chemistry*, VI:436–448, 2005.

[13] C.A. Beam, P.M. Layder, and D.C. Sullivan. Variability in the Interpretation of Screening Mammograms by US Radiologists. *Arch Intern Med*, 156:209–213, 1996.

[14] P. Belhomme, A. Elmoataz, P. Herlin, and D. Bloyet. Generalized region growing operator with optimal scanning: application to segmentation of breast cancer images. *Journal of Microscopy*, 186:41–50, 1997.

[15] S. Beucher. *Segmentation d'images et morphologie mathématique*. PhD thesis, Ecole National Supérieur des Mines de Paris, 1990.

[16] S. Beucher and F. Meyer. Mathematical Morphology in Image Processing, Chapter 12. *Marcel Dekker, New York*, pages 433–481, 1992.

[17] J.C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.

[18] C.M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer–Verlag, New York, NY, 2007.

[19] H.J.G. Bloom and W.W. Richardson. Histological Grading and Prognosis in Breast Cancer. *British Journal of Cancer*, 11:359–377, 1957.

[20] R. Bondugula, O. Duzlevski, and D. Xu. Profiles and fuzzy k-nearest neighbor algorithm for protein secondary structure prediction. In *Proceeding of the 3rd Asia-Pacific Bioinformatics Conference, APBC*, pages 85–94, 2005.

[21] M.J Bottema and J.P. Slavotinek. Detection and Classification of Lobular and DCIS (small cell) Microcalcifications in Digital Mammograms . *Pattern Recognition Letters*, 21(13-14):1209–1214, 2000.

[22] L. Breiman, J. Friedman, C.J. Stone, and R.A. Olshen, editors. *Classification and Regression Trees*. Chapman & Hall, Boca Raton, 1993.

[23] N.A. Campbell and J.B. Reece. *Biology.* Benjamin Cummings, San Francisco, sixth edition, 2002.

[24] J.F. Canny. Finding edges and lines in images. *Technical report, Massachusetts Institute of Technology*, no. 720, 1983.

[25] J.F. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[26] J.R. Carr and F.P. de Miranda. The semivariogram in comparison to the co-occurrence matrix for classification of image texture. *IEEE Transactions on Geoscience and Remote Sensing*, 36(6):1945–1952, 1993.

[27] W. Chamers, T. J. Fellers, and M. W. Davidson. Illumination for stereomicroscopy: Darkfield illumination. *Nikon MicroscopyU*, 2008. http://www.microscopyu.com/articles/stereomicroscopy/stereodarkfield.html.

[28] Y.L Chang. *Fast adaptive region growing for image segmentation.* PhD thesis, University of Alberta, 1993.

[29] C.C. Chen and R.C. Dubes. Experiments in fitting discrete markov random fields to texture. *IEEE Computer Vision and Pattern Recognition*, pages 298–303, 1989.

[30] P.C. Chen and T.E. Pavlidis. Segmentation by texture using a co–occurrence matrix and split–and–merge algorithm. *Computer Vision, Graphics and Image Processing*, 10:172–182, 1979.

[31] S.Y. Chen, W.C. Lin, and C.T. Chen. Split–and–merge segmentation based on localized feature analysis and statistical tests. *Computer Vision, Graphics and Image Processing*, 53:457–475, 1991.

[32] H.D. Cheng and M. Cui. Mass Lesion Detection with a Fuzzy Neural Network. *Pattern Recognition*, 37:1189–1200, 2004.

[33] H.D. Cheng, X.Q. Li, D. Riodan, and Scrimger J.N. A Parallel Approach to Tubule Grading in Breast Cancer Lesions and its VLSI Implementation. *Computer-Based Medical Systems: Fourth Annual IEEE Symposium*, pages 322–329, 1991.

[34] H.D. Cheng, X.J. Shi, R. Min, X.P. Cai, and Du H.N. Approaches for Automated Detection and Classification of Masses in Mammograms. *Pattern Recognition*, 39(4):646–668, 2006.

[35] H.D. Cheng, J. Wang, and X. Shi. Microcalcification Detection using Fuzzy Logic and Scale Space Approaches. *Pattern Recognition*, 37(2):363–375, 2004.

[36] H.D. Cheng, C.Y. Wu, and D.L. Hung. VLSI for Moment Computation and its Application to Breast Cancer Detection. *Pattern Recognition*, 31(8):1391–1406, 1998.

[37] Y.L. Cheng and Xiaobo. Fast image region growing. *Image and Vision Computing*, 13(7):559–571, 1995.

[38] R.S. Choras. *Computer Vision*. Akademicka Oficyna Wydawnicza EXIT, 2005. Komputerowa Wizja.

[39] Zh.V. Churakova, I.B. Gurevich, I.A. Jernova, and *et al.* Selection of diagnostically valuable features for morphological analysis of blood cells. *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications*, 13(2):381–383, 2003.

[40] D. Clausi and M. Jernigan. A fast method to determine co–occurrence texture features and its applicability to image interpretation. In *26th Symposium on Remote Sensing of Environment and 18th Annual Symposium of the Canadian Remote Sensing Society*, pages 298–300, 1995.

[41] A. Collins, A. Zomorodian, G. Carlsson, and L.J. Guibas. A barcode shape descriptor for curve point cloud data. *Computers & Graphics*, 28:881–894, 2004.

[42] W.J. Conover. *Practical Nonparametric Statistics*. Wiley, New York, NY, U.S.A, 1980.

[43] J.D. Crissman, R. Makuch, and M. Budhraja. Histopathologic grading of squamous cell carcinoma of the uterine cervix. an evaluation of 70 stage ib patients. *Cancer*, 55(7):1590–1596, 2006.

[44] Cross–validation, 2008. http://research.cs.tamu.edu/prism/lectures.htm.

[45] E.R. Davis. A survey of edge detection techniques. *Computer Graphics and Image Processing*, 4:248–270, 1975.

[46] A.D.G. de Roos. Origins of introns based on the definition of exon modules and their conserved interfaces. *Bioinformatics*, 21:2–9, 2005.

[47] A.D.G. de Roos. The origin of the eukaryotic cell based on conservation of existing interfaces. *Artificial Life*, 12(4):513–523, 2006.

[48] A.D.G. de Roos. Modelling evolution on design-by-contract predicts an origin of life through an abiotic double-stranded rna world. *Biology Direct*, 2:12, 2007.

[49] M. De Santo, M. Molinara, F. Tortorella, and M. Vento. Automatic Classification of Clustered Microcalcifications by a Multiple Expert System . *Pattern Recognition*, 36(7):1467–1477, 2003.

[50] H.J. de Voogt. Rapid urinary cytology by phase contrast microscopy a preliminary report. *Urological Research*, pages 113–119, 1973.

[51] P.A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice–Hall, London, 1982.

[52] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition (Stochastic Modelling and Applied Probability)*. Springer–Verlag, New York, NY, 1997.

[53] M.F. Dixon, I.G. Martin, H.M. Sue-Ling, J.I. Wyatt, P. Quirke, and D. Johnston. Goseki grading in gastric cancer: comparison with existing systems of grading and its reproducibility. *Histopathology*, 25(4):309–316, 2007.

[54] R.C Dubes and A.K. Jain. Random field models in image analysis. *Journal of Applied Statistics*, 16(2):131–164, 1989.

[55] R.O Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley Interscience Publishers, New York, 2nd edition, 2000.

[56] C.R. Dyer and A. Rosenfeld. Thinning algorithms for grayscale pictures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(1):88–89, 1979.

[57] J. Estevez, S. Alayon, and L. Moreno. Cytological Breast Cancer Fine Needle Aspirate Images Analysis with a Genetic Fuzzy Finite State Machine. *Conference Board of the Mathematical Sciences, CBMS 2002*, pages 21–26, 2002.

[58] L.M.J. Florack and A. Kuijper. The topological structure of scale–space images. *Journal of Mathematical Imaging and Vision*, 12(1):65–80, 2000.

[59] D.A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice–Hall, London, 2008.

[60] Y. Freund and L. Masin. The alternating decision tree learning algorithm. In *Proceedings of the 16th International Conference on Machine Learning*, pages 124–133, 1999.

[61] T.T. Friess, N. Cristianini, and C. Campbell. The kernel adatron algorithm: a fast and simple learning procedure for support vector machines. In *15th International Conference on Machine Learning, Morgan Kaufman Publishers*, 1998.

[62] X. Game, M. Souli, A.M. Fontanilles, J.M. Benoit, J.X. Corberand, and P. Plante. Comparison of red blood cell volume distribution curves and phase–contrast microscopy in localization of the origin of hematuria. *Urology*, 61(3):507–511, 2003.

[63] M.F. Garcia-Parajoy, J.A. Veerman, S.J.T van Noort, B.G. de Grooth, J. Greve, and N.F van Hulst. Near-field optical microscopy for dna studies at the single molecular level. *Bioimaging*, 6:43–53, 1998.

[64] R.W. Giard and J. Hermans. The Value of Aspiration Cytologic Examination of the Breast. A Statistical Review of the Medical Literature. *Cancer*, 69:2104–2110, 1992.

[65] G. Glab, K. Florczak, J. Jaronski, and T. Licznerski. *Ginecological Cytodiagnosis in Contrast–Phase Microscope*. Blackhorse, 2001. Cytodiagnostyka Ginekologiczna w Mikroskopie Kontrastowofazowym.

[66] W.B. Goh. Strategies for shape matching using skeletons. *Computer Vision and Image Understanding*, 110(3):326–345, 2008.

[67] W.B. Goh and K.Y. Chan. Shape description using gradient vector field histograms. In *Scale space methods in computer vision: ('Isle of Skye', 10-12 June 2003)*, volume 2695, pages 713–728. 2003.

[68] R.C. Gonzalez and R.E. Woods. *Digital Image Processing (3rd Economy Edition)*. Prentice–Hall, London, 2008.

[69] J. Grabska-Chrzastowska, W. Libuszowski. and W. Tomalak. Comparison of the KNN Method and the Neural Network Technique for Spirometry Classification. *Bio-Algorithms and Med-Systems*, 1(1/2):9–12, 2005. Porownanie Metody KNN i Techniki Sieci Neuronowych w Klasyfikacji Badan Spirometrycznych.

[70] W.M. Grohman and A.P. Dhawan. Fuzzy Convex Set-based Pattern Classification for Analysis of Mammographic Microcalcifications. *Pattern Recognition*, 34(7):1469–1482, 2001.

[71] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer. An knn model-based approach and its application in text categorization. *Lecture Notes in Computer Science*, 2945/2004:559–570, 2004.

[72] I. Gurevich, D. Kharazishvili, D. Murashov, O. Salvetti, and I. Vorobjev. Technology for automated morphologic analysis of cytological slides. methods and results. In *The 18th International Conference on Pattern Recognition. ICPR'06. Hong Kong, China*, pages 711–714, 2006.

[73] I. Gurevich and D. Murashov. Method for early diagnostics of lymphatic system tumors on the basis of the analysis of chromatin constitution in cell nucleus images. In *The 17th International Conference on Pattern Recognition. ICPR'04, Cambridge, UK*, pages 806–809, 2004.

[74] J. Hallinan and P. Jackway. Detection of Malignancy Associated Changes in Thionin Stained Cervical Cells. *Digital Image Computing: Techniques and Applications, DICTA95*, pages 426–431, 1995.

[75] R.M. Haralick. Statistical and structural approaches to texture. *IEEE Trans. Syst. Man Cyber*, 67(5):786–804, 1979.

[76] R.M. Haralick. Ridge and valley on digital images. *Computer Vision, Graphics and Image Processing*, 22:28–38, 1983.

[77] R.M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Trans. Syst. Man Cyber*, 3(6):610–621, 1973.

[78] S. Haykin. *Neural Networks and Learning Machines*. Prentice Hall, 2008.

[79] T. Hothorn and B. Lausen. Bagging tree classifiers for laser scanning images: a data- and simulation-based strategy. *Artificial intelligence in medicine*, 27(1):65–79, 2003.

[80] P.V.C. Hough. Method and Means for Recognizing Patterns, US Patent 3969654, 1962.

[81] M.K. Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, pages 179–187, 1962.

[82] H.J. Huang and C.N. Hsu. Bayesian classification for data from the same unknown class. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 32(2):137–145, 2002.

[83] P.A. Humphrey. Gleason grading and prognostic factors in carcinoma of the prostate. *Modern Pathology*, 17:292–306, 2004.

[84] L. Jeleń, T. Fevens, and A. Krzyżak. Classification of breast cancer malignancy using cytological images of fine needle aspiration biopsies. *Int. J. Math. Comput. Sci.*, 18(1):75–83, 2008.

[85] L Jeleń, T. Fevens, and A. Krzyżak. Influence of nuclei segmentation on breast cancer malignancy classification. In *Proceedings of SPIE*, volume 7260, 2009. in press.

[86] L Jeleń, T. Fevens, A. Krzyżak, and M. Jeleń. Discriminatory power of cells grouping features for breast cancer malignacy classification. In *Proceedings of the International Federation for Medical and Biological Engineering*, volume 21(3), pages 559–562. Springer Berlin/Heidelberg, 2008.

[87] L. Jeleń, A. Krzyżak, and T. Fevens. Automated feature extraction for breast cancer grading with Bloom-Richardson scheme. *International Journal of Computer Assisted Radiology and Surgery*, 1(1):468–469, 2006.

[88] L Jeleń, A. Krzyżak, and T. Fevens. Comparison of pleomorphic and structural features used for breast cancer malignancy classification. *Lecture Notes in Computer Science, Advances in Artificial Intelligence*, 5032/2008:138–149, 2008.

[89] M. Jeleń. Personal communication. 2005.

[90] G. H. John and P. Langley. Estimating continous distributions in bayesian classifiers. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, Montreal, Canada*, pages 338–345, 1995.

[91] M. Jun, W. Cheng-Ke, and L. Xin-Ru. A fast shape descriptor. *Computer Vision, Graphics and Image Processing*, 34(3):282–291, 1986.

[92] K. Kavukcuoglu, M.A. Ranzato, R. Fergus, and Y. LeCun. Learning invariant features through topographic filter maps. In *Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR'09)*. IEEE, 2009.

[93] J.V. Kittler, R. Marik, M. Mirmehdi, M. Petrou, and J. Song. Detection of defects in colour texture surfaces. In *Machine Vision Application*, pages 558–567, 1994.

[94] G.J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, New Jersey, 1995.

[95] R. Kohavi. Scaling up the accuracy of naive–bayes classifiers: a decision tree hybrid. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, pages 202–207, 1996.

[96] R. Kohavi. A study of cross–validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the fourteenth international joint conference on artificial intelligence, Morgan Kaufman, San Mateo*, pages 1137–1143, 2006.

[97] R. Kohavi and F. Provost. Glossary of terms. *Machine Learning*, 30(2–3):271–274, 1998.

[98] T. Kohonen. The self–organizing map. In *Proceedings of IEEE*, volume 78, pages 1464–1480, 1990.

[99] T.M. Kolb, J. Lichy, and Newhouse J.H. Occult cancer in women with dense breasts: detection with screening US–diagnostic yield and tumor characteristics. *Radiology*, 207:191–199, 1998.

[100] R. Kondepudy and G. Healey. Use of invariants for recognition of three-dimensional color textures. *Journal of the Optical Society of America A*, 11(11):3037–3049, 1994.

[101] A. Lapedes and R. Farber. Nonlinear signal processing using neural networks: prediction, and system modelling. *Technical Report*, LA–UR–87–2662, 1987.

[102] L.J. Latecki, L. Lakamper, and E. Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 424–429, 2000.

[103] Y. LeCun and Y. Bengio. Convolutional networks for images, speech, and time-series. In M. A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*. MIT Press, 1995.

[104] K.M. Lee and W.N. Street. A Fast and Robust Approach for Automated Segmentation of Breast Cancer Nuclei. In *Proceedings of the Second IASTED International Conference on Computer Graphics and Imaging*, pages 42–47, Palm Springs, CA, 1999.

[105] K.M Lee and W.N. Street. Model-based Detection, Segmentation and Classification for Image Analysis using On-line Shape Learning. *Machine Vision and Applications*, 13(4):222–233, 2003.

[106] Y. Lee and C.K. Lee. Classification of Multiple Cancer Types by Multicategory Support Vector Machines Using Gene Expression Data. *Bioinformatics*, 19(9):1132–1139, 2003.

[107] A. Lehmussola, J. Selinummi, P. Ruusuvuori, A. Niemisto, and O. Yli-Harja. Simulating fluorescent microscope images of cell populations. In *27th Annual International Conference of the Engineering in Medicine and Biology Society, IEEE-EMBS 2005, Shanghai, China*, pages 3153–3156, 2005.

[108] Leica confocal microscope, 2009. http://www.core-facility.uni-freiburg.de/intro

[109] O. Lezoray, A. Elmoataz, H. Cardot, G. Gougeon, M. Lecluse, and M. Revenu. Segmentation of cytological images using color and mathematical morphology. *European Conference on Stereology, Amsterdam, Netherlands*, page 52, 1998.

[110] C. Li, C. Xu, C. Gui, and M.D. Fox. Level Set Evolution Without Re-initialization: A New Variational Formulation. *IEEE Proceedings of Conference on Computer Vision and Pattern Recognition 2005*, pages 430–436, 2005.

[111] L. Li, D.M. Umbach, P. Terry, and J.A. Taylor. Application of the ga/knn method to seldi proteomics data. *Bioinformatics*, 20(10):1638–1640, 2004.

[112] X. Li and N. Ye. Decision tree classifiers for computer intrusion detection. *Real-time system security*, pages 77–93, 2003.

[113] Z.S. Li. Markov random fields models in computer vision. *European Conference on Computer Vision, Stockholm, Sweden. ECCV 1995*, pages 361–370, 1994.

[114] D. Liu, T. Shi, J.A. DiDonato, J.D. Carpten, J. Zhu, and Z.H. Duan. Application of genetic algorithm/k-nearest neighbor method to the classification of renal cell carcinoma. In *IEEE Proceedings of the Computational Systems Bioinformatics Conference, CSB, Stanford, CA, USA*, pages 558–559, 2004.

[115] M.A. MacAulay, J.N. Scrimger, D. Riodan, and H.D. Cheng. An Interactive Graphics Package with Standard Examples of the Bloom and Richardson Histological Grading Technique. *Computer-Based Medical Systems: Fourth Annual IEEE Symposium*, pages 108–112, 1991.

[116] P. Maillard. Comparing texture analysis methods through classification. *American Society for Photogrammetry and Remote Sensing*, 69(4):357–367, 2003.

[117] W. Malina, S. Ablameyko, and W. Pawlak. *Basics of the Digital Image Processing*. Akademicka Oficyna Wydawnicza EXIT, Warszawa, Poland, 2002. Podstawy cyfrowego przetwarzania obrazów.

[118] O.L. Mangasarian, R. Setiono, and W.H. Wolberg. Pattern Recognition via Linear Programming: Theory and Application to Medical Diagnosis. *Large-Scale Num. Opt., Philadelphia: SIAM*, pages 22–31, 1990.

[119] O.L Mangasarian, W.N. Street, and W.H. Wolberg. Breast Cancer Diagnosis and Prognosis via Linear Programming. *Operations Research*, 43(4):570–576, 1994.

[120] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London*, 207:187–217, 1980.

[121] S. McKenna. *Automated Analysis of Papanicolaou Smears*. PhD thesis, University of Dundee, 1994.

[122] A. McWilliams, C. MacAulay, A.F. Gazdar, and S. Lam. Innovative molecular and imaging approaches for the detection of lung cancer and its precursor lesions. *Oncogene*, 21:6949–6959, 2002.

[123] Phase contrast microscopes – microbus, 2009. http://www.microscope-microscope.org/advanced/phase-contrast-microscope.htm.

[124] G. Miniello. Phase contrast microscopy: A new dimension to pathophysiology of cervicovaginal epithelium. *New Frontiers in Obstetric & Gynaecology*, 1995.

[125] G. Miniello. Integrated colposcopy and phase contrast microscopy. *Przeglad Lekarski*, 53:28–33, 1996.

[126] M. Minsky. Memoir on inventing the confocal scanning microscope. *Scanning*, 10:128–138, 1988.

[127] P. Mirowski, Y. LeCun, D. Madhavan, and R. Kuzniecky. Comparing svm and convolutional networks for epileptic seizure prediction from intracranial eeg. In *Proceedings of Machine Learning and Signal Processing (MLSP'08)*. IEEE, 2008.

[128] C. Mori, H. Hashimoto, and K. Hoshino. Fluorescence microscopy of nuclear dna in oocytes and zygotes during in vitro fertilization and development of early embryos in mice. *Biology of Reproduction*, 39:737–742, 1988.

[129] D.B. Murphy, R. Oldfield, and S. Schwartz. Phase contrast microscopy. *Nikon MicroscopyU*, 2008. http://www.microscopyu.com/articles/phasecontrast/phasemicroscopy.html.

[130] S. Naicker, V. Poovalingaimy, K. Mlisana, N. Christopher, D. Nadesan, and Y.K. Seedat. Comparative assessment of phase contrast microscopy and coulter counter measurements in localizing the site of haematuria. *South African Medical Journal*, 82:183–185, 1992.

[131] S. Naik, S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski. Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. In *5th IEEE International Symposium of Biomedical Imaging: From Nano to Macro, ISBI 2008. Paris, France*, pages 284–287, 2008.

[132] V.S. Nalva. *A Guided Tour of Computer Vision*. AddisonWesley, 1993.

[133] P. Nathangle, W. Chambers, and M. W. Davidson. Introduction to stereomicroscopy. *Nikon MicroscopyU*, 2008. http://www.microscopyu.com/articles/stereomicroscopy/stereointro.html.

[134] Neuro Solution, 2008. http://www.neurosolutions.com.

[135] R. Nezafat, A. Tabesh, S. Akhavan, C. Lucas, and M.A. Zia. Feature Selection and Classification for Diagnosing Breast Cancer. *Proceedings of International Association of Science and Technology for Development International Conference*, pages 310–313, 1998.

[136] O. Noriko, Y. Katsumi, K. Mitsuo, and S. Hiroshi. Comparison between optical microscopic examination and phase contrast microscopic examination for diagnosing the origin of urinary bleeding. *Japanese Journal of Nephrology*, 48(5):401–406, 2006.

[137] H. Ogawa and K. Taniguchi. Thinning and stroke segmentation for handwritten chinese recognition. *Pattern Recognition*, 15:229–308, 1982.

[138] E. Oja. A simplified neuron modelled as a principal component analyzer. *Journal of mathematical biology*, 15:267–273, 1982.

[139] Optics of a fluorescence microscope, 2009.
http://www.fluorescence.com/tutorial/fm-optic.htm.

[140] S. Osher and J.A. Sethian. Fronts Propagating with Curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *J Comp. Phys.*, 79:12–49, 1988.

[141] F. Pernkopf. Bayesian network classifiers versus selective k-nn classifier. *Pattern Recognition*, 38(1):1–10, 2005.

[142] M. Peura and J. Iivarinen. Efficiency of simple shape descriptors. *Aspects of Visual Form*, pages 443–451, 1997.

[143] D.W. Piston. Imaging living cells and tissues by two-photon excitation microscopy. *Trends in Cellular Biology*, 9(2):66–69, 1999.

[144] K. Polat and S. Güneş. A novel hybrid intelligent method based on c4.5 decision tree classifier and one-against-all approach for multi-class classification problems. *Expert Systems with Applications: An International Journal*, 36(2):1587–1592, 2009.

[145] R. Poulsen and I. Pedron. Region of Interest Finding in Reduced Resolution Colour Imagery–Application to Cancer Cell Detection. *Pattern Recognition*, 28(11):1645–1655, 1995.

[146] W.K. Pratt. *Digital Image Processing: PIKS Scientific Inside.* Wiley, Hoboken, NJ, 2007.

[147] QinetiQ. Automated Histopathology Breast Cancer Analysis and Diagnisis System. *Data Sheet*, 2005. http://www.qinetiq.com/.

[148] J.R. Quinlan. *C4.5: Programs for Machine Learning.* Morgan Kaufmann, San Mateo, CA, 1993.

[149] S. Racki, M. Grzetic, Z. Prodan-Merlak, S. Vuksanovic-Mikulicic, B. Sladoje-Martinovic, and S. Zivcic-Cosic. Clinical use of phase-contrast microscopy in the differential diagnosis of microhematuria. *Acta Medica Croatia*, 57(1):11–16, 2003.

[150] R.M. Rangayyan. *Biomedical Image Analysis (Biomedical Engineering).* CRC Press, Boca Raton, FL, 2004.

[151] M.L. Raymer, T.E. Doom, L.A. Kuhn, and W.F. Punch. Knowledge discovery in medical and biological datasets using a hybrid bayes classifier/evolutionary algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 33(5):802–813, 2003.

[152] J.W. Reagan and Y.S. Fu. Histologic types and prognosis of cancers of the uterine cervix. *International Journal of Radiation Oncology, Biology, Physics*, 5:1015–1020, 1979.

[153] N. Rezai. Confocal microscopy – a visual slice of the cellular world. *The Science Creative Quarterly*, 3, 2004. http://www.scq.ubc.ca/confocal-microscopy-a-visual-slice-of-the-cellular-world.

[154] T.W. Ridler and S. Calvard. Picture thresholding using an iterative selection. *IEEE Transactions on System, Man and Cybernetics*, 8:630–632, 1978.

[155] B.D. Ripley. *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge, 2008.

[156] K. Rodenacker. Applications of topology for evaluating pictorial structures. *Theoretical Foundations of Computer Vision, Akademie–Verlag, Berlin*, pages 35–46, 1993.

[157] K. Rodenacker. Quantitative microscope image analysis for improved diagnosis and prognosis of tumors in pathology. *Creaso Info Medical Imaging, Creaso GmbH, Gilching*, 22, 1995.

[158] K. Rodenacker and E. Bengtsson. A feature set for cytometry on digitized microscopic images. *Anal Cell Pathol*, 25(1):1–36, 2003.

[159] C.J. Roe, R.H. Whiteheads, R.G. Walker, J.P. Dowling, I. Birchall, and P.S. Kincaid-Smith. Morphology of renal parenchymal culture in collagen gels. *Pathology*, 23(1):25–29, 1991.

[160] A. Rosenfeld. *Picture Processing by Computer*. Academic Press, London, 1969.

[161] A. Rosenfeld and L.S. Davis. A note on thinning. *IEEE Transactions on Systems, Man, and Cybernetics*, 6:226–228, 1976.

146

[162] P.L. Rosin. Measuring shape: Ellipticity, rectangularity, and triangularity. In *15th International Conference on Pattern Recognition*, pages 179–187, 2000.

[163] K.F.A. Ross. *Phase Contrast and Interference Microscopy for Cell Biologists*. Edward Arnold, 1967.

[164] D.E. Rumelhart, G.E. Hinton, and R.J Williams. Learning internal representatons by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1, pages 318–362, 1986.

[165] S.R. Safavian and D. Landgrebe. A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(3):660–674, 1991.

[166] E. Salari and P. Siy. The ridgeseeking method for obtaining the skeleton of digital images. *IEEE Transactions on Systems, Man, and Cybernetics*, 14(3):524–528, 1984.

[167] T. Sanger. Optimal unsupervised learning in a single layer linear feedforward network. *Neural Network*, 12:459–473, 1989.

[168] F. Schnorrenberg, C.S. Pattichis, K. Kyriacou, and C.N. Schizas. Detection of Cell Nuclei in Breast Cancer Biopsies using Receptive Fields. *IEEE Proceedings of Engineering in Medicine and Biology Society*, pages 649–650, 1994.

[169] F. Schnorrenberg, C.S. Pattichis, K. Kyriacou, and C.N. Schizas. Content-based Description of Breast Cancer Biopsy Slides. *Proc. Intl. EuroPACS Mtg.*, pages 136–140, 1996.

[170] F. Schnorrenberg, C.S. Pattichis, K. Kyriacou, M. Vassiliou, and C.N. Schizas. Computer-aided Classification of Breast Cancer Nuclei. *Technology & Health Care*, 4(2):147–161, 1996.

[171] F. Schnorrenberg, N. Tsapatsoulis, C.S. Pattichis, C.N. Schizas, S. Kollias, M. Vassiliou, A. Adamou, and K. Kyriacou. A modular neural network system for the analysis of nuclei in histopathological sections. *IEEE Engineering in Medicine and Biology Magazine*, 19:48–63, 2000.

[172] S. Schüpp, A. Elmoataz, J. Fadili, P. Herlin, and D. Bloyet. Image segmentation via multiple active contour models and fuzzy clustering with biomedical applications.

In *The 15th International Conference on Pattern Recognition, ICPR'00, Barcelona, Spain*, volume 1, pages 622–625, 2000.

[173] J.A. Sethian and D. Adalsteinsson. An Overview of Level Set Methods for Etching, Deposition, and Lithography Development. *IEEE Transactions on Semiconductor Manufacturing*, 10(1):167–184, 1997.

[174] L. Shapiro and G. Stockman, editors. *Computer Vision*. Prentice Hall, 2002.

[175] H. Sherman. A quasitopological method for the recognition of line patterns. In *IEEE International Conference on Image Processing, ICIP59*, pages 232–238, 1959.

[176] A.C. Silva, A.C. de Paiva, and A.C.M. de Oliveira. Comparison of FLDA, MLP and SVM in Diagnosis of Lung Nodule. *MLDM 2005*, pages 285–294, 2005.

[177] A.C. Silva, A.C. Paiva, P.C.P Carvalho, and M. Gattass. Semivariogram and sgldm methods comparison for the diagnosis of solitary lung nodule. *Lecture Notes in Computer Science*, 3523:479–486, 2005.

[178] M. Smereka. *Computer Based Analysis of Cytological Images from Contrast–Phase Microscopy*. PhD thesis, Wroclaw University of Technology, Poland, 2006. Komuterowa analiza cyfrowych obrazow cytologicznych pochodzacych z mikroskopii kontrastowo–fazowej.

[179] R.W. Smith. Computer processing of line images: a survey. *Pattern Recognition*, 20:7–15, 1987.

[180] American Cancer Society. Breast Cancer Facts & Figures 2007–2008. *Atlanta: American Cancer Society, Inc.*, 2004.

[181] K.Y. Song, J.V. Kittler, and M. Petrou. Defect detection in random color textures. *Image and Visual Computing*, 14(9):667–683, 1996.

[182] N.W. Street. *Cancer Diagnosis and Prognosis via Linear-Programming-Based Machine Learning*. PhD thesis, University of Wisconsin, 1994.

[183] N.W. Street. Xcyt: A System for Remote Cytological Diagnosis and Prognosis of Breast Cancer. In L.C. Jain, editor, *Soft Computing Techniques in Breast Cancer*

Prognosis and Diagnosis, pages 297–322, Singapore, 2000. World Scientific Publishing.

[184] W. N. Street, W. H. Wolberg, and O. L. Mangasarian. Nuclear Feature Extraction for Breast Tumor Diagnosis. In *Imaging Science and Technology/Society of Photographic Instrumentation Engineers 1993 International Symposium on Electronic Imaging: Science and Technology*, volume 1905, pages 861–870, San Jose, California, 1993.

[185] T. Suk and J. Flusser. Graph method for generating affine moment invariants. In *17th International Conference on Pattern Recognition, ICPR '04*, pages 192–195, 2004.

[186] J. Suri, L. Liu, S. Singh, S. Laxminarayan, X. Zeng, and L. Reden. Shape recovery algorithms using levels sets in 2-D/3-D Medical Imagery: A state-of-the-art review. *IEEE Transactions on Information Technology in Biomedicine*, 6(1), 2002.

[187] H. Tamura. A comparison of line thinning algorithms from digital geometry viewpoint. *International Joint Conference on Pattern Recognition, ICPR '78*, pages 715–719, 1978.

[188] S.C. Tan and J.V. Kittler. Color texture analysis using color histogram. *IEE Proceedings Vision, Image & Signal Processing*, 141(6):403–412, 1994.

[189] R.W. Taylor, M. Savani, and A.P. Reeves. A fast method to determine co–occurrence texture features and its applicability to image interpretation. *Computer Vision, Graphics and Image Processing*, 45:42–60, 1989.

[190] Techniques in the microscopy suite, 2009. http://www.itg.uiuc.edu/exhibits/gallery/images/confocal/xenopus.jpg.

[191] N. Theera-Umpon. Patch–Based white blood cell nucleus segmentation using fuzzy clustering. *ECTI Transactions on electrical eng., electronics and communications*, 3(1):15–19, 2005.

[192] S. Theodoridis and K. Koutroumbas. *Pattern Recognition, Fourth Edition*. Academic Press, London, 2008.

[193] V. Torre and T.A. Poggio. On edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(2):147–163, 1986.

[194] E. Trucco and A. Verri. *Introductory Techniques 3-D Computer Vision.* Prentice-Hall, London, 1998.

[195] S.E. Umbaugh. *Computer Imaging: Digital Image Analysis and Processing.* CRC Press, New York, 2005.

[196] S.E. Umbaugh, Y. Wei, and M. Zuke. Feature extraction in image analysis. *IEEE Engineering in Medicine and Biology*, 16(4):62–73, 1997.

[197] H.L Jr. Walker, L. Albertelli, Y. Titkov, P. Kaltsatis, and G. Seburyano. Evolution of Neural Networks for the Detection of Breast Cancer. *Proceedings of International Joint Symposia on Intelligence and Systems*, pages 34–40, 1998.

[198] H.L Jr. Walker and L.E Albertelli. Breast Cancer Screening Using Evolved Neural Networks. *IEEE International Conference on Systems, Man, and Cybernetics*, 2:1619–1624, 1998.

[199] B. Weyn, G. Van de Wouwer, M. Koprowski, and *et al.* Value of morphometry, texture analysis, densitometry and histometry in the differential diagnosis and prognosis of malignant mesthelioma. *Journal of Pathology*, 4(189):581–589, 1999.

[200] B. Weyn, G. van de Wouwer, A. van Daele, P Scheunders, D. van Dyck, E. van Marck, and W. Jackob. Automated Breast Tumor Diagnisis and Grading Based on Wavelet Chromatin Texture Description. *Cytometry*, 33:32–40, 1998.

[201] Fluorescence microscopy – wikipedia, 2009.
http://en.wikipedia.org/wiki/Fluorescence_microscopy.

[202] Phase contrast microscopy – wikipedia, 2009.
http://en.wikipedia.org/wiki/Phase_contrast_microscopy.

[203] W. H. Wolberg, W. N. Street, and O. L. Mangasarian. Breast Cytology Diagnosis Via Digital Image Analysis. *Analytical and Quantitative Cytology and Histology*, 15(6):396–404, December 1993.

[204] W. H. Wolberg, W. N. Street, and O. L. Mangasarian. Machine Learning Techniques to Diagnose Breast Cancer from Image-Processed Nuclear Features of Fine Needle Aspirates. *Cancer Letters*, 77:163–171, 1994.

[205] W.H Wolberg and O.L. Mangasarian. Multisurface Method of Pattern Separation for Medical Diagnosis Applied to Breast Cytology. *Proceedings of National Academy of Science, USA*, 87:9193–9196, 1990.

[206] Q. Xu, J. Yang, and S. Ding. Color texture analysis using the wavelet-based hidden markov model. *Pattern Recognition Letters*, 26(11):1710–1719, 2005.

[207] I.T. Young, P. Verbeek, and B.H. Mayall. Characterization of chromatin distribution in cell nuclei. *Cytometry*, 7(5):467–474, 1986.

[208] P. Zhang, B. Verma, and K. Kumar. Neural vs. Statistical Classifier in Conjunction with Genetic Algorithm Based Feature Selection. *Pattern Recognition Letters*, 26(7):909–919, 2005.

[209] H. Zinser and H. Runge. *PhasencontrastMicroscopie. Klinische Fortschritte in der Gynacologie.* Urban und Schwartzenberg, 1954.

[210] M. Zribi. Unsupervised bayesian image segmentation using orthogonal series. *Journal of Visual Communication and Image Representation*, 18(6):496–503, 2007.

[211] S.W. Zucker. Survey, region growing: childhood and adolescence. *Computer Vision, Graphics and Image Processing*, 5:382–399, 1976.

[212] J.Y Zunic and P.L. Rosin. A Convexity Measurement for Polygons. *British Machine Vision Conference*, 24:173–182, 2002.

Appendix A

Database of Fine Needle Aspirates

This section provides a set of images that were used in this research. The database was divided into two subsets depending on the magnification of the slides. The first subset consists of the image that were recorded with 100x magnification. The second subset contains of 400x magnification images.

A.1 100x Magnification Images

A.1.1 Intermediate Malignancy (G2)
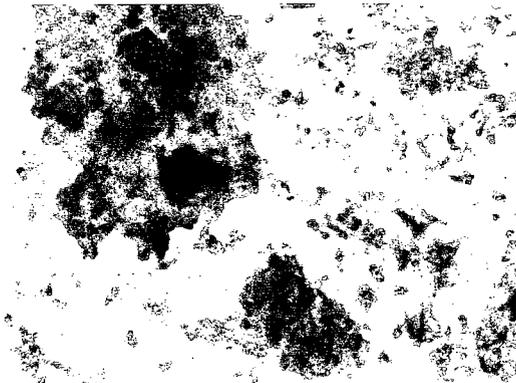
12071

152
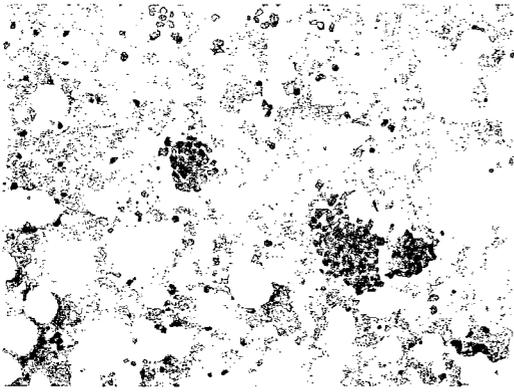
170
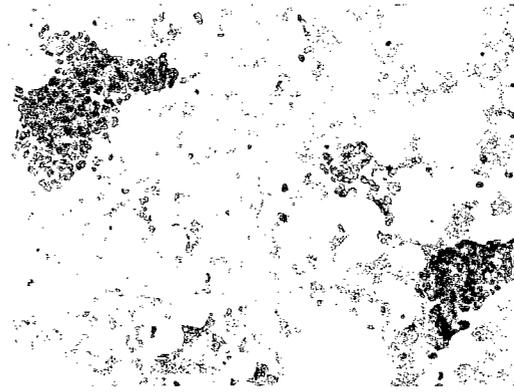
170-1
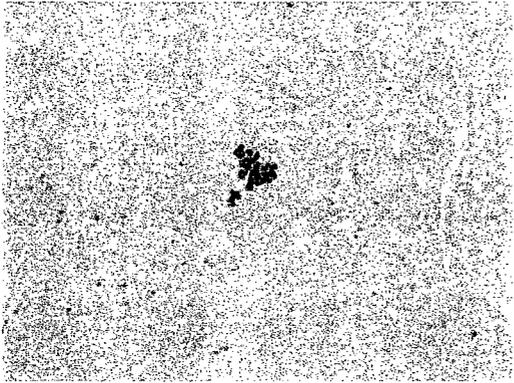
176
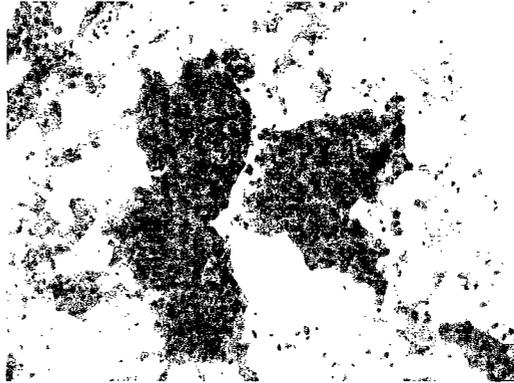
195
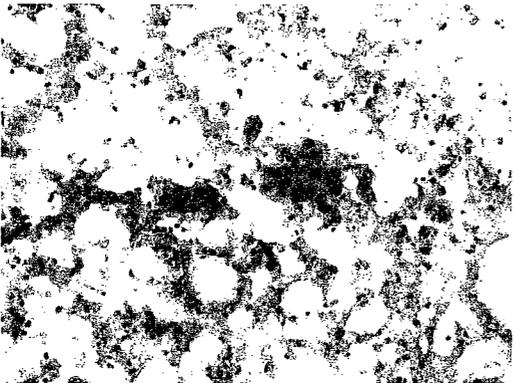
195-1
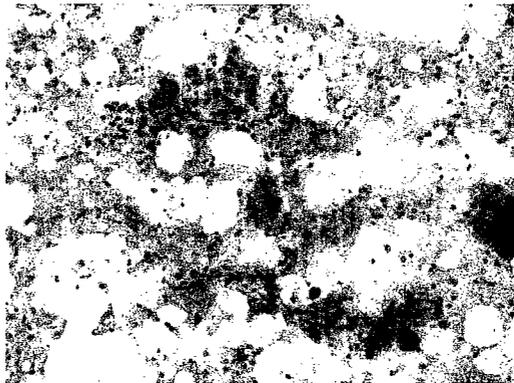
195-2
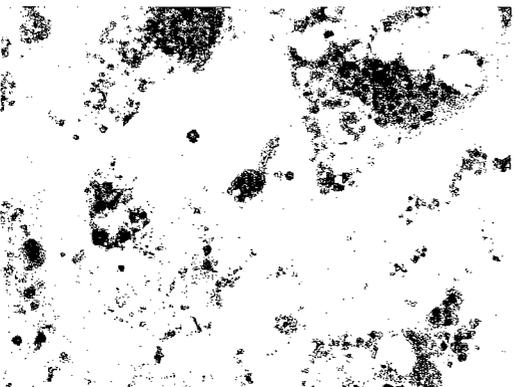
223
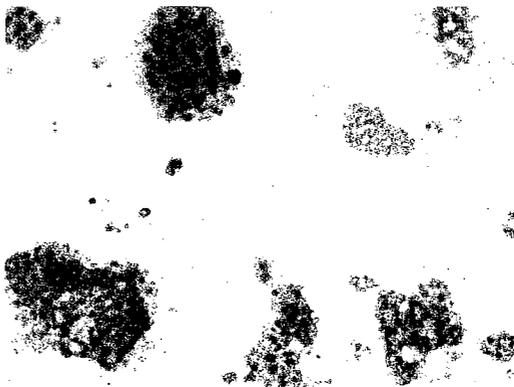
223-1
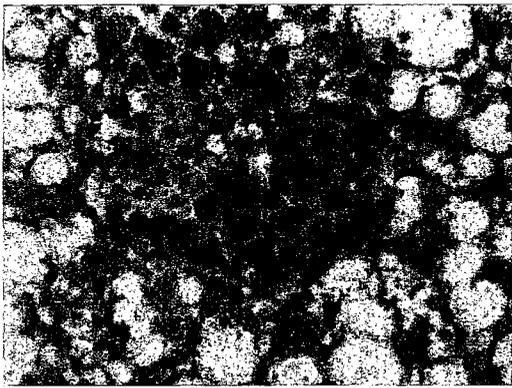
234
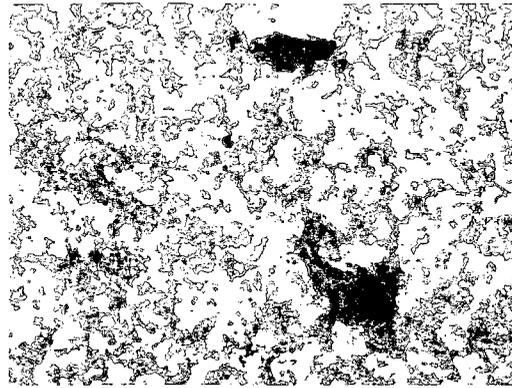
34
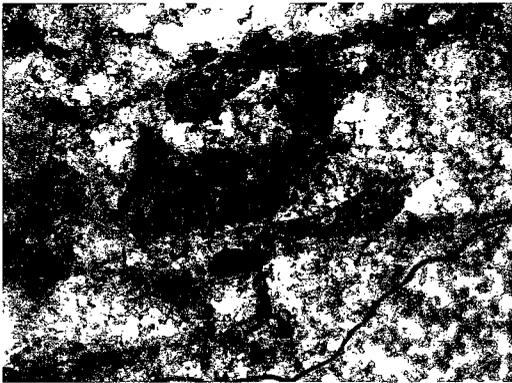
34-1
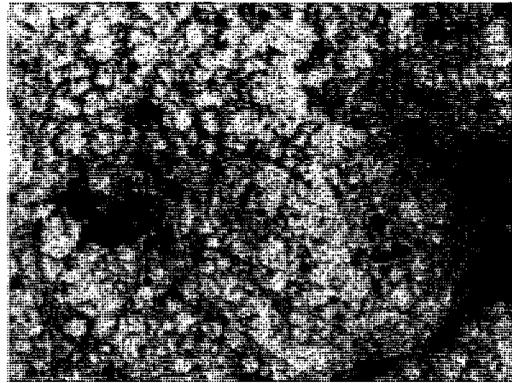
365
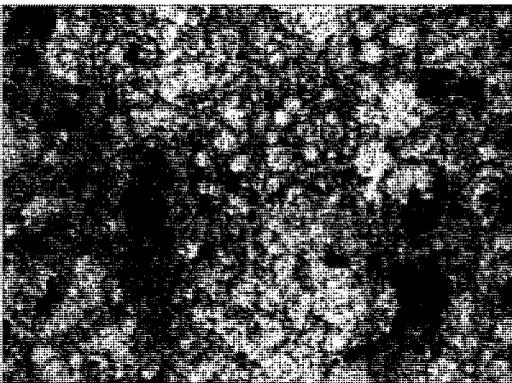
365-1
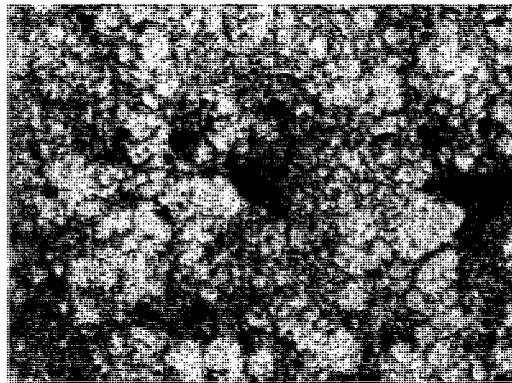
365-2
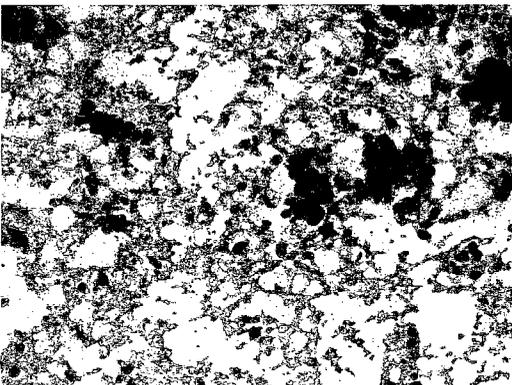
365-3
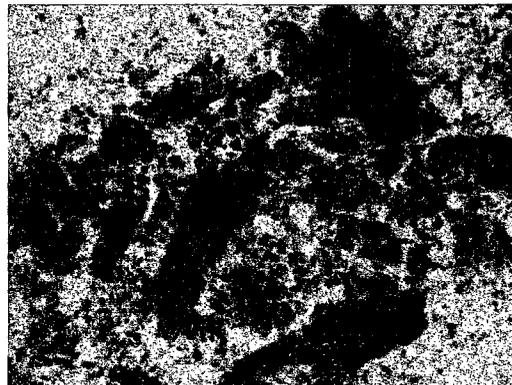
385
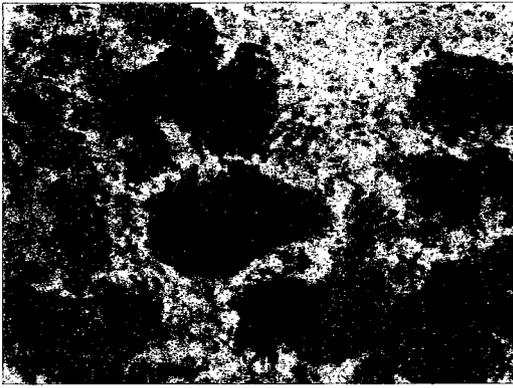
385-1
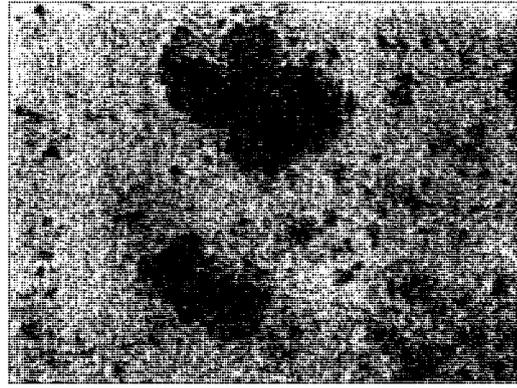
385-2
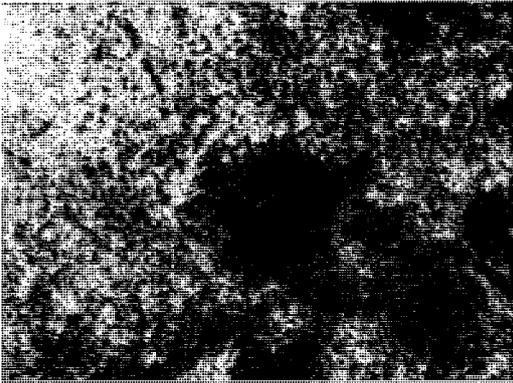
385-3
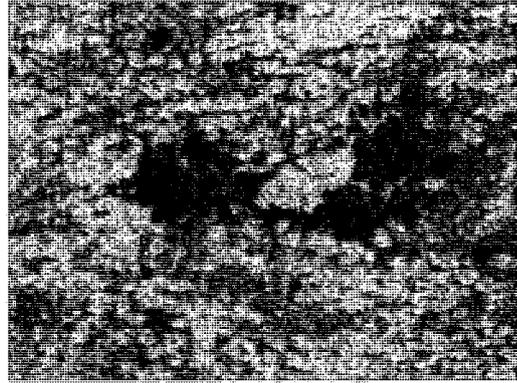
400
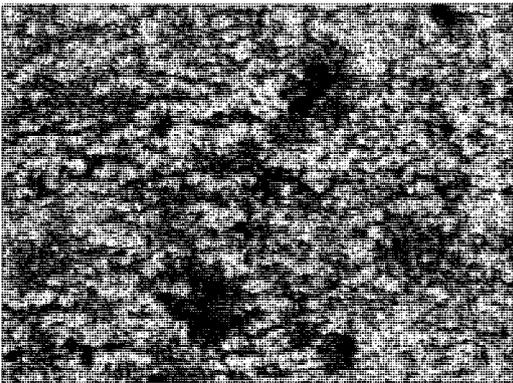
400-1
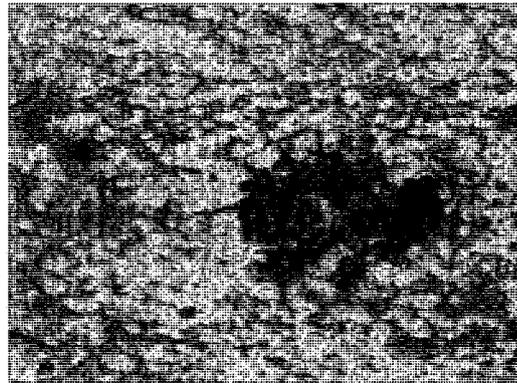
400-2
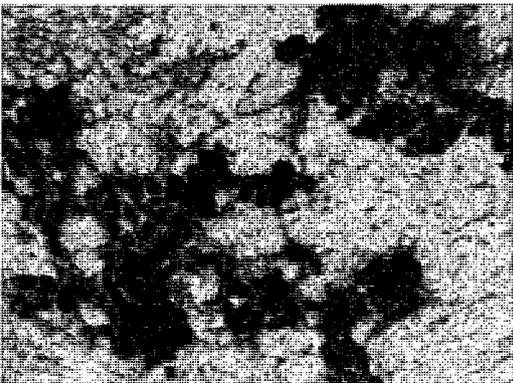
400-3
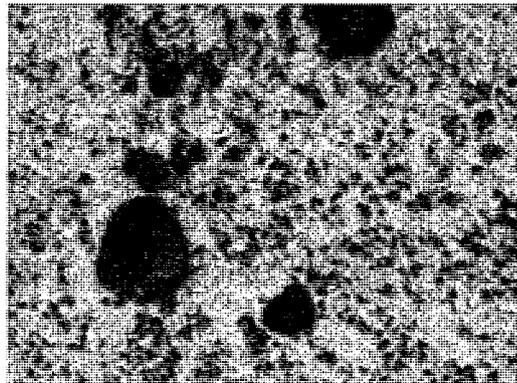
408
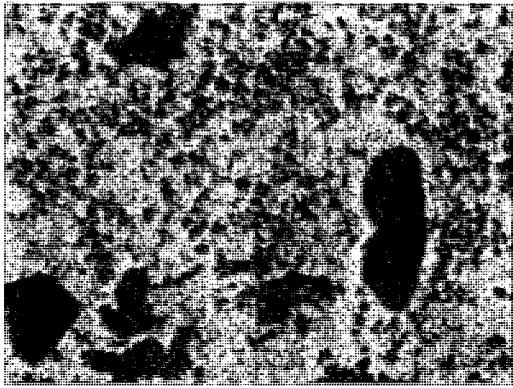
408–1
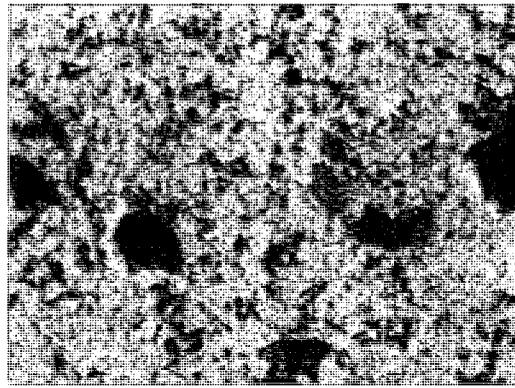
408–2
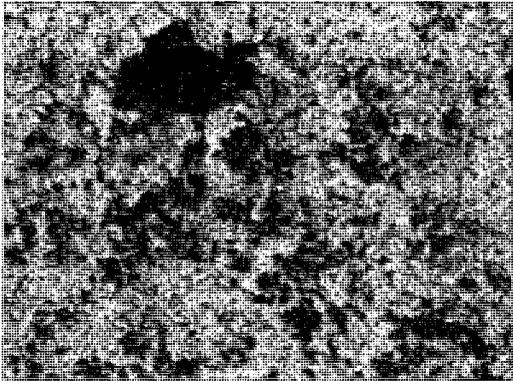
408–3
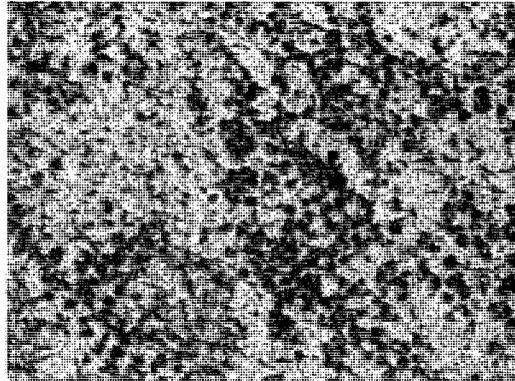
409
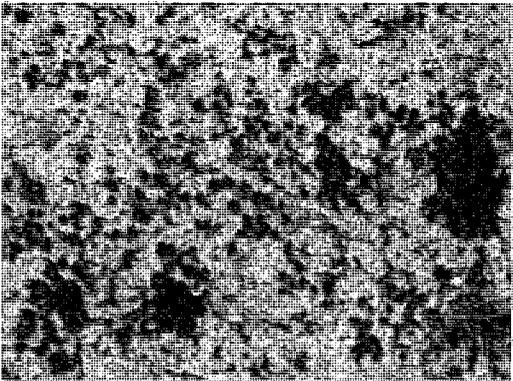
409–1
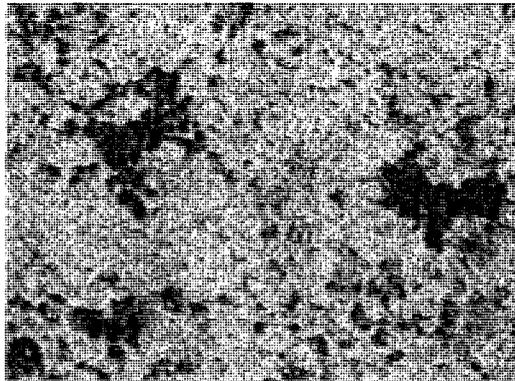
409–2
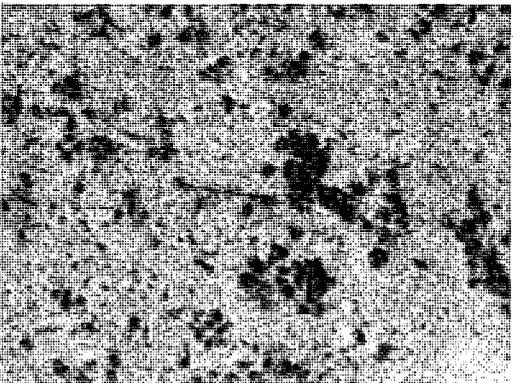
409–2b
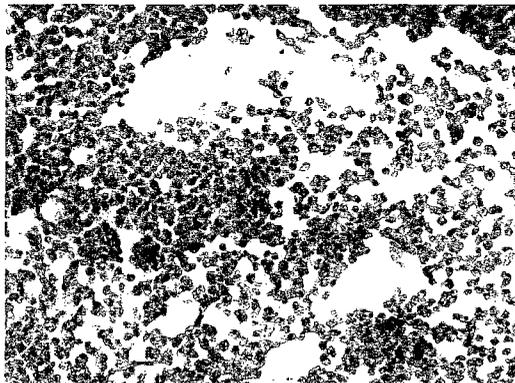
43444

A.1.2 High Malignancy (G3)
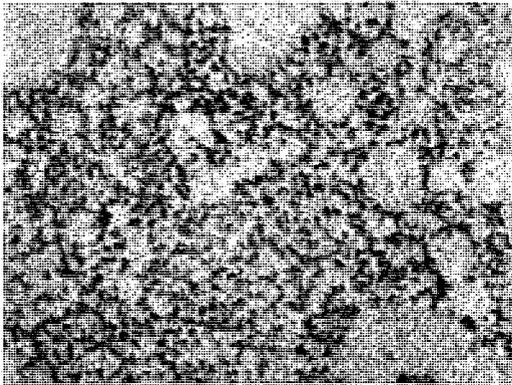
171
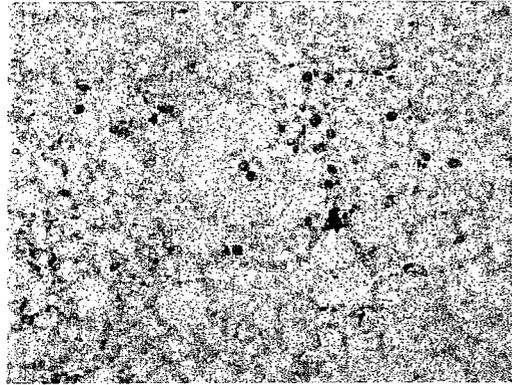
171-1
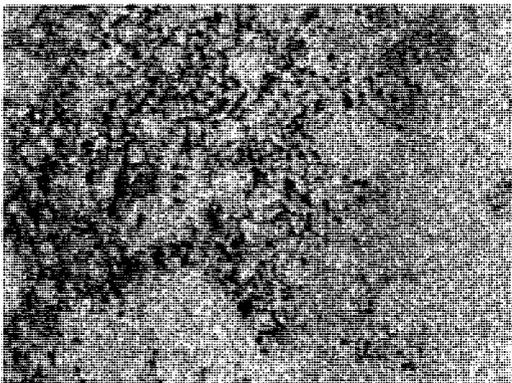
179
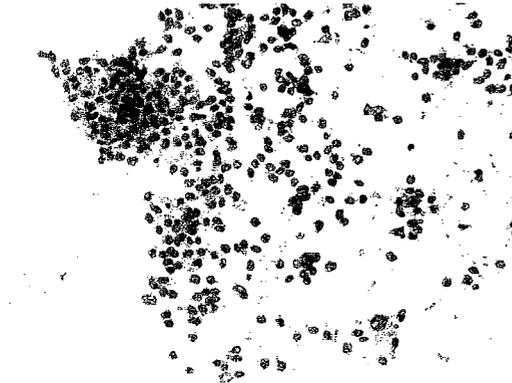
204
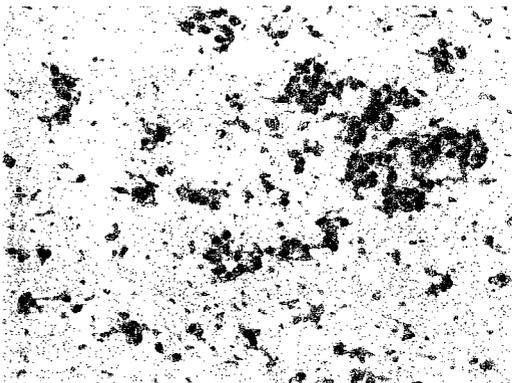
204-1
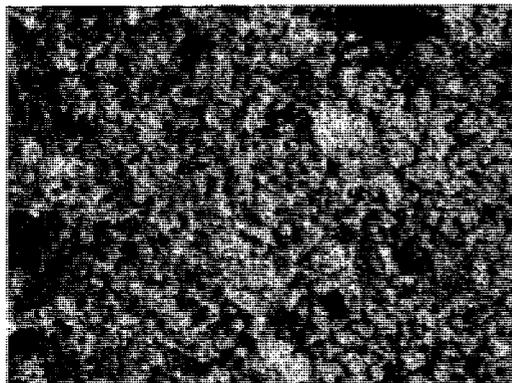
251
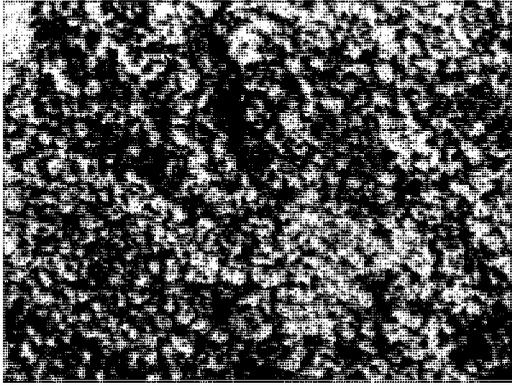
251-1
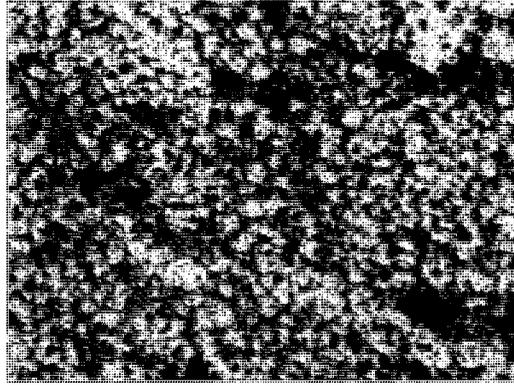
251-2
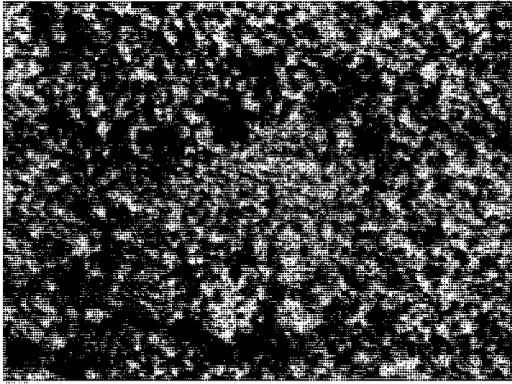
251-3
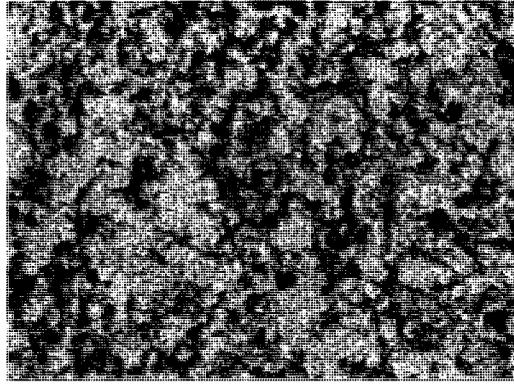
43213
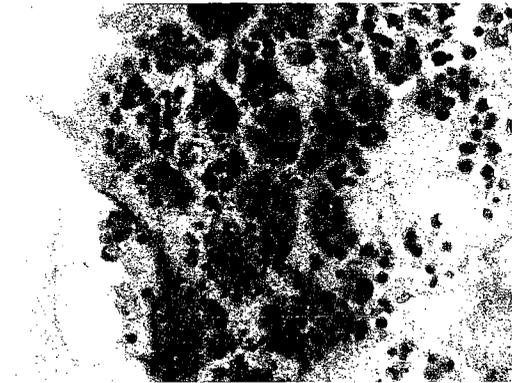
43246
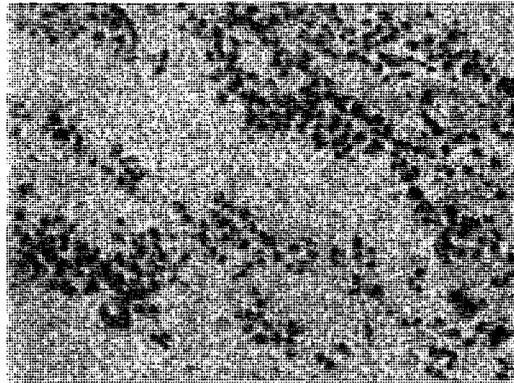
7403
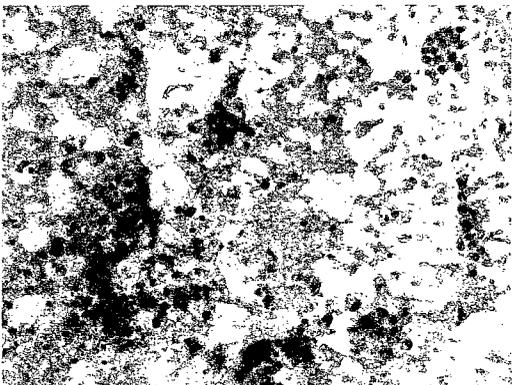
771

A.2 400x Magnification Images

A.2.1 Intermediate Malignancy (G2)
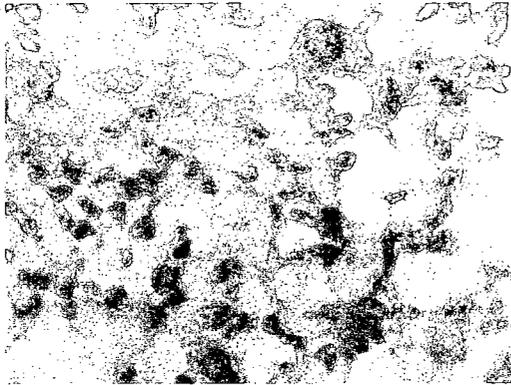
12071-a
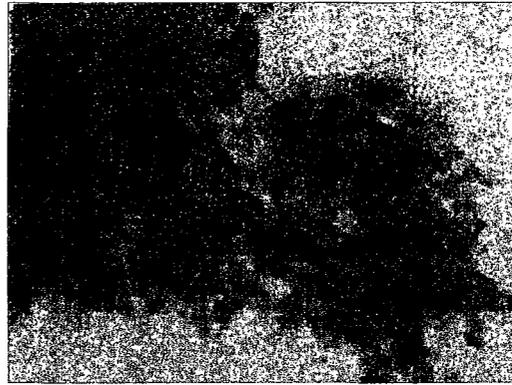
12071

152
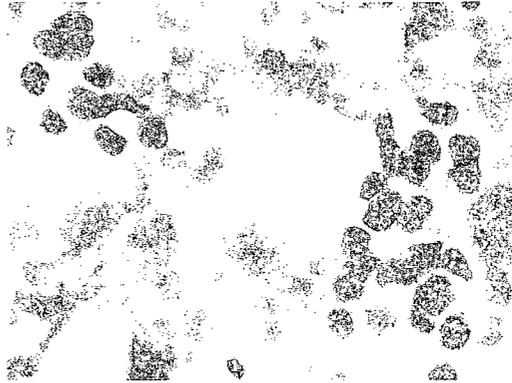
170-1
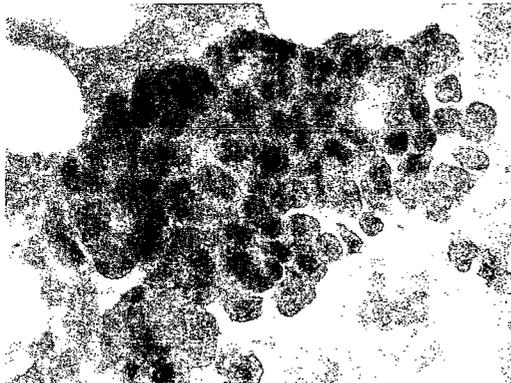
170
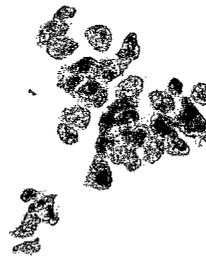
176
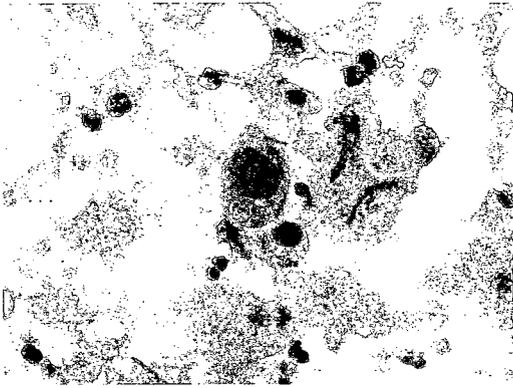
195–1
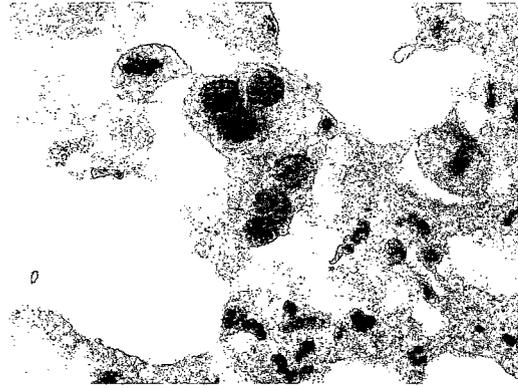
195–2
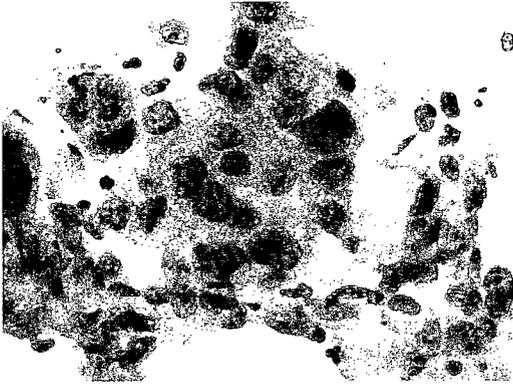
195

223–1
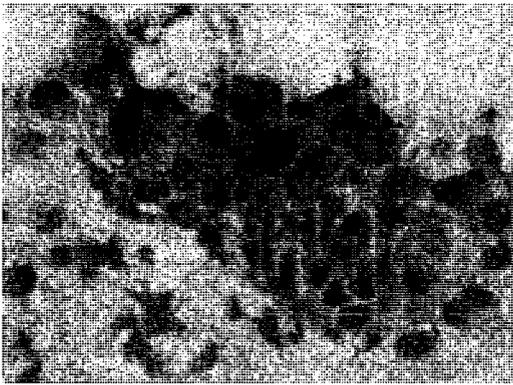
223
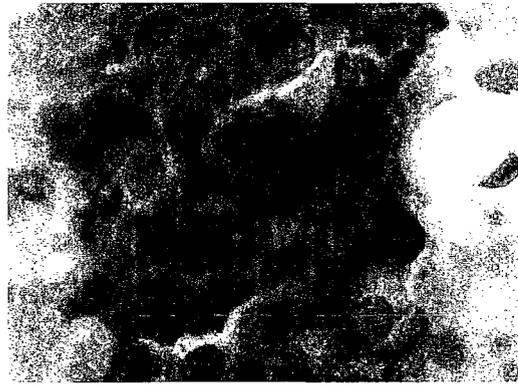
234
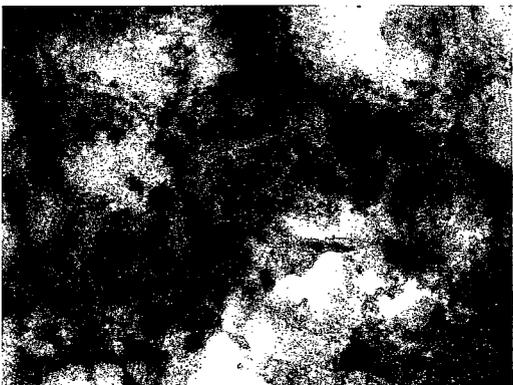
34–1
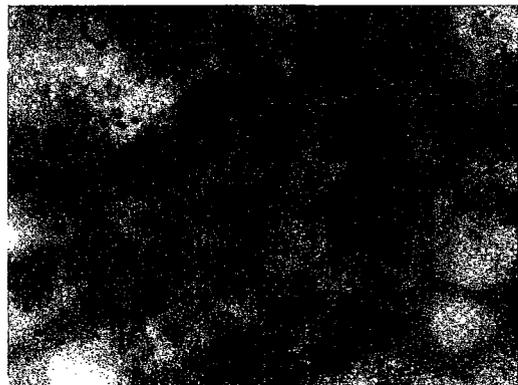
34
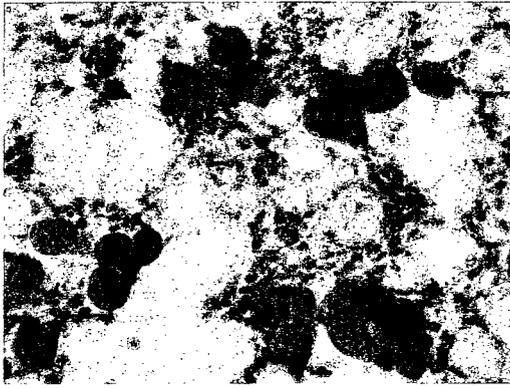
365–1

365–2
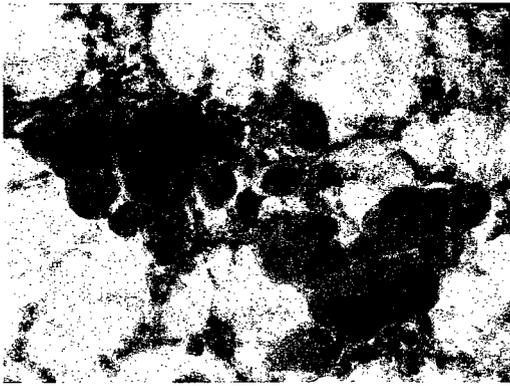
365–2
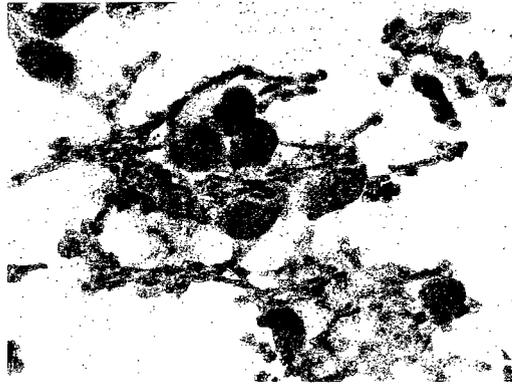
365
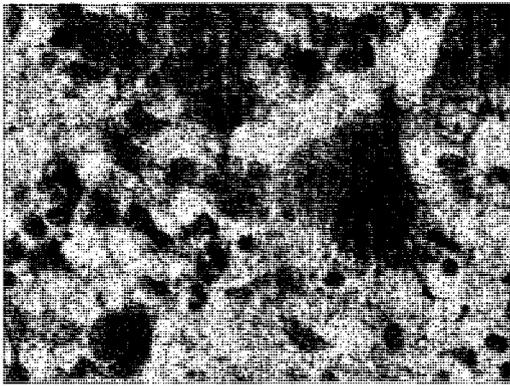
385–1
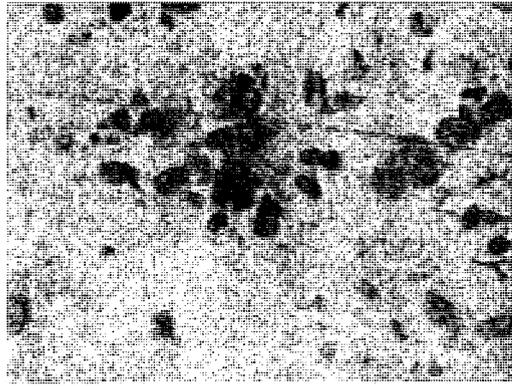
385–2
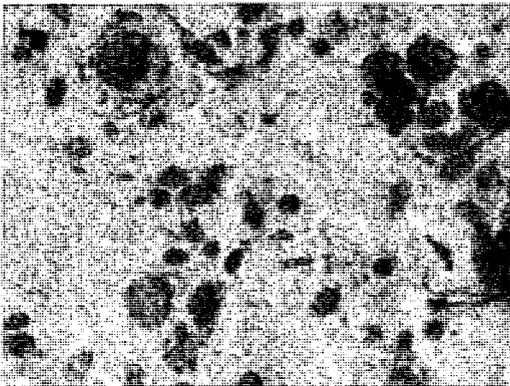
385–3
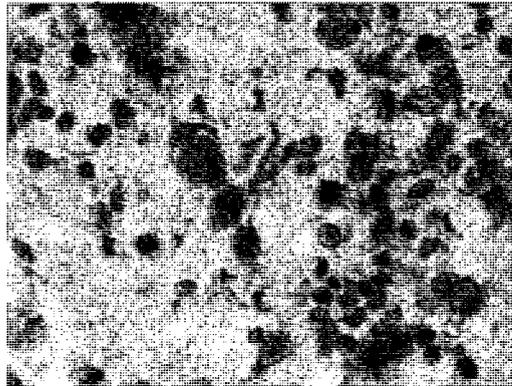
385
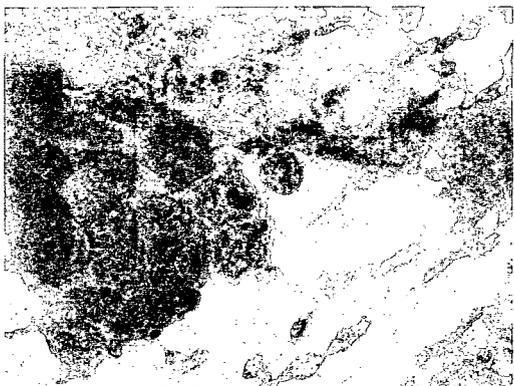
400–1
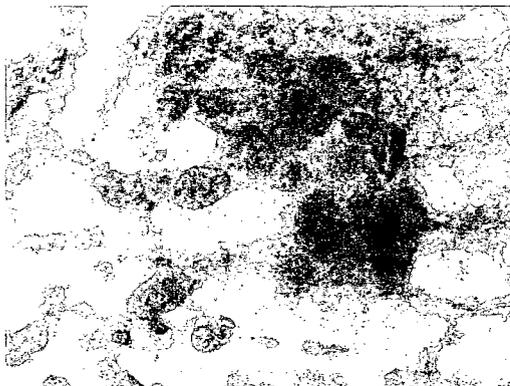
400–2

400–3
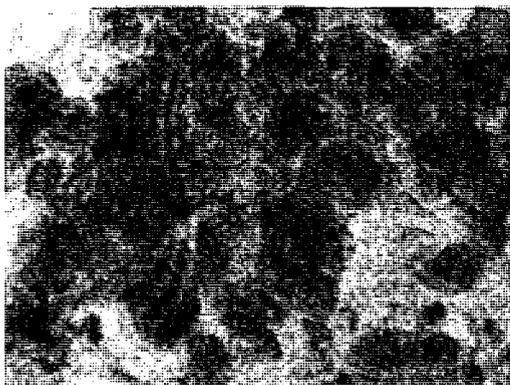
400
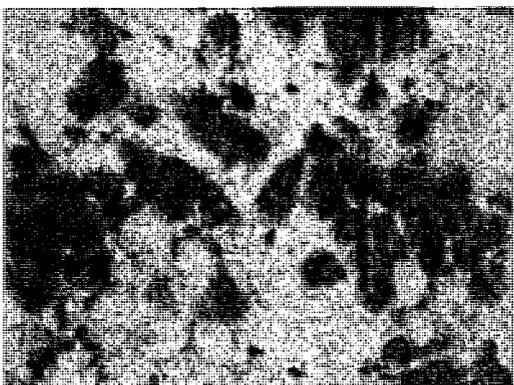
408–1
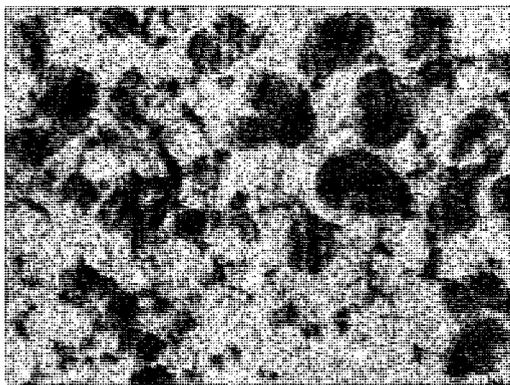
408–2
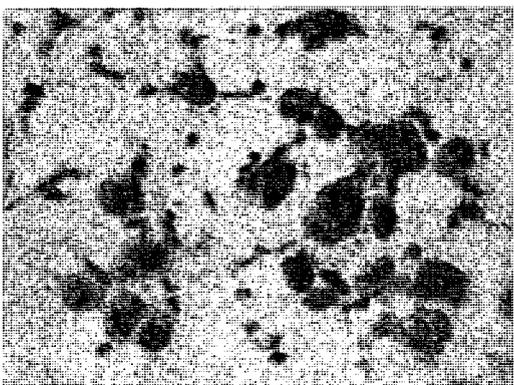
408–3
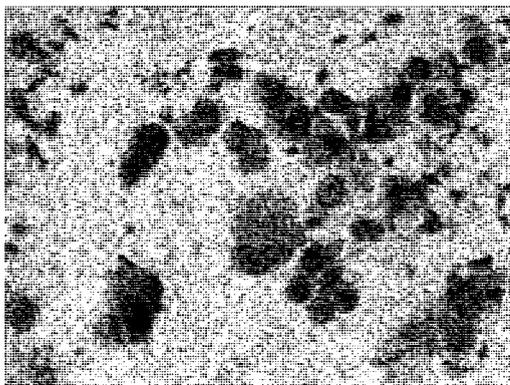
408
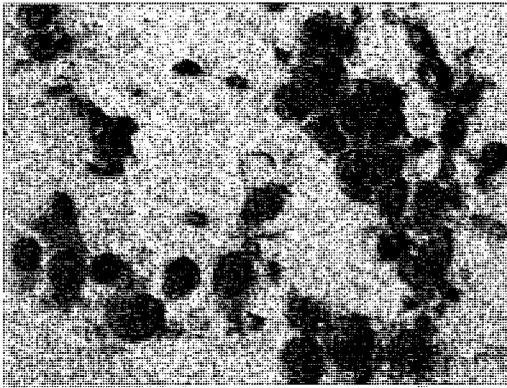
409-1
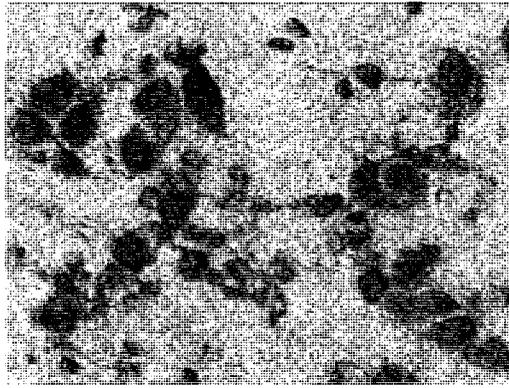
409-2
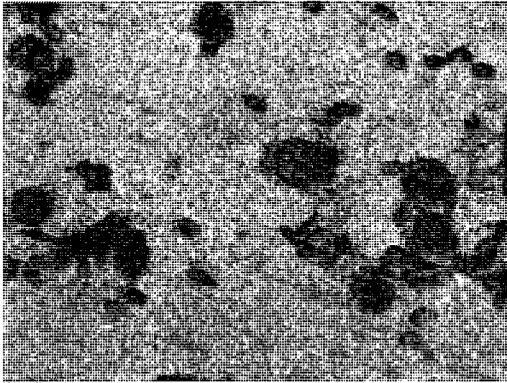
409
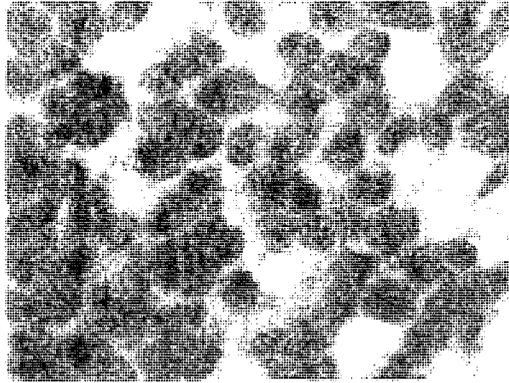
43444

A.2.2 High Malignancy (G3)
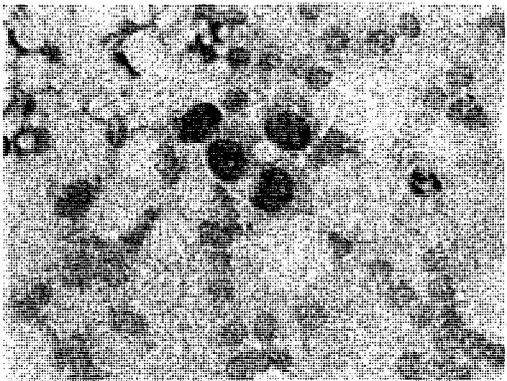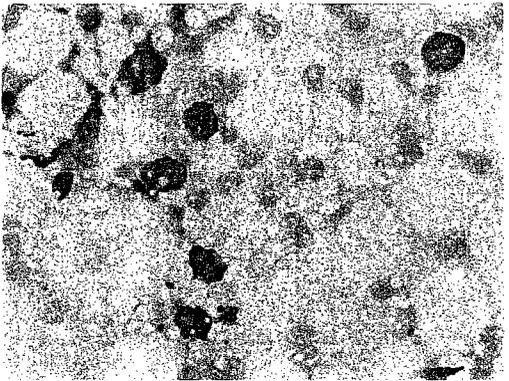
171-1

171
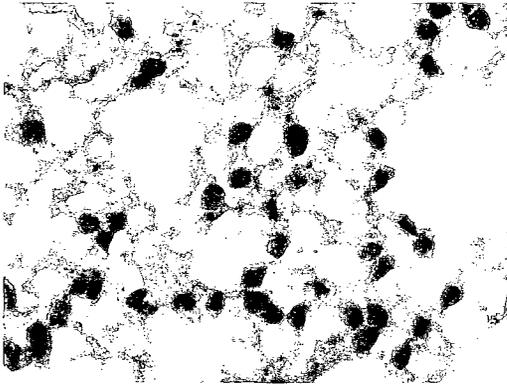
179-a
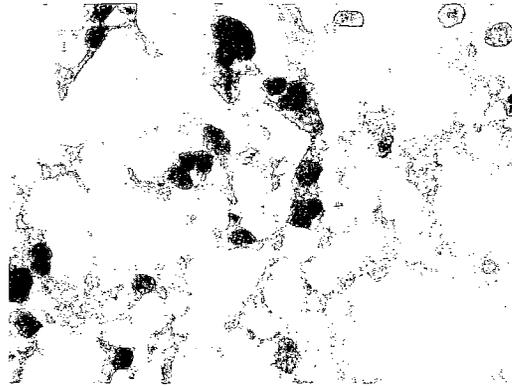
179
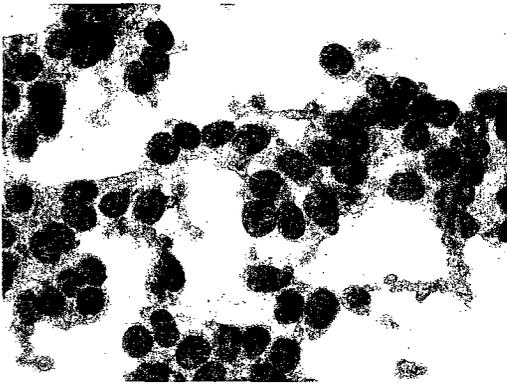
204-1
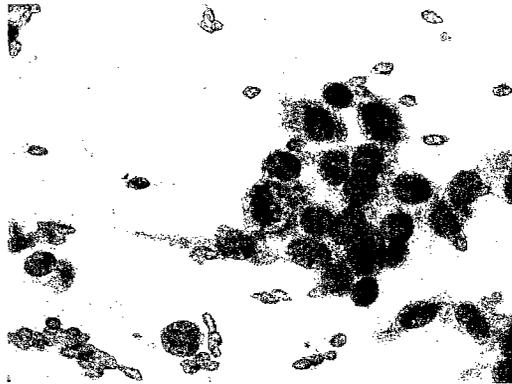
204
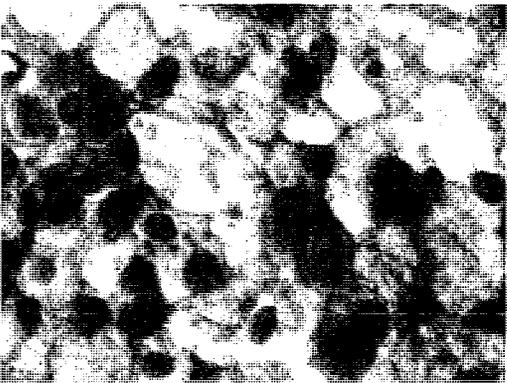
251-1

251-2

251-3

251-3a

164
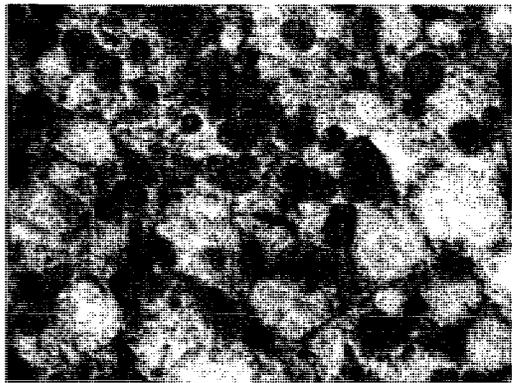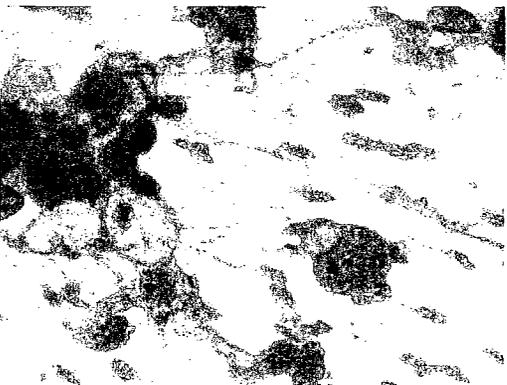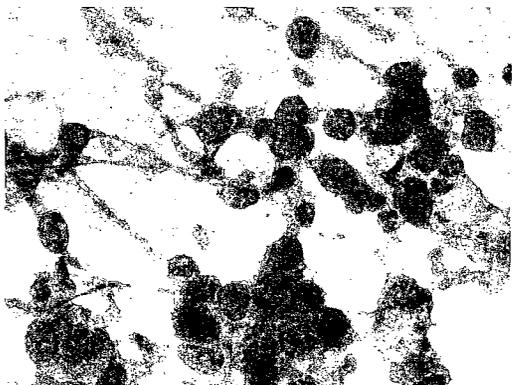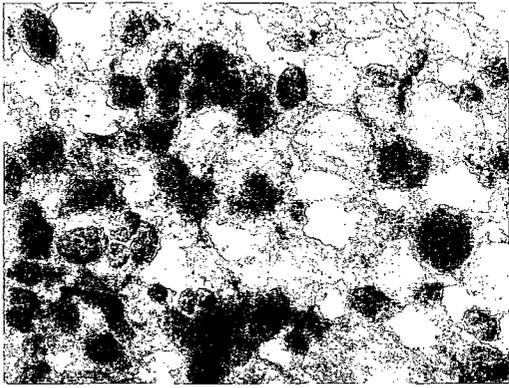
251
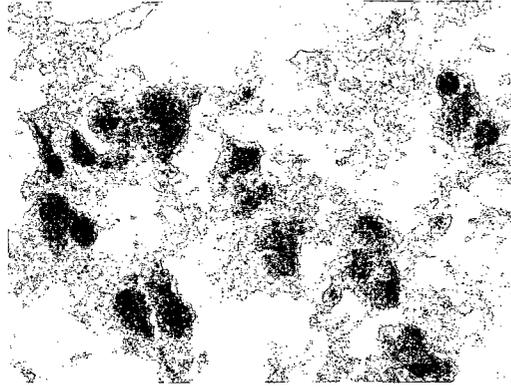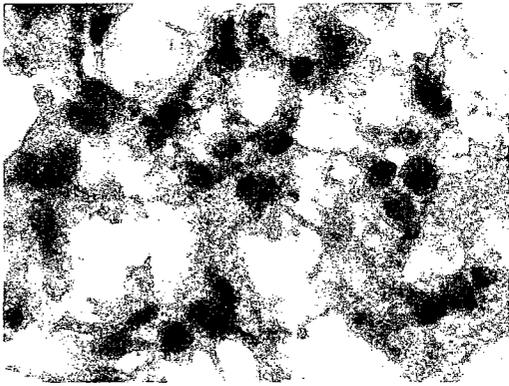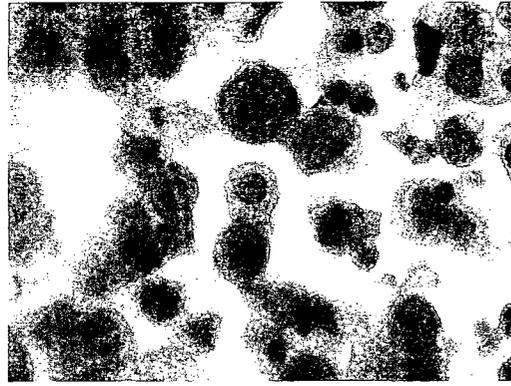
43213–a

43213

43246–a

43246–b
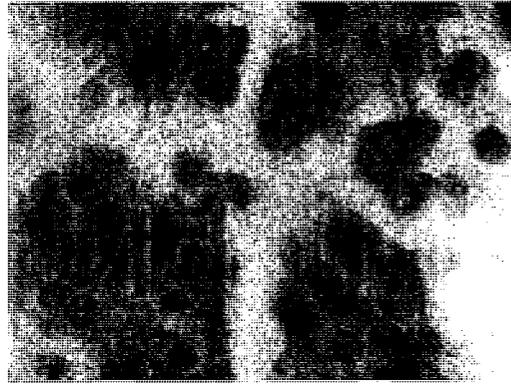
43146
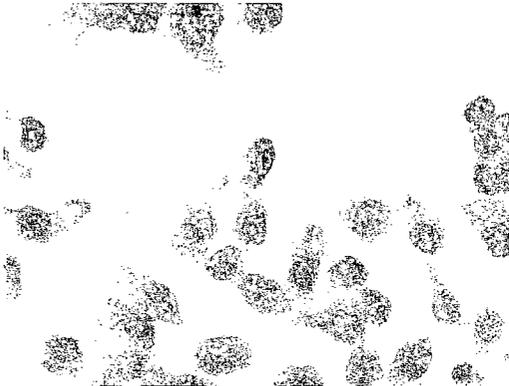
7403–a
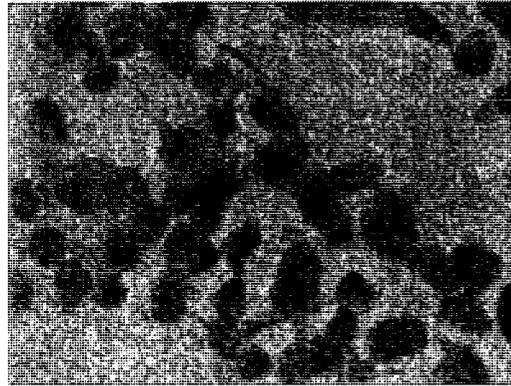
7403
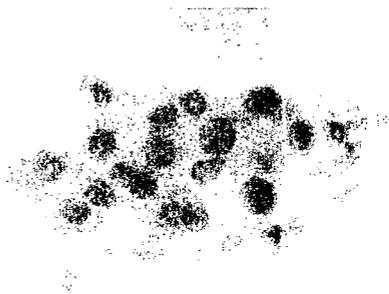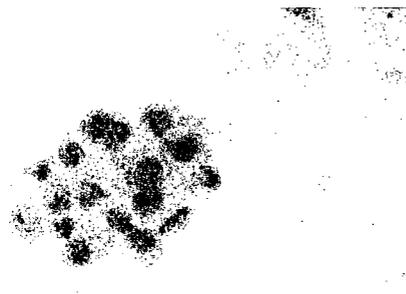
771-a   771

A.3 Cases with multiple 400x images taken for one 100x image

100x	400x
179	179
	179–a
251–3	251–3
	251–3a
409–2	409–2
409–2b	
771	771
	771–a
7403	7403
	7403–a
43213	43213
	43213–a
43246	43246
	43246–a
	43246–b
12071	12071
	12071–a

Table A.1: 100x cases for which multiple 400x images were taken.

Glossary

$(\overline{x}, \overline{y})$	Image centroid, 50
(x, y)	Integers that act as indexing for the pixels of the image, 38
A	Area, 52
A_i	Area of the ith nucleus, 77
A_{100}	Average nuclei area at 100x magnification, 74
C	Convexity, 53
$CH(S)$	Convex hull of the shape S, 53
$COMP$	Compactness, 54
CV	Circular variance, 54
C_i	Convexity of the ith nucleus, 78
C_{ij}	Graylevel conditional joint probabilities, 63
Cir	Circularity, 52
Cov	Covariance matrix, 53
Ctr_i	Centroid of the ith nucleus, 78
D	Dispersion, 75
$D(A, B)$	Dilation of a figure A with a structuring element B, 43
E	Histogram energy, 42
$E(A, B)$	Erosion of a figure A with a structuring element B, 43
EV	Elliptical variance, 54
E_E	External energy, 60
E_l	Cluster area, 75
E_I	Internal energy, 60
Ecc	Eccentricity, 52
Ecc_i	Eccentricity of the ith nucleus, 78
El	Elongation, 53

167