Effective improvement of cancer diagnostics and prognostics by computer-assisted cell image analysis

Von der Fakultät für Elektrotechnik und Informationstechnik der Rheinisch-Westfälischen Technischen Hochschule Aachen zur Erlangung des akademischen Grades eines Doktors der Ingenieurwissenschaften genehmigte Dissertation

vorgelegt von

Dipl.-Math.

David Friedrich

aus Saarburg

Berichter:

Prof. Dr.-Ing. Dietrich Meyer-Ebrecht
Prof. Dr. Bastian Leibe
Prof. Dr. med. Alfred Böcking

Tag der mündlichen Prüfung: 15. Juli 2015

Abstract

DNA Image Cytometry is a method for the early diagnosis and prognosis of cancer. It exploits, as a biomarker for cancer, the DNA content of morphologically suspicious nuclei measured from digital images. Therefore, the identification of these suspicious nuclei in a microscopic inspection is a crucial step of the method.

Until now, this task had to be performed by a pathological expert who required, on the average, 40 minutes per slide—prohibitive for a wide-spread routine application. This thesis presents image processing algorithms for accomplishing this task automatically, the core component being classifiers which are capable of distinguishing morphologically abnormal nuclei from normal nuclei, other types of nuclei, and artifacts. These algorithms were integrated into a software package, and a workflow which loads the tedious work onto the machine leaving only critical tasks to the responsible expert. This provides an overall solution, which was evaluated in three clinically relevant applications: the identification of cancer cells in nuclei from serous effusions and from brush biopsies of the oral cavity, and grading the malignancy of prostate cancer biopsies.

The developed solution reduces the workload for the expert to 5 minutes per slide. As compared the previous visual selection of nuclei, in addition both the diagnostic accuracy and prognostic validity are increased.

Acknowledgments

First, I want to honor Prof. Dr.-Ing. Til Aach†, who prematurely passed away during the time of this thesis. I would like to thank you for your recruiting me for this challenging project, and your confidence in me. As a young researcher, I could learn many important aspects from you. Prof. Dr.-Ing. Dietrich Meyer-Ebrecht, I would like to thank you for starting the DNA Image Cytometry project and managing it for such a long time. Thank you for standing in when needed, for the supervision of this thesis and your excellent feedback. You are a great adviser and I admire your wisdom. Prof. Dr. med. Alfred Böcking I would like to thank for the uncountable hours you invested in this project: brainstorming for new ideas, providing priceless annotations of data, teaching the medical background needed for this thesis, and so forth. I doubt that there are many professors who work for the good cause with such a passion, and who are that open to other disciplines as you are. A further acknowledgment goes to Prof. Dr. Bastian Leibe for also supervising this thesis, and for many valuable inputs to this work. Additionally, I thank Prof. Dr.-Ing. Dorit Merhof for supporting the continuation of this project.

I would like to thank Motic Asia for funding the DNA Image Cytometry project. I am very grateful to the R&D team of Motic, namely Dr. Chen Jin, Yu Zhang, Li Yuang, and Chen Demin for working together successfully for such a long time, and for implementing many of my ideas in their software. The employees at the Division Cytology, Heinrich-Heine University Düsseldorf and the Institute of Pathology, Düren, deserve my great acknowledgment for continuous cooperation and for providing the slides for this work. Anna Feider, Silke Meier, and Christiane Knops I would like to thank for performing the manual DNA measurements for this work.

For their continuous support and sharing of ideas I would like to thank my colleagues and students at the Institute of Imaging and Computer Vision, RWTH Aachen University: André Bell, Thomas Würflinger, Simon Koppers, Anna Jasinski, Christoph Haarburger, Anton Travinski, Adrian Luna-Cobos, Marta Solabre-Martínez, Peter Faltin, Marcin Kopaczka, Maria Sagrebin-Mitzel, Sebastian Groß, Joschka zur Jakobsmühlen, Matthias Breier, Julie Klein, Dorian Schneider, and Wei Li. Also, I am very grateful to Daniel Brückner and his team for the IT support, and fulfilling many of my (special) requests.

Finally, I want to thank all my friends and my family for constant encouragement and prayer support. Especially, my wife Pirjo, I would like to thank you for your continuous backing, for your love, and for taking many loads of from me so that I could focus on this thesis. And Jonatan, my dear son, thank you for the joy that

you brought into our life. Your smile, when coming home after a challenging day, was a great motivation for continuing. I hope that at least you may see a world where cancer has lost its deadly threat.

Contents

1 Introduction

Most cancers can be cured if treated early [1]. Since cancer is the uncontrolled growth of cells, it can be diagnosed by the microscopic study of cells, called cytology. Cells can be acquired non- or minimal-invasively, for example by brushing mucosal membranes such as the oral cavity or the uterine cervix. This allows for the use of cytology in screening for cancer. Cancer is found earlier [2], which increases the probability of survival of patients. Additionally, the compliance for the non- or minimal-invasive extraction of cells is higher than obtaining tissue for histological examinations, which requires a biopsy or operation [3]. For a cytological examination, the extracted cells are deposited on glass slides and stained to make cells and their nuclei visible by microscopic inspection. Conventional cytology is based on the *visual assessment* of the morphology of cells and their nuclei by a cytopathologist. The accuracy of this diagnosis, however, essentially depends on the experience of the pathologist. Another pathologist might judge the same cells differently, thus the interobserver reproducibility of cytological diagnoses often is only moderate [4]. Considering that an expert might see up to two million cells per day, cytology is also prone to errors due to fatigue. For about 4 % of the cases the cytological diagnosis remains unclear, since conventional cytology often cannot definitely decide between benign and malignant cells [5].

The most common and consistent property of solid tumors is an abnormal nuclear DNA content [6, 7]. DNA Image Cytometry is a cytological examination technique, which exploits the DNA content of morphologically abnormal (dyscariotic) nuclei as biomarker for cancer. The DNA content of individual nuclei is measured from digital microscopic images of nuclei, employing a special staining technique and dedicated image processing algorithms. Up to now, an expert systematically scans a slide, visually identifies morphologically abnormal nuclei and selects them for measurement. Finally, the DNA distribution of the selected nuclei is computed, and the expert derives a diagnosis or prognosis based on the deviation of this distribution from that of healthy cells [8, 9].

Often DNA Image Cytometry is used in addition to conventional cytological methods for the following purposes:

- Identification of cancer cells, e.g. in nuclei from brush biopsies of the oral cavity.
- Assessment of cytological or histological specimens suspicious of cancer, in case a definite diagnosis cannot be assigned by these conventional methods (adjuvant DNA Image Cytometry).

- Grading the malignancy of tumors from tissue or cell specimens, e.g. from prostate cancer biopsies.

The diagnostic accuracy and prognostic validity of DNA Image Cytometry are in most cases superior to conventional cytology alone [10–13]. This is for two reasons: First, it is based on the *measurement* of a *specific and valid biomarker*. Second, the DNA distribution is *solely* based on the DNA content of abnormal nuclei [8, 9]. In small tumors, there might be as little as several hundred abnormal among tens of thousands of normal nuclei. Only by focusing exclusively on the abnormal nuclei, they become detectable in the DNA distribution, which increases the diagnostic sensitivity.

A major disadvantage of DNA Image Cytometry, however, is the time needed for manually scanning a slide and visually identifying sufficient abnormal nuclei for computing their DNA distribution. On the average, a pathological expert requires 40 minutes per slide. With most of the schools for cytotechnicians in Germany being closed, soon there will be not enough skilled personnel available for this task. Finally, the reimbursement by the German healthcare system does not cover the expenses of DNA Image Cytometry for material and personnel. Thus despite of its advantages, up to now the use of DNA Image Cytometry had been limited to a few specialized institutions.

One approach to reduce the human workload is to automate time-consuming tasks: A microscope with motorized stage and a digital camera systematically scans a slide and acquire digital images of nuclei. Subsequently, a nucleus classifier identifies clinically relevant nuclei in these images. For the sake of automation however, most of the approaches presented in literature derive a diagnosis based on statistics of nucleus features or the frequency of occurrence of nuclei with particular properties [14–22]. None is able to distinguish morphologically abnormal nuclei from normal nuclei and all other objects of the slide [23]. Therefore, these approaches cannot make use of the essential enhancement of sensitivity due to the restriction on abnormal nuclei. As a consequence, they suffer diagnostic accuracy and are not in line with international guidelines of DNA Image Cytometry [8, 9, 17, 20].

1.1 Contributions

It is the aim of this thesis to support the pathologists by automating time-consuming and tedious steps of manual DNA Image Cytometry and henceforth pave the way to a broad application of DNA Image Cytometry in clinical routine. To this end, this thesis provides the following contributions:

- Image processing algorithms which accomplish time-consuming steps of DNA Image Cytometry: The main contributions are classifiers for nuclei originating from body cavity effusions, prostate biopsies and cervical/oral brush biopsies. The classifiers are able to distinguish abnormal from all other nucleus classes

and artifacts. Since the classification is mainly based on features quantifying the morphology, these classifiers perform a morphometric preselection of abnormal nuclei, and the DNA distribution for diagnosis can exclusively be derived from these nuclei. Another contribution is a segmentation pipeline for the precise, unsupervised segmentation of nuclei stained for DNA Image Cytometry.

- The translation of the developed algorithms into clinical routine: This is achieved by the integration into a software package which can be used by clinical end users, and by the implementation into a clinical workflow which loads the tedious work onto the machine, leaving only the critical tasks to the responsible expert.
- The evaluation of the developed algorithms and workflow as an integrated solution in three applications: the identification of cancer cells in nuclei from serous effusions and from brush biopsies of the oral cavity, and grading the malignancy of prostate cancer biopsies. The evaluation was performed on specimens from 203 patients.
- A proof-of-concept that DNA Image Cytometry can also be performed on Whole Slide Imagers: These system scan glass slides considerably faster than conventional microscopes with motorized stage and are equipped with a slide loader for the automated batch processing of slides. Thus they have the potential to increase the throughput of DNA Image Cytometry systems further.
- Image processing algorithms for the detection of felt tip marked ROIs from an overview scan of the glass slide: These algorithms are required for the automated batch processing of slides.

By means of these contributions, the following results were obtained:

- The presented solution reduces the workload for the expert from 40 down to 5 minutes.
- At the same time, the solution increases the diagnostic accuracy and prognostic validity as compared to manual DNA Image Cytometry.

1.2 Organization

This thesis is organized as follows:

Chapter 2 presents the relevant medical background information for this thesis. The first part is about conventional histological and cytological methods for cancer diagnostics and prognostics. Subsequently, the preparation of glass slides and the principles of DNA Image Cytometry are explained. The latter method is then compared to the conventional methods by quality criteria such as diagnostic accuracy, reproducibility, and costs. The main hindrance for a broad application of DNA Image Cytometry is the time needed to perform it manually. State-of-the-art approaches to reduce this workload are presented and analyzed in more detail.

In **Chapter 3**, image processing algorithms which were developed for accomplishing time-consuming steps are presented. First, a brief overview of the image acquisition processes is given. Subsequently, the segmentation pipeline for the precise, unsupervised segmentation of nuclei is explained. The development, optimization and evaluation of classifiers for the identification of abnormal nuclei are presented next. The chapter ends with an algorithm for the automated identification of ROIs from overview scans of glass slides.

Chapter 4 addresses topics related to the translation of the algorithms into clinical routine. For that purpose, most of them were integrated into a commercially available software package. Its user interface is briefly presented. Next, guidelines for integrating the algorithms into the clinical workflow are described. Finally, the developed algorithms and the workflow are compared to manual DNA Image Cytometry in three clinically relevant applications, the identification of cancer cells in nuclei from serous effusions and from brush biopsies of the oral cavity, and grading the malignancy of prostate cancer biopsies.

The potential to apply Whole Slide Imagers to DNA Image Cytometry is studied in **Chapter 5**. First, methods for the optical characterization of these systems are presented. Next, the precision of the DNA measurement algorithms on Whole Slide Imaging scans is evaluated.

In **Chapter 6**, the main conclusions from this work are drawn and remaining limitations are listed. Finally, perspectives for further research are given.

Note: Parts of this work are based on the author's publications. These publications are listed in the motivation part of the corresponding algorithm, while this work also presents significant enhancements. Also, results from bachelor and master's theses, supervised by Prof. Dr.-Ing. Jens-Rainer Ohm or Prof. Dr.-Ing. Dorit Merhof, and the author, are derived. In that case the corresponding works are cited.

2 Background

This chapter presents the medical and technical background of this thesis. Its first part presents criteria for assessing the quality of methods for the diagnosis and prognosis of cancer. Current state-of-the-art methods are analyzed with respect to these quality criteria. Subsequently, DNA Image Cytometry, a method which can be applied for both diagnosis and prognosis of cancer, is presented. In some aspects it is superior to conventional methods or adds significant diagnostic or prognostic information. However, due to the long interaction time of medical experts and the lack of skilled personnel, its use is limited to a few specialized institutions. In order to make the advantages available to a broader range of patients, time-consuming steps of DNA Image Cytometry can be automated. The chapter ends with current state-of-the-art methods for this purpose and an analysis of their limitations.

2.1 Diagnosis and prognosis - key factors for fighting cancer

2.1.1 Diagnosis

Most cancers can be cured if diagnosed and treated early [1]. Figure 2.1 depicts the probability of survival of patients with oral cancer, depending on the stage at which the cancer was found. For the earliest stage T1, about 90 % of the patients are still alive 10 years after surgical treatment. The probability of survival decreases if the cancer is found in later stages. It is below 40 % for a cancer found at the latest stage T4. This applies for almost all types of cancer, similar results have for instance been shown for cervical cancer [24], lung cancer [25] or pancreatic cancer [26]. Hence, the early diagnosis of cancer is a key factor for successfully fighting this disease.

A diagnostic examination for cancer clarifies the question "Does the patient have cancer or not?". Choosing an adequate diagnostic method depends on the type and size of the cancer, the clinical questions and consequences. For instance, a screening method must be cheap and should not yield too many false-positive diagnoses. On the contrary, if a patient shows up in the clinic with symptoms, the overall diagnostic performance must be high, but the method may be more expensive than for screening. After these kinds of requirements have been clarified, the following criteria can be considered for assessing the quality of a diagnostic method:

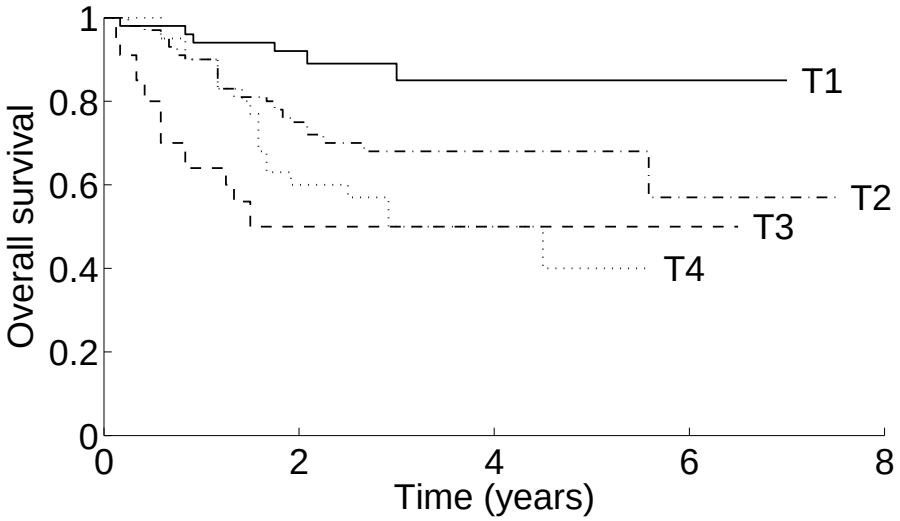


Figure 2.1: Overall survival, depending on the stage in which oral cancer was diagnosed. Data from [1].

- **Diagnostic accuracy:** A diagnostic method should yield a correct result. If a cancer is overlooked (false-negative), it may be detected too late for curative therapy. Diagnosing cancer in a healthy person (false-positive) will unnecessarily create major psychological discomfort and physical harm.
- **Early diagnosis:** The point in time of cancer diagnostics heavily influences the probability of survival. Therefore, methods for diagnosing cancer should detect it at an early stage.
- **Reproducibility:** A diagnostic method should yield the same result when performed by different examiners and institutions to be reliable.
- **Invasiveness and side effects:** Harming the patient clearly is a disadvantage of a diagnostic method.
- **Costs:** A method must be affordable, otherwise it cannot be applied for a broad range of patients.

2.1.2 Prognosis

A second key factor for adequately fighting cancer is a reliable prognosis. The task of a prognostic index is to make a prediction about the future development of a cancer at the current point in time. This information guides type and invasiveness of the treatment. For instance, for an early low-grade ovarian cancer with a good prognosis it is sufficient to remove the affected ovary by an operation which usually preserves fertility. For ovarian cancers with a bad prognosis removing both ovaries and the uterus, and additionally treating the patient with chemotherapy, is necessary [27].

For a prognostic index it is important to predict the patient relevant endpoints, such

as overall survival, disease free survival, occurrence of metastases, or relapses, as precisely as possible. If the prognosis is too low, the cancer is not treated effectively enough. If the prognosis is too high, this will lead to overtreatment, with unnecessary harm and side effects for the patient and avoidable costs for the health care system.

2.1.3 Active surveillance of prostate cancer

A special case among all types of cancer is prostate cancer. It often grows so slowly that most patients die *with*, not *of* the cancer [28]. Most of the patients are already around 70 years old when their cancer is diagnosed [28].

The current practice of the German health care system is that every man older than 45 years has the chance for an annual checkup for prostate cancer. Besides a rectal examination, the patient is offered a blood test for the prostate specific antigen (PSA). If the PSA value is above a certain threshold, a biopsy is advised. Patients with cancer then mostly undergo a radical prostatectomy (removal of the prostate) or a radiation therapy [29].

The latter practice has, however, two drawbacks: it often leads to *overdiagnosis*—cancers which would not have harmed the patient and which would never have been found during his lifetime are identified [30]. Overdiagnosis increases *overtreatment*—patients undergo unnecessary treatments. The associated side-effects of radical prostatectomy, for instance, are severe: Between 59.9 % - 72 % of the patients report impotence, 8.4 % - 10 % incontinence, and 41.9 % have moderate to big problems in their sexuality [31, 32]. Welch et al. provide evidence that overdiagnosis is actually happening [30]. Bangma et al. argue that the consequences of overdiagnosis are so severe that a nationwide screening for prostate cancer is not recommended [33]. In 2012, the U.S Preventive Services Task Force recommended against PSA-based screening for prostate cancer, as its harm does not outweigh the benefits [34].

Instead, the concept of active surveillance manages non-aggressive cancers conservatively. If a clinically insignificant prostate cancer has been diagnosed, an invasive treatment is postponed until the cancer shows progression. To this end, the progression of the cancer is assessed in regular intervals. If no treatment is required at lifetime, this spares the patient a radiation or operation and their associated complications and discomfort. A study at the University of Göteborg, Sweden, recently estimated that about 60 % of all prostate cancers are suited for active surveillance, as they are very low or low risk cancers [35]. Thus active surveillance could overcome the drawback of overtreatment associated with PSA-based screening, while the advantage, the earlier detection of aggressive cancer, is retained.

Essential for the concept of active surveillance is a valid prognostic test which precisely predicts if the cancer will be life-threatening or not. Otherwise, anxiety among patients and their physicians for postponing active therapy is too large. Yet, the current predictors need improvement [28, 33].

2.1.4 Histology - the gold standard

The current gold standard for the diagnosis and prognosis of most cancers is histology, the analysis of tissue sections. Even if modalities like tomographic imaging, molecular imaging, or endoscopic examinations arouse the suspicion for cancer, the definite diagnosis usually needs to be confirmed by histology. For a histological examination, tissue is extracted with a punch biopsy, a forceps biopsy, scalpel biopsy, or operation. Tissue sections can be stained with Hematoxilin and Eosin (H&E), which stains nuclei blue and cytoplasm purple. The pathologist then analyzes the structure of the tissue, growth patterns, and morphology of cells with a brightfield microscope.

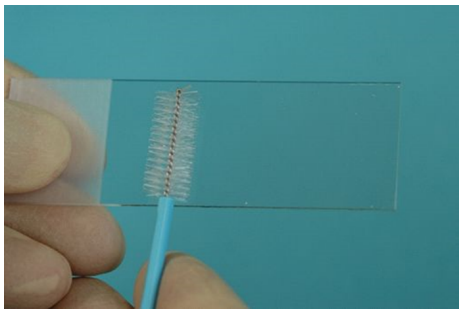
Histology is used for diagnosing literally all cancers which derive from tissues. Another application is the prognostication of individual cancers. For prostate cancers, the Gleason score is used for this task: the pathologist assesses the deviation of cancer tissue from normal tissue and assigns grades for the different growth patterns. These grades range from 1 (closest to normal tissue) to 5 (strongest deviation from normal tissue) [36]. The Gleason score in punch biopsies is then generated as the sum of growth pattern grades from the predominant and the worst growth pattern [37]. The Gleason score is related to the prognosis of a prostate cancer patient [38]. Patients with a Gleason score lower than 7 and appropriate PSA statistics come into consideration for active surveillance [39].

2.1.5 Cytology - a non-invasive alternative

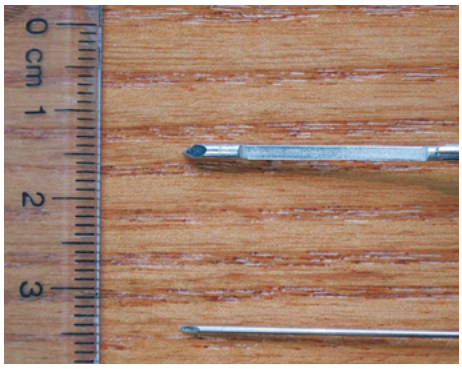
Cytology is the study of cells. Since cancer is the uncontrolled growth of cells, it is possible to diagnose cancer on the cellular level. Cytology has the advantage over histology that the cells for the examination can be extracted non- or minimal-invasively by a wide range of methods:

- **Brush biopsies from mucosal membranes:** A small brush is turned on a suspicious mucosal lesion. This gently extracts superficial cells from tissue. Brush biopsies are for instance possible for the uterine cervix or the oral mucosa (see Figure 2.2(a)).
- **Body fluids:** Body fluids contain cells from epithelium of the organ where they originate from. If these cells include cancer cells, they derive from the respective epithelial outlines. Often body fluids are directly accessible, for example urine or sputum.
- **Needle aspiration of body cavity effusions:** Metastatic cancer cells might block lymph vessels which are necessary for the outflow of fluid from body cavities. Due to the block, the fluid is collected in the cavity, which is called a serous effusion. Effusions caused by a metastasis contain cancer cells, which can be gained by aspirating the effusion with a thin needle. Therefore the identification of cancer cells in these aspirates assists to detect cancer. Often effusions are the first clinical symptoms of tumors or their metastases [10].

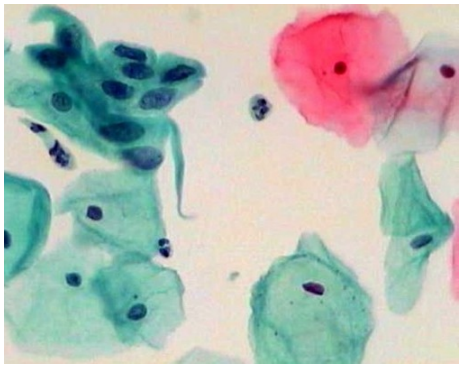
(a) Brush biopsy of the oral cavity.



(b) Deposition of cells on a glass slide.



(c) Difference between core- and fine-needle.



(d) Cervical cells stained acc. to Papanicolaou.

Figure 2.2: Extraction, preparation and microscopical analysis of cells: Cells can be gained for instance by a brush biopsy (a) or a fine-needle biopsy (c). After deposition on a glass slide (b), the cells are stained and analyzed by brightfield microscopy (d).

- **Fine-needle biopsies from lesions:** A fine-needle can be used to extract cells from tissue. To this end, a thin needle is inserted into the tissue region, if required guided by ultrasound imaging. The cells are then sucked out using vacuum. Compared to the core-needle biopsy used for histology, which has a diameter of 1 mm - 2 mm, a fine-needle only has a diameter of 0.7 mm (see Figure 2.2(c)). And whereas a core-needle cuts out tissue, the fine-needle only removes cells and leaves connective tissue at its place. Consequently a fine-needle aspiration is less invasive and the wound heals more quickly. Applications of fine-needle aspiration are tumors of the prostate, thyroid, salivary glands, lymph nodes, pancreas, and thyroid.

- **Enzymatic cell separation from tissues:** It is possible to obtain individual cells from tissue of existing biopsies using enzymatic cell separation. Thus histological material which is already available can additionally be used for cytological analysis.

After extraction, cells are deposited on a glass slide and fixed (see Figure 2.2(b)). In order to make the cells and nuclei visible for a microscopic inspection, they are stained.

The most common application of diagnostic cytology is screening of women for cervical cancer using the PAP-test. Cells are extracted by brush biopsy and stained according to Papanicolaou (see Figure 2.2(d)). The diagnosis is based on the judgment of the cell and nucleus morphology by a cytotechnician and a pathologist.

2.1.6 Quality of histological and cytological methods

Histopathological and for some cancers also cytopathological diagnoses currently are the gold standard of cancer diagnostics. Nevertheless, they reveal some shortcomings:

- **Diagnostic accuracy:** Histology is the gold standard in cancer diagnosis. Yet, neither its sensitivity nor its specificity reach 100 % [40]. For cytology, the sensitivity of the oral brush cytology is 91.35 % and specificity 95.1 % [12]. About 4 % of the diagnoses are suspicious [5]. This means that the method often cannot definitely decide between benign and malignant cells. Mean sensitivity for the cervical Pap-test is only 58 % [41]. Screening for cancer cells in serous effusions yields a sensitivity of only 50 %, with a specificity of 97 % [10].
- **Early diagnosis:** In conventional histology and cytology, the diagnosis is based on the assessment of tissue growth patterns and cell morphology respectively. A change in cell- and tissue morphology, however, is only an epiphenomenon of cancer. And as just mentioned for about 4 % of the oral smears no definite diagnosis can be assigned, which also reflects the difficulties in detecting subtle cancerous changes based on morphology. Thus it might be possible to diagnose cancer earlier using other biomarkers.
- **Prognostic validity:** As an example, the Gleason score fails in predicting the progression of locally confined prostate cancers with scores 6 to 7a with sufficient validity [42–45]. But especially these cancers come into consideration for active surveillance.
- **Reproducibility:** As the diagnostic or prognostic result is based on a human assessment of tissue patterns or cell morphology, histology and cytology reveal an interobserver variability. The reproducibility of the Gleason score is as low as 68.3 %. Compared to an expert in Gleason scoring, 22 out of 29 pathologists under-graded cancers [46], which might have fatal consequences in active surveillance. The reproducibility of the Pap test for screening for cervical cancer is only moderate ($\kappa = 0.46$, [4]).
- **Invasiveness:** Histology requires invasive procedures to obtain tissue, which may cause pain, bleeding, or infection.

These shortcomings should be overcome by methodological innovations.

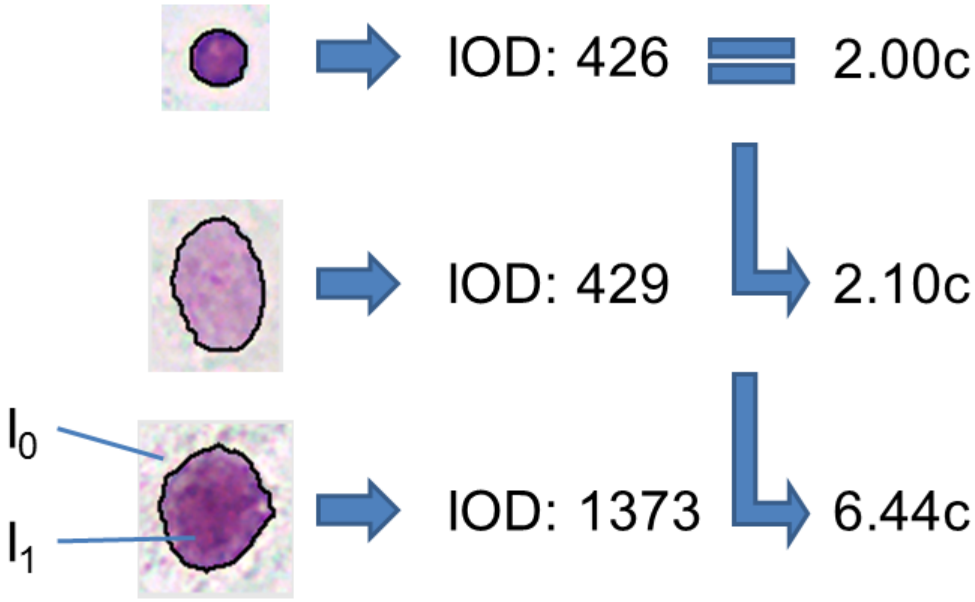


Figure 2.3: Computation of DNA values in DNA Image Cytometry.

2.2 DNA Image Cytometry

Abnormal amount of nuclear DNA is the most common property of human cancers [6, 7]. In DNA Image Cytometry, the DNA content of nuclei is measured from digital images of these nuclei. The diagnosis or prognosis of cancer is then based on the DNA distribution of hundreds to thousands of abnormal nuclei (DNA ploidy analysis). In the following, the principles for measuring DNA and performing a diagnosis or prognosis based on the DNA distribution are explained. Finally, the advantages and disadvantages of DNA Image Cytometry are discussed with respect to the quality criteria described the previous section.

2.2.1 Measuring the DNA content of nuclei

The measurement of the DNA content of cells is enabled by staining them according to Feulgen. For this staining protocol, the uptake of stain in a nucleus is proportionally to its DNA content (stoechiometric staining). During the Feulgen reaction, a dye is needed to actually make the nucleus visible. Either the thionin or pararosanlin dye can be used for this purpose. For this work the pararosanilin dye was applied, which stains nuclei purple (Figure 2.3).

The DNA content is then measured from the overall attenuation of light passing through the stained nucleus. Digital images of nuclei are acquired by a brightfield microscope and a digital camera. The Lambert-Beer law describes the attenuation

of light passing through semi-transparent matter:

$$I_1 = I_0 \cdot \mathrm{e}^{-c \cdot d \cdot \epsilon_\lambda}, \tag{2.1}$$

where I_0 is the incoming light, I_1 the transmitted light, c and d are the concentration and thickness of the matter, and ϵ_λ is a wavelength dependent constant. Consequently, the higher the DNA content, the higher is the concentration of stain, and the darker nuclei of the same size appear in digital images (see the two lower nuclei in Figure 2.3). The incoming light intensity I_0 is the pixel value of a digital image at an empty position, and the transmitted light intensity I_1 the pixel values inside an object attenuating light. Due to the 2D imaging, the thickness d cannot be considered, instead the optical density (OD) is defined as

$$\mathrm{OD} := -\ln\left(\frac{I_1}{I_0}\right) = c \cdot d \cdot \epsilon_\lambda. \tag{2.2}$$

The OD is linearly related to the concentration of the stain. After the delineation of a nucleus has been found by a segmentation algorithm, all pixel values inside the nucleus are summed up and yield the integrated optical density (IOD), a quantity reflecting the overall attenuation of the nucleus. In order to relate the integrated optical density to DNA content, the system needs to be calibrated for normal DNA content by the IOD of nuclei with normal DNA content (reference nuclei). Nuclei from the immune system (granulocytes, lymphocytes), connective tissue (fibroblasts), or normal epithelial cells can be used as reference nuclei[1]. The DNA content is then computed from the IODs by the rule of proportions (see Figure 2.3). In practice, however, error correction procedures need to be applied for ensuring correct results. Effects like glare, diffraction, background artifacts, or shading effects change the true value of I_1. These effects need to be compensated using dedicated image processing algorithms [48–50], which are explained in appendix A.1.

A typical workflow of a conventional DNA Image Cytometry is depicted in Figure 2.4. At first, the pathological expert selects the regions of interest (ROI) which should be scanned. Either, the whole slide is scanned, or a pre-examination has identified regions with abnormal cells. In this case, the ROIs have been marked by felt tip marker directly on a slide. These regions are then scanned in a structured manner. The pathologist examines the nuclei under the microscope and on a live video on screen. For each field of view (FOV), the expert ensures that the cells are in focus so that they are digitized correctly. The expert selects normal nuclei for the calibration of the DNA measurement algorithms and abnormal nuclei for measuring the DNA content by clicking on the nuclei in the live video. These nuclei are automatically segmented, and the expert assigns the corresponding nucleus class. The system computes the DNA content of the nuclei as described in the previous

[1] Since these nuclei are contained in the specimen, they are also denoted as internal reference nuclei—opposed to external reference nuclei such as diploid rat liver hepatocytes, which are artificially added during the preparation of the slide [47]. In this work, only internal reference nuclei are used.

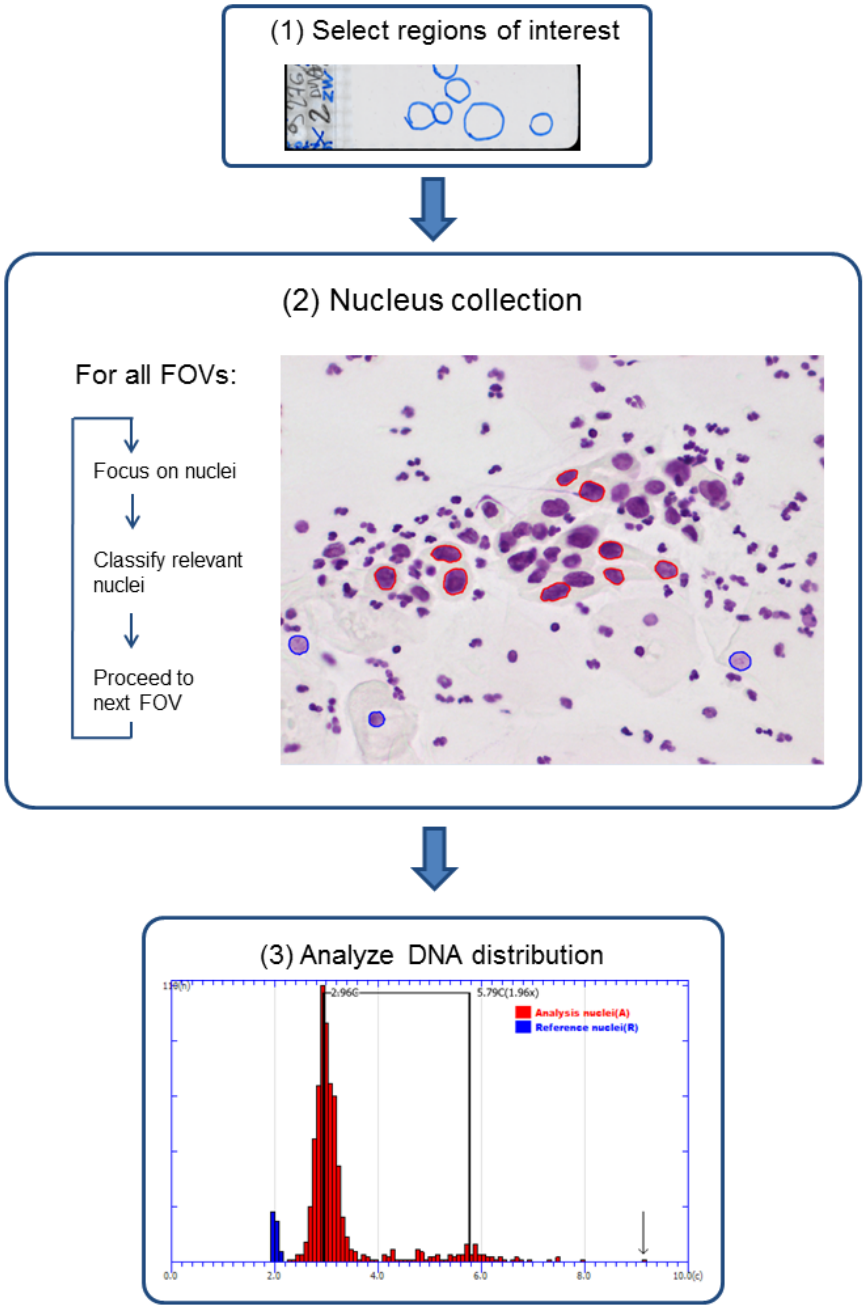
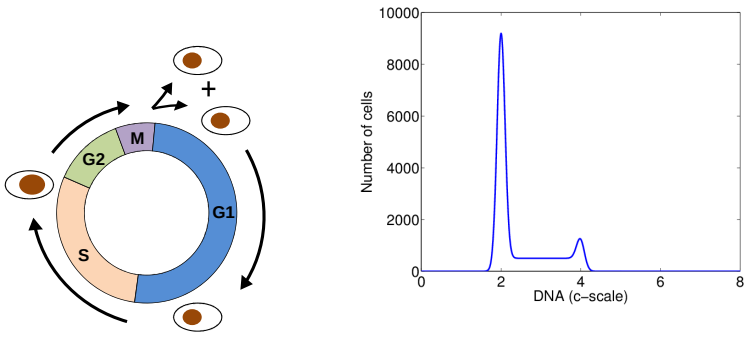
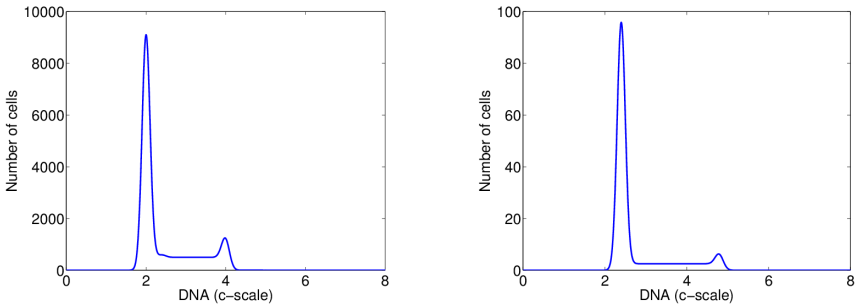


Figure 2.4: Workflow of DNA Image Cytometry: (1) The regions of interest for scanning are obtained from felt tip markings on the slide and scanned in a structured manner. (2) For each field of view available, the pathologist ensures that the present nuclei are focused. Reference and analysis cells are identified and added to the measurement by clicking on their nuclei (red=analysis, blue=reference, example from cervical smear). (3) Finally the pathologist analyzes the DNA distribution.



(a) Cell cycle **(b)** DNA distribution

Figure 2.5: (a) Cell cycle (b) DNA distribution of a stemline with healthy nuclei.



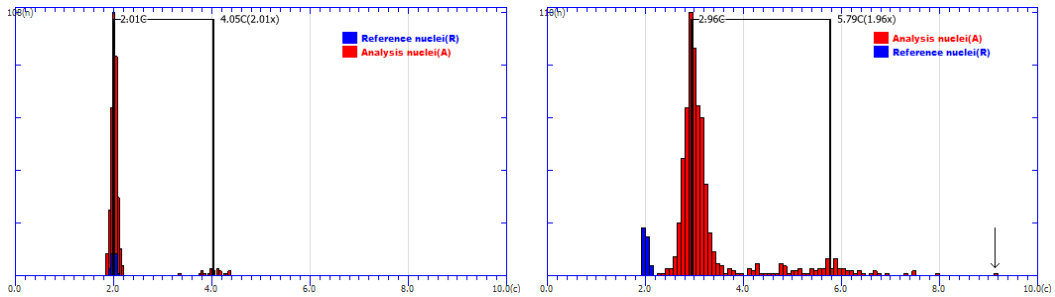
(a) **(b)**

Figure 2.6: Schematic DNA-distributions (a) derived from both normal and abnormal epithelial nuclei. The 100 abnormal nuclei are not detectable among the 10,000 normal nuclei. (b) derived from the abnormal nuclei only. Only by focusing exclusively on these nuclei, a significant peak with abnormal DNA content (2.4 c) becomes visible.

paragraph. After all regions of interest have been scanned and at least 30 reference nuclei and about 300 abnormal nuclei have been collected [8, 9, 47, 51], the expert analyzes the DNA distribution of the abnormal nuclei for diagnosis or prognosis [8].

2.2.2 Diagnosis and prognosis

The diagnosis and prognosis of DNA Image Cytometry is based on the occurrence of stemlines with abnormal DNA content (DNA-aneuploid stemline). A stemline is a proliferating cell population including all genetically identical siblings which originate from this population. If such a population is found, this implies that a population of cancerous cells which is able to proliferate exists. In the following, it is explained why the diagnosis or prognosis cannot rely on the DNA content of individual nuclei but the DNA-distribution of several hundred nuclei, why this distribution must be derived exclusively from abnormal nuclei, and how the final



(a) DNA-euploid (diploid) case—no exceeding event or abnormal stemline.

(b) DNA-aneuploid (peritriploid) case—one exceeding event and an abnormal stemline close to 3c.

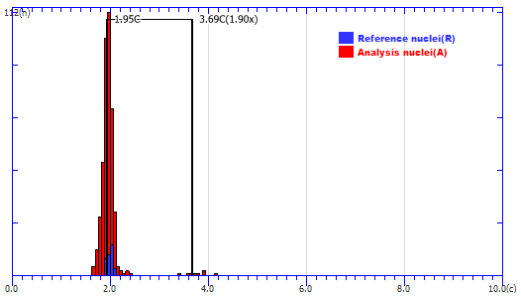
Figure 2.7: DNA histograms for diagnosis in DNA Image Cytometry. Stemlines are connected by vertical and horizontal black lines, furthermore their position is given. Exceeding events are marked by arrows.

diagnosis or prognosis is performed.

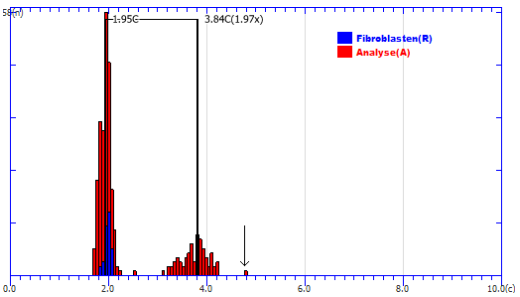
The diagnosis or prognosis in DNA Image Cytometry is based on the DNA distribution for distinguish proliferating populations with normal DNA content from those with an abnormal DNA. Figure 2.5 depicts the cell cycle of and DNA distribution of a healthy stemline. If a cell from a stemline with normal DNA content proliferates, it doubles its DNA content, as each of the siblings needs the full amount of DNA. This process of doubling the DNA is called S-phase of the cell cycle. After the DNA has been doubled, the cell grows further (G2-phase) and starts to split (mitosis, also called M-phase). The M-phase is longer than the S-phase, therefore in a healthy DNA distribution a peak at normal DNA content can be observed (the resting part of the population, G0/1-phase), a few cells in S-phase, and then again a peak at twice the DNA content from cells in the G2/M-phase (see Figure 2.5(b)). For an individual nucleus, however, its current phase in the cell cycle is not known. The DNA content of a nucleus may differ from normal DNA content due to a malignant transformation, or because the cell is in the S-phase of the cell cycle. This information, however, is revealed by considering the DNA distribution of several hundreds of nuclei, where a few cells with abnormal DNA content are explained by the S-phase, but a significant peak in this region by a stemline with abnormal DNA content.

This distribution must be derived exclusively from abnormal nuclei to achieve a high sensitivity for diagnosing cancer. In this work, the term abnormal nucleus refers to dysplastic, i.e. morphologically suspicious but not yet definitely malignant, nuclei and cancer cell nuclei. In small tumors, there might be as little as several hundred abnormal among tens of thousands of normal nuclei. If the distribution of all nuclei is considered, the S-phase of the normal nuclei might dominate over the G0/1 phase of abnormal nuclei (Figure 2.6(a)). Only by focusing exclusively on the abnormal nuclei, they become detectable in the DNA distribution (Figure 2.6(b)).

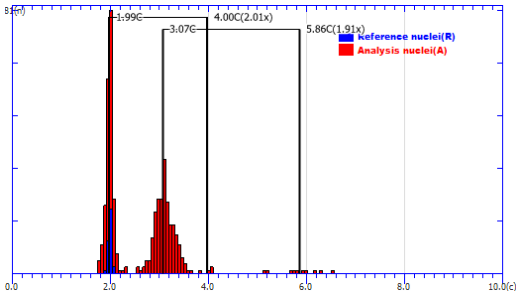
The final diagnosis and prognosis is thus performed based on the analysis of the



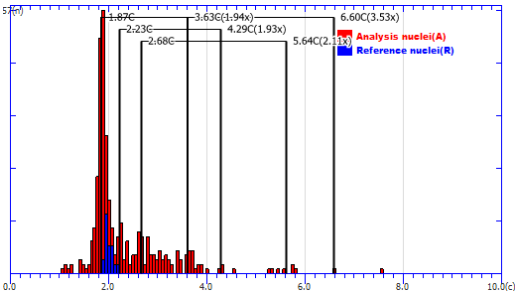
(a) Peri-diploid cancer (stemline at 2c)



(b) Peri-tetraploid cancer (stemline at 4c or exceeding events $\geq 4.4c$)



(c) X-ploid cancer (one abnormal stemline)



(d) Multiploid cancer (several abnormal stemlines)

Figure 2.8: DNA histograms for prognostication of prostate cancer by DNA Image Cytometry, from good (a) to bad (d). Stemlines are connected by vertical and horizontal black lines, furthermore their position is given. Exceeding events are marked by arrows.

DNA distribution. Healthy cases are those with a DNA distribution comparable to the one in Figure 2.5(b) (DNA-euploid stemlines). Furthermore, it is possible that healthy cells also double their DNA content, and if such a cell proliferates, it might gain quadruple DNA content. This is still considered non-malignant. However, as mentioned above a significant peak in the S-phase region, in combination with a peak at twice the DNA content, indicates a proliferating stemline with abnormal DNA content (DNA-aneuploid stemline). Likewise, nuclei exceeding certain DNA thresholds (exceeding events) can only be explained by a proliferating cell population with abnormal DNA content. Figure 2.7 gives an overview over the main histogram classes which are assigned in diagnosis of cancer (DNA-euploid and DNA-aneuploid) and Figure 2.8 the histogram classes in prognosis (peri-diploid, peri-tetraploid, x-ploid, multiploid) for cancers such as prostate or breast cancer. Note that it is common practice in DNA Image Cytometry that the histograms are represented in the c-scale, where 2 c corresponds to the full set of chromosomes and thus normal DNA content [9].

Applications of DNA Image Cytometry are:

- **Identification of cancer cells:** Cancer is diagnosed if at least one stemline

with abnormal DNA content has been identified (DNA stemline aneuploidy), or if at least one exceeding event has been found (DNA single cell aneuploidy).

- **Assessment of suspicious lesions:** In case a definite diagnosis cannot be assigned by cytological or histological examination, DNA Image Cytometry can be used for further assessment. To this end, the pathologist marks the regions on the glass slide which contain morphologically suspicious cells with a felt tip marker. The slide is restained from the original stain to Feulgen stain, and the DNA distribution of nuclei in these regions is measured.
- **Grading the malignancy of tumors:** The existence of more stemlines with abnormal DNA content or the occurrence of exceeding events may correlate with a higher grade of malignancy and thus worse prognosis.

2.2.3 Discussion

DNA Image Cytometry has been verified in numerous studies with the following performance:

- **Diagnostic accuracy:** For oral cancer, a sensitivity of 95.5 % and a specificity of 100 % have been reported, compared to the 91.3 % and 95.1 % for conventional cytology. And if both methods are combined, the sensitivity even reaches 98.1 % and the specificity 100 % [12]. For assessing suspicious lesion of the cervix, DNA Image Cytometry increases the positive predictive value of the Pap test from 35.2 % to 65.9 % [13]. On pleural effusions, the sensitivity of 50.0 % and specificity of 97.0 % of conventional cytology are outperformed by 75.0 % sensitivity and perfect specificity when additionally using DNA Image Cytometry [10, 11].
- **Early diagnosis:** Seven cases of oral cancer have been reported in which cancer was diagnosed up to two years earlier using DNA Image Cytometry compared to bioptic histology [2].
- **Prognostic value:** DNA Image Cytometry predicts the probability of survival for prostate cancer with Gleason score 7 more precisely than the Gleason score itself [42]. DNA ploidy adds significant prognostic information for low-grade low-stage prostate cancers [52–54], which are the cancers suited for active surveillance.
- **Reproducibility:** DNA Image Cytometry reaches almost perfect reproducibility, with a concordance of 94.1 percent and κ=0.87 in cervical brush biopsies [55]. The concordance is around 20 % higher than for conventional cytological methods like the PAP test [55].
- **Invasiveness and side effects:** DNA Image Cytometry performed in cytological specimen is non- or minimal-invasive (see chapter 2.1.5).
- **Costs:** While the costs for preparation and equipment are rather low (about EUR 2 per specimen, about twenty-five thousand euros for a DNA Image

Cytometry system), a DNA Image Cytometry measurement is labor intensive. As mentioned above, at least 30 reference nuclei for calibrating the system and at least 300 abnormal nuclei have to be identified among thousands to ten thousands of objects [13, 55, 56]. This usually takes more than 40 minutes per slide. Therefore, the costs do not cover the reimbursements of the German health care system (EBM 19330, 27.83 euros per case).

The high diagnostic and prognostic performance of DNA Image Cytometry is explained by five properties:

1. Different to conventional histological and cytological methods, the diagnosis and prognosis for DNA Image Cytometry is not based on *judging* the morphology of tissue, cells, or nuclei, but on *measuring* a chemical quantity.
2. By focusing the measurement exclusively abnormal nuclei, they become detectable in the DNA distribution even if there are only several hundred abnormal among tens of thousands of normal nuclei. Therefore, the method becomes more sensitive without decreased specificity [56].
3. Concerning the diagnostic accuracy, a highly relevant biomarker is measured for diagnosis—an abnormal DNA content is the most common property of cancer cells [6,7,57]. It is even proposed that abnormal DNA content is actually the cause of cancer [58]. This implies that the crucial event that transforms a cell into a malignant cancer cell is detected for the diagnosis of cancer.
4. Concerning the prognostic validity, according to Duesberg, a higher amount and variability of chromosomal sets correlates with a worse prognosis [59]. A change in chromosomal sets is linked with a change in the DNA content, thus the DNA measured by DNA Image Cytometry is a surrogate marker for the malignant potential of cancer. Highly malignant tumors, in turn, have a worse prognosis.
5. DNA Image Cytometry is highly standardized by the European Society of Analytical Cellular Pathology (ESACP) in four international consensus reports [8, 9, 47, 51]. Standardization is an effective mean to reduce errors [60].

These items ensure early diagnosis, high reproducibility, high diagnostic accuracy, prognostic validity, and trustworthiness of the results. But despite these advantages, DNA Image Cytometry is hardly used in current practice. The main reasons why the advantages have not yet reached the patient is the long time needed for one measurement, and that the reimbursement do not cover the expenses. So far, diagnostic and prognostic DNA Image Cytometry has been limited to a few specialized institutions.

2.3 State of the art

Most time in a DNA Image Cytometry examination is needed for the systematic scanning of the slide for the identification of relevant nuclei. Additionally it is vital

for DNA Image Cytometry that high resolution digital images of the relevant objects are acquired, as the DNA content is computed from these. DNA Image Cytometry is less labor intensive for the pathological expert if parts of the measurement are carried out by machines:

- **Digitization of the slide:** can be realized by a motorized microscope with digital camera and autofocus system.
- **Identification of diagnostically relevant nuclei:** Can be realized by a filtering approach or a pattern recognition system. For filtering, a filter is designed which yields a high response for objects of interest, and those are then considered for diagnosis and DNA measurements. A pattern recognition system aims to learn the patterns an expert uses to classify nuclei. All objects in a field of view are segmented and the pattern recognition system then assigns one of the occurring classes.

The following approaches for increasing the productivity of DNA Image Cytometry have been proposed in literature:

2.3.1 Leiden Image Cytometer

The Leiden Image Cytometer, also known as LEYTAS, was developed by Ploem et al. from the University of Leiden. it has been used for screening for cervical cancer [16], identification of urothelial cancer cells in urine diagnosis of bladder cancer from urine [17] and the prognositfication of ovarian cancers and soft tissue malignancies [61, 62].

The Leiden Image Cytometer utilizes a motorized Leitz Orthoplan microscope with a 40× objective (NA 1.3) and a TV camera (Bosch Fernseh GmbH) for the digitization of nuclei [15]. Abnormal nuclei are detected by filtering the absorption image with a circular filter, which detects large and dark objects. Using two different parameter combinations, two classes of objects are detected: *high level objects* and *low level objects*. High level objects are even larger and have a higher absorption than low level objects. An artifact rejection routine excludes dirt, overlapping nuclei and other undesired artifacts from the high and low level nuclei [14]. The system offers the possibility to verify the detected objects by rapid visual inspection by a human expert. The final diagnosis for a slide is based on the number of high and low level objects detected: A slide is considered positive if at least one high level object is present, when the absolute amount of low level objects is higher than 30 or if the relative amount of low level nuclei is above 0.1 % [63]. In later versions of the system, the DNA content is computed by calibrating the system with lymphocytes, again detected by filtering. Then, the diagnosis can also be based on the DNA distribution or existence of 4.5c exceeding events [17, 62].

For the screening of cervical cancer, the diagnostic accuracy of the fully automated system is 0.3 % false-negatives and 12 % false-positives [16]. For bladder cancer, 2 % false-negatives and 33 % false-positives are diagnosed [17].

However, for evaluating the performance of the filtering algorithm for detecting the high and low level objects, only the number of detected objects is given. The number of relevant objects which *should have been detected* remains unclear. It has to be assumed that many objects are overlooked, as the number of objects which are detected by the system is relatively low: Only 273 nuclei among 940000 objects from 18 slides have been detected. This corresponds to 0.03 % [14] and is much lower than the 12.5 % of abnormals which have been identified by a pathological expert among thirty-five thousand objects from cervical brush biopsies (see Section 3.3.5). Additionally, using the statistics of high and low level nuclei as biomarker tends to produce many false-positive results: In the fully automated scenario, 33 % of the urine samples are falsely classified as positive. Even after manual correction, this number remains as high as 12 %. The extended system, where also the DNA content can be computed, seems to lack precision in DNA measurements. In [64], it is stated that it is difficult to discriminate diploid and near diploid DNA stemlines.

2.3.2 Cyto-Savant

In a joint cooperation Oncometrics Imaging Corporation and the Cancer Imaging Department of the British Columbia Cancer Agency developed the Cyto-Savant system. This system was initially developed for two applications: Firstly, as an automated pre-screener for cervical cancer, which would reject 50 % of all negative slides without need for manual processing. Secondly, dysplastic lesions of the cervix, where no definite diagnosis can be achieved using conventional methods [20], should be assessed. Later, the Cyto-Savant system has been extended with a nucleus classifier for sputum and used for the detection of lung cancer [22].

Both the microscope and the camera are specially designed for the intended application in image cytometry (a solid state microscope [18] using a PlanApo 20X objective from Nikon and a MicroImagager 1400 camera from Xillix Technologies Corp. with a CCD sensor, 1038x1350 resolution [65]). The Cyto-Savant uses a pattern recognition system approach. Its gold standard consists of a compilation of about one million objects from 1100 cervical slides. These objects have been classified by technologists. A set of 150 features is used [20, 21]. For classifying nuclei, a decision tree is used, which applies either thresholds or linear discriminant functions at the nodes for decision making [19]. The Cyto-Savant scans about two to five percent of the area of the slide and then performs a diagnosis. It based on the number of objects in certain cell classes and statistics of the features of these objects.

In a prospective study, these statistics were adapted on a training set of about 1100 slides, so that 50 % of the negative smears in screening for cervical cancer are rejected. Applied on the test set, 26 % mild, 11 % of moderate and 4 % of severe dysplasias were wrongly rejected as well. For the detection of stage 0 and stage 1 lung cancer, a sensitivity of 45 % and a specificity of 90 % is achieved [22].

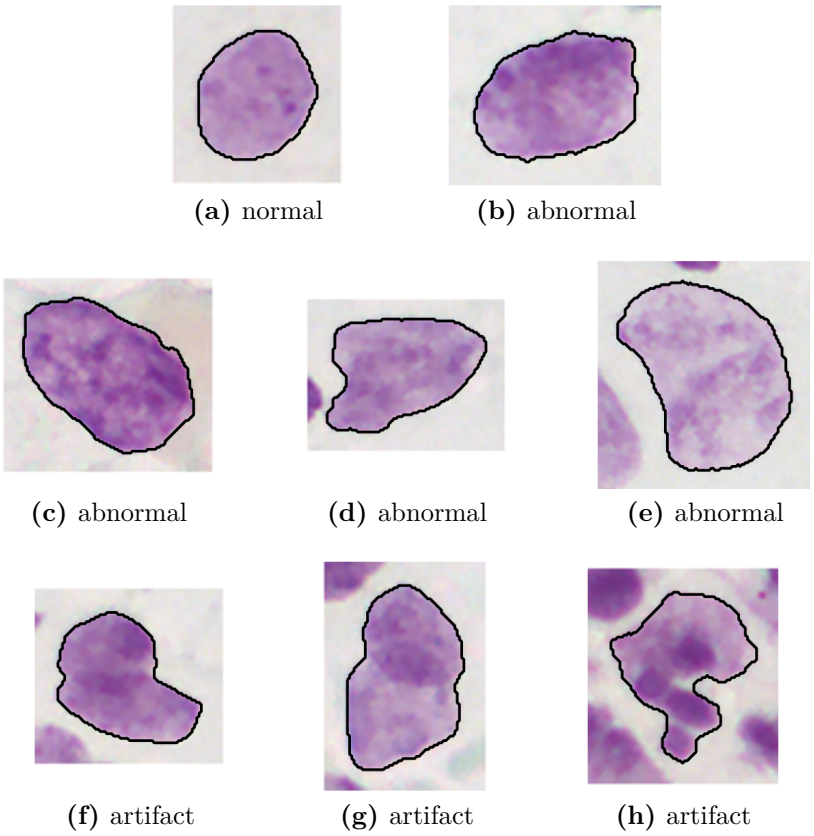
The performance of the nucleus classifier was, however, either not evaluated [20] or is only about 75 % [66]. Only decision trees were used for classifying nuclei, and a statistician is needed to manually set up and tune the classifier [21]. The most severe limitation of the Cyto-Savant system is classification of normal and abnormal nuclei: Instead of morphology, their classification is based on the DNA content only: Nuclei are classified abnormal if they have a DNA content higher than 2.6c [67] or 3c [21]. Another aspect is that the final diagnosis is based on nucleus numbers and feature statistics of objects from a small fraction of the slide (2 % - 5 % [20]). Different from the usual approach, the identification of DNA stemlines with abnormal DNA content, these statistics are hard to comprehend by the user. In summary, neither the classification of abnormal nuclei nor the way the diagnosis is made are in conformity with international guidelines for DNA Image Cytometry [9].

2.3.3 Schneider et al.

Schneider et al. continued research mainly for pattern recognition systems for Feulgen-stained nuclei from serous effusions and the oral mucosa [23, 68–70].

Nuclei were digitized using a 3 chip CCD camera and a 63x oil immersion objective. In a first experiment, they set up a pattern recognition system for discriminating several cells from serous effusions (mesothelial cells, lymphocytes and granulocytes) [68]. Their gold standard contained 3100 objects. From a set of 112 features described in literature, they selected a subset by feature selection strategies. A kNN or Fuzzy kNN classifier was used for classification, which reached an overall correct classification rate of up to 97.5 %. More importantly, in an extension of their research they showed that the discrimination of normal and abnormal nuclei of the oral mucosa is possible with a correct classification rate of up to 95 %. This was achieved by enlarging the feature set to 203 features, again selecting a subset using feature selection, and by using kNN, Fuzzy kNN and Support Vector machines as classifiers [23, 69, 70].

The fact that normal and abnormal nuclei are discriminated based on morphology would overcome a severe limitation of the Cyto-Savant system. But these results only describe the differentiation of abnormal and normal nuclei. However, objects from other nucleus classes such as artifacts or nuclei from the immune system are also present on the slide and can make up more than 75 % of all objects on oral slides [71]. The pattern recognition system will assign these objects to one of the two classes present in the gold standard. Thus all artifacts and nuclei from the immune system will be misclassified and considerably impurify the nucleus classes. Therefore, this approach is limited to the evaluation on a training and a test set containing only normal and abnormal nuclei, but experiments in clinically relevant applications have not been published.



(a) normal **(b)** abnormal

(c) abnormal **(d)** abnormal **(e)** abnormal

(f) artifact **(g)** artifact **(h)** artifact

Figure 2.9: Nuclei and artifacts from a brush biopsy of the uterine cervix. First row: A normal epithelial nucleus compared to an abnormal nucleus. Second row: Abnormal nuclei which demonstrate the variation in appearance of abnormal nuclei. Third row: Artifacts which demonstrate their variation in appearance and the similarity to abnormal nuclei.

2.3.4 Conclusions

To conclude, state-of-the-art approaches for automating DNA Image Cytometry lack the ability to identify morphologically abnormal nuclei among all other objects on the slide [14–22]. Therefore, these approaches cannot make use of the essential enhancement of sensitivity due to the restriction on abnormal nuclei (see section 2.2). As a consequence, they suffer diagnostic accuracy [17, 20] and are not in line with international guidelines of DNA Image Cytometry [8, 9].

The challenge for distinguishing abnormal from normal nuclei is that in case the malignant transformation to cancer is in its beginning, the visual difference between normal and abnormal nuclei is slight, but needs to be detected (see Figure 2.9(a) and 2.9(b)). In case the malignant transformation has progressed further, cancer cell nuclei have a huge variation in their appearance (see Figure 2.9(c) to 2.9(e)), since cancer evolution is a chaotic process [72]. Classifying such highly variable data is naturally a difficult task [73, page 12].

Schneider et al. demonstrated that it is in principle possible to discriminate abnormal from normal nuclei in brush biopsies from the oral cavity [23, 69, 70]. However, their approach is only restricted to these two classes of nuclei. Further types of objects are present on the slide and need to handled appropriately by the classifier. Therefore a system suitable for routine must be able to distinguish abnormal nuclei from all other nucleus classes present on the slide, and exclude artifacts such as overlapping nuclei and defocused objects. Such a system has not yet been developed. Since the difficulty in pattern recognition usually increases with the number of classes [74], despite the advances achieved by Schneider et al. this remains a challenging problem. Besides discriminating normal and abnormal nuclei, especially discriminating artifacts and abnormal nuclei is a task yet to be solved. Similar to abnormal nuclei, artifacts have a huge variation in their visual appearance. Furthermore, artifacts and abnormal nuclei share similar visual properties such as a ragged contour or a variation in intensity values (see Figure 2.9(f) to 2.9(h)).

3 Image processing algorithms

The time-consuming steps of manual DNA Image Cytometry are the systematic scanning of regions of interest (ROIs) and the identification of relevant nuclei. This chapter presents algorithms for accomplishing these steps automatically.

The first three sections are devoted to the automated acquisition and processing of individual fields of view (FOVs): First, an overview of the image acquisition for manual and automated DNA Image Cytometry is given. Next, an algorithm for the precise, unsupervised segmentation of nuclei within these FOVs is presented. The subsequent chapter presents the core component of this work, classifiers for automatically identifying relevant nuclei among the automatically acquired and segmented objects. Finally, in addition to the processing of the individual FOVs, the last section is about an algorithm for the automated identification of ROIs in overview scans of slides. This identification unburdens the operator from manually selecting the ROIs prior to scanning. Figure 3.1 is a flow chart illustrating the essential steps for the automated identification of nuclei.

3.1 Image acquisition

In this section, the image acquisition for manual and automated DNA Image Cytometry is described. Since the DNA content of nuclei is measured on digital images of the nuclei (see chapter 2.2.1), high quality digital images of nuclei in focus have to be acquired to guarantee accurate measurement results. For manual DNA Image Cytometry measurements, the following setup was used throughout this work: the microscope utilized is a Motic BA410 with a 40× objective (NA=0.65). The system is equipped with a MotiCam Pro 285A camera (2/3" Sony ICX 285 CCD Color Sensor, 1360 × 1024 pixel, 8 bit). The camera is mounted by means of a 0.87× C-mount adapter. For this optical configuration, one pixel in the image corresponds to $0.18 \times 0.18\,\mu m^2$ on the slide. Thus one image displays $245\,\mu m \times 184\,\mu m$ of the slide, and for manually scanning an area of $1\,cm \times 1\,cm$ about 2.200 FOVs need to be analyzed. The microscope has a beam splitter, which distributes the light to the eyepiece (20 %) and to the camera (80 %). The live image acquired by the camera is displayed on a computer screen. The expert can select relevant nuclei by mouse-clicking, an image of the selected nucleus is then stored digitally. For the automated acquisition of nuclei, a Motic BA600 microscope is used. It has the same optical components and the same camera as the manual system, but in addition a motorized scanning stage and objective revolver. For the automated workflow, first

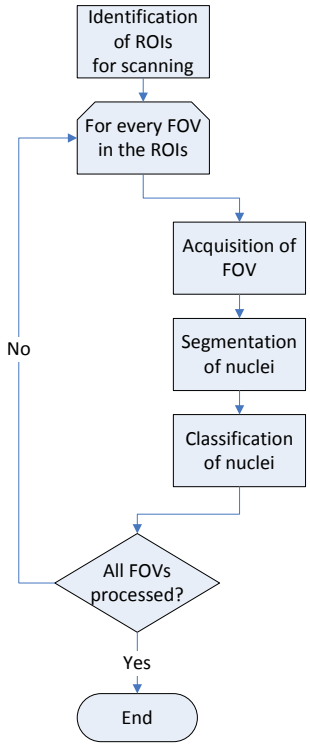


Figure 3.1: Flow chart illustrating the automated acquisition and classification of nuclei.

an overview scan is obtained at low microscope magnification. The user then selects the ROIs for scanning, for instance areas with high cell density, or ROIs identified in a previous examination. Next, the autofocus procedure of the Motic MMS platform is used for scanning these ROIs [71, 75]. First a set of uniformly distributed focus points is determined, and for each FOV to be scanned an initial z-position is interpolated from the z-positions of the neighboring focus points. A z-stack of images is acquired around this position, and the scene of best focus is chosen based on a focus score. As focused images contain more sharp transitions than defocused ones, the filter response of the Roberts edge detector is used for this purpose. After the scene of best focus has been found, all objects in the scene are segmented (see chapter 3.2), and the digital images are stored. Digitizing an area of $1\,\mathrm{cm}\times 1\,\mathrm{cm}$ at $40\times$ magnification takes about 31 minutes, but does not require manual interaction.

Before the actual acquisition of digital images, the system is calibrated. The following calibration protocol for was used for every measurement of this work, be it manual or automated:

1. The system is set up for **Köhler illumination**. This method for specimen illumination, named after its inventor August Köhler, produces a homogeneous illumination of the specimen and reduces glare [76].
2. The **exposure time** of the camera is adjusted. First, the user controls the light intensity of the light source of the microscope so that a visual inspection

through the eyepiece is convenient. Subsequently, the exposure time of the camera is adapted so that the regions without any cells yield an average sensor response of 90 % of the maximum sensor response in the green channel.

3. A **shading correction** is applied on every FOV acquired. This procedure eliminates further irregularities in the illumination, for example shading effects caused by uneven illumination, small misalignments of microscope components, impurities on optical surfaces or thermal noise of the sensor [77]. The original image I_{in} is corrected via

$$I_{\text{out}} = g\frac{I_{\text{in}} - b}{w - b}, \tag{3.1}$$

where w is the white reference image acquired at an empty position, b is the black reference image which has been acquired without any light impinging on the sensor, and g_{max} a correction factor to scale the image to the original range.

The benefit of this calibration is two-fold. First, it reduces errors by the microscopic image formation. Second, nuclei are acquired under the same imaging conditions. This standardization is beneficial for the segmentation and classification algorithms, which then only have to deal with nuclei which are evenly illuminated and have similar pixel intensities.

3.2 Segmentation of nuclei

After the automated acquisition of a FOV, the next step is to find the contour of clinically relevant nuclei in these FOVs as precisely as possible. Figure 3.2 shows an example of an automatically collected FOV of an oral smear.

Both the algorithms for measuring DNA as well as the classification of nuclei depend on this segmentation: in case a part of a nucleus is not included within the segmented boundary, it does not contribute to the DNA value of the nucleus. In the worst case, this can lead to a false-negative diagnosis. If a nucleus is segmented imprecisely, this influences the morphological features extracted for the nucleus classifier, thus causing misclassifications. Finally, should a relevant nucleus not be segmented at all, it will not be available for the subsequent analysis steps.

The software for manual DNA Image Cytometry employs a thresholding algorithm in HSV colorspace to separate nuclei from background, with a median filtering as preprocessing and morphological operations as postprocessing [78]. It has been validated and used in clinical practice for several years. However, applying it for the automated scenario has two drawbacks. First, too many irrelevant objects are segmented. In case the user selects an object in manual DNA Image Cytometry, only the object containing the position of the mouse click is extracted. If this segmentation is applied to a whole FOV, objects such as dirt particles or glass splinters

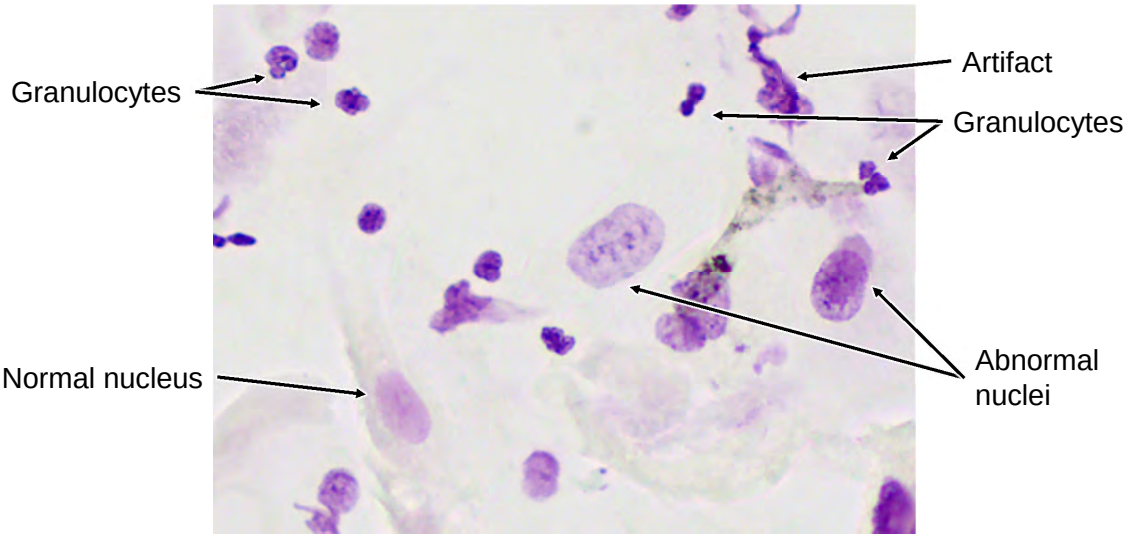


Figure 3.2: Parts of an FOV of an oral smear.

are also segmented—objects the user never would select. Second, missegmentations such as imprecise contours or the joint segmentation of touching nuclei occur. In the manual scenario, the user can call more complex refinement or splitting methods. However, user interaction should be as much as possible in the automated scenario, and these algorithms are too complex to apply them on every object.

As the segmentation of nuclei or cells is usually one of the first tasks in any automated detection system in digital pathology, a rich literature on this topic is available. General overviews can be found in [79–81]. Nuclei are for instance segmented by (adaptive) thresholding [78, 82–84], Bayesian classification of pixel values [85], graph cut approaches [86, 87] or Mean-Shift-Segmentation [88]. Besides these general approaches, many algorithms to correct missegmentations have been published. For imprecise initial segmentations, *contour refinement algorithms* can be applied. These algorithms are for instance based on Active Contours [84,89,90], level sets [85], morphological operations [91], dynamic programming [78, 92], or a maximum likelihood approach [93]. In [94], the low frequency Fourier descriptors of the contour are optimized for finding a contour which is optimal with respect to a cost function. Another type of missegmentation is the segmentation of several individual touching nuclei as one object (undersegmentation). In that case, the segmentation masks of the individual nuclei should be found by a *splitting algorithm*. Many splitting algorithms use the watershed algorithm [95, 96], applied on the distance transform or gradient images [97–103]. Another approach uses the Top-hat filter to highlight background pixels which are surrounded by darker pixels [78], as it is the case for the space between two touching nuclei. A completely other group of algorithms are those solely based on the geometry of the object. Concavity points or dominant concave regions are detected and iteratively split along the most probable split lines [104–108].

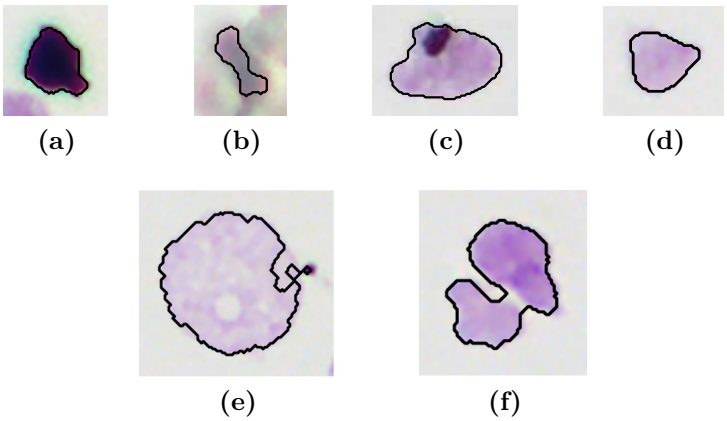


Figure 3.3: Examples of irrelevant objects (first row) and missegmentations (second row) of the algorithm from manual DNA Image Cytometry. (a) dirt, (b) glass splinter, (c) nucleus with dirt, (d) nucleus fragment, (e) imprecise contour, (f) touching nuclei.

As mentioned above, the segmentation algorithm of the manual DNA-Image Cytometry finds too many irrelevant objects (see Figure 3.3). Missegmentations are corrected by manually calling more complex algorithms, which is not applicable for the automated scenario. Many of the state-of-the-art contour refinement or splitting algorithms are applied to all objects, but no or only little effort is spend to identify the objects with need for improving the segmentation. The contribution of this work is an integrated framework for automated segmentation of all nuclei in automatically acquired FOVs. This segmentation is more specific by not only separating objects from background, but by learning the stain color occurring for nuclei. Nuclei with dirt and nucleus fragments however, cannot be distinguished based on color values of pixels - they are rejected based on features which can be rapidly computed. Finally, several features for detecting the need for refinement/splitting as well as the actual algorithms for these tasks are optimized and compared with each other. Parts of this work have been published in [94, 109].

3.2.1 Material

Nuclei or whole FOVs were acquired using the microscope and camera configuration for manual DNA measurements (8 bit RGB images, see Section 3.1). All nuclei were stained with the purple pararosanilin stain.

Several data sets are used for training, optimizing and evaluating the segmentation algorithm; an overview can be found in Table 3.1. A training set for learning the color of nuclei was extracted from databases of annotated nuclei segmented with the algorithm of manual DNA Image Cytometry. These nuclei are used for training nucleus classifiers (see section 3.3). The advantage of this approach is that all objects in this set which are not classified as artifact, but into one of the nucleus classes,

Table 3.1: The data sets used for training, optimizing and evaluating the algorithms.

Set	#nuclei	Type
Color training set	19739	Thumbnails
Nuclei with dirt	5370	Thumbnails
Nucleus fragments	16996	Thumbnails
Contour refinement	536	Thumbnails and manual segmentation
Splitting	2013	Thumbnails
FOV training set	1072	From 80 FOVs with manual segmentation
FOV test set	812	From 80 FOVs with manual segmentation

have correct segmentation verified by a cytopathologist. Thus a large database is readily available. From this color training set, further subsets were extracted to train algorithms for detecting nuclei with dirt or nucleus fragments, for the refinement of contours and for splitting touching nuclei. As a second training set, a manual gold standard segmentation was delineated for all nuclei within whole FOVs. In total 80 FOVs were annotated. Finally, again 80 FOVs with manual gold standard segmentation are used as test set. The images for the FOV sets originate from prostate, cervix, oral and effusion specimen, uniformly distributed among the four modalities. Under no circumstances two sets contain data from the same patient.

3.2.2 Methods

The developed segmentation pipelines contains algorithms for the initial segmentation, the rejection of irrelevant objects, the conditional refinement of nuclei, and the splitting of undersegmented touching nuclei (see Figure 3.4). To motivate the order of these steps, the rejection of irrelevant objects is performed directly after the initial segmentation to reduce the number of objects which need to be processed subsequently by the more complex algorithms. The refinement is performed before the splitting, as the refinement possibly could merge formerly unconnected objects which then need to be split again. Moreover, the decision for splitting nuclei is based on the morphology, and therefore the contour should already be as precise as possible.

Before describing the details of each step of the pipeline, two general remarks are made: In case an algorithm requires a gray image instead of a color image, the green channel is used if not stated otherwise. This is because the pararosanilin stain has the highest absorption in this channel, thus background and nucleus can be discriminated best in this channel. Algorithms for the detection of a certain object type are optimized by a Receiving Operator Characteristics (ROC) analysis: the threshold for a detection task based on a single feature is set such that all objects of interest are detected (100 % sensitivity) and the specificity is highest.

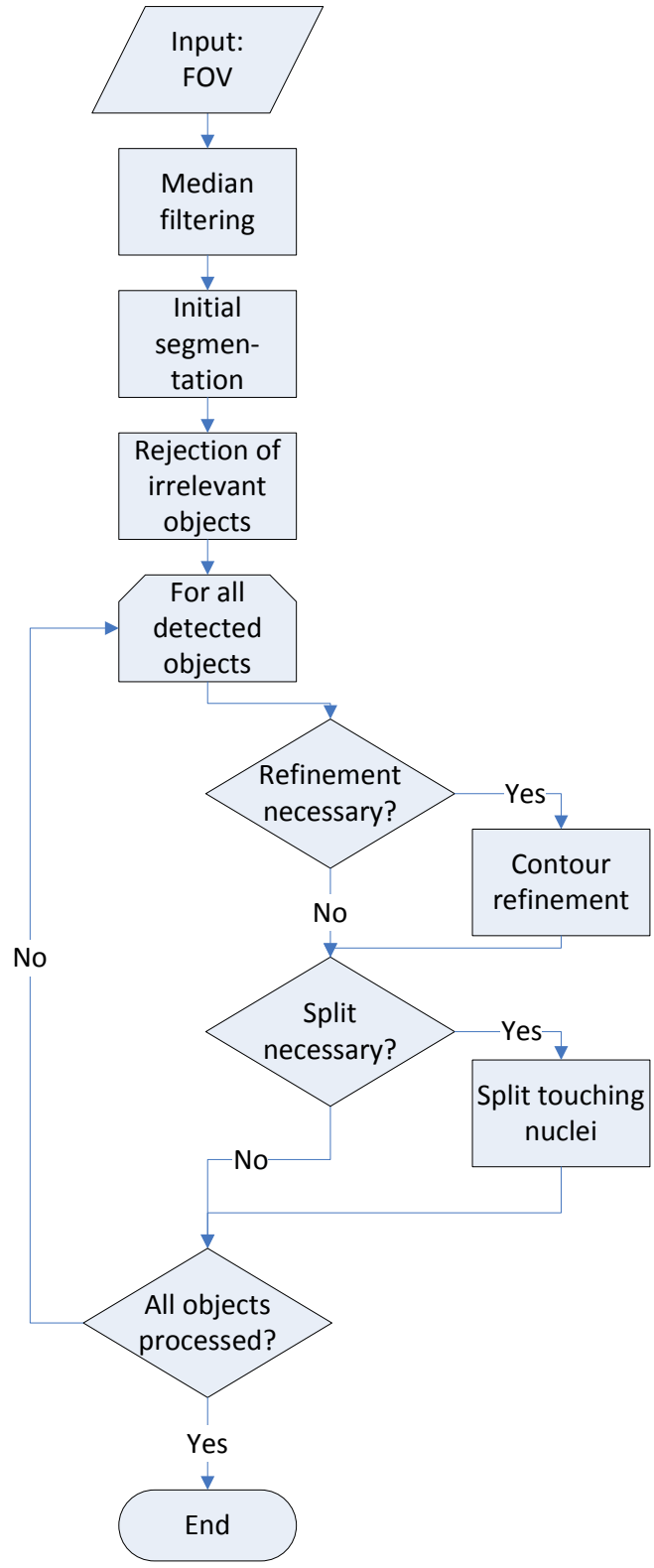


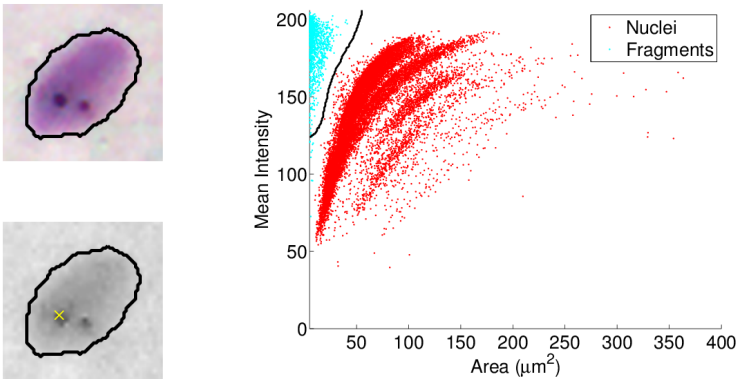
Figure 3.4: Flow chart illustrating the segmentation pipeline developed in this work.

3.2.2.1 Initial segmentation

To make the initial segmentation more specific for nuclei than the segmentation from manual DNA Image Cytometry, it is based on the color values of nuclei learned from the color training set. All initial segmentation methods are pixel based approaches, that is for each pixel in the image it is decided based on its color values whether it is labeled as nucleus. Previous experiments demonstrated that it is beneficial to consider the HSV colorspace rather than the RGB colorspace [110], thus the work presented here is limited to this colorspace. Three approaches were compared:

- **Two-sided thresholding:** In addition to the thresholds for separating background from objects as in the manual algorithm, further thresholds are introduced for separating nuclei from irrelevant objects. These thresholds were found by determining the intervals which contain 99 % of all pixels belonging to nuclei in the color training set by a ROC analysis. The analysis yielded the following intervals: all pixels with a hue component in $[0.69, 0.87]$, a saturation component in $[0.05, 1]$, and a value component in $[0.59, 0.90]$ are segmented as nucleus.
- **Maximum likelihood classification:** In the color training set, 37.61 % of all pixels belong to nuclei, whereas 62.39 % belong to background or irrelevant objects. For each possible HSV-combination, it is counted how often it occurs in the color training set and how often it occurs inside or outside the nucleus. If the percentage being inside a nucleus is above 37.61 %, it is more likely that this combination belongs to a nucleus. Thus pixels with this HSV-combination are labeled as nucleus.
- **Classification:** The maximum likelihood approach yields a good initial segmentation for known pixel values, however it classifies all unknown pixels as background. This might be too restrictive, considering that there is a small inter-slide variability for staining intensity and color. To allow for a better generalization, classifiers are trained using the hue, saturation and value component as features, and the classes 'belonging to a nucleus' and 'not belonging to a nucleus'. In [110], a k Nearest Neighbor (kNN), Adaboost, and a Support Vector Machine (SVM) classifier and their parameters were analyzed. As a kNN with k=41 turned out to yield the best performance, this classifier is used in this work.

For speed-up, the output of these methods is stored in lookup tables. That is, for every possible combination of RGB-values of the image, it is stored whether this combination is labeled as nucleus (including the conversion to HSV colorspace). To perform an initial segmentation, for every pixel of the image the corresponding lookup table entry is transferred to the segmentation mask. As a preprocessing step, a median filter is applied. The width of this filter has been optimized on the FOV training set (0.54 µm to 2.7 µm, steps of 0.36 µm). For postprocessing, an opening and a closing operation are performed, using a circular structure element of radius 0.36 µm as proposed in [110]. Holes within objects are filled using a flood-fill



(a) (b)

Figure 3.5: (a) Top: Nucleus with dirt, bottom: Blue channel, the position with minimal filter response of the DarkParticle feature is marked with a cross, (b) area and mean intensity of nuclei and fragments. The black line is the decision boundary of the kNN classifier used to separate the two classes.

operation. Finally, objects with an area smaller than 4 µm^2 and all objects touching the boundary of the FOV are removed.

3.2.2.2 Rejection of irrelevant objects

Whereas the initial segmentation effectively prevents the segmentation of dirt or glass splinters, nuclei containing dirt (Figure 3.3(c)), or nucleus fragments (Figure 3.3(d)) are still segmented. As these objects cannot be distinguished solely based on their color values, they are rejected based on a set of features.

The DNA content of nuclei containing dirt cannot be measured precisely, thus there is no reason to include them in the further analysis. In the subtractive image formation of light microscopy, the blue light is almost not absorbed by the pararosanilin stain of nuclei, whereas dirt strongly absorbs light in all color channels. Therefore dirt particles are dark in the blue channel, but nuclei are hardly visible (see Figure 3.5(a)). The presence of a dirt particle within an object is quantified by the feature DarkParticle: The object is filtered with a circular filter and the minimal filter response is computed. The feature thus yields low feature values for objects containing dirt. It has two parameters, the size of the filter element and whether the feature should be normalized with the mean intensity of the nucleus. Objects are then rejected if the feature DarkParticle is below a certain threshold. The threshold for this decision as well as the best parameters are found on the training set by the ROC analysis described at the beginning of the methods-section. Here the threshold is set such that all nuclei *without* dirt are kept and as many objects with dirt are rejected.

During the preparation of the slide, mechanical stress is induced on the nuclei which

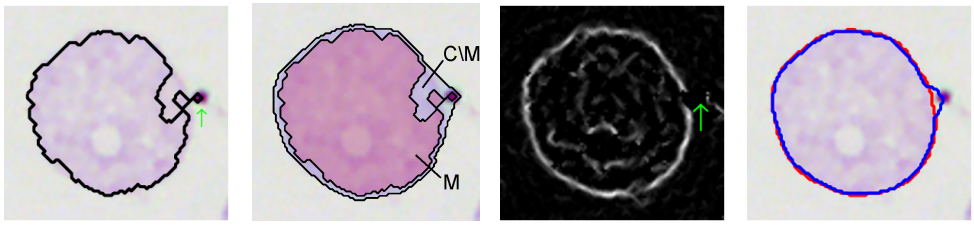
might disintegrate them into several small fragments. As these fragments originate from nuclei, they have the same color. But they contain only a small fraction of the nucleus. But as they contain only a small fraction of the nucleus and its DNA, they are both smaller in size and brighter in the stoechiometric staining. Therefore, they can be distinguished based on their area and their mean intensity in the green channel. Figure 3.5(b) shows a scatter plot of these two features for fragments and intact nuclei. The two classes are well separated and can be distinguished straightforwardly by a kNN classifier (k=40, euclidean distance).

3.2.2.3 Detection and refinement of imprecise contours

Occasionally, nuclei are segmented imprecisely (Figure 3.6(a)). This occurs most frequently for very dim nuclei, which have close color values to background. As there is no need to apply a contour refinement algorithm to correctly segmented nuclei, it is first checked if refining the segmentation is necessary (conditional refinement): the refinement is only performed if a certain feature value is below a threshold. As refinement methods, an Active Contour approach and a level set approach are examined.

Two features were developed for detecting the need for refinement. For correctly segmented nuclei, the contour of the segmentation mask coincides with the transition between background and nucleus, which usually is an area with a higher gradient compared to the rest. Therefore the gradient magnitude along the contour is exploited as a first feature, the feature GradientOnContour. The final feature value is computed as the p %-quantile of these magnitude values in the green channel. As a second feature, the morphology of the nucleus and the possible presence of dye in the neighborhood of the initial segmentation are considered: Nuclei are usually elliptical and thus convex. Imprecisely segmented nuclei, however, have a ragged and non-convex contour. The same is, in fact, also true for artifacts, but a visual analysis of nuclei from the training set with an incorrect initial segmentation revealed that usually some dim dye is visible within the convex hull (see Figure 3.6(a) for an example). The feature ConvexHullIntensityRatio quantifies this observation by first computing the convex hull C for the initial segmentation mask M. Then, the feature value is computed as the ratio of the mean green channel intensity in the regions $C \setminus M$ and S. For very dim nuclei with imprecise segmentation, the segmentation might not reach the boundary of the nucleus (see Figure 3.6(b)). As the presence of dye outside the initial segmentation and its convex hull is also an indicator for an imprecise contour, it turned out to be beneficial to dilate the convex hull slightly before computing the mean intensity.

Each of this features has one parameter which needs to be optimized. For the feature GradientOnContour, this is the quantile p of the gradient magnitude values along the contour (0 to 1, in steps of 0.1), and for the feature ConvexHullIntensityRatio the extend of the dilation of the convex hull (0, 0.18, 0.36, 0.54 and 0.72 µm). These parameters are optimized on the training set for the contour refinement. The



(a) (b) (c) (d)

Figure 3.6: Refinement of nuclei: (a) Nucleus needing refinement (b) Computation of the ConvexHullIntensityRatio feature, which is the fraction of the mean intensities within $C \setminus M$ and M (c) Cost function for snakes. The gradient of the dirt particle, marked by the green arrow in (a) and (c), is removed, (d) results of refinement (red=GVF snakes, blue=level set)

threshold is set such that all objects which need refinement are detected, and the rate of objects for which the refinement is performed without an actual need is minimized.

For the refinement itself, one approach studied is the gradient-vector-flow (GVF) snake proposed by Xu and Prince [111]. Generally, snakes start from an initial contour and then iteratively adapt the contour to minimize a tradeoff between external energies ("How good does the contour fit to the image data?") and internal energies encoding prior knowledge about the morphology of the contour (Kass et al. [112]). The original snakes however, have a limited capture range; they fail in finding the correct segmentation if the initial segmentation is far away from the optimal solution [111]. The extension to the GVF snake overcomes this drawback by employing a vector field as external energy instead of using image data. In this vector field, the vectors point to desired points in the image, for example the edges of a nucleus. By using a regularization method, reasonable estimates for these vectors are also computed in regions further away from edges. This increases the capture range effectively. For the problem at hand, the basis for the external energy is a cost function with high values at the edges of a nucleus. First, the green channel is median filtered to remove noise (filter size 0.54 µm). Then, the gradient magnitude is computed. Other nuclei in the image and dirt particles might distract the snake, as they also have edges with a strong gradient. To remove the influence of other nuclei, only gradients pointing towards the centroid of the initial segmentation are considered. As demonstrated for the feature DarkParticle, dirt can be identified in the blue channel. To remove the influence of dirt particles, the gradient magnitude is also set to zero for all pixel where the blue channel is below 205 (see Figure 3.6(c)). Finally, the gradient-vector-flow field is computed from this cost function. As internal energies the first and second derivatives are used (weighted with scalars α and β). Additionally, as an imprecise segmentation usually is located in the interior of the object, a balloon force δ is added. This balloon force pushes the contour in the direction of the outward contour normal and thus into the direction of the correct

Table 3.2: Parameters of the contour refinement methods.

Method	Parameter	Range
GVF snake	α	0:0.4:0.8
	β	0:0.5:2
	δ	0:0.1:0.4
	Number of iterations	{50, 100, 200, 400}
Level set	median filter size	0.54:0.36:2.7
	μ	0:20:100
	Number of iterations	{30,50,100}

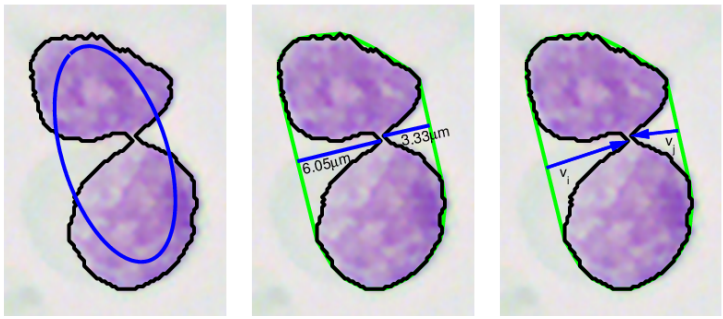
segmentation.

Another refinement approach used is the level set approach by Chan and Vese [113]. The aim of this approach is to partition the image into regions with homogeneous intensity by minimizing an energy functional depending on the contour $\mathcal{C}$, its length, and the deviation of the image data from the mean intensity inside and outside of $\mathcal{C}$ (see [113, p. 268]). This functional is reformulated as a level set approach, that is, the contour is implicitly represented by a function $\phi(x, y)$, where $\phi(x, y) = 0$ is the contour, $\phi(x, y) > 0$ are the regions inside, and $\phi(x, y) < 0$ are the regions outside $\mathcal{C}$. This reformulated functional can then be optimized by a discretization and linearization of the corresponding Euler-Lagrange equations and regularization methods. As input image, the green channel is used, and as the method works best for homogenous regions it is preprocessed by median filtering. A weight μ is used to balance the influence between the length of the contour and image information.

A pre- and a postprocessing step are identical for both the GVF snake and the level set approach. As preprocessing, it is beneficial to use the convex hull of the initial segmentation as initial contour, as it is usually closer to the desired segmentation than the initial segmentation. As postprocessing, tests of the refinement methods on the training set revealed that the segmentation is improved in most of the cases, but that there are a few cases where the segmentation becomes much worse. Therefore the refined contour is only kept if the criterion for the detection of the refinement is higher for the refined contour than for the initial segmentation. Both refinement methods have parameters which need to be optimized. This optimization is performed on the training set for refinement by an exhaustive search of the parameters. As an optimization criterion, the maximum distance between the gold standard segmentation and the refined segmentation (Hausdorff distance) is used. Table 3.2 lists the parameters which are optimized and their ranges. Figure 3.6(d) exemplarily shows the refinement of an imprecisely segmented nucleus.

3.2.2.4 Detection and splitting of touching nuclei

Similar to the refinement of contours, the algorithms for the splitting touching nuclei are only applied conditionally. In each case three methods for the detection and the



(a) (b) (c)

Figure 3.7: Splitting of nuclei: green: convex hull, (a) Best fit ellipse, in blue (b) Computation of the depth of a concavity, blue: line segments for the distance computation, plus the depth measured (c) Alignment of a concavity, the concavity is well aligned as the vectors v_i and v_j, starting from the midpoint of the convex hull segmented to the deepest point of the concavity, point towards each other.

splitting were analyzed.

If two elliptical nuclei which are close to each other are segmented as one object, the resulting contour is often similar to the shape of the digit 8. Thus the object has a non-elliptical shape and concave regions which are usually oppositely aligned. Three features, EllipticDeviation, ConcavityDepth, and ConcavityAlignment are used to quantify these visual observations. The two concavity features are based on the concavity analysis framework of Kumar et al. [105],which is summarized in Appendix A.2.

- **Elliptic deviation:** For this feature, the best fit ellipse E for the contour is computed using Fourier Descriptors (see Figure 3.7(a)). Denoting the initial segmentation of an object as S, the elliptic deviation is then defined as

$$\text{EllipticDeviation} = \frac{\text{Area}(E)}{\text{Area}(E \cup S).} \tag{3.2}$$

  The elliptic deviation is 1 for elliptical objects and smaller otherwise. This feature does not have any parameters.

- **Concavity depth:** In this work the depth of the concavity is measured as the maximum distance between the boundary of the object and the chord of the corresponding convex hull segment ([114], and see Figure 3.7(b)). Several strategies for computing a final feature value are considered. Either the concavity depth of the deepest concave region, the second deepest concave region, the mean depth of all concavities, or its sum is returned. The computation strategy is the parameter of this feature.

- **Concavity alignment:** Two concavities are considered as well aligned if the concavity vectors v_i and v_j point towards each (Concavity-Concavity Align-

ment, see Figure 3.7(c) for an illustration, and appendix A.2 for an exact definition). If more than two concavities are present, the alignment is computed for all pairs of concavities, and the lowest alignment value is returned as the feature value.

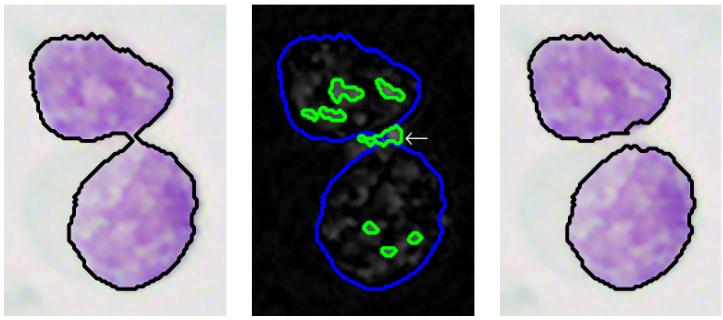
The only parameter of these features which needs optimization is the computation strategy of the concavity depth feature; the best strategy is found on the training set for touching nuclei to detect all touching nuclei but including as little correctly segmented objects as possible. Likewise, the thresholds for the detection were set such that all objects which need to be split are identified.

State-of-the-art methods for splitting touching nuclei can be categorized into three groups: Watershed approaches, Top-hat filtering and approaches based on the geometry of the object. One algorithm from each of these groups was chosen and adapted for the problem at hand.

As first approach, a watershed algorithm is employed. The watershed algorithm interprets a gray image as a topographic relief. This relief is filled with water, and the ridges where individual water basins meet are taken as segmentation boundaries. In this work, the watershed algorithm was applied to the distance transform of the touching nuclei. In this distance transform, for every pixel the distance to the closest background pixel is computed. As proposed in [115], the chessboard distance is used as distance metric, since it reduces over-segmentation, a common problem of the watershed algorithm.

The second approach studied is based on the Top-hat filter. For the Top-hat filter, a morphological opening is applied to the green channel and then subtracted from the green channel. The opening operation closes the gap between dark objects which are close to each other. Thus the subtraction from the original image highlights the path between these objects (see Figure 3.8(b)). Candidates for bright paths between nuclei are extracted by thresholding the Top-hat image, with the aim to remove these bright regions from the original mask and thus split the touching nuclei. These candidates are those regions with a gray value above 40 in the Top-hat image, a threshold determined empirically on the training set. First experiments revealed that these candidates also occur in the interior of the object. Therefore, only candidate regions which lead to a splitting of the touching nuclei are removed from the original segmentation mask. The result of this approach is depicted in Figure 3.8(c).

Third, the approach of Kumar et al. [105] is used from the group of geometric approaches. Briefly, a split line candidate is a line segment connecting two concave regions. If these two regions are properly aligned, a measure of split is computed from the depth of the two concave regions as well as their distance to each other (for a detailed algorithmic description the reader is referred to A.2). The split line with the highest measure of split is the best split line. Touching objects are then iteratively split along the best split lines, until no split line can be found. This approach was extended, because when splitting touching nuclei across a straight



(a) (b) (c)

Figure 3.8: Splitting of undersegmented touching nuclei: (a) before the splitting (b) Result of the Top-hat filter, blue: initial segmentation, green: candidate regions for splitting, white arrow: Only this candidate leads to a splitting when removed from the initial segmentation mask.

line the segmentation of the emerging objects usually does not coincide with the edges of the nuclei. To perform a split coinciding along these edges, the bright path between two nuclei is extracted by an iterative thresholding procedure on the green channel. Starting from the average background value, all pixels brighter than this value are extracted and subtracted from the initial segmentation mask. This is repeated until it leads to a splitting of the initial segmentation mask. However, this split is only kept if the midpoint of the split line is within the bright region and if the threshold is above the average green channel value of the touching nuclei.

3.2.2.5 Evaluation and optimization

The developed algorithms are optimized and evaluated on the manually annotated FOVs. For evaluating the detection performance of the algorithm, the sensitivity is used; an object of the gold standard is counted as detected if it shares at least one pixel with a machine segmentation. The relevance of the detected objects is assessed by the positive predictive value (PPV), the number of detected nuclei divided by the number of all detected objects. The precision of the segmentation is quantified by the Hausdorff distance, the maximum distance between the gold standard and the algorithm's segmentation. For a precise segmentation this distance should be as low as possible. For evaluating the ability of the splitting algorithms, the number of undersegmented touching nuclei is counted; these are the objects which share pixels with at least two gold standard objects.

The segmentation algorithms are optimized on the manually annotated FOVs from the training set. Choices to be made are the method for the initial segmentation, the size of the median filter for the preprocessing, the contour refinement algorithm, and the splitting algorithm. As a first criterion, the sensitivity must be as high as possible. In case of a tie, the Hausdorff distance is used for decision. For the splitting

Table 3.3: Optimization of the segmentation algorithms on the training and test set.

Training set				
	Sensitivity (%)	PPV (%)	HD (µm)	#touching
Baseline	99.72	44.88	1.066	27
Initial segmentation	99.91	48.03	1.138	18
+ Median filtering	99.91	47.64	0.960	23
+ Rejection	99.91	60.07	0.960	23
+ Refinement	99.91	60.06	0.811	23
+ Split touching nuclei	99.91	57.86	0.648	2
Test set				
Final algorithm	99.63	53.65	0.447	2

of touching, the number of remaining touching nuclei is the primary criterion for choosing a method, and again the Hausdorff distance is the secondary criterion.

3.2.3 Results

In the following, the optimization of individual parts of the segmentation pipeline as well as their influence on the segmentation performance on the training set are presented. Finally, the results of the best algorithm applied on the test set are given.

As a baseline for the segmentation performance, the segmentation algorithm from manual DNA-ICM is used, which detects nuclei with a sensitivity of 99.72 % and a PPV of 44.88 % (see Table 3.3). Already the initial segmentation of the newly developed segmentation pipeline finds more relevant objects and segments less irrelevant objects. If the image is median filtered before segmentation, also the contour precision is better. Hereby, no matter which initial segmentation is used, the most precise contours are reached for a median filter of size 1.26 µm. As a side effect of the median filtering, more individual nuclei are jointly segmented (without median filter 18 touching nuclei, with median filter 23 touching nuclei). From the three initial segmentation methods, two-sided thresholding, maximum likelihood approach, and the classification of color values with a kNN classifier, the classifier approach is best. The PPV of the segmentation is increased further by the rejection algorithms for nuclei with dirt and nucleus fragments. The initial segmentation finds 2236 objects in the 80 FOVs, from which 465 objects (20.80 %), are rejected without removing any relevant object. For detecting the need for refinement, the feature GradientOnContour yielded a higher specificity than the feature ConvexHullIntensity (84.55 % compared to 70.33 %, reached for $p = 0.4$). All objects with GradientOnContour below 14.86 are refined, which is the case for 471 objects in the training set. Considering the contour refinement algorithm itself, the GVF snakes yield a slightly more precise contour than the level set method (a Hausdorff distance of 0.811 µm

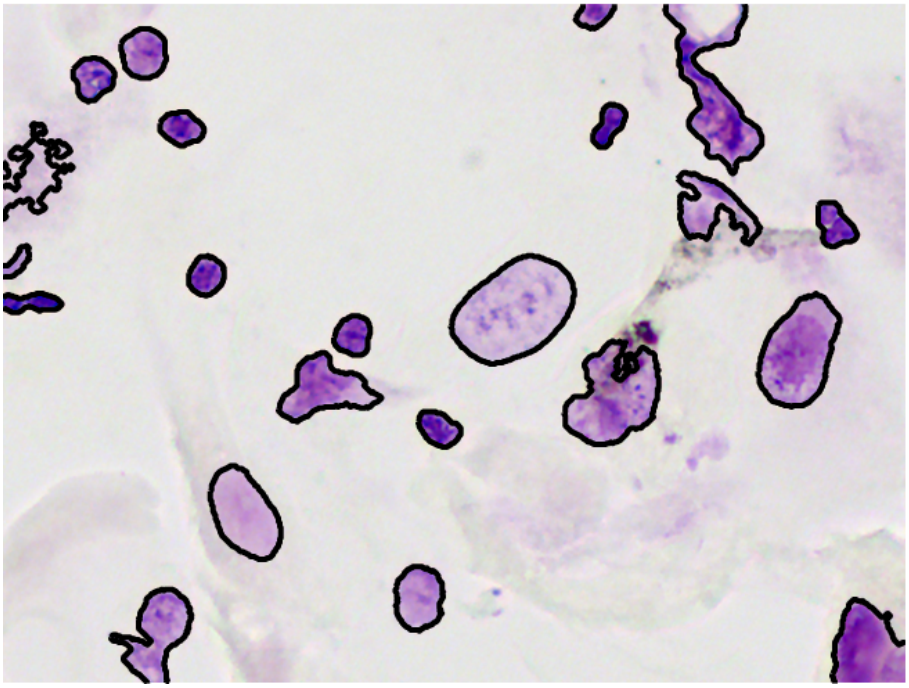


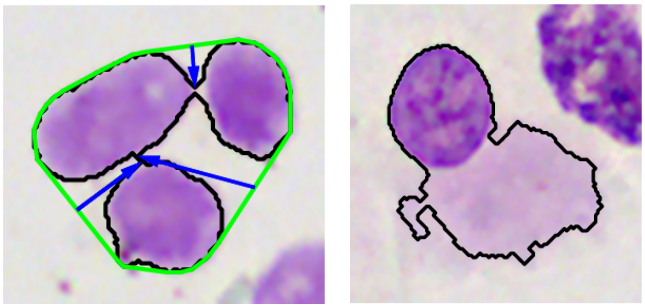
Figure 3.9: Segmentation results for the FOV from Figure 3.2.

compared to 0.831 µm on the training set for contour refinement, using $\alpha = 0.4$, $\beta = 1$, $\delta = 0.1$ and 200 iterations). Detecting the need for splitting an object should be done with the feature ConcavityDepth, and only the second deepest concavity should be considered. A splitting algorithm is called if this concavity is deeper than 1.2 µm, which is the case for 236 objects (touching nuclei as well as artifacts). From the splitting algorithms, the Top-hat filter performed best: it splits 21 of the 23 undersegmented touching nuclei in the training set, whereas the watershed splits only 16 and the morphology based approach only 8.

To summarize, the final configuration for the segmentation algorithm is:

- Preprocessing: median filtering with filter size 1.26 µm
- Initial segmentation: based on kNN classifier, trained on the color values
- Detection refinement: GradientOnContour
- Refinement: GVF snakes
- Detection splitting: depth of second deepest concavity
- Splitting: Top-hat filter

Compared to the baseline classification, on the training set the proposed segmentation pipeline yields a sensitivity of 99.91 % instead of 99.72 %, a PPV of 57.86 % instead of 44.88 %, a Hausdorff-Distance 0.648 µm instead of 1.066 µm, and only two instead of 27 touching nuclei remain. Figure 3.9 exemplarily shows the contours found by this segmentation algorithm.



(a) (b)

Figure 3.10: (a) Three touching nuclei, segmented as one object. The segmentation mask is given in black, the convex hull is green and the blue arrows are the vectors used for computing the alignment of concavities. The assumption that the concavities are aligned oppositely does not hold anymore, (b) Joint segmentation of an intact and a disintegrated nucleus.

Applied on the unseen 80 FOVs of the test set, nuclei are detected with a sensitivity of 99.63 %, a PPV of 53.65 %, and a Hausdorff Distance of 0.447 µm.

3.2.4 Discussion and conclusions

The results on the training set clearly motivate the application of the individual parts of the segmentation pipeline: Each step increases either the PPV or the contour precision (see Table 3.3). Different from many state-of-the-art algorithms, in this pipeline complex contour refinement and splitting algorithms are only applied conditionally. Applied on the test set, the detection performance and the PPV are very similar compared to the training set, and the Hausdorff distance to the manual gold standard segmentation is even better. In the following, the influences of individual parts of the segmentation are discussed and explanations why certain algorithms or features perform better are given.

The median filter as a preprocessing step decreases the Hausdorff distance, and a visual analysis revealed that the contours are much smoother and coincide better with the gold standard segmentation. It leads, however, to an increased number of undersegmented touching nuclei, as the median filter replaces the background pixels between two nuclei which are close to each other with dye color. But this drawback is clearly compensated by the benefit of a more precise segmentation, and the touching nuclei can later be split again by the splitting algorithm. For the initial segmentation, the lookup table trained from the kNN classifier is best. Compared to the two-sided thresholding approach, it allows a more flexible partition of the HSV colorspace into nucleus colors and non-nucleus colors. Compared to the maximum likelihood approach, it generalizes better for data which is not part of the color training set. The contour refinement method improves the contour precision from

0.960 µm down to 0.811 µm. Both GVF snakes and level set methods perform well, with a slight advantage for the GVF snakes. Interestingly, if the GVF snakes are applied to all objects, this not only increases the computational burden, also the Hausdorff distance is slightly larger (0.816 µm). For detecting objects which need to be split, the depth of the second deepest concavity is most specific. For the feature EllipticDeviation, this is because its feature value is not only diminished by deep concavities, but also if the contour is ragged everywhere, thus leading to more false-positives. For the feature ConcavityAlignment, a visual analysis of touching nuclei revealed that the concavities are not always aligned, especially if more than two nuclei are jointly segmented (see Figure 3.10(a)). Therefore this feature is also inferior to the ConcavityDepth feature. The best splitting algorithm is the Top-hat filter. The watershed algorithm, does not split as many touching nuclei, and besides leads to an over-segmentation of artifacts. The main reason why the morphological approach does not split as many objects is the same drawback as for the ConcavityAlignment feature: Especially if the initial segmentation contains several objects, the concavities are not aligned, and only one candidate point per concavity is available (see Figure 3.10(a)). Thus the algorithm does not find the desired split lines. The splitting algorithm slightly decreases the PPV, as a split artifact, which was formerly counted as one false-positive object, is counted as two or even more false-positive objects after the splitting. However, again the benefit of a more precise contour compensates for this drawback. On the average, the segmentation algorithm from manual DNA-ICM segments 16.41 irrelevant objects per FOV on the training set, whereas the proposed segmentation pipeline only has 9.75 false-positives per FOV. The remaining objects are mainly artifacts or defocused objects, which need to be identified by the subsequent classifiers. However, considering that up to 4000 FOVs have to be processed when scanning a slide, more than 20.000 irrelevant objects can already be rejected at this stage. Altogether, the conditional contour refinement and splitting decrease the Hausdorff distance, which is very sensitive for missegmentations, from 1.066 µm of the manual DNA-ICM segmentation down to 0.648 µm on the training set. This performance is confirmed on the test set.

To conclude, a segmentation pipeline for the precise, unsupervised segmentation of nuclei has been presented. It decreases the number of irrelevant objects during an initial segmentation step by learning color values which are specific for nuclei. Further irrelevant objects of this segmentation are rejected based on feature values. Furthermore, it improves the contour precision by conditionally applying a contour refinement and a splitting algorithm.

As a future research task, Markov-Random-fields will be tested as initial segmentation. For Markov-Random-fields not only the pixel intensities but also the relationship to the neighboring pixels is examined; this can improve the initial segmentation. An alternative for Markov-Random-fields is to adapt the initial segmentation to the stain intensity at the first FOVs which are analyzed. Especially dim nuclei need refinement, therefore the adaption can reduce the number of objects needing refine-

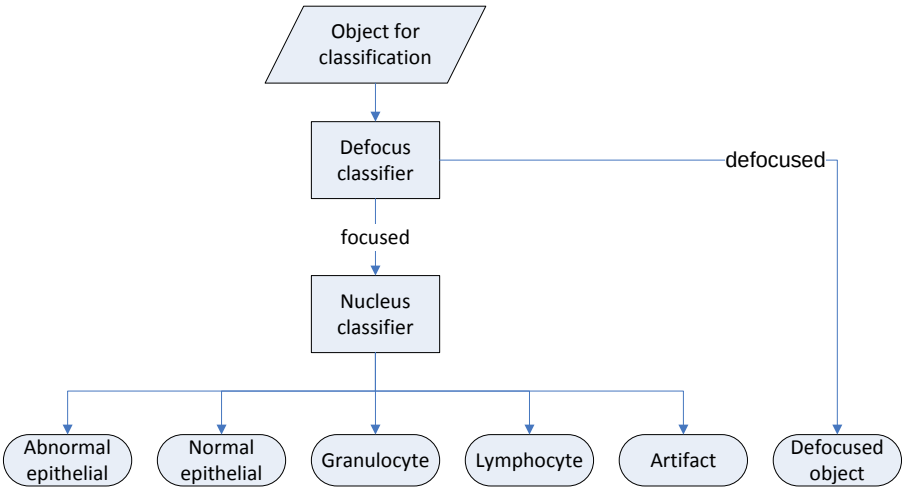


Figure 3.11: Flow chart of the classifier cascade, which first decides whether an object is focused and subsequently classifies the focused objects into one of the nucleus classes or as artifact. The nucleus classes vary depending on the type of specimen, in this example the classes in oral brush biopsies are shown.

ment. Currently, only a single feature and a threshold are used to identify the need for refinement or splitting. The combination of several features and more complex classifiers can be used for this purpose. However, in that case, the computational burden should not exceed the effort for unnecessary applications of the algorithm itself. The main reason for an imprecise segmentation of the proposed segmentation pipeline is the joint segmentation of nuclei and larger regions of disintegrated nuclei (see Figure 3.10(b)). A discriminative criterion for these artifacts and nuclei is the textural information and the stain intensity—similar to the refinement and splitting algorithms for both the detection of this missegmentation as well as its correction need to be developed.

3.3 Classification of nuclei

The automated image acquisition and segmentation yields up to 100,000 objects (nuclei and artifacts) per slide. Next, classifiers are employed to assign to each object either one of the nucleus classes (four to five different classes), or the class “artifact” or the class “defocused”. They are a core component for increasing the productivity of diagnostic and prognostic DNA Image Cytometry, as they automatically identify the relevant nuclei within a large amount of objects.

This section presents classifiers for distinguishing nuclei originating from effusions,

prostate cancer biopsies and cervix/oral[1], with a special focus on classifying morphologically abnormal nuclei. Since the nucleus classes and the preparation technique are different for the three modalities, individual solutions are needed for each one. The classification itself is realized as a two-stage classifier cascade (see Figure 3.11). First, it is decided whether an object is focused. Only the focused objects are then handed over to the second stage, the nucleus classifier. The reason for this is that focused and defocused objects mainly differ by texture, whereupon nuclei from different classes mainly differ by morphology. The organization as a classifier cascade allows the individual selection of appropriate features for each task. For both stages of the cascade, the development includes the systematic optimization of classifier models, the choice of state-of-the-art features or the development of novel features, and the use of feature selection techniques to find the most relevant combination.

The organization of this section is as follows: First, an overview of general concepts used in this work is given. The second part is about classifiers for focused and defocused objects. Third, the nucleus classifiers are presented. Since the concepts for focused/defocused objects are very similar for each type of specimen, they are presented together. The nucleus classifiers, however, are presented separately.

3.3.1 General concepts

The classifiers were trained to classify nuclei in accordance with a cytopathological expert using a supervised learning approach. Starting point for supervised learning is a gold standard database, a set of nuclei for which an expert has labeled their corresponding nucleus class. A set of features which quantify characteristic properties of nuclei from different is computed for each object from the gold standard. Based on the gold standard and the features, a classifier algorithm "learns" how the expert classifies objects and applies this knowledge to classify new nuclei. In the following, the general concepts applied throughout this work for collecting gold standard databases, selecting and developing features, and evaluating the performance are presented. This chapter presents the general concepts applied for the development of all classifiers—the types of nuclei and nucleus statistics for the individual classifiers are presented in the corresponding sections.

Gold standard: Prior to collecting gold standard data, the system was set up according to the standardized image acquisition of section 3.1. To ensure that the objects in the gold standard cover all classes and their real frequency of occurrence, all objects from a diagnostically representative part of a slide were acquired by segmenting all objects in the FOVs. The inter-patient variability is accounted for by

[1] As brush biopsies of the oral cavity and the uterine cervix are both taken from mucosa which represents squamous epithelium, and the preparation of specimens is identical, the cell types derived from cervical and oral smears are partly the same and the cells have a very similar appearance. Therefore, classifiers trained on cervical mucosa data can be used for the classification of oral nuclei.

collecting objects from four and nine different patients for the defocus and nucleus classifiers respectively. A cytopathologist with more than 35 years of experience labeled the objects' classes in the gold standards for the nucleus classifiers. The classes of objects in the gold standard for the defocus classifier were assigned by the author. Additionally, each gold standard was reviewed by checking the initial classifications of all objects and correcting for misclassifications.

Features The features used for classification should quantify characteristic properties of objects by values which are similar for objects in the same class, but different from objects in other classes [73]. Therefore, choosing or developing features comprises analyzing and understanding the visual properties of different types of nuclei, and the creative process of identifying ways to quantify these differences. The features used in this work are either taken from the rich literature of nucleus features [21, 80, 116], or novel features were developed. In case the novel features have parameters which need to be optimized, an exhaustive parameter search is performed on a small training set (objects from one slide of the gold standard). The parameter which separates the classes of this set best, quantified by a separability criterion, is chosen for this specific feature. As a separability criterion, the rate of objects which have a neighbor from the same class is used (1NN-classification); for features with a high discriminative power this value is high.

Feature normalization: The values from different features might span different ranges. For classifier algorithms such as kNN or SVM, this can be problematic as, depending on the range, some features dominate over others and reduce their influence on the classification result [117]. Therefore it is beneficial to normalize the features to similar ranges. The normalization strategies from [118] were studied for the classification of nuclei from prostate biopsies [117]. These strategies are: linear scaling to $[0, 1]$, rank normalization, linear scaling to unit variance, and transformation into a uniform random variable via the cumulative distribution function. As the experiments in [117] showed that the normalization by variance is best suited for the problem at hand, this normalization method is used throughout this work. To account for outliers, all values below the 2.5 % and above the 97.5 % quantile are not considered for computing the mapping.

Feature selection The task of feature selection is to automatically find the combination of features which discriminates the individual classes best. Therewith, it has to be considered that the combination of several weak features might be better than a single strong feature [119, 1165]. As the number of possible combinations raises exponentially with the number of features, testing all features ("brute force") quickly becomes infeasible. In that case, sequential forward feature selection (SFFS) was used for feature selection. Starting from an empty feature set, this method iteratively adds features by selecting the combination with the previously selected features which has the highest classification rate [120].

Table 3.4: Classifiers used in this work, including the parameters tested.

Classifier	Parameter	Range
kNN	Number k of neighbors	1:1:30
	Distance metrics	Cityblock, Euclidean, Cosine Correlation, Mahalanobis
Neural Network	Number of Layers	10:10:200
Decision Tree	Min. branch node observations	10:10:300
	Prune criterion	Deviance, Towing, Gini
	Split criterion	error rate, impurity
SVM	Kernel function	Radial basis function
	C	$2^{-5}, \ldots 2^{15}$
	γ	$2^{-5}, \ldots 2^{15}$
Random Forest	Number of classifiers	5:5:300
	Number of features at nodes	4,8,12,16, all
Adaboost	Number of classifiers	100:50:2500

Classifier algorithms: For each classification task, six classifier algorithms and their parameters were systematically evaluated: Neural networks, Decision trees, k Nearest Neighbors classifiers (kNN), and Support Vector Machines (SVM) [121], AdaBoost, and Random Forest. The latter two are so called ensemble classifiers, for which several weak classifiers are combined to a stronger one. The classifier algorithms and their parameters are listed in Table 3.4, detailed descriptions of these classifier algorithms can be found in [117].

Evaluation: The classifiers were evaluated using a leaving-one-out strategy on a slide basis. That is, a classifier is trained with data from all but one slide of the gold standard, and the remaining slide is classified. The outcome is compared against the annotations of the expert. This process is repeated until each slide is classified once. Opposed to an approach where the gold standard data is split into training set and test set, the leaving-one-out- strategy gives a better estimate for the classifiers performance (see [122, p. 312]). And opposed to randomly splitting the gold standard (cross-validation), by leaving out the data from a whole slide it is ensured that no data from the same patient is in the training and test set at the same time, which would lead to a bias. An overall correct classification rate is then computed as an estimate of the classifiers performance. Since the requirement for this work is to develop classifiers which are able to identify morphologically abnormal nuclei, additional statistics reflecting the performance to classify abnormal nuclei are given.

3.3.2 Classification of defocused nuclei

The first step of the classifier cascade is to sort out defocused objects. In case the autofocus system fails to correctly focus on nuclei, they are acquired out of focus.

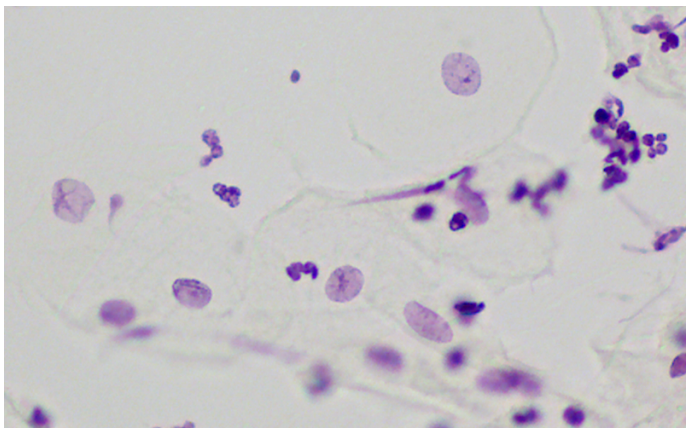
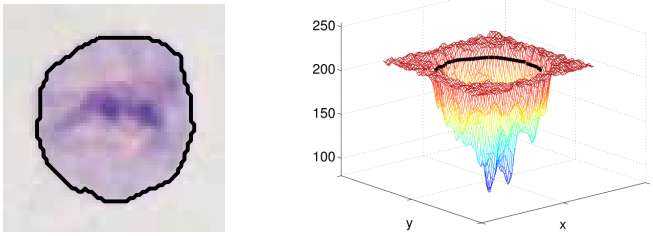


Figure 3.12: Optimally focused scene of an oral smear, as determined by the autofocus procedure. Still, some objects are out of focus.

Since the depth of field of the microscope objectives at high magnification is narrow, defocused objects will even be present in the FOV with the highest focus score during the autofocus procedure (see Fig. 3.12). Sorting out defocused objects is necessary for two reasons: First, the DNA content of nuclei can only be measured precisely if the nuclei are acquired in focus. The reason for this is that the light intensities in the nucleus image, which are used for DNA computation, are altered when the nucleus is out of focus. Second, the correct nucleus class can only be determined if a nucleus is in focus; for defocused objects important morphological and textural information is missing.

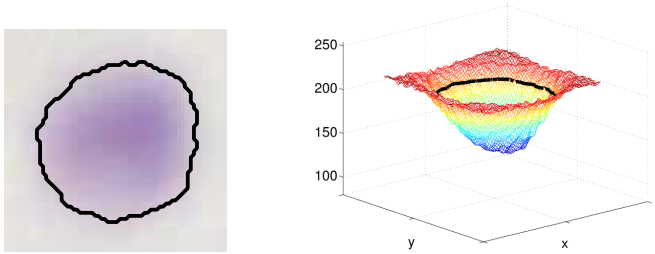
The nucleus classifiers cannot be used to sort out defocused objects, because they mainly concentrate on the morphology of a nucleus which hardly changes for slightly defocused nuclei (see Fig. 3.13). Therefore, this classifier misclassifies defocused objects into non-artifact classes. Numerous autofocus algorithms for brightfield microscopy have been presented in literature. These algorithms select the optimally focused FOV from a set of images with different focusing distances to the slide (z-stack); comprehensive overviews can be found in [123, 124]. However, these approaches have in common that select the FOV with the highest focus score out of a set of images. But in this case, only *one* image of the nucleus is available—and from this image it has to be decided whether the nucleus is defocused or not.

Therefore, individual classifiers for automatically sorting out defocused nuclei were developed. Based on a visual analysis of the changes between focused and defocused images of the same nucleus, a set of novel features was developed. Gold standards were collected by acquiring focused and defocused versions of several nuclei. These gold standards were used for optimizing the feature parameters and selecting the most relevant ones. Further training sets were used for adapting the sensitivity for the detection of defocused nuclei as well as choosing the best classifier algorithm. Finally, the optimized classifiers were evaluated on a test set. Parts of this work have been published in [125].



(a) Focused nucleus

(b) 3D intensity plot of 3.13(a)



(c) Defocused nucleus

(d) 3D intensity plot of 3.13(c)

Figure 3.13: Focused and defocused images of the same nucleus (3.13(a) and 3.13(c)) and 3D intensity plots of their gray images. The black line is the contour found by the segmentation algorithm.

3.3.2.1 Equipment and material

A Motic BA410 manual microscope was used for collecting the gold standard sets with the following procedure: First, an image of a nucleus in focus was acquired. Subsequently, the microscope objective was moved in z-direction in 10 µm steps, acquiring five further images of the nucleus. By doing so, a focused image and defocused at several levels are available.

Training and test sets were automatically collected by using a Motic BA600 motorized microscope for scanning areas of about $0.5 \times 0.5\,\mathrm{cm}^2$. To enrich the set with defocused nuclei, one third of the focus points were set out of focus. After acquisition, the collected nuclei were classified manually into the classes “defocused” and “focused”.

Gold standard, training set, and test set were collected for nuclei from prostate biopsies, oral smears, and serous effusions. DNA Image Cytometry specimens might obey an inter-slide variability, for example with respect to the stain intensity. To take these differences into account during the development and evaluation of the system, four slides for each set and type of specimen were used. Table 3.5 shows the number of nuclei.

Table 3.5: Number of nuclei in the gold standard, training and test sets.

	Gold standard	Training set	Test set
Effusions	3762	3547	3200
Oral	2249	2871	2878
Prostate	2110	3614	3564

3.3.2.2 Methods

An object which is out of focus is a low pass filtered version of the original object [126]. Figure 3.13 displays a focused nucleus and its defocused counterpart, as well as a 3D intensity plots of the gray image of the these nuclei. By visually analyzing this kind of images, the following characterizations for distinguishing focused and defocused objects were made:

(1) Defocused objects have higher intensity values at the boundary of the object.

(2) The variation of intensity pixels in the interior of the object is lower for defocused objects.

(3) Defocused objects have less variation in the derivatives of pixel intensities.

(4) The transition from background to nucleus perpendicular to the nucleus contour is less steep for defocused objects.

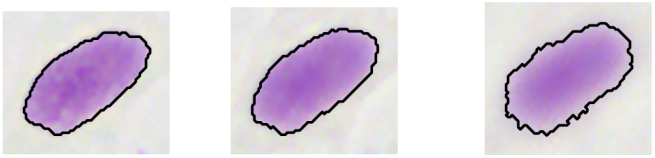
(5) Defocused nuclei are brighter than their focused counterparts.

Based on this analysis, features for quantifying these observations have been developed.

Feature development All features are based on a gray image of the nucleus. Briefly, observation (1) is quantified by computing the relationship of boundary intensity values to intensity values in the interior of the object (feature Rbiii). As for observation (2), the original segmentation mask is shrunk down to a percentage p by morphological erosion, and the coefficient of variation within this mask is used to quantify the variation (feature VarInterior). The mean absolute filter response of a Laplacian filter is used for measuring the variation in the derivatives within the pixel intensities (feature Laplacian), as a consequence of observation (4). For quantifying the transition from object to background, the absolute difference of pixel intensity along the objects contour normal is computed. The final value of the feature ADNormal is the p % quantile of all of these difference values. Finally, the feature MeanIntensity is computed for reflecting how bright a nucleus is. A detailed description of these features and their parameters can be found in the appendix A.3. Table 3.6 shows how the feature values change if an object is moved out of focus.

Feature optimization and feature selection All features, except for the average intensity feature, have parameters which are optimized to achieve the highest discriminance between focused and defocused objects. As an optimization criterion,

Table 3.6: Change of feature values when acquiring a nucleus out of focus.



	focused	defocused 10 µm	defocused 20 µm
ADNormal	38	26	14
Laplacian	0.018	0.013	0.012
MeanIntensity	157.509	159.245	169.893
Rbiii	1.332	1.320	1.243
VarInterior	0.064	0.050	0.052

the classification rate of a 1NN classifier, evaluated with a leaving-one-out strategy on slide basis, was used (see Section 3.3.1). This criterion was computed for all parameters on the gold standard data set, and the parameter combination with the highest 1NN classification rate was chosen as optimum. Subsequently, the best combination of features was found by using a 10NN classifier as baseline classifier and testing all possible combinations of the five features. Again, the gold standard set and the leaving one out strategy on slide basis were used for optimization. All these optimizations were performed separately for prostate, oral and effusion specimens.

Classifier optimization and final evaluation Using the optimal subset of optimized features, the classifiers' sensitivities were tuned. This was done by including or excluding the slightly defocused nuclei in class "defocused" of the gold standard. The gold standard was created by first acquiring a focused image of the nucleus, and then bringing it more and more out of focus by moving the microscope objective 10, 20, 30, 40 and 50 µm in z-direction. Usually the values of the developed features show a monotonic trend when going more and more out of focus. Therefore, if the classifier uses all of these nuclei in the class "defocused", objects whose feature values are closer to the 10 µm defocused class than to the focused class will be considered as defocused. If only the 20 to 50 µm defocused nuclei are used, this will be the case if the object is closer to the 20 µm defocused nuclei as the focused ones, thus shifting the decision boundary and being less sensitive. Moreover, the best classifier model was selected by exhaustive search of classifier models and their parameters. Both optimizations were performed on the training set. Since classifying a focused abnormal nucleus as defocused object has more severe consequences than vice versa, a weighted error rate was computed as optimization criterion. Classifying a focused object as defocused was punished with a five times higher weight than classifying a defocused object as focused. Finally, the gold standard and classifier algorithm with the lowest weighted error rate were chosen for training the final classifier and used to classify the test set.

Table 3.7: Results from the parameter optimization of the defocus classifier for effusions. The highest 1NN correct classification rate and the corresponding parameters are given.

Feature	Parameter	Best parameter	1NN CCR (%)
ADNormal	Normal length nl	0.18 µm	88.81
	Quantile p	100 %	
Laplacian	Resize percentage p	110 %	90.01
MeanIntensity			76.87
Rbiii	Resize percentage p	90 %	77.01
	Extend outside	yes	
VarInterior	Resize percentage p	30 %	81.13

3.3.2.3 Results

Since the optimal feature values are very similar for the different modalities, table 3.7 shows the optimal parameters only for effusions. The feature AD Normal discriminated best if the maximum difference from all absolute pixel differences is used. For the Laplacian feature, it is beneficial to extend the original segmentation mask slightly (+10 % added to the original size). The relationship between boundary and interior intensity should be computed along a thin ring around the nucleus contour. The variation of the intensity values should be assessed in the center of the nucleus (30 % of the original size of the segmentation mask).

The features are used with the optimized parameters in the second step, the feature selection process described in section 3.3.2.2 to find the optimal combination of features. The parameter optimization of the features already indicates that the Laplacian and ADNormal feature are most discriminative features. For effusions and prostate, these two features and the VarInterior feature give the highest classification rates of 99.13 % and 93.13 % respectively. For oral, the mean intensity feature was chosen as well. However, as the gain was only marginal (97.64 % with average intensity, 97.47 % without), this feature is not used to achieve consistency of features between the three modalities. In conclusion, the ADNormal, Laplacian and Rbiii feature are used for all three types of specimen.

The final step in the optimization of the classifiers is the adaption of the sensitivity and the choice of the classifier model. For effusions, the objects acquired 20 to 50 µm out of focus are used as class " defocused" in the gold standard. The optimal classifier algorithm is a kNN classifier, using $k = 26$ and the cityblock distance. For oral, again the objects 20 to 50 µm out of focused are used as defocused objects. This time, the SVM classifier is best ($C = 4$, $\gamma = 512$). For prostate, the optimization selects a more sensitive classification scheme, as objects 10 to 50 µm out of focus are used. Similar to oral, the SVM classifier is best, but different parameters are used ($C = 2^{-5}$, $\gamma = 16$). Table 3.8 shows the application of these classifiers on the data from the test set. The nuclei are classified with correct classification rates of

Table 3.8: Classification performance on the test sets.

	Effusions	Ground truth	
		Focused	Defocused
Cls.	Focused	1992	60
	Defocused	0	1148

	Oral	Ground truth	
		Focused	Defocused
Cls.	Focused	2143	86
	Defocused	7	642

	Prostate	Ground truth	
		Focused	Defocused
Cls.	Focused	2373	5
	Defocused	16	1170

98.13 %, 96.77 % and 99.30 % for effusions, oral and prostate.

3.3.2.4 Discussion and conclusion

Taking all types of specimen together and defining the defocused objects as "positive" class, the positives are detected with an overall sensitivity of 95.15 % and an overall specificity of 99.64 %. Thus most of the defocused objects are removed. Only 23 out of 6531 focused objects are misleadingly rejected, and a detailed analysis of these misclassifications revealed that none of them would change the diagnosis or prognosis[2]. Being representative for all types of specimen, Figure 3.14 shows the classification result of the oral defocus classifier.

One key factor for reaching this performance of detection is the development and optimization of novel features. From the five features examined, the features MeanIntensity and VarInterior were rejected by the feature selection. Whereas the defocused version of a nucleus is brighter than its focused version, it might be that the focused version of a brighter nucleus has the same mean intensity as the defocused version of a darker nucleus. Therefore, the feature MeanIntensity is not suited for deciding if an object is focused or not. The feature VarIntensity, assesses the variation of intensities in the center of an object. However, objects like aggregates of nuclei or granulocytes might not have any stain in their center, so in this case the feature does not provide discriminative information. The three features chosen by the feature selection are ADNormal, Laplacian and Rbiii. Although the feature Rbiii is, like the MeanIntensity feature, based on intensities, it can be used for deciding if an object is defocused or not because it relates the intensities close to the nucleus boundary to the interior intensities. All three features are relevant: the ADNormal feature is the most discriminative one, however when using only this feature the correct classification rate is 92.56 % compared to 99.13 % (effusion), 96.00 % compared to 97.47 % (oral) and 87.73 % compared to 93.13 % (prostate). To conclude,

[2]For oral, the 7 rejected focused nuclei were bright normal epithelial cells which are used for calibration but not diagnosis. For prostate, one focused fibroblast and 15 focused near-diploid abnormal nuclei were rejected. Only the rejection of non-diploid abnormal nuclei can be critical.

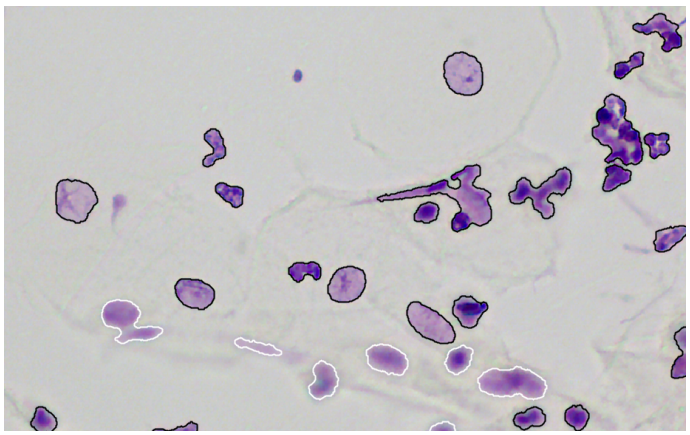


Figure 3.14: Field of view from Figure 3.12, where the nuclei classified as focused have a black contour and the contour of defocused nuclei is white.

the sharpness of the transition of intensity values from background to object, the variation in derivatives and the relationship between boundary and interior intensity contain sufficient information to decide if an object is focused or not.

The classification rates on oral nuclei was slightly lower than for the other types of specimen. The detection of defocused objects in oral specimen is more difficult because around oral cells more cytoplasm is visible in the microscopic images. Therefore, the difference of intensity values between nucleus and surrounding background, which influences the feature value for all three features used for classification, is less sharp. This could be overcome by normalizing the features with respect to the light intensity in a small reference ring around the segmentation mask of the nucleus.

As a conclusion, the developed classifiers can be used for automatically removing defocused objects from DNA Image Cytometry measurements prior to the classification into nucleus classes. Alternative applications of these classifiers are speeding up the autofocus routine and quality control of the automated digitization. The best focus position for neighboring FOVs usually differs only slightly. Therefore, the same z-position of the objective could be kept instead of acquiring a whole z-stack for each FOV. Such a stack would only be required if too many objects are classified as defocused. Additionally, the number of defocused nuclei of a whole scan allows a quality check for automatically scanned slides. If this number is above a predefined threshold, the slide needs to be rescanned with modified focus settings.

3.3.3 Classification of nuclei from effusions

This section presents a first classifier for distinguishing the different nucleus classes, namely for nuclei originating from effusion specimen. Body cavity effusions might be caused by a metastasizing cancer. Therefore, for every body cavity effusion is is necessary to investigate if it contains cancer cells (see chapter 2.1.5). The presented classifier is a step towards the automated screening of fine-needle aspirates of body

cavity effusion for cancer cells using DNA Image Cytometry. Parts of this work have been published in [127].

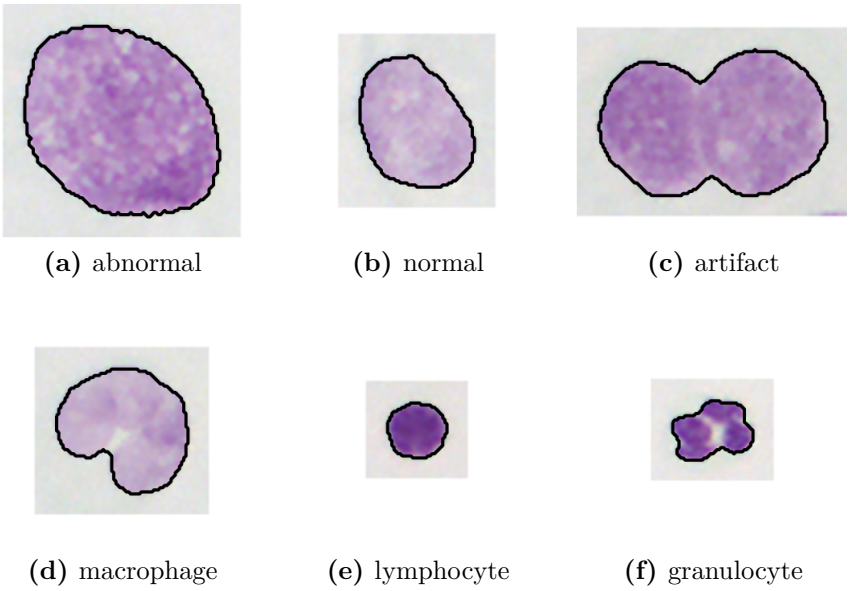
3.3.3.1 Material

A gold standard was collected comprising 54,374 nuclei from nine different patients. The effusions were gained by a fine-needle aspiration biopsy of body cavity effusions (see chapter 2.1.5). The cells contained in these effusions were deposited on glass slides by smearing effusion sediments and air-dried. The slides were first stained using the May-Grünwald-Giemsa scheme and then restained specific for DNA according to Feulgen. For six of the patients, DNA-aneuploidy was found with manual DNA Image Cytometry. Figure 3.15 shows the six classes of nuclei which are present on slides from effusions, which briefly are:

- **Normal mesothelial nuclei:** Cells originating from the cellular lining of the pleural or peritoneal cavity. They are elliptically shaped with a finely granular texture and without morphological or textural abnormalities. Normal mesothelial nuclei are used as reference nuclei for calibrating the algorithms for DNA measurement.
- **Abnormal mesothelial nuclei:** Mesothelial nuclei with morphological and textural abnormalities, suspicious for malignancy. Only these nuclei are used for calculating a diagnostically meaningful DNA distribution.
- **Macrophages, lymphocytes, and granulocytes:** Cells from the immune system. Macrophage nuclei are shaped like a kidney, lymphocyte nuclei are small and round, and granulocyte nuclei usually consist of three to four segments.
- **Artifacts:** Objects for which the DNA content cannot be used for diagnosis and which do not belong to any nucleus class, for example overlapping nuclei, disintegrating (lytic) nuclei, and missegmented objects.

3.3.3.2 Methods

In total 18 features from the literature about cell features for classification [21,80,116] were selected under the advice of a cytopathologist: these features mainly describe the morphology of the objects (11 features), their light intensity (4 features) or textural information (3 features). A list of these features can be found in appendix A.4.1, in the following these features are denoted as *state-of-the-art features*. As classifier algorithms, a kNN classifier (k=15, euclidean distance) and a Random Forest classifier (with 100 trees and randomly select 4 features at each node) were tested initially. The performance of these classifiers is evaluated by the leaving-one-out strategy on a slide basis (see 3.3.1). The evaluation is based on the overall correct classification rate, the percentage of detected abnormal nuclei and the percentage of abnormal nuclei among all objects classified as abnormal (relevance classified as abnormal).



(a) abnormal **(b)** normal **(c)** artifact

(d) macrophage **(e)** lymphocyte **(f)** granulocyte

Figure 3.15: Classes of objects from the effusion gold standard.

3.3.3.3 Results

The kNN classifier achieved a correct classification rate of 86.15 % and the Random Forest classifier a rate of 88.11 %. The Random Forest classifier was able to identify 95.24 % of all abnormal nuclei, and 94.04 % of all objects classified as abnormal are correct (see Table 3.9). The correct classification rates for individual slides during the leaving-one-out training varied between 84.41 % to 94.61 %, with a coefficient of variation of 3.80 %. A detailed list for the performance of each slide as well as the the performance for identifying abnormal nuclei, divided into DNA-exceeding events and abnormal stemlines, can be found in Appendix A.5.0.1.

3.3.3.4 Discussion and conclusions

More than 95 % of all abnormal nuclei are identified correctly by Random Forest classifier, employing the state-of-the-art features. Also, the number of objects which were misclassified as abnormal is low, therefore it can be concluded that this classifier is able to discriminate abnormal nuclei from all other types of nuclei and artifacts. A further important implication for the application in practice is that over 90 % of the normal mesothelial nuclei were classified correctly. They are used as reference cells for converting the integrated optical densities into DNA values, thus enough nuclei are available for this purpose.

Considering the classification rates of the individual nucleus classes, the class with the worst classification performance are macrophages. However, since these cells are neither used as reference nor as analysis cells and have normal DNA content, no

Table 3.9: Classification table of the Random Forest classifier for effusions.

		Ground truth					
		Artifact	Abnormal	Macrophage	Normal	Lymphocyte	Granulocyte
Classification	Artifact	14437	390	527	626	94	190
	Abnormal	470	10635	11	193	0	0
	Macrophage	226	26	955	338	4	15
	Normal	914	98	758	14708	641	26
	Lymphocyte	118	16	13	396	5840	73
	Granulocyte	166	2	48	10	72	1338
	Total	16331	11167	2312	16271	6651	1642
	Error (%)	11.60	4.76	58.69	9.61	12.19	18.51

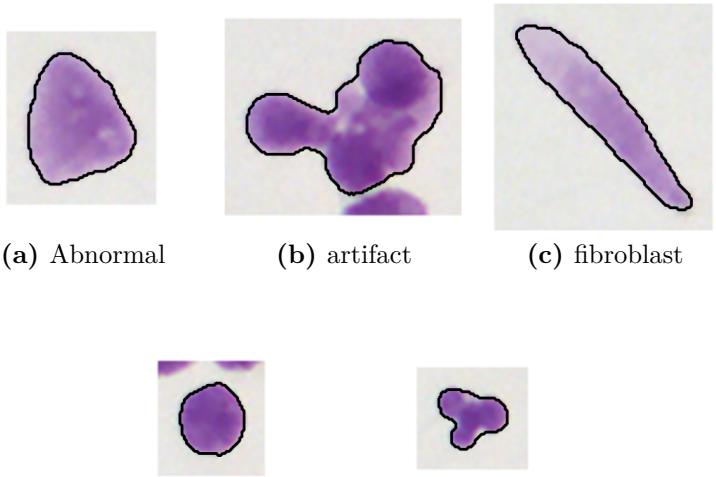
Correct classification rate:	88.11 %
Abnormals detected:	95.24 %
Relevance classified as abnormal:	94.04 %

influence on the final diagnosis is expected. Macrophages are cells of the immune system, therefore an increased amount is found for example in case of an inflammation. The variation of the classification rates for the individual slides is tied to the presence of macrophages: the five slides with the lowest classification rates are also the five slides where the most macrophages are present.

Selecting and optimizing the classifier models clearly has the potential to increase the classifier performance—the choice of a Random Forest classifier instead of kNN improved the classification rate by almost 2 percentage points. First experiments with the Random Forest classifier under clinical conditions were promising: They revealed that it is possible to confirm the diagnosis of manual DNA Image Cytometry by means of automatically collected and classified nuclei [127]. Thus a larger set of slides was evaluated using the Random Forest classifier. For consistency during the evaluation of this study, which is presented in section 4.3.1, the same classifier had to be used for all slides. This is why further research on classifier models was not pursued at that point. It should, however, be performed as future work.

3.3.4 Classification of nuclei from prostate cancer biopsies

A further application of DNA Image Cytometry is the grading of malignancy of an already diagnosed prostate cancer on enzymatic cell-separation specimens from Formalin-fixed and Paraffin embedded biopsies. In this section, a classifier for distinguishing nuclei found in these specimens is developed. The development includes feature selection, optimization and selection of the classifier, and the development of novel features. As the results of subsequent development steps build on the previous results, the results of an individual step are first presented and discussed before proceeding to the next step. Parts of this work have been published in [128, 129].



(a) Abnormal **(b)** artifact **(c)** fibroblast

(d) lymphocyte **(e)** granulocyte

Figure 3.16: Classes of objects from the prostate biopsy gold standard.

Table 3.10: Results of the feature selection on the gold standard from prostate cancer biopsies.

Feature set	Number of features	CCR (%)
Baseline	18	89.83
Forward feature selection	24	89.90

3.3.4.1 Material

Nuclei from nine patients were extracted out of prostate cancer foci from core needle biopsies. These were disintegrated by enzymatic cell separation, centrifuged, deposited on a glass slide and stained according to Feulgen. From these slides, 47.982 objects were collected and annotated. For grading the malignancy of prostate cancer, five classes of objects need to be distinguished (see Figure 3.16(e)). Besides artifacts, granulocytes and lymphocytes, which have already been described for the effusion classifier, "cigar-shaped" nuclei of fibroblasts which originate from connective tissue are present on these slides. As the cell material exclusively originates from cancer foci in core needle biopsies, there are no normal epithelial but only abnormal epithelial nuclei on these slides.

3.3.4.2 Feature selection

Like for the effusion classifier, the analysis of the prostate classifier was started using the state-of-the-art features and a kNN classifier with k=15 and euclidean distance, and the performance of the classifier was evaluated using the leaving-one-out evaluation on a slide basis. As a first approach for improving the classifiers performance,

Table 3.11: Performance of classifier algorithms on prostate data. For each classifier model, also the best parameter configuration is given.

Classifier	Parameters	Best CCR (%)
kNN	$k = 19$, Cityblock distance	90.24
Neural Network	80 layers	90.39
Decision Tree	131 min. branch node observations, splitting by error rate, pruning by deviance	88.15
SVM	C=32, $\gamma = 1$	91.34
Random Forest	170 classifiers, randomly select 4 features	91.12
Adaboost	1750 classifiers	84.94

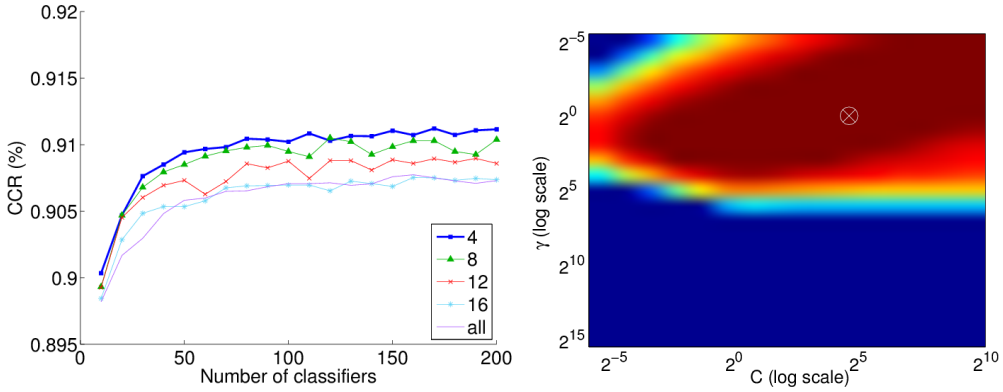
further features from literature were implemented to extend the original set to 75 features (see [130]). A sequential forward selection was performed on this extended set using the same kNN classifier.

The baseline for the classifier optimization is set by a kNN classifier with a correct classification rate of 89.83 % (k=15, euclidean distance, state-of-the-art features). Feature selection increases the performance by 0.07 % (see table 3.10), but also needs six more features. Employing other feature selection strategies such as backward selection or oscillating search strategies did not yield an noticeable improvement either (see [130]). Since less features are needed for almost the same classification performance, the further analysis was continued using the state-of-the-art features.

3.3.4.3 Optimization of classifiers

Next, different classifier algorithms were tested employing the state-of-the-art features. The tests included kNN, Neural Network, Decision Tree, SVM, Random Forest and Adaboost classifiers. Their parameters, as listed in table 3.4, were optimized by an exhaustive parameter search. Concerning the performance evaluation during the optimization, the SVM did not use the whole gold standard, but an internal five-fold cross validation procedure of the libsvm-package on five of the nine slides (see [117]). The reason is that training a single SVM is already very time-consuming, and in total 441 different parameter configurations needed to be tested; the evaluation on the reduced set cuts down the time needed for the optimization to one half. The best parameter configuration of this analysis is then evaluated on the full set of slides. For the other classifiers all parameter combinations were tested on the whole set, the leaving-one-out tests being applied on a slide basis.

The rate of correctly classified nuclei is raised above 91 % by a Random Forest classifier or an SVM (91.12 % and 91.34 % respectively). The results from the parameter optimization of these two classifiers is shown in Figure 3.17, and the result from the classifier optimization in Table 3.11.



(a) Random Forest **(b)** Support Vector Machine

Figure 3.17: Parameter optimization for (a) the Random Forest classifier (b) the SVM classifier. As the SVM classifier was evaluated using another optimization strategy to reduce the computational burden, only qualitative results are given for this classifier. On the full set, the best parameter combination ($C = 32$, $\gamma = 1$, white cross) achieves a classification rate of 91.34 % correctly classified nuclei.

3.3.4.4 Requirement analysis for novel features

It was concluded from the the research on feature selection and classifier optimization that the features available currently do not allow a better discrimination of the five classes and that more discriminant features are needed [117,130]. Prior to the actual development of the features, a requirement analysis was performed by analyzing the nearest neighbors in the feature space of the misclassifications of the best kNN classifier (k=19, cityblock distance). The rationale behind this approach is that the nearest neighbors of an object are those which are considered similar with respect to the current features. If an object is misclassified, and among the nearest neighbors are more objects from *another class* with a *visually distinguishable* property, this means that the current features do not quantify this property well. A feature needs to be developed which assumes different values whether or not this property is visible. Ideally, if this feature is also used for classification, only the objects of the correct class are considered as similar and the misclassification is corrected. Thus this analysis yields starting points for developing new features.

In total 4684 misclassified objects were analyzed, and three common patterns of misclassifications were observed:

1. **Lymphocytes vs. abnormal nuclei**: For low-grade abnormal nuclei, the changes in morphology are still subtle, and these still have an elliptical shape similar to those of lymphocytes. The classifier's decision is strongly based on the objects morphology (11 out of 18 features), consequently these classes are frequently mixed. Opposed to lymphocytes, abnormal nuclei contain nucleoli, mostly a circular nuclear structure which does not contain DNA and thus

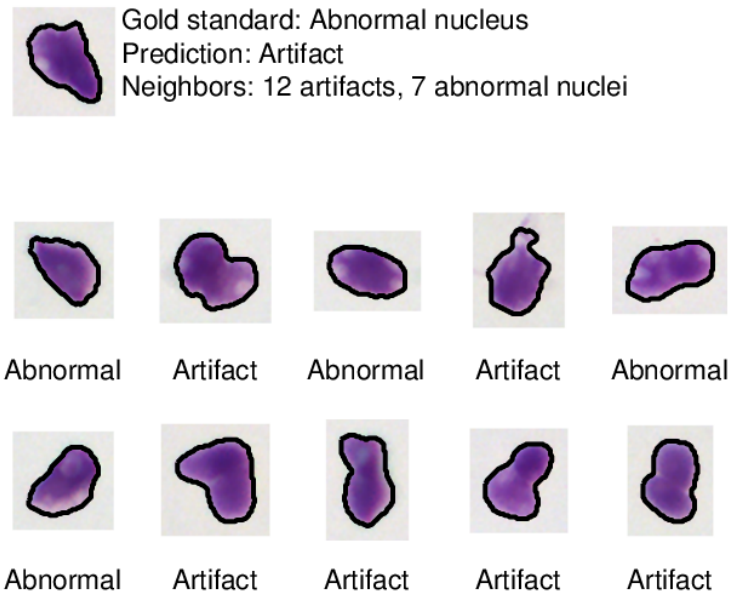


Figure 3.18: Requirement analysis. For the object to be classified (top), the nearest neighbors in the feature space are analyzed (for the sake of clarity limited to 10 neighbors in this example). Abnormal and overlapping nuclei are still considered as similar, since the current features cannot distinguish the ragged contour of abnormal nuclei from those of overlapping nuclei.

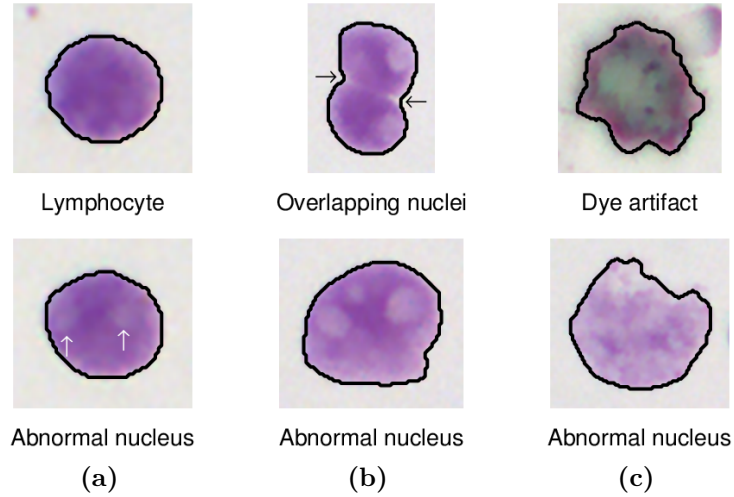


Figure 3.19: Patterns of misclassifications identified by the requirement analysis. (a) Lymphocyte and abnormal. Only abnormal nuclei contain nucleoli (white arrows), (b) Overlapping nuclei and abnormal nuclei. Overlapping nuclei usually have two oppositely aligned concavities (black arrows in this Figure, see also Figure 3.18), (c) Dye artifacts and nucleus.

is brighter after Feulgen staining (see Figure 3.19). This pattern occurred in 10.8 % of the cases. Features which quantify whether or not a nucleus contains nucleoli are needed.

2. **Overlapping vs. abnormal nuclei:** A subtype of artifacts are overlapping nuclei. Frequently, the neighbors of overlapping nuclei frequently were abnormal nuclei, and vice versa (5.5 % of the misclassifications). The reason for this is that both abnormal nuclei and overlapping nuclei have an irregular contour. For overlapping nuclei, however, the contour is irregular because of (usually two) deep concave regions, whereas for abnormal nuclei the whole contour is ragged. Therefore, features quantifying if the irregularities of the contour are due to concavities or not are required.
3. **Dye artifacts vs. abnormal nuclei** In 4.8 % of the cases nuclei contained dye artifacts which come from particles introduced during the preparation of the slide. All state-of-the-art features exploiting the intensity information only work on a gray image, therefore features depending on color information are needed.

For other misclassifications, no specific pattern could be observed.

3.3.4.5 Development of novel features

The requirement analysis provides the foundation for the actual development of features. First, several options for features quantifying the observed differences were developed and implemented. Second, the parameters of these features were optimized on a small training set. Third, the optimized features were integrated into the final classifier.

In the following, the features for the three patterns of misclassifications identified by the requirement analysis are presented. For a detailed description of the features, the range of parameters tested, and the optimization results, the reader is referred to [131]. The first class of features presented aims to quantify the existence of nucleoli, i.e. bright ellipse-shaped regions within the nucleus (observation 1). Two preprocessing steps are possible prior to computing the feature. First, in case a gray image is needed, either the green channel, a weighted combination of RGB channels, or the optical density image[3] are tested. Second, diffraction of light at the boundary is accounted for. Diffraction brightens up up boundary pixels, since this can disturb the quantification of bright regions, the segmentation mask M is downsized by morphological erosion to a smaller mask M_{erode}. For the features itself, eight options are examined:

- **MaxIntensity:** The gray image is filtered with a circular average filter, and the maximum filter response is returned. The only parameter of this feature is the radius of the filter.

[3] A bright region in an image has a low optical density. Therefore, the optical density image is inverted to yield high values for bright regions like the other gray images.

- **GradientSum:** At the boundary of a nucleoli, the gradient is increased. This feature is the sum of the gradient of the gray image of an object, normalized by its area.
- **BWCount:** The area of the bright regions of a gray image, found by Otsu's thresholding [132].
- **BWCircleLikelihood:** Again the bright regions from Otsu's thresholding are the starting point. This time, the computation is limited to the smaller mask M_{erode}, and single pixels are removed by morphological image processing. For all remaining connected components, a circularity measure is computed based on the mean intensity, their eccentricity and the ratio of the area of the nucleus to the area versus the convex hull. The final feature value is the maximum circularity criterion.
- **MeanShiftCircleLikelihood:** Similar to the BWCircleLikelihood feature, but this time the homogenous regions found by Mean Shift segmentation [133] are used.
- **HoughCircle:** This feature is the maximum response of the Hough transformation for circles [134] on the gray image within the segmentation mask. To reduce the computational burden, the Hough transformation is limited to circles with a radius within a small range.
- **BWHoughCircle:** The maximum response of the Hough transformation, on the image segmented by Otsu's thresholding.
- **MeanShiftHoughCircle:** The maximum response of the Hough transformation, on the image segmented by the Mean Shift algorithm.

Parameters which need optimization are the computation strategy for the gray image, the radius of the filter of the MaxIntensity feature, the bandwidth parameters of the Mean Shift algorithm, the range of permitted circles for the Hough transformation, and the computation strategy of the circularity feature for the BWCircleLikelihood and MeanShiftCircleLikelihood feature.

For discriminating overlapping nuclei and abnormal nuclei (observation 2), the features are based on the concavity analysis framework of Kumar et al. [105], which is described in detail in appendix A.2. The following features were tested:

- **ConcavityDepth:** The depth of an individual concavity is defined as the distance from the deepest concavity point to the corresponding convex hull segment. A final feature value is computed as a combination of the depth values of all concavities. Six combinations are examined: the mean depth, the added depth, the deepest concavity, the second deepest concavity, and the sum or the mean of the two deepest concavities.
- **NumberOfConcavities:** The number of concavities which exceed a minimum depth.
- **ConcavityAngle:** The angle between the lines connecting the deepest concavity point and the end points of the corresponding convex hull segment is

defined as the angle of an individual concavity. Similar to the ConcavityDepth feature, in the case of several concavities, the final feature value is computed as combination of the individual angles. Options for this are: The largest angle, the second largest angle, or the mean of all angles.

- **CCAlignment:** This feature is a first approach to quantify if two concavities are "oppositely aligned" (Concavity-Concavity Alignment) by considering vectors describing the orientation of a concavity. Either the lowest or the mean CCAlignment of all pairs of concavities is returned as feature value.
- **CLAlignment:** Another alignment approach considers angle between the line connecting two concavities and the orientation vectors (Concavity-Line Alignment). Either the lowest or the mean CLAlignment of all pairs is returned.
- **MeasureOfSplit:** The measure of split for a pair of concavities is computed from the depth of the concavities and their distance, or artificially set to a low value if the CCAlignment and CLAlignment are too high. The final feature value is the largest measure of split found by testing all pairs of concavities.

For these features, the thresholds (NumberOfConcavities, MeasureOfSplit) and the computation strategies (ConcavityDepth, ConcavityAngle, CCAlignment, CLAlignment) need optimization.

The existence of dye artifacts (observation 3) is quantified by the following features:

- **HSVLookUp**: This feature value is the percentage of pixels within S which are labeled as "belonging to nucleus" in the lookup table used for segmentation (see section 3.2.2.1).
- **HSVHueMax**: The predominant hue value within S, after an HSV conversion of the image.
- **MaxWeightedGray** The maximum value within S of a gray image, computed with different weights from the RGB image.

From this class of features, only the weights for the MaxWeightedGray feature need optimization.

For the optimization of the features parameters, three subsets from one slide of the gold standard were extracted. These subsets contain only the relevant classes which should be discriminated: abnormal nuclei and lymphocytes for the first pattern, abnormal nuclei and overlapping nuclei for the second, and dye artifacts for the third pattern. The parameters were optimized by an exhaustive parameter search and by means of the 1NN criterion as described in section 3.3.1.

Finally, the novel features were integrated into the existing classifier. This was accomplished by adding the 17 optimized novel features to the set of 18 state-of-the-art features and then performing a Sequential-Forward-Feature selection on this set. The selection was performed with two different classifiers, an SVM and a kNN classifier, both with the parameters from the classifier optimization. The SVM was chosen because it was the best classifier from the optimization. The kNN was chosen because for this classifier each feature has an equal weight for the final classifications;

Table 3.12: Performance of the kNN and SVM classifier, on the gold standard from prostate cancer biopsies, employing the feature set found by a sequential forward selection on the novel and state-of-the-art features. The features are given in the order by which they were selected by the Feature Selection. The novel features are in bold.

Classifier	Features used	Best CCR (%)
kNN	Perimeter, MeanRadius, **MaxWeightedGray**, CircularForm, **ConcavityDepth**, ClusterShade, Area, MinRadius, VarianceRadius, **MaxIntensity**, MarkovEntropy, **NbrConcavities**, ClusterProminence, **HSVLookUp**, MaxRadius	91.39
SVM	Perimeter, MeanRadius, MeanLuminance, Eccentricity, ClusterShade, MinRadius, **CLAlignment**, LuminanceVariance, MarkovEntropy, **MaxIntensity**, Circularity, **ConcavityDepth**, MaxRadius, Area	91.77

classifiers such as SVM or the Decision-Tree based Random Forests can diminish the influence of weak features during the optimization or induction of the Decision Trees. Thus the kNN classifier gives a more direct feedback for the discriminative power of the selected features than the other classifiers. Additionally, the kNN classifier allows to check how the patterns of misclassifications changes from the features used for the requirement analysis to the newly selected features.

For the SVM classifier, the classification rate raises from 91.34 % to 91.77 % (see Table 3.12). The best result is reached with 14 instead of 18 features. Among these features are three novel features. For the kNN classifier, the rate raises from 90.24 % to 91.39 %. Among the 15 features selected are 5 novel features. Concerning the patterns of misclassifications, which have again been analyzed using the same procedure as for the requirement analysis, the pattern of misclassified abnormals/lymphocytes due to a nucleoli lowered from 10.8 % to 5.1 %, misclassifications due to overlapping nuclei from 5.5 % to 2.2 %, and misclassifications due to color from 4.8 % down to 2.1 %. Table 3.13 shows the classification table of the SVM on the selected features. The correct classification rates for individual slides during the leaving one out training varied between 88.56 % to 96.04 %, with a coefficient of variation of 2.40 %. Appendix A.5.0.2 displays the classification performance limited to the prognostically relevant exceeding events (above 4.4 c) and DNA-aneuploid stemlines as well as the classification performance of individual slides.

Developing novel features is a labor-intensive process, however it leads to an increase of the classification rate by 0.43 percentage points for the SVM and 1.15 percentage points for the kNN classifier. The fact that the feature selection leads to less features for both the SVM and the kNN classifier furthermore shows that the discriminative power of the features has been increased. The developed classifier always identified

Table 3.13: Classification table for the SVM classifier on prostate, using the 14 features found by a sequential forward selection on the novel and state-of-the-art features.

		Ground truth				
		Artifact	Abnormal	Fibroblast	Lymphocyte	Granulocyte
Classification	Artifact	14992	527	146	102	37
	Abnormal	654	22679	232	1181	15
	Fibroblast	157	83	1152	9	2
	Lymphocyte	105	657	10	5210	29
	Granulocyte	1	1	0	0	1
	Total	15909	23947	1540	6502	84
	Error (%)	5.76	5.30	25.19	19.87	98.81

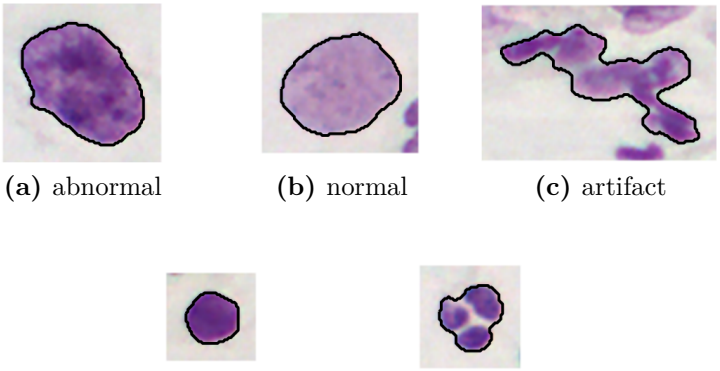
Correct classification rate:	91.77 %
Abnormals detected:	94.70 %
Relevance classified as abnormal:	91.59 %

at least thirty reference nuclei for the calibration on each slide. In total 94.70 % of the abnormal nuclei were identified correctly, and 91.59 % of all objects classified as abnormal are correct.

Concerning the performance of the individual novel features, MaxIntensity and ConcavityDepth are selected both in the feature selection for the kNN and SVM classifier. The feature MaxIntensity outperforms more complex features for the quantification of nuclei, for instance those based on circularity measures of bright regions or the Hough transform. The reason for this is that nucleoli are relatively small and span only a few pixels. Determining if such a small object is circular is therefore difficult. Furthermore, nucleoli might also be ellipse-shaped. The MaxIntensity feature, however, employs a circular average filter, and small deviations from a circular shape still lead to a high maximum filter response for objects with nucleoli and lower maximum filter response for objects without nucleoli. The ConcavityDepth feature, which already proved to be powerful for the detection of touching nuclei during the segmentation, is also powerful for the discrimination of overlapping nuclei from abnormal nuclei. Compared to the features quantifying the alignment or the feature MeasureOfSplit, it still can distinguish overlapping nuclei from abnormal ones even if the concavities are not well aligned. Features for dye artifacts are only chosen by the kNN classifier.

3.3.4.6 Conclusions

The best classifier developed detects 94.70 % of all abnormal nuclei, and 91.59 % of all objects classified as abnormal are correct. Since the cell material exclusively originates from cancer foci in core needle biopsies, there was no need to distinguish



(a) abnormal **(b)** normal **(c)** artifact

(d) lymphocyte **(e)** granulocyte

Figure 3.20: Classes of objects from the gold standard collected from cervical smears.

abnormal epithelial nuclei from normal epithelial nuclei—the main challenges were to distinguish low-grade abnormal epithelial nuclei from lymphocytes and high-grade abnormal nuclei from overlapping nuclei, and to identify dye artifacts. To this end, in total 17 novel features were developed to quantify visual differences between distinct classes. The development furthermore included identifying the best classifier algorithms and a feature selection on the novel and state-of-the-art features. The best classifier was an SVM, employing 3 novel and 11 state-of-the-art features. This classifier detects abnormal nuclei with the aforementioned performance and has an overall correct classification rate of 91.77 %.

3.3.5 Classification of nuclei from mucosal membranes

The third nucleus classifier presented is for nuclei from mucosal membranes. Applications for this classifier are the identification of cancer cells in brush biopsies of the uterine cervix or of suspicious lesions in the oral cavity. The two development steps of the prostate classifier which yielded the highest increase in classification rate are also used for this classifier, the optimization of classifier algorithms and the combination of state-of-the-art and novel features.

3.3.5.1 Material

Nuclei were obtained from the uterine cervix of nine women by a brush biopsy. The brush was smeared onto a glass slide, fixed in alcohol and subjected to staining according to Papanicolau. Subsequently, the slides were restained according to Feulgen. From these slides, a gold standard containing 35,920 objects was created. The gold standard contains nuclei from five classes: Artifacts, abnormal epithelial nuclei, normal epithelial nuclei, lymphocytes, and granulocytes (see Figure 3.20).

Table 3.14: Performance of different classifiers on the cervix gold standard. For each classifier model, also the best parameter configuration is given.

Classifier	Parameters	Best CCR (%)
kNN	$k = 14$, Cityblock distance	89.02
Neural Network	100 layers	89.01
Decision Tree	91 min. branch node observations, splitting by error rate, pruning by deviance	86.67
SVM	C=32, $\gamma = 1$	**91.31**
Random Forest	140 classifiers, randomly select 8 features	90.43
Adaboost	115 classifiers	88.87

3.3.5.2 Methods

For the prostate classifier, a systematic optimization of the parameters of classifier algorithms and a selection of state-of-the-art and novel features yielded the highest increase of the correct classification rate. Therefore, these two steps are also applied for optimizing the classifier for nuclei from mucosal membranes. The optimization of classifier algorithms is performed first, employing the state-of-the art features. Second, a feature selection is performed on the state-of-the-art and novel features with the best classifier algorithm identified in the previous step.

3.3.5.3 Results

As it is the case for the prostate classifier the SVM is best, it achieves a correct classification rate of 91.31 %. Table 3.14 displays the correct classification rates for the classifier algorithms tested. The feature selection increases the classification rate to 91.71 %, using 3 novel and 17 state-of-the-art features (see Table 3.15). This classifier identified abnormal nuclei with a detection performance of 93.85 %, and 96.28 % of all objects classified as abnormal are correct (see Table 3.16). The correct classification rates for individual slides during the leaving one out training varied between 89.69 % to 95.21 %, with a coefficient of variation of 2.00 % (see Appendix A.5.0.3).

3.3.5.4 Conclusions

The classifier for nuclei from mucosal membranes is, like its counterparts for effusions and prostate cancers, able to identify abnormal nuclei from all other types of objects—to this end, the development process used for the prostate classifier was also successfully applied to this classifier. More than 92 % of all normal nuclei are detected, sufficient to use these nuclei as reference cells for DNA computations.

To summarize, classifiers which are able to discriminate abnormal nuclei from all other types of nuclei and artifacts were developed for nuclei from effusions, prostate

Table 3.15: Performance of the SVM classifier on the cervix gold standard, on the feature set found by a sequential forward selection on the state-of-the-art and novel features. The features are given in the order by which they were selected by the Feature Selection. The novel features are in bold.

Classifier	Features used	Best CCR (%)
SVM	**HSVLookUp**, MinRadius, Perimeter, LuminanceMean, ClusterProminence, **ConcavityDepth**, Entropy, ClusterShade, RadiusMax, Eccentricity, MinFilter, RadiusMean, Inertia, BendingEnergy, CircularForm, RadiusVariance, Area, LuminanceVariance, Sphericity, **NbrConcavities**	91.71

cancers and mucosal membranes. The detection performances for abnormal nuclei of these types of specimen 95.24 %, 94.70 % and 93.85 % respectively. Key factors for reaching these performances was a careful choice of state-of-the-art features (all types of specimen), as well as a systematic optimization of classifier algorithms and the development of new features (prostate cancers and mucosal membranes). The performance statistics were computed on gold standard sets of several tens of thousands annotated nuclei, therefore they are valid estimators of the classifiers performance to discriminate the individual types of nuclei. The performance of these classifiers in clinical routine as well as the resulting benefits for practice will be studied in chapter 4.

3.4 Detection of regions of interest for scanning

The algorithms presented so far have been developed for the automated image acquisition and processing of individual FOVs. In many cases DNA Image Cytometry is performed only in regions of interest which are relatively small compared to the whole slide to save time during the scanning. This section presents algorithms for the automated detection of these ROIs.

If DNA Image Cytometry is used to assess suspicious cases from conventional cytology (see chapter 2.2.2), the pathologist has previously identified suspicious nuclei and requests to measure their DNA content. The corresponding regions are marked with a felt tip pen on the slide. Alternatively, some slides already specify a ROI by an imprint on the slide, for example if a certain preparation technique concentrates the cells in a predefined area. Figure 3.21 shows two slides with felt tip marked ROIs and one ROI imprint. The slide is restained for DNA Image Cytometry, and the measurement is carried out only on nuclei in these regions.

To consider an alternative to the felt tip marker procedure, if the microscope for the

Table 3.16: Classification table for the SVM classifier on cervical nuclei, employing the 20 features found by a sequential forward selection on the state-of-the-art and novel features.

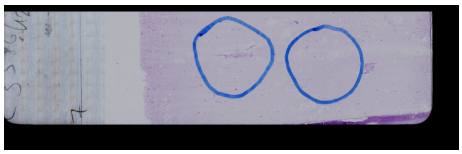
		Ground truth				
		Artifact	Abnormal	Normal	Lymphocyte	Granulocyte
Classification	Artifact	19929	175	284	194	756
	Abnormal	139	4270	23	1	2
	Normal	267	77	3989	33	9
	Lymphocyte	50	20	29	524	14
	Granulocyte	759	8	1	36	3087
	Total	21144	4550	4326	788	3868
	Error (%)	5.75	6.15	7.79	33.50	20.19

Correct classification rate:	91.71 %
Abnormals detected:	93.85 %
Relevance classified as abnormal:	96.28 %

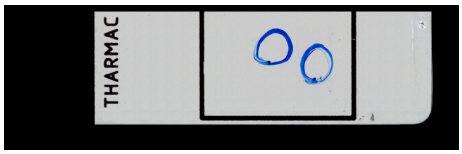
initial analysis would be equipped with a motorized stage positioning system, the relevant cells could be automatically repositioned after the restaining. This could be done, for instance, using the algorithms of the Multimodal Cell Analysis framework [83, 135, 136]. In practice however, the initial analysis is often performed on a microscope from a different vendor, and it cannot be assumed that this microscope is equipped with a positioning system with an open interface. Therefore, marking the ROIs on the slide allows to encode this information without needing a positioning system. If a slide with felt tip markings is scanned manually, the pathological expert takes care that only the ROIs are scanned. If the slide is scanned automatically, one possibility would be to obtain an overview scan either by an additional camera or by scanning the slide at low microscope magnification. Then the user would manually select ROIs in these overview scans, and the slide would only be scanned in the regions selected. This procedure is time-consuming. Even worse, it requires manual interaction *every time before* the slide can be scanned, which for instance hampers the automated scanning of several slides overnight.

To allow for scanning of slides without user interaction, algorithms for automatically detecting ROIs from overview scans are presented. These algorithms contain methods to identify slide types of different manufacturers, to automatically identify felt tip markings, close incomplete markings and extract the ROIs. So beyond sparing the user the time-consuming step of manually selecting the ROIs for scanning, these algorithms enable the automated *batch scanning* of slides, since no manual interaction is required before scanning each slide. Scanning overnight will be possible, which considerably increases the utilization of the device.

Parts of this work have been published by the author in [137].



(a) Standard slide



(b) A slide from Tharmac with felt tip markings

(c) A slide from Tharmac without felt tip markings

Figure 3.21: Slides with felt tip markings.

3.4.1 Material

A MotiCyte BA600 Microscope and a MotiCam 285A RGB camera (1360×1024 pixel) were used for imaging. Overview scans of microscope slides were acquired by stitching together 280 images acquired at 2× magnification. This process takes approximately 40 seconds. In the overview scans, one pixel in the image corresponds to 51×51 µm^2 on the slide. In total 174 slides with annotated ROIs were scanned. The set includes six different slide types and was divided into a training and a test set, each containing 87 slides. The ROIs are circumscribed by red, blue, green, and black felt tip pens with a width of approximately 0.6 mm. For every scan in the training set and set set, the slide type and the ROIs which should be detected by the algorithm were manually annotated.

3.4.2 Methods

In the following, the image processing algorithms for the detection of ROIs are presented, an overview is shown in Figure 3.22.

3.4.2.1 Detection of the slide type

In clinical routine, a pathological institute uses different slide types and receives slides from different ordering institutions. The variation of slide types makes it difficult to develop a general approach for the detection of ROIs: slides might contain the name of the manufacturer, and parts of the names might be confused with felt tip markings (for instance the “R” in “Tharmac” in Fig. 3.21(b)). Some slides already specify a region of interest, felt tip markings should only be detected in these regions. Therefore, the slide type and a region for detecting felt tip markings are identified first.

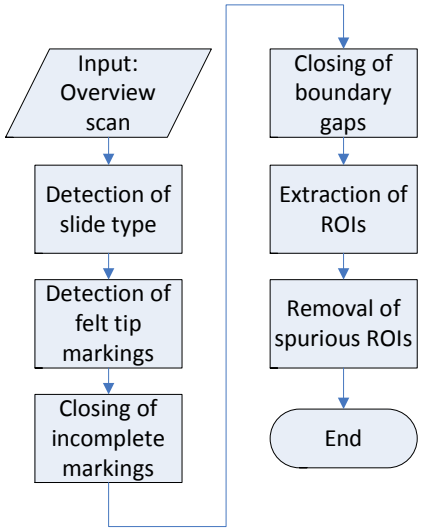


Figure 3.22: Flow chart for the detection of ROIs from overview scans.

Table 3.17: Parameters for post processing.

	Low	Medium	High
Dilation/erosion detected markings (mm):	0.15/0.075	0.25/0.1	0.3/0.2
Remove markings smaller than (mm^2):	2	3	5
Remove markings with contour variation larger:	0.2	0.15	0.12
Remove ROIs smaller than (mm^2):	0.5	0.75	1

Detection of the slide type is realized by scanning empty reference slides once for each type. An expert then annotates two kinds of regions:

- The *recognition region*: A region comprising the features of the slide which are characteristic for the slide type. This region contains for instance the manufacturer's name and markings for depositing cells. However the label region containing the case number is excluded, since this region is specific for the individual slide, but not for the slide type.
- The *detection region*: Region in which felt tip markings should be detected.

These annotated slides constitute a slide reference database. For determining the type of an analysis slide, the normalized correlation of this slide with each slide from the reference database is computed, and the slide type which yields the highest correlation is assigned. The detection region of the reference slide is then transferred to the slide to be analyzed.

3.4.2.2 Detecting felt tip markings

The detection of felt tip markings consists of three steps: gray-level image conversion, detection of felt tip markings, and post processing.

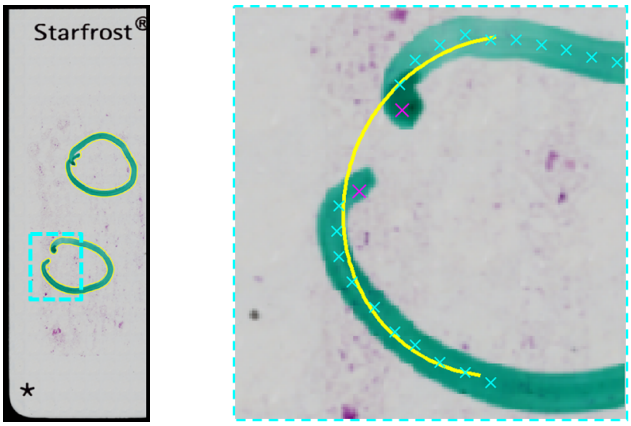
Two methods for the conversion of color images into a gray-level image are used. A first approach is to compute the gray-level image as weighted combination of the color channels ($r = 0.299$, $g = 0.587$, $b = 0.114$). A second approach is based on the HSV colorspace and considers that for the current task black markings or markings of highly saturated colors need to be discriminated from background and pale stain color. The pale stain is visible in densely populated slides (see Fig. 3.21(b)), which has a lower saturation as colored felt tip markings. Black markings, which also are used in practice, also have a low saturation, but differ from stain and background by a lower light intensity. To achieve a gray-level image with low values for felt tip markings and high values for background and stain color, the gray-level image M is computed as

$$M = (1 - s) \cdot v,$$

where s is the saturation component and v the light intensity of the HSV color space with range $[0, 1]^3$.

For discriminating felt tip markings from background and stain, two adaptive thresholding methods are applied on the gray-level image. The first method is Otsu's thresholding method, which separates two distributions based on the gray level histogram and by finding a threshold which maximizes inter-class variance and minimized intra-class variance [132]. However, for the problem at hand, the percentage of felt tip markings compared to the whole detection area is relatively low, usually four to five percent. Therefore, the statistical estimators of Otsu's method might have problems estimating the distributions correctly. For the second approach, a Expectation Maximization (EM) approach, the background values are excluded from the analysis by only considering gray level values below 0.6. A Gaussian distribution is fit to this model by the EM algorithm and all gray level values in the 1.96σ-surrounding of the mean are segmented as felt tip markings. The corresponding thresholds have been found by a gray level histogram analysis on the training set. In some slides, no felt tip marking is present in the slide, which is detected as follows: if the area of the detected felt tip markings is less than $5\,\mathrm{mm}^2$ or if the gray level found by Otsu's algorithm is above 0.75 (gray image) or 0.7 (saturation and value), this might indicate that no felt tip markings have been put on the slide. In that case, the whole detection region is used as ROI.

Post processing is applied to smooth the detection results and remove spurious elements. Smoothing of the detected felt tip markings is performed by a morphological dilation, followed by erosion. In order to create a small gap between the felt tip markings and the final ROI, the structure element for dilation is larger than for erosion. All elements which are smaller than a certain threshold are removed. Also if the distance from the centerline to the contour varies too much, an element which is too ragged to be a valid felt tip marking has been segmented and is therefore removed. Three levels of post processing are applied (low, medium and high level). Table 3.17 lists the corresponding parameters and thresholds.



(a)

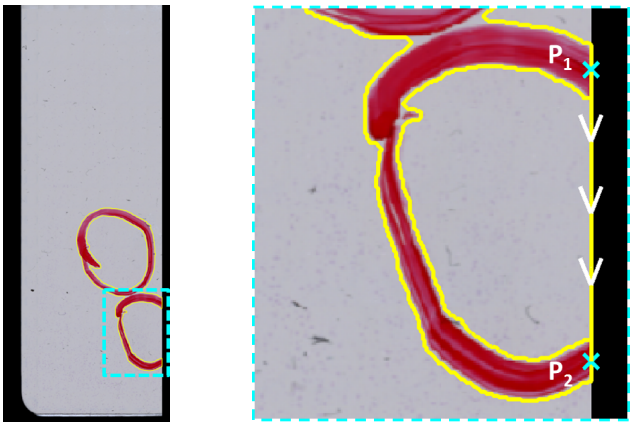
(b) Ellipse fit for incomplete felt tip markings

Figure 3.23: Closing incomplete felt tip markings: In (b) it is shown how the gap within the square region of (a) is closed. The end points of the centerline of the detected felt tip markings are marked by magenta crosses. From there, the centerline is traced back and points on the trajectory are extracted (cyan crosses). An ellipse (yellow) is fit to these points and used for closing the detected felt tip markings.

3.4.2.3 Closing incomplete felt tip markings

Gaps in detected felt tip markings occur if the color of the felt tip marking is too weak to be detected or if the manual drawing is incomplete (see Figure 3.23(a)). As later the ROIs are extracted from circumscribed regions, these gaps need to be closed. Three algorithms have been developed for this purpose:

- **Morphological approach:** The detection mask is dilated by 1mm and the morphological skeleton of the dilated mask is computed. If the dilation connects two disjoint objects, they remain connected in the skeleton. By setting all pixels of the skeleton as true in the detection mask, the gaps are closed without expanding the detection mask into regions where no connection was established.
- **Linear approach:** A line is fit to the end points of the detected felt tip markings. The line is extended by 1mm, and if another part of the felt tip marking is found in this region, the pixels of the line are marked as true in the detection mask
- **Ellipse fit:** If two endpoints of the centerline of detected felt tip markings are closer to each other then 2mm, they are connected via an ellipse fit. Starting from the two end points, the corresponding felt tip markings are traced back by 5mm, and every 0.5mm the coordinates of the current tracing point are extracted. An ellipse is fit to these points using the algorithm from [138]. The pixels of the ellipse between the two end points are marked as true in the



(a) (b)

Figure 3.24: Closing boundary gaps: The algorithm identified two points P_1 and P_2 where felt tip markings touch the boundary (cyan crosses). Starting from point P_1, the boundary is traced into direction of P_2. As they are close enough to each other, they are connected (yellow line).

detection mask (see Figure 3.23(b)).

As each of these approaches might be advantageous in a certain situation, the combination of these algorithms is also examined.

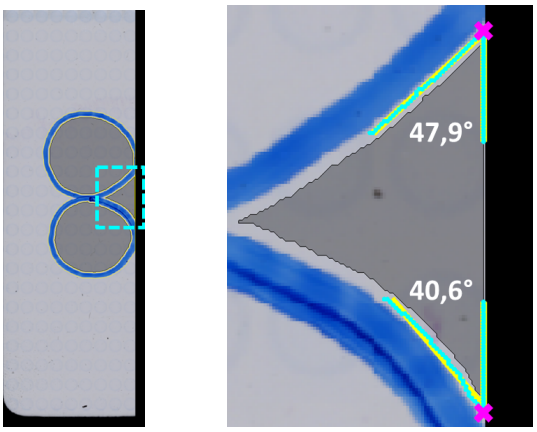
3.4.2.4 Closing boundary gaps

In clinical routine it might happen that cells on the very border of the slide need to be scanned. In these cases, the ROIs are not fully circumscribed by felt tip markings, but also by the slide boundary (see Figure 3.24(a)). Just as the gaps due to incomplete felt tip markings, these gaps need to be closed for extracting ROIs.

Felt tip marking segments which touch the boundary are identified by finding pixels which are marked as true in the detection mask and are, at the same time, boundary pixels of the detection region. For all possible pairs of such points, it is checked if they should be connected. To this end, starting from the first point, the boundary of the detection region is traced for at maximum 2cm, and if the second point is within this range, they are connected in either of two ways. In the first method, only segments of the same detected object can be connected. In the second method, any object can be connected with any other object touching the boundary. The pixels which connect these two objects are also marked as true in the detection mask (see Figure 3.24(b)).

3.4.2.5 Extracting ROIs

After detecting the felt tip markings and closing gaps between adjacent felt tip markings or along the slide boundary, it is expected that the ROIs are fully circum-



(a) (b)

Figure 3.25: Removing spurious ROIs: The gap between two felt tip markings has been closed, which causes the detection of a spurious ROI (dark gray). This case is detected by finding the corner points (magenta crosses), fitting lines to the ROI boundaries (yellow lines) and computing the angle between them. If both angles are below 50°, the ROI is removed.

scribed by detected objects. The ROIs themselves are extracted by applying a flood fill operation on the detected objects and subtracting the detected objects.

3.4.2.6 Removing spurious ROIs

Another post processing step is applied to remove ROIs which are too small to be an intentional ROI. Table 3.17 lists the thresholds for removing ROIs based on their area. Similar to the post processing for detected felt tip markings, three levels of post processing are examined.

Although small spurious ROIs can straightforwardly be removed based on their area, a region can be circumscribed by two touching felt tip markings and the boundary region. In that case, the algorithm for closing boundary gaps causes the detection of a spurious ROI above the area threshold (see Figure 3.25(a)). These regions differ from relevant ROIs by a triangular shape, and this fact is used for removing them. For every ROI which touches the slide boundary, the corner points are detected by finding the end points of the joint pixels of the slide and ROI boundary. Subsequently, starting from a corner point the slide boundary is traced back and a line model is fit to points on the way. The same is repeated, again starting from a corner point, but this time tracing the boundary of the detected felt tip markings. The angle between the two line models at a corner point is computed. If the angles at both corner points are smaller than 50°, the region is removed. This angle has been heuristically found on the training set. Figure 3.25(b) displays examples of line fits and the angles between the fits.

Table 3.18: Detection performance on training set and test set.

	Sensitivity (%)	PPV (%)
Training set	99.23	100.00
Test set	98.45	97.65

3.4.2.7 Evaluation and optimization of parameters

The first step, the determination of the slide type, is assessed by the correct classification rate. For the second step, the detection of ROIs, a region is considered as detected (true-positive) if it fully covers the manual annotation of this region. A detected region is considered as false-positive if it does not fully cover any manual annotation. Two kinds of errors are possible:

- A relevant ROI is not detected: in that case, important diagnostic information might be missing and the slide needs to be scanned again.
- A spurious ROI is detected: this increases the scanning time unnecessarily.

Sensitivity and positive predictive value (PPV) are used for evaluating the detection performance. The sensitivity is the percentage of correctly identified ROIs. The positive predictive value is the percentage of correctly identified regions among all regions detected and thus reflects the relevance of detected regions.

The optimization of parameters is performed by an exhaustive search on the training set. As a scalar measure is needed for the optimization, the F-Score, which is the harmonic mean of sensitivity and positive predictive value, is used for this purpose.

3.4.3 Results

The slide type could be identified correctly on all slides of the training set. For detecting ROIs the highest performance with an F-Score of 99.65 % was reached using the following parameter combination:

- Gray-level image: combination of saturation and value component
- Detection of felt tip markings: EM-Algorithm
- Post processing level: medium
- Closing incomplete felt tip markings: ellipse fitting and linear approach
- Closing boundary gaps: connect same objects
- Removing spurious ROIs: activated

Sensitivity for this parameter configuration is 99.23 % and the positive predictive value 100 %.

For the test set, the slide types were identified with a correct classification rate of 98.85 %. Sensitivity and positive predictive value are 98.45 % and 97.65 % respectively. Table 3.18 summarizes the results.

Table 3.19: Influence of the extensions developed in this work.

	Sensitivity (%)	PPV (%)
Otsu's algorithm	70.63	95.28
EM algorithm	71.33	100.00
+ closing gaps in felt tips	85.60	100.00
+ closing gaps at boundary	99.30	98.61
+ removing triangle ROIs	99.30	100.00

3.4.4 Discussion and conclusion

Algorithms for automatically detecting felt tip markings on microscope slides have been presented. These algorithms are required to unburden the operator of an automated DNA Image Cytometry system of manually selecting the ROIs for scanning. Thus they enable batch processing of slides.

A straightforward approach for detecting felt tip markers would be to use Otsu's thresholding method on a gray-level image. However, this approach only achieves a sensitivity of 70.63 % and a positive predictive value of 95.28 % on the training set. Several extensions needed to be developed for achieving an acceptable detection performance. In order to motivate these extensions and quantify their influence, Table 3.19 shows how these extensions improve the performance on the training set. Using a gray-level image composed of a saturation and value component and using an EM-approach, sensitivity is raised by 0.7 percentage points, and a perfect PPV is reached. Sensitivity considerably increases if gaps in felt tip markings and at the slide boundary are closed. However, closing gaps at the boundary also causes the detection of spurious ROIs (see chapter 3.4.2.6), which reduces the formerly perfect PPV. The PPV can be raised back to perfect PPV by estimating the angle at the corner points and removing ROIs which are too acute.

After optimization on a training set, the algorithm achieves a sensitivity of 98.45 % and a PPV of 97.65 %. The decrease compared to the training set is low (sensitivity −0.87 %, PPV −2.35 %), which means that the the algorithm generalizes well on unseen data. Two ROIs were missed on the test set. The first region was missed as a consequence of a misclassification of the slide type, and the second region because a big gap within a felt tip marking was not closed correctly. In practice, these slides would need to be rescanned. A few very small spurious regions have been detected, which lowered the PPV. However, these are very small regions, with an average area of 3.66mm^2, which would only slightly increase scanning time. As these regions are either empty or contain non-relevant nuclei which are identified by the classifier, this has no impact on the clinical result. In conclusion, batch processing of slides is empowered by the presented algorithms. Compared to scanning the whole relevant area of all slides, only 30.3 % need to be scanned when limited to the detected ROIs.

4 Clinical application

Plenary talks like "Medical Imaging Computing for translational biomedical research" (BVM 2013) or "CAD - success or failure?" (SPIE 2014) on recent medical image processing conferences demonstrate that there is an increasing awareness in the scientific community that more effort must be spend to translate their algorithms into practice. This chapter is devoted to the translational research of the work described in the previous chapter. An easy to operate user-interface is one indispensable precondition for a successful transfer of the research results into clinical application. The other is a careful integration of its operation into a routine workflow. These aspects are presented in the first two sections. Subsequently, the evaluation of the developed algorithms and workflow as an integrated solution in three applications is presented: the identification of cancer cells in nuclei from serous effusions and from brush biopsies of the oral cavity, and grading the malignancy of prostate cancer biopsies.

4.1 User interface

A software for the clinical end users cannot rely on command line tools or laboratory prototypes, which have been used for research. Instead, a software which is easy to use, robust, and fast is desired (see "Ron's rules for tools" [139]). Most of the algorithms from chapter 3 have been included into a commercially available software package (MotiCyte) in cooperation with Motic Asia[1]. The user's view of the interface is composed of the following interactive segments called "panels" (see Figure 4.1):

1. **Video panel:** This panel displays the live image from the camera on a screen. After segmentation of nuclei, their delineation is superimposed. The assignment to different nucleus classes after classification by the nucleus classifier is encoded by color.
2. **Gallery:** The gallery shows the nuclei which have been collected and classified.
3. **Group panel:** In the group panel, the user can assign names for nucleus classes, select the nucleus classes for calibration of the system and the final diagnosis, or define which nucleus classes are displayed in the gallery. In

[1]The authors contributions include: implementation of the developed image processing algorithms, conceptual design of user interfaces in cooperation with cytopathologists, testing of the software, distribution of software packages for the end users and integrating their feedback.

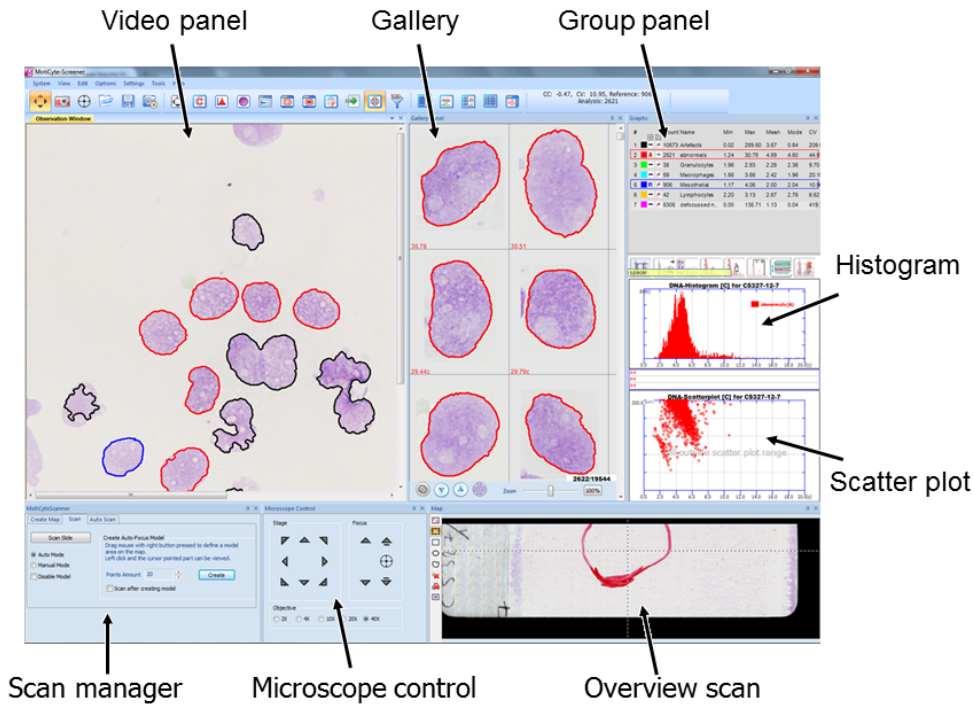


Figure 4.1: User Interface of the MotiCyte screener software.

addition, basic statistics based on the DNA values of the objects of each class are given.

4. **Histogram:** The DNA histogram which is used for the final diagnosis or prognosis. Also, a scatter plot is displayed showing the relationship between the DNA value and the area of nuclei.
5. **Scan manager:** The scan manager can be used to performing an overview scan of a slide, setting the focus points for the focus model and starting the automated scanning process (see section 3.1).
6. **Microscope control:** This panel contains buttons for moving the microscope stage in x, y and z-direction and for changing the microscope objective.
7. **Overview scan:** The overview scan of the slide, acquired as described in section 3.4, is shown in this panel.

The software can be used for both manual and automated DNA Image Cytometry. For manual DNA Image Cytometry, a camera mounted to a manually operated microscope is used, and scan manager, microscope control and overview scan are deactivated. For automated DNA Image Cytometry, the camera is mounted to a motorized microscope and all modules are activated. Besides the image processing algorithms for ROI detection, segmentation, and classification described in previous chapters, the following algorithms provide further support for the user: The interpretation of the DNA histogram for the final diagnosis or prognosis is facilitated

by tools highlighting exceeding events, and an automated DNA-stemline detection. For handling patient information and documenting the diagnostic findings, a patient database and a measurement report are included. The report can be flexibly adapted and layouted to meet the specifications of the user site. The software employs the algorithms for DNA measurement developed by Haroske et al. [49] and Würflinger at al. [50], which include the correction of diffraction, background, and glare.

4.2 Workflow

The integration of the developed algorithms and software into a clinical workflow provides the clinical end users with a clear sequence of steps how to use both in practice. Two workflow aspects are discussed in this section: The first aspect is about beneficially dividing the work between machine and expert for minimizing the workload for the expert and maximizing diagnostic accuracy. The second aspect is about the extraction of clinically relevant nuclei from the automatically collected objects to reduce the number of objects which need to be handled by the software.

The nucleus classifiers have correct classification rates between 88.11 % to 91.77 % compared to a pathological expert (see Section 3.3). Whereas some misclassifications are irrelevant for the outcome of the measurement, for example the classification of a reference nuclei as artifact, especially the classification of objects in the DNA range of exceeding events is critical. Here, the presence of one exceeding event can already change the diagnosis or prognosis: Classifying an artifact with high DNA content as abnormal nucleus, in the presence of no other abnormal nuclei in this range, leads to a false-positive diagnosis or an overgrading. Overlooking all abnormal nuclei with high DNA content yields a false-negative diagnosis or an undergrading. As demonstrated in the previous chapter, the classification of these nuclei is not perfect, and due to the high variability of abnormal nuclei and artifacts possibly it will never be perfect. To compensate for this, a *verification step* is integrated into the workflow: after the automated digitization and classification of nuclei, the user first checks the reference cells to ensure a correct calibration of the DNA measurement algorithms. Then, the user is requested to check all objects above the exceeding event threshold which have been classified as abnormal, artifacts, or defocused. All misclassifications are reclassified. After this, the diagnosis or prognosis is performed based on the DNA distribution of the abnormal nuclei. The conventional workflow and new one are compared in Figure 4.2.

The benefit of this verification step is twofold. First, it accounts for the difficulty to classify artifacts and abnormal nuclei, which have a very high variability. Second, the verification step increases the confidence of the pathologist into the system; it is no “black-box” solution, but the pathologist has seen all nuclei before making a diagnosis or prognosis. The effort for the verification step is readily justified. As the nuclei are compactly presented in a gallery, the verification is usually completed within five minutes.

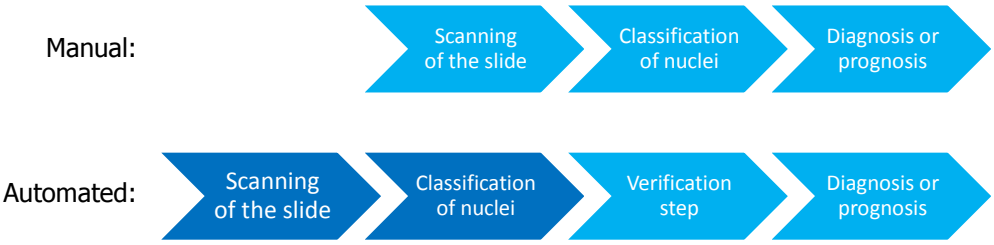


Figure 4.2: Comparison of the workflows of manual and automated DNA Image Cytometry. Parts which are performed manually are in light blue, whereas parts which are performed without user interaction are in dark blue.

Table 4.1: Extraction of clinically relevant data of an oral smear.

Classes	Full measurement	Reduced set
Normal epithelials (reference)	10515	50
Abnormals	166	166
Lymphocytes	254	0
Granulocytes	3294	0
Artifacts	44220	1794
Defocused objects	2547	66
Total	60996	2076

A microscope slide may contain up to 100,000 objects ore more. To give an example, table 4.1 shows the object statistics of an automatically collected slide from an oral smear. However, a huge fraction of these data are irrelevant for the clinical outcome. If these data is still kept it will unnecessarily increase the processing time and storage requirements. Therefore, the collected objects are reduced to three classes:

1. **Safe reference nuclei:** According to ESACP requirements, thirty reference nuclei are needed for the calibrating of the the system [8]. Especially for oral smears and effusions, where normal epithelial cells are used for the calibration, there are often several thousands of reference cells, which would need a check during the verification step. To reduce the number of objects for this step, 50 reference nuclei (the required 30 plus 20 supplementary nuclei) are extracted by selecting the ones with the highest classification confidence. By doing so, it is very probable that these classifications are correct. If there would be misclassifications, the user would delete them in the verification step, but as 20 supplementary nuclei are available there will be still enough nuclei for the calibration.
2. **Abnormal nuclei:** All abnormal nuclei are kept.
3. **Artifacts for verification:** As the user only has to verify all artifacts and defocused nuclei above the exceeding event threshold, all artifacts with a lower DNA content are rejected.

This procedure typically reduces the number of objects by 50 % to 70 %.

4.3 Clinical studies

The developed algorithms and workflow were evaluated as an integrated solution in three applications: the identification of cancer cells in nuclei from serous effusions and from brush biopsies of the oral cavity, and the grading of malignancy of prostate cancers under active surveillance. The assessment covered the comparison of the overall clinical outcome with manual DNA Image Cytometry, the detection performance for the clinically relevant DNA-criteria (exceeding events and aneuploid stemlines), and the effort for the verification step.

Manual and automated measurements were performed with the MotiCyte software. The ROIs for scanning were selected still manually, since the algorithm for ROI detection was not yet available in the software package. Also, the study on effusions used the segmentation algorithm by [78]. For all other cases, the segmentation pipeline and the best classifier for the corresponding type of specimen from chapter 3 were used.

4.3.1 Identification of cancer cells in serous effusions

In a first study, the performance of manual and automated DNA Image Cytometry for effusions were compared to each other, with the following research questions:

1. Which of the method is superior in finding DNA-aneuploidy?
2. By which of the method more exceeding events or DNA-aneuploid stemlines are discovered?
3. How many objects need to be analyzed and reclassified during the verification step?

Additionally, it was retrospectively analyzed if a statistic based on the morphometric selection of abnormal nuclei can be exploited as an additional diagnostic marker. The aim of this approach is to identify cancer-cell positive cases where the changes in the amount of DNA are below the detection limit, but morphological changes of nuclei already indicate a malignant transformation.

4.3.1.1 Material

Body cavity effusions of 122 patients were punctuated by fine needle biopsy. Cells contained in the extracted liquid were sedimented, smeared onto a glass slide and air-dried. First, the slides were stained according to the May-Grünwald-Giemsa scheme and analyzed due to their morphology. Cytological diagnoses concerning the presence or absence of malignant cells of different types were derived. Subsequently, the slides were restained stoichiometrically for DNA according to Feulgen. In order to reduce the processing time, the scanning was limited to areas with high cell density. These areas were marked with a felt tip pen on the slide prior to scanning.

Table 4.2: Comparison of manual and automated DNA Image Cytometry for the detection of DNA-aneuploidy in serous effusions.

Effusions		Manual	
		DNA-euploid	DNA-aneuploid
Auto	DNA-euploid	73	0
	DNA-aneuploid	4	37

A ground truth was established for clarifying discrepancies between both manual and automated DNA Image Cytometry and evaluating the performance of automated measurement for the detection of cancer. The slides were classified into positive for cancer-cells under the consensus of a morphological diagnosis of two experienced cytopathologists. In all diagnostically suspicious or doubtful cases, immunocytochemistry using the BerEp-4 antibody against an epithelial-specific antigen was performed (14 cases). The study population contained 54 cancer-cell positive and 60 negative cases.

4.3.1.2 Experiments

Manual DNA Image Cytometry measurements of these slides were performed at the Institute of Cytopathology, Heinrich-Heine University Düsseldorf, Germany. For the automated version, the system was calibrated according to the standardized image acquisition of section 3.1. After about 20 focus points had been set manually, the automated scanning of the slide was started. A cytopathologist then performed the verification step. Subsequently, for both the manually and automatically measured cases, the number of exceeding events and DNA-aneuploid stemlines were counted. The thresholds for exceeding events in effusions specimen is set at 9.0 c, and an DNA-aneuploid stemline is defined as a stemline outside the intervals $[1.8\text{c}, 2.2\text{c}]$ and $[3.6\text{c}, 4.4\text{c}]$. A case with an exceeding event or a DNA-aneuploid stemline is classified as "DNA-aneuploid", otherwise as "DNA-euploid" [8, 9].

4.3.1.3 Results

In 114 out of the total 122 cases, an automated measurement could be successfully performed. In three cases, there were insufficient nuclei on the slide for the autofocus algorithm. In one case, the staining of the nuclei was too pale and thus nuclei could not be segmented correctly. In four other cases, the diagnostically relevant nuclei were too close to the boundary of the slide. Opposed to the manual microscope, the stage of the BA600 used for automated DNA Image Cytometry does not allow the acquisition of nuclei at the boundary of the slide.

Table 4.2 compares the results from manual and automated measurements for the detection of DNA-aneuploidy. In 110 out of 114 cases the diagnoses coincide. However in four cases, automated DNA Image Cytometry detected DNA-aneuploidy

Table 4.3: Detection performance of the clinically relevant DNA-criteria in serous effusions. In the upper part, the number of 9 c-exceeding events and aneuploid DNA-stemlines on the whole study are given. In the lower part are the number of cases where more, as many as, or less of these criteria were found in the manual measurement.

	9 c-exceeding events	DNA-aneuploid stemlines
#manual (total)	242	53
#auto (total)	435	45
#cases manual > auto	17	17
#cases manual = auto	86	86
#cases manual < auto	11	11

while the manual did not. A comparison to the ground truth revealed that all these four cases were positive for cancer cells. The number of 9 c-exceeding events and DNA-stemlines detected by both methods and the number of cases where one or the other method found more clinically relevant objects are shown in table 4.3. During the verification step, on average 164.63 artifacts larger 9.0 c and 3.38 abnormal nuclei above this threshold had to be analyzed per slide. It was necessary to correct one artifact in every seventh slide, which had been misclassified as abnormal, and one abnormal nuclei in two out of three slides which had been misclassified as artifact.

4.3.1.4 Discussion and conclusions

The presented study proved the effectiveness of the nucleus classifier for effusions for identifying morphologically abnormal nuclei, integrated into an overall solution and evaluated on 122 cases from clinical routine. The identification of abnormal nuclei is mainly based on features quantifying morphological properties. This quantification, also called morphometry, is exploited first for identifying morphologically suspicious nuclei. Subsequently, the DNA distribution can be exclusively derived from these nuclei, which is a key factor for reaching a high diagnostic sensitivity (see Section 2.2.2). Both morphology and DNA content of nuclei are *measured* from image data, these quantifications are a key component for the reliable outcome of the presented solution. To conclude, a *bimodal cytometric approach* has been established and validated in clinical routine.

This automated bimodal cytometry identified four cancer-cell positive cases which were missed by manual DNA Image Cytometry. Also, it did not make any false-positive diagnoses. The higher diagnostic accuracy of this method is for two reasons. First, automated cytometry detects more abnormal nuclei. On the average, the classifier identified 791 abnormal nuclei in the DNA-aneuploid cases, whereas only 288 nuclei were selected in the manual examination. Therefore, the DNA distribution derived from these nuclei is more representative. Second, the cytological expert might suffer from fatigue and miss relevant objects such as exceeding events.

Table 4.4: Comparison of automated cytometry and the ground truth of the effusions study (left), and the same comparison when the abnormal rate is exploited for classifying cases as suspicious (right).

		Ground truth	
		negative	positive
Auto	negative	60	13
	positive	0	41

		Ground truth	
		negative	positive
Auto	negative	42	0
	suspicious	18	13
	positive	0	41

For automated cytometry, the temporal effort of the expert is reduced to the verification of selected nuclei. This task is usually completed within 5 minutes, therefore the effort is justifiable. Compared to the average number of 30,684 objects per slide which are collected initially, the number of objects which need to be checked is much lower (168 objects). And on average, less than one object per case needs to be reclassified.

4.3.1.5 Morphometry as additional diagnostic marker

DNA-aneuploidy of cancer cells in effusion is not as frequent as for example in oral cancer [10–12]. Therefore it may happen that the changes in the amount of DNA remain below the detection threshold of DNA Image Cytometry. This could be an explanation why the sensitivity of DNA Image Cytometry to detect cancer cells in effusions is low (75 % for effusions compared to 95.5 % for oral cancer, see section 2.2.3). For applying automated cytometry for the screening of body cavity effusions, a higher sensitivity is desired.

Although a malignant transformation associated with chromosomal aneuploidy might not yet be detectable by a change of the nuclear DNA content, it is possible that small changes in texture and morphology are already visible. Since the nucleus classifier is able to discriminate morphologically abnormal from normal mesothelial nuclei, such a transformation leads to an increased number of detected abnormal nuclei. Therefore it was analyzed retrospectively if the rate of abnormal mesothelial nuclei among all mesothelial nuclei can be exploited to detect these transformations. The rational of this approach is to classify all cases as *suspicious* which are DNA-euploid, but have an abnormal rate above 0.75 %. Suspicious cases need to be analyzed in depth.

Using this threshold on the data from the presented study, all 54 cancer-cell positive cases were either classified as DNA-aneuploid or suspicious (see table 4.4). If applied in screening, this means that all cancer-cell positive cases are detected either by DNA-aneuploidy or morphometry. Indeed, also 18 out of 60 cancer-cell negative cases need in-depth analysis. But this still implies that for 72 % of all cases a diagnosis is assigned by automated cytometry alone, either because the case is DNA-aneuploid, or DNA-euploid and unsuspicious—a significant reduction of the

workload in screening.

As a future work, the presented approach should be validated on a separate test set.

4.3.2 Identification of cancer cells in oral brush biopsies

As a second application, manual and automated DNA Image Cytometry were compared for identifying cancer cells in brush biopsies of the oral cavity. Since the clinical application for this study is the same as for serous effusions, the same research questions were considered (see 4.3.1).

4.3.2.1 Material

The cell material for this study had been collected by brush biopsies of visually suspicious lesions (leukoplakias) in the oral cavity of 31 patients. The microscope slides were prepared using liquid based cytology: The head of the brush was broken off into a small glass vial containing containing an alcoholic preservative fluid, which brings cells into suspension. This suspension was centrifuged to remove mucus and other artifacts, and a thin layer of cells was deposited onto a slide. Subsequently, the slides were stained according to Papanicolaou, analyzed due to their morphology, and finally restained according to Feulgen.

4.3.2.2 Experiments

The same protocol for DNA measurements as for the study on effusions was employed (see section 4.3.1.2). The Institute of Cytopathology, Heinrich-Heine University Düsseldorf, Germany, performed the manual examinations. For automated cytometry, the nucleus classifier from section 3.3.5 was employed, which was trained on a gold standard collected from cervical smears.

A slide with 9 c-exceeding events or a DNA-aneuploid stemline is classified as "DNA-aneuploid", otherwise as "DNA-euploid". In case of a discrepancy between manual and automated measurement, a cytopathologist verified the nuclei of both measurements in the nucleus gallery of the software. If the nuclei which lead to the diagnosis "DNA-aneuploid" were found to be morphologically intact and well segmented, this diagnosis was considered as the correct, since cancer cell nuclei had been identified.

4.3.2.3 Results

Manual and automated measurements could be performed for all cases. In 30 out of 31 cases, the diagnosis coincide (see Table 4.5). For one slide, automated cytometry identified a single atypical nucleus with 9 c-exceeding DNA content, whereas no 9 c-exceeding event was found by manual DNA Image Cytometry. Since this nucleus

Table 4.5: Comparison of manual and automated DNA Image Cytometry for the detection of DNA-aneuploidy in brush biopsies from the oral cavity.

Oral cavity		Manual	
		DNA-euploid	DNA-aneuploid
Auto	DNA-euploid	14	0
	DNA-aneuploid	1	16

Table 4.6: Detection performance of the relevant DNA-criteria in brush biopsies from the oral cavity.

	9 c-exceeding events	DNA-aneuploid stemlines
#manual (total)	88	9
#auto (total)	312	7
#cases manual > auto	3	5
#cases manual = auto	18	23
#cases manual < auto	10	3

was verified by a cytopathologist, automated cytometry identified one more DNA-aneuploid case than manual DNA Image Cytometry. Furthermore, it identified 312 9 c-exceeding events, compared to 88 events for manual DNA Image Cytometry (see Table 4.3). During the verification step, on the average 8.47, 504.62, and 3.26 objects classified as abnormal, artifact and defocused respectively needed to be checked. It was necessary to correct three artifacts per slide which had been misclassified as abnormal, and two abnormal nuclei per slides which had been misclassified as artifact. All defocused objects were classified correctly.

4.3.2.4 Discussion and conclusion

The classifier for nuclei from mucosal membranes is, like the classifier for serous effusions, capable of identifying abnormal nuclei in routine application. Thus in the same manner it performs a bimodal cytometric assessment of nuclei. The consequence is a higher detection performance for DNA-aneuploidy, which was also confirmed in this study: for one case, the only nucleus with a 9 c-exceeding DNA content was missed by manual DNA Image Cytometry, which would mean a false-negative diagnosis. This nucleus was, however, identified by automated cytometry. Additionally, automated cytometry found significantly more 9 c-exceeding events.

The sensitivity of manual DNA Image Cytometry is 95.5 % [12]. Since the presented solution is superior in identifying DNA-aneuploidy, it is likely that it can increase this sensitivity further. As a future work, the diagnoses from automated cytometry should be compared to the follow-up of the patients examined.

4.3.3 Grading the malignancy of prostate cancer biopsies

Currently, DNA Image Cytometry is studied as a prognostic tool for the active surveillance of prostate cancer as part of the DNA-ProKo study[2]. The aim of the DNA-ProKo study is to investigate whether the the 2005 modified Gleason score [37] or the grading of malignancy by DNA Image Cytometry are better suited to predict non-progression of a prostate cancer under active surveillance. Prostate cancer has a worse prognosis if its DNA distribution changes from peri-diploid (only one stemline around 2c) to peri-tetraploid (one stemline around 4c) or even to x-ploid (at least one stemline with abnormal DNA content). Considering further information like the age of the patient, usually an urologist might recommend a therapeutic intervention therapy if a peri-tetraploid stemline is found.

The potential of both manual and automated DNA Image Cytometry to identify patients in active surveillance which need therapy will be analyzed when the clinical follow-up study is available—but already now it is possible to compare manual and automated DNA Image Cytometry with the following research questions:

- Which of the methods is superior in finding prognostically relevant DNA-criteria (exceeding events or DNA-aneuploid stemlines)?
- How much effort has to be spend for the verification step? Which verification steps are crucial for a correct grading and which parts can be omitted?

4.3.3.1 Material

The patients included in the DNA-ProKo study were selected according to the inclusion criteria for active surveillance of the German S3 guidelines for diagnosis and therapy of prostate cancer [29]. A thin slice of tissue was cut off from the core needle biopsies taken for Gleason scoring and nuclear specimen were prepared through enzymatic cell separation specifically from cancer cell foci of remaining tissue within paraffin blocks. For the DNA-ProKo study, in total 83 slides were prepared and analyzed with DNA Image Cytometry up to now.

4.3.3.2 Experiments

On these 83 slides, automated DNA Image Cytometry measurements including the verification step as well as manual DNA Image Cytometry were performed. The manual measurements were carried out at the Institute of Pathology, Düren, Germany.

For both types of measurements, the grades as described in the background chapter (see section 2.2.2) were assigned. The grade peri-tetraploid states that cells at 4 c

[2]This abbreviation is derived from the German study name “DNA-Bildzytometrie zum Ausschluss eines Progresses bei Mikrokarzinomen der Prostata - Prospektive validierende Kohorten-Studie mit gutem Referenz-Standard”

Table 4.7: DNA-Grades of prostate cancers found by manual and automated DNA Image Cytometry.

		Manual		
		p.-diploid	p.-tetraploid (EE)	p.-tetraploid (STL)
Auto	p.-diploid	18	1	0
	p.-tetraploid (EE)	15	21	0
	p.-tetraploid (STL)	1	2	0

exist and start to proliferate. This grade is either assigned if an exceeding event above 4.4 c or if a whole DNA-stemline (a significantly high number of nuclei at 4 c and cells at the doubling) are found. To allow a more detailed analysis, it is differentiated during the evaluation whether this grade is assigned due to 4.4 c-exceeding events or a DNA-stemline.

As this study is a prospective one, the final outcome is not yet available, so no gold standard can be set up at the current point in time. In case manual and automated DNA Image Cytometry yielded different grades, a cytopathologist verified the classification of the reference nuclei for the calibration of the system and of the cancer cell nuclei which contributed to the higher grade. If these cells were correct, the higher grade was considered as true.

4.3.3.3 Results

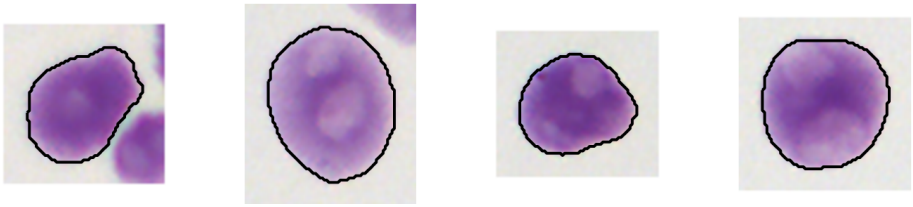
From the 83 slides in this study, 25 could not be evaluated. The reasons for this were: staining too dim (3 slides), not enough reference cells for a calibration according to ESACP guidelines (10 slides), not enough analysis cells (9 slides), or not enough cells for an automated scanning of the slide (3 slides). For the last three cases, however, a manual measurement could be performed.

Table 4.7 shows the grades for the remaining 58 slides. Only peri-diploid and peri-tetraploid grades were found. In 16 cases, automated DNA Image Cytometry found cells which lead to the higher peri-tetraploid instead of the peri-diploid grade. In all cases these 4.4 c-exceeding events and DNA-stemlines were verified and found valid by a cytopathologist. This way it was confirmed that the corresponding cells were present on the slide and that higher grades were correctly assigned. In three cases, a peri-tetraploid DNA-stemline was discovered by automated DNA Image Cytometry, whereas no stemline at all was found by manual DNA Image Cytometry (see Table 4.8).

The verification step requires that abnormal nuclei, artifacts and defocused objects above the exceeding-event threshold are checked. On the average, per case 9.29 cancer cell nuclei and 598.12 artifacts needed to be checked and 3.52 respectively 0.26 objects from the corresponding classes needed to be reclassified. Also 17.66 defocused objects per case needed to be verified, but no reclassification was necessary. In 9 cases, all objects above the exceeding-event threshold were artifacts, thus just

Table 4.8: Detection performance of the clinically relevant events in prostate cancer biopsies.

	4.4 c-exceeding events	Peri-tetraploid stemlines
#manual (total)	100	0
#auto (total)	658	3
#cases manual > auto	5	0
#cases manual = auto	21	55
#cases manual < auto	32	3



(a) 4.29 c **(b)** 4.30 c **(c)** 4.45 c **(d)** 5.26 c

Figure 4.3: Nuclei from prostate cancer biopsies, (a)-(b) with a DNA content below the exceeding-event threshold of 4.4 c, (c)-(d) with a DNA content above this threshold. Both nuclei above and below the exceeding-event threshold have a similar visual appearance, and this might be one explanation why these exceeding events are overlooked more easily in a manual examination.

because a verification was performed the correct grades were assigned. Likewise, the verification of artifacts corrected the grade in one case.

4.3.3.4 Discussion and conclusions

The active surveillance of prostate cancer requires highly reliable prognostic markers (see section 2.1.3). This study proved that the classifier for nuclei from prostate cancer biopsies performs a reliable morphometric selection of abnormal nuclei prior to assessing the grade of malignancy based on their DNA distribution. In all but one case automated cytometry was superior in identifying prognostically relevant DNA-criteria for the grading of malignancy of prostate cancer (see Table 4.7): it was the only method to detect peri-tetraploid DNA-stemlines and discovered 6.5 times more 4.4 c-exceeding events. This led to a higher grade of malignancy for 16 of the 58 cases, and all higher grades were verified by a pathological expert.

The reason for the high discrepancy between manual and automated DNA Image Cytometry is that the detection task in prostate specimens is more difficult than for the other modalities: since low-grade cancers are examined and since the exceeding-event threshold for prostate (4.4 c) is lower than for effusions or oral (9.0 c), the exceeding-event nuclei are visually very similar to those below the exceeding-event

Table 4.9: Time estimates for manual and automated DNA Image Cytometry. Steps requiring user interaction are in bold. CTA=cytotechnician.

Method	Performed by	Time estimate
Manual DNA Image Cytometry	CTA/pathologist	**40-60 min**
Autom. DNA Image Cytometry		
Determining focus points	CTA	**5 min**
Scanning and classification of nuclei	Machine	40-90 min
Verification step	pathologist	**5 min**

threshold (see Figure 4.3). Detecting a relevant object among very similar object is more difficult than in case of a larger visual difference. Thus these exceeding events are overlooked more easily. On the contrary, during automated DNA Image Cytometry, all objects are segmented, classified and compactly presented in a gallery. During the verification step, the expert is aware of analyzing highly relevant nuclei, and less time is needed compared to scanning a full slide. Therefore the expert overlooks less objects and makes less errors due to fatigue.

Verifying the cancer cell nuclei changed the diagnosis in 9 cases, since all exceeding-event objects actually were artifacts. No object classified as defocused needed to be reclassified, thus it can be concluded that this part of the verification step is not needed. Verifying the artifacts changed the diagnosis in 1 out of the 58 cases from a peri-diploid to peri-tetraploid grade, as the only cancer cell nucleus with 4.4 c-exceeding DNA content was classified as artifact. Under-grading in active surveillance can have fatal consequences, thus this verification is mandatory. However, if an exceeding event has been identified already among the original classifications, finding more events will not change the prognosis. Therefore it is the final recommendation to verify all cancer cell nuclei above the exceeding-event threshold, and only in the case no cells are above this, to also verify the artifacts. Consequently in this study only 19 instead of 58 cases the artifacts need verification. This reduced the total number of revised artifacts from 34,691 down to 5,258.

It will be the objective of future work to compare the grades of automated DNA Image Cytometry alone and in combination with the Gleason score to predict progression of prostate cancer. This comparison can be performed when the clinical follow-up is available, which is anticipated for 2015.

4.3.4 Conclusions

For the first time, classifiers which are able to identify abnormal nuclei among all other types of nuclei and artifacts were successfully applied in clinical routine. They were applied for the identification of cancer cells in serous effusions and oral brush biopsies, and for grading the malignancy of prostate cancer. In total they were evaluated on 203 cases. The achievements for practice are:

- A **morphometric preselection of abnormal nuclei:** The classifiers are able to identify abnormal, i.e. dysplastic and malignant, nuclei originating from epithelium or mesothelium among their normal counterparts, other types of cells, and artifacts. The DNA distribution for the final diagnosis can therefore be derived exclusively from abnormal nuclei, which increases the sensitivity of DNA Image Cytometry and is therefore prescribed by international guidelines [8, 9]. The identification is mainly based on a quantification of the morphology of nuclei. This effectively combines two biomarkers, morphology and abnormal DNA content. Both markers are *measured* from digital images of nuclei, therefore a *bimodal diagnostic cytometry* has been established and validated in clinical routine. This was proven for effusions and oral. For prostate cancer biopsies, the task of the classifier is to identify abnormal nuclei from other types of nuclei and artifacts. For these types of specimen, the classifier contributes to a reliable morphometric selection of abnormal nuclei which is directly linked to a more reliable grading.
- A **higher diagnostic accuracy and prognostic validity:** Automated DNA Image Cytometry is superior to its manual version in identifying DNA-aneuploidy (5 cases, effusions and oral) and higher grades of malignancy (16 cases, prostate cancer) for two reasons: First, it identifies more abnormal nuclei—therefore the final diagnosis or prognosis is based on a more representative DNA distribution. Second, it does not suffer from fatigue and therefore misses less clinically relevant nuclei.
- A **reduction of interaction time of a cytopathological expert or cytotechnician:** For automated DNA Image Cytometry, the workload for the experts is reduced to setting focus points prior to scanning and the verification step (see Table 4.9). On the average, 50 reference nuclei, 6 abnormal nuclei, and 195 artifacts need verification, which is usually accomplished within 5 minutes. This allows a more efficient application of DNA Image Cytometry in clinical routine.
- Utilization of **morphometry as an additional diagnostic marker**: In case of a malignant transformation, the fraction of abnormal nuclei is increased, therefore the rate between abnormal mesothelial nuclei and all mesothelial nuclei was exploited to detect cancer-positive cases which would have been missed by DNA Image Cytometry alone. This was proven for effusions.

5 DNA Image Cytometry on Whole Slide Imagers

A bottleneck in the solution presented so far is the time needed for digitizing a slide with a motorized microscope and a conventional camera sensor. Depending on the size of the ROIs for scanning, this might take up to 90 minutes for one slide. Although no user interaction is needed during this time, a speedup is still desirable for increasing the throughput. A recent development in Digital Pathology are Whole Slide Imagers. These Imagers digitize a whole slide within five minutes and they use a slide loader for batch processing of several hundred slides [140, 141].

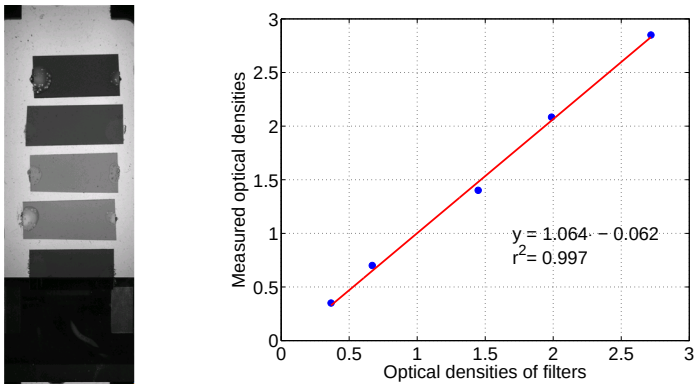
So far only conventional microscopes were used for DNA Image Cytometry. This chapter analyzes the potential of Whole Slide Imagers for DNA Image Cytometry. A prerequisite is the precise measurement of the DNA content of nuclei from digital microscopic images. The final diagnosis or prognosis in DNA Image Cytometry essentially depends on a correct measurement of the DNA content, however several optical effects can distort the result of this measurement. To explore which errors occur and need correction, first the optical characteristics of a Whole Slide Imaging system are analyzed. Subsequently, adequate correction and measurement algorithms are applied and evaluated.

5.1 Optical characterization

The optical characterization of a commercially available Whole Slide Imager presented in this section comprises the measurement of the pixelsize of the system, measurement of the opto-electronic conversion function, and a quantification of the errors due to diffraction and glare.

5.1.1 Materials and methods

A Hamamatsu NanoZoomer 2.0-HT is used as an example for a Whole Slide Imager in this work. The slide loader of the NanoZoomer can load up to 210 slides. This system acquires RGB images by using three time-delay and integration sensors (TDI), one for each color channel. For acquiring microscopical images using a TDI sensor, the slide is moved through the optical path. In synchronization with the velocity of the slide, the charge induced by individual sensor elements is “coupled” from row to



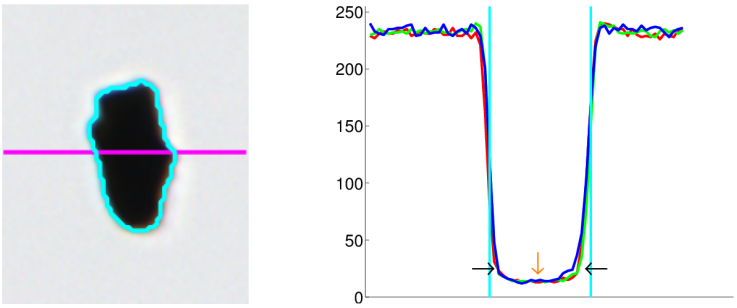
(a) Test slide **(b)** Results

Figure 5.1: Linearity of the opto-electronic conversion functions. (a) Test slide containing filters with predefined optical densities [141]. (b) Optical density of the filters in relationship to the measured optical densities.

row. The signal integrates and thus yields a stronger signal than a conventional line sensor. The TDI sensors of the NanoZoomer have $4,096 \times 64$ pixels, whereas the signal is integrated over 64 rows and scans "lanes" of 4,096 pixels. For producing an image of the whole slide, multiple lanes are acquired and stitched. All scans in this work were performed at a 40× mode[1]. The focus points for the autofocus were set manually, and the image data was stored without compression.

The image processing algorithms developed in this work often depend on measures such as the area, perimeter, or other morphological properties of an object. The size of objects can be derived from the number of pixels, given the relation between sizes of objects and pixels in the image. To establish this relationship, a test slide containing circles with known diameter was scanned. The measurement of the DNA content of a nucleus is based on the sum of the optical densities measured in each pixel of an image of the object. The optical density is computed from the light intensities of incoming and outgoing light, which are in turn derived from the response of the camera sensor (see section 2.2.1). However, these responses may only be used for DNA computation if the opto-electronic conversion functions (OECF)—the relationship between incident light intensity and sensor response—is linear [142]. This however, is not always the case [143, Figure 3]. One approach to measure the OECF is to acquire images of uniformly illuminated patches with known reflectance [144]. Alternatively, the distance between a uniform light source and the sensor can be varied. With larger distance, the light intensity impinging on the sensor is reduced, and can be derived from a physical model [143, 145–147]. However, for WSI scanners, the sensor is mounted stationary in the system. The only possibility to acquire

[1]In the strict sense, terms from conventional microscopy cannot be transferred one by one to Whole Slide Imagers, as the image formation process is different. The slides were scanned with a resolution which roughly corresponds to images acquired at 40× magnification in conventional microscopy, which is therefore denoted as 40× mode.



(a) Coal particle **(b)** Intensity profile

Figure 5.2: Errors due to diffraction and glare. (a) Image of a coal particle, acquired with a WSI system. The segmentation mask is delineated by a cyan contour. (b) Intensity plot along the purple line in (a), for the red, green and blue color channel. The boundary of the segmentation is depicted by the vertical cyan lines. As the coal particle is opaque, the light intensity should be zero. Diffraction brightens up the boundary regions (black arrows), and glare the interior of the object (orange arrow).

images is to bring slides into the optical path. Therefore, optical density filters with known attenuation were fixed on a glass slide (see Figure 5.1(a)). This slide was scanned, and the optical densities of the filters were compared with those derived from the scan.

Two sources of error can brighten up nuclei and thus lead to an underestimation of their DNA content: diffraction, the "bending of light rays" at the boundary of objects, and glare, the reflection of light at surfaces of the optical system. To evaluate the influences of diffraction and glare, a slide containing small carbon particles was scanned (see Figure 5.2(a)). These particles are opaque, thus analyzing the light intensities at the boundary of these objects allows conclusions about the influence of diffraction. For a quantification of glare, the light intensity at the center of the particles was measured. The particles were first segmented by thresholding. The segmentation masks were shrunk down to 30 % of the area of the original segmentation mask by iteratively applying morphological erosion, which resulted in a "central region" which is not affected by diffraction at the boundary of the particle. The mean intensity of this central region was divided by the mean intensity of the incident light. This fraction reflects how much of the incoming light is distributed to other regions due to glare. The measurement was performed on the green channel and restricted to carbon particles which approximately have the size and shape of nuclei and which were acquired in focus [141].

5.1.2 Results

A circle of 600 µm diameter is 2669 pixels wide in the scanned data. Thus for the Hamamatsu NanoZoomer 2.0-HT in 40× mode, one pixel in the image corresponds

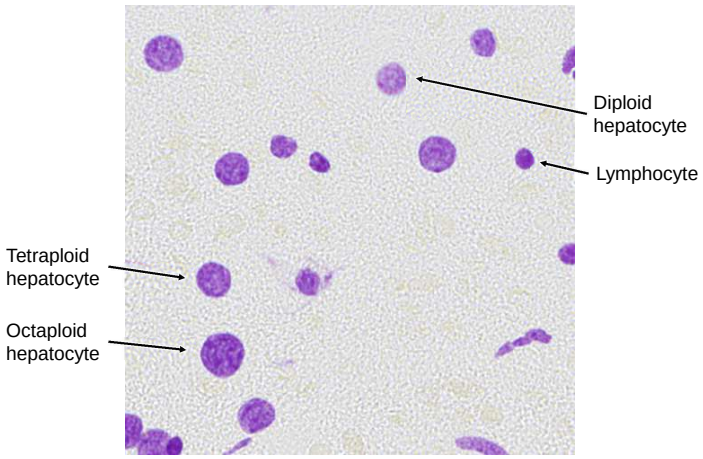


Figure 5.3: Rat liver nuclei acquired with a Hamamatsu NanoZoomer 2.0-HT.

to $0.2248 \times 0.2248\,\mu m^2$ on the slide. Figure 5.1(b) depicts the optical densities of the filters related to the optical densities measured from the images. They follow a linear relationship (coefficient of determination $r^2 = 0.997$). The light intensity profile of a carbon particle is shown in Figure 5.2(b). Clearly diffraction occurs at the boundaries of an object. Glare was quantified on 118 coal particles, it brightened up the interior of these objects by 6.44 % of the incident light.

5.1.3 Conclusions

Since the OECF of the sensor is linear, there is no need to apply a correction for non-linearities of the sensor, but diffraction and glare need correction.

5.2 DNA measurements

Algorithms for the correction of diffraction and glare in DNA Image Cytometry have been presented, among others, by Haroske et al. [49] and Würflinger et al. [50]. So far, these algorithms were only used in combination with conventional microscopes—this section evaluates their accuracy on Whole Slide Imager data. The algorithms are tested according to the quality control guidelines for DNA Image Cytometry [51], established by the European Society for Analytical Cellular Pathology (ESACP).

5.2.1 Material and methods

For evaluating the accuracy of the DNA algorithms, the DNA content of healthy rat liver cells (hepatocytes) is measured, using lymphocytes as reference nuclei. From hepatocytes it is known that they occur as diploid, tetraploid or octaploid nuclei, thus the measured DNA values of hepatocytes should be at 2.0 c, 4.0 c or 8.0 c.

Table 5.1: Accuracy of the DNA measurements on Whole Slide Imager data. The first row of the tables gives the quality level prescribed by the ESACP report, the second and third row the mean statistics without and with correction. The number of cases above the allowable limit are given brackets.

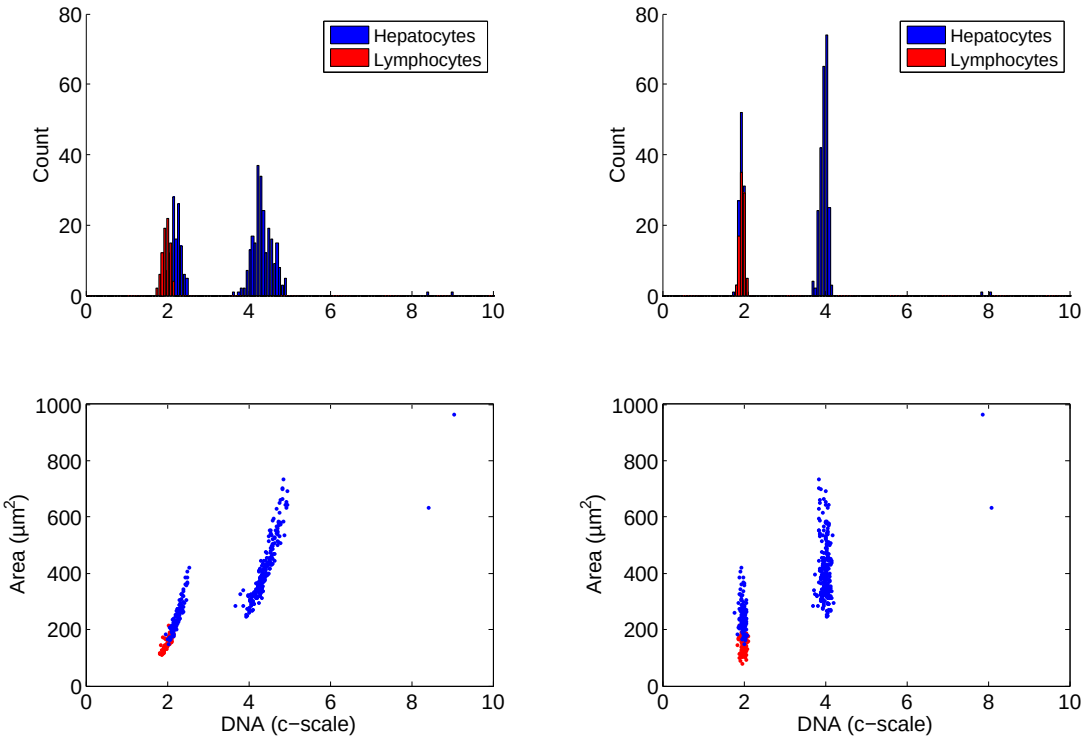
	CV-ref.	|CC|-ref	CV-dipl.	CV-tetrapl.	CV-octapl.
ESACP Guidelines	< 5	< 0.4	< 5	< 5	< 5
No correction	4.90 (6)	0.81 (16)	3.85 (3)	3.53 (0)	2.46 (4)
With correction	3.54 (0)	0.09 (0)	3.28 (0)	2.90 (0)	1.81 (0)

According to the ESACP guidelines, the coefficient of variation (CV) of the DNA values of the reference nuclei (CV-ref) and the hepatocytes at each ploidy peak (CV-diploid, CV-tetraploid, CV-octaploid) must not exceed 5 %. Additionally, the coefficient of correlation r between the DNA values of the reference cells and their area (|CC|-ref) must fulfill $|r| < 0.4$ [51]. For this work, two liver slides were prepared by cutting a liver of a healthy rat in half, and pressing the cut surface on a glass slide. The cells were then stained according to Feulgen and scanned with a Hamamatsu NanoZoomer. In total 16 cell collections for the ESACP rat liver test were collected from different parts of the slide. Figure 5.3 shows rat liver nuclei acquired with a Hamamatsu NanoZoomer 2.0-HT.

For measuring the DNA content of the collected nuclei, the basic measurement principle presented in section 2.2.1 was followed. Additionally, corrections for objects in the background, diffraction and glare were applied. Briefly, objects in the background of a nucleus also attenuate the incoming light. This attenuation is estimated from a reference ring around the nucleus and compensated for. The diffraction correction according to Haroske et al. [49] iteratively increases the integrated optical density of the region affected by diffraction until it is close to the mean IOD of the nucleus. The adaptive glare correction by Würflinger et al. [50] models glare as a uniform redistribution of light, and exploits this model to find a glare estimate which minimizes the correlation between area and IOD values of reference nuclei—in theory this correlation should be 0. A detailed description of these algorithms can be found in appendix A.1

5.2.2 Results and discussion

The ESACP statistics of the rat liver measurement before and after correction are given in Table 5.1. Additionally, this table presents the number of cases which fail the ESACP test; failed tests are those where at least one statistic exceeds the permitted limit. Clearly the correction algorithms improve the accuracy: The average CV and correlation values, which ideally should be zero, are significantly lower when applying the correction. Without correction, the ESACP test failed for all cases, because the coefficient of correlation between DNA and area values of the reference nuclei is too



(a) Without correction **(b)** With correction

Figure 5.4: DNA histogram and scatterplots (DNA vs. area) of the same rat liver (a) without correction, (b) with correction.

high. Furthermore, the CV of the DNA values of diploid and octaploid hepatocyes as well as the reference nuclei are too high for some cases. With correction however, all cases pass the test. Figure 5.4 shows the DNA measurement results (DNA histograms and scatterplots) of nuclei from a rat liver before and after correction. As required, after correction the DNA values of diploid, tetraploid and octaploid hepatocytes are around 2.0 c, 4.0 c and 8.0 c after correction, and the variation of the DNA values in each group is low.

The accuracy of the DNA values measured from Whole Slide Imager data was close to that of conventional microscopes. Berger-Fröhlich performed 40 rat liver tests with a Motic BA410 microscope, a MotiCam Pro 285A camera, and the same correction algorithms as in this work. The mean CV-ref was 2.77 (instead of 3.54) and the mean |CC|-ref was 0.05 (instead of 0.09) [148].

5.2.3 Conclusions and future work

If the diffraction, glare, and the attenuation of light by objects in the background of nuclei are corrected for, Whole Slide Imagers can be used for accurately measuring the DNA content of Feulgen-stained nuclei. As a future work, the segmentation and classification algorithms from chapter 3 should be adapted to Whole Slide Imager data by collecting new gold standards. This will allow the automated scanning and identification of relevant nuclei using Whole Slide Imagers and thus yield a significant increase of the throughput of automated DNA Image Cytometry.

6 Conclusions and future work

The motivation for this thesis is the fact that DNA Image Cytometry has the potential to improve cancer diagnostics and prognostics, but when performed manually it is too time-consuming for a broad application in practice. Therefore, this thesis presented image processing algorithms for automatically accomplishing time-consuming steps of DNA Image Cytometry. These algorithms were integrated into a software package and a workflow and compared to manual DNA Image Cytometry in three applications.

The key result of this thesis is that the developed solution reduces the interaction time of a cytopathological expert from 40 to 5 minutes per slide. At the same time, it has a higher diagnostic accuracy or prognostic validity than manual DNA Image Cytometry.

Dedicated segmentation and classification algorithms were developed for automatically identifying abnormal nuclei in images acquired by a microscope with motorized stage. The segmentation pipeline of this work segments less irrelevant objects and yields a more precise contour than the algorithm commonly used in manual DNA Image Cytometry. The precise contour is achieved by refining the contour of imprecise segmentations and splitting touching nuclei which have been segmented as one object. To reduce the computational burden, the developed pipeline automatically detects the need for refinement or splitting prior to applying the complex correction algorithms. The segmented objects are then classified by a nucleus classifier, which are capable of identifying morphologically abnormal nuclei and cancer cell nuclei among artifacts and all other classes of nuclei present on a slide. Such classifiers were developed for three types of specimens: effusions, prostate cancers and mucosal membranes. For these types of specimen they detect 95.24 %, 94.70 %, and 93.85 % of all abnormal nuclei, and 94.04 %, 91.59 %, and 96.28 % of all objects classified as abnormal are correct. The decision of the classifiers is mainly based on features describing the morphology of objects and large gold standards which have been annotated by a cytopathologist with more than 35 years of experience. Consequently the classifiers perform a morphometric identification of abnormal nuclei in accordance with an expert.

To facilitate the translation of the developed algorithms into practice, they were integrated into a software package with appropriate user interface and a clinical workflow. The workflow beneficially divides the labor between machine and expert. The machine automatically scans a slide and classifies nuclei. Subsequently, the expert verifies the classification of exceeding events, objects with a DNA content

above a threshold. These events are critical since the misclassification of one exceeding event can already change the diagnostic or prognostic result. The effort for the expert for verifying the objects above the exceeding event threshold was reasonable, usually less than 260 objects needed to be checked per slide. Since these objects are compactly presented by the user interface of the software, this is possible in less than 5 minutes.

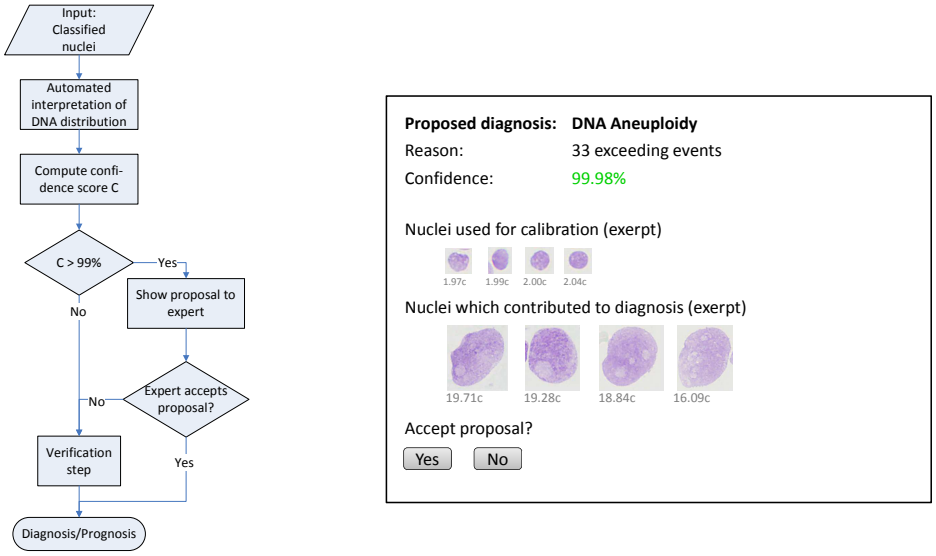
The developed solution was compared to manual DNA Image Cytometry in three applications: the identification of cancer cells in nuclei from serous effusions and from brush biopsies of the oral cavity, and grading the malignancy of prostate cancer biopsies. The evaluation was performed on 203 cases. Since the nucleus classifiers are capable of identifying abnormal nuclei, the DNA distribution can be derived exclusively from these nuclei. This effectively combines a quantification of the morphology and the DNA content into bimodal cytometry. For the identification of cancer cells, automated DNA Image Cytometry detected 5 cancer-cell positive cases which were missed by manual DNA Image Cytometry, and it assigned higher grades of malignancy in 16 cases. The reason for the higher diagnostic accuracy and prognostic validity is that this is that automated DNA Image Cytometry identified more abnormal nuclei, thus the DNA distribution for the final diagnosis or prognosis is more representative than those of manual DNA Image Cytometry. Furthermore, the machine finds objects which the expert might miss due to fatigue. Finally, the morphometric identification of abnormal nuclei could be exploited as additional diagnostic marker. For serous effusions, DNA-aneuploidy of cancer cell nuclei is not as strong as in many other types of specimen. In case of a malignant transformation, the fraction of abnormal nuclei is increased. This allows identifying cancerous cases where the changes in the amount of DNA were below the detection threshold of DNA Image Cytometry, but a change in morphology already indicated a malignant transformation.

A proof-of-concept showed that it is possible to perform DNA Image Cytometry using Whole Slide Imagers. Compared to conventional microscopes with motorized stage, these systems are able to scan a slide in 5 instead of 60 minutes. Therefore this proof-of-concept opens a new perspective for increasing the throughput of automated DNA Image Cytometry.

6.1 Future work

One future research goal is to develop classifiers for further relevant applications, such as the detection of lung cancer from sputum, of bladder cancer from urine, and of thyroid cancer from aspiration biopsies. The process of developing such classifiers presented in section 3.3.4 can also be followed for the new applications.

The solution presented in this work enable the automated processing of individual slides. Two steps during the image acquisition still require user interaction, placing a slide on the microscope stage and identifying a set of representative focus points



(a) **(b)**

Figure 6.1: (a) Flow chart for processing unambiguous cases (b) Example of an automatically generated proposal, including the proposal, a score reflecting the systems confidence in it, and a few representative nuclei.

for the focus model. By automating these steps, the batch scanning of a set of slides is enabled: a slide loader can be used to automatically place slides on the microscope stage. The algorithms from Bell et al. [149] for conventional microscopy or Lahrmann et al. [150] for Whole Slide Imagers can automatically digitize cytology slides without user interaction. In combination with the algorithms for the detection of ROIs developed in this work (section 3.4), scanning overnight will be possible, which will considerably increase the utilization of the device.

Another research goal is to further reduce the expert's effort of verifying the classification of exceeding events after scanning. Besides reducing the rate of misclassifications for these objects, statistical modeling and testing theory can be used to safely identify unambiguous cases for which the verification step might be skipped. As an example, in one case from the effusion study the pattern recognition system initially classified 122 objects above the exceeding event threshold into the class "abnormal". Only if all of these objects were artifacts instead of abnormal nuclei, this would change the diagnosis. The probability for this is, however, very low. After the automated scanning and classification of objects, the system would:

- use algorithms for the automated interpretation of DNA distributions to propose a diagnosis or prognosis [9],
- compute a confidence score based on statistical modeling of the factors which contributed to the proposal,
- select a few representative nuclei.

In case of a high confidence, the proposal and the selected nuclei are presented to the expert (see Figure 6.1(a)). The expert can either accept the proposal, or perform a manual diagnosis or prognosis. The benefit of this strategy is twofold. First, it allows a quick decision for unambiguous cases. Second, the computer's confidence and rationale for the decision are presented, which is essential for the credibility in the system's proposal.

The results of this work motivate the development of a system for high throughput DNA Image Cytometry—the vision for such a system is: slides will be digitized using a Whole Slide Imager, which will be equipped with a slide loader and an autofocus routine suited for cytology slides. For speed-up, the scanning will be limited to the ROIs detected by the algorithm from section 3.4 from a macro image of the slide. Clinically relevant nuclei will be automatically detected by means of the segmentation and classification approaches presented in this work (see sections 3.2 and 3.3). After the automated batch processing of a set of slides, all cases will presented to the expert for verification, whereas a rapid diagnosis or prognosis for unambiguous cases will be performed based on the statistical modeling and confidence score presented in the previous paragraph. The expert will be supported in the automated generation of a report as the end result of a DNA Image Cytometry examination. With this solution, one expert may be able to perform more than 100 bimodal cytometric diagnoses or prognoses per day.

A Appendix

A.1 Measuring the DNA content of nuclei

This section describes the algorithms for measuring the DNA content from digital images of nuclei which have been stained according to Feulgen. First, the basic measurement principle is described. Next, correction procedures for objects in the background of nuclei, diffraction, and glare are presented.

In Feulgen-stained nuclei the total amount of the stain in the nucleus is proportional to its DNA content. The attenuation of light passing through a nucleus, in turn, is stronger for higher concentrations, which is described by Lambert-Beer's law:

$$I_1 = I_0 \cdot \mathrm{e}^{-c \cdot d \cdot \epsilon_\lambda}, \tag{A.1}$$

where I_0 is the incoming light, I_1 the transmitted light, c and d are the concentration and thickness of the matter, and ϵ_λ is a wavelength dependent constant. In practice, I_0 is obtained as the pixel value of a digital image at an empty position, and the transmitted light intensity I_1 the pixel values inside a segmented nucleus. Due to the 2D image formation, the thickness d cannot be considered, instead the optical density is defined as

$$\mathrm{OD} := -\ln\left(\frac{I_1}{I_0}\right) = c \cdot d \cdot \epsilon_\lambda. \tag{A.2}$$

It is linearly related to the concentration of the stain. By summing up optical density values (equation A.2) of all pixels of a nucleus, an overall measure for the attenuated of light by the nucleus is computed (integrated optical density, IOD). The basic measurement principle is to measure the attenuation of light by each nucleus, and then use a calibration with reference nuclei to transform it to a DNA value. The attenuation of the pararosanilin dye, the dye used for staining the nuclei during the Feulgen reaction, is strongest in the green channel. For numerical stability this channel is used for the IOD computations. The IOD values have to be converted to DNA values. This is done by dividing the IOD values by the median IOD value of the reference nuclei, which have normal DNA content. The median instead of the mean is used for the calibration because it is more robust against outliers. However, instead of measuring the true IOD, certain sources of errors during the microscopic image artificially darken or brighten up the pixels. Therefore, corrections for non-transparent objects in the background, for diffraction, and for glare are applied. These corrections will be described in the following.

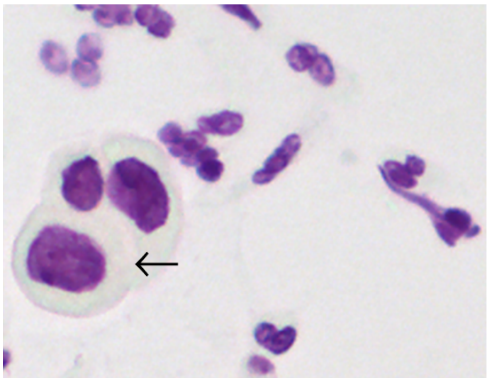


Figure A.1: Motivation for local background correction. For some cells, like the epithelial cells on the left, the unstained cytoplasm is also visible (black arrow). Therefore, it is beneficial to correct for this for every individual object.

The incident light is attenuated not only by the dye of the nucleus, but also by objects in the background. The cells are located between two glass slides and fixative, which are not perfectly transparent. Also diffraction at unstained cell structures like the cytoplasm occurs (see Figure A.1). Since the attenuation by the background can change from cell to cell, a local background correction is applied. To this end, a small reference ring outside each nucleus is computed by morphological image processing. The mean optical density of this ring is subtracted from the OD values.

Diffraction brightens up the pixel values at the boundary of an object and thus decreases the IOD (see also Figure 5.2). To compensate for this, the diffraction correction by Haroske et al. is applied [49]. The area affected by diffraction, a small region at the interior of the boundary, is extracted by morphological operations (diffraction region). The IOD is then iteratively increased by adjusting the IOD of the diffraction region to fit to the mean IOD of the nucleus. As this adjustment, in turn, changes the mean IOD, this procedure is iteratively repeated until the change is smaller than 5 %.

Glare brightens up opaque objects to about 6 % of the incident light. One possibility would be to subtract this light intensity from all pixel values (fixed glare correction). However, the influence of glare depends on the illumination settings, alignment of the optical path, and further factors. Therefore the adaptive glare correction by Würflinger et al. [50] is used. Glare is modeled as the uniform redistribution of light, according to the Goldstein model [151]. This means that all pixel intensity values in the image are artificially brightened up by the same intensity value k. Thus glare is straightforwardly corrected by subtracting k prior to the computation of the optical densities. Then, Würflingers correction exploits the fact that reference nuclei all have a standard DNA content, but different sizes due to their biological variability. Thus ideally there is is *no* correlation between the IOD and area values. Different values for k are tested; the one which minimizes the correlation between

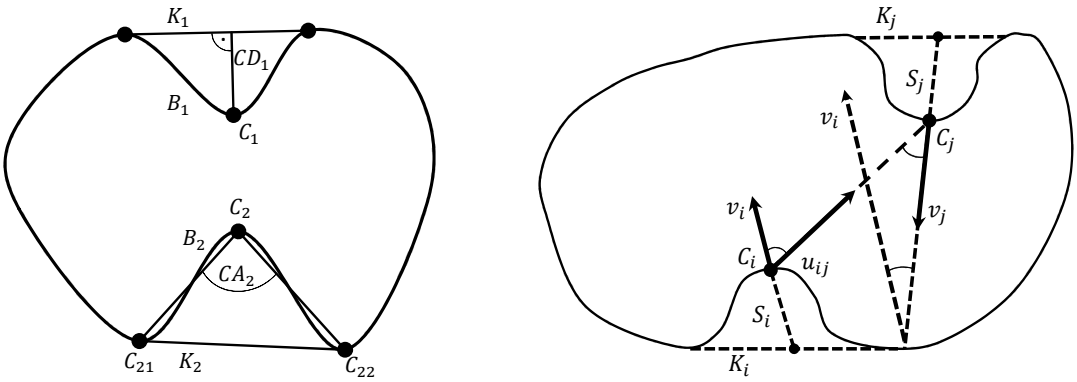


Figure A.2: (a) Computation of the concavity depth and concavity angle, (b) computation of the concavity-concavity alignment and concavity-line alignment.

IOD and the area of the reference nuclei is then used as adaptive glare estimate $\widetilde{k}$. Finally, the IODs of all objects are computed again, correcting for glare by means of $\widetilde{k}$.

A.2 Quantification of concavities according to Kumar et al.

In this work several features used by the segmentation algorithm or the nucleus classifier are based on the concavity analysis framework of Kumar et al. [105], which is summarized in this section. Following the definitions of the paper, for a concavity with index i, the boundary arc is denoted as B_i, the chord of the corresponding hull segment as K_i, and the concavity pixel C_i is the point on B_i with the largest distance to K_i (see also Figure A.2(a)). Several metrics for quantifying the size of a concavity, the alignment of a pair of concavities, and a figure of merit for split lines are presented:

- **Concavity Depth:** As measure for the depth of an individual concavity, the distance from C_i to K_i is used.
- **Concavity Angle:** The angle CA_i of a concavity is the angle between the two lines pointing from C_i to the endpoints C_{i1} and C_{i2} of K_i.
- **Concavity-Concavity Alignment:** This feature quantifies if two concavities are oppositely aligned. The orientation of a concavity is given by the vector v_i, starting from the midpoint of the convex hull segment K_i through the concavity pixel C_i (see Figure A.2(b)). The Concavity-Concavity alignment (CCAlignment) of two concavities is computed as $\pi - \cos^{-1}(v_i \cdot v_j)$.

- **Concavity-Line Alignment:** Another alignment measure also considers the line u_{ij} connecting two concavities. The Concavity-Line Alignment (CLAlignment) is the maximum angle between v_i and u_{ij} or v_j and u_{ij}.
- **Measure Of Split:** If the CCAlignment and CLAlignment of two concavities with index are below certain thresholds (CCAlignment $< 105°$ and CCAlignment $< 70°$), they are considered as well aligned and are thus candidates for split lines of a clump. The measure of split between two concavities with index i and j is computed based on the depth of the concavities CD_i and CD_j, and the distance $d\,(C_i, C_j)$ of the two corresponding concavity pixels as

$$\chi = \frac{c_1\mathrm{CD}_i \cdot c_1\mathrm{CD}_j + c_2}{d\,(C_i, C_j) + c_1\mathrm{CD}_i \cdot c_1\mathrm{CD}_j + c_2.}$$

A.3 Features for the defocus classifiers

This section presents the four features ADContour, Laplacian, Rbiii and VarInterior used for pattern recognition systems for distinguishing between focused and defocused objects (see section 3.3.2). All features for the defocus classifier are based on a gray image of an object, which is computed as weighted combinations of the color channels ($R = 0.299$, $G = 0.587$ and $B = 0.114$).

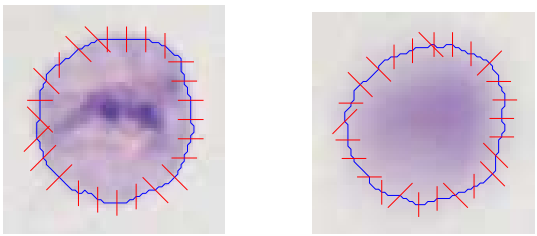
A.3.1 Absolute difference along contour normals

Motivation and description: Focused nuclei have a sharper transition from background to nucleus. This is quantified based on the absolute difference of neighboring pixels along the contour normals. An example is shown in A.3(b).

Algorithm:

1. For each contour pixel c_i find the vector c_n orthogonal to c_i and its predecessor by solving $\langle c_i - c_{i-1}, c_n\rangle = 0$.
2. Consider the contour normal with a length $2 \cdot n_l$, pointing by n_l both to the interior and exterior of the nucleus. For each pair p_j and p_{j+1} of neighboring pixels within the contour normal, compute the absolute difference $d(c_i, p_j, p_{j+1}) = |p_j - p_{j+1}|$.
3. Store the maximum value $d_{max}(c_i) = \max_j(c_i, p_j, p_{j+1})$
4. `ADNormal` $:= \mathcal{Q}_p\{d_{max}(c_i)|1 \le i \le N\}$, where $\mathcal{Q}_p$ is the p% quantile of all $d_{max}(c_i)$.

Parameter	**Range**
Normal length n_l	$0.18 : 0.18 : 1.8\mu$m
Quantile p	$0.25 : 0.25 : 1$



(a) Focused nucleus

(b) Defocused nucleus

Figure A.3: Contour normals for a focused and defocused nucleus. For clarity reasons, only every sixth normal is shown. The `ADNormal` feature is 23 for the focused nucleus and 9 for the defocused nucleus, using $n_l = 0.72\mu$m and $p = 1$.

A.3.2 Laplacian

Motivation and description: Defocused objects have less variation in the derivatives of pixel intensities. This is quantified using the Laplace filter, which is the sum of the second derivatives in x- and y-direction.

Algorithm:

1. Resize the segmentation mask M to a percentage p of the original area to get a mask M_s, using morphological erosion ($p < 100$) or dilation ($p > 100$).
2. Filter the image within M_a using the laplacian filter

$$\frac{1}{6} \cdot \begin{bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix}$$

3. `Laplacian`:=Mean absolute filter response in M_s.

Parameter	Range
Percentage p for resizing the mask	$5 : 5 : 150\%$

A.3.3 Relationship boundary intensity and interior intensity

Motivation and description: Defocused nuclei have higher intensity values close to the nucleus boundary compared to focused nuclei. This is quantified by computing the fraction of the mean intensity around the boundary region and the mean intensity in the interior of the nucleus. An example is shown in A.4.

Algorithm:

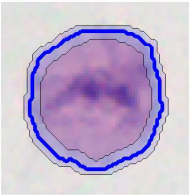


Figure A.4: Feature Rbiii with $p = 0.7$ and `extend_outside=true`

1. Shrink down the segmentation mask M using morphological erosion, down to a percentage p of the original area of M to get the interior mask M_i. The boundary mask is defined as $M_b = M \setminus M_i$. If the parameter `extend_outside` is `true`, M is first dilated up to an area of $2-p$ the original area to get a mask M_d. Then, M_b is computed as $M_d \setminus M_s$.
2. Compute the mean intensities m_b in M_b and m_i in M_i.
3. `Rbiii` $:= \frac{m_b}{m_i}$

Parameter	**Range**
Percentage p for resizing the mask	$5 : 5 : 95\%$
`extend_outside`	`{true, false}`

A.3.4 Variation in the interior

Motivation and description: The variation of intensity values in the interior of a nucleus is smaller for defocused objects. This is quantified by considering pixel intensities in the interior of the nucleus and computing a measure for variance. To gain robustness against changes in the stain intensity, the coefficient of variation is used for this purpose instead of the conventional variation.

Algorithm:

1. Shrink down the segmentation mask M using morphological erosion, down to a percentage p of the original area of M. This yields a new mask M_s.
2. Compute the mean intensity M_i and standard deviation σ_i.
3. `VarInterior` $:= \frac{M_i}{\sigma_i}$

Parameter	**Range**
Percentage p for shrinking the mask	$5 : 5 : 95\%$

A.4 Features for the nucleus classifiers

A.4.1 State-of-the-art features

Feature	Description
Area	Area of the segmentation mask
Perimeter	Perimeter of the segmentation mask in
MinRadius	Smallest distance from the centroid to the contour
MaxRadius	Largest distance from the centroid to the contour
MeanRadius	Average distance from the centroid to the contour
VarianceRadius	Variance of the distances from the centroid to the contour
Sphericity	Fraction of Radius Min and Radius Max
Eccentricity	Ratio of the major to minor axis of the best fit ellipse
Inertia	Squared distance of all object's pixels to the centroid, normalized by squared area
Compactness	$P^2/(4\pi \cdot A)$, where P is the perimeter and A the area
BendingEnergy	Energy needed to bend the contour to its current shape
Background	Average intensity of all pixel values larger than 150 in a small reference region around the nucleus
MeanLuminance	Average intensity of the gray image
VarLuminance	Variance of intensity values of the gray image
MinFilter	Minimum response of a square filter on the gray image
Entropy	Entropy of gray image
ClusterShade	Contrast between dark clumps and light background
ClusterProminence	"Darkness" of clusters

A.5 Additional statistics for the nucleus classifiers

A.5.0.1 Classification of nuclei from effusions

Table A.1: Number of nuclei and correct classification rate (CCR) for the individual slides of the effusion gold standard.

Slide Index	Number of nuclei	CCR
1	5664	87.50 %
2	6268	87.30 %
3	6025	90.08 %
4	5563	94.61 %
5	5673	85.99 %
6	5899	85.13 %
7	6859	86.70 %
8	6233	91.82 %
9	6190	84.41 %

Table A.2: Classification performance for artifacts and abnormal nuclei above the 9.0 c-exceeding-event threshold, and abnormal nuclei belonging to a DNA-aneuploid stemline (effusion classifier).

		Ground truth		
		EE artifact	EE abnormal	STL abnormal
classification	Artifact	1420	24	136
	Abnormal	23	185	5853
	Macrophage	0	0	8
	Normal	0	0	53
	Lymphocyte	0	0	1
	Granulocyte	0	0	0
	Total	1443	209	6051
	Error (%)	1.59	11.48	3.27

A.5.0.2 Classification of nuclei from prostate cancer biopsies

Table A.3: Number of nuclei and correct classification rate (CCR) for the individual slides for nuclei of prostate cancer biopsies.

Slide Index	Number of nuclei	CCR
1	5609	91.21 %
2	5160	91.32 %
3	5371	91.77 %
4	5887	90.49 %
5	5407	96.23 %
6	5321	93.82 %
7	4716	89.91 %
8	6120	91.54 %
9	4391	88.29 %

Table A.4: Classification performance of the prostate classifier, exclusively for artifacts and abnormal nuclei above the 4.4 c-exceeding-event threshold exceeding event threshold, and abnormal nuclei belonging to a DNA-aneuploid stemline.

		Ground truth		
		EE artifact	EE cancer cell	STL cancer cell
classification	Artifact	3260	36	0
	Cancer cell	18	284	43
	Fibroblast	1	0	0
	Lymphocyte	0	1	0
	Granulocyte	0	0	0
	Total	3279	321	43
	Error (%)	0.58	11.53	0

A.5.0.3 Classification of nuclei from mucosal membranes

Table A.5: Number of nuclei and correct classification rate (CCR) for the individual slides of the cervix gold standard.

Slide Index	Number of nuclei	CCR
1	4127	94.33 %
2	3658	91.09 %
3	2478	89.35 %
4	5688	90.58 %
5	2963	92.67 %
6	2932	91.13 %
7	3423	90.73 %
8	3565	92.53 %
9	3221	89.69 %

Table A.6: Classification performance for artifacts and abnormal nuclei above the 9.0 c-exceeding-event threshold, and abnormal nuclei belonging to a DNA-aneuploid stemline (classifier for mucosal membranes).

		Ground truth		
		EE artifact	EE abnormal	STL abnormal
classification	Artifact	947	11	34
	Abnormal	5	382	719
	Normal	0	0	20
	Lymphocyte	0	0	17
	Granulocyte	0	0	2
	Total	952	393	792
	Error (%)	0.52	2.80	9.21

Bibliography

[1] C. S. Wellen, *Prognostische Relevanz von Staging sowie histologischem und DNA-zytometrischem Malignitäts-Grading des Plattenepithelkarzinoms des Kopf-Hals-Bereiches*. PhD thesis, RWTH Aachen, 1996.

[2] A. Böcking, C. Sproll, N. Stöcklein, C. Naujoks, R. Depprich, N. R. Kübler, and J. Handschel, "Role of brush biopsy and DNA cytometry for prevention, diagnosis, therapy, and followup care of oral cancer," *Journal of Oncology*, 2011.

[3] S. Zunt, "Transepithelial brush biopsy: an adjunctive diagnostic procedure," *Journal of the Indiana Dental Association*, vol. 80, no. 2, p. 6, 2001.

[4] M. H. Stoler and M. Schiffman, "Interobserver reproducibility of cervical cytologic and histologic interpretations," *JAMA: The Journal of the American Medical Association*, vol. 285, no. 11, pp. 1500–1505, 2001.

[5] T. W. Remmerbach, H. Weidenbach, N. Pomjanski, K. Knops, S. Mathes, A. Hemprich, and A. Böcking, "Cytologic and DNA-cytometric early diagnosis of oral cancer," *Analytical Cellular Pathology*, vol. 22, no. 4, pp. 211–221, 2001.

[6] H. Rajagopalan and C. Lengauer, "Aneuploidy and cancer," *Nature*, vol. 432, no. 7015, pp. 338–341, 2004.

[7] G. J. Kops, B. A. Weaver, and D. W. Cleveland, "On the road to cancer: aneuploidy and the mitotic checkpoint," *Nature Reviews Cancer*, vol. 5, no. 10, pp. 773–785, 2005.

[8] A. Böcking, F. Giroud, and A. Reith, "Consensus report of the ESACP task force on standardization of diagnostic DNA image cytometry," *Analytical Cellular Pathology*, vol. 8, no. 1, pp. 67–74, 1995.

[9] G. Haroske, J. Baak, H. Danielsen, F. Giroud, A. Gschwendtner, M. Oberholzer, A. Reith, P. Spieler, and A. Böcking, "Fourth updated ESACP consensus report on diagnostic DNA image cytometry," *Analytical Cellular Pathology*, vol. 23, no. 2, pp. 89–95, 2001.

[10] H. Motherby, B. Nadjari, P. Friegel, J. Kohaus, U. Ramp, and A. Böcking, "Diagnostic accuracy of effusion cytology," *Diagnostic Cytopathology*, vol. 20, no. 6, pp. 350–357, 1999.

[11] H. Motherby, N. Pomjanski, M. Kube, A. Boros, T. Heiden, B. Tribukait, and A. Böcking, "Diagnostic DNA-flow-vs.-image-cytometry in effusion cytology," *Analytical Cellular Pathology*, vol. 24, no. 1, pp. 5–15, 2002.

[12] T. Remmerbach, S. Mathes, H. Weidenbach, A. Hemprich, and A. Böcking, "Noninvasive brush biopsy as an innovative tool for early detection of oral carcinomas," *Mund-, Kiefer-und Gesichtschirurgie: MKG*, vol. 8, no. 4, p. 229, 2004.

[13] H. J. Grote, H. V. Nguyen, A. G. Leick, and A. Böcking, "Identification of progressive cervical epithelial cell abnormalities using DNA image cytometry," *Cancer Cytopathology*, vol. 102, no. 6, pp. 373–379, 2004.

[14] F. Meyer, "Iterative image transformations for an automatic screening of cervical smears," *Journal of Histochemistry & Cytochemistry*, vol. 27, no. 1, pp. 128–35, 1979.

[15] J. Ploem, N. Verwoerd, J. Bonnet, and G. Koper, "An automated microscope for quantitative cytology combining television image analysis and stage scanning microphotometry," *Journal of Histochemistry & Cytochemistry*, vol. 27, no. 1, pp. 136–143, 1979.

[16] C. Cornelisse, A. Driel-Kulker, F. Meyer, and J. Ploem, "Automated recognition of atypical nuclei in breast cancer cytology specimens by iterative image transformations," *Journal of Microscopy*, vol. 137, no. 1, pp. 101–110, 1985.

[17] H. Tanke, J. Brussee, A. van Driel-Kulker, M. van der Burg, C. Schelvis-Knepfle, U. Jonas, and J. Ploem, "Results of the automated analysis of 328 bladder specimens using the leyden television analysis system (LEYTAS)," *World Journal of Urology*, vol. 1, no. 2, pp. 77–81, 1983.

[18] B. Jaggi, M. Dean, and B. Palcic, "Design of a solid-state microscope," *Optical Engineering*, vol. 28, no. 6, pp. 286675–286675, 1989.

[19] B. Palcic, D. Garner, C. MacAulay, J. Matisic, and G. Anderson, "Oncometrics Imaging Corporation and Xillix Technologies Corporation: Use of the Cyto-Savant in quantitative cytology," *Acta Cytologica*, vol. 40, no. 1, p. 67, 1996.

[20] G. Anderson, C. Macaulay, J. Matisic, D. Garner, and B. Palcic, "The use of an automated image cytometer for screening and quantitative assessment of cervical lesions in the British Columbia Cervical Smear Screening Programme," *Cytopathology*, vol. 8, no. 5, pp. 298–312, 1997.

[21] B. Palcic, C. MacAulay, S. Harrison, S. Lam, P. Payne, D. Garner, and A. Doudkine, "System and method for automatically detecting malignant cells and cells having malignancy-associated changes," Feb. 15 2000. US Patent 6,026,174.

[22] B. Palcic, D. M. Garner, J. Beveridge, X. R. Sun, A. Doudkine, C. MacAulay, S. Lam, and P. W. Payne, "Increase of sensitivity of sputum cytology using high-resolution image cytometry: Field study results," *Cytometry*, vol. 50, no. 3, pp. 168–176, 2002.

[23] T. Schneider, D. Wissing, A. Böcking, D. Meyer-Ebrecht, and T. Aach, "Cytological cancer diagnostics: Support vector machine based separation of epithelial cells from smears of the oral mucosa into normal and abnormal cells for

DNA image cytometry," in *41. DGBMT-Jahrestagung Biomedizinische Technik (BMT 2007)*, (Aachen), September 26–29 2007.

[24] D. Saslow, D. Solomon, H. W. Lawson, M. Killackey, S. L. Kulasingam, J. Cain, F. A. Garcia, A. T. Moriarty, A. G. Waxman, D. C. Wilbur, *et al.*, "American cancer society, american society for colposcopy and cervical pathology, and american society for clinical pathology screening guidelines for the prevention and early detection of cervical cancer," *CA: A Cancer Journal for Clinicians*, vol. 62, no. 3, pp. 147–172, 2012.

[25] M. Melamed, B. Flehinger, and M. Zaman, "Impact of early detection on the clinical course of lung cancer.," *The Surgical Clinics of North America*, vol. 67, no. 5, p. 909, 1987.

[26] T. A. Brentnall, M. P. Bronner, D. R. Byrd, R. C. Haggitt, and M. B. Kimmey, "Early diagnosis and treatment of pancreatic dysplasia in patients with a family history of pancreatic cancer," *Annals of Internal Medicine*, vol. 131, no. 4, pp. 247–255, 1999.

[27] B. Schmalfeldt and J. Pfisterer, "Interdisziplinäre S2k-Leitlinie für die Diagnostik und Therapie maligner Ovarialtumoren," tech. rep., Kommission Ovar der Arbeitsgemeinschaft Gynäkologische Onkologie e.V. in der Deutschen Gesellschaft für Gynäkologie und Geburtshilfe e.V. sowie in der Deutschen Krebsgesellschaft e.V., 2007.

[28] S. Roemeling, M. J. Roobol, S. H. de Vries, T. Wolters, C. Gosselaar, G. J. van Leenders, and F. H. Schröder, "Active surveillance for prostate cancers detected in three subsequent rounds of a screening trial: Characteristics, PSA doubling times, and outcome," *European Urology*, vol. 51, no. 5, pp. 1244 – 1251, 2007.

[29] M. Wirth, L. Weißbach, R. Ackermann, *et al.*, "Interdisziplinäre Leitlinie der Qualität S3 zur Früherkennung, Diagnose und Therapie der verschiedenen Stadien des Prostatakarzinoms," tech. rep., Deutsche Gesellschaft für Urologie e. V, 2009.

[30] H. G. Welch and W. C. Black, "Overdiagnosis in cancer," *Journal of the National Cancer Institute*, vol. 102, no. 9, pp. 605–613, 2010.

[31] D. F. Penson, D. McLerran, Z. Feng, L. Li, P. C. Albertsen, F. D. Gilliland, A. Hamilton, R. M. Hoffman, R. A. Stephenson, A. L. Potosky, *et al.*, "5-year urinary and sexual outcomes after radical prostatectomy: results from the prostate cancer outcomes study," *Journal of Urology*, vol. 173, no. 5, pp. 1701–1705, 2005.

[32] J. L. Stanford, Z. Feng, A. S. Hamilton, F. D. Gilliland, R. A. Stephenson, J. W. Eley, P. C. Albertsen, L. C. Harlan, and A. L. Potosky, "Urinary and sexual function after radical prostatectomy for clinically localized prostate cancer," *JAMA: The Journal of the American Medical Association*, vol. 283, no. 3, pp. 354–360, 2000.

[33] C. Bangma, S. Roemeling, and F. Schröder, "Overdiagnosis and overtreatment of early detected prostate cancer," *World Journal of Urology*, vol. 25, no. 1, pp. 3–9, 2007.

[34] V. A. Moyer, "Screening for prostate cancer: US preventive services task force recommendation statement," *Annals of Internal Medicine*, vol. 157, no. 2, pp. 120–134, 2012.

[35] R. A. Godtman, E. Holmberg, A. Khatami, J. Stranne, and J. Hugosson, "Outcome following active surveillance of men with screen-detected prostate cancer: Results from the göteborg randomised population-based prostate cancer screening trial," *European Urology*, vol. 63, no. 1, pp. 101 – 107, 2013.

[36] D. Gleason, "Classification of prostatic carcinomas," *Cancer chemotherapy reports. Part 1*, vol. 50, pp. 125–128, 1966.

[37] J. I. Epstein, W. C. Allsbrook Jr, M. B. Amin, L. L. Egevad, *et al.*, "The 2005 international society of urological pathology (ISUP) consensus conference on gleason grading of prostatic carcinoma," *The American Journal of Surgical Pathology*, vol. 29, no. 9, pp. 1228–1242, 2005.

[38] P. A. Humphrey, "Gleason grading and prognostic factors in carcinoma of the prostate," *Modern Pathology*, vol. 17, no. 3, pp. 292–306, 2004.

[39] M. A. Dall'Era, M. R. Cooperberg, J. M. Chan, B. J. Davies, P. C. Albertsen, L. H. Klotz, C. A. Warlick, L. Holmberg, D. E. Bailey, M. E. Wallace, *et al.*, "Active surveillance for early-stage prostate cancer," *Cancer*, vol. 112, no. 8, pp. 1650–1659, 2008.

[40] L. B. Rorke, "Pathologic diagnosis as the gold standard," *Cancer*, vol. 79, no. 4, pp. 665–667, 1997.

[41] M. T. Fahey, L. Irwig, and P. Macaskill, "Meta-analysis of Pap test accuracy," *American Journal of Epidemiology*, vol. 141, no. 7, pp. 680–689, 1995.

[42] M. E. Pretorius, H. Wæhre, V. M. Abeler, B. Davidson, L. Vlatkovic, R. A. Lothe, K.-E. Giercksky, and H. E. Danielsen, "Large scale genomic instability as an additive prognostic marker in early prostate cancer," *Analytical Cellular Pathology*, vol. 31, no. 4, pp. 251–259, 2009.

[43] A. Billis, M. S. Guimaraes, L. Freitas, L. Meirelles, L. A. Magna, U. Ferreira, *et al.*, "The impact of the 2005 international society of urological pathology consensus conference on standard gleason grading of prostatic carcinoma in needle biopsies.," *The Journal of Urology*, vol. 180, no. 2, p. 548, 2008.

[44] W. K. Lau, M. L. Blute, D. G. Bostwick, A. L. Weaver, T. J. Sebo, and H. Zincke, "Prognostic factors for survival of patients with pathological gleason score 7 prostate cancer: differences in outcome between primary gleason grades 3 and 4," *The Journal of Urology*, vol. 166, no. 5, pp. 1692–1697, 2001.

[45] J. Adolfsson, L. Ronstrom, P. Hedlund, T. Lowhagen, J. Carstensen, and B. Tribukait, "The prognostic value of modal deoxyribonucleic acid in low

grade, low stage untreated prostate cancer," *The Journal of Urology*, vol. 144, no. 6, pp. 1404–6, 1990.

[46] M. Burchardt, R. Engers, M. Müller, T. Burchardt, R. Willers, J. Epstein, R. Ackermann, H. Gabbert, A. de la Taille, and M. Rubin, "Interobserver reproducibility of gleason grading: evaluation using prostate cancer tissue microarrays," *Journal of Cancer Research and Clinical Oncology*, vol. 134, no. 10, pp. 1071–1078, 2008.

[47] G. Haroske, F. Giroud, A. Reith, and A. Böcking, "1997 ESACP consensus report on diagnostic DNA image cytometry - Part I: basic considerations and recommendations for preparation, measurement and interpretation," *Analytical Cellular Pathology*, vol. 17, no. 4, p. 189, 1998.

[48] D. Kindermann and C. Hilgers, "Glare-correction in DNA image cytometry.," *Analytical cellular pathology: the journal of the European Society for Analytical Cellular Pathology*, vol. 6, no. 2, p. 165, 1994.

[49] G. Haroske, W. D. Meyer, F. Theissig, K. D. Kunze, *et al.*, "Increase of precision and accuracy of DNA cytometry by correcting diffraction and glare errors," *Analytical cellular pathology: the journal of the European Society for Analytical Cellular Pathology*, vol. 9, no. 1, p. 1, 1995.

[50] T. Würflinger, L. Höck, A. Böcking, D. Meyer-Ebrecht, and T. Aach, "Accurate DNA image cytometry for automated screening scenarios," *Cellular Oncology, ISCO Congress*, vol. 30, no. 2, p. 114, 2008.

[51] F. Giroud, G. Haroske, A. Reith, and A. Böcking, "1997 ESACP consensus report on diagnostic DNA image cytometry - Part II: Specific recommendations for quality assurance," *Analytical Cellular Pathology*, vol. 17, pp. 201–208, 1998.

[52] B. Tribukait, "Nuclear deoxyribonucleic acid determination in patients with prostate carcinomas: clinical research and application," *European Urology*, vol. 23, p. 64, 1993.

[53] M. Borre, M. Høyer, B. Nerstrøm, J. Overgaard, *et al.*, "DNA ploidy and survival of patients with clinically localized prostate cancer treated without intent to cure," *The Prostate*, vol. 36, no. 4, pp. 244–249, 1998.

[54] S. Vesalainen, S. Nordling, P. Lipponen, M. Talja, and K. Syrjänen, "Progression and survival in prostatic adenocarcinoma: a comparison of clinical stage, gleason grade, S-phase fraction and DNA ploidy.," *British Journal of Cancer*, vol. 70, no. 2, p. 309, 1994.

[55] V. Q. H. Nguyen, H. J. Grote, N. Pomjanski, K. Knops, and A. Böcking, "Interobserver reproducibility of DNA-image-cytometry in ASCUS or higher cervical cytology," *Analytical Cellular Pathology*, vol. 26, no. 3, pp. 143–150, 2004.

[56] A. Pindur, S. Chakraborty, D. Welch, and T. Wheeler, "DNA ploidy measurements in prostate cancer: differences between image analysis and flow cy-

tometry and clinical implications," *The Prostate*, vol. 25, no. 4, pp. 189–198, 1994.

[57] P. Duesberg, "Chromosomal chaos and cancer," *Scientific American*, vol. 296, no. 5, pp. 52–59, 2007.

[58] P. Duesberg, D. Rasnick, R. Li, L. Winters, C. Rausch, R. Hehlmann, *et al.*, "How aneuploidy may cause cancer and genetic instability.," *Anticancer Research*, vol. 19, no. 6A, p. 4887, 1999.

[59] P. Duesberg, C. Iacobuzio-Donahue, J. A. Brosnan, A. McCormack, D. Mandrioli, and L. Chen, "Origin of metastases: subspecies of cancers generated by intrinsic karyotypic variations," *Cell Cycle*, vol. 11, no. 6, pp. 1151–1166, 2012.

[60] L. L. Leape, "Error in medicine," *JAMA: The Journal of the American Medical Association*, vol. 272, no. 23, pp. 1851–1857, 1994.

[61] H. Pape, C. Pöttgen, J. Ploem, A. Van Driel-Kulker, R. Wurm, and G. Schmitt, "The prognostic value of DNA content measured by image cytometry in soft tissue sarcomas," *Annals of Oncology*, vol. 3, no. suppl 2, pp. S89–S92, 1992.

[62] T. Wagner, A. Adler, P. Sevelda, I. Assmann, C. Knepflé, K. Czerwenka, H. Heinzl, and J. Ploem-Zaaijer, "Prognostic significance of cell DNA content in early-stage ovarian cancer (figo stages I and II/A) by means of automatic image cytometry," *International Journal of Cancer*, vol. 56, no. 2, pp. 167–172, 1994.

[63] A. M. van Driel-Kulker and J. S. Ploem, "The use of LEYTAS in analytical and quantitative cytology," *Biomedical Engineering, IEEE Transactions on*, vol. BME-29, no. 2, pp. 92–100, 1982.

[64] M. Stöckle, H. Tanke, W. Mesker, J. Ploem, U. Jonas, and R. Hohenfellner, "Automated DNA-image cytometry: a prognostic tool in infiltrating bladder carcinoma?," *World Journal of Urology*, vol. 5, no. 2, pp. 127–132, 1987.

[65] B. Jaggi and B. Palcic, "Charge-coupled device requirements in the design of a quantitative microscope for image cytometry," in *Proceedings 6th Ann. Electronic Imaging West '89, Pasadena CA, Waltham, MA*, pp. 177–82., 1989.

[66] B. Palcic, C. MacAulay, S. Shlien, W. Treurniet, H. Tezcan, and G. Anderson, "Comparison of three different methods for automated classification of cervical cells.," *Analytical Cellular Pathology*, vol. 4, no. 6, p. 429, 1992.

[67] B. Palcic, D. Garner, and C. MacAulay, "Image cytometry and chemoprevention in cervical cancer," *Journal of Cellular Biochemistry*, vol. 59, no. S23, pp. 43–54, 1995.

[68] T. Schneider, A. Bell, D. Meyer-Ebrecht, A. Böcking, and T. Aach, "Computer aided cytological cancer diagnosis: Cell type classification as a step towards fully automatic cancer diagnosis on cytopathological specimens of serous effusions," in *Medical Imaging 2007: Computer-Aided Diagnosis* (M. L. Giger and

N. Karssemeijer, eds.), (San Diego, USA), pp. 65140G–1–10, SPIE Vol. 6514, February 17–22 2007.

[69] T. Schneider, A. Bell, G. Herberich, D. Meyer-Ebrecht, A. Böcking, and T. Aach, "Chromatinmuster-basierte Zellklassifizierung für die DNS-Bildzytometrie an Mundschleimhaut-Abstrichen," in *Bildverarbeitung für die Medizin 2007* (H. Handels, J. Ehrhardt, A. Horsch, H.-P. Meinzer, and T. Tolxdorff, eds.), (München), pp. 257–261, Springer, 2007.

[70] T. E. Schneider, "Automated classification of analysis- and reference cells for cancer diagnostics in microscopic images of epithelial cells from the oral mucosa," *Acta Polytechnica: Journal of Advanced Engineering*, vol. 47, no. 4-5, pp. 86–90, 2007.

[71] A. Böcking, D. Friedrich, C. Jin, A. Bell, T. Würflinger, D. Meyer-Ebrecht, Y. Zhang, S. Biesterfeld, N. Pomjanski, L. Berynskyy, R. Yeung, J. Handschel, B. Palcic, and T. Aach, *Oral Cytology: A Concise Guide*, ch. Diagnostic Cytometry, pp. 125–145. Springer, 2012.

[72] I. P. Janecka, "Cancer control through principles of systems science, complexity, and chaos theory: A model," *International Journal of Medical Sciences*, vol. 4, no. 3, pp. 164–173, 2007.

[73] R. Duda, P. Hart, and D. Stork, *Pattern classification*. Pattern Classification and Scene Analysis: Pattern Classification, Wiley, 2001.

[74] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *Neural Networks, IEEE Transactions on*, vol. 13, no. 2, pp. 415–425, 2002.

[75] S. Jia, J. Kang, J. Chen, and Z. Gao, "A method for autofocus during digital slide scanning," 2014.

[76] A. Köhler, "Gedanken zu einem neuen Beleuchtungsverfahren für mikrophotographische Zwecke," *Zeitschrift für wissenschaftliche Mikroskopie*, vol. X, pp. 433–440, 1893.

[77] B. Jähne, *Digital Image Processing*. Springer, 6 ed., 2005.

[78] L. Höck, "Entwicklung robuster Verfahren für die automatische Segmentierung und Identifikation geeigneter zellkerne in der DNA-Bildzytometrie," Master's thesis, RWTH Aachen University, 2008.

[79] E. Bengtsson, C. Wahlby, and J. Lindblad, "Robust cell image segmentation methods," *Pattern Recognition and Image Analysis*, vol. 14, no. 2, pp. 157–167, 2004.

[80] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *Biomedical Engineering, IEEE Reviews in*, vol. 2, pp. 147–171, 2009.

[81] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation and classification in digital histopathology: A review," *IEEE Reviews in Biomedical EngineeringBiomedical Engineering*, 2013.

[82] M. Gao, P. Bridgman, and S. Kumar, "Computer-aided prostrate cancer diagnosis using image enhancement and jpeg2000," in *Optical Science and Technology, SPIE's 48th Annual Meeting*, pp. 323–334, International Society for Optics and Photonics, 2003.

[83] T. Würflinger, J. Stockhausen, D. Meyer-Ebrecht, and A. Böcking, "Robust automatic coregistration, segmentation, and classification of cell nuclei in multimodal cytopathological microscopic images," *Computerized Medical Imaging and Graphics*, vol. 28, pp. 87–98, 2004.

[84] B. Nielsen, F. Albregtsen, and H. E. Danielsen, "Automatic segmentation of cell nuclei in Feulgen-stained histological sections of prostate cancer and quantitative evaluation of segmentation results," *Cytometry Part A*, vol. 81, no. 7, pp. 588–601, 2012.

[85] 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 *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pp. 284–287, IEEE, 2008.

[86] V.-T. Ta, O. Lézoray, A. Elmoataz, and S. Schüpp, "Graph-based tools for microscopic cellular image segmentation," *Pattern Recognition*, vol. 42, no. 6, pp. 1113–1125, 2009.

[87] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images," *Biomedical Engineering, IEEE Transactions on*, vol. 57, no. 4, pp. 841–852, 2010.

[88] D. Friedrich, J. zur Jacobsmühlen, T. Braunschweig, A. Bell, K. Chaisaowong, R. Knüchel-Clarke, and T. Aach, "Detection of immunocytological markers in photomicroscopic images," in *SPIE Medical Imaging 2012: Computer-Aided Diagnosis* (B. van Ginneken and C. L. Novak, eds.), vol. 8315, p. 83153D, SPIE, 2012.

[89] W. F. Clocksin, "Automatic segmentation of overlapping nuclei with high background variation using robust estimation and flexible contour models," in *Image Analysis and Processing, 2003. Proceedings. 12th International Conference on*, pp. 682–687, IEEE, 2003.

[90] D. Glotsos, P. Spyridonos, D. Cavouras, P. Ravazoula, P.-A. Dadioti, and G. Nikiforidis, "Automated segmentation of routinely hematoxylin-eosin-stained microscopic images by combining support vector machine clustering and active contour models," *Analytical and Quantitative Cytology and Histology*, vol. 26, no. 6, pp. 331–340, 2004.

[91] C. MacAulay and B. Palcic, "An edge relocation segmentation algorithm," *Analytical and quantitative cytology and histology/the International Academy of Cytology and American Society of Cytology*, vol. 12, no. 3, p. 165, 1990.

[92] P. Bamford and B. Lovell, "Unsupervised cell nucleus segmentation with active contours," *Signal Processing*, vol. 71, no. 2, pp. 203–213, 1998.

[93] T. Mouroutis, S. J. Roberts, and A. A. Bharath, "Robust cell nuclei segmentation using statistical modelling," *Bioimaging*, vol. 6, no. 2, pp. 79–91, 1998.

[94] D. Friedrich, A. L. Cobos, S. Biesterfeld, A. Böcking, and D. Meyer-Ebrecht, "Konturverfeinerung über Fourierdeskriptoren," in *Bildverarbeitung für die Medizin 2013*, pp. 211–216, Springer, 2013.

[95] S. Beucher and C. Lantuejoul, "Use of watersheds in contour detection," in *International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation*, 1979.

[96] S. Beucher, "The watershed transformation applied to image segmentation," *Scanning Microscopy*, pp. 299–299, 1992.

[97] N. Malpica, C. Ortiz de Solorzano, J. J. Vaquero, A. Santos, I. Vallcorba, J. M. Garcia-Sagredo, and F. d. Pozo, "Applying watershed algorithms to the segmentation of clustered nuclei," *Cytometry, Part A*, vol. 28, no. 4, pp. 289–297, 1997.

[98] C. Wählby, J. Lindblad, M. Vondrus, E. Bengtsson, and L. Björkesten, "Algorithms for cytoplasm segmentation of fluorescence labelled cells," *Analytical Cellular Pathology*, vol. 24, no. 2, pp. 101–111, 2002.

[99] C. Wählby, I.-M. Sintorn, F. Erlandsson, G. Borgefors, and E. Bengtsson, "Combining intensity, edge and shape information for 2d and 3d segmentation of cell nuclei in tissue sections," *Journal of Microscopy*, vol. 215, no. 1, pp. 67–76, 2004.

[100] F. Cloppet and A. Boucher, "Segmentation of overlapping/aggregating nuclei cells in biological images," in *19th International Conference on Pattern Recognition*, 2008.

[101] O. Schmitt and M. Hasse, "Radial symmetries based decomposition of cell clusters in binary and gray level images," *Pattern Recognition*, vol. 41, no. 6, pp. 1905–1923, 2008.

[102] J. Cheng and J. C. Rajapakse, "Segmentation of clustered nuclei with shape markers and marking function," *Biomedical Engineering, IEEE Transactions on*, vol. 56, no. 3, pp. 741–748, 2009.

[103] S. Di Cataldo, E. Ficarra, A. Acquaviva, and E. Macii, "Achieving the way for automated segmentation of nuclei in cancer tissue images through morphology-based approach: a quantitative evaluation," *Computerized Medical Imaging and Graphics*, vol. 34, no. 6, pp. 453–461, 2010.

[104] T. Yeo, X. Jin, S. Ong, Jayasooriah, and R. Sinniah, "Clump splitting through concavitiy analysis," *Pattern Recognition Letters*, vol. 15, pp. 1013–1018, 1994.

[105] S. Kumar, S. Ong, S. Ranganath, T. Ong, and F. Chew, "A rule-based approach for robust clump splitting," *Pattern Recognition*, vol. 39, Issue , June, pp. 1088–1098, June 2006.

[106] O. Schmitt and S. Reetz, "On the decomposition of cell clusters," *Journal of Mathematical Imaging and Vision*, vol. 33, no. 1, pp. 85–103, 2009.

[107] X. Bai, C. Sun, and F. Zhou, "Splitting touching cells based on concave points and ellipse fitting," *Pattern Recognition*, vol. 42, no. 11, pp. 2434–2446, 2009.

[108] H. Kong, M. Gurcan, and K. Belkacem-Boussaid, "Partitioning histopathological images: An integrated framework for supervised color-texture segmentation and cell splitting," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 9, pp. 1661–1677, 2011.

[109] D. Friedrich, C. Haarburger, A. Luna-Cobos, D. Meyer-Ebrecht, A. Böcking, and D. Merhof, "Segmentierung von Zellkernen für die Hochdurchsatz-DNA-Bildzytomerie," in *Bildverarbeitung für die Medizin 2014*, pp. 216–221, Springer, 2014.

[110] C. Haarburger, "Segmentierung von Zellkernen für die Hochdurchsatz-DNA-Bildzytomerie," Master's thesis, RWTH Aachen University, 2013.

[111] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *Image Processing, IEEE Transactions on*, vol. 7, no. 3, pp. 359–369, 1998.

[112] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[113] T. F. Chan and L. A. Vese, "Active contours without edges," *Image processing, IEEE transactions on*, vol. 10, no. 2, pp. 266–277, 2001.

[114] A. Rosenfeld, "Measuring the sizes of concavities," *Pattern Recognition Letters*, vol. 3, pp. 71–75, 1985.

[115] Q. Chen, X. Yang, and E. M. Petriu, "Watershed segmentation for binary images with different distance transforms," in *Proc. The 3rd IEEE International Workshop on Haptic, Audio and Visual Environments and Their Applications*, pp. 111–116, 2004.

[116] K. Rodenacker and E. Bengtsson, "A feature set for cytometry on digitized microscopic images," *Analytical Cellular Pathology*, vol. 25, no. 1, pp. 1–36, 2003.

[117] A. Jasinski, "Klassifikation von Zellkernen für die DNA-Bildzytometrie," Master's thesis, RWTH Aachen University, 2013.

[118] S. Aksoy and R. M. Haralick, "Feature normalization and likelihood-based similarity measures for image retrieval," *Pattern Recognition Letters*, vol. 22, no. 5, pp. 563–582, 2001.

[119] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[120] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Academic Press, 1999.

[121] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[122] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Academic Press Limited, 1990.

[123] L. Firestone, K. Cook, K. Culp, N. Talsania, and K. Preston, "Comparison of autofocus methods for automated microscopy," *Cytometry*, vol. 12, pp. 195–206, 1991.

[124] A. Kimura, M. G. F. Costa, C. F. F. C. Filho, L. B. M. Fujimoto, and J. Salem, "Evaluation of autofocus functions of conventional sputum smear microscopy for tuberculosis," in *32nd Annual International Conference of the IEEE EMBS*, 2010.

[125] D. Friedrich, "Eliminating defocused objects from a cell collection used for early cancer diagnosis," in *Proceedings of the 17th International Student Conference on Electrical Engineering POSTER 2013*, 2013.

[126] T. Darrell and K. Wohn, "Pyramid based depth from focus," in *Computer Vision and Pattern Recognition, 1988. Proceedings CVPR'88., Computer Society Conference on*, pp. 504–509, IEEE, 1988.

[127] D. Friedrich, C. Jin, Y. Zhang, C. Demin, L. Yuan, L. Berynskyy, S. Biesterfeld, D. Meyer-Ebrecht, and A. Böcking, "A pattern recognition system for identifying cancer cells in serous effusions," in *Proceddings of Pathology Informatics 2012, Journal of Pathology Informatics*, vol. 3, pp. S28–S29, 2012.

[128] D. Friedrich, C. Jin, Y. Zhang, C. Demin, L. Yuan, L. Berynskyy, S. Biesterfeld, T. Aach, and A. Böcking, "Identification of prostate cancer cell nuclei for DNA-grading of malignancy," in *Bildverarbeitung für die Medizin 2012*, pp. 334–339, Springer, 2012.

[129] A. Jasinski and D. Friedrich, "Classifier analysis for the classification of prostate nuclei," in *Proceedings of the 18th International Student Conference on Electrical Engineering POSTER 2014*, 2014.

[130] M. S. Martínez, "Feature selection for improved classification in DNA image cytometry," Master's thesis, RWTH Aachen University, 2012.

[131] S. Koppers, "Merkmale für die Klassifikation von Zellkernen in der DNA-Bildzytometrie," Master's thesis, RWTH Aachen University, 2014.

[132] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, pp. 62–66, January 1979.

[133] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[134] E. Davies, "A modified hough scheme for general circle location," *Pattern Recognition Letters*, vol. 7, no. 1, pp. 37–43, 1988.

[135] D. Meyer-Ebrecht, A. Böcking, A. Bell, T. Klein, T. Schneider, J. Stockhausen, and T. Würflinger, "Multimodal cell analysis (MMCA): A new method for

highly accurate and early non-invasive cancer diagnoses," in *Biomedical Engineering (BMT 2004)*, vol. 1, pp. 410–411, 2004.

[136] T. Remmerbach, D. Meyer-Ebrecht, T. Aach, T. Würflinger, A. Bell, T. Schneider, N. Nietzke, B. Frerich, and A. Böcking, "Towards a multi modal cell analysis of brush biopsies for early detection of oral cancer," *Cancer Cytopathology*, vol. 117, no. 3, pp. 228–235, 2009.

[137] D. Friedrich, D. Meyer-Ebrecht, A. Böcking, and D. Merhof, "Detection of felt tip markers on microscope slides," in *SPIE Medical Imaging*, pp. 904105–904105, International Society for Optics and Photonics, 2014.

[138] R. Halır and J. Flusser, "Numerically stable direct least squares fitting of ellipses," *Proc. 6th International Conference in Central Europe on Computer Graphics and Visualization. WSCG*, vol. 98, pp. 125–132, 1998.

[139] R. Kikinis, "Ron's rules for tools." observed 6th of October, http://www.slicer.org/slicerWiki/index.php/Documentation-Rons-Rules-For-Tools.

[140] M. G. Rojo, G. B. García, C. P. Mateos, J. G. García, and M. C. Vicente, "Critical comparison of 31 commercially available digital slide systems in pathology," *International Journal of Surgical Pathology*, vol. 14, no. 4, pp. 285–305, 2006.

[141] A. Travinsky, "DNA content measurement with high throughput digital microscopy," Master's thesis, RWTH Aachen University, 2014.

[142] A. C. Ruifrok, D. A. Johnston, *et al.*, "Quantification of histochemical staining by color deconvolution.," *Analytical and quantitative cytology and histology/the International Academy of Cytology and American Society of Cytology*, vol. 23, no. 4, p. 291, 2001.

[143] A. Bell, J. Kaftan, D. Meyer-Ebrecht, and T. Aach, "An evaluation framework for the accuracy of camera transfer fuctions estimated from differently exposed images," in *7th IEEE Southwest Symposium on Image Analysis and Interpretation*, 2006.

[144] "ISO 14524:2009 photography – electronic still-picture cameras – methods for measuring opto-electronic conversion functions (OECFs)."

[145] B. Jähne, *Practical handbook on image processing for scientific and technical applications*. CRC Press, 2004.

[146] A. Bell, J. Brauers, J. N. Kaftan, D. Meyer-Ebrecht, A. Böcking, and T. Aach, "High dynamic range microscopy for cytopathological cancer diagnosis," *IEEE Journal of Selected Topics in Signal Processing (Special Issue on: Digital Image Processing Techniques for Oncology)*, vol. 3, no. 1, pp. 170–184, 2009.

[147] D. Friedrich, J. Brauers, A. A. Bell, and T. Aach, "Towards fully automated precise measurement of camera transfer functions," in *Image Analysis & Interpretation (SSIAI), 2010 IEEE Southwest Symposium on*, pp. 149–152, IEEE, 2010.

[148] B. Berger-Fröhlich, *Verbesserung der Messpräzision der diagnostischen DNA-Bildzytometrie*. PhD thesis, Heinrich-Heine University Düsseldorf, Germany, to appear.

[149] A. Bell, T. Schneider, D. Müller-Frank, D. Meyer-Ebrecht, A. Böcking, and T. Aach, "Fully automatic screening of immunocytochemically stained cytopathological specimens for early cancer detection," in *Medical Imaging 2007: Computer-Aided Diagnosis* (M. L. Giger and N. Karssemeijer, eds.), (San Diego, USA), pp. 65131–1–11, SPIE Vol. 6514, February 17–22 2007.

[150] B. Lahrmann, N. A. Valous, U. Eisenmann, N. Wentzensen, and N. Grabe, "Semantic focusing allows fully automated single-layer slide scanning of cervical cytology slides," *PloS One*, vol. 8, no. 4, p. e61441, 2013.

[151] D. Goldstein, "Aspects of scanning microdensitometry I. Stray light (glare)," *Journal of Microscopy*, vol. 92, no. 1, pp. 1–16, 1970.