

PhD thesis

Stage-specific detection and characterisation of the malaria parasites

Katharina Preißinger

Supervisor: Dr. Béata G. Vértessy, Dr. István Kézsmárki

Department of Physics
Budapest University of Technology and Economics
University of Augsburg
(2023)



Contents

1	Motivation	1
2	Introduction	4
2.1	What is malaria?	4
2.1.1	Morphological properties of malaria parasites	5
2.1.2	Optical properties of malaria parasites	7
2.2	Diagnostics	10
2.2.1	State of the art in the diagnosis of malaria	10
2.2.2	Neural networks as boosters for microscopic diagnosis	12
3	Applied methodology	14
3.1	In vitro cultivation of <i>Plasmodium falciparum</i>	14
3.2	Atomic force microscopy	16
3.2.1	Interactions between tip and sample	16
3.2.2	Probe characteristics	19
3.2.3	Topography measurements	20
3.3	Fluorescence microscopy	21
3.3.1	Basic principles of fluorescence emission	21
3.3.2	Fluorescence microscopy techniques	22
3.4	Neural networks	23
3.4.1	The process of machine learning	24
3.4.2	Principles and architectures of neural networks	25
3.4.3	Convolutional neural networks	29
3.4.4	Evaluation of neural networks	29
3.5	Experimental equipment and methods	31
4	Results	34
4.1	High-parasitemia cultures for in vitro analysis of <i>P. falciparum</i> under physiological conditions	34
4.2	Imaging, identification, and manual classification of malaria	37
4.2.1	Imaging of the morphological and autofluorescent properties of <i>P. falciparum</i> -infected RBCs	37
4.2.2	Identification of RBCs in microscopy images	38
4.2.3	Manual classification of intra-erythrocytic stages	41
4.3	Exploring characteristic features of <i>P. falciparum</i>	43

4.3.1	Statistical analysis for the characterisation of RBCs	44
4.3.2	Morphological properties of malaria-infected erythrocytes . . .	46
4.3.3	Optical properties of malaria-infected erythrocytes	53
4.3.4	Intensity profiles of Giemsa-stained erythrocytes	60
4.4	Reduction of dimensionality as a tool for feature selection in RBCs .	62
4.4.1	Dimension reduction of two-dimensional RBC images	62
4.5	Stage-specific classification of <i>P.falciparum</i>	68
4.5.1	Neural network architecture	68
4.5.2	Stage-specific classification of two-dimensional RBC images . .	69
4.5.3	Evaluation of the stage-specific classification on characteristic features of single-RBC images	71
4.6	The malaria stage classifier	77
4.6.1	Design and Implementation	77
4.6.2	User manual	78
5	Summary and thesis points	84
5.1	Publications related to the thesis points	86
	Bibliography	87
	A Experimental data	99
A.1	T-test	100
A.2	Network architecture	104
A.3	Performance parameters of the stage-specific detection of malaria . .	107
	B Documentation of the malaria stage classifier	116
	Acronyms	149
	List of figures	152
	List of tables	154
	Acknowledgements	155

Motivation

Malaria, an infectious disease more ancient than humanity, still imposes a global burden with 240 million annual cases worldwide, as estimated in 2020 [1]. Due to the interplay of various factors, such as the ongoing COVID-19 pandemic, the disruption of measures against malaria and the global climate change, the number of cases has been increasing from 227 million in 2019 again [1]. This circumstance has a large influence on the mortality, if the infection is not promptly and effectively treated, which is often the case in endemic regions [2].

This vector borne disease in humans is caused by five species of the *Plasmodium* genus, where *P. falciparum* is the most widespread and mainly responsible for severe malaria [2]. Following the mosquito bite and the symptom-free liver stage of the infection, the parasites burst out into the blood stream to start an asexual life cycle. In this 48 hours long intra-erythrocytic cycle, they mature through the ring, trophozoite and schizont forms. At the end of the cycle, they multiply and the merozoites begin the next cycle by invading new red blood cells (RBCs) [3]. The intra-erythrocytic cycle has been the subject of intense research because it causes the main clinical symptoms and is the major target of diagnostics and antimalarial treatment [3, 4].

Nowadays, the gold-standard diagnostic tool is still light microscopy on Giemsa-stained thin blood films, which relies on human performance both in terms of sample preparation and visual inspections, making this method time-consuming and subjective [5]. The persistence of microscopy, that is rooted in its long worldwide tradition and the general concept of *seeing is believing*, defers the spread of other methods. Currently, rapid diagnostic tests represent the most competitive alternative of light microscopy due to their easy usability and affordable price. However, they are not quantitative and impose a compromise on sensitivity and species specificity [6, 7]. There have also been other methods like polymerase chain reaction (PCR) [8], flow cytometry [8] and magneto-optical detection [9, 10] developed for malaria diagnosis, however, currently they are not available in a point-of-care format required for high-throughput in-field diagnostics.

During maturation, the parasites change the topography [11–14] and the optical properties of the host RBCs [15]. In fact, it is well-established that malaria-infected RBCs can be characterised by their morphological properties, that can also be correlated with their characteristic fluorescence patterns [16, 17]. These new insights and the widespread use of microscopy triggered new approaches for the automatised recognition of malaria infected RBCs in thin blood films, which can lead to a breakthrough in malaria diagnosis. These efforts are well exemplified by recent works reporting on neural network (NN)-based algorithms for the recognition of infected RBCs [18–23], which can run even on standard cell phones. While these pioneering studies clearly demonstrate the potential of NNs for automatised malaria diagnosis, this approach still requires substantial improvement in terms of sensitivity, specificity to different malaria species and stages, as well as robustness against the imperfectness of microscopy images. The other challenge is to improve the performance while keeping the computational costs low.

As an important new direction, it has been shown recently that NNs do not only work on images of Giemsa-stained thin blood films but also perform well on label-free light microscopy images [17]. This also gives the hint that malaria diagnosis can benefit from the advantages of NNs in analysing images recorded by microscopy methods other than light microscopy. In fact, recently there is an increasing number of techniques successfully applied for high-resolution imaging of malaria infected RBCs, including topographic imaging by atomic force microscopy (AFM) [13, 14, 16] and infrared nano-imaging [12].

Despite the rapid progress in treatment and diagnosis, malaria is still a burden in African, American, Eastern Mediterranean, South-East Asian, and Western Pacific regions [1]. The recent increase in deaths by 13% compared to 2019 due to the COVID-19 pandemic requires a better understanding of the alterations in RBC structure during the maturation of malaria parasites. Tracing the morphological and optical changes may contribute to the development of new and more effective antimalarials. If such changes are characteristic to the intra-erythrocytic stages of malaria, they could improve diagnosis by NNs.

This thesis aims to investigate the structural alterations in RBCs during the intra-erythrocytic cycle of malaria, using a combination of imaging techniques. The experimental work is complemented by the development of a neural network-based stage-specific detection method for the diagnosis of malaria in microscopy images. Along this line, I prepared data base of microscopy images. This step required a detailed investigation of morphological and optical changes of infected RBCs upon the maturation of the malaria parasites by combining multiple microscopic imaging tools, including atomic force, optical, and fluorescence microscopy. Based on these results, I developed a NN-based application for the identification of infected RBCs in microscopy images and for their further classification according to the parasite age.

The outline of the thesis is as follows. In the introduction, the biology and diagnosis of malaria are explained. After this, the experimental methods used in this work are described in chapter three, focusing on the culturing of *P. falciparum* and the principles of atomic force and fluorescence microscopy. The methodology is concluded by the introduction of the basic concepts of NNs. Next, the results are discussed following the phases of the model described above, where the main focus lies on the creation of a data base for the neural network-based stage-specific detection of malaria. The results section is concluded with the presentation of an application for the classification of intra-erythrocytic malaria stages in microscopy images, which may help to improve malaria diagnosis on the global scale and is a prerequisite for the elimination of this infectious disease.

Introduction

In 2020, an estimated number of 241 million malaria cases occurred worldwide, most of them in the African regions, South East Asia, and South America. Due to the ongoing COVID-19 pandemic and the disruption of measures against malaria, the number of cases and deaths increased between 2019 and 2020 [1]. The infection is caused by the malaria parasite, a mosquito-transmitted protozoan. Malaria parasites belong to the *Plasmodium* genus. Five species infecting humans of the genus *Plasmodium* are known: *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *P. knowlesi*. Among them, *P. falciparum* causes the most deaths. It is widely spread in Africa, New Guinea and in Haiti [24]. Treatment and elimination of the disease require fast and reliable diagnosis, as the detection of malaria in patient blood samples enables specific and effective treatment. Since most endemic areas suffer from low resources, they require simple diagnostic methods. While the standard methods, such as light microscopy of stained samples and PCR offer reliable diagnosis, they are time-consuming and require trained experts. For this reason, the application of neural networks for detection and classification of malaria parasites in RBCs is rising. Starting from simple categorisation into healthy and infected cells, newly developed NNs are now able to classify all malaria blood stages.

The first part of this chapter presents the well-explored course of the malaria infection inside the human body and the induced morphological and optical changes of the host RBC. The effective treatment of the disease requires reliable and fast diagnostic methods, which will be highlighted in the subsequent section. In the last part, the focus rests on the latest application of NNs in malaria diagnosis and research.

2.1 What is malaria?

The course of the infection that is depicted in figure 2.1 starts with the female *anophe-line* mosquito transmitting the plasmodial sporozoites into the human body. From the place of injection, they enter the bloodstream directly or via lymph channels, targetting the hepatic cells, where they begin to reproduce asexually. After approximately 6-10 days, the sporozoites have proliferated to form merozoites, which are

released into the circulation from the hepatocyte [4]. The merozoites are oval shaped forms of the parasite, circulating in the blood stream and invading erythrocytes by boring or wriggling into the membrane. In this phase, invasion of a single RBC by multiple parasites can happen.

The blood cycle of malaria is divided into three stages, corresponding to the asexual forms: ring, trophozoite, and schizont. Engulfed by the erythrocyte cytosol, the parasites form a circular ring of cytoplasm around their central food vacuole in the first 12 hours. The ring appears with a darkly stained chromatin in the nucleus, as shown in figure 2.1. Inside the cell, the parasites are moving freely. During maturation, the main focus is to uptake nutrients from the host erythrocyte, as the haemoglobin provides the energy and resources for replication and division of the parasite. Since the heme monomer, freed as a result of haemoglobin digestion, is toxic to the parasite, it is polymerised to hemozoin, an inert crystalline substance also called malaria pigment. This metabolic by-product is assembled inside the food vacuole. While growing, the width of the cytoplasm increases until it has reached the size of the nucleus (0 - 26 h). The stage between 26 and 36 hours is called trophozoite. To prevent invasion by other merozoites, the parasite exports antigens, increasing the stiffness of the RBC. For further growth, the parasite needs nutrients, which it draws from outside of the erythrocyte. To make the cell more permeable, it inserts proteins, which disrupt the membrane. In case of *P. falciparum*, the parasites exhibit antigens, causing the RBC to stick to the blood vessel walls. To support the adhesion, other exported proteins form knob-like bulges on the cell membrane [4]. Upon further growth, the amount of genomic Deoxyribonucleic acid (DNA) increases due to the replication, which is visible as blue stain under the microscope, when stained Giemsa and causes a morphological change in the RBC. After 36 hours, the replicated DNA is separated into individual daughter cells called schizogony. While the nuclei are already present in the trophozoite stage, they are not yet separated visually. After approximately 48 hours, the reproduced parasites burst out of the RBC and the cycle begins anew. To transmit the infection, a subpopulation of parasites develop into gametocytes, the sexual forms of *P. falciparum*, also depicted in figure 2.1. If a female *anopheline* mosquito bites a human that carries gametocytes, the male and female gametocytes picked up by the mosquito, form a zygote within the latter host. Invading the stomach of the mosquito, the zygote matures to an oocyst, expands and bursts to free sporozoites. These move to the salivary glands to await the next blood meal of the mosquito, releasing them into the next human. These phases of the infection are also indicated in figure 2.1 [3,4].

2.1.1 Morphological properties of malaria parasites

With maturation the morphology of the *P. falciparum* parasite changes. Figure 2.2 shows asexual (Phases 2 - 26) and sexual (Phases 27 - 30) blood stages of *P. falciparum* parasites within the RBCs. The corresponding age of the parasites in hours is also listed in the caption. In the ring stage, the parasite adopts the form of circular discs

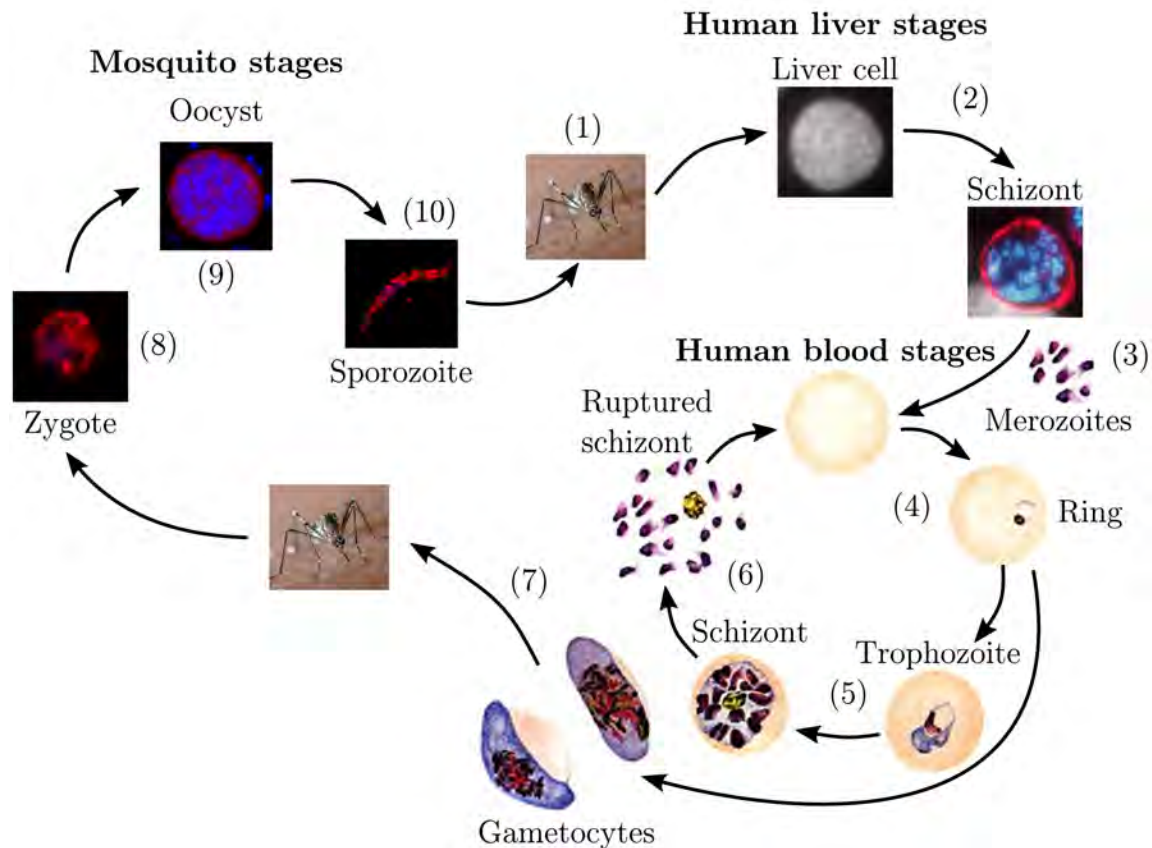


Figure 2.1: **Schematic of the malaria parasite life cycle.** The female *anopheline* mosquito transmits plasmodial sporozoites into the human body (1). The sporozoites infect liver cells and mature into schizonts (2), which rupture and release merozoites, which invade the RBCs (3). As part of the human blood stage, the parasites mature from the ring stage (4) into trophozoites (5) and schizonts, which release new merozoites into the blood stream (6). The infection is transmitted to a mosquito by gametocytes (7), which form a zygote (8) inside its new host. After maturing to an oocyst (9), new sporozoites (10) are freed, which are transmitted during the mosquito's next blood meal. Reproduced with some modifications from [25–28].

(Phases 2 - 11 in figure 2.2). In this phase, the parasites are motile in the cell. Upon transition from ring to trophozoite, the parasites develop a central cavity and become cup-shaped. This cavity increases towards the end of the trophozoite phase. During digestion, the parasite forms a food vacuole, hosting the hemozoin crystals. From several small food vacuoles in the ring stage, it becomes a prominent feature in the trophozoite stage, when they fuse to a single one. Over the progress of maturation, the food vacuole settles to one side of the parasite (Phases 12 - 18). As in the former phase, the parasite is still moving around inside the cell, showing an irregular, rounder shape (Phases 19 - 20). When reaching the final phase of the trophozoite stage, the cavity inside the parasite disappears and the parasite increases in size. The food vacuole then moves to a central position (Phases 23 - 24). A few hours before egress, the cell shows a ruffled appearance and is completely filled by the parasite (Phase 25). The food vacuole has by then assumed a compact, fully-connected layer form [11, 29]. As the hemozoin crystals assemble inside the food vacuole during infection, it appears as a protrusion inside the parasite [12]. After bursting out, the merozoites leave the hemozoin and the remains of the cell behind, ready to invade new RBCs.

A closer look into the morphology of infected erythrocytes reveals knobs on the surface of infected RBCs, appearing in the early trophozoite stage, which are not present in the ring stage [13]. The number of knobs increases towards the schizont stage and was found to be directly proportional to number of parasites per host RBC. Upon formation of the knobs on the host cell membrane in the early trophozoite stage, the cell becomes stiffer, less deformable and more fragile. This development is accompanied by a collapse of the cytoskeletal network of the infected erythrocyte [14].

2.1.2 Optical properties of malaria parasites

The protein fraction of a RBC consists up to approximately 95% of haemoglobin. During the malaria blood cycle, the parasites digest 60 - 80% of the haemoglobin inside its host cell. The released heme is stored as malaria pigment. In the early stage of the blood cycle, the haemoglobin ingestion is limited. The parasite only digests small portions of the cytoplasm. Upon maturation, the parasite takes up more of the haemoglobin and forms one large digestive vacuole, containing the metabolised hemozoin. The highest amount of haemoglobin is digested in the trophozoite stage [15]. By optical microscopy, the depletion of haemoglobin can be observed. When bound to oxygen in the blood, its light absorption reaches its maximum at 415 nm, while the maximum is shifted to a slightly higher wavelength, 432 nm, in its free form, as seen in figure 2.3).

Figure 2.4 shows microscopy images of cells excited at 405 nm [17]. At this wavelength, the haemoglobin inside the RBC absorbs most of the light and the healthy cell appears dark in the microscopy image, as clear from figure 2.4 (a). Since this is the dominant effect in the ring stage parasite, which contains no to little haemoglobin, these young parasites appear bright on a dark background, as seen in figure 2.4 (b). The centre of the ring stage is filled with a central food vacuole, which looks dark in

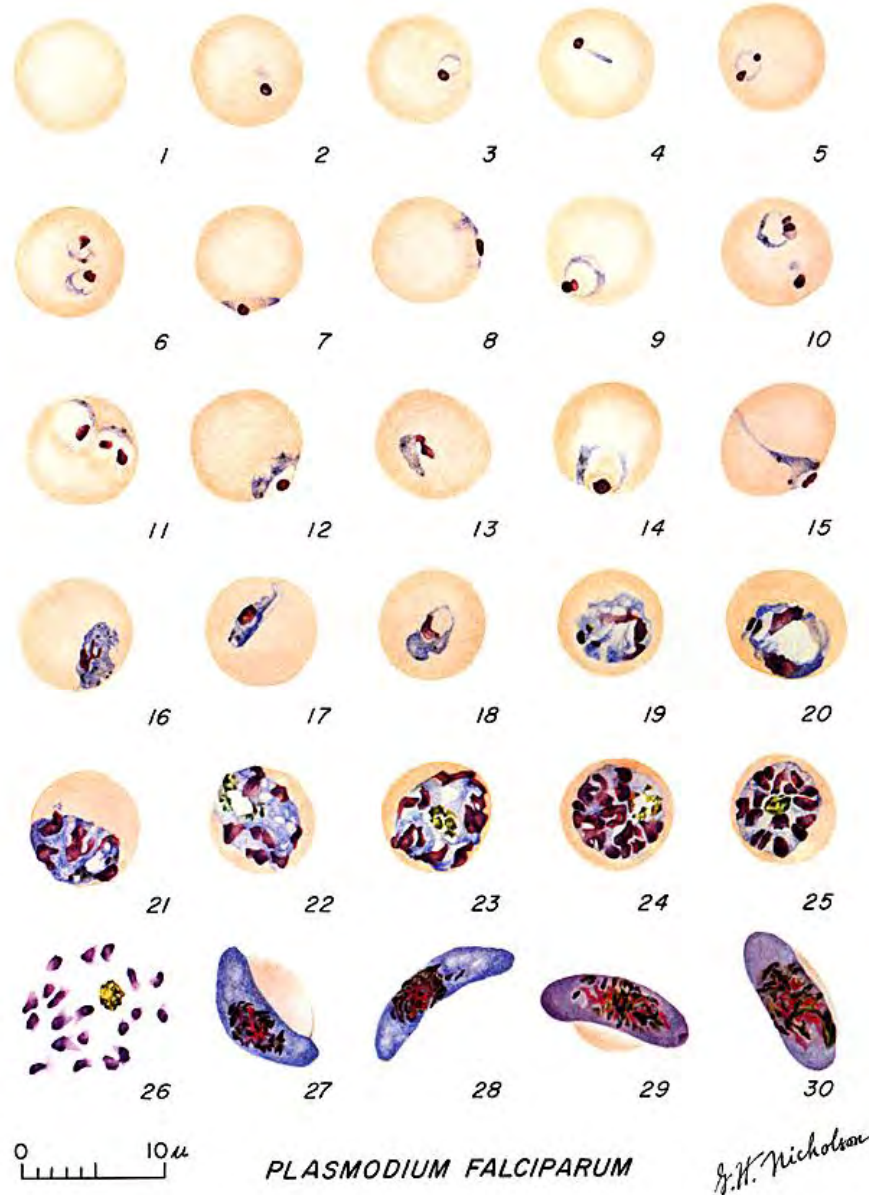


Figure 2.2: **Representative drawings of microscopy images of RBCs infected with *P. falciparum*.** The colours are based on Giemsa-stained thin blood films detected with simple light microscopy. To enhance the features of the parasites, the images were coloured artificially. 1: healthy RBCs, 2 - 5: early ring stage 0 - 6 h, 6 - 7: mid ring stage 6 - 16 h, 8 - 11: late ring stage 16 - 26 h, 12 - 15: early trophozoite stage 26 - 30 h, 16 - 18: mid trophozoite stage 30 - 34 h, 19 - 20: late trophozoite stage 34 - 38 h, 21 - 23: early schizont stage 38 - 44 h, 24 - 25: late schizont 44 - 48 h, 26: ruptured merozoites, 27 - 28: mature macrogametocytes, 29 - 30: mature microgametocytes. Reproduced from [30]

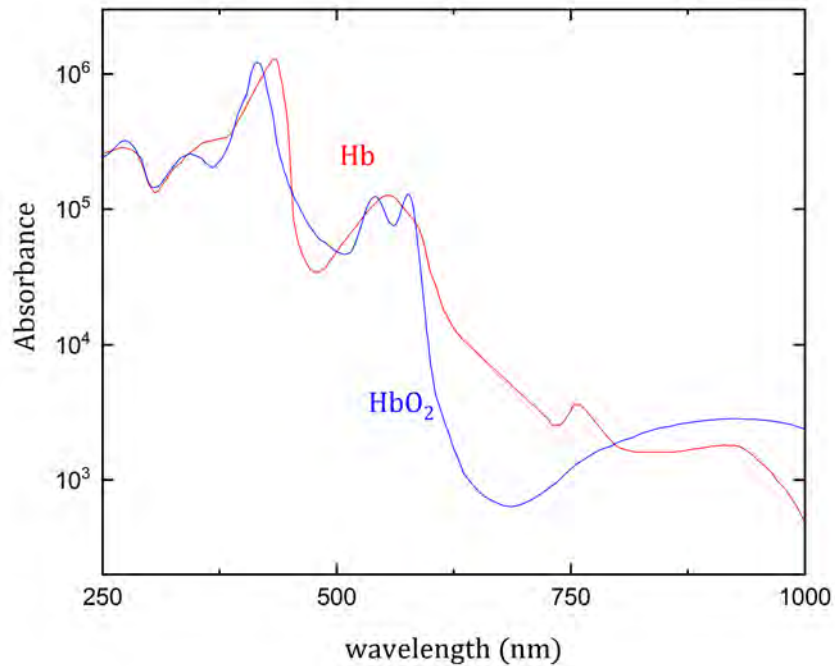


Figure 2.3: **Optical absorption of haemoglobin.** Absorption spectrum of haemoglobin (Hb, red) and haemoglobin bound to oxygen (HbO₂, blue). Reproduced from [31]

the image. During maturation, more haemoglobin is ingested and the RBC loses its absorption capability, appearing brighter than the healthy cell and the ring stage. As the crystals inside the food vacuole absorb the light, they are visible as a dark spot inside the parasite. Investigated by UV-Vis spectroscopy, hemozoin shows a maximum in absorbance at 384 nm [32], which is slightly shifted to 415 nm for the heme group [33]. Due to the spatial reordering of the hemozoin crystals during maturation, the trophozoite and schizont stages, shown in Figs. 2.4 (c) and (d) are easily distinguishable from earlier stages [17].

While the infection can be monitored with visible light, studies on fluorescing *P. falciparum* parasites strongly rely on fluorescence markers binding to specific cell components, which emit fluorescence upon excitation [34, 35]. These techniques also enable live-cell imaging, but they require a set of fluorescing dyes to visualise the parasite [36]. The occurrence of autofluorescence in malaria-infected RBCs due to the presence of zinc protoporphyrin IX (ZnPPIX) with a fluorescence excitation maximum of ZnPPIX at 425 nm [37–39], which binds to the surface of hemozoin crystals during their formation, shows that it is also possible to distinguish malaria parasites based on their fluorescence properties [40].

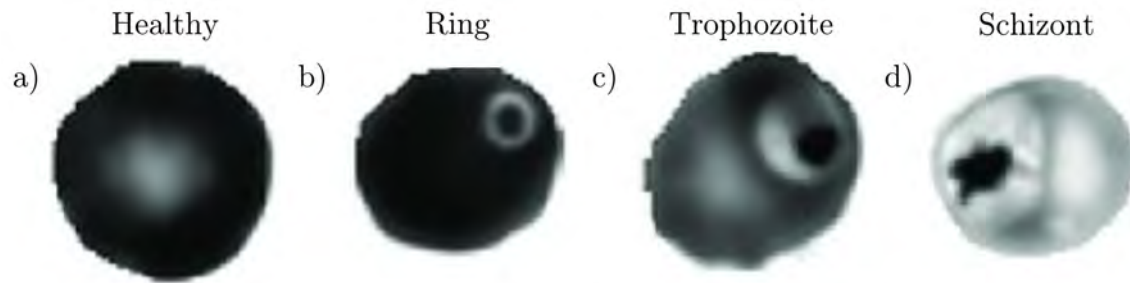


Figure 2.4: **Optical properties of malaria-infected RBCs** excited at 405 nm. (a) - (d): representative images of a healthy cell, and RBCs infected with ring-, trophozoite-, and schizont-stage parasites. Reproduced from [17].

2.2 Diagnostics

From optical diagnosis to rapid diagnostic tests, current diagnostic techniques contribute to the fight against malaria. Recently, newer techniques such as the rotating-crystal magneto-optical detection (RMOD) method have been developed, which offer a sensitivity close to nucleic acid amplification test (NAAT)s. Since most of these methods rely on trained experts, in the last few years, NNs were introduced to facilitate and to speed up diagnosis.

2.2.1 State of the art in the diagnosis of malaria

Optical microscopy-based diagnosis

Malaria is normally diagnosed by optical microscopic analysis of dried thick and thin blood films on microscopy slides, which requires a blood sample from a fingertip. While thick blood films are used to detect the infection, thin blood films provide information about the species and possible mixed infections. To investigate thin blood films, the slides are exposed to the stain for 8-10 m, after drying for approximately 2 min. The stain is then flushed from the slide by adding drops of buffered water at neutral pH of 7.0 to ensure good staining and left to dry. As an alternative to Giemsa's stain, this procedure can be done with KwikDiff [41], where the components are in separate tubes for faster staining. The slide is first dipped into a fixing agent, followed by an exposure of 2s to Eosin (red stain), and 1 min to Methylene blue. After dipping the slide in red and blue stain, the parasites exhibit a strong contrast compared to the RBCs. Each part of the parasite is coloured differently. Eosin binds to positively charged organic matters such as the membrane of the RBC, while Methylene blue binds to negatively charged molecules, e.g. DNA. If the parasite already contains a visible food vacuole, the malaria pigment appears in dark colours from golden-brown to black [5]. Figure 2.2 shows representative drawings of Giemsa-stained asexual (2 - 26) and sexual (27 - 30) blood stages of *P. falciparum*-infected RBCs with corresponding duration of the cycle in hours. For visualisation, the slides

are examined with a $100\times$ oil immersion objective in a systematic manner. There are two parameters, which are used to quantify the number of parasites in the peripheral blood: parasite density and parasitemia. To determine the parasite density, the number of white blood cells and infected RBCs is counted in several fields in parallel until a number of 200 white blood cells and more than 100 parasites or 500 white blood cells and less than 100 parasites is reached. This number then serves as a proxy for the blood volume examined based on the assumption that $1\ \mu\text{l}$ of blood contains approx. 8000 white blood cells. Thus, the number of infected RBCs can be calculated. As the other measure for the level of infection, the parasitemia is defined by

$$\text{parasitemia} = \frac{\text{infected RBCs}}{\text{healthy RBCs}}. \quad (2.1)$$

Microscopic diagnosis is nowadays the gold-standard to identify malaria and is able to provide details about the species, age, and number of parasites. Depending on the expertise of the examiner, the detection levels are around 50-100 parasites/ μl [42]. Moreover, it is not only labour-intensive and time-consuming but also requires an effective management system. This contains a central coordinator to manage the quality assurance, a reference group of microscopists with expertise in training and slide validation, training systems, supply management and maintenance of microscopes, standard operating procedures, and funding for malaria case management [43]. In spite of these drawbacks, optical microscopy is still considered to be the gold-standard approach for malaria diagnosis. Due to the widespread use of microscopy for malaria diagnosis, further development of this method, e.g. via NN-based recognition and classification, may be more desired than complete replacement by an alternative approach.

Rapid diagnostic tests

Especially in regions with limited access to microscopes, rapid diagnostic tests greatly improve the detection of malaria as they require limited training and rapidly provide results [43]. There are various types of rapid diagnostic tests. By targeting *Plasmodium* species-specific proteins and antigens, mixed malaria infections and non-*P. falciparum* infections can be distinguished from *falciparum* infections using a drop of blood within 15-30 minutes. The principle of rapid diagnostic malaria tests works similar to COVID-19 tests. Dye-labelled antibodies are captured and produce a visible band on a strip of nitro-cellulose encased by a plastic housing. It contains a square hole for blood and a round hole for the buffer. When the antibody binds to a parasite antigen, it forms a complex and is captured on the nitro-cellulose, forming a visible line in the results window. To determine the validity, a control line appears [43]. While rapid diagnostic tests offer a fast, simple way to detect malaria, they vary in sensitivity. Currently, they need to be supplied by other methods to confirm the results, characterise infection, and monitor treatment, which is not always possible in remote areas [42].

Nucleic acid amplification-based diagnostics

While microscopy and rapid diagnostic tests fail for low density malaria infections, nucleic acid amplification-based diagnostics are able to detect infections below 1 parasite/ μl . NAATs are able to detect sexual- and asexual-stage parasites as well as all five malaria species: *Plasmodium falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *P. knowlesi*. The first step of NAATs is the extraction of nucleic acid, which is followed by amplification. One of the key NAATs is PCR. The detection or quantification can be visually performed, done on an agarose gel, through examination of reaction tubes, or through incorporation of fluorescent labels in the reaction [43]. While NAATs are sensitive, they require high expertise for the analysis of the blood samples [43].

Polymerase chain reaction

PCR-based techniques are one of the most sensitive diagnostic methods. They are very sensitive and able to detect a parasitemia level of 1-5 parasites/ μl , which is $10\times$ lower than the sensitivity of microscopy or rapid diagnostic tests. They also allow for identification of infective species, mixed infections, and drug-resistant parasites. Although PCR is an accurate tool for malaria diagnosis, it is cost- and labour-intensive and requires trained experts, which makes this method unsuitable for remote endemic areas [42].

Potential diagnostic approaches under development

Given the drawbacks of current methods used for the diagnosis of malaria, many new tools are being developed. Based on molecular biology, loop-mediated isothermal amplification, microarray, and flow cytometric assay techniques offer a way to characterise the parasite by targeting its DNA [42, 44]. Other methods exploit the optical and magnetic properties of hemozoin crystals, such as the RMOD, which is able to quantify hemozoin crystals by magnetically induced rotation [9, 45]. Further techniques are magnetic field enriched Raman spectroscopy [46], magnetic purification of infected cells [47], and the breath analyser, which is able to detect malaria-associated biomarkers in the breath of infected patients [48].

2.2.2 Neural networks as boosters for microscopic diagnosis

The main goal of NN-based algorithms developed for malaria infected RBC recognition in light microscopy images is to increase the throughput of diagnosis by reducing the image processing time and to enhance its reliability by eliminating human errors [49]. The most efficient approaches typically apply a two-step scheme, where healthy and infected RBCs are distinguished in the first step, followed by a stage-specific classification of infected RBCs in the second step [50–53]. Most of the efforts have been focused so far on Giemsa-stained images, where several attributes of the infected RBCs have been successfully exploited in the recognition process, including the

characteristics of the colour scheme [51], the morphology of the RBCs [52] and their other statistical features [53]. The accuracy of these methods ranges between ~ 80 - 98% . Similar performance levels have recently been achieved by an unbiased convolutional neural network (CNN) with minimal or no pre-processing of microscopy images [54–57].

All the above mentioned imaging techniques have two important limitations for NN-based recognition. They can provide limited training sets with typically a few thousands images, where healthy RBCs and RBCs containing the different stages are rather unevenly represented. This is due to the labour-intensive imaging and manual categorisation. Even when working with parasite-enriched cultures [58–60], the number of uninfected RBC images is approx. 10 times larger than that of the infected ones. Such imbalance in the data set can result in the so-called overfitting of the images [61, 62], lowering the accuracy of the NN-based recognition. As a common solution for this problem, data augmentation can be used to equalise the number of images in the different categories (ring-, trophozoite-, schizont-stage, and healthy RBCs) by generating more data for the training set [54, 55]. This has been demonstrated to improve the performance of the NN [56, 57] and to further increase the generalisation ability.

Even after the data augmentation process, the image set is usually too small to properly train a two-dimensional CNN for the stage-specific classification of RBCs due to the large number of parameters to be fitted by the NN. One way to overcome this challenge is to reduce the dimension of the input data, namely the dimension of the images in the present case. It is known that, under such conditions, dimension reduction may increase the performance of NNs [62]. The reduction of dimension naturally comes with a massive loss of data, unless characteristic features are selected, as was done by Das [52], Kareem [63], Pinkaew [53], and their co-workers.

Despite the challenges, the application of neural networks gains more and more attention, especially for the analysis of large data sets. Neural networks can not only speed up the data evaluation but also improve the classification compared to humans.

Applied methodology

From molecules to entire populations, biological systems present one of the most complex processes in our life, such as the origin and course of diseases. Understanding those requires detailed investigation of the underlying cellular processes. To access information on a cellular level, magnifications beyond the limitations of human vision are necessary. Light microscopy provides a way to investigate cells in a non-invasive manner. While this is a clear advantage for imaging of living systems, morphological properties remain hidden. In biological studies, not only the analysis but also the sample preparation plays an important role, as reproducible and reliable protocols for sample preparation lay the foundation for successful data acquisition. Especially in case of cellular mechanisms, culturing the biological systems in controlled conditions is essential for representative and high-quality measurements.

This chapter introduces the fundamentals of *in vitro* cultivation of malaria, which provide the basis for optical and morphological measurements by fluorescence microscopy and AFM. A general introduction to both techniques is presented. Subsequently, an overview about the basics, applications, and evaluation of neural networks is given, followed by the routines for sample preparation and measurement setup.

3.1 In vitro cultivation of *Plasmodium falciparum*

In 1976, Traeger and Jensen [64] developed a protocol for the continuous cultivation of *P.falciparum* under laboratory conditions, which was designed to reproduce the conditions in the human body during the intra-erythrocytic cycle. This method enabled the in-depth exploration of cell modifications during parasite maturation.

Figure 3.1 shows the workflow of culturing. Cultures, which are stored in liquid nitrogen for cryopreservation [65], can be initiated using preserved stocks of laboratory-adapted parasite strains or field isolates. Thawing is done by adding sodium chloride (NaCl) solutions of various concentrations to the culture in a stepwise manner. This step is performed in a flow box to ensure sterile environment. After thawing, nutrient medium and human RBCs as hosts for parasites are added to the culture. To ensure

optimal growth, the erythrocytes are kept in a shallow layer covered by Roswell Park Memorial Institute cell medium (RPMI 1640) supplemented with HEPES buffer (25 mM), human serum or serum replacements such as Albumax, Hypoxanthine for purine biosynthesis, and gentamicin. The cells are placed in a petri dish with a layer depth of approximately 2 mm. To maintain the growth of the parasites, they are incubated at 37°C in an atmosphere of 5% carbon dioxide (CO₂) and 5% oxygen (O₂) [64]. The continuous cultivation requires keeping the culture at 5% haematocrit and 1% parasitemia, where haematocrit denotes the ratio of RBCs and medium and parasitemia is a measure for the percentage of infected cells. To provide optimal parasite gain, the cells should be kept in an extracellular pH of 7.2 to 7.45 and a lactate concentration below 12 mM [66]. Monitoring the growth of the parasites is achieved through thin smearing a drop of blood from the culture on a glass slide, which is stained with KwikDiff [41], see section 2.2.1. The parasitemia is then estimated by counting approximately 20 fields of 200 cells. By calculating the percentage of infected cells in the thin blood film, the parasitemia can be adjusted to < 1% through diluting the culture with uninfected RBCs. Repeating this procedure every day ensures optimal growth of the parasites [67].

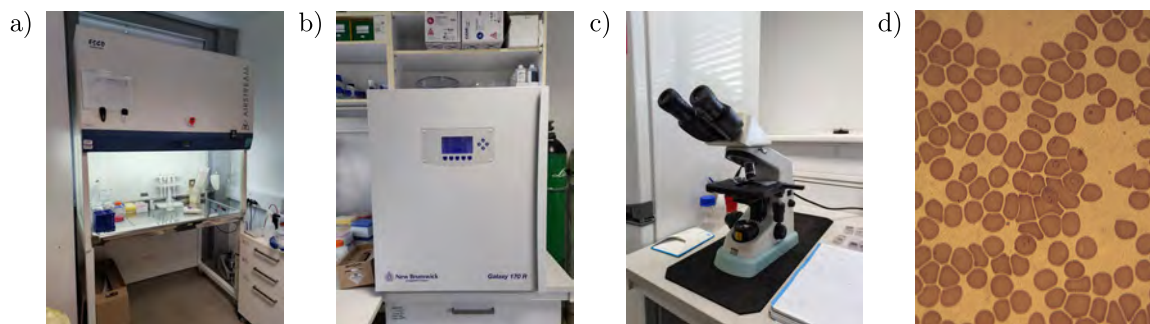


Figure 3.1: **Workflow of culturing.** The malaria parasites are stored in liquid nitrogen. To start a new culture, they are thawed in a flow box (a), which ensures a sterile environment. Optimal growth is provided by an incubator (b) at 37°C in an atmosphere of 5% CO₂ and 5% O₂. To check the state of the culture, thin blood films are investigated under a light microscope (c), showing the healthy and infected RBCs (d).

To preserve cultures, the parasitised RBCs can be frozen with the Stockholm Sorbitol method. This method works best with a parasitemia of > 5%. The erythrocytes are centrifuged and the remaining supernatant discarded. Then, freezing solution (28% glycerol, 3% sorbitol, 0.65% NaCl, distilled water) is added drop-wise to the pellet and the resulting solution is stored in a liquid nitrogen tank. To reuse the frozen cultures, they are thawed and then diluted with two NaCl solutions consecutively 1.6% NaCl and 12% NaCl. After washing the blood with culture medium, the solution is centrifuged and the remaining pellet resuspended in a flask, ready to use for cultivation [68].

In most experiments, the synchrony of malaria parasites plays an important role. The most common method is the use of sorbitol treatment. By adding a 5% sorbitol solution (in distilled water), the RBCs infected with late stages are lysed, while the RBCs infected with ring stages are preserved. Centrifuging and resuspending the pellet in medium returns a synchronised culture with parasites in the ring stage [60]. To separate the late stage parasites, the cells are sorted by sedimentation in Percoll [58, 59]. The culture is layered on top of a 70% Percoll solution in RPMI 1640. After centrifugation in tubes, the fractions of erythrocytes are collected and evaluated. Since the density of parasitised cells varies with the maturity of the parasite, the schizont stage can be recovered from the top layer, while the ring stage and healthy cells remain in the bottom of the tube. The isolated late schizont-stage parasites are then washed with culture medium and diluted with 50% uninfected RBCs to a 5% haematocrit culture [67].

3.2 Atomic force microscopy

Studying the intracellular structures of biological samples requires high resolution in the range of a few hundred nanometers. Among other methods, e.g. transmission electron microscopy [69] or scanning near-field optical microscopy [70, 71], AFM is a common tool to investigate the morphology and mechanical properties of RBCs [12, 14, 72]. AFM relies on the interaction between the sample surface and a sharp tip, which also qualifies this technique for other applications in physics, chemistry and nanotechnology [73–75].

The fundamental concept of AFM is based on measuring atomic forces between probe and sample surface. This is achieved by scanning the surface with a cantilever of known spring constant, which acts as a force sensor over a sample. The deflection of the cantilever is proportional to the measured forces. One of the most common methods to detect the deflection is by a laser beam [76]. By focusing a laser on the backside of the cantilever, the reflected light is monitored using a segmented photodiode via the relative photocurrent produced in the segments, see figure 3.2. This method allows for distinguishing between lateral forces (torsion) and deflection (up or down) of the cantilever.

To move the tip across the sample surface, piezoelectric actuators are used. There are three actuators, two for scanning in xy plane and one for regulating the z height. The sample can be either positioned on a piezo tube or a piezo-stack. An additional piezo actuator drives the oscillation of the cantilever.

3.2.1 Interactions between tip and sample

AFM enables the measurement of forces between atoms at the tip and the sample, which can be classified into attractive and repulsive. Among the attractive forces are van der Waals interaction, electrostatic, and chemical forces, while the repulsive forces

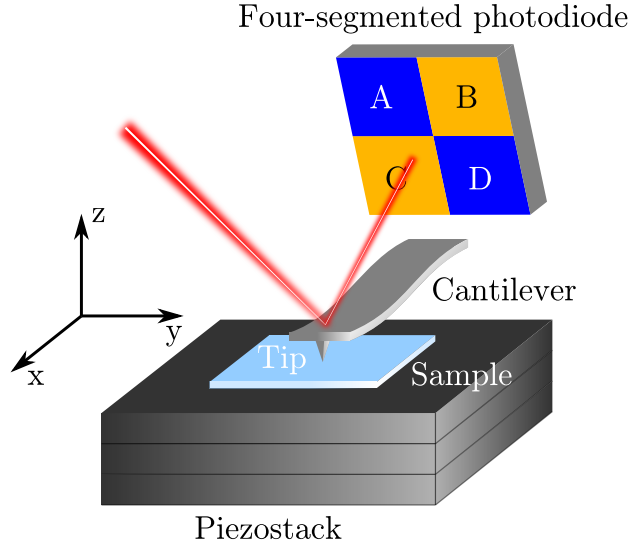


Figure 3.2: **Schematic setup of AFM.** The cantilever is scanned over the sample surface by piezo elements. Interaction between sample and tip is monitored by beam-reflection. Reproduced from [77].

result from Pauli-exclusion and electron-electron Coulomb interaction. Generally, these are short-ranged and decay exponentially with distance. When the AFM is used in liquid, capillary forces contribute to the attractive forces. Assuming that the AFM is operated in vacuum and only considering two neutral atoms, the interaction between tip and sample is defined by the Lennard-Jones potential.

$$V_{\text{Lennard-Jones}} = -4\epsilon \left(\frac{\sigma^6}{z^6} - \frac{\sigma^{12}}{z^{12}} \right), \quad (3.1)$$

where ϵ is the depth of the potential well and σ the distance at which the force is zero. The first term results from the van der Waals force due to interaction between dipoles and chemical interactions, while the other originates from repulsive forces caused by the overlapping of electron clouds of the tip and sample atoms. Figure 3.3 shows the tip-sample interaction and the contributions from the attractive and repulsive forces individually [78, 79].

Upon approaching the surface, the tip is pulled down due to the attractive forces. When it is close enough, the tip jumps into contact. Moving closer to the surface and into the repulsive regime bends the cantilever away from the surface. Upon retracting, the tip remains in contact with the surface until the elastic force of the cantilever overcomes the attractive forces between tip and sample [78]. These characteristics form the basis for the two types of AFM: static and dynamic. While the static type is operated in contact mode, in case of dynamic AFM, the cantilever is oscillated at its resonance frequency f_0 in non-contact mode. In contact mode, the interacting force between tip and sample is proportional to the cantilever deflection. In case of soft samples, e.g. biological materials, the resolution is compromised by elastic and

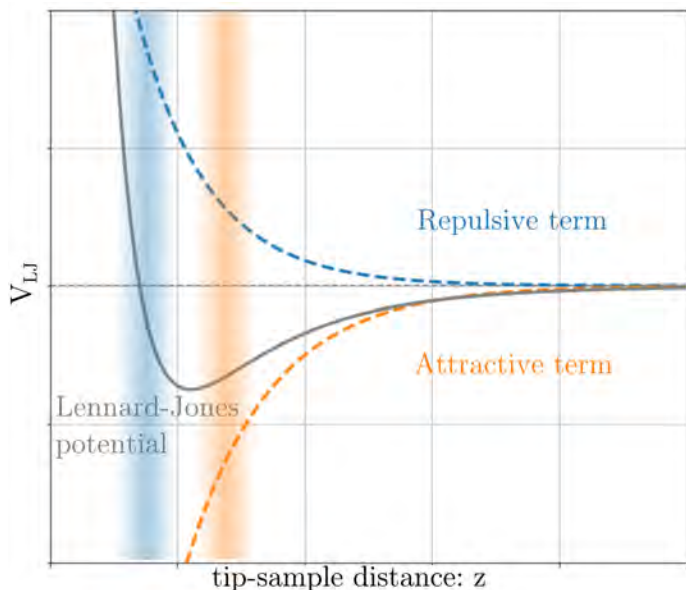


Figure 3.3: **Tip-sample interaction.** Schematic plot of the Lennard-Jones potential with attractive and repulsive contributions. In the attractive regime (orange), the AFM is operated in non-contact mode and in the repulsive regime (blue) in contact mode.

plastic deformations [80]. This limitation is overcome by non-contact mode AFM. The cantilever is vibrated at a fixed frequency, which is close to f_0 . Upon approaching the sample, the interactions between tip and surface cause changes in amplitude, phase, and resonance frequency of the cantilever, which are used as a feedback signal for topographic imaging. This mode enables imaging with atomic resolution, which can be limited by the geometry and chemical properties of the tip [81].

In static AFM, the measurement is either controlled by height or force. If the height is kept constant during the measurement, the deflection Δz is monitored, which can be directly translated into the interaction force F_{hook} by Hooke's law

$$F_{\text{hook}} = k \cdot \Delta z, \quad (3.2)$$

where k denotes the spring constant of the cantilever. The constant height mode enables high-speed atomic resolution images. In constant force mode, the height of the sample is adjusted by a feedback loop during scanning to maintain a constant deflection of the cantilever. While providing not only information about topography but also other characteristics, e.g. friction forces, the response time of the feedback system imposes limitations to the scanning speed.

For non-contact measurements, the AFM is generally used in dynamic mode, which is further divided into two modes, amplitude-modulation or tapping mode and frequency-modulation mode, depending on the feedback parameter. In tapping mode, the tip is moved close enough to the surface to be partially influenced by the repulsive

regime, leading to a reduction in amplitude, which is then adjusted to be constant by a feedback circuit. The interactions between tip and sample shift the resonance frequency of the cantilever, which is proportional to the gradient of the interaction force F_{int} , f_0 , and k

$$\Delta f = -\frac{f_0}{2k} \frac{\partial F_{\text{int}}}{\partial z}. \quad (3.3)$$

Furthermore, the induced changes in phase describe if attractive or repulsive forces govern the interaction between tip and sample, which qualifies the phase as an indicator for the current mode of the system. This mode enables qualitative imaging of a large variety of surfaces. While amplitude-modulation is unsuitable for measuring in vacuum, as the feedback is too slow, frequency-modulation reacts instantaneously to tip-sample interactions. In this mode, the cantilever is oscillated at its resonant frequency. When a shift is induced by forces between tip and sample, the probe-sample distance is adjusted, which allows for measuring the chemical bond formation of single molecules [81, 82]. The sensitivity of the method depends on the Q-factor, which is a measure for the changes in amplitude due to the frequency shift. In response to variations in sample topography, a cantilever with a high Q-factor causes a larger change in amplitude than a low Q cantilever, which increases the sensitivity, allowing high resolution imaging. Working with high Q cantilevers is a compromise between fast scan speeds and resolution [83, 84].

3.2.2 Probe characteristics

The accuracy of AFM measurements depends on the shape and spring constant of the cantilever probing the sample. Probes are generally microfabricated from silicon wafers by lithography [85]. The spring constant is defined by the geometry of the cantilever and ranges from 0.1-100 N/m, which also influences the Q factor and hence plays a crucial part for the sensitivity and resolution. Figure 3.4 shows the probe geometry of an OTESPA-R3 cantilever, which was used for the AFM measurements performed in this work.

Spring constants are chosen slightly stiffer than the stiffness of the probed sample to prevent adhesion of the cantilever. For cell studies, the cantilever stiffness typically ranges between 0.01 - 0.6 N/m [88].

Resolution is further limited by size and shape of the tip, e.g. if the tip width is larger than the lateral dimensions of surface features, which causes artefacts due to the dilation of the sample surface. Optimum tip geometries depend on their similarity to the measured surface. In case of biological samples, such as cells, the choice of tips is determined by the aim of the measurement. For understanding the cellular mechanical response, large colloidal probes are preferable, while sharp tips enable high-resolution mapping of cell topography and mechanics [89].

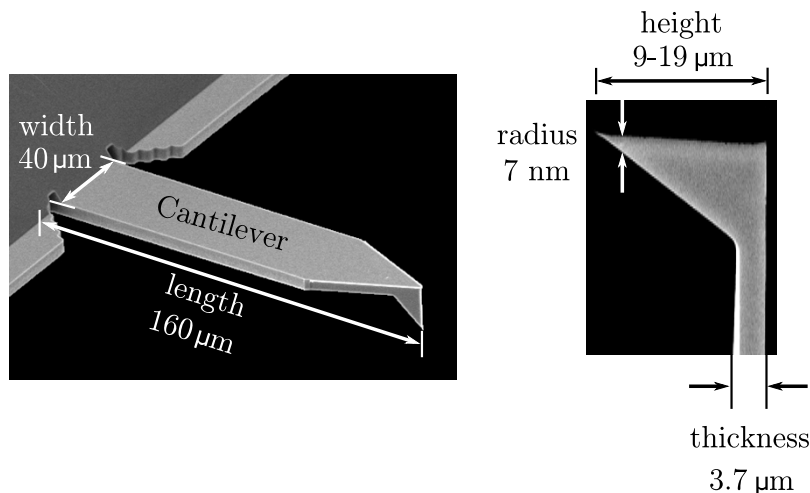


Figure 3.4: **Typical cantilever geometries.** Dimensions of the OTESTPA-R3 cantilever tip. Reproduced from [86,87]

3.2.3 Topography measurements

High-resolution topography images are acquired by smart combination of the parameters mentioned above. In case of biological samples the topography is measured by tapping mode, which reduces the interaction between sample and tip, preserving soft matter, such as cells. The drive frequency of the cantilever during the measurement is determined from the resonance frequency of the cantilever in air, as shown in figure 3.5. For imaging in fluid environment, the resonance frequency is determined by applying a range of oscillation frequencies to the cantilever. The frequency with the largest response is then the resonance frequency. For amplitude-modulated mode imaging, the drive amplitude is used as feedback parameter for the tip-sample distance and remains constant during the scan. The tip is approached towards the surface until it reaches the set point amplitude. Deviations in amplitude are recorded as error signal. Additionally, the phase monitors if the interaction is governed by attractive or repulsive forces.

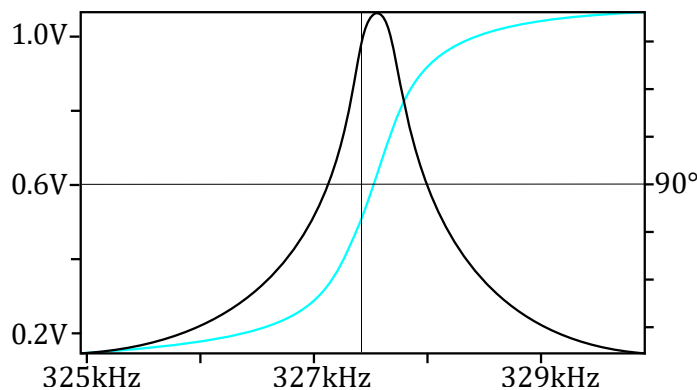


Figure 3.5: **Resonance frequency in air.** The resonance frequency is denoted by a peak, and the black, vertical line shows the drive frequency. The teal line shows the phase, which is set to 90° at the resonance frequency.

3.3 Fluorescence microscopy

In the field of biology and medical sciences, fluorescence microscopy offers a non-invasive way to investigate biological matter, e.g. cells or tissue. A wide range of organic fluorescent dyes and proteins enables sensitive and selective imaging. Due to the large spectral range of fluorophores, simultaneous imaging of multiple cellular, subcellular, and molecular components is possible, and even imaging of living systems.

3.3.1 Basic principles of fluorescence emission

Fluorescence is the emission of light due to absorption of light with typically shorter wavelength. By filtering the excitation wavelength, the contrast between fluorescing objects and background is strong enough to visualise even single fluorescent molecules. Organic substances with intrinsic fluorescence (autofluorescence) are called fluorophores. When a fluorophore is hit by a photon, all of its energy is absorbed and the molecule is excited from its ground state to higher energy levels, which include triplet and singlet states. In the triplet state, the spin orientation is flipped upon excitation, while it is preserved in the singlet state. Depending on the wavelength, the electrons of the fluorophore can be excited to one or more energy levels. The absorbed energy is released by fluorescence emission, which takes a few nanoseconds. The physical process can be illustrated by the Jablonski diagram [90, 91], see figure 3.6. The absorption process spans a time range of 10^{-18} s. In its excited state, the molecule loses energy by transitioning to the lowest vibrational level. This vibrational relaxation requires approximately $10^{-12} - 10^{-14}$ s. Energy loss can also occur through intersystem crossing from singlet to triplet state. In the latter case, when the molecule relaxes to its ground state, the light emission is referred to as phosphorescence, which occurs within a timescale of $10^{-3} - 1$ s [92]. Internal conversion is much slower and results in the relaxation of the molecule into the lowest excited state or into the lowest

ground state. Due to the energy loss caused by vibrational relaxation or intersystem crossing, the wavelength of the emission is higher and the frequency lower relative to the absorption (Stokes shift) [93].

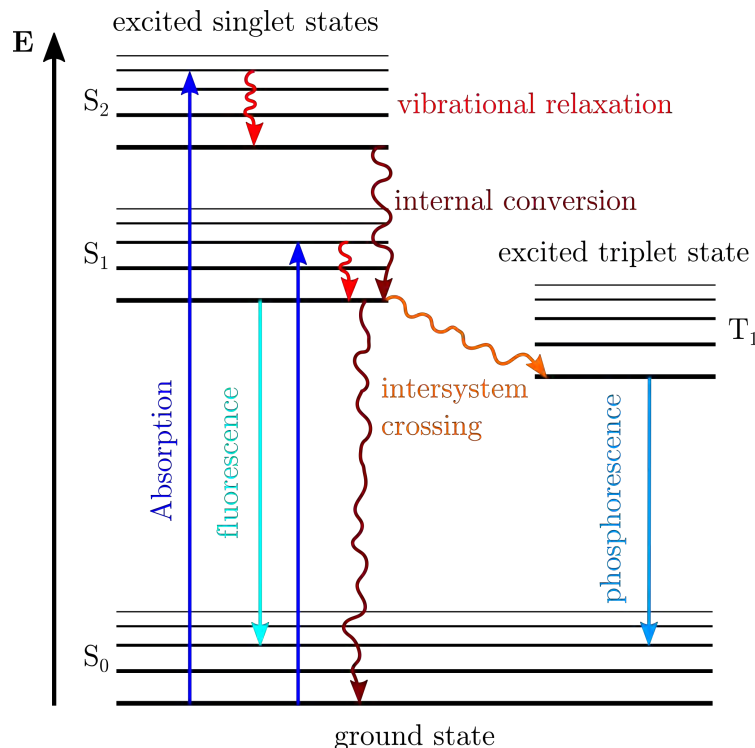


Figure 3.6: **Jablonski diagram.** The energy states of a molecule are divided into the ground state S_0 and excited singlet S_1 , S_2 and triplet T_1 states. Upon absorption, the energy is either re-emitted as fluorescence or phosphorescence or dissipated by relaxation processes (vibrational relaxation, internal conversion or intersystem crossing).

3.3.2 Fluorescence microscopy techniques

The application of fluorescence microscopy techniques depends on their field of use. While wide-field microscopy is fast, allowing direct image observation, image acquisition with confocal microscopes is time-consuming, but contrast and resolution are improved [94]. In wide-field microscopy, the sample is completely exposed to the light source. In confocal microscopy, the specimen is illuminated with a focused scanning laser beam. Additionally, a pinhole aperture is placed in the image plane in front of an electronic photon detector. The incoming light signal is converted into a continuously changing voltage by the detector, see figure 3.7 [95].

The wide-field microscope is equipped with a fluorescence objective, a system of emission and excitation filters, a dichromatic mirror and a light source. After passing through an excitation filter, the light beam from the source is refracted on a dichroic

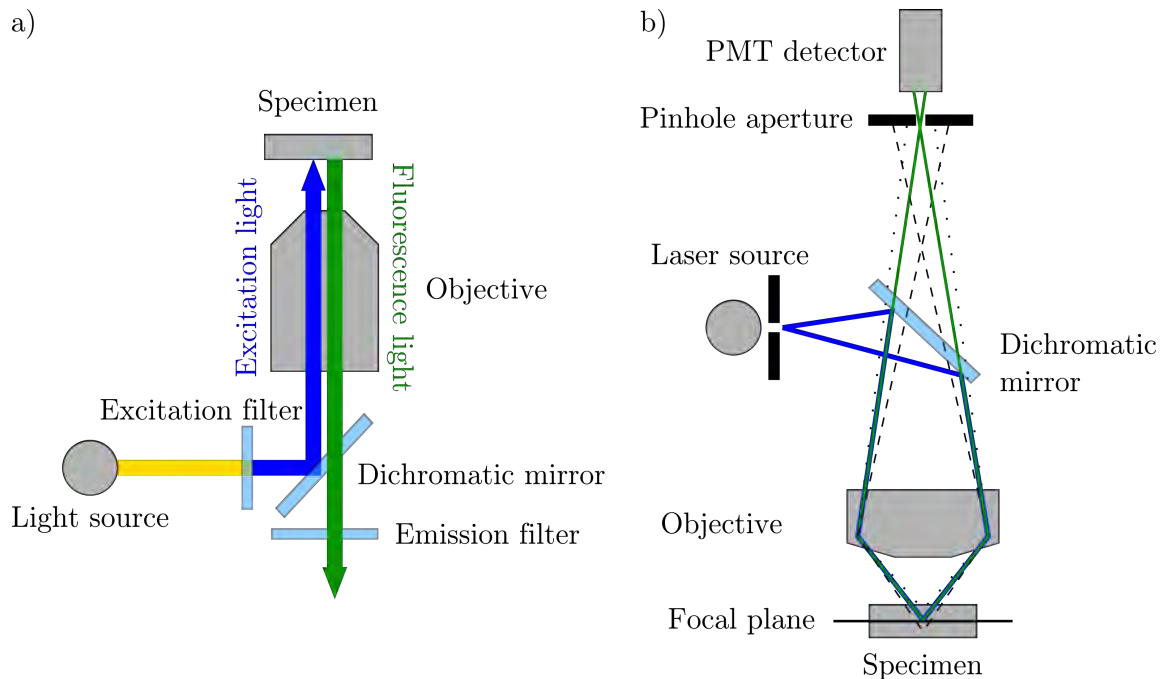


Figure 3.7: **Illumination in wide-field and confocal fluorescence microscopy.** (a) Light path in wide-field microscopy, illuminating the whole sample. (b) Illumination of the specimen with a focused laser beam through a pinhole aperture by using confocal microscopy, which filters out fluorescence light emitted below and above the focal plane. Reproduced from [95,96].

mirror, which is oriented in such a way that the emission wavelengths pass through, while the excitation wavelengths are reflected. The fluorescence signal emitted from the sample is collected by a camera. This setup has some inherent limitations. As the entire sample is illuminated, regions above and below the focal plane are excited by the light source. The additional background fluorescence decreases resolution and contrast in the image [96]. In confocal microscopy, out-of-focus light is removed by a pinhole aperture, see figure 3.7(b). The sample is excited by a laser beam spot, which is moved over the sample in a raster. The emitted fluorescence signal is spatially filtered by the pinhole aperture and received by a photomultiplier tube (PMT) detector for different fluorescent wavelengths [95]. Despite the improved image quality, raster-scanning comes with long exposure times, which can influence the fluorescence response of the sample, e.g. by photobleaching, and increases the image acquisition time compared to wide-field microscopy [97].

3.4 Neural networks

Machine learning algorithms are inspired by the human brain. They are designed to find features in a data set and to generalise from these experiences, such as comparing cell images for the diagnosis of diseases in blood samples. The term machine learning

is used for a process describing the field of predictive modelling, where the goal is to make the most accurate prediction with the help of algorithms from many different fields, such as statistics. These search for relationships between variables and predict an outcome based on these connections. One prominent example used for data sets with high complexity are deep learning algorithms. They are widely used in the fields of computer vision, speech recognition, and bioinformatics, where they are able to produce results comparable to human experts and even to surpass human performance.

3.4.1 The process of machine learning

The development of a machine learning model requires the assignment of possible data sources and conditions, as well as the choice of suitable algorithms for learning. These are divided into four categories: supervised, unsupervised, semi-supervised, and reinforcement learning. In supervised models, the algorithm develops a prediction based on predefined samples, while it autonomously clusters the data in unsupervised models. As a combination of both methods, in semi-supervised learning only a part of the data is predefined and the algorithm clusters the rest based on the predefined samples. The fourth method is inspired by human learning, as the learning process is controlled by reward and punishment, teaching the algorithm a strategy to react in known and unknown situations.

As not all data is fit to be used as input for a machine learning algorithm, it requires modifications in order to be applicable for models, such as NNs. In most cases, the data has to be extracted from data bases, signals, images etc., to be converted into numerical input for the network. These kind of data normally contain a high amount of features and hence high dimensionality, such as 2D RBC images, which requires measures to reduce the dimension in the data set. In case of working with more-dimensional data, the “curse of dimensionality” often degrades the performance of the network. If the data set contains too many dimensions, and therefore too many features, the network is not able to form meaningful clusters. One way out of this dilemma is dimensionality reduction, which refers to methods for projecting the data to a lower dimensional subspace, while still capturing the essence of the data [62].

The performance of the machine learning model not only depends on the dimensionality but also on the amount and distribution of the input data. In case of supervised learning, where the algorithm detects features based on a predefined categorisation of the data, these categories should be evenly distributed to guarantee accurate predictions on unknown data. To increase the data set or to equalise the number of images for each category, data augmentation can be employed, which generates new data from already existing data by applying minor geometric transformations, such as rotation or shearing. This step not only improves the performance and results of the machine learning model but also saves time in terms of data collection and labelling. Especially in case of NNs, the size of the input data set plays an important

role for accurate predictions. One of the most popular data augmentation methods is position augmentation. It includes operations like cropping, reflection, rotation, and scaling. While there are many ways to increase the data set, the application of data augmentation requires careful selection of operations as some are not fit for a specific set of images [98].

3.4.2 Principles and architectures of neural networks

The idea of machine learning laid the foundation for artificial neural networks (ANNs), which are based on the processes observed in biological NNs, where the brain is the control unit with different subunits for vision, hearing, movement and senses. Through the brain, the human body reacts to stimuli, which are processed by the nervous system. It contains approximately 10^9 neurons, receiving and transmitting the stimuli as signals. Each neuron consists of three major parts, a cell body with extensions, the dendrites, and the axon, see figure 3.8.

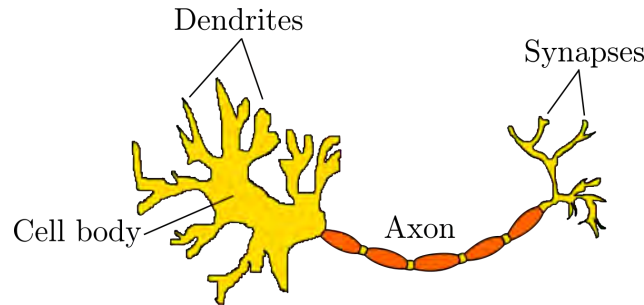


Figure 3.8: **Components of a biological NN.** The cell body of the neuron receives signals from the dendrites. Received information is transmitted as electric signals by axons, which are connected to other cells by synapses. Reproduced from [99]

The signals are collected by dendrites. Informations between the nervous cells are transferred by electric impulses through the axon, which are connected to other cells by synapses [100]. The concept of information transmission between neurons has formed the basis for creating mathematical models for ANNs, designed to learn and adapt to situations and inputs.

The predecessors to artificial neurons are perceptrons, which imitate the biological neurons using mathematical operations. In this simplified model, the perceptron receives an input information \mathbf{x} , where x_i indicates the i -th element of the vector \mathbf{x} , through the axon connection between neurons. Each input element is multiplied by weights w_i , which define the importance of the connections. The weighted inputs are then given to the cell body by the dendrites, where all contributions are summed up with an additional weight b (bias). The sum $s = \sum_i w_i x_i + b$ is translated into a binary output y by the activation function a , see figure 3.9.

$$y = b + a \left(\sum_i w_i x_i \right) \quad (3.4)$$

This process is evaluated by comparing the output with the desired value d . The difference between both scalar values (y and d) is used to adapt the weights accordingly. With these basic concepts, perceptrons are able to perform logic operations, such as AND, OR, and NAND [101].

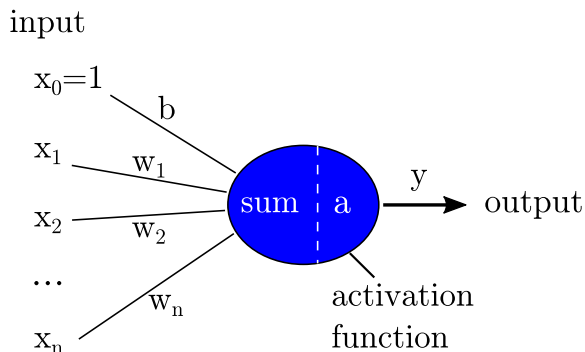


Figure 3.9: **Schematics of a simple perceptron.** The input x_i is provided to the neurons by the axons, which are connected to the dendrites by synapses. The connections between neurons are weighted and fed to the cell body. The sum of all connections is then converted to an output by an activation function. Reproduced from [99, 102]).

In case of samples, which are not linearly separable, the perceptron is not able to divide the data into two classes. A possible solution for this problem is the combination of multiple perceptron layers, which contain an input layer, an arbitrary number of intermediate or hidden layers, and an output layer [103]. This new approach requires a learning rule to repeatedly adjust the weights of the connections between neurons in the network in order to minimise the difference between the actual output y and the desired output d . This difference is defined as the total error E :

$$E = (y - d)^2, \quad (3.5)$$

where c runs over all input samples and j represents the number of output neurons. The minimisation of E is performed by gradient descent, which computes the partial derivative of E with respect to each weight in the network. These are calculated in two ways: forward and backward. In the first step, y is determined by giving the input x_m from one layer to the next, see figure 3.10. In the first layer, f_k is defined as

$$f_k = t_{k0} + a \left(\sum_m t_{km} x_m \right), \quad (3.6)$$

where t_{k0} is the bias and x_m the input. In the next layer, g_j is calculated by

$$g_j = u_{j0} + a \left(\sum_k u_{jk} f_k \right) \quad (3.7)$$

with the weight u_{jk} and bias u_{j0} , then h_i by

$$h_i = v_{i0} + a \left(\sum_j v_{ij} g_j \right) \quad (3.8)$$

and finally y by

$$y = w_0 + a \left(\sum_i w_i h_i \right). \quad (3.9)$$

The difference E between y and the desired output d (Equation 3.4.2) is then used to calculate the error on each layer, which is expressed by its partial derivative in the respective layer

$$\frac{\partial E}{\partial g_j} = \sum_i a'(h_i) v_{ij} \frac{\partial E}{\partial h_i} \quad (3.10)$$

and equivalent for the other layers in figure 3.10. $\frac{\partial E}{\partial g_j}$ shows if g_j should be higher or lower and if h_i was too high or too low. With the partial derivative the error in the weight affecting g_j is computed:

$$\frac{\partial E}{\partial u_{jk}} = \frac{\partial E}{\partial g_j} a'(g_j) f_k \quad (3.11)$$

In the final step, the weight is updated by

$$\Delta u = -\eta \frac{\partial E}{\partial u_{jk}}, \quad (3.12)$$

where η describes the learning rate [104]. In this way the error propagates backwards through the layers, as denoted by the blue arrows in figure 3.10.

The concept of backpropagation has some limitations, as the partial derivative of the error is calculated separately for each input sample. For a large data set, this is quite time consuming. The stochastic gradient descent provides a solution for this problem. Computing the gradient for a small sample of randomly chosen inputs gives an estimation of the true gradient and speeds up the learning. When the algorithm has run through all the input data, one epoch of training is completed [106].

One important element of neural networks is the activation function, which was introduced for the perceptron but not discussed in detail. It serves as a measure for the importance of a connection between neurons and decides, which neurons are activated. Depending on the complexity of the problem given to the network, the activation function can be a binary step function, a linear activation function or non-linear, see figure 3.11. The most common non-linear functions are sigmoid, tanh, ReLU, and softmax. While the sigmoid function is applied for classification problems, which require the probability of the respective category as output, the tanh function is mainly used for binary problems. In contrast to sigmoid and tanh, the rectified linear unit (ReLU) does not activate all neurons simultaneously. The function is not

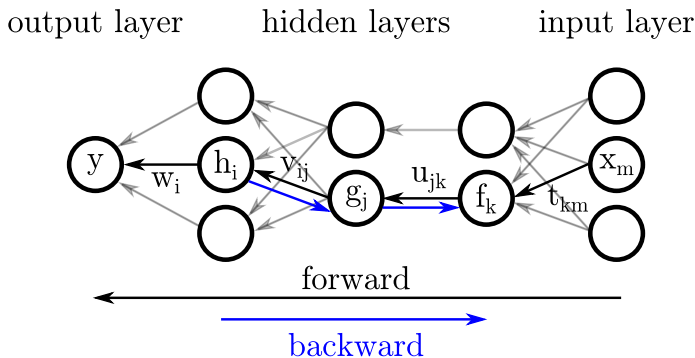


Figure 3.10: **Schematics of backpropagation in a neural network.** Multilayered network with output layer, three hidden layers, and input layer. The grey arrows show the connections between the neurons and the black arrows a sample path for the forward calculation of weights and output. The path of the backpropagation for adjusting the weights based on error minimisation is shown by blue arrows. Reproduced from [105].

only more efficient but also accelerates the learning process, qualifying this function for use in CNNs or deep learning. In case of multi-class classifications, the softmax function – a combination of multiple sigmoids – maps the output as a probability distribution for all possible classes [107, 108].

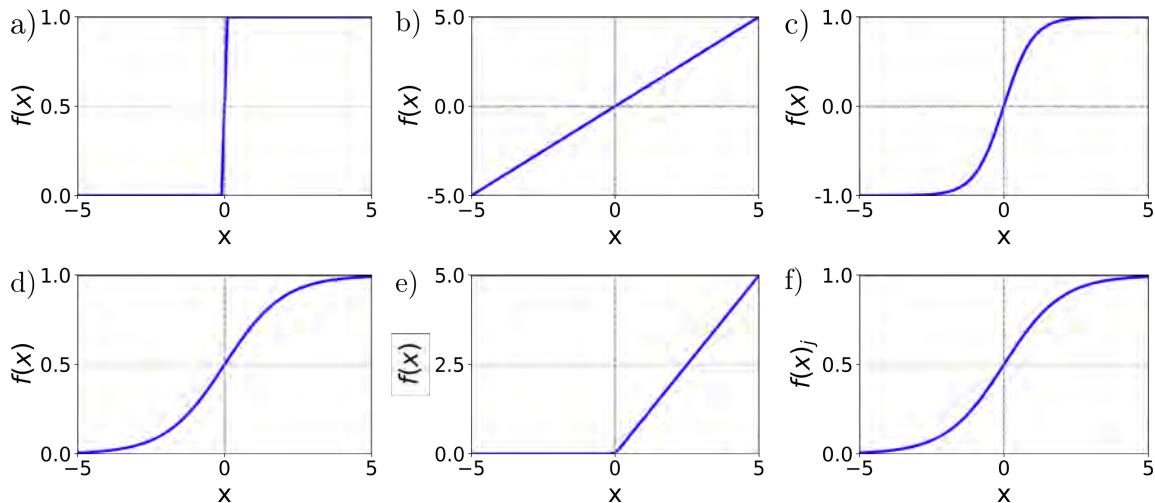


Figure 3.11: **Common activation functions.** (a) Binary step function, (b) linear activation function, (c) sigmoid activation function, (d) tanh activation function, (e) ReLU, and (f) softmax.

When working with NNs, a typical problem is overfitting, which describes the inability of the network to generalise, i.e. the network performs well on the training set but fails on unknown data. The easiest way to overcome this is to increase the training data set, which is not always feasible, if the data is difficult to acquire or expensive.

Another possibility is to introduce regularisation techniques, which add an extra term to the error E , such as $L2$ regularisation:

$$\frac{\lambda}{2n} \sum_w w^2, \quad (3.13)$$

where λ is the regularisation parameter, n the size of the input data set, and w the weight. The idea is to make the network learn on small weights, only allowing large weights, if they improve the cost function, which is regulated by λ [109]. The concepts of back propagation and refinement of the learning process by using suitable error and activation functions pave the way for more complex networks with multiple layers of hidden neurons, such as deep neural networks. Due to their complexity, they are able to understand edges, simple geometric shapes, and can learn hierarchies of knowledge, which is especially useful for image recognition.

3.4.3 Convolutional neural networks

The most prominent example of deep neural networks are CNNs. Their purpose is to learn features from a sample by applying convolving filters to the input, which are connected to a small region in the sample. For image recognition, the CNN is composed of several filters, the feature maps, each reacting to a different property. Each neuron in the feature map is connected to a field in the image, which enables the network to find features, independent of their scaling factor or location. This concept is inspired by the visual system of monkeys and allows the network to learn how to see [110]. To improve the efficiency of the CNN, pooling layers are used between convolutional layers. Their purpose is to return information about the presence of a feature, while discarding its exact location, which reduces the dimensionality. After the network has successfully learned the features, their representations are fed to fully-connected layers for classification, where each neuron from one layer is connected to the next, see figure 3.12.

While CNNs provide a stable and fast way for image recognition, the performance of the algorithm is compromised by vanishing or exploding gradients, i.e. the speed of learning decays towards earlier hidden layers or by large error gradients, which in turn results in an unstable network [112]. Despite these problems, deep neural networks have gained importance in a wide range of applications, especially in diagnosis of diseases, such as malaria.

3.4.4 Evaluation of neural networks

One way to evaluate the quality of a NN is the confusion matrix, which displays information about the actual and predicted classifications of a model. It shows the distribution of correctly and incorrectly classified data. In case of a binary problem, the data is divided into true positive (TP) for the number of correct predictions and false positive (FP) for incorrect predictions for the positive output, true negative

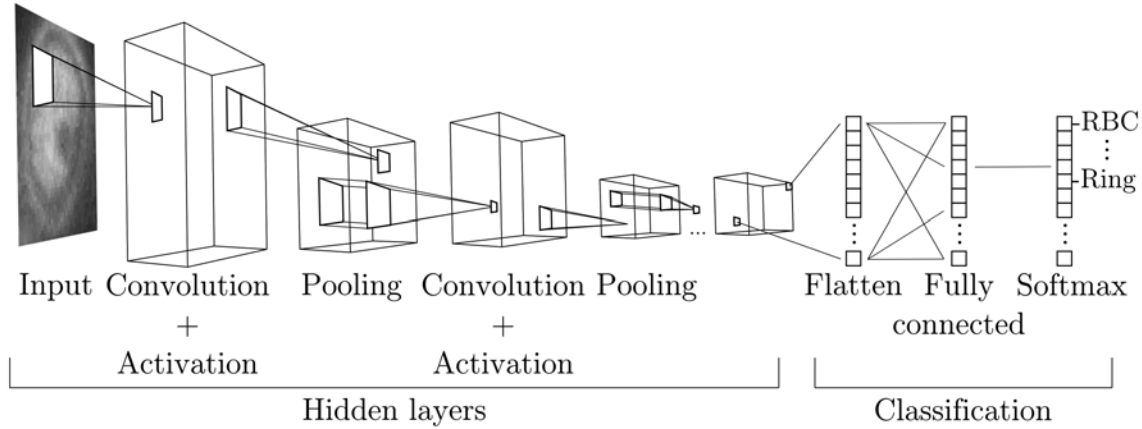


Figure 3.12: **Representation of a CNN.** The network learns characteristic features of the the input image by convolution, which are classified by a regular NN. Reproduced from [111].

(TN) for correct predictions, and false negative (FN) for incorrect predictions for the negative output. To represent the performance of NNs, *precision* and *recall* are introduced. The *precision* is defined as the ratio of TP and all results classified as positive, while *recall* is a measure of the ratio of TP and all predictions for the positive output [113, 114].

$$precision = \frac{TP}{TP + FP}. \quad (3.14)$$

$$recall = \frac{TP}{TP + FN}. \quad (3.15)$$

		Assigned classes	
		<i>True positive (TP)</i>	<i>False positive (FP)</i>
Predicted classes	<i>False negative (FN)</i>		<i>True negative (TN)</i>
	<i>True positive (TP)</i>		

Figure 3.13: **Confusion matrix for a binary classification problem.** The fields in the table show the number of data denoted as TP, FP, TN, FN.

For multi-class problems, the definitions from eq. 3.4.4 and eq. 3.4.4 are modified to

$$precision_i = \frac{C_{ii}}{\sum_{j=1}^N C_{ij}} \quad (3.16)$$

$$recall_i = \frac{C_{ii}}{\sum_{i=1}^N C_{ij}}. \quad (3.17)$$

The overall accuracy of the classification is calculated as the ratio of correct predictions, independent of the categories, to the total of values in the confusion matrix [115].

$$accuracy = \frac{\sum_{i=1}^N C_{ii}}{\sum_{i=1}^N \sum_{j=1}^N C_{ij}}, \quad (3.18)$$

where N the number of classes, i the row index and j the column index. Figure 3.14 shows a four category classification, containing the classes A, B, C, and D, which is later used in the thesis, as confusion matrices are a powerful tool for the evaluation of the performance of the NN on the data sets.

		Assigned classes				
		A	B	C	D	
Predicted classes	A	C_{11} r_{11}	C_{12} r_{12}	C_{13} r_{13}	C_{14} r_{14}	C_{11} $\frac{C_{11}}{\sum_{j=2}^4 C_{1j}}$
	B	C_{21} r_{21}	C_{22} r_{22}	C_{23} r_{23}	C_{24} r_{24}	C_{22} $\frac{C_{22}}{\sum_{j=1}^4 C_{2j}}$ $j \neq 2$
	C	C_{31} r_{31}	C_{32} r_{32}	C_{33} r_{33}	C_{34} r_{34}	C_{33} $\frac{C_{33}}{\sum_{j=1}^4 C_{3j}}$ $j \neq 3$
	D	C_{41} r_{41}	C_{42} r_{42}	C_{43} r_{43}	C_{44} r_{44}	C_{44} $\frac{C_{44}}{\sum_{j=1}^3 C_{4j}}$
		C_{11} $\frac{C_{11}}{\sum_{i=2}^4 C_{i1}}$	C_{22} $\frac{C_{22}}{\sum_{i=1}^4 C_{i2}}$ $i \neq 2$	C_{33} $\frac{C_{33}}{\sum_{i=1}^4 C_{i3}}$ $i \neq 3$	C_{44} $\frac{C_{44}}{\sum_{i=1}^3 C_{i4}}$	$\frac{C_{11}}{\sum_{i=1}^4 \sum_{j=1}^4 C_{ij}}$
		Recall				

Figure 3.14: **Confusion matrix for a four category classification.** The diagonal elements correspond to the number of objects correctly classified by the NN, while the off-diagonal elements show false classifications. In each field, the number of counts as well as the corresponding percentage with respect to the total set r_{ij} are shown. *Precision* and *recall* are displayed in the green column and the row. The *accuracy* is displayed in the grey field in the bottom right corner.

3.5 Experimental equipment and methods

The optical and morphological measurements of RBCs in this work were performed with several microscopes. In the following, each microscope and the setup of the experiment are described.

AFM All morphological measurements of RBCs on Giemsa-stained and unstained thin blood films were carried out with a MFP-3D AFM (Asylum Research, Oxford

Instruments). The set up and acquisition of the AFM was performed with using Igor Pro 6.37. For the morphological measurements, the AFM was operated in AC mode with amplitude as feedback parameter (Amplitude-modulation) in air at a scanning speed of 0.25 Hz. Each scan was carried out on an area of 90 x 90 μm with a resolution of 512 x 512 pixels. For the measurements, an OTESPA-R3 cantilever from Bruker with a rectangular shape, a tip radius of 7 - 10 nm, and a spring constant of 26 N/m, was operated at a frequency of 280 - 300 kHz with a drive amplitude of 250 - 300 mV. To prepare the images for further analysis, they were preprocessed by using the image analysis provided by the Igor Pro software. By applying the "Flatten" function in first order, the background tilt was removed, increasing the contrast of the images. The height information was then saved as a matrix in text format with corresponding x- and y-values. The measurements of RBCs in liquid were performed with the same AFM in AC mode in phosphate-buffered saline (PBS) at a scanning speed of 0.25 Hz. Images were obtained by operating a BL AC 40 cantilever from Bruker with a rectangular shape, a tip radius of 8 - 15 nm, a spring constant of 0.09 N/m, at a frequency of 38 - 41 kHz with drive amplitude of 300 - 600 mV.

Fluorescence microscope The emission of autofluorescence in malaria-infected RBCs, was observed on thin blood films on coverslips (nominal thickness 150 μm). The samples were excited by an OMICRON (Rodgau-Dudenhofe, Germany) LaserHub, which contained four independent laser sources (405 nm 120 mW CW diode, 488 nm 200 mW CW diode, 561 nm 156 mW CW diode, 642 nm 140 mW CW diode), operated at 5 - 20% power. Wavelength and power were controlled by the OMICRON Control Center software (v.3.3.19). To record the fluorescence microscopy images, an Olympus IX81 inverted microscope combined with Olympus TIRF objective (UApo N, 100x, 1.49NA) was used. The RBCs were observed through immersion oil. An EM-CCD camera (Andor iXon DU-885KCSO-VP, Oxford instruments) collected the images through a quad-band dichroic mirror and emission filter set (TRF89901-EM-ET-405/488/561/640, Chroma Technology, Bellows Falls, VT USA). For high resolution, the exposure time of the samples was set to 0.5 s. To identify the stage of the RBCs, each image was supplemented by a light microscopy image of the same area.

Light microscope For the investigation of Giemsa-stained thin blood films, a Nikon Eclipse E 100 with a Nikon BE Plan 100x 1.25 NA oil objective was used. The stained thin blood films were prepared following the standard procedure [5].

Maintenance of *P. falciparum* cultures b The cultures of *P. falciparum* parasites, used for the experiments described in this work, were from the laboratory adapted strain 3D7, which were provided by Dr. Kai Wengelnik, University of Montpellier, Montpellier, France. As the parasites require host erythrocytes, RBC concentrate was purchased from the National Blood Services of Hungary. Following the recommendations of the Malaria Research and Reference Reagent Resource Center [116], the culture medium was prepared by supplementing RPMI 1640 (Gibco, thermo Fisher

Scientific, Waltham, MA USA) with Albumax I (ThermoFisher Scientific, Waltham, MA, USA), Hypoxanthine and Gentamicin (VWR International, Radnor, PA, USA). Before use, the medium was sterile filtered. The detailed composition is shown in table 3.1. For continuous growth, the cultures were kept at 1% parasitemia with a

Medium composition	concentration
Albumax I	1.25 g/l
D-Glucose	11.1 mM
Gentamicin	0.05 mg/ml
Hepes	25 mM
Hypoxanthine	μ M
KCl	5.4 mM
L-Glutamine	2 mM
NaCl	85.6 mM
NaHCO ₃	23.8 mM

Table 3.1: Components of the culture medium used for the experiments on RBCs in this work.

5% haematocrit value in a New Brunswick Eppendorf 170R incubator at 37°C in an atmosphere of 5% CO₂, 5% O₂, and 90% N₂. The investigation of the three main erythrocytic asexual stages required synchronised cultures. Therefore, the ring-stage parasites were sorted out by supplementing the cultures with 5% Sorbitol and kept at 37°C for 10 min. Subsequently, RPMI 1640 was added, the mixture centrifuged for 5 min with 800 g and the supernatant removed. These three steps complete one cycle of washing. After two more washing steps, the cultures were set to 5% haematocrit. To access the schizont-stage parasites, 70% Percoll was layered on top of a RBC pellet with supernatant and centrifuged for 11 min with 1500 g [67]. The schizont-stage parasites were collected from the top layer and the ring-stage from the bottom of the tube. Trophozoite-stage parasites were collected after entering the early trophozoite phase based on optical microscopic analysis of Giemsa-stained thin blood films. To obtain high parasitemia in the samples for microscopy analysis, the synchronised cultures were grown up to 5% - 10% parasitemia for approx. three cycles and harvested the parasites with their host RBCs upon reaching either ring, trophozoite or schizont stage. The samples were then prepared by collecting 1 μ l of the pellet, which was smeared on a glass slide, following the standard procedure for preparing thin blood films [5]. For light and atomic force microscopy, VWR microscope slides with 90° ground edges and nominal thickness 0.8 - 1.0 mm were used and precision cover glasses of a thickness of 170 μ m (Superior Marienfeld, No. 1.5H) for the fluorescence microscopy measurements to guarantee high-quality imaging.

Results

4.1 High-parasitemia cultures for in vitro analysis of *P. falciparum* under physiological conditions

Addressing the challenge of efficient malaria treatment and diagnosis requires in-depth understanding of the parasite maturation during the intra-erythrocytic cycle. Exploring the structural and functional changes of the parasite through the intra-erythrocytic stages and their impact on RBCs is a cornerstone of antimalarial drug development. Especially for diagnosis, such insights are relevant for developing autonomous and reliable techniques.

Upon the five species, *P. falciparum* is responsible for the vast majority of severe cases in Africa [43]. Due to its high infection rate and multidrug resistance [43, 117], antimalarial research is mostly focused on this malaria species. Studying the changes and mechanisms during parasite growth requires preparation protocols for dried and liquid samples of *P. falciparum*-infected RBCs. This section presents a routine for the preparation and measurement of samples for atomic force and fluorescence microscopy under physiological conditions.

Preparation of high-parasitemia samples for studies of RBCs in culture medium

Microscopy studies on dried and fixed thin blood films have some inherent limitations. Under such conditions, the shape of the RBCs is deformed, i.e. the cells are flattened and their mechanical properties altered by partial dehydration [118]. Furthermore, thin blood films only show a snapshot of the parasite life cycle. This section introduces a protocol for the preparation of liquid samples together with the setup requirements for the AFM and fluorescence microscope. Additionally, an outlook on the investigation of RBCs with AFM in liquid environment is given with the purpose of gaining a more detailed and realistic view about the maturation of *P. falciparum*.

Studies of living cells with AFM impose major challenges to the sample preparation and microscope settings. Due to the elasticity of the RBC, tip-induced deformation significantly reduces the resolution and complicates mapping of the actual cell surface [119]. The proper analysis of individual RBCs in liquid environment requires a single cell layer. To determine the necessary cell concentration, the volume V_{cells} of all cells needed to cover the complete area of the chamber used for the experiments (BioHeaterTM closed fluid cell [120]) was calculated. With the average values for radius and thickness of RBCs proposed by Diez-Silva *et al.* [121] ($r_{\text{cell}} \approx 4 \mu\text{m}$, $h_{\text{cell}} \approx 2 \mu\text{m}$), the area occupied by a single cells was estimated to $A_{\text{cell}} \approx 50 \mu\text{m}^2$. From this follows that $\approx 10^7$ are required to cover the complete surface of the closed fluid cell chamber with a diameter of $25 \cdot 10^3 \mu\text{m}$, which equals a volume of $V_{\text{cell}} \approx 10^8 \mu\text{m}^3 = 10^{-1} \mu\text{l}$ and 0.1% of the holding capacity of the chamber. Based on this estimation, various concentrations of RBCs, ranging from 0.1% to 0.5% were tested. Figure 4.1 shows representative images for the cell coverage observed at 0.1%, 0.25%, and 0.5% of the volume of the chamber, revealing the optimal cell coverage at a RBC concentration of 0.1%.

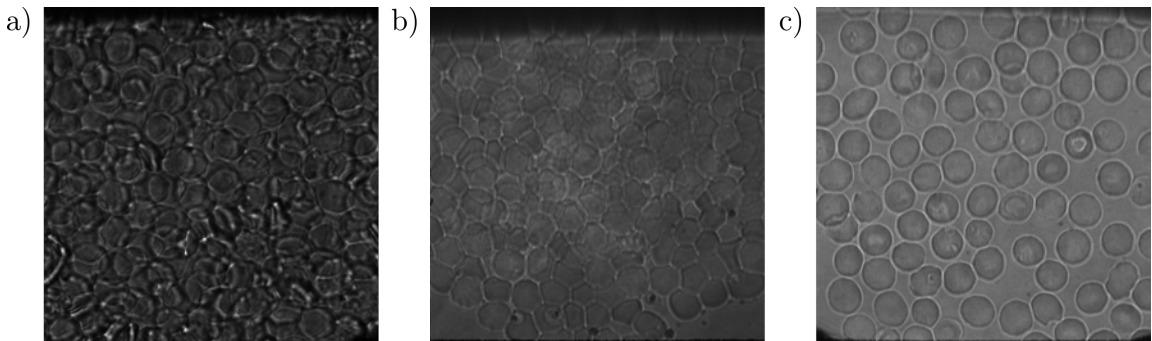


Figure 4.1: **Cell coverage at various RBC concentrations:** 0.5% (a), 0.25% (b), and 0.1% (c). The images were taken via simple light microscopy on unstained cells.

Prior to scanning, the cells need to be immobilised on the glass surface to prevent them from being moved by the cantilever of the AFM. For this purpose, various cantilevers were tested on cells immobilised either with poly-L-lysine (0.5 mg/ml) for 20 min [122] or an additional coating of glutaraldehyde for another 20 min [123]. The RBCs were then left to settle for 30 min on the glass, which ensures cross-linking between the cells and the glass surface. Table 4.1 shows the used cantilevers with corresponding spring constant and an additional column for the success of RBC imaging. The first and second row display the settings and success on samples only coated with poly-L-lysine. In both settings, the RBCs were moved by the cantilever and detached from the surface. With the additional coating of glutaraldehyde, when using pnp-TR, MSCT, and BL-RC-150 the cells remained attached to the surface but were distorted by the cantilever, making it impossible to image the RBCs (Figure 4.2 (a) - (c)). When scanning the sample with the BL-AC-40 cantilever, the RBCs were clearly visible in the image, see figure 4.2 (d) - (e).

Cantilever	Spring constant	Imaging success
BL-RC-150B [124]	0.006 N/m	no
BL-AC-40 [125]	0.09 N/m	no
BL-RC-150B	0.006 N/m	no
pnp-TR [126]	0.0.08 N/m	no
MSCT [127]	0.07 N/m	no
BL-AC-40	0.09 N/m	yes

Table 4.1: Cantilever type and spring constant with corresponding success of RBC imaging. First and second row: samples coated with poly-L-lysine. Third to sixth row: samples coated with poly-L-lysine and glutaraldehyde.

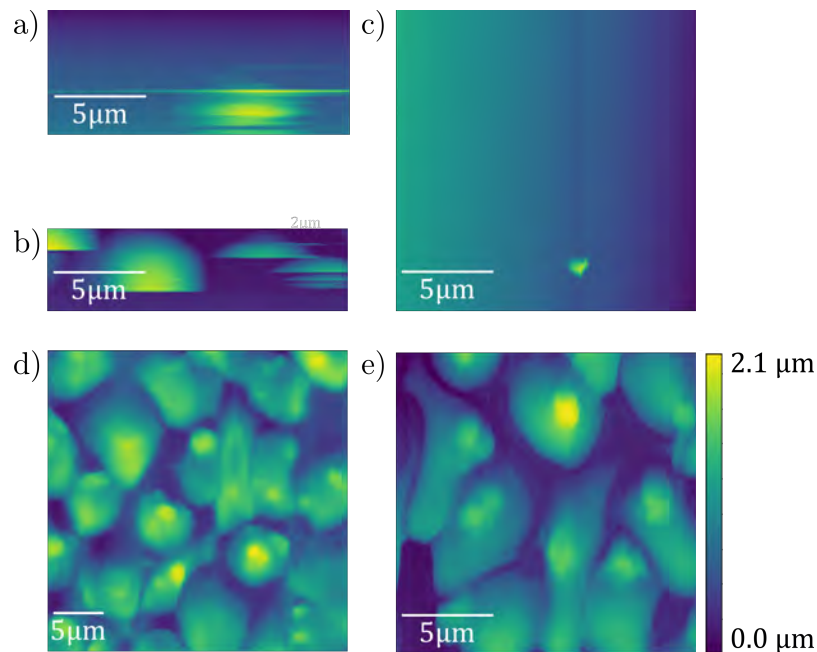


Figure 4.2: **AFM image of a healthy RBCs in liquid environment.** (a) - (c): Single RBCs measured with BL-RC-150, pnp-TR, and MSCT. (d) - (e): Multiple RBCs imaged with BL-AC-40 and corresponding colour bar showing the height of the cells.

The AFM measurements of RBCs in liquid environment revealed the best result for sample surfaces coated with both poly-L-lysine and glutaraldehyde. Testing cantilevers with various spring constants showed that RBCs are deformed and can be displaced upon contact with the cantilever due to their high compliance. While the cells remained attached to the surface of the substrate, they also showed a tendency to adhere to the cantilever tip. Choosing the BL-AC-40 cantilever significantly reduced these effects and enabled the imaging of healthy RBCs in liquid environment, as demonstrated in figure 4.2. However, the current data set obtained on RBCs in liquid environment is limited and systematic studies are planned in the future. The sample preparation routine for AFM measurements under physiological conditions can straightforwardly be adapted for imaging RBCs with fluorescence microscopy. Thus, the experiments shown in this chapter provide the basis for the characterisation and detection of *P. falciparum*-infected RBCs presented in this work.

4.2 Imaging, identification, and manual classification of malaria

The growth of malaria parasites comes with alterations of the RBC structure, which show up as modifications of the cytoskeleton and composition. While light microscopy captures the modifications based on dyes binding to specific parasite components, mechanical alterations inside the cell cannot be understood by the optical observation. For this matter, AFM qualifies as an ideal tool to image the mechanical changes during the intra-erythrocytic cycle. As AFM and light microscopy alter the RBCs during or prior to measurement, fluorescence microscopy is employed as a non-invasive technique to monitor changes based on their optical properties. The throughput of cell imaging requires a compromise between resolution and the duration of image acquisition. Instead of individual cells, a microscopy image typically contains 20-50 RBCs with lower resolution, i.e. single cells are only accessible through image processing. This plays a large role for applications in NNs, where a high amount of single RBC images is required. To overcome this challenge and to prepare the images for the NN, a detection algorithm is employed. This allows for identification and classification of individual cells, forming the basis for the stage-specific detection of malaria in RBCs. Starting with a method for cell detection, this section elucidates a step-wise approach for the processing of microscopy images with the aim to extract single RBCs for further analysis, such as stage classification.

4.2.1 Imaging of the morphological and autofluorescent properties of *P. falciparum*-infected RBCs

Employing atomic force and fluorescence microscopy to resolve the alterations in RBCs during parasite growth provides an in-depth view of the intra-erythrocytic stages. To build a data base for the stage-specific morphological and optical characteristics, RBC

images were recorded by all three imaging methods. Throughout the experiments, light microscopy was used as a control and to evaluate the alternative imaging methods. In table 4.2, the distribution of healthy and infected RBCs is listed, according to the respective imaging method. For a data base that can be used as input for a

Imaging method	number of cells	parasitemia level
AFM	7151	23%
Fluorescence microscopy	41287	12%
Light microscopy	2025	11%

Table 4.2: Number of recorded RBCs and ratio of infected cells (parasitemia) for atomic force, fluorescence, and light microscopy

NN that detects malaria-infected RBCs irrespective of the imaging technique, the number of cells should be comparable for each microscopy method. As a way out of this dilemma, I wrote an algorithm to process each image prior to feeding it to the NN.

4.2.2 Identification of RBCs in microscopy images

Image processing was performed in two steps: contrast adjustment and cell detection. From the microscopy measurements, the algorithm received two kinds of input, text files from AFM and images from fluorescence and light microscopy, see figure 4.3. To enhance the cell features in the images, atomic force and fluorescence microscopy images are artificially coloured. Both inputs were treated as matrices and converted to greyscale for further processing.

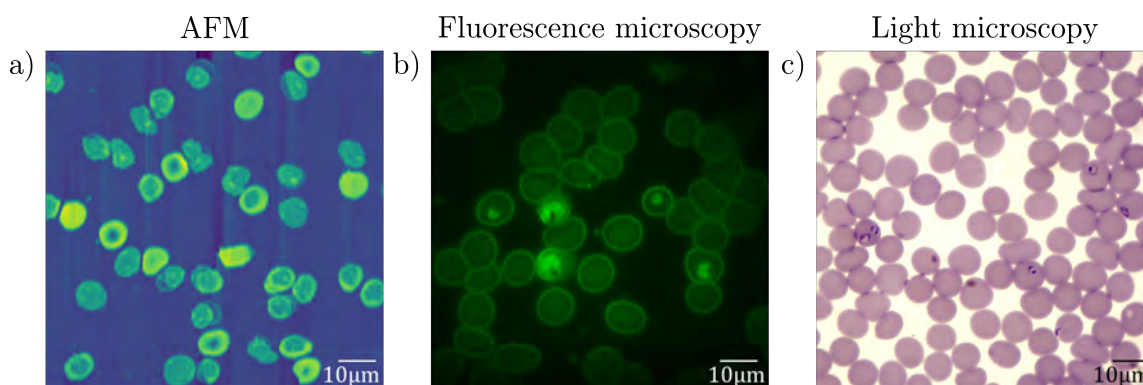


Figure 4.3: **Visualisation of microscopy images with artificial colouring.** Input data from atomic force (a), fluorescence (b), and light microscopy images (c).

In case of AFM and some fluorescence and light microscopy images, the contrast between background and RBCs is not strong enough to locate single cells, which

impacts the accuracy of the detection. Hence, the images were binarised based on pixel intensities by Otsu's method [128], see figure 4.4.

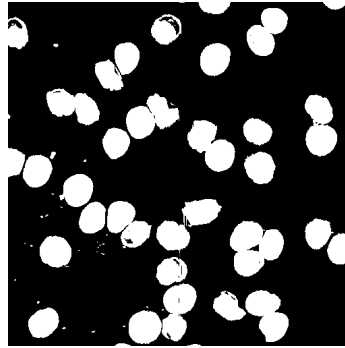


Figure 4.4: **Binarised AFM image.** The pixels in the image are set to 0 (black), if their value is below or equal to the global threshold and to 1 (white), if it is above.

While the processing of AFM images requires an additional step, it is sufficient to enhance brightness, sharpness, and contrast in the fluorescence and light microscopy images. Python offers a module for automatic enhancement of images by a manually chosen factor [129], which was applied to highlight the RBCs.

Subsequent to contrast adjustment, the Hough gradient method was employed. Objects, such as cells are detected in three steps: localising edges, finding the centre of the object, and calculating its radius [130]. The method is controlled by five parameters, $mDist$, $par1$, $par2$, $minR$, and $maxR$. For a graphical representation of the parameters see figure 4.5.

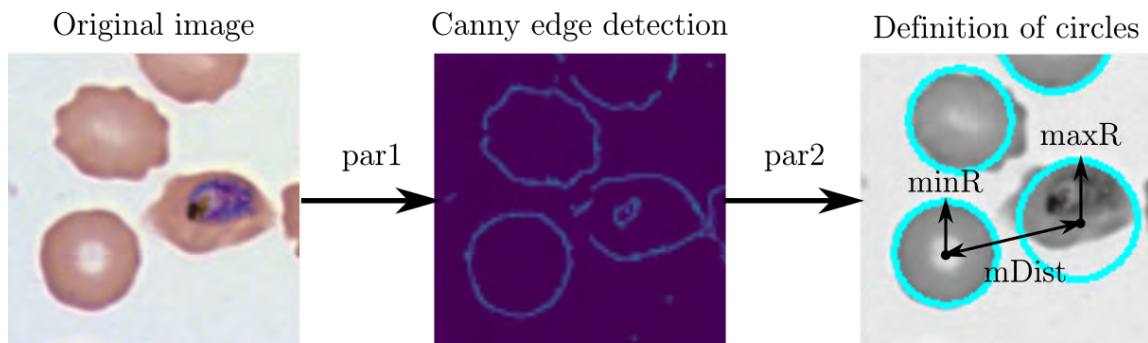


Figure 4.5: **Graphical representation of the Hough circle method.** $par1$ is used to define the threshold for the Canny edge detector, while $par2$ denotes the threshold for the centre detection. $mDist$, $minR$, and $maxR$ are shown in the right image:

$mDist$ defines the minimum pixel distance between the centres of two objects. $par1$ is the threshold value for edge detection by the Canny edge detector [131]. $par2$ sets

the threshold for the number of edge points to declare the object a circle. $minR$ and $maxR$ set the minimum and maximum size of the radius in pixels. To find the optimum value for each parameter, a number of images was analysed with the Hough gradient method. The quantitative results for all microscopy techniques are shown in table 4.3. These results highlight the influence of strong contrast between RBC

Imaging method	$mDist$ [px]	$par1$ [px value]	$par2$ [px value]	$minR$ [px]	$maxR$ [px]
AFM	28	10	8	14	30
Fluorescence/light microscopy	28	20	18	14	30

Table 4.3: Optimum values for the sensitivity of the Hough gradient method for various imaging methods.

and background, i.e. the greyscale fluorescence and light microscopy images require stronger threshold parameters ($par1$, $par2$) than the binarised AFM images. As the Hough gradient method was designed for the detection of circles, it is unsuitable for not circular shapes. The cells depicted in figure 4.6 for example, strongly deviate from the average shape of RBC. Hence, such images are excluded from the analysis.

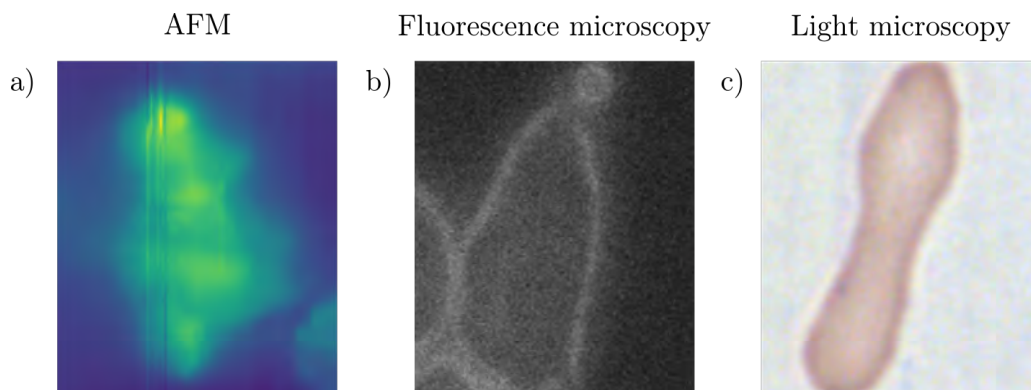


Figure 4.6: **Representative images of not circular RBCs**, recorded by different imaging techniques.

Figure 4.7 shows images of detected RBCs. The cyan circle denotes a successful identification, which was assessed for all imaging techniques. A visual comparison of the detected cells and the actual number of cells in the images revealed that more than 95% of all RBCs (approx. 7000 for each imaging technique) were localised with the cell detection algorithm, which worked best on the fluorescence microscopy images, see table 4.4. The centre and radius of the cells, detected by the Hough gradient method, define the image coordinates associated with the RBC, which are used for the calculation of the geometric and gravitational centre in sections 4.3.1 and 4.4.1.

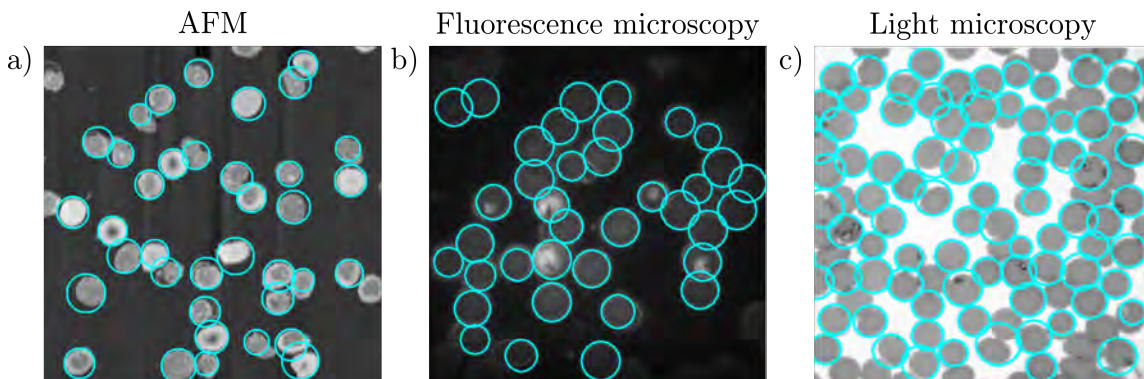


Figure 4.7: **Cell detection on microscopy images.** RBCs detected by the Hough gradient method are marked by cyan circles.

Imaging method	Detection probability
AFM	96%
Fluorescence microscopy	97%
Light microscopy	95%

Table 4.4: Probability of cell detection for different imaging techniques.

Despite the high performance, RBCs strongly deviating from a circular shape impose a challenge to my approach. However, only a negligible amount of those cases occurs in the field samples. Additionally, the detection accuracy is reduced by overlapping cells, as the algorithm is not always able to separate them. While the presented detection method has some limitations, especially if the RBCs are not circular, it reflects a compromise between computation time and detection accuracy. Other approaches, such as neural networks, require more complicated calculations at the expense of time, while reaching a comparable detection accuracy of 70-95% [132–135].

4.2.3 Manual classification of intra-erythrocytic stages

To build a database for the NN, I manually classified the cell images into four categories: healthy, ring, trophozoite, schizont. In case artefacts, such as overlapping RBCs occurred, they were excluded from the analysis. The accuracy of the classification was approved by two experts, Richard Izrael and Petra Molnár. Figure 4.8 shows microscopy images and individual RBCs classified and coloured according to their intra-erythrocytic stage: red (healthy), blue (ring), yellow (trophozoite), green (schizont).

The complete classification of RBCs in the recorded images, see table 4.2, forms the full data base for the stage-specific detection of malaria. Compared to atomic force and fluorescence microscopy, the data base contains a low number of RBC images

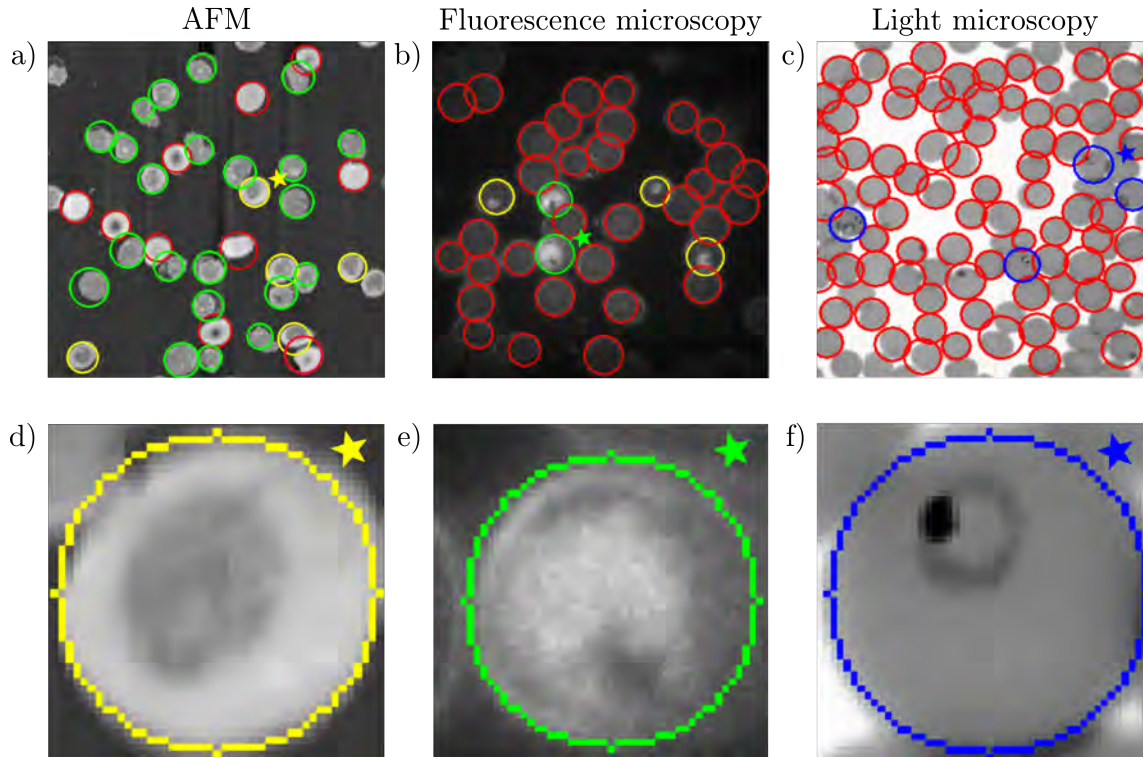


Figure 4.8: **Images of manually classified RBCs.** (a)-(c) Microscopy images, red: healthy RBCs, blue: ring-stage parasites, yellow: trophozoite-stage parasites, green: schizont-stage parasites. (d) - (f): Individual cells, the star marks their position in the microscopy image.

from light microscopy samples. Therefore, the set was supplemented with publicly accessible images from Abbas *et al.* [136]. The distribution of the categories, including the added microscopy images, is shown in table 4.5.

Imaging method	total	healthy	ring	trophozoite	schizont
AFM	7386	5679	350	558	799
Fluorescence microscopy	45726	40874	1946	1598	1265
Light microscopy	26223	24586	502	575	560

Table 4.5: Distribution of healthy RBCs and intra-erythrocytic stages in the recorded microscopy images.

As obvious from these numbers, the stages are inadequately represented. The amount of healthy RBCs exceeds all other categories. Such imbalance in the data set can lead to overfitting and reduce the performance of the network [98]. To equalise the number of images for each category, I augmented the single cell images by three

geometrical operations: rotation, scaling, and reflection. Rotation angle and scaling factor were chosen randomly (between 0° and 180° and 0.5 - 2) for each augmentation to guarantee a large variety of images. The reflection was performed along the horizontal axis, see figure 4.9.

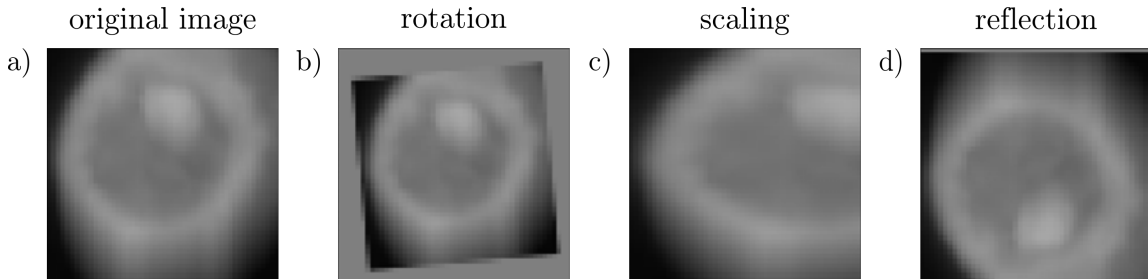


Figure 4.9: **Geometric augmentations on single RBC images.** (a) Original, (b) rotated, (c) scaled, and (d) reflected image

In this section, I presented a step-wise approach for image processing with the goal to prepare data for the application in NNs. The first step contained the detection of objects in the recorded images, which was shown by the example of RBCs. Its success required a compromise of accuracy and expenditure of time. I opted for the Hough gradient method, which was fast but at the cost of precision. Nevertheless, apart from minor limitations, such as overlapping or not circular cells, my algorithm was able to find more than 95% of all measured RBCs. In the second step, I classified the detected cells according to figure 2.2 into the four categories: healthy, ring, trophozoite, and schizont. Since the sample preparation, culturing, and imaging are time consuming and difficult, I could only record a limited number of images, which provide small data sets, where healthy RBCs and intra-erythrocytic stages are rather unevenly represented. Thus, in the final step, I employed geometric operations to augment the data set and hence equalised the number of RBC images in the categories. Despite some limitations, the introduced three step approach allows fast preprocessing to prepare microscopy images for their application in NNs.

4.3 Exploring characteristic features of *P. falciparum*

For the realisation of a reliable classification algorithm, feature engineering is a crucial step, as predictions are based on prominent features of the input data. Therefore, they are required to be characteristic to the intra-erythrocytic stages and should be accessible through statistical analysis or image processing.

Atomic force and fluorescence microscopy enable an in-depth investigation of the alteration in RBCs during the intra-erythrocytic cycle. These techniques allow probing

of the morphology and the observation of autofluorescence. In what sense those properties are specific to the malaria stages is the subject of this section.

4.3.1 Statistical analysis for the characterisation of RBCs

The main goal of this step is to present a statistical analysis of characteristic RBC features. For this purpose, I determined cross-sections through the single RBC images. Their orientation and length were defined by the following calculations. Using the circles fitted around the RBC by the Hough gradient method introduced in section 4.2.2 as contour, I determined the geometric centre \vec{r}_{geo} based on the conventional definition:

$$\vec{r}_{\text{geo}} = \frac{1}{n} \sum_{i,j} \vec{r}_{i,j}, \quad (4.1)$$

where $r_{i,j}^{\vec{}}$ denotes the position of the pixel (i, j) with the x and y coordinates i and j , respectively, and the summation goes over all n pixels within the interior of the RBC. In order to determine characteristic features in the radial profile of infected RBCs, as a first approach, I calculated 18 equally (10°) spaced cross-sections through the geometric centre of the cells. Subsequently, the length of the cuts was restricted by the intersections between cut and contour. A representation of the cross-sections is displayed in figure 4.10.

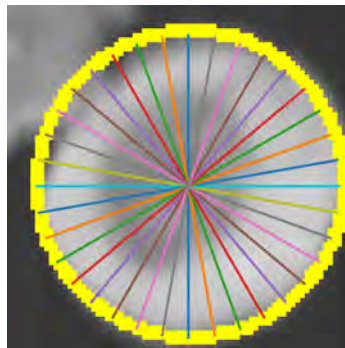


Figure 4.10: **Illustration of radially spaced cross-sections** using the example of a single cell AFM image. The yellow circle represents the contour of the cell.

Figure 4.11 shows the height- and intensity-profile curves corresponding to these cross-sections as well as their average for representative sample images with insets displaying the corresponding RBC. The height is denoted as h_{RBC} and the intensity as I_{px} and their average as \bar{h}_{RBC} and \bar{I}_{px} . For visualisation, the height-profile curves are displayed with the same colour as the corresponding cuts in figure 4.11. Four cross-sections, namely the horizontal, vertical and two diagonal ones, are also shown in the insets of figure 4.11 and the corresponding height-profile curves are highlighted as thick lines in the plots. While AFM represents the height profile, fluorescence and light microscopy demonstrate how the pixel intensity changes. As both are connected to the parasite maturation in RBCs, they serve as the basis for finding prominent features characteristic to the intra-erythrocytic stages.

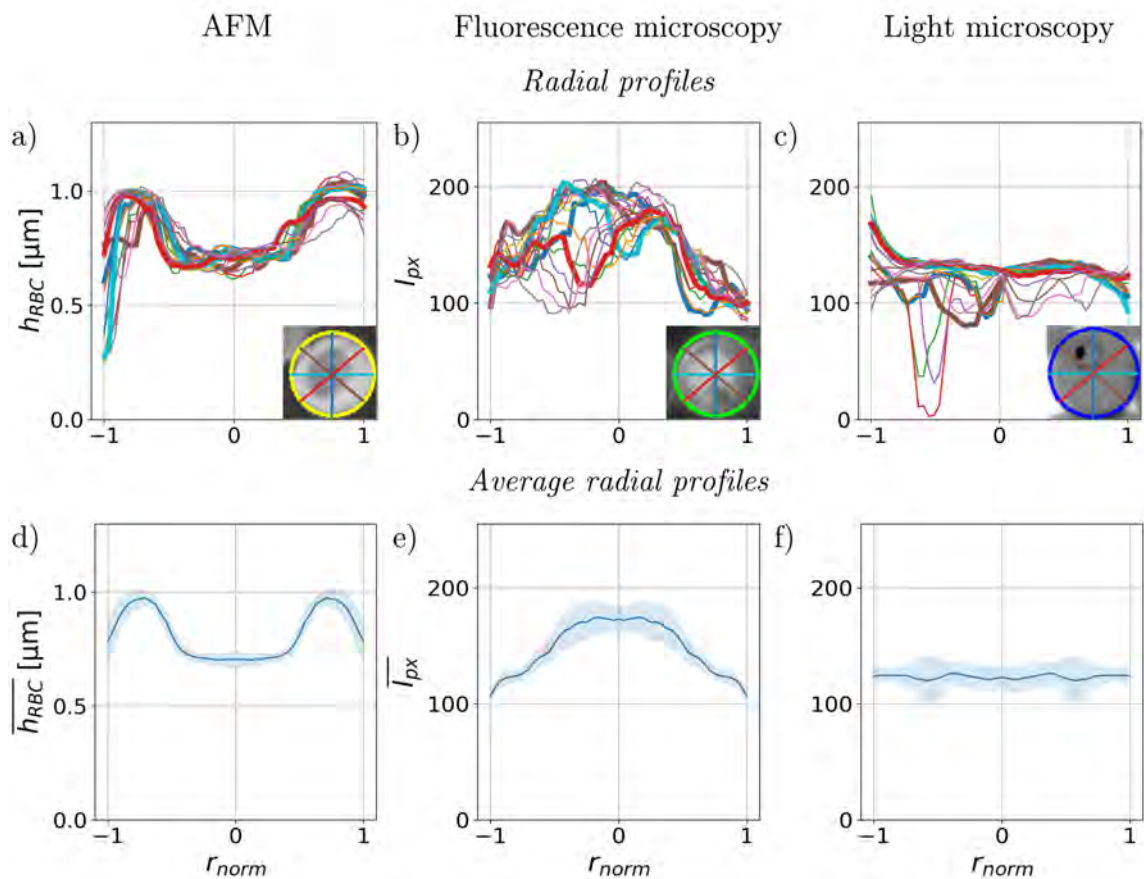


Figure 4.11: **Radial cross-sections through representative RBC images.** (a): Cross-sections through an AFM image representing the height profile of the RBC. (b) and (c): Pixel intensity profiles of RBCs in fluorescence and light microscopy images. Panels (d), (e) and (f) display the average radial height profile corresponding to panels (a), (b) and (c), respectively. The shaded area around the average cut shows the standard deviations from the mean values. Small insets with coloured lines show the RBCs related to the profiles. The height- and intensity-profile curves corresponding to the four cuts, depicted in the insets, are highlighted as thick lines.

4.3.2 Morphological properties of malaria-infected erythrocytes

AFM is a surface sensitive technique and provides information about the morphology of an object. As such, this method is perfectly suitable for monitoring the alterations in RBCs caused by *P. falciparum*. While AFM enables measurements with high resolution, image acquisition is time consuming, which may impose limitations to routine use. However, AFM is performed without contrast material, e.g. Giemsa's stain, which is a clear advantage over light microscopy, as the cells are not altered prior to the measurements. In this section, the topographical alteration in RBCs during the intra-erythrocytic cycle is quantitatively analysed with the goal of assigning characteristic features to the malaria stages. Additionally, the influence of Giemsa's stain on the morphology of RBC is explored.

Topographical alterations in RBCs

Figure 4.12 shows the one representative image for each of the intra-erythrocytic stages, recorded on a thin blood film. The radial and average height profiles reveal the detailed topographical changes in each cell. For all images, the absolute height was measured relative to the surface of the substrate. Healthy RBCs show only minor surface deflection and appear almost flat. Following the invasion of the unperturbed RBC, which serves as reference for infected cells, the parasite induces the development of a cavity in the ring stage, showing an indentation in the height profile at its position. Upon maturation to the trophozoite-stage parasite, this cavity manifests as a sharp drop in height. In the further course of the cycle, the cavity widens, occupying up to half of the RBC diameter. In the final stage, the RBC develops a flat and featureless surface, similar to the healthy one, albeit displaying strong roughness. In addition, the schizont-stage parasite deforms the cell contour that strongly contrasts with the cylindrical shape observed for healthy RBCs and earlier intra-erythrocytic stages.

To quantify the relation between the growth of *P. falciparum* and the topographical alterations in RBCs, I analysed the set of radial height- and intensity-profile curves and their average, which are shown in figure 4.12 (i) - (p) for a representative set of RBCs. Notably, the height profiles of the healthy RBC and the schizont-stage parasite are symmetric with respect to the centre of the cell. In case of ring- and trophozoite-stage parasites, an asymmetric behaviour was observed. Figure 4.13 shows typical RBCs infected with ring and early trophozoite. The radial cross-sections covering the parasite are displayed as solid lines, indicating the parasite as reason for the asymmetry inside the cell.

A quantitative analysis of the measured RBC images revealed that the majority of ring-stage parasites was off-centred (60%), whilst the opposite was true for the trophozoite (39%). This observation is supported by data in the literature, see chapter 2.1.1, stating that malaria parasites are motile inside the RBC during the

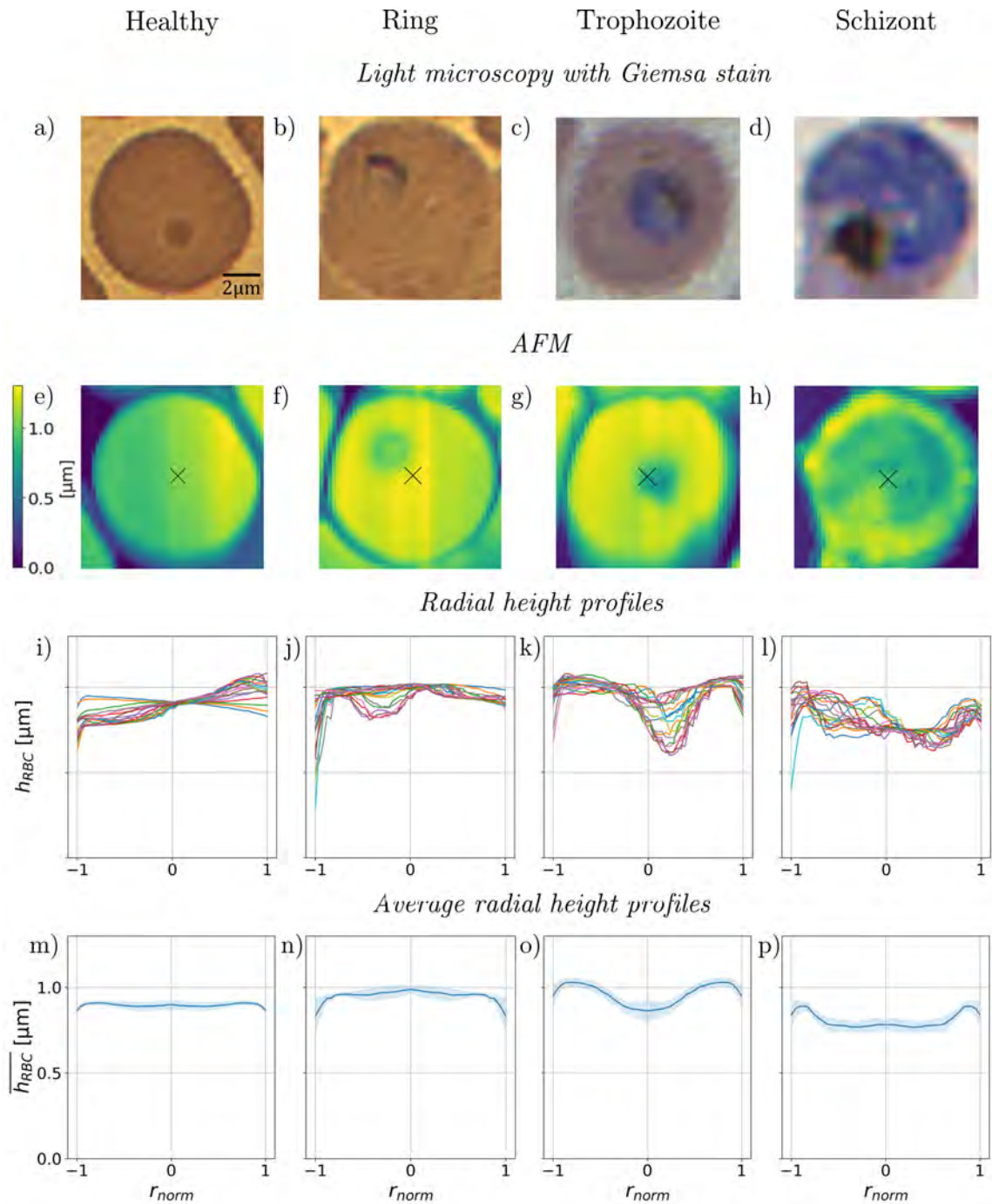


Figure 4.12: **Topographical alterations of *P. falciparum*-infected erythrocytes.** (a) - (d): Light microscopy images of Giemsa-stained RBCs. (e) - (h): Corresponding AFM topography images recorded over the same areas as the light microscopy images. To provide guidance for the features displayed by the radial cross-sections, the centre of the RBCs is indicated by a black cross. (i) - (l): Radial height profiles of RBCs, as obtained by following the description in section 4.3.1. (m) - (p): Mean height profiles with standard deviations from the mean values corresponding to the radial cross-sections in panels (i) - (l).

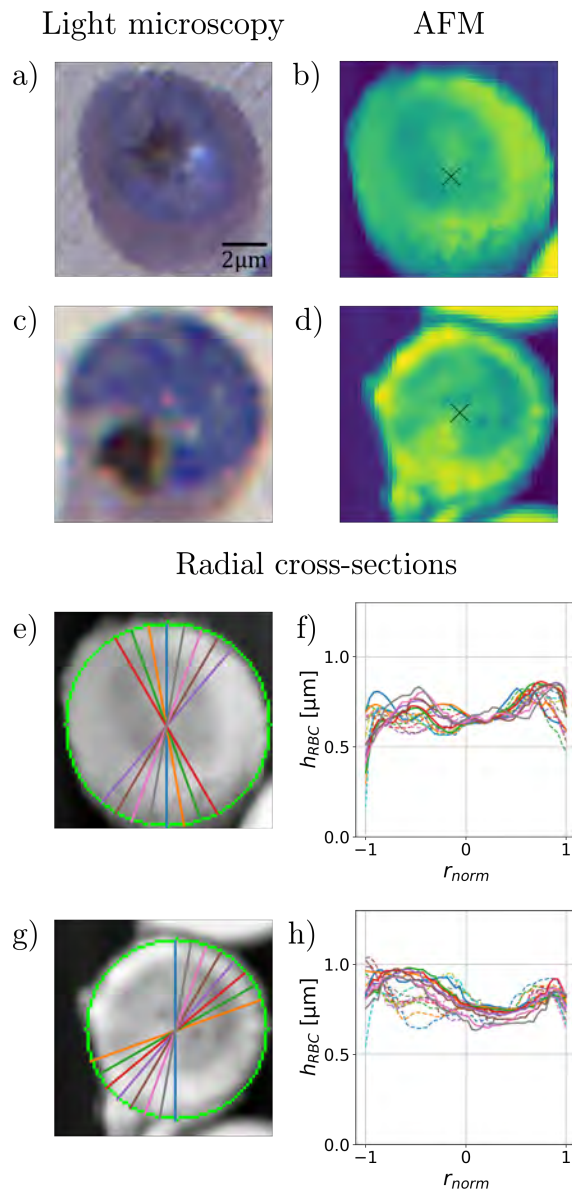


Figure 4.13: **Asymmetric height profiles of *P. falciparum*-infected RBCs.** First and second row: Light microscopy images of erythrocytes infected with ring and early trophozoite stage parasites in stained thin blood films and AFM topography images of the respective cells. The black cross indicates the centre of each RBC. Third and fourth row: Corresponding radial cross-sections showing the cross-sections covering the parasite, indicated as solid lines in (d) and (h).

ring and early trophozoite stage. Upon growth, they lose their ability to move around and settle in a more central position. With this approach, it is also possible to resolve food vacuole in trophozoite- and schizont-stage parasites, which is visible as a broad peak in the height profile. Representative images for each stage are presented in figure 4.14. The cross-sections showing the profile of the food vacuole, where the hemozoin crystals are accumulated, are indicated as solid lines in the cell images and in the plot.

While the radial cross-sections provide a lot of information about the features of the intra-erythrocytic stages, it is not straightforward how to systematically sort out the relevant characteristics of the RBCs. As my first attempt to simplify this representation, the 18 height-profile curves were averaged to a single radial height-profile for each RBC, as shown in figures 4.12 (m)-(p). With these curves, I was able to reproduce the features observed in the two-dimensional images: the flat, featureless, discoid shape of the healthy RBC, the development of the cavity in the ring- and trophozoite-stage parasite, and the rough landscape, characteristic to the schizont-stage parasite. Even though the asymmetry of the cells is not directly visible in the curve, it is partially captured by the standard deviation of the mean values, as represented by the shaded area.

Characteristic height profiles of intra-erythrocytic stages

To quantify the topographical alterations observed for single RBCs, I sorted the cells according to the parasite stages into healthy, ring, trophozoite, and schizont by the corresponding light microscopy images. The next part of this section contains a thorough analysis of the key features provided by the radial height profiles. As part of this analysis, the influence of Giemsa-staining on the morphology of the RBCs is explored. In the final part, an outlook on the investigation of RBCs in liquid environment is given.

To obtain representative curves for each intra-erythrocytic stage, I determined the mean height along the cell radius by averaging the radial cross-sections of the RBCs belonging to the respective stage. Figure 4.15 shows the representative curves obtained for the stained, marked by a star, and unstained samples. Their distribution is displayed in table 4.6.

Imaging method	total	healthy	ring	trophozoite	schizont
AFM	7151	5505	355	540	751
Giemsa-stained	1743	1217	122	209	195
unstained	5408	4288	233	331	556

Table 4.6: Distribution of Giemsa-stained and unstained RBCs in the AFM images.

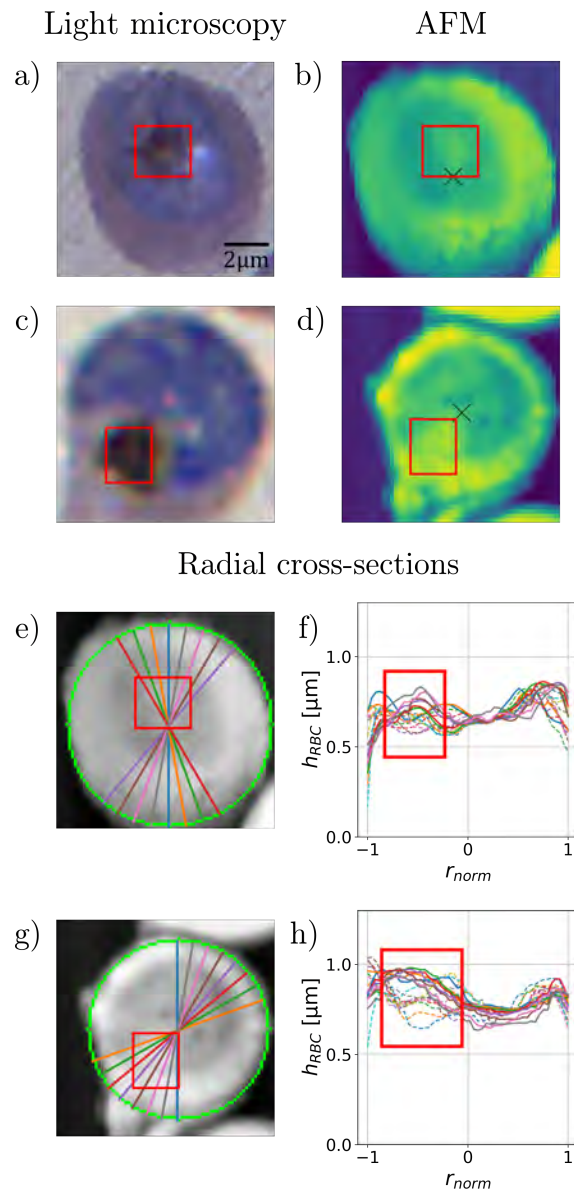


Figure 4.14: **Characteristic cross-sections of infected RBCs indicating the location of the food vacuole.** The food vacuole and its corresponding height profile are surrounded by a red rectangle. First and second row: Light microscopy images of RBC containing a trophozoite- and schizont-stage parasite with corresponding AFM images. Third and fourth row: Radial cross-sections representing the height profile of the food vacuole are shown as solid lines.

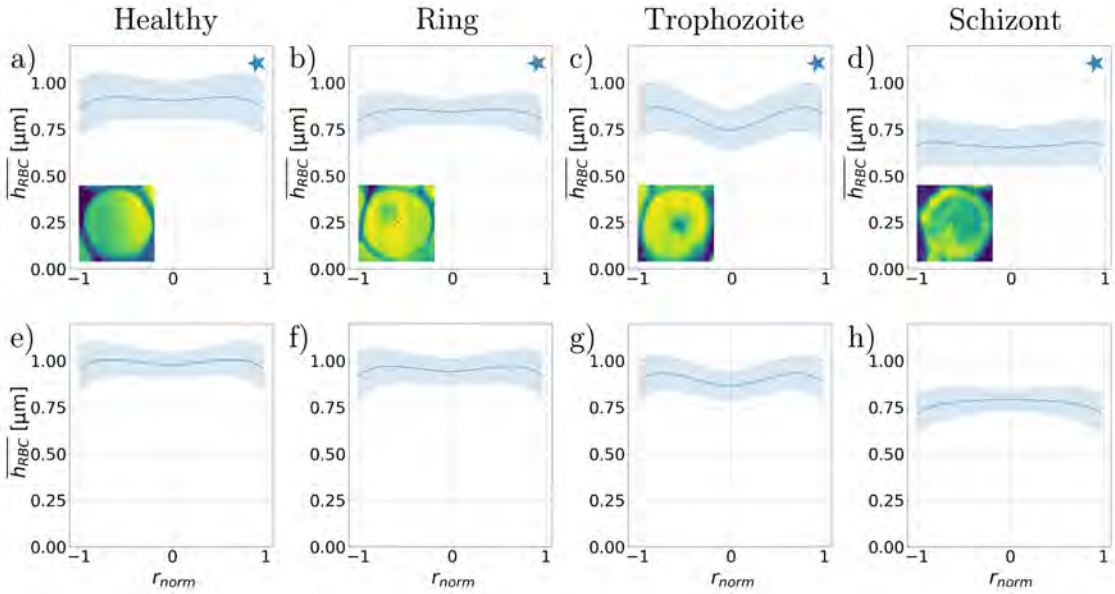


Figure 4.15: **Height profiles of Giemsa-stained and unstained *P. falciparum*-infected RBCs.** First row: Mean height profiles of stained cells, marked by a blue star. Second row: Mean height profiles of unstained cells.

A comparison with figure 4.12 reveals similar features as observed for individual RBCs. To provide a statistical analysis of these features, the mean height h_{mean} of the cell and the height drop h_{diff} from the maximum at the cell edge to the minimum near the centre were calculated for each intra-erythrocytic stage and separately for Giemsa-stained (*) and unstained cells. Statistical significance was proven with a t-test between the stages (Table A.3-A.4). The results are shown in table 4.7, revealing a reduction in mean height from healthy RBC to the schizont-stage parasite. As

	healthy	ring	trophozoite	schizont
h_{mean}	0.91*, 0.99	0.84*, 0.95	0.83*, 0.90	0.67*, 0.76
h_{diff}	-0.02*, -0.03	-0.02*, -0.03	-0.12*, -0.07	-0.03*, 0.04

Table 4.7: Characteristic values from the analysis of mean height profiles, representing the healthy RBC and each intra-erythrocytic stage. h_{mean} describes the average height of the RBCs and h_{diff} the height drop from the maximum at the cell edge to the minimum near the centre of the cell, which is denoted by a minus sign, if it describes a local minimum and by a positive sign in case of a local maximum. The values obtained for the Giemsa-stained images are marked with a star.

second parameter, the height drop remains similar throughout the stages but shows a significant increase in the trophozoite stage. While the radial variation of the height is approximately 14% of the mean height for the trophozoite-stage parasite in the

stained case and 8% in the unstained, the variations for the other stages are negligible. The numbers in table 4.7 clearly show that Giemsa-staining reduces the height of the RBC and enhances the morphological features of the characteristic profile. This observation is supported by an independent t-test (A.1) proving that the h_{mean} and h_{diff} of Giemsa-stained images are significantly different from unstained ones. For this reason, the Giemsa-stained RBCs are qualified for the study of morphological alterations. Therefore, in the further course of the thesis, both stained and unstained samples are merged into a single data set.

Due to the limitations of the cell identification method presented in section 4.2.2, background noise, such as edge effects caused by overlapping cells or a tilt of the sample, may be introduced. These are reduced by weighting down the periphery of the RBCs using a Gaussian filter. Compared to the characteristic values obtained for the unprocessed data set, the radial variation of the height increased to 20% of the mean height in the trophozoite, to 14% in the schizont stage, to 7% in the ring stage and to 8% in case of the healthy cell, see table 4.8. This also leads to an increase in the standard deviation of the average profiles. As the features in the radial height profiles become more pronounced, the differences between the single cells increase. These

	healthy	ring	trophozoite	schizont
h_{mean}	0.97	0.95	0.87	0.76
h_{diff}	-0.06	-0.03	-0.07	0.05
$h_{GF\text{mean}}$	0.99	0.95	0.90	0.76
$h_{GF\text{diff}}$	-0.08	-0.07	-0.18	0.11

Table 4.8: Characteristic values from the analysis of mean height profiles of RBCs before and after applying a Gaussian filter to the images. The data set contains both Giemsa-stained and unstained images.

results show that the Gaussian filter not only preserves the observed topographical alterations but also significantly enhances the characteristic features, see figure 4.16 and tables A.2 and A.6.

In summary, the topographical alterations obtained *P. falciparum*-infected RBCs for thin blood films reveal characteristic features for the intra-erythrocytic stages. In the first phase of the cycle, I observed small indentations caused by the presence of the parasites. Due to their motility, they are not restricted to a certain location inside the RBC as reflected in the asymmetry of the height profiles. For the trophozoite-stage parasite, a cavity was observed, which proved to be a characteristic feature to this stage. Depending on the age of the schizont, the RBC is mostly filled by the parasite, showing as a ruffled, rather flat surface. Capturing the hemozoin crystals, the height profiles also serve as an indicator for the location of the food vacuole. With the statistical analysis of height profiles based on the calculation of cross-

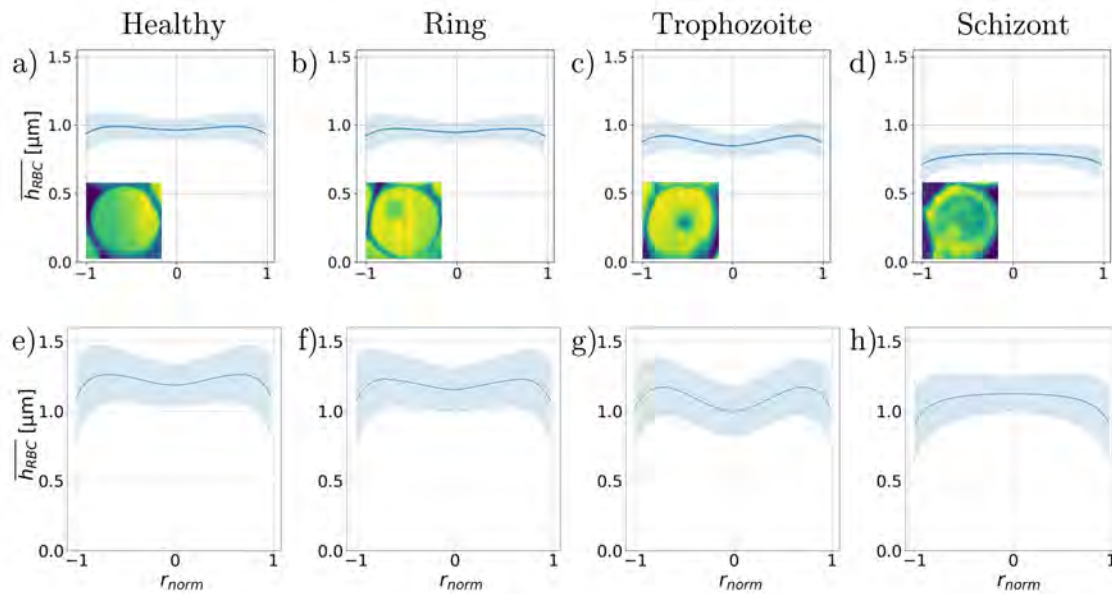


Figure 4.16: **Comparison of averaged height profiles of RBCs infected with the intra-erythrocytic stages of *P. falciparum* before and after image processing.** (a)-(d): Mean height profiles of unprocessed images. (e)-(h): Mean height profiles after removing background noise by applying a Gaussian filter.

sections through single RBC images, I was able to corroborate these observations with the experiments of Grüning and Bannister *et al.* [11,29]. Moreover, my statistical evaluation provided characteristic mean height profiles and parameters for healthy RBCs and the intra-erythrocytic stages.

4.3.3 Optical properties of malaria-infected erythrocytes

As observed above, the morphology of RBCs is strongly affected by *P. falciparum*. Such changes are connected with the restructuring of the RBC and the formation of hemozoin. While AFM provides information about the topography, it may alter the cytoskeleton during image acquisition, as it is an invasive technique. In addition, AFM is not suitable for generating a high throughput of images. As an alternative approach, in this section, optical microscopy, specifically fluorescence microscopy, is introduced as a tool for the investigation of alterations in RBCs during the intra-erythrocytic cycle. Autofluorescence emission is known to emerge in RBC at around 400 nm. How this wavelength can be used to uncover malaria parasites inside their host erythrocytes is explored in this section. Starting with the analysis of autofluorescence patterns, the investigation is further supplemented with a comparison to the emergence of fluorescence in living parasites.

Fluorescence patterns in single red blood cells

The effect of laser excitation on *P. falciparum*-infected RBCs was studied on thin blood films, employing various wavelengths. Samples were prepared according to section 3.5 and the microscope operated with the setting described in section 3.5. Figure 4.17 shows autofluorescence patterns of a healthy RBC and each of the intra-erythrocytic stages supplemented with light microscopy images as reference, excited at four different wavelengths. Upon excitation wavelengths of 405 and 488 nm, the parasites are clearly visible inside the cell, while the RBCs and the parasites are barely distinguishable from the background when excited with the wavelengths 561 nm and 642 nm. In addition, the trophozoite and schizont-stage parasite display weak autofluorescence at 642 nm.

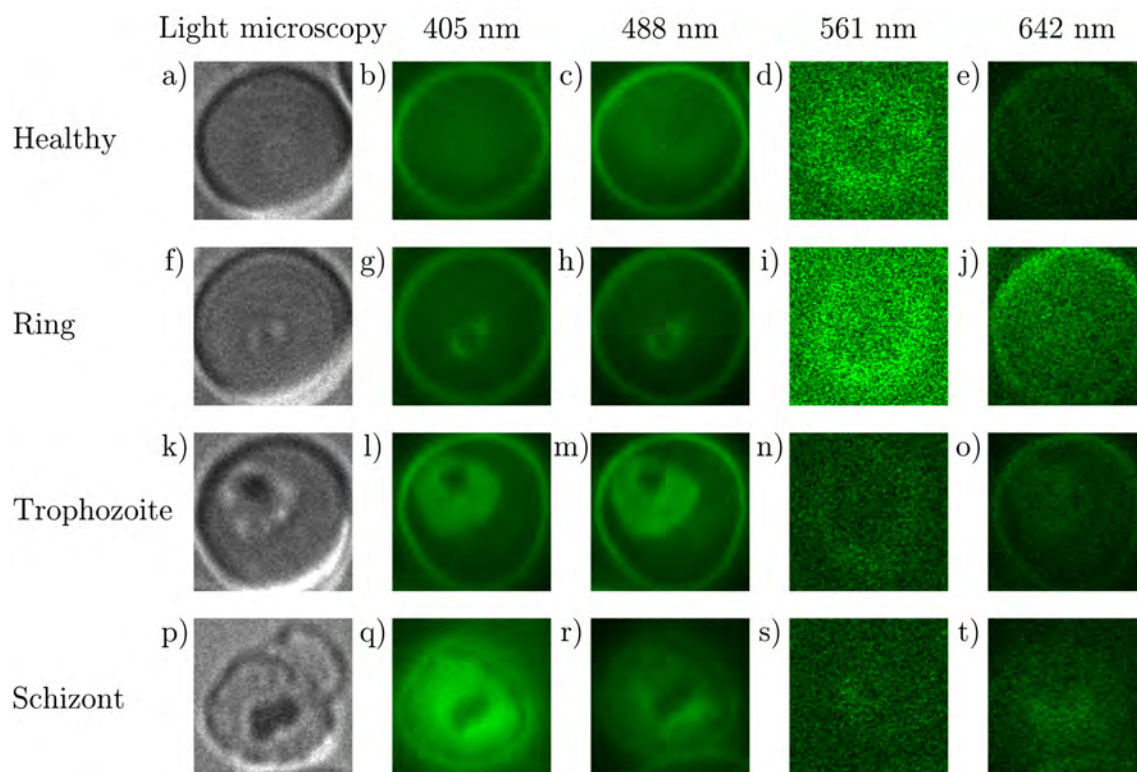


Figure 4.17: Fluorescence maps of *P. falciparum*-infected erythrocytes at various excitation wavelengths. First column: Light microscopy images. Second - fifth column: Autofluorescence patterns of the corresponding cells excited at various wavelengths, indicated on the top of the columns.

Based on these observations, all additional experiments were performed with an excitation wavelength 405 nm. For the identification of the intra-erythrocytic stage, I recorded light microscopy images, while simultaneously monitoring the emergence of autofluorescence. Figure 4.18 shows representative RBCs with characteristic autofluorescence patterns.

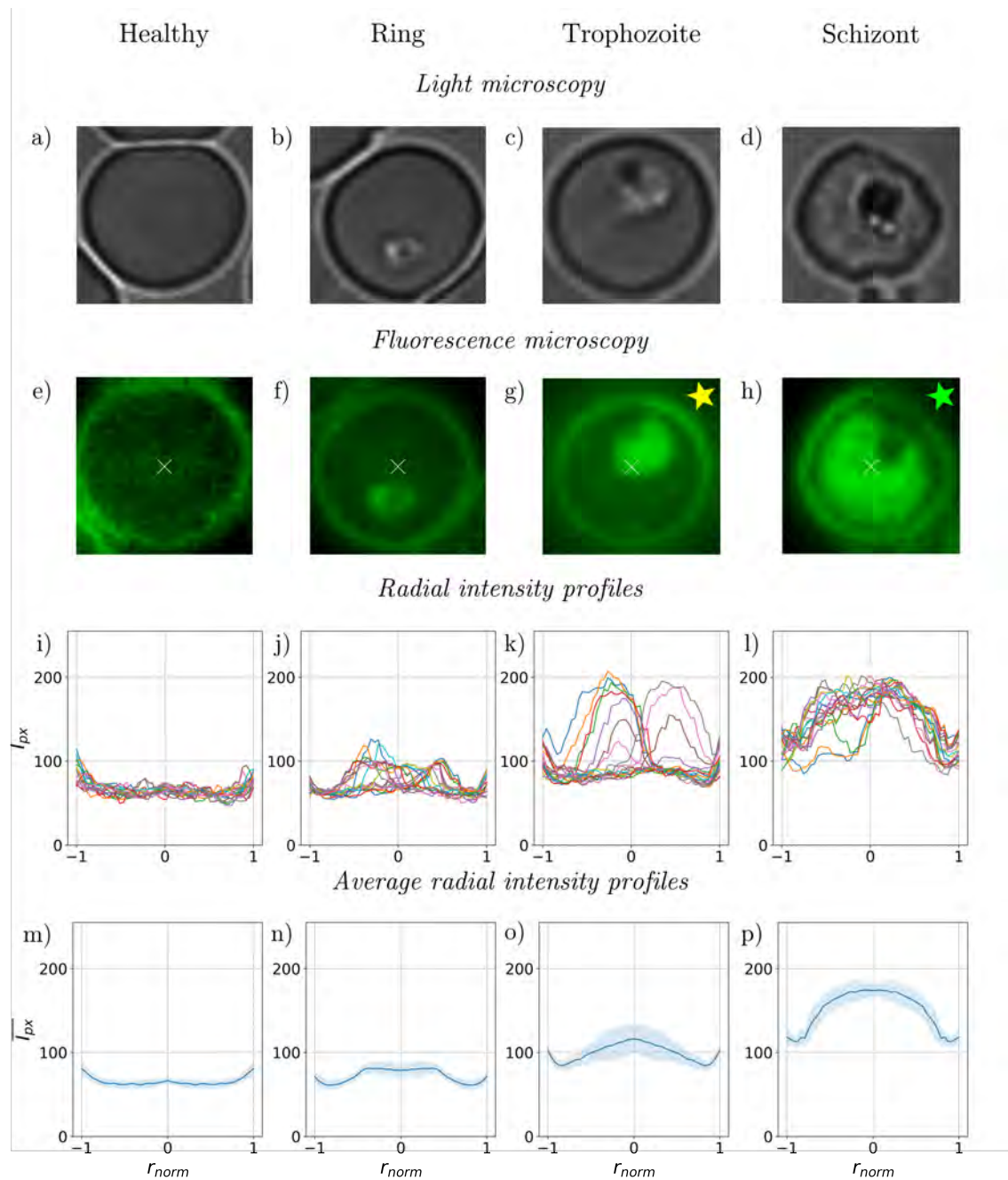


Figure 4.18: **Autofluorescence maps of *P. falciparum*-infected erythrocytes.** The fluorescence images are artificially coloured to enhance the contrast. First and second row: Light microscopy with corresponding fluorescence microscopy images showing the fluorescence properties of a healthy RBC together with the intra-erythrocytic stages, as studied at an excitation wavelength of 405 nm. The white cross marks the centre of the RBCs. Third row: Radial fluorescence intensity profiles. Colouring of the curves follows the colour convention introduced in figure 4.10. Fourth row: Average intensity profiles, obtained by averaging the 18 intensity-profile curves in the panels above with the standard deviation from the mean values.

As observed in figure 4.17, the malaria parasite clearly stands out from the non-fluorescent RBC. Following the concept established in section 4.3.1, I analysed the radial fluorescence pattern and determined characteristic average curves for the intra-erythrocytic stages. As the haemoglobin inside the healthy RBC absorbs most of the light at 405 nm, the radial and mean intensity profiles display a low and uniform signal. For the ring stage, a distinct peak at the position of the parasite below the white cross was observed, manifesting as a small elevation in the average profile. With maturation to the trophozoite stage, the fluorescence intensity increases, followed by the formation of a region with low fluorescence emission, appearing as a dark spot on top of the parasite, visible as a drop in the intensity profile in figure 4.19 (b), represented by the blue and orange solid lines. Parasite maturation leads to an increasing and more distinct elevation in the mean intensity profile. As the schizont-stage parasite fills up the whole cell, broad fluorescence signal is visible in the image and in the intensity profiles. In addition, the region with low fluorescence has increased, which can be seen in the radial intensity profile, displayed in figure 4.19(d) by solid lines.

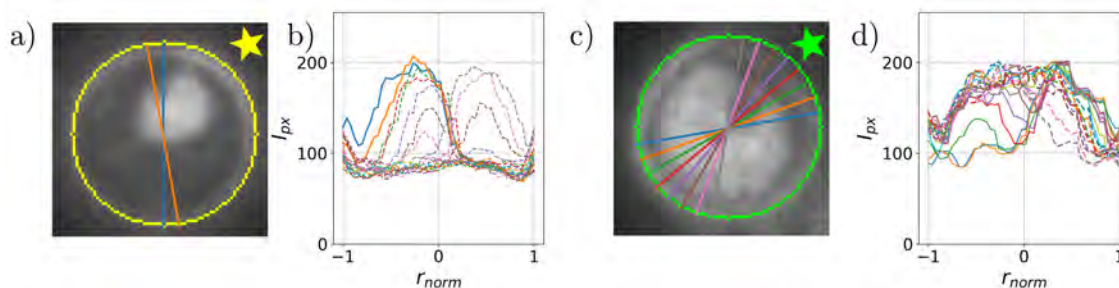


Figure 4.19: **Radial intensity profiles corresponding to the trophozoite- and schizont-stage parasites in figure 4.18.** The matching images are marked with stars. Cross-sections covering the regions of low fluorescence are shown as solid lines in the microscopy images and intensity profiles.

An additional autofluorescence study on hemozoin crystals extracted from infected RBCs, presented in figure 4.20 (d), revealed that hemozoin exhibits no fluorescence upon excitation at 405 nm. This observation strongly suggests that the regions in question, see figure 4.20 (e) and (f), show the food vacuole.

Characteristic fluorescence maps of intra-erythrocytic stages

Stage-specific autofluorescence patterns were determined as shown in section 4.3.2 for the topographical alterations in RBCs. The fluorescence maps presented in figure 4.21 reveal distinct patterns, close to those observed for individual cells in figure 4.18, which are further enhanced after removing background noise from the images with a Gaussian filter.

Equivalent to the definitions of the characteristic height profile values in section 4.3.2, I determined the mean intensity I_{mean} of the cell and the intensity drop I_{diff} from

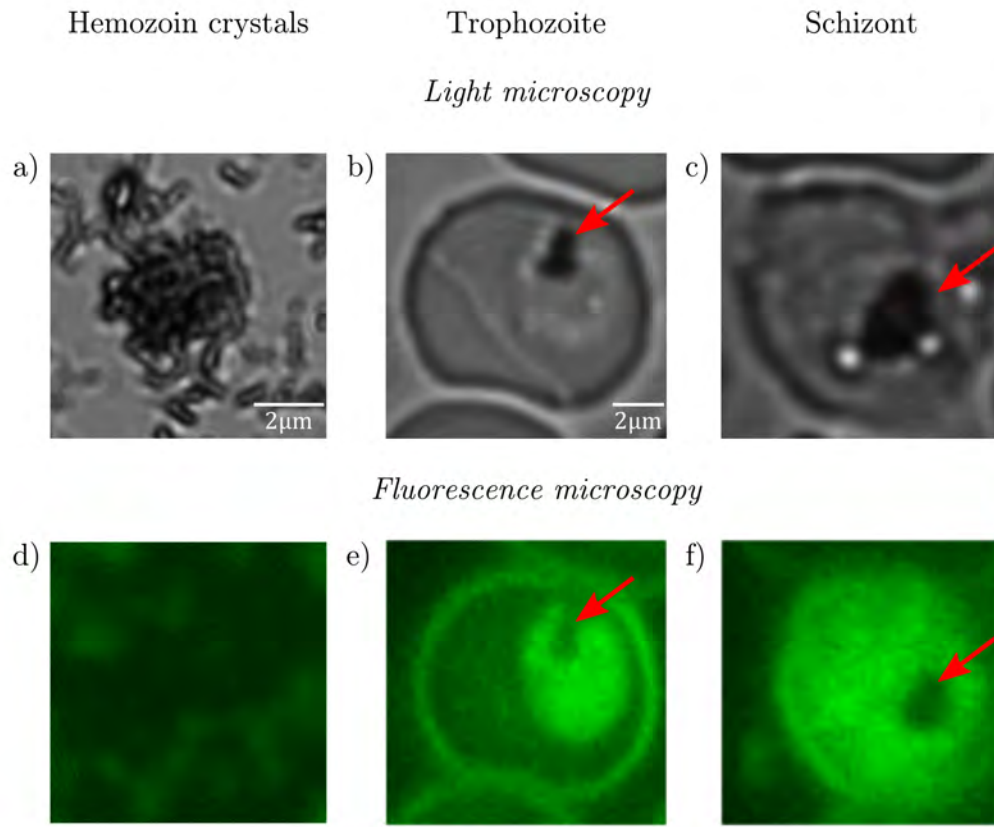


Figure 4.20: **Autofluorescence maps of hemozoin crystals**, as observed at an excitation wavelength of 405 nm. Light microscopy with corresponding fluorescence microscopy images of extracted hemozoin crystals, and RBCs containing trophozoite- and schizont-stage parasites. Regions of high absorbance are marked with a red arrow.

the maximum at the cell edge to the minimum near the centre. The analysis of the intensity patterns, see table 4.9, reveals a significant increase of the mean fluorescence intensity towards the schizont stage, which goes along with a change in radial variation of the intensity from 9% of the mean height in the healthy RBC to 20% of the mean height in RBCs infected with schizont-stage parasites. This dependency further increases by 20-40% in case of the intra-erythrocytic stages for Gaussian filtered images. For proof of statistical significance, see tables A.7- A.9.

Autofluorescence studies of living parasites in culture medium

Imaging of living malaria parasites was carried out on samples prepared according to the protocol presented in section 4.1. The emergence of fluorescence was monitored with the microscope settings described in section 3.5. With this setup, I was able to record images of living parasites inside RBCs, see figure 4.22.

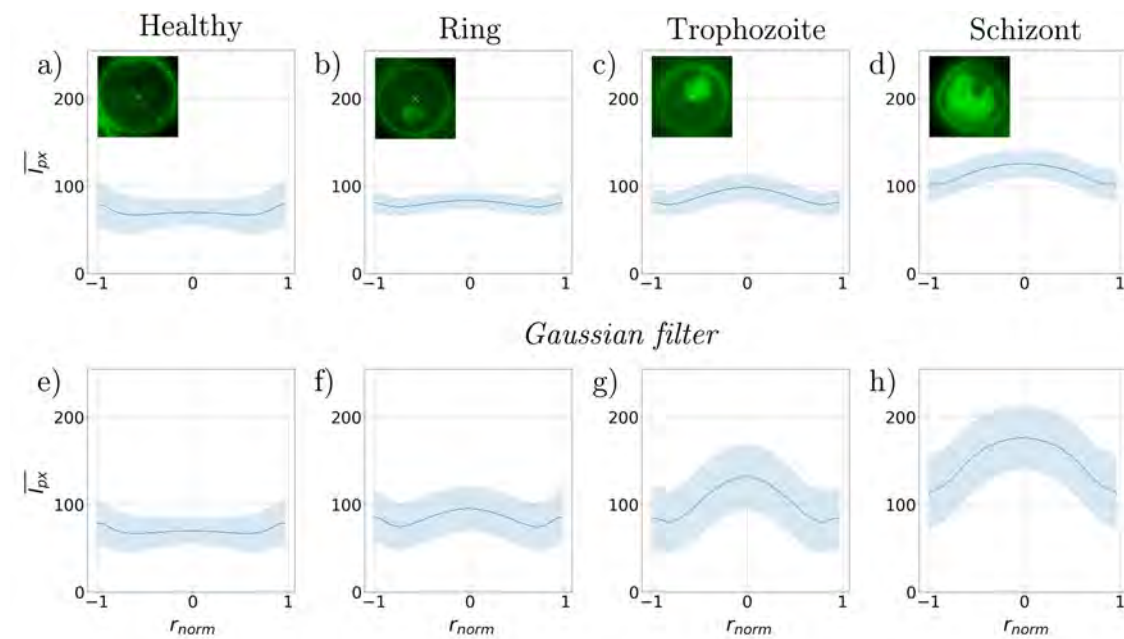


Figure 4.21: **Characteristic autofluorescence maps of *P. falciparum*-infected RBCs.** Mean intensity profiles with standard deviation from the mean values for healthy RBCs and the intra-erythrocytic stages before (first row) and after removing background noise (second row).

	healthy	ring	trophozoite	schizont
I_{mean}	69.52	79.00	84.58	111.06
I_{diff}	-6.09	7.53	19.58	22.35
$I_{GF\text{mean}}$	69.89	79.00	84.58	111.06
$I_{GF\text{diff}}$	-6.14	20.92	51.58	54.31

Table 4.9: Analysis of stage-specific intensity patterns obtained from fluorescence microscopy images of individual RBCs. Characteristic values for images before and after removing background noise.

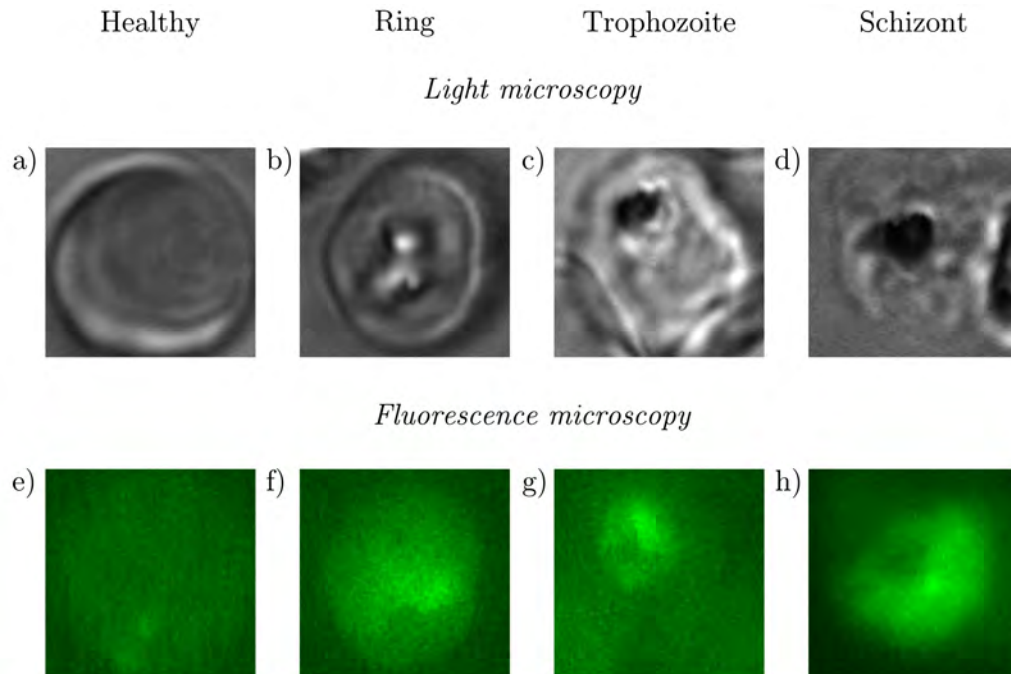


Figure 4.22: **RBC images showing the emergence of autofluorescence in living parasites.** Light microscopy images of *P. falciparum*-infected erythrocytes in cell culture medium with corresponding fluorescence maps.

In agreement with the observations on thin blood films, the parasites exhibit visible autofluorescence. While the healthy RBC emits no signal, the intra-erythrocytic stages and also the food vacuole could be resolved. Compared to the thin blood films, the resolution and contrast of the images was reduced. Additionally, the power of the excitation laser was increased to detect the autofluorescence signal, suggesting that the presence of the culture medium is the cause for the higher absorbance of light during imaging. As a component in the culture medium, phenol red shows a peak in absorbance at the excitation wavelength of 415 nm [137–140], used for the experiments, which supports the above suggestion.

In this section, I demonstrated that *P. falciparum* not only alters the morphology of RBCs but also results in the emergence of autofluorescence upon excitation at 405 nm. Monitoring the fluorescence emission in thin blood films revealed patterns specific to the parasites, which turned out to be characteristic to the intra-erythrocytic stages. Furthermore, the low fluorescence signal of hemozoin crystals at 405 nm, enables the detection of the food vacuole in RBCs. While the optical properties of ZnPPIX hint that it plays a role in the emergence of autofluorescence in malaria-infected RBCs, the revelation of the molecular origin of fluorescence goes beyond the scope of this work.

Imaging the emergence of autofluorescence in a label-free manner offers a way to follow the intra-erythrocytic development without disturbing the environment of the parasites. Thus, I was able to observe fluorescence in living parasites, which showed close correlations with the patterns presented for thin blood films. The imaging routine established for the live measurements provides the basis for recording the autofluorescence of *P. falciparum* during a complete intra-erythrocytic cycle.

4.3.4 Intensity profiles of Giemsa-stained erythrocytes

To complete the investigation of optical properties in RBCs infected with *P. falciparum*, this section finishes the analysis of cell features with a thorough study of single Giemsa-stained RBCs, supplemented by a statistical evaluation of the intra-erythrocytic stages and the calculation of characteristic intensity profiles.

The representative single RBC images in figure 4.23 show a clear variation of the intensity profile with progressing parasite maturation. While the healthy RBC is almost featureless, the presence of ring-stage parasite in the bottom and the trophozoite-stage parasites on the left side of the cell leads to a drop in intensity. Caused by hemozoin crystals located in the middle of the schizont, the intensity profile forms a broad minimum in the centre. Upon averaging the radial cross-sections, these distinct features appear less pronounced.

Figure 4.24 presents the analysis of the Giemsa-stained RBC images. In case of the healthy RBC and ring-stage parasite, a broad local maximum is observed, while trophozoite- and schizont-stage show a local minimum caused by the presence of the parasite. In case of the ring-stage, the effect of the parasite on the overall intensity is lower than was observed for the single image. The image processing by the Gaussian filter leads to smoothing of the curves in the trophozoite and schizont stage, while also introducing a large standard deviation from the mean intensity values. From the healthy RBC to the schizont-stage parasite, the average intensity decreases by 26%, going along with a change in curvature direction from ring to trophozoite, denoted by a negative sign in table 4.10. For the definition of I_{mean} and I_{diff} see section 4.3.3. Despite the occurrence of a large standard deviation in the curves caused by the application of the Gaussian filter, after removing the background noise, the prominent features observed for each intra-erythrocytic stage remain visible.

As already shown for RBCs imaged with AFM and fluorescence microscopy, the statistical evaluation of Giemsa-stained thin blood films revealed distinct patterns and parameters for healthy RBCs and the intra-erythrocytic stages. These results prove that the approach, presented in this section, is a valid tool for exploring the characteristic features of malaria-infected RBCs and further enables a stage-specific classification of RBCs.

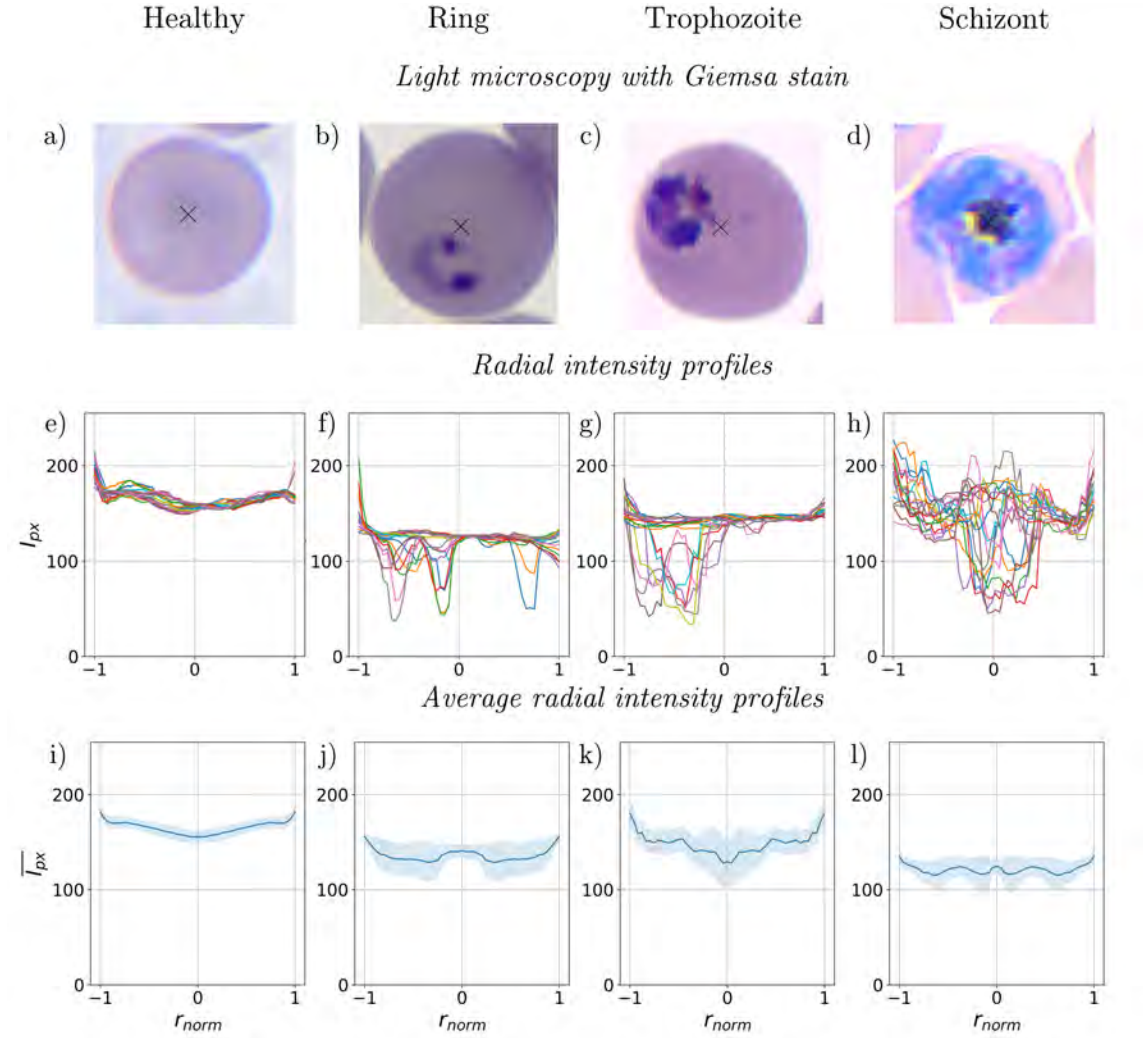


Figure 4.23: **Light microscopy images of Giemsa-stained erythrocytes infected with *P. falciparum*.** (a) - (d): Giemsa-stained images of the intra-erythrocytic stages with a black cross indicating the centre of each cell. (e) - (h): Corresponding radial intensity profiles. (i) - (l): Average intensity profiles with the standard deviation from the mean intensity values indicated by the shaded area.

	healthy	ring	trophozoite	schizont
I_{mean}	170.86	163.76	145.22	121.76
I_{diff}	10.56	7.63	-14.38	-22.39
$I_{GF\text{mean}}$	170.86	163.76	132.43	126.48
$I_{GF\text{diff}}$	31.42	25.82	-9.16	-10.47

Table 4.10: Characteristic statistical values for the intensity curves of Giemsa-stained RBCs with and without background noise.

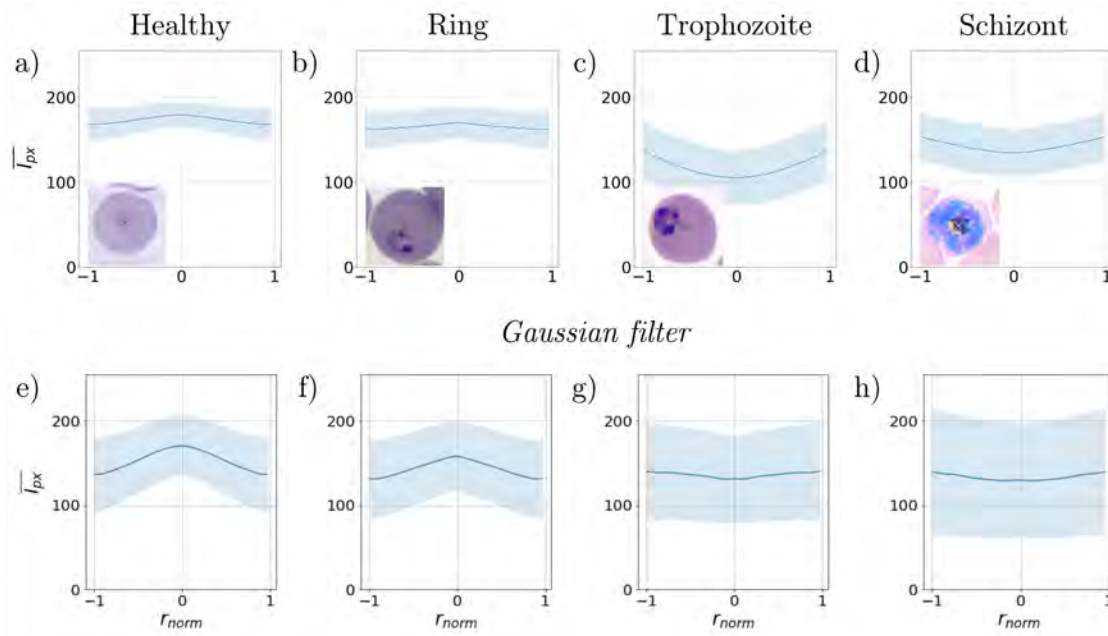


Figure 4.24: **Statistical analysis of Giemsa-stained erythrocytes infected with *P. falciparum*.** Characteristic intensity curves with standard deviation from the mean values of intra-erythrocytic stages with ((a) - (d)) and without background noise ((e) - (h)).

4.4 Reduction of dimensionality as a tool for feature selection in RBCs

In machine learning applications, dimension reduction is often employed to increase the performance NNs and to improve their generalisation ability. Since a reduction comes with a massive loss of information, careful selection of features is required in order to keep the essence of the data.

The studies on morphological and optical properties of the intra-erythrocytic stages presented in section 4.3 revealed the parasite as the most prominent feature of RBCs infected with *P. falciparum*. In this respect, there are various possible choices of measures or indicators, for instance, the sum of local changes over the whole RBC or its asymmetry.

This section introduces a method that yields cross-sections containing the characteristic features of RBCs, which is followed by the assessment of its efficiency.

4.4.1 Dimension reduction of two-dimensional RBC images

A suitable measure for the characteristic features of the intra-erythrocytic stages should be insensitive to external noise, which is often present in AFM and fluorescence

microscopy images due to tip contamination or background illumination, leading to low contrast, as shown in figure 4.25.

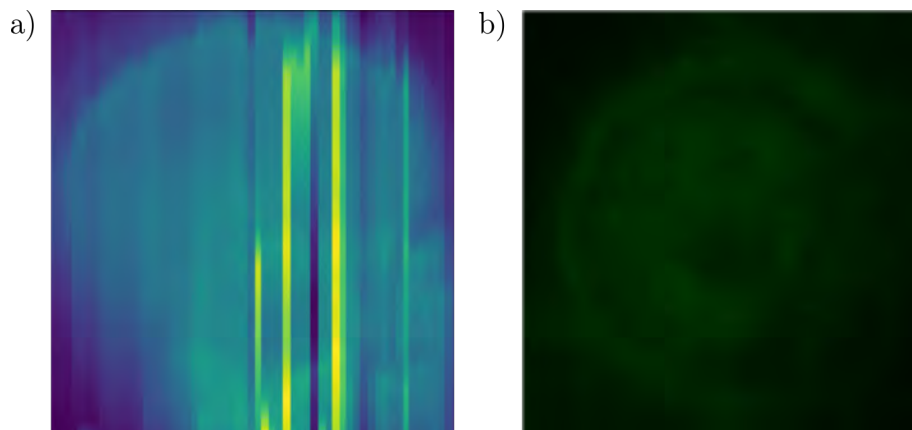


Figure 4.25: **Artefacts in atomic force and fluorescence microscopy images due to external noise.** (a) AFM image (b) Fluorescence microscopy image.

To obtain the characteristic cross-sections of RBC images, I determined two measures: the geometric centre, see section 4.3.1, and the centre of gravity, which is defined by:

$$\vec{r}_g = \frac{1}{n} \sum_{i,j} h(\vec{r}_{ij}) \vec{r}_{ij}, \quad (4.2)$$

where $h(\vec{r}_{i,j})$ denotes the height (AFM) or intensity (fluorescence and light microscopy) of a pixel at position $\vec{r}_{i,j}$. The coordinates inside the RBCs were calculated as described in section 4.2.2. The idea behind this approach is to localise the parasites inside the cell, i.e. the coordinates are weighted by their corresponding height or intensity values. Figure 4.26 shows the effect of a region with lower intensity on two images of cylinders. The presence of the hole inside the cylinders leads to a shift of the gravitational centre, either away or towards the hole, which is indicated by an arrow. In case of the dark cylinder, the shift is higher, as the image values of black pixels are 0 and the values of white pixels 255. Therefore, the presence of the white hole has a larger effect on the position of the gravitational centre. Similarly, this shift can be observed for RBC images. From this follows that the height or intensity values along the straight line through both centres represent a cross-sections of the parasite. In the further course of the thesis, this line is defined as the parasite cut.

To evaluate the efficiency of the presented measure, I visually examined, if the parasite cut detects the presence of the parasite. The percentage of successful detections is shown in table 4.11. The results reveal that the parasite cut only captures 25-50% of the ring-stage parasites in light and atomic force microscopy images, while reaching a significantly higher rate of 88% on fluorescence images. For trophozoite-stage parasites, the rate is much higher with more than 77% for atomic force and fluorescence images, and 81% for light microscopy images. The schizont-stage parasites typically fill up

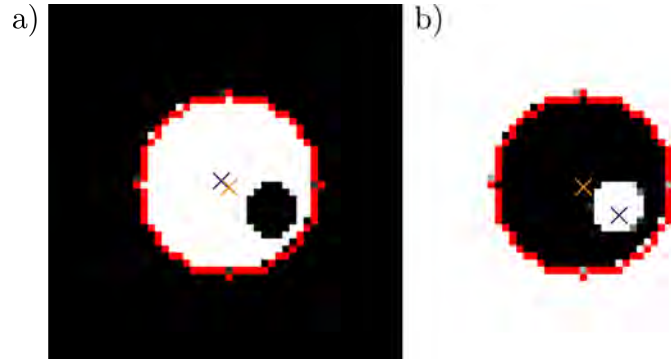


Figure 4.26: **Location of the geometric and gravitational centre in simple cylinders with holes.** Geometric and gravitational centre are shown as an orange and blue cross. The effect of the hole on the location of the gravitational centre is shown for a black (a) and white cylinder.

Imaging method	ring	trophozoite	schizont
AFM	55%	77%	100%
Fluorescence microscopy	88%	79%	100%
Light microscopy	25%	81%	100%

Table 4.11: Percentage of successful parasite detection with the parasite cut.

most of the RBC, thus parasite cuts going through the cell always intersect the parasite. Background noise, such as edge effects like overlapping cells or misaligned contours caused by the limitations of the cell identification method, see section 4.2.2, may compromise the detection of parasites. As shown for the mean height profiles in section 4.3, the application of a Gaussian filter successfully removes noise and additionally enhances the cell features. The effect of this procedure on the parasite cut can be seen in table 4.12. The successful detection of ring-stage parasites in

Imaging method	ring	trophozoite
AFM	59%	86%
Fluorescence microscopy	88%	95%
Light microscopy	51%	88%

Table 4.12: Percentage of successful parasite detection with the parasite cut after removing background noise from the image.

atomic force and light microscopy images improved to 50-60% and the rate for trophozoite-stage parasites to more than 85% and even 95% for the fluorescence microscopy images. Despite of capturing the parasite, the parasite cut may not

contain enough information to fully characterise the features of the intra-erythrocytic stage. For this reason, I determined three additional cuts spanning 90° and $\pm 45^\circ$ with it, see figure 4.27. The parasite cut is shown in blue, connecting the geometric and gravitational centre. Additional cuts are coloured in orange, cyan, and red. The three displayed combinations of cuts are later used in section 4.5.3 for the stage-specific classification of RBCs. In the further course of the thesis, the combinations of cuts in figure 4.27 are denoted as follows: four cuts (parasite cut, 90°), parasite plus cuts (parasite cut, 90°), and parasite cut. For the purpose of comparing the influence of a controlled dimensionality reduction on the stage-specific classification of intra-erythrocytic stages with an arbitrary choice of cuts, I calculated randomly oriented cuts for each RBC image. A sample RBC is shown in figure 4.27(g)-(l) with corresponding intensity profiles in greyscale colours.

In the following, the features captured by the parasite cut and the additional cuts are explained. Figure 4.28 shows a set of RBCs with corresponding cross-sections. The cross-sections clearly reveal the presence of the parasite. In case of the AFM images, see figure 4.28 (a) - (h), the healthy RBC shows a uniform height profile. In the ring stage, a drop occurs in the parasite cut, which is reproduced by all other cuts but the orange, which indicates the asymmetry of the cell. As the trophozoite-stage parasite is located in the centre of the RBC, the cuts uniformly represent the characteristic cavity. In the schizont stage, the height profile again looks rather uniform, albeit revealing the uneven surface of the cell. These observations are reproduced for the fluorescence and light microscopy images. In addition, the orange cut serves as measure for the asymmetry of the RBC, i.e. it shows a uniform profile, if the parasite is off-centred and follows the parasite cut, if the parasite is centred. A comparison of figure 4.28 to the radial cross-sections through the same cells (see figures 4.12, 4.18, 4.23) reveals that the cross-sections determined for each RBC capture the features associated with the presence of the parasite.

In this section, I introduced an approach for dimensionality reduction, which provides a simple measure for detecting the characteristic features of malaria parasites. The method localises the parasite by calculating the centre of gravity. As the presence of a parasite introduces changes in height or intensity, the centre is shifted relative to the geometric centre of the RBC. The straight through both centres then represents the cross-sections of the parasite. To evaluate the efficiency of my approach, I visually examined the detection rate of the parasite cut, which revealed a hit rate of $> 50\%$ for ring- and $> 85\%$ for trophozoite-stage parasites after removing background noise from the images. As the symmetry properties of the RBCs profiles contain information about the stage and also the location of the parasite, I supplemented the parasite cut with three additional cuts. In agreement with the observations of characteristic cell features in section 4.3, my method for dimensionality reduction successfully compresses two-dimensional images into a set of single cuts.

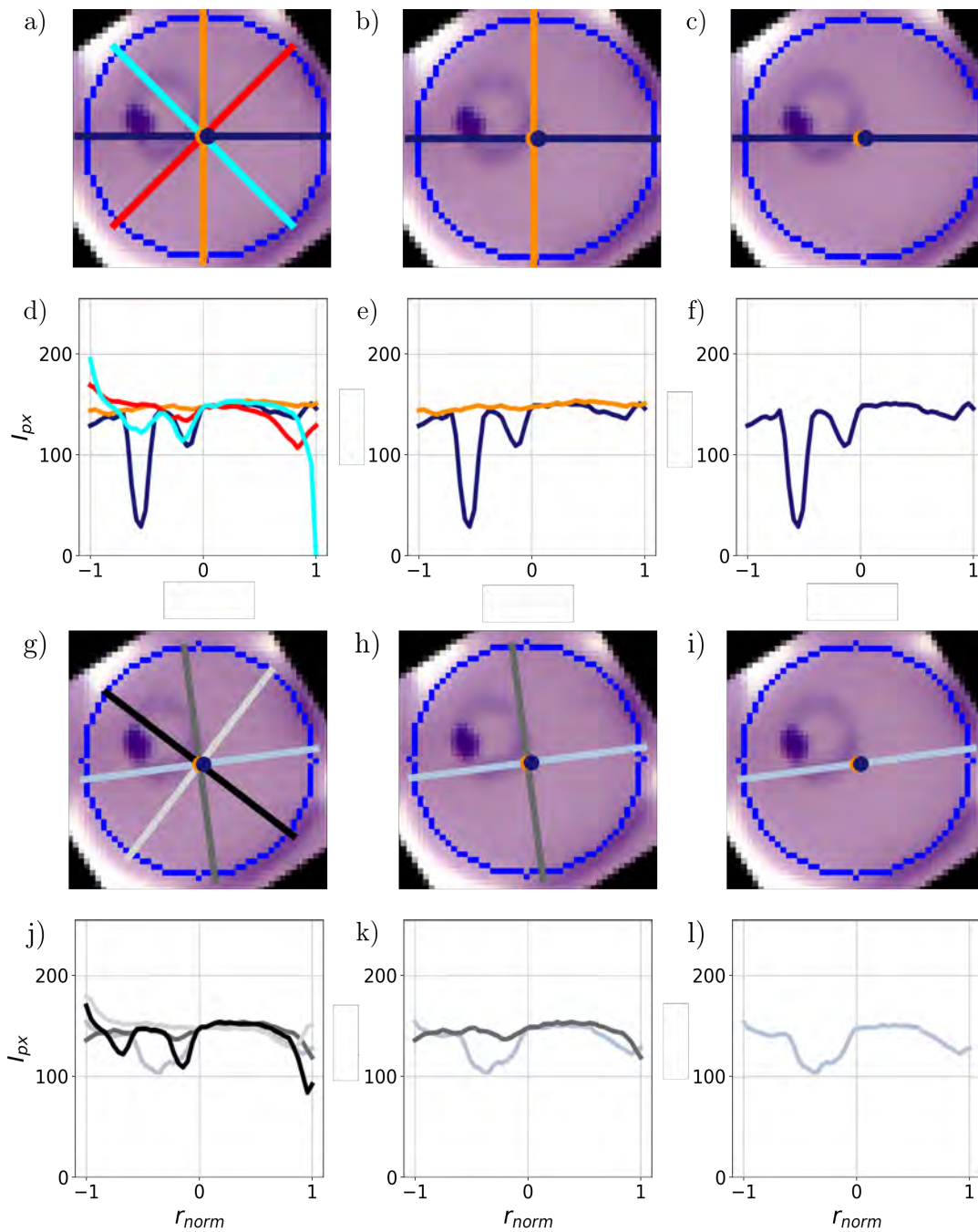


Figure 4.27: **Characteristic cross-sections of a parasite.** Light microscopy image of a ring-stage parasite with the four cuts (a), parasite plus cuts (b), and parasite cut (c). (d) - (e): Corresponding intensity profiles for each cut. (g)-(l): Light microscopy image with four randomly oriented cuts and corresponding intensity profiles in greyscale colours.

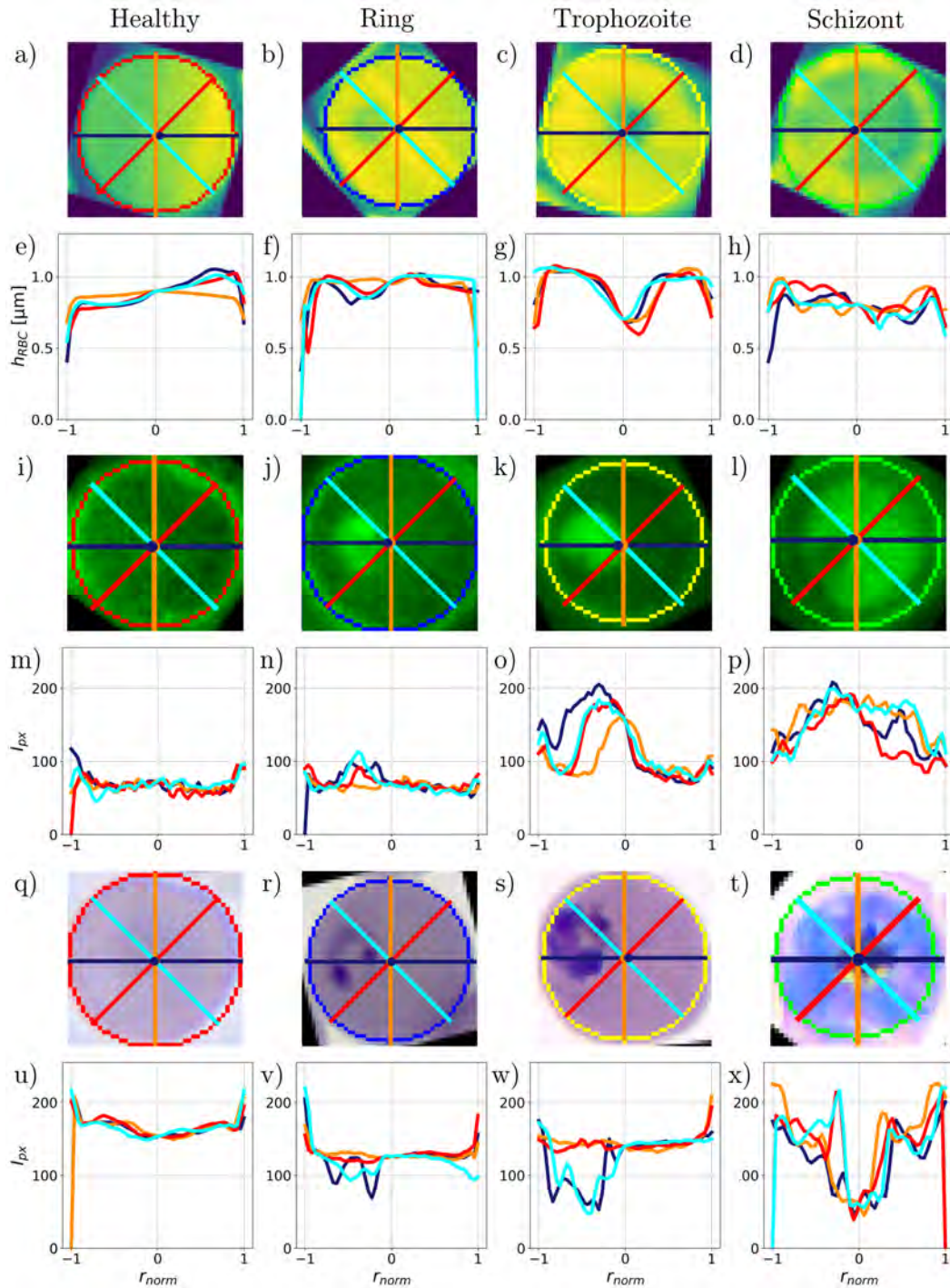


Figure 4.28: **Cross-sections of malaria-infected RBCs imaged with atomic force, fluorescence, and light microscopy.** Atomic force (a) - (h), fluorescence (i) - (p), and light microscopy images (q) - (x) of representative intra-erythrocytic stages with corresponding stages. The coloured circles show the cell contour, each colour denoting the intra-erythrocytic stage, as defined in section 4.2.3. The parasite cut is shown in blue and the other cuts in orange, cyan, and red. Geometric and gravitational centre are denoted in orange and blue.

4.5 Stage-specific classification of *P.falciparum*

For the establishment of neural networks as diagnostic tools for disease control, they are required to work fast and reliable in order to support or replace human perception. One of the central points is the increase in throughput of diagnosis, whilst ensuring low image processing time. In addition, such networks are designed to eliminate human errors to improve the reliability of disease detection and surveillance. In case of malaria diagnosis, it is helpful, if the network is not only able to distinguish between healthy and infected RBCs but also between the intra-erythrocytic stages for effective treatment. Creating new networks based on these guidelines is a important for the fight against malaria, especially in regions with low resources. As preparation for the implementation of a neural network-based stage-specific classification, the previous sections focused on understanding and categorising morphological and optical features of RBCs and compressing those into one dimension. How this influences the success of CNNs compared to two-dimensional images, will be explored in this section. Starting with the architecture of the NN, the second part focuses on how the appropriate selection of RBC features influences the performance of the classification. The networks were run on a standard laptop with windows 10, x64-based processor, and a processor speed of 2.50 GHz.

4.5.1 Neural network architecture

The success of NNs not only depends on the input data but also on its architecture. When dealing with images, the most suitable models are CNNs. As described in section 3.4.3, they contain convolutional, pooling, and fully-connected layers. These layers should be combined in such a way as to fulfil the following conditions: applicability to RBC images and stage-specific cuts, classification into four categories: healthy, ring, trophozoite, schizont, minimum performance of $> 90\%$, smallest possible amount of fit parameters to guarantee short computation time. The models were tested on a set of malaria-infected RBC images, recorded by AFM and balanced with data augmentation by rotation. For a representative result, all models were run for 20 epochs. In each model, the convolutional layers are followed by a pooling layer and two fully-connected layers at the end of the network. In table 4.13, the influence of the amount of feature maps is shown on a selection of the tested architectures. Model M4 provides the best compromise between the amount of fit parameters and classification accuracy. As the network was prone to overfitting, I added the kernel regulariser $L2$ with a hyperparameter of 0.001 to the first fully-connected layer. The complete architecture for the two possible inputs is given in figures A.1 and A.2. Instead of a 2D convolutional layer, the network for the classification of single cuts contains a one-dimensional convolutional layer, which reduces the number of fit parameters to 34494. Thereby, the network fulfils all conditions imposed at the beginning of the section. The network not only applicable to images and cuts but also able to classify the input into four categories. In addition, the model performed with 96% accuracy

Network	feature map	performance	parameters
M1	4×32	80%	81566
M2	$2 \times 32, 2 \times 64$	87%	68446
M3	4×64	92%	114878
M4	$1 \times 32, 3 \times 64$	96%	96126

Table 4.13: Classification accuracy on a test set of RBCs for various network architectures.

on the test data with a relatively low amount of parameters. In the further course of the thesis, both network architectures are used for classification.

4.5.2 Stage-specific classification of two-dimensional RBC images

Neural networks have proven to successfully classify RBCs in microscopy images. With a few exceptions [17], most of the approaches deal with Giemsa-stained light microscopy images [54–57]. Inspired by these works, I designed my network in such a way as to recognise not only light but also atomic force and fluorescence microscopy images. For the evaluation of the network performance, the RBC images were divided into a random training and test set at a ratio of 90% to 10%. The first test of the NN was performed on the set of single RBC images, obtained in section 4.2.2. These cut-outs were classified into four categories: healthy, ring, trophozoite, and schizont. The results are summarised in confusion matrices, presented in figure 4.29(a), (c), and (e), showing the performance of the classification for each microscopy technique on the test set. The labels of the rows are the categories predicted by the NN-based classifier, while the labels of the columns indicate the classification by human experts. Correspondingly, the sum over a given column/row gives the number of cases falling into that category by human/NN classification. The confusion matrix further shows the precision, recall, and overall accuracy values defined in section 3.4.4. For visual guidance, the diagonal fields with percentages $>18\%$ are highlighted by a blue background, while fields with the precision values (last column) have background colours according to the common colour bar on the right. The overall accuracy of 83-95%, as shown in the grey box is the result of the extremely high success rate on the healthy cells. As a known consequence of a strongly imbalanced set, when 80-90% of the data belongs to one category, the neural network is prone to overfitting, which significantly reduces the success rate on the minor categories. Therefore, the network is not able to reach a performance beyond the percentage of data belonging to the largest category. Thus, in the present case, the NN always predicts a healthy RBC.

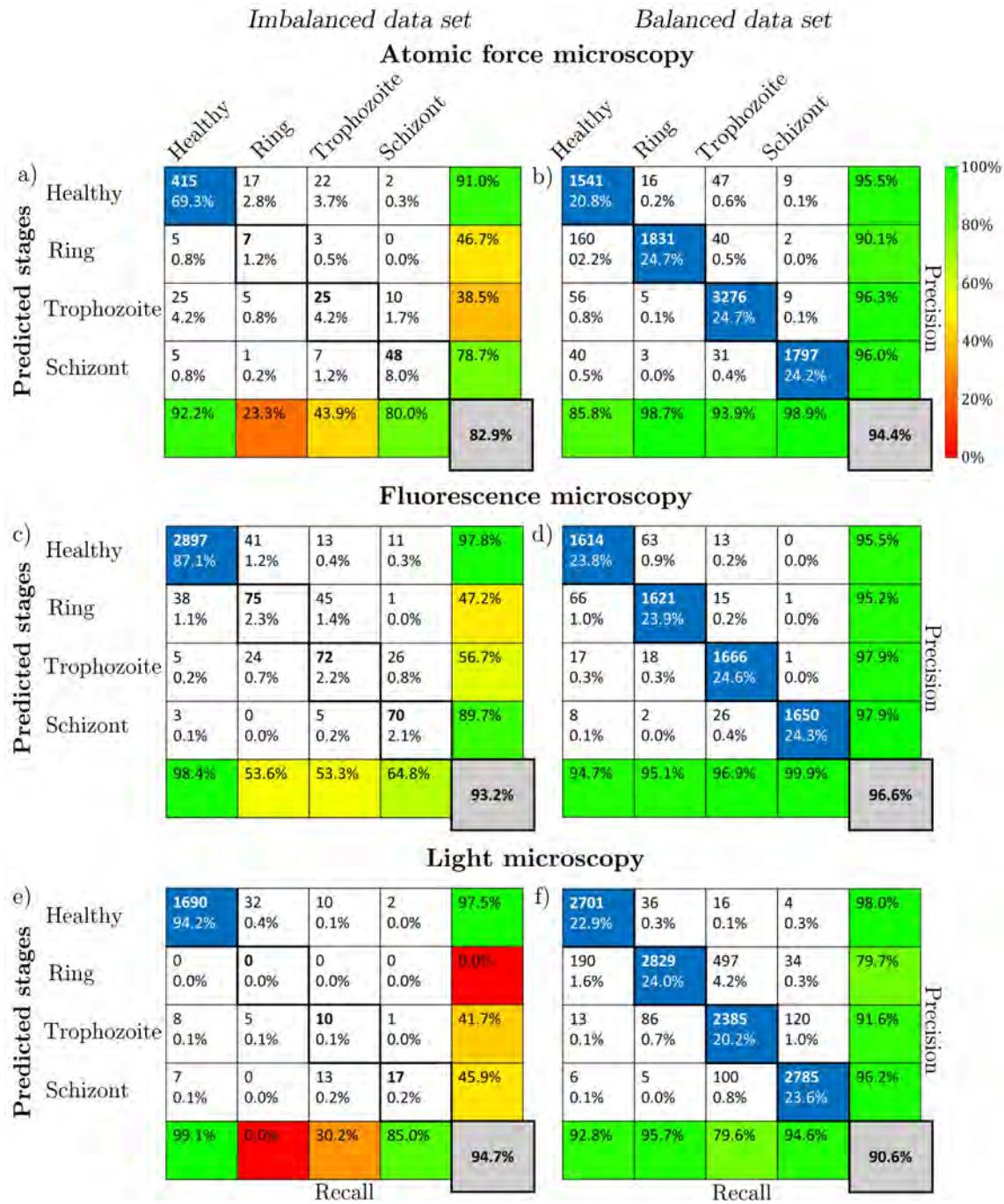


Figure 4.29: **Classification performance of the NN on microscopy images.** The performance parameters, as obtained on imbalanced (non-augmented) sets of atomic force, fluorescence, and light microscopy images, are respectively summarised in the confusion matrices of panels (a), (c), and (e). The confusion matrices obtained for the corresponding balanced (augmented) data sets are shown in panels (b), (d), and (f).

For this reason, I implemented a data augmentation method to balance the data set, see section 4.2.3. Figure 4.29(b), (d), and (f), shows the confusion matrices of the tests, performed by using the same NN but a balanced data set. The results reveal a large improvement with respect to the non-augmented data set. Except for the prediction of trophozoite stages in fluorescence microscopy images, the precision of the prediction (last column of the confusion matrix) is above 90% for all four RBC categories and imaging techniques. Overall, the NN achieved a high classification accuracy for all microscopy techniques, 94.4% for atomic force, 90.6% for fluorescence, and 96.6% for light microscopy.

A dive into the stage-specific precision on the augmented data sets shows that early stages of malaria, i.e. rings and trophozoites, are best recognised in Giemsa-stained light microscopy images, while the schizont stage is captured most in atomic force and fluorescence microscopy images. The precision levels further reveal a high sensitivity to the quality and contrast of the images. Due to the high amount of training parameters for the NN, the analysis of the two-dimensional images required high computation time, ranging from 117 ms per RBC for atomic force to 49 ms for fluorescence, and 48 ms for light microscopy images. The performance reached for the classification of each microscopy technique is similar to that achieved, e.g., in reference [57], showing that the NN presented in this section qualifies well for the purpose of a stage-specific detection of malaria. As the balancing of the categories significantly improved the performance of the NN, throughout the thesis, the balanced data set will be used as input and for the dimension reduction in the next section.

4.5.3 Evaluation of the stage-specific classification on characteristic features of single-RBC images

Dimension reduction is a well established tool to improve the performance of NNs. Following the thorough analysis of morphological and optical properties of RBC, presented in section 4.3, I compressed the cell features into one-dimensional cuts, thus capturing the presence or absence of the parasite. Therefore, these cuts contain the most important features of the RBC. In the following, the NN introduced for the classification of two-dimensional images is tested on various combinations of the 1D cuts. The section is concluded with a comparison to a set of randomly chosen cuts.

Radial cross-sections

As shown in section 4.3, the radial cross-sections contain information about the presence of the parasite, its asymmetry and also the location of the food vacuole. Therefore, they are well suited as input for the stage-specific classification. The test results, displayed in figure 4.30, range from 95.4% for AFM to 97.4% for light microscopy, revealing a significant improvement compared to the 2D images. Moreover,

the training time per RBC was significantly reduced to 4.0 ms for atomic force, 4.1 ms for fluorescence, and 4.2 ms for light microscopy.

A detailed investigation of the stage-specific predictions revealed the lowest precision for the ring-stage parasites. As their height or intensity profile closely resembles that of a healthy RBC, it can lead to false detection.

Characteristic cuts

Based on the observations in section 4.4, the one-dimensional cuts can be used to represent the characteristic cell features. In contrast to the radial cross-sections, only the most important features are captured. In the following, the performance of the NN on three different combinations of the cuts is evaluated: four cuts, the parasite plus cuts, and the parasite cut, as calculated in section 4.4. Table 4.14 shows the most relevant performance parameters for all three combinations. More details are presented in figures A.3-A.5. The results show that the reduction of dimension

Imaging technique	Cuts	Network performance
AFM	Four cuts ($4n$)	98.8%
	parasite plus cuts ($2n$)	99.3%
	parasite cut ($1n$)	98.0%
Fluorescence microscopy	Four cuts ($4n$)	97.5%
	parasite plus cuts ($2n$)	98.0%
	parasite cut ($1n$)	95.3%
Light microscopy	Four cuts ($4n$)	97.7%
	parasite plus cuts ($2n$)	98.7%
	parasite cut ($1n$)	96.9%

Table 4.14: Classification accuracy on the four cuts, parasite plus cuts, parasite cut.

to four cuts ($4n$) significantly improved the performance of the NN with respect to the classification of the 2D images and the radial cross-sections, increasing the accuracy to 98.9% for the atomic force, 97.5% for the fluorescence, and 97.7% for the light microscopy images. With further reduction of the input dimension to two cuts ($2n$), the performance increases to 98.0% - 99.3%. This proves to be the maximum possible accuracy, as the performance on the single parasite cut is lowered to 98.0% for the atomic force, 95.3% for the fluorescence, and 97.7% for the light microscopy images. Furthermore, for all combinations of cuts, the computation time was reduced to 3.7 ms for atomic force, 3.9 ms for fluorescence, and 3.8 ms for light microscopy. An additional test series on the characteristic cuts calculated without removing background noise revealed a lower classification accuracy on all three cut

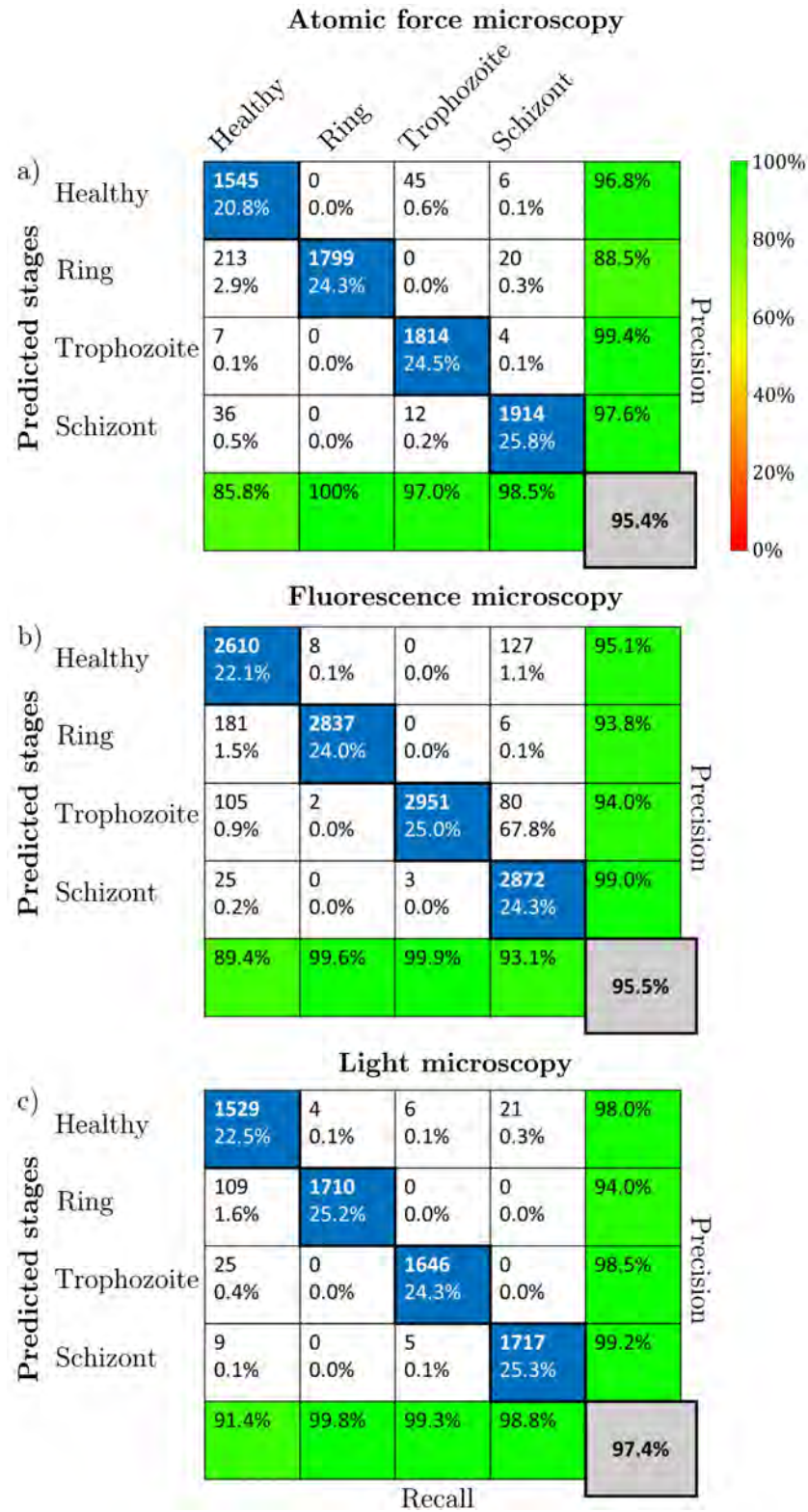


Figure 4.30: **Classification of the NN on radial cuts.** The confusion matrices show the summary of performance parameters for atomic force (a), fluorescence (b), and light microscopy (c), as obtained using radial cuts through single-RBC images. Notations, labels and colours follow the same convention as in Figure 4.29.

combinations, see table 4.15, as compared to the cuts extracted from processed images (Table 4.14). Corresponding performance parameters are presented in figures A.6-A.8. The limitations discussed for the success rate of the parasite cut influence

Imaging technique	Cuts	Network performance
AFM	Four cuts ($4n$)	98.5%
	parasite plus cuts ($2n$)	97.3%
	parasite cut ($1n$)	97.7%
Fluorescence microscopy	Four cuts ($4n$)	95.4%
	parasite plus cuts ($2n$)	96.7%
	parasite cut ($1n$)	93.9%
Light microscopy	Four cuts ($4n$)	97.3%
	parasite plus cuts ($2n$)	98.6%
	parasite cut ($1n$)	96.6%

Table 4.15: Classification accuracy on the four cuts, parasite plus cuts, parasite cut. The orientation of the cuts was calculated without removing background noise.

the performance of the NN, as it gets the most characteristic features from this step. While incorrectly classified cells mostly originate from this issue, the NN still outperforms the success rate, as the relation between both cuts also contributes to the classification. This statement is supported by tests on the sensitivity of the NN to the order of the parasite cut and its perpendicular counterpart by swapping both cuts in the test set, see table 4.16, which revealed a significant reduction of the performance to $< 53\%$. These observations further corroborate the parasite plus cuts are the best choice for the stage-specific classification. To conclude the evaluation of the network

Imaging technique	Accuracy
AFM	37.7%
Fluorescence microscopy	52.7%
Light microscopy	37.2%

Table 4.16: Prediction accuracy of swapped parasite plus cuts ($2n$).

performance, the three combinations of cuts were replaced with randomly oriented cuts, spanning the dimensions $4n$, $2n$, and $1n$. Representative images are shown in figures A.9-A.11. Figure 4.31 shows the results on random cuts with relation to the parasite plus cuts (ppc) for atomic force (AFM, black), fluorescence (FM, green), and light microscopy (LM, orange).

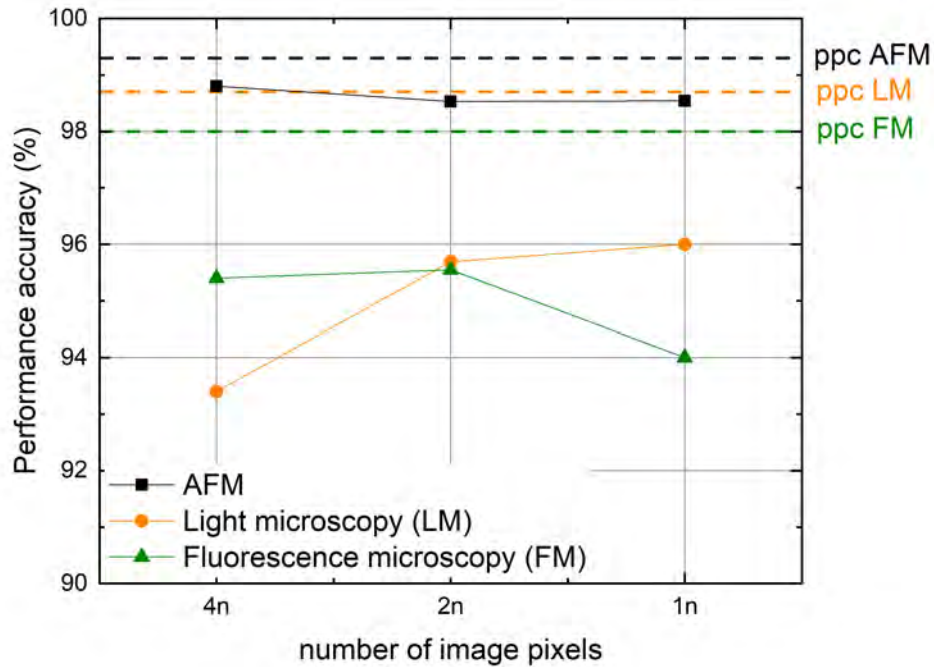


Figure 4.31: **Classification performance of the NN on randomly oriented cuts.** The random cuts span the dimensions $4n$, $2n$, and $1n$. Their performance is shown for atomic force (black), fluorescence (green), and light microscopy (orange) in relation to the parasite plus cuts (ppc) ($2n$), denoted by the dashed lines, separately labelled for each microscopy technique.

The combination of parasite plus cuts clearly exceed the performance of the NN on randomly oriented cuts in all three tested dimensions, irrespective of the imaging technique. Ranging from 98.5% to 98.8%, the classification accuracy for atomic force microscopy is lowered by 0.5%. For fluorescence microscopy, the reduction is 2.1% and 2% for light microscopy. As well as the previous observations, the tests on random cuts again prove that the parasite plus cuts provide the highest classification accuracy and therefore best represent the characteristic features of healthy and malaria-infected RBCs.

In this section, I designed a neural network for the stage-specific classification of malaria-infected RBCs. On this basis, I conducted a test series to determine the performance of the NN on a data set of RBC images. In the further course of the section, I showed how the input dimension influences the classification. The performance results are summarised in figure 4.32, which shows the overall accuracy of the malaria stage detection versus the step-wise reduction of input data supplied to the NN from n^2 to $1n$ for all three imaging methods. These results are further supplemented with the computation time per RBC.

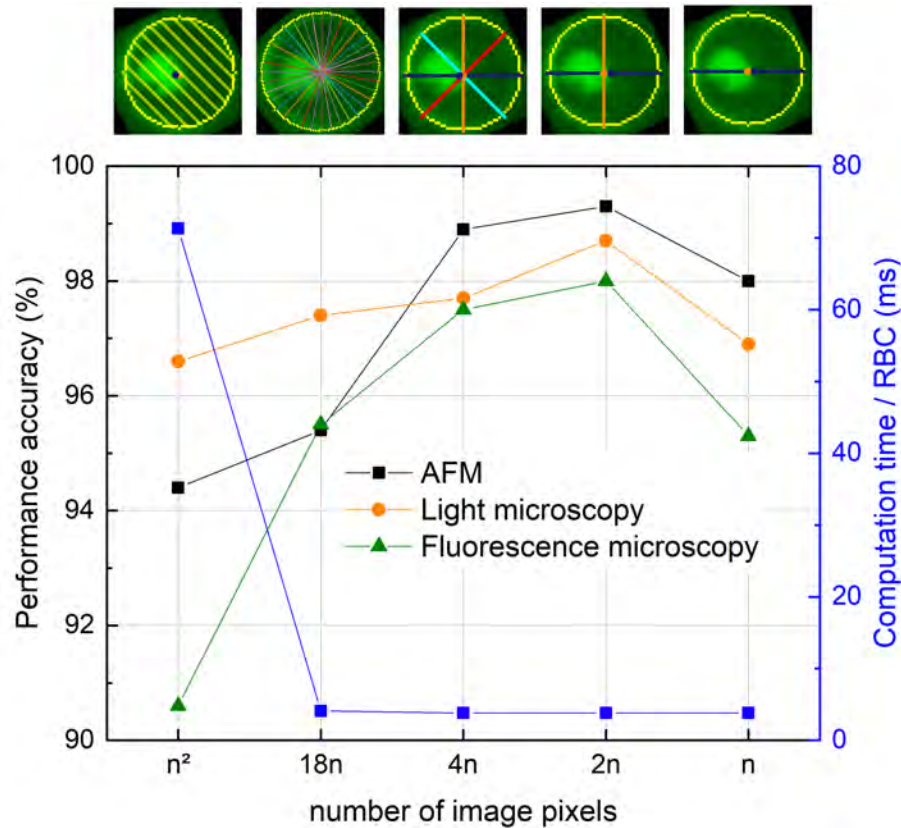


Figure 4.32: **Classification performance of the convolutional NN and computation time per RBC versus the amount of input data** for light microscopy, AFM, and fluorescence microscopy images. The NN was trained and tested on 2D RBC images (n^2 pixels), four cross-sections ($4n$), two characteristic cross-sections ($2n$), and one single cross-section (n), the asymmetric cut. These four cases are illustrated by exemplary fluorescence images on the top. Asymmetric cuts are indicated with blue lines, while the 90° , $+45^\circ$ and -45° cuts are shown in orange, cyan and red, respectively.

The reduction of n^2 to $18n$ significantly increased the performance of the NN to $> 95\%$, providing more stability to the network. While the four characteristic cuts boosted the classification accuracy to $> 97.5\%$, the optimum is achieved at $2n$ for the parasite plus cuts. With further reduction to $1n$, the performance of the NN drops to $< 98\%$. This observation suggests that the additional cut helps the system to work on a differential basis and not be affected by the actual values of the height or intensity. As another important effect discernible in figure 4.32, the reduction of dimension not only improves the classification but also reduces the computation time per RBC by a factor of 18 from 71.3 ms to 3.8 ms. While the calculation of the gravitational centre requires 115 ms per RBC, the overall computation time for 2D images already exceeds this additional time span by 12%, when training for more than one epoch. Furthermore, the lower performance on random cross-sections shows that

the orientation of the parasite cut plays a large role in the stage-specific classification. As shown in section 4.4, the success rate of the parasite cut was significantly lower for images with background noise. With tests on cuts extracted from such unprocessed images, I was able to show that this leads to a reduction in the performance of the NN. Together with the observed dependence of the classification accuracy on the order of the cuts, the test results corroborate the parasite plus cuts as the best representation for two-dimensional images of healthy RBCs and the intra-erythrocytic stages. Additionally, I proved that the concept of reducing images to two characteristic cuts is applicable on various imaging methods and therefore has the potential to be extended for the analysis of general objects.

4.6 The malaria stage classifier

The final thesis point was to create an efficient and reliable method for the stage-specific detection of malaria blood stages, which is able to handle images from multiple microscopy techniques. In this section, the results presented in the four preceding sections, are combined into an application. Following the structure of the thesis, the application is divided into four main steps: (i) imaging, (ii) cell identification, (iii) dimension reduction, and (iv) classification. Starting with the an overview of the algorithm, the architecture of the application is described. Subsequently, the section ends with a detailed explanation of how the application is used.

4.6.1 Design and Implementation

The Malaria Stage Classifier is designed to facilitate and accelerate the staging of malaria infected RBCs in microscopy images. Due to its robustness against imaging platform-specific features, it is applicable to a wide range of light microscopy images. The interface of the application is arranged in tabs, which makes it easy to follow the image processing steps. The Malaria Stage Classifier further offers the possibility to manually optimise the cell detection and classification.

From the microscopy measurement, the algorithm receives two kinds of input, text files from atomic force microscopy, as the evaluation software (Igor Pro 6.37) used for the AFM measurements only allows for exporting text files, and images from fluorescence and light microscopy. Both inputs are treated as matrices and converted to greyscale for further processing. Due to the strong contrast between cell and parasite, the conversion did not influence the classification. In case of atomic force and some light and fluorescence microscopy images, the contrast between background and RBCs is not strong enough to locate single cells, which impacts the accuracy of the detection. Hence, the images are binarised based on pixel intensities by Otsu's method [128]. While the processing of atomic force microscopy images requires an additional step, it is sufficient to enhance brightness, sharpness, and contrast in the light and fluorescence microscopy images. Python offers a module for automatic enhancement of images by a manually chosen factor [129], which can be applied to

highlight the RBCs in contrast to the image background. The processed images are then used for cell detection, employing the Hough gradient method. While the detection parameters are preset, they can be manually adjusted by the user. This step is then followed by the reduction of the image dimension to the parasite plus cuts, containing the characteristic features of the cells. In the final part, the detected RBCs are classified by the NN presented in section 4.5.1.

The underlying network was taken from the classification of the parasite plus cuts, see section 4.5.1, which showed the best performance on the RBCs. The training, test and validation data are provided within the software package for each of the microscopy techniques. As the data base is relatively small for a neural network, the data set can be extended by adding more cross-sections after the stage classification step. The prediction for each cell is returned in form of an image, where all cells are assigned to one of the four categories. In case of false classifications, the algorithm offers the option to change the assigned category. The detected cells can be added to the pre-trained NN to increase the training data set, which is then retrained with the new input. After the classification of the microscopy images, the results can be saved as a csv or text-file, returning the amount and ratio of healthy RBCs and intra-erythrocytic stages.

The Malaria Stage Classifier is accessible as a GUI, which is arranged in tabs. It has been developed using Python 3.7 with the following dependencies: *numpy* [141] and *pandas* [142] for the data analysis. For the cell detection, *OpenCV* [143] together with *skimage* [144] and *matplotlib 3.5.2* [145] for the data visualisation are needed. To integrate the pre-trained NN, *tensorflow* [146] with *keras* has to be included in the algorithm. The interface further requires *tkinter* [147] and the libraries *os*, *sys*, *csv*, *traceback* for handling errors and the output files, as well as the library *webbrowser* to open system folders and external links.

Common errors are prevented by messages, detailing the problem and providing a solution. Furthermore, the errors saved in a log file to allow for bug-fixing by any developer. While the user is able to modify a few parameters through the GUI, suitable values are suggested for each case and all parameters can be set back to the default values.

4.6.2 User manual

The package is available open source on git hub repository https://github.com/KatharinaPreissinger/Malaria_stage_classifier or the archive <https://zenodo.org/record/7261800>. In the following, the program structure is explained on a sample atomic force microscopy image. The interface of the package is built with five tabs, where each performs one step of the stage-specific classification of RBCs. Starting with the general settings, the algorithm only allows text or image files as input. Depending on the file type, the user can set the number of header lines the input text file contains. In case of the sample image, the value is set to three and

confirmed by the button “Set file format”. As the program is designed for multiple imaging methods, the respective technique determines the possible actions on the input file. The settings are completed by selecting an output path and the output file type, see figure 4.33.

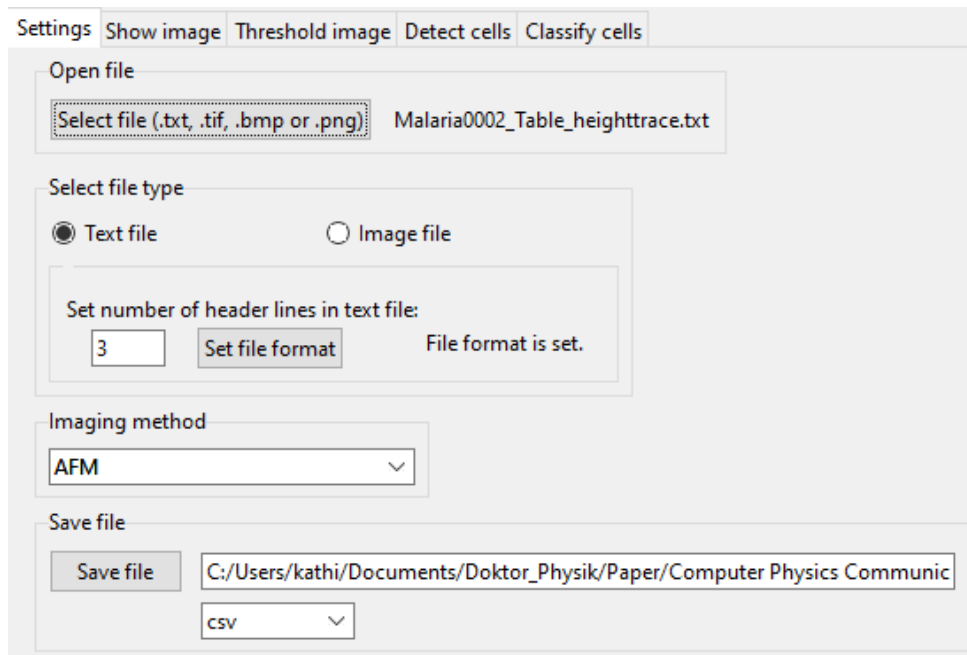


Figure 4.33: **Setting of initial parameters:** input image, its format, imaging method, memory location of the results

In the second tab, the input image is displayed to show the location and stage of the cells. A screenshot of the interface is presented in figure 4.34.

If the image was recorded by AFM the next step requires thresholding to enhance its contrast. While the value can be adapted by the user, the algorithm suggests a pre-calculated number, as shown figure 4.35. In case of the light microscopy techniques, this step is skipped.

The next tab triggers the cell detection algorithm. Depending on the size, contrast, and brightness of the image, the algorithm offers the option to control the detection by manually changing the calculation parameters and further allows contrast enhancement. Optionally, the parameters can be set back to the default value. During this step, the dimension of the input data is reduced to the two characteristic cuts to capture the strongest features associated with the presence or absence and the stage of the malaria parasites, see figure 4.36.

In the final part of the program, the user can load the pre-trained NNs and start the stage-specific prediction of the detected RBCs. This triggers the option to change false predictions accordingly. The algorithm further provides the possibility to add new data and to retrain each NN. The last step then returns the statistics of the

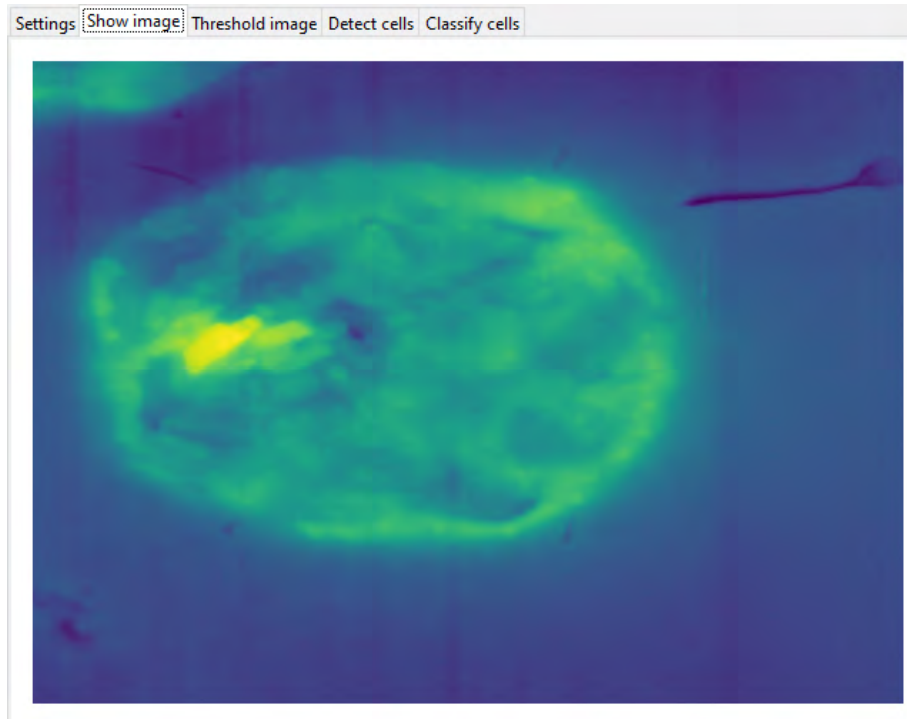


Figure 4.34: **Show image.** The input file is shown as image.

analysed RBCs in form of a text file or table. The corresponding tab is shown in figure 4.37.

In this section, I presented a software, which I developed for the stage-specific detection of malaria in RBC images. The detection is based on dimensionality reduction, using only two characteristic cuts to classify the cells. The program is freely accessible and can handle text data and images from various microscopy techniques as input. The user interface is divided into tabs, making it easy to follow the step-wise analysis of the data from input to cell detection and finally the classification of the RBCs. It also allows manual corrections of the predictions, which are saved in a table. Due to its simple concept, the method can easily be extended beyond the evaluation and classification of parasites and cells to include the analysis of arbitrary objects. Given the rising number of techniques, successfully applied for the imaging of RBCs, the algorithm can be augmented for any method with high contrast, formatted as text or image file. Moreover, it offers the opportunity to increase the data set used to train the NN to improve the classification accuracy. After loading the NN and visually verifying the accuracy of the classifications, the characteristic single-RBC cuts can be added to the original images to retrain the NN with the new data set. The algorithm is available on https://github.com/KatharinaPreissinger/Stage_specific_classification_of_RBCs. Details about the implementation can be read in the appendix B or on <https://stage-specific-classification-of-rbc.readthedocs.io/en/latest/index.html#>



Figure 4.35: **Threshold image.** This tab provides the option to enhance the image contrast by thresholding (disabled for light and fluorescence microscopy)

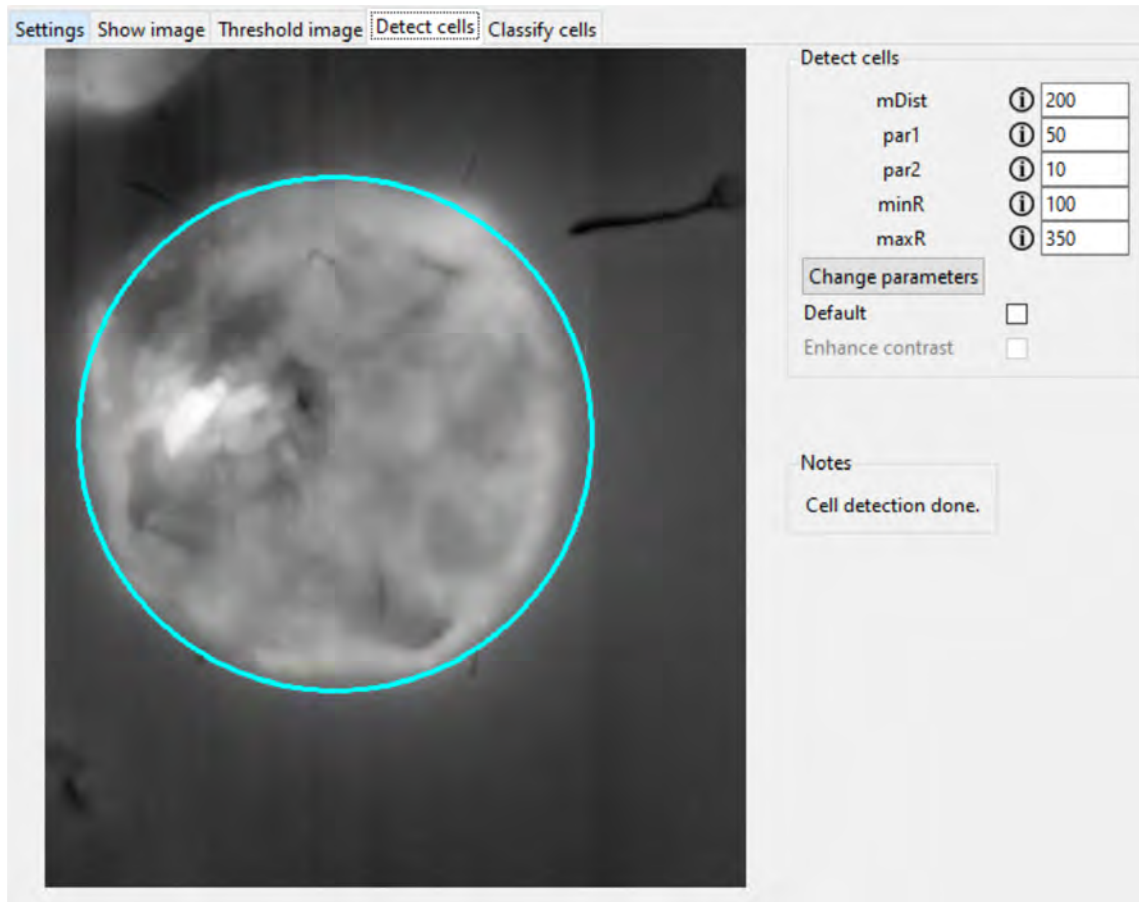


Figure 4.36: **Detect cells.** RBCs in the images are detected by the cell detection algorithm. To fine-tune, the algorithm offers the possibility to enhance the image and to set the parameters for the detection manually.

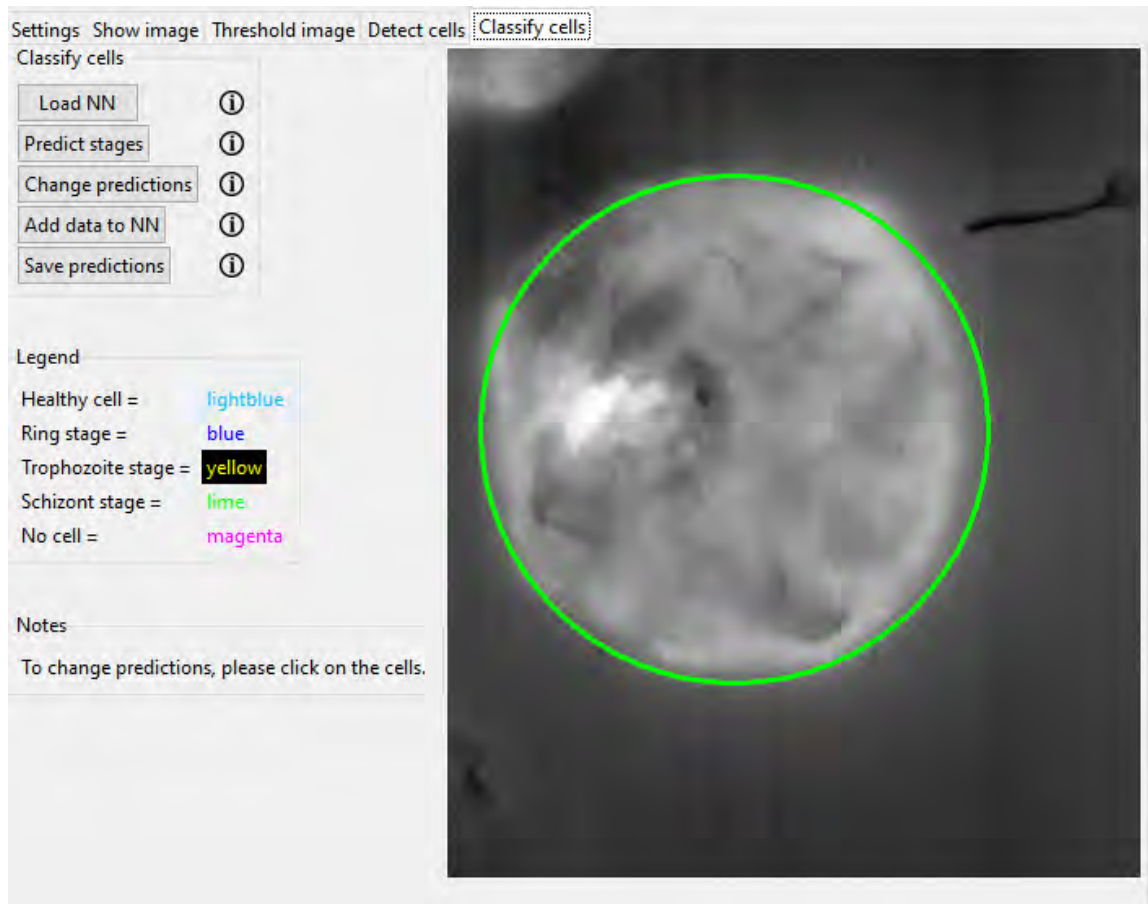


Figure 4.37: **Cell classification.** The intra-erythrocytic stages are predicted and can be altered manually. Optionally, the neural network can be retrained with new data.

Summary and thesis points

Caused by global climate change and the ongoing COVID-19 pandemic, the numbers of malaria infections are increasing again. This development imposes numerous challenges for effective malaria treatment and diagnosis. As the intra-erythrocytic cycle of the malaria parasite causes the main clinical symptoms, gaining insight into the mechanical and optical alterations in RBCs inhibited by *P. falciparum* is the major target of antimalarial treatment and diagnostics. Moreover, especially in areas with low resources, early detection is difficult and closely linked with access to laboratories and trained scientists. Due to this, the application of NNs has become one major area in the field of malaria research, as it has the potential to increase the throughput in diagnosis and eliminate human error. During my PhD work, I focused on these two major fields of malaria research in order to provide a deeper understanding of the changes in RBCs inhibited during the intra-erythrocytic cycle and to present a new neural network-based approach to improve and facilitate the diagnosis of the disease. The most important results obtained in this work are summarised in the following thesis points.

1. High-parasitemia cultures for in vitro analysis of *P. falciparum* under physiological conditions

I tested two methods for the immobilisation of RBCs on a glass surface in liquid environment by exposing them to cantilevers of different spring constant. Furthermore, I calculated the required concentration of RBCs to obtain a single cell layer on the sample surface and confirmed the calculated value by tests with various concentrations. With the determined settings, I successfully imaged healthy RBCs under physiological conditions. Based on these experiments, I derived a sample preparation routine for imaging RBCs under physiological conditions, which can also be applied for imaging of the autofluorescence in living cells. These results provide the basis for the studies of optical properties in RBCs infected with *P. falciparum*.

2. Exploring characteristic features of malaria-infected RBCs

To create a data base for studies of morphological and optical properties of malaria-infected RBCs, I recorded all atomic force and fluorescence microscopy images used in this thesis. I designed an algorithm to extract single cells from microscopy images and

evaluated its performance. I found that the algorithm can detect cells with a precision of more than 95%. Furthermore, I showed that the presented detection method is not limited to a certain type of imaging technique. This was a crucial step for the analysis of single RBC images, which showed that the morphological and optical changes during the maturation of the parasite are reflected in the height and fluorescence intensity profiles of the host cell. Furthermore, I revealed diagnostic patterns in the topographical structure of infected erythrocytes, showing close correlation with their fluorescence map. I showed that atomic force and fluorescence microscopy can be used to locate hemozoin crystals based on the topographical and optical features. These observations revealed connections between the structure and the hemozoin content of RBCs, and their consequences on the optical properties. I further confirmed statistic significance between average profiles characteristic to the intra-erythrocytic stages. Additionally, I showed that malaria-infected RBCs under physiological conditions emit fluorescence patterns similar to the patterns I observed in thin blood films.

Publication 1

3. Reduction of dimensionality as a tool for feature selection in RBCs boosts the stage-specific classification of *P.falciparum*

I showed that in malaria-infected RBCs, the most important features are associated with the presence or absence of the parasite. I determined the position of the parasite by using the gravitational centre, which is shifted by the presence of the parasite. I further tested the influence of background noise on the position of the gravitational centre and showed that removing background significantly improved the localisation method. In the next step, I used this method to reduce the two-dimensional images of single RBCs to the parasite cut, which goes through the parasite, and three additional cuts spanning 90° and $\pm 45^\circ$ with it. The experiments revealed that this method for parasite localisation works with high accuracy on atomic force, fluorescence, and light microscopy images, providing the characteristic properties of RBC in the form of four one-dimensional cuts.

To find a suitable network for the stage-specific classification of RBC, I tested various architectures of networks. In the next step, I used the network with the highest performance for the classification of single RBC images and the cuts characteristic to the parasite features. I found that the network reaches a sensitivity of $> 87\%$ on the 2D images. By the smart reduction of data dimension to $2n$ with a careful selection of features, I significantly boosted the performance of the NN-based classification to $> 96\%$, independent of the microscopy technique. I demonstrated that the characterisation method I developed in my work captures the most important features of extracted, single RBCs from microscopy images and provides a reliable tool for an automatised stage-specific recognition of malaria. **Publication 2**

4. The malaria stage classifier

I developed a software package for the neural network-based stage-specific detection of malaria. I wrote a user friendly interface and implemented dimension reduction method introduced in the previous point. In the algorithm, I implemented the cell detection introduced in the second point. I further wrote a documentation, detailing

the use and implementation of the package. In addition, the package provides a pre-trained NN for the blood stage classification and further offers the possibility to retrain the NN with additional data. **Publication 3**

5.1 Publications related to the thesis points

1. **Preißinger, K.**, Molnár, P., Vértessy, B. G., Kézsmárki, I. & Kellermayer, M. Stage-Dependent Topographical and Optical Properties of Plasmodium Falciparum-Infected Red Blood Cells. *J Biotechnol Biomed* 4 (3): 132-146 (2021). doi: 10.26502/jbb.2642-91280040 **IF: 0**
2. **Preißinger, K.**, Kellermayer, M, Vértessy, B. G. Kézsmárki, I. & Török, J. Reducing data dimension boosts neural network-based stage-specific malaria detection. *Sci. Rep.* 12(1):1-14 (2022). doi: 10.1038/s41598-022-19601-x **IF: 5.5**
3. **Preißinger, K.**, Kézsmárki, I. & Török, J. An automated neural network-based stage-specific malaria detection software using dimension reduction: The malaria microscopy classifier. *MethodsX* Accepted **IF: 1.9**

Bibliography

- [1] Tedros Adhanom Ghebreyesus. World malaria report 2021. Technical report, World health organisation, 2021.
- [2] Organization World Health. Malaria. <https://www.who.int/news-room/fact-sheets/detail/malaria>, 2019.
- [3] J. Farrar and Et al. *Manson's tropical infectious diseases*. Saudner Ltd., 2014.
- [4] Lawrence H Bannister and Irwin W Sherman. Plasmodium. In *Encyclopedia of Life Sciences (ELS)*. John Wiley & Sons, 2009.
- [5] World Health Organization. *Basic malaria microscopy*. World Health Organization, Geneva, 2nd edition, 2010.
- [6] S.A. Harvey, S. Incardona, and N. Martin. Quality issues with malaria rapid diagnostic test accessories and buffer packaging: findings from 5-country private sector project in Africa. *Malaria Journal*, 16(160), 2017.
- [7] Araia Berhane, Mulugeta Russom, Iyassu Bahta, Filmon Hagos, Michael Ghirmai, and Selam Uqubay. Rapid diagnostic tests failing to detect Plasmodium falciparum infections in Eritrea: an investigation of reported false negative RDT results. *Malaria Journal*, 16(105), 2017.
- [8] Dave M Newman, John Heptinstall, Raphael J Matelon, Luke Savage, M Lesley Wears, Jamie Beddow, Martin Cox, Henk D F H Schallig, and Petra F Mens. A Magneto-Optic Route toward the In Vivo Diagnosis of Malaria : Preliminary Results and Preclinical Trial Data. *Biophysical Journal*, 95:994–1000, 2008.
- [9] Agnes Orban, Maria Rebelo, Petra Molnar, Inês S. Albuquerque, Adam Butykai, and Istvan Kezsmarki. Efficient monitoring of the blood-stage infection in a malaria rodent model by the rotating-crystal magneto-optical method. *Scientific Reports*, 6:1–9, 2016.
- [10] A. Butykai, A. Orbán, V. Kocsis, D. Szaller, S. Bordács, E. Tátrai-Szekeres, L. F. Kiss, A. Bóta, B. G. Vértessy, T. Zelles, and I. Kézsmárki. Malaria pigment crystals as magnetic micro-rotors: Key for high-sensitivity diagnosis. *Scientific Reports*, 3:1–10, 2013.

- [11] Christof Grüning, Arlett Heiber, Florian Kruse, Johanna Ungefehr, Tim-wolf Gilberger, and Tobias Spielmann. Development and host cell modifications of *Plasmodium falciparum* blood stages in four dimensions. *Nature communications*, 2(165), 2011.
- [12] David Perez-guaita, Kamila Kochan, Mitchell Batty, Christian Doerig, Jose Garcia-bustos, Shirly Espinoza, Don Mcnaughton, Phil Heraud, and Bayden R Wood. Multispectral Atomic Force Microscopy-Infrared Nano-Imaging of Malaria Infected Red Blood Cells. *Analytical Chemistry*, 90:3140–3148, 2018.
- [13] Eriko Nagao, Osamu Kaneko, and James A Dvorak. Plasmodium falciparum-Infected Erythrocytes : Qualitative and Quantitative Analyses of Parasite-Induced Knobs by Atomic Force Microscopy. *Journal of Structural Biology*, 44:34–44, 2000.
- [14] Hui Shi, Zhuo Liu, Ang Li, Jing Yin, Alvin G.L. Chong, Kevin S.W. Tan, Yong Zhang, and Chwee Teck Lim. Life Cycle-Dependent Cytoskeletal Modifications in *Plasmodium falciparum* Infected Erythrocytes. *PLoS ONE*, 8(4):1–10, 2013.
- [15] Susan E. Francis, David J. Sullivan, and Daniel E. Goldberg. Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. *Annual Review of Microbiology*, 51:97–123, 1997.
- [16] Katharina Preißinger, Petra Molnar, Beata Vertessy, Istvan Kezsmarki, and Miklos Kellermayer. Stage-Dependent Topographical and Optical Properties of *Plasmodium Falciparum*-Infected Red Blood Cells. *Journal of Biotechnology and Biomedicine*, 4(3):132–146, 2021.
- [17] Paul Lebel, Rebekah Dial, Venkata N P Vemuri, and Valentina Garcia. Label-free imaging and classification of live *P. falciparum* enables high performance parasitemia quantification without fixation or staining. *Plos computational biology*, 17(8):1–29, 2021.
- [18] Arash Mehrjou. Automatic Malaria Diagnosis System. *International Conference on Robotics and Mechatronics*, 2013:205–211, 2013.
- [19] Mahdiah Poostchi, Kamolrat Silamut, Richard J Maude, Stefan Jaeger, George Thoma, Oxford Tropical, and Global Health. Image analysis and machine learning for detecting malaria. *Translational Research*, 194:36–55, 2019.
- [20] Mark C Mushabe, Ronald Dendere, and Tania S Douglas. Automated detection of malaria in Giemsa-stained thin blood smears. *Annual international conference of the IEEE engineering in medicine and biology society*, 2013(3698-701), 2013.
- [21] Naveed Abbas, Saba Tanzila, Dzulkifli Mohamad, and Amjad Rehman. Machine aided malaria parasitemia detection in giemsa-stained thin blood smears. *Neural Computing and Applications*, 29(3):803–818, 2018.

- [22] Katherine Torres, Christine M. Bachman, Charles B. Delahun, Jhonatan Alarcon Baldeon, Freddy Alava, Dionicia Gamboa Vilela, Stephane Proux, Courosh Mehanian, Shawn K. McGuire, Clay M. Thompson, Travis Ostbye, Liming Hu, Mayoore S. Jaiswal, Victoria M. Hunt, and David Bell. Automated microscopy for routine malaria diagnosis: A field comparison on Giemsa-stained blood films in Peru. *Malaria Journal*, 17(1):1–11, 2018.
- [23] Ishan R. Dave and Kishor P. Upla. Computer aided diagnosis of Malaria disease for thin and thick blood smear microscopic images. *2017 4th International Conference on Signal Processing and Integrated Networks, SPIN 2017*, 2017:561–565, 2017.
- [24] Nicholas J. White. Malaria. *Manson’s tropical infectious diseases*, (23):532–600, 2014.
- [25] U.S. Department of Health and Human Services. Biology-Lifecycle, 2020.
- [26] Kara Manke. Malaria-causing parasite manipulates liver cells to survive, 2022.
- [27] Zhenke Yang, Yang Shi, Huiting Cui, Shuzhen Yang, Han Gao, and Jing Yuan. A malaria parasite phospholipid flippase safeguards midgut traversal of ookinetes for mosquito transmission. *Science Advances*, 7(30), 2021.
- [28] BBC news. Glasgow scientists develop simple test for malaria in mosquitoes, 2022.
- [29] L H Bannister, J M Hopkins, R E Fowler, S Krishna, and G H Mitchell. A Brief Illustrated Guide to the Ultrastructure of Plasmodium falciparum Asexual Blood Stages. *Parasitology Today*, 16(10):427–433, 2000.
- [30] Contacos PG Coatney GR, Collins WE, Warren M. *The primate malarias*. U.S. Departmen of health, education, and welfare, Atlanta, 1 edition, 2003.
- [31] Scott Prahl. Tabulated molar extinction coefficient for haemoglobin in water, 1998.
- [32] Liliana González-Linares, Víctor Esteban Reyes-Cruz, María Aurora Veloz-Rodríguez, Gustavo Urbano-Reyes, José Luis Imbert-Palafox, and José Angel Cobos-Murcia. New method of production and characterization of haemozoin and B-haemozoin from meccus longipennis. *Iranian Journal of Parasitology*, 14(1):59–67, 2019.
- [33] Amit V. PANDEY, Himani BISHT, Vinod K. BABBARWAL, Jaya SRIVASTAVA, Kailash C. PANDEY, and Virander S. CHAUHAN. Mechanism of malarial haem detoxification inhibition by chloroquine. *Biochemical Journal*, 355(2):333–338, 2001.
- [34] Miranda S. Oakley, Noel Gerald, Vivek Anantharaman, Yamei Gao, Victoria Majam, Babita Mahajan, Phuong Thao Pham, Leda Lotspeich-Cole, Timothy G. Myers, Thomas F. McCutchan, Sheldon L. Morris, L. Aravind, and

- Sanjai Kumar. Radiation-induced cellular and molecular alterations in asexual intraerythrocytic plasmodium falciparum. *Journal of Infectious Diseases*, 207(1):164–174, 2013.
- [35] Frank Wissing, Cecilia P. Sanchez, Petra Rohrbach, Sigrid Ricken, and Michael Lanzer. Illumination of the malaria parasite *Plasmodium falciparum* alters intracellular pH. Implications for live cell imaging. *Journal of Biological Chemistry*, 277(40):37747–37755, 2002.
- [36] Sabrina Yahiya, Sarah Jordan, Holly X. Smith, David C.A. Gaboriau, Mufliat T. Famodimu, Farah A. Dahalan, Alisje Churchyard, George W. Ashdown, and Jake Baum. Live-cell fluorescence imaging of microgametogenesis in the human malaria parasite *Plasmodium falciparum*. *PLoS Pathogens*, 18(2):1–23, 2022.
- [37] Jayasree K Iyer, Lirong Shi, Anuraj H Shankar, and David J Jr Sullivan. Zinc Protoporphyrin IX Binds Heme Crystals to Inhibit the Process of Crystallization in *Plasmodium falciparum*. *Molecular Medicine*, 9(5):175–182, 2003.
- [38] Georg Hennig, Christian Gruber, Michael Vogeser, Herbert Stepp, Stephan Dittmar, Ronald Sroka, and Gary M Brittenham. Dual-wavelength excitation for fluorescence-based quantification of zinc protoporphyrin IX and protoporphyrin IX in whole blood. *Journal of Biophotonics*, 524(7):514–524, 2014.
- [39] Georg Hennig, Christian Homann, Ilknur Teksan, Uwe Hasbargen, Stephan Hasmüller, Lesca M. Holdt, Nadia Khaled, Ronald Sroka, Thomas Stauch, Herbert Stepp, Michael Vogeser, and Gary M. Brittenham. Non-invasive detection of iron deficiency by fluorescence measurement of erythrocyte zinc protoporphyrin in the lip. *Nature Communications*, 7:1–8, 2016.
- [40] Vadivel Masilamani, Sandhanasamy Devanesan, Mani Ravikumar, Kantharaj Perinbam, Mohamad Saleh Alsalhi, Saradh Prasad, Siddanna Palled, and Kadirampatti Mani Ganesh. Fluorescence spectral diagnosis of malaria – a preliminary study. *Diagnostic Pathology*, (1):1–7, 2014.
- [41] Thermo Fisher Scientific Inc. Thermo Scientific Rapid-Chrome Kwik-Diff Staining System. *Thermo Scientific*, (1):1, 2015.
- [42] Noppadon Tangpukdee, Chatnapa Duangdee, Polrat Wilairatana, and Srivicha Krudsood. Malaria diagnosis: A brief review. *Korean Journal of Parasitology*, 47(2):93–102, 2009.
- [43] World Health Organization. Diagnostic testing, 2022.
- [44] Shouki Yatsushiro, Shohei Yamamura, Yuka Yamaguchi, Yasuo Shinohara, Eiichi Tamiya, Toshihiro Horii, Yoshinobu Baba, and Masatoshi Kataoka. Rapid and highly sensitive detection of malaria-infected erythrocytes using a cell microarray chip. *PLoS ONE*, 5(10), 2010.

- [45] Surasak Kasetsirikul, Jirayut Buranapong, Werayut Srituravanich, Morakot Kaewthamasorn, and Alongkorn Pimpin. The development of malaria diagnostic techniques: A review of the approaches with focus on dielectrophoretic and magnetophoretic methods. *Malaria Journal*, 15(1):358, 2016.
- [46] Clement Yuen. Magnetic field enriched surface enhanced resonance Raman spectroscopy for early malaria diagnosis. *Journal of Biomedical Optics*, 17(1):017005, 2012.
- [47] Charles C. Kim, Emily B. Wilson, and Joseph L. Derisi. Improved methods for magnetic purification of malaria parasites and haemozoin. *Malaria Journal*, 9(1):1–5, 2010.
- [48] Chad L. Schaber, Nalin Katta, Lucy B. Bollinger, Mwawi Mwale, Rachel Mlotha-Mitole, Indi Trehan, Barani Raman, and Audrey R. Odom John. Breathprinting Reveals Malaria-Associated Biomarkers and Mosquito Attractants. *Journal of Infectious Diseases*, 217(10):1553–1560, 2018.
- [49] Kristofer E. delas Penas, Pilarita T. Rivera, and Prospero C. Naval. Malaria Parasite Detection and Species Identification on Thin Blood Smears using a Convolutional Neural Network. In *International Conference on Connected Health*, 2017.
- [50] Qazi Ammar, Arshad Mohsen, Ali Saeed-ul Hassan, Chen Chen, Ayisha Imran, and Ghulam Rasul. A dataset and benchmark for malaria life-cycle classification in thin blood smear images. *Neural Computing and Applications*, 6, 2021.
- [51] Vishnu V Makkapati and Raghuvver M Rao. Segmentation of malaria parasites in peripheral blood smear images. *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, (1):1361–1364, 2009.
- [52] Devkumar Das and Madhumala Ghosh. Probabilistic Prediction of Malaria using Morphological and Textural Information. In *2011 International Conference on Image Information Processing (ICIIP 2011)*. IEEE, 2011.
- [53] Aeggarut Pinkaew, Tulaya Limpiti, Akraphon Trirat, Idoflsdupx Vxe, Lpdjvh Ylyd, and V X E Lpdjvh. Automated classification of malaria parasite species on thick blood film using support vector machine. *The 2015 Biomedical Engineering International Conference*, (1):8–12, 2015.
- [54] Mira S Davidson, Clare Andradi-brown, Sabrina Yahiya, Jill Chmielewski, Aidan J O Donnell, Pratima Gurung, Myriam D Jeninga, Parichat Prommana, Dean W Andrew, Michaela Petter, Chairat Uthaipibull, Michelle J Boyle, George W Ashdown, Jeffrey D Dvorin, Sarah E Reece, Danny W Wilson, Kane A Cunningham, D Michael Ando, Michelle Dimon, and Jake Baum. Automated detection and staging of malaria parasites from cytological smears using convolutional neural networks. *Biological Imaging*, 1(2):1–13, 2021.

- [55] Aimon Rahman, Hasib Zunair, M Sohel Rahman, and Jesia Quader Yuki. Improving Malaria Parasite Detection from Red Blood Cell using Deep Convolutional Neural Networks. *arXiv*, 1907.10418:1–33, 2019.
- [56] Krit Sriporn, Cheng-Fa Tsai, Chia-En Tsai, and Paohsi Wang. Analyzing Malaria Disease Using Effective Deep Learning Approach. *Diagnostics*, 10(10):1–22, 2020.
- [57] David Opeoluwa Oyewola, Emmanuel Gbenga Dada, Sanjay Misra, and Robertas Damaševičius. A Novel Data Augmentation Convolutional Neural Network for Detecting Malaria Parasite in Blood Smear Images. *Applied Artificial Intelligence*, 2022:1–22, 2022.
- [58] Enid M. Rivadeneira, Moises Wasserman, and Carlos T. Espinal. Separation and Concentration of Schizonts of Plasmodium falciparum by Percoll Gradients. *The Journal of Protozoology*, 30(2):367–370, 1983.
- [59] Håkan Pertoft. Fractionation of cells and subcellular particles with Percoll. *Journal of Biochemical and Biophysical Methods*, 44(1-2):1–30, 2000.
- [60] Chris Lambros and Jerome P Vanderberg. Synchronization of Plasmodium falciparum Erythrocytic Stages in Culture. *The Journal of Parasitology*, 65(3):418–420, 1979.
- [61] Connor Shorten and Taghi M. Khoshgoftaar. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1), 2019.
- [62] Jason Brownlee. Introduction to Dimensionality Reduction for Machine Learning, 2020.
- [63] S. Kareem, I. Kale, and R. C.S. Morling. Automated malaria parasite detection in thin blood films:-A hybrid illumination and color constancy insensitive, morphological approach. *IEEE Asia-Pacific Conference on Circuits and Systems, Proceedings, APCCAS*, 2012:240–243, 2012.
- [64] W Trager and J.B. Jensen. Human malaria parasites in continuous culture. *Science*, 193(4254):673–675, 1976.
- [65] G. J. Christofinis and H. Miller. A simplified method for cryopreservation of Plasmodium falciparum from continuous in vitro cultures. *Annals of Tropical Medicine and Parasitology*, 77(2):123–126, 1983.
- [66] Frederick L. Schuster. Cultivation of plasmodium spp. *Clinical Microbiology Reviews*, 15(3):355–364, 2002.
- [67] Kirsten Moll, Akira Kaneko, Arthur Scherf, and Mats Wahlgren. *Methods in malaria research*, volume 11. EVIMalaR, 6th edition, 2013.

- [68] R. J.M. Wilson, J. Farrant, and C. A. Walter. Preservation of intraerythrocytic forms of malarial parasites by one-step and two-step cooling procedures. *Bulletin of the World Health Organization*, 55(2-3):309–315, 1977.
- [69] Om Prakash Choudhary, Rupan Sarkar, Priyanka, G. E. Chethan, Probal Jyoti Doley, Pranab Chandra Kalita, and Arup Kalita. Preparation of blood samples for electron microscopy: The standard protocol. *Annals of Medicine and Surgery*, 70(September):102895, 2021.
- [70] Anatoli Ianoul, Melissa Street, Donna Grant, John Pezacki, Rod S. Taylor, and Linda J. Johnston. Near-field scanning fluorescence microscopy study of ion channel clusters in cardiac myocyte membranes. *Biophysical Journal*, 87(5):3525–3535, 2004.
- [71] Christiane Höppener. Scanning near-field optical microscopy for investigations of bio-matter. In *Super-resolution microscopy techniques in the neurosciences*, volume 86, pages 189–223. Humana Totowa, NJ, 2014.
- [72] Inga Morkvenaite-Vilkončiene, Almira Ramanavičiene, and Arunas Ramanavičius. Atomic force microscopy as a tool for the investigation of living cells. *Medicina (Lithuania)*, 49(4):155–164, 2013.
- [73] Franz J. Giessibl, S. Hembacher, H. Bielefeldt, and J. Mannhart. Subatomic features on the silicon (111)-(7x7) surface observed by atomic force microscopy. *Science*, 289(5478):422–425, 2000.
- [74] Nader Jalili and Karthik Laxminarayana. A review of atomic force microscopy imaging systems: Application to molecular metrology and biological sciences. *Mechatronics*, 14(8):907–945, 2004.
- [75] Ricardo García and Rubén Pérez. Dynamic atomic force microscopy methods. *Surface Science Reports*, 47(6):197–301, 2002.
- [76] Gerhard Meyer and Nabil M. Amer. Novel optical approach to atomic force microscopy. *Applied Physics Letters*, 53(12):1045–1047, 1988.
- [77] Peggy Schönherr. *Topological Structures in Magnetic and Electric Materials*. PhD thesis, ETH Zürich, 2018.
- [78] Yongho Seo and Wonho Jhe. Atomic force microscopy and spectroscopy. *Reports on Progress in Physics*, 71(1), 2008.
- [79] Sascha Sadewasser and Thilo Glatzel. Experimental technique and working. *Kelvin probe force microscopy*, (48):7–24, 2012.
- [80] E. Meyer. Atomic force microscopy. *Progress in surface science*, 41(1):3–49, 1992.

- [81] Andre Schirmeisen, Boris Anczykowski, and Harald Fuchs. Dynamic modes of atomic force microscopy. *Nanotribology and Nanomechanics (Second Edition): An Introduction*, (1):235–277, 2008.
- [82] Yaxin Song and Bharat Bhushan. Modeling of Tip-Cantilever Dynamics in Atomic Force Microscopy. *Applied scanning probe methods V: scanning probe microscopy techniques*, (1):149–223, 2007.
- [83] Matthew W. Fairbairn, Philipp Müller, and S. O.Reza Moheimani. Active piezoelectric shunt control of an Atomic Force Microscope micro-cantilever. *2013 3rd Australian Control Conference, AUCC 2013*, (1):257–262, 2013.
- [84] T. R. Albrecht, P. Grütter, D. Horne, and D. Rugar. Frequency modulation detection using high-Q cantilevers for enhanced force microscope sensitivity. *Journal of Applied Physics*, 69(2):668–673, 1991.
- [85] Roland Guerre, Ute Drechsler, Daniel Jubin, and Michel Despont. Low-cost AFM cantilever manufacturing technology. *Journal of Micromechanics and Microengineering*, 18(11), 2008.
- [86] Ted pella inc. Silicon AFM Tapping Mode Probes, 2022.
- [87] Bruker. OTESPA-R3, 2022.
- [88] Núria Gavara. A beginner’s guide to atomic force microscopy probing for cell mechanics. *Microscopy Research and Technique*, 80(1):75–84, 2017.
- [89] Andrew A. Vekinis and Vassilios Constantoudis. Quantifying geometric tip-sample effects in AFM measurements using certainty graphs. *Micro and Nano Engineering*, 8(1):100067, 2020.
- [90] Jeff W. Lichtman and José Angel Conchello. Fluorescence microscopy. *Nature Methods*, 2(12):910–919, 2005.
- [91] Bernard Valeur and Nuno Mário Berberan-Santos. *Molecular Fluorescence: Principles and Applications*. John Wiley & Sons, 2012.
- [92] Hans H. Jaffé and Albert L. Miller. The fates of electronic excitation energy. *Journal of Chemical Education*, 43(9):469–473, 1966.
- [93] Joseph R. Lakowicz. *Fluorescence spectroscopy*. Plenum Publishing Corporation, 1983.
- [94] Andreas Stemmer, Markus Beck, and Reto Fiolka. Widefield fluorescence microscopy with extended resolution. *Histochemistry and Cell Biology*, 130(5):807–817, 2008.
- [95] Douglas B Murphy and Michael W. Davidson. *Fundamentals of light microscopy and electronic imaging*. John Wiley & Sons, 2012.
- [96] Olympus. Olympus IX71 inverted microscopes. *Olympus*, page 30, 2015.

-
- [97] J. B. Pawley. Fundamental and practical limits in confocal light microscopy. *Scanning*, 13(2):184–198, 1991.
- [98] Deval Shah. The essential guide to data augmentation in deep learning, 2022.
- [99] Rukshan Pramoditha. The concept of artificial neurons (perceptrons) in neural networks, 2021.
- [100] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of mathematical biophysics*, 5:115–133, 1943.
- [101] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.
- [102] Roland Schwaiger and Joachim Steinwender. *Neuronale Netwerke programmieren mit Python*. Rheinwerk Verlag, Bonn, 1. edition, 2019.
- [103] Marvin Minsky and Seymour Papert. *Perceptrons; an Introduction to Computational Geometry*. MIT Press, 1969.
- [104] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.
- [105] Jorge Leonel. Backpropagation, 2018.
- [106] Bernard Widrow. Adaptive "adaline" neuron using chemical "memistors". *Stanford electron Labs*, 2(1553), 1960.
- [107] Sagar Sharma. Activation functions in neural networks.
- [108] Pragati Baheti. Activation functions in neural networks, 2022.
- [109] Keras API reference. Layer weight regularizers, 2022.
- [110] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1):215–243, 1968.
- [111] Sumit Saha. A comprehensive guide to convolutional neural networks - the ELI5 way, 2018.
- [112] Zikun Liu, Hongzhen Wang, Lubin Weng, and Yiping Yang. Understanding the difficulty of training deep feedforward neural networks. *IEEE Geoscience and Remote Sensing Letters*, 13(8):1074–1078, 2016.
- [113] Ron Kohavi and Foster Provost. Glossary of terms, editorial for the special issue on applications of machine learning and the knowledge discovery process. *Machine learning*, (1):271–274, 1998.
- [114] Vaibhav Jayaswal. Performance Metrics: Confusion matrix, Precision, Recall, and F1 Score, 2020.

- [115] Hynek Prasanna, S. R. Mahadeva ; Yegnanarayana, B. ; Pinto, Joel Praveen ; Hermansky. Analysis of Confusion Matrix to Combine Evidence for Phoneme Recognition. Technical report, IDIAP Research institute Switzerland, 2007.
- [116] Kirsten Moll, Inger Ljungström, Hedvig Perlmann, Artur Scherf, and Mats Wahlgren. *Methods in malaria research*. Malaria Research and Reference Reagent Resource Center (MR4), 2008.
- [117] M. Isabel Veiga, Satish K. Dhingra, Philipp P. Henrich, Judith Straimer, Nina Gnädig, Anne Catrin Uhlemann, Rowena E. Martin, Adele M. Lehane, and David A. Fidock. Globally prevalent PfMDR1 mutations modulate Plasmodium falciparum susceptibility to artemisinin-based combination therapies. *Nature Communications*, 7, 2016.
- [118] Yong Chen and Jiye Cai. Membrane deformation of unfixed erythrocytes in air with time lapse investigated by tapping mode atomic force microscopy. *Micron*, 37(4):339–346, 2006.
- [119] Robert Nowakowski, Paul Luckham, and Peter Winlove. Imaging erythrocytes under physiological conditions by atomic force microscopy. *Biochimica et Biophysica Acta - Biomembranes*, 1514(2):170–176, 2001.
- [120] Oxford Instrument. Bio Heater Closed Fluid Cell Accessory for the MFP-3D™ AFM. *Oxford Instruments*, 4(6):3–4, 2014.
- [121] Monica Diez-Silva, Ming Dao, Jongyoon Han, Chwee-Teck Lim, and Subra Suresh. Shape and Biomechanical Characteristics of Human Red Blood Cells in Health and Disease. *MRS bulletin / Materials Research Society*, 35(5):382–388, 2010.
- [122] Elena Kozlova, Aleksandr Chernysh, Ekaterina Manchenko, Viktoria Sergunova, and Viktor Moroz. Nonlinear biomechanical characteristics of deep deformation of native RBC membranes in normal state and under modifier action. *Scanning*, 2018, 2018.
- [123] H.-J. Butt, E.K. Wolff, S.A.C. Gould, B. Dixon Northern, C.M. Peterson, and P.K. Hansma. Imaging cells with the atomic force microscope. *J. Struct. Biol.*, 105:54–61, 1990.
- [124] Olympus. BioLever, 2018.
- [125] Bruker. AC40, 2022.
- [126] Nanoworld. PNP-TR, 2022.
- [127] Bruker. MSCT, 2022.
- [128] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-9(1):62–66, 1979.
- [129] Frederik Lundh and Alex Clark. Pillow (PIL Fork) Reference, 2022.

- [130] Tutorials OpenCV. Hough Circle Transform, 2021.
- [131] John Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986.
- [132] N. H.M. Daud, R. A.A. Raof, M. K. Osman, and N. H. Harun. Segmentation Technique for Nucleus Detection in Blood Images for Chronic Leukaemia. *Journal of Physics: Conference Series*, 1755(1), 2021.
- [133] Gulpreet Kaur Chadha, Aakarsh Srivastava, Abhilasha Singh, Ritu Gupta, and Deepanshi Singla. An Automated Method for Counting Red Blood Cells using Image Processing. *Procedia Computer Science*, 167(2019):769–778, 2020.
- [134] Engin Avci, Resul çötelı, Turker Tuncer, and Esin Dogantekin. A New Blood Cell Recognition Algorithm Based On Directed Vector Method. *Journal of Applied Biological Sciences*, 9(3):18–20, 2015.
- [135] Jane Hung, Allen Goodman, Deepali Ravel, Stefanie C.P. Lopes, Gabriel W. Rangel, Odailton A. Nery, Benoit Malleret, Francois Nosten, Marcus V.G. Lacerda, Marcelo U. Ferreira, Laurent Rénia, Manoj T. Duraisingh, Fabio T.M. Costa, Matthias Marti, and Anne E. Carpenter. Keras R-CNN: Library for cell detection in biological images using deep neural networks. *BMC Bioinformatics*, 21(1):1–7, 2020.
- [136] Syed Saiden Abbas and Tjeerd M H Dijkstra. Detection and stage classification of Plasmodium falciparum from images of Giemsa stained thin blood films using random forest classifiers. *Diagnostic Pathology*, 15(130):1–11, 2020.
- [137] Dan Wang, Bin Xue, Xianggui Kong, Langping Tu, Xiaomin Liu, Youlin Zhang, Yulei Chang, Yongshi Luo, Huiying Zhao, and Hong Zhang. 808 nm driven Nd³⁺-sensitized upconversion nanostructures for photodynamic therapy and simultaneous fluorescence imaging. *Nanoscale*, 7(1):190–197, 2015.
- [138] Marta Orłowska, Tatiana Koutchma, Michael Grapperhaus, John Gallagher, Raymond Schaefer, and Cheryl Defelice. Continuous and Pulsed Ultraviolet Light for Nonthermal Treatment of Liquid Foods. Part 1: Effects on Quality of Fructose Solution, Apple Juice, and Milk. *Food and Bioprocess Technology*, 6(6):1580–1592, 2013.
- [139] Narsingh R. Nirala, Vinita, and Rajiv Prakash. Quick colorimetric determination of choline in milk and serum based on the use of MoS₂ nanosheets as a highly active enzyme mimetic. *Microchimica Acta*, 185(4):1–8, 2018.
- [140] Yi Wen Zheng, Shao Fang Chen, Bi Yuan Zheng, Mei Rong Ke, and Jian Dong Huang. A silicon(IV) phthalocyanine-folate conjugate as an efficient photosensitizer. *Chemistry Letters*, 43(11):1701–1703, 2014.
- [141] Travis Oliphant and Jarrod k. Millma. A guide to NumPy. *Trelgol Publishing*, 2006.

- [142] Wes McKinney. Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, 1:56–61, 2010.
- [143] OpenCV Team. OpenCV, 2022.
- [144] Scikit image development team. Scikit-image: image processing in python, 2022.
- [145] Matplotlib development team. Matplotlib 3.5.2 documentation, 2022.
- [146] NVIDIA. TensorFlow, 2022.
- [147] Python Software Foundation. Graphical user interfaces with Tk, 2022.
- [148] San Jose State University. T-table, 2007.
- [149] William Sealy Gosset. The probable error of a mean. *Biometrika*, 6(1):1–25, 1908.

A

Experimental data

A.1 T-test

T-tests are employed to show statistical significance between the mean values of two samples (two-sample test) or the deviation of a mean value of a sample from a defined value (one-sample test). Two-sample tests can either be dependent or independent. If the same sample is compared in two different states, such as the mean height profiles of RBCs before and after removing background noise, the sample are dependent. In the other case, two different samples, such as average height profiles of stained and unstained cells are compared. The test value for the statistical significance is called t . Depending on the test type, t is defined as

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \quad (\text{A.1})$$

for the one-sample test, as

$$t = \frac{d}{s_d \cdot \frac{1}{\sqrt{n}}} \quad (\text{A.2})$$

for the dependent two-sample test, and as

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (\text{A.3})$$

for the independent two-sample test. The mean value of the single sample is expressed by \bar{x} , the difference between the mean values of the dependent samples by d and the mean values of the independent samples by \bar{x}_1 and \bar{x}_2 . s determines the standard deviation of the sample, s_1 and s_2 the standard deviation of independent samples, and s_d the standard deviation of the difference between the dependent samples. To determine the statistical significance, the t value calculated for the samples is compared to the t-value distribution [148] at the degree of freedom corresponding to the samples and the significance level chosen for the test. For a dependent test, the degree of freedom is calculated by $df = n - 1$ and by

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}{n_1 - 1} + \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}{n_2 - 1}} \quad (\text{A.4})$$

for an independent test. If the determined value for t is larger than the t-value distribution, there is an intrinsic difference between the groups. In the other case, no statistical significance exists between the samples [149].

In this thesis, the t-test was employed to compare the mean height h_{mean} of the cell and the height drop h_{diff} from the maximum at the cell edge to the minimum near the centre of the average height profiles of healthy RBCs and the intra-erythrocytic stages. Tables A.1 - A.12 show the comparison between the stages healthy (h), ring (r), trophozoite (t), and schizont (s), stained and unstained samples, and images with and without background noise for each microscopy technique.

Imaging method	AFM			
Group	stained - unstained			
Stage	h	r	t	s
Significant? $h_{\text{mean}}/h_{\text{diff}}$	yes/yes	yes/yes	yes/yes	yes/yes

Table A.1: Results of the t-test with a significance level of 95% for the stage-wise comparison of h_{mean} and h_{diff} for stained and unstained samples.

Imaging method	AFM			
Group	background - no background			
Stage	h	r	t	s
Significant? $h_{\text{mean}}/h_{\text{diff}}$	yes/yes	no/yes	yes/yes	no/yes

Table A.2: Results of the t-test with a significance level of 95% for the stage-wise comparison of h_{mean} and h_{diff} calculated with and without background.

Imaging method	AFM					
Group	stained					
Stage	h-r	h-t	h-s	r-t	r-s	t-s
Significant? $h_{\text{mean}}/h_{\text{diff}}$	yes/yes	yes/yes	yes/yes	yes/yes	yes/yes	yes/yes

Table A.3: Results of the t-test with a significance level of 95% for the comparison of h_{mean} and h_{diff} between the stages in case of stained samples.

Imaging method	AFM					
Group	unstained					
Stage	h-r	h-t	h-s	r-t	r-s	t-s
Significant? $h_{\text{mean}}/h_{\text{diff}}$	yes/yes	yes/yes	yes/yes	yes/yes	yes/yes	yes/yes

Table A.4: Results of the t-test with a significance level of 95% for the comparison of h_{mean} and h_{diff} between the stages in case of unstained samples.

Imaging method	AFM					
Group	with background					
Stage	h-r	h-t	h-s	r-t	r-s	t-s
Significant? $h_{\text{mean}}/h_{\text{diff}}$	yes/yes	yes/yes	yes/no	yes/yes	yes/yes	yes/yes

Table A.5: Results of the t-test with a significance level of 95% for the comparison of h_{mean} and h_{diff} between the stages in case of samples with background.

Imaging method	AFM					
Group	without background					
Stage	h-r	h-t	h-s	r-t	r-s	t-s
Significant? $h_{\text{mean}}/h_{\text{diff}}$	yes/yes	yes/yes	yes/yes	yes/yes	yes/yes	yes/yes

Table A.6: Results of the t-test with a significance level of 95% for the comparison of h_{mean} and h_{diff} between the stages in case of samples without background.

Imaging method	Fluorescence microscopy			
Group	background - no background			
Stage	h	r	t	s
Significant? $h_{\text{mean}}/h_{\text{diff}}$	yes/yes	no/yes	no/yes	no/yes

Table A.7: Results of the t-test with a significance level of 95% for the stage-wise comparison of h_{mean} and h_{diff} calculated with and without background.

Imaging method	Fluorescence microscopy					
Group	with background					
Stage	h-r	h-t	h-s	r-t	r-s	t-s
Significant? $h_{\text{mean}}/h_{\text{diff}}$	yes/yes	yes/yes	yes/yes	yes/yes	yes/yes	yes/yes

Table A.8: Results of the t-test with a significance level of 95% for the comparison of h_{mean} and h_{diff} between the stages in case of samples with background.

Imaging method	Fluorescence microscopy					
Group	without background					
Stage	h-r	h-t	h-s	r-t	r-s	t-s
Significant? $h_{\text{mean}}/h_{\text{diff}}$	yes/yes	yes/yes	yes/yes	yes/yes	yes/yes	yes/yes

Table A.9: Results of the t-test with a significance level of 95% for the comparison of h_{mean} and h_{diff} between the stages in case of samples without background.

Imaging method	Light microscopy			
Group	background - no background			
Stage	h	r	t	s
Significant? $h_{\text{mean}}/h_{\text{diff}}$	no/yes	no/yes	yes/yes	yes/yes

Table A.10: Results of the t-test with a significance level of 95% for the stage-wise comparison of h_{mean} and h_{diff} calculated with and without background.

Imaging method	Light microscopy					
Group	with background					
Stage	h-r	h-t	h-s	r-t	r-s	t-s
Significant? $h_{\text{mean}}/h_{\text{diff}}$	yes/yes	yes/yes	yes/yes	yes/yes	yes/yes	yes/yes

Table A.11: Results of the t-test with a significance level of 95% for the comparison of h_{mean} and h_{diff} between the stages in case of samples with background.

Imaging method	Light microscopy					
Group	without background					
Stage	h-r	h-t	h-s	r-t	r-s	t-s
Significant? $h_{\text{mean}}/h_{\text{diff}}$	yes/yes	yes/yes	yes/yes	yes/yes	yes/yes	yes/yes

Table A.12: Results of the t-test with a significance level of 95% for the comparison of h_{mean} and h_{diff} between the stages in case of samples without background.

A.2 Network architecture

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 48, 48, 32)	320
max_pooling2d (MaxPooling2D)	(None, 24, 24, 32)	0
conv2d_1 (Conv2D)	(None, 22, 22, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 11, 11, 64)	0
conv2d_2 (Conv2D)	(None, 9, 9, 64)	36928
max_pooling2d_2 (MaxPooling2D)	(None, 4, 4, 64)	0
conv2d_3 (Conv2D)	(None, 2, 2, 64)	36928
max_pooling2d_3 (MaxPooling2D)	(None, 1, 1, 64)	0
flatten (Flatten)	(None, 64)	0
dense (Dense)	(None, 50)	3250
dense_1 (Dense)	(None, 4)	204
Total params: 96,126		
Trainable params: 96,126		
Non-trainable params: 0		

Figure A.1: Neural network architecture for the stage-specific classification of RBCs of two-dimensional input images cuts.

Model: "sequential"

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 48, 32)	128
max_pooling1d (MaxPooling1D)	(None, 24, 32)	0
conv1d_1 (Conv1D)	(None, 22, 64)	6208
max_pooling1d_1 (MaxPooling1D)	(None, 11, 64)	0
conv1d_2 (Conv1D)	(None, 9, 64)	12352
max_pooling1d_2 (MaxPooling1D)	(None, 4, 64)	0
conv1d_3 (Conv1D)	(None, 2, 64)	12352
max_pooling1d_3 (MaxPooling1D)	(None, 1, 64)	0
flatten (Flatten)	(None, 64)	0
dense (Dense)	(None, 50)	3250
dense_1 (Dense)	(None, 4)	204
=====		
Total params: 34,494		
Trainable params: 34,494		
Non-trainable params: 0		

Figure A.2: Neural network architecture for the stage-specific classification of RBCs of one-dimensional cuts.

A.3 Performance parameters of the stage-specific detection of malaria

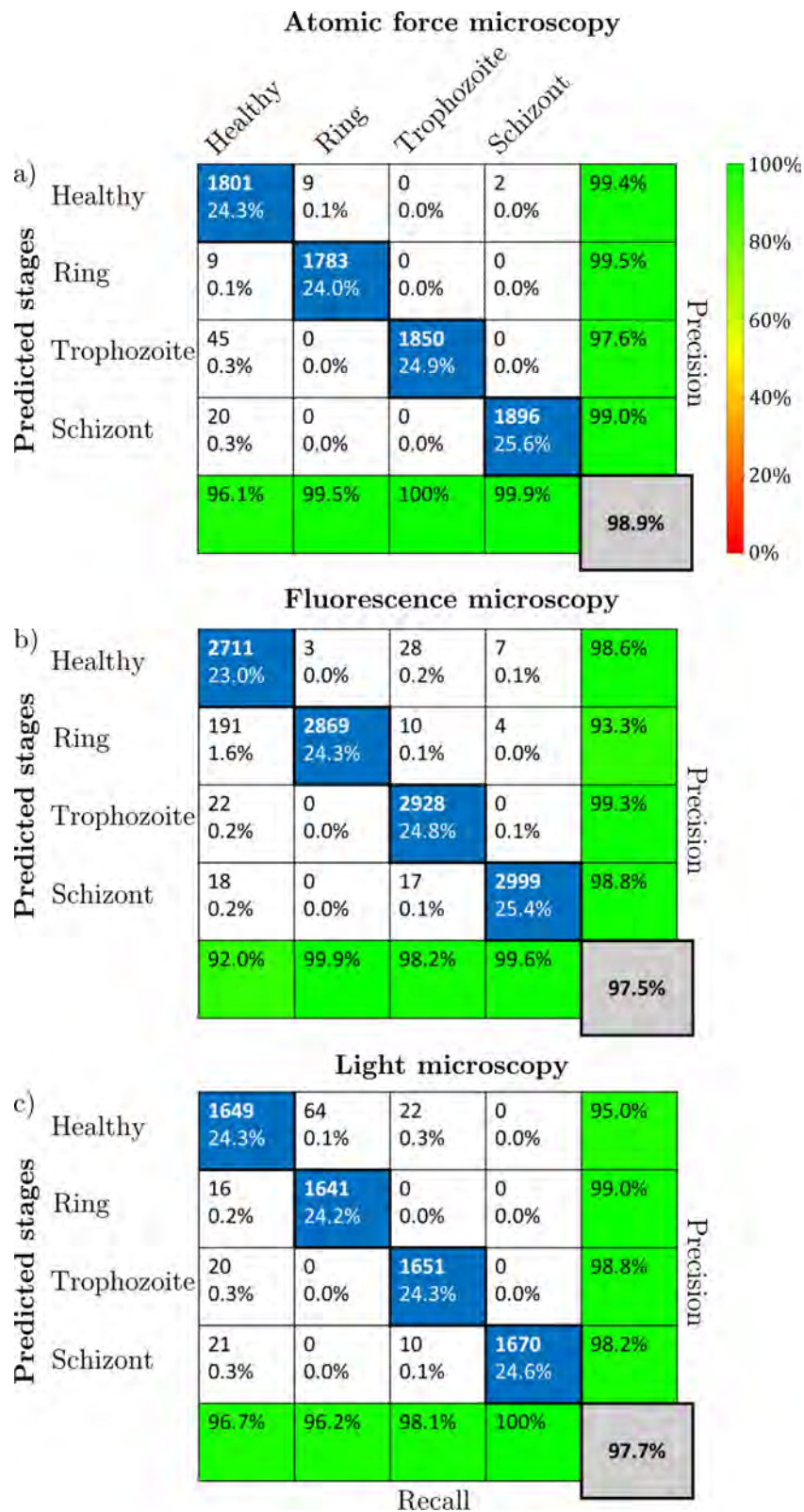


Figure A.3: Classification performance of the NN on four cuts.

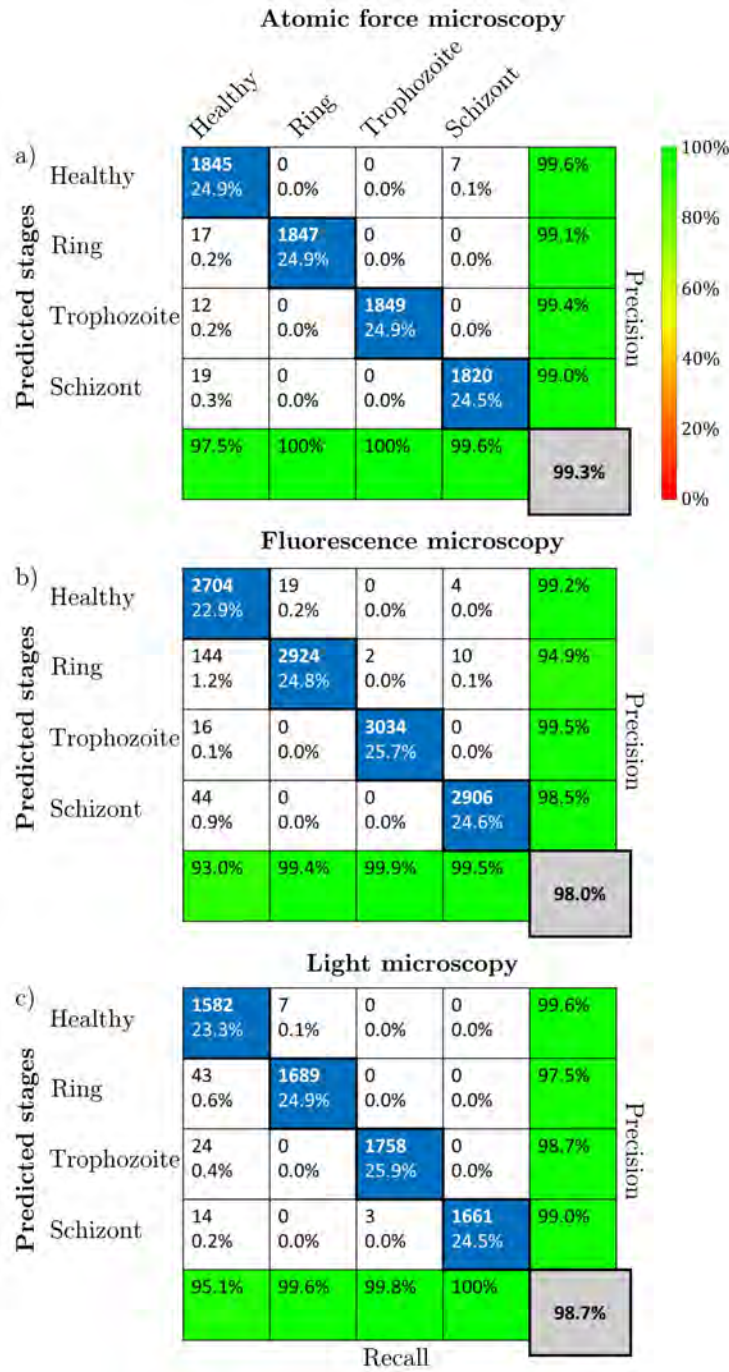


Figure A.4: Classification performance of the NN on the parasite plus cuts.

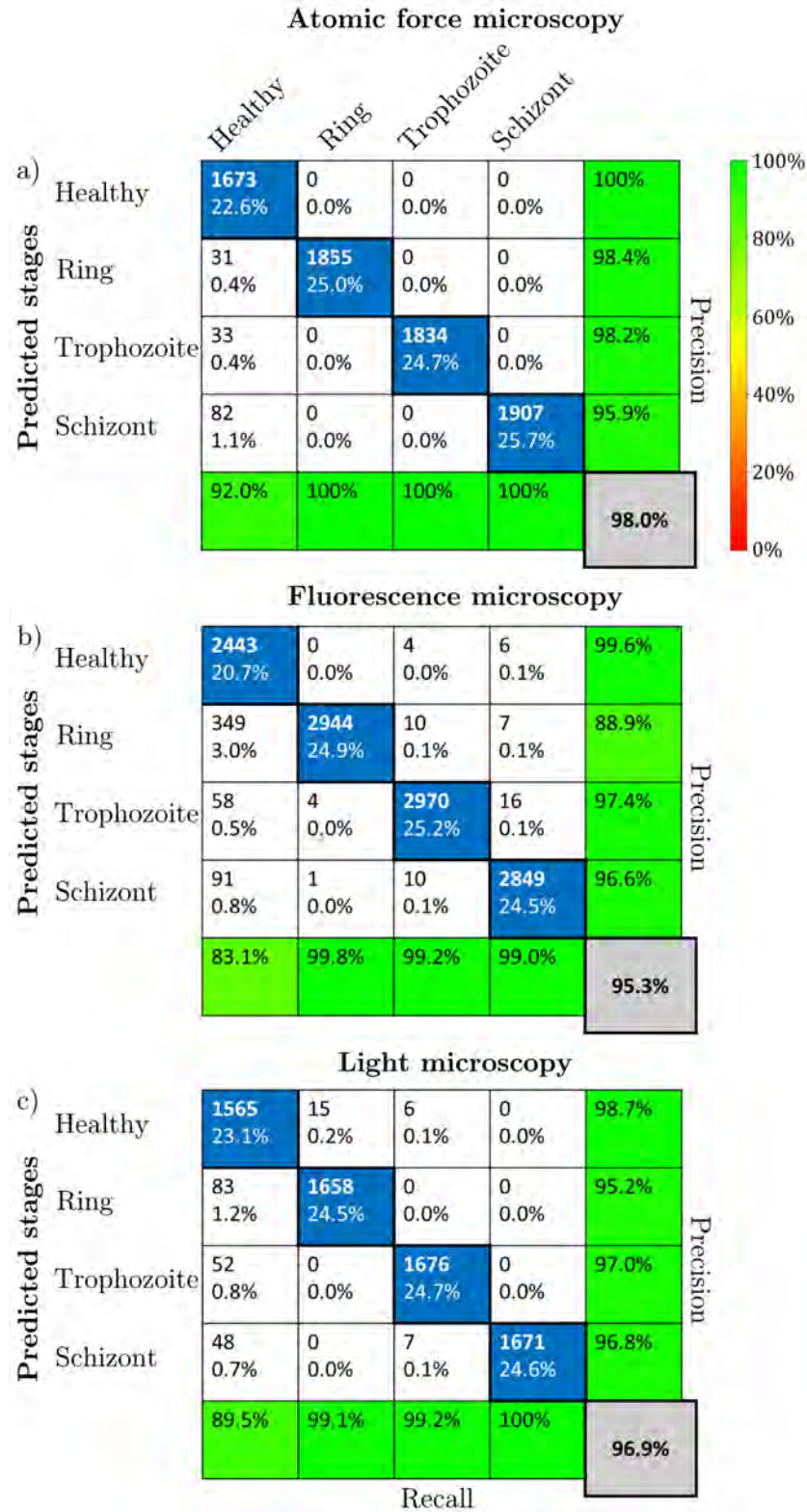


Figure A.5: Classification performance of the NN on the parasite cut.

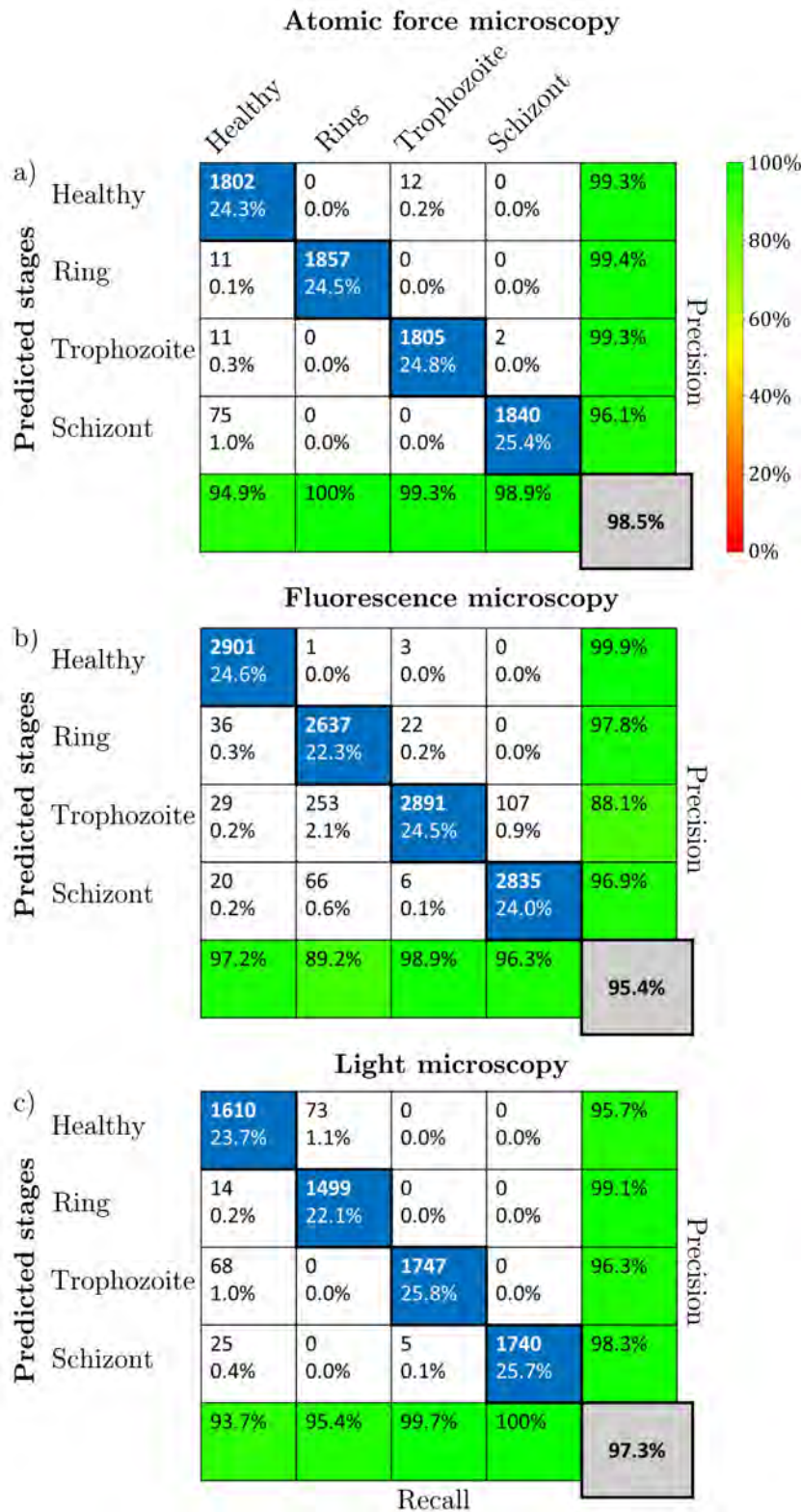


Figure A.6: Classification performance of the NN on the four cuts. The orientation of the four cuts was calculated without removing background noise.

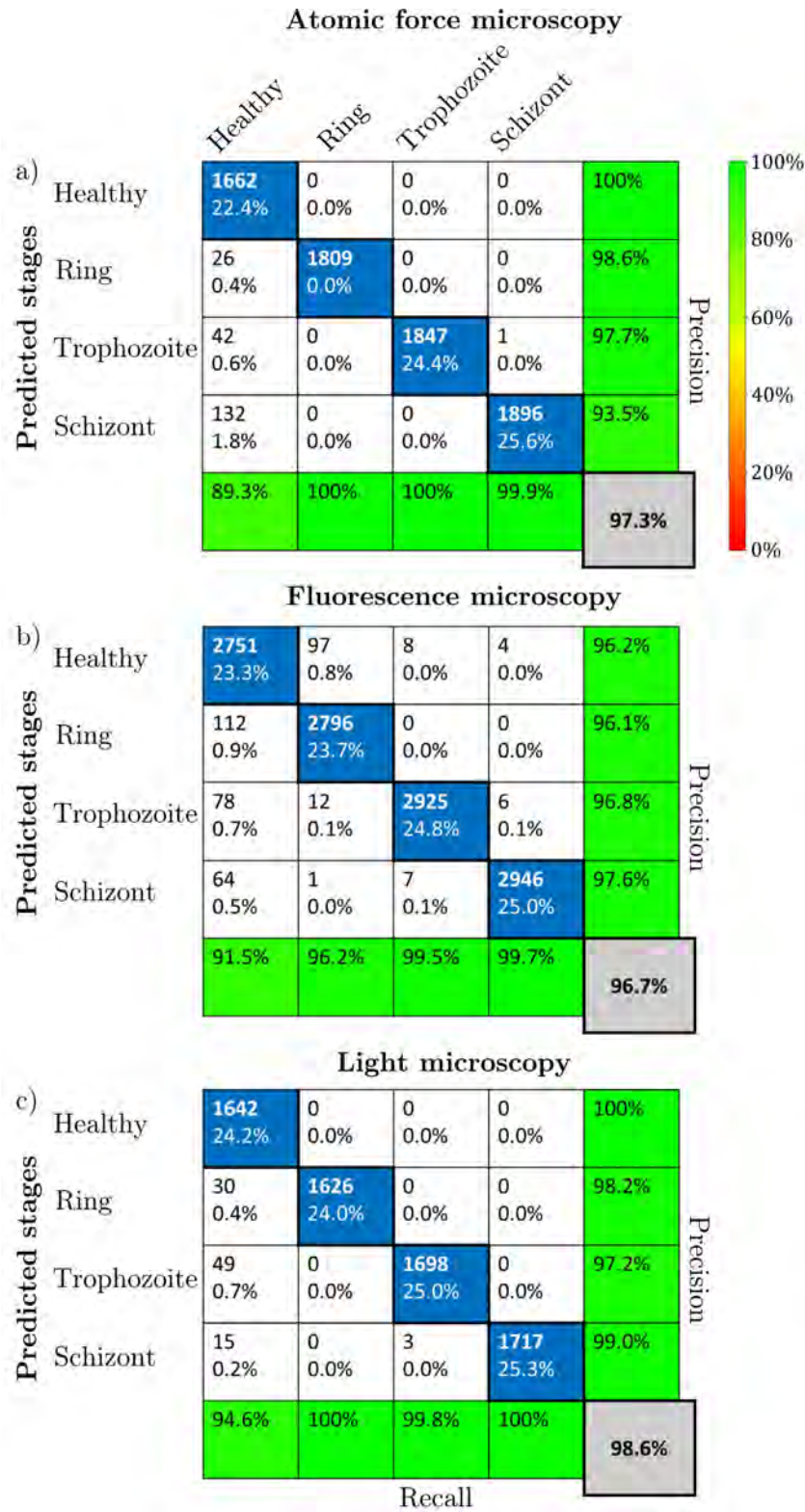


Figure A.7: Classification performance of the NN on the parasite plus cuts. The orientation of the parasite plus cuts was calculated without removing background noise.

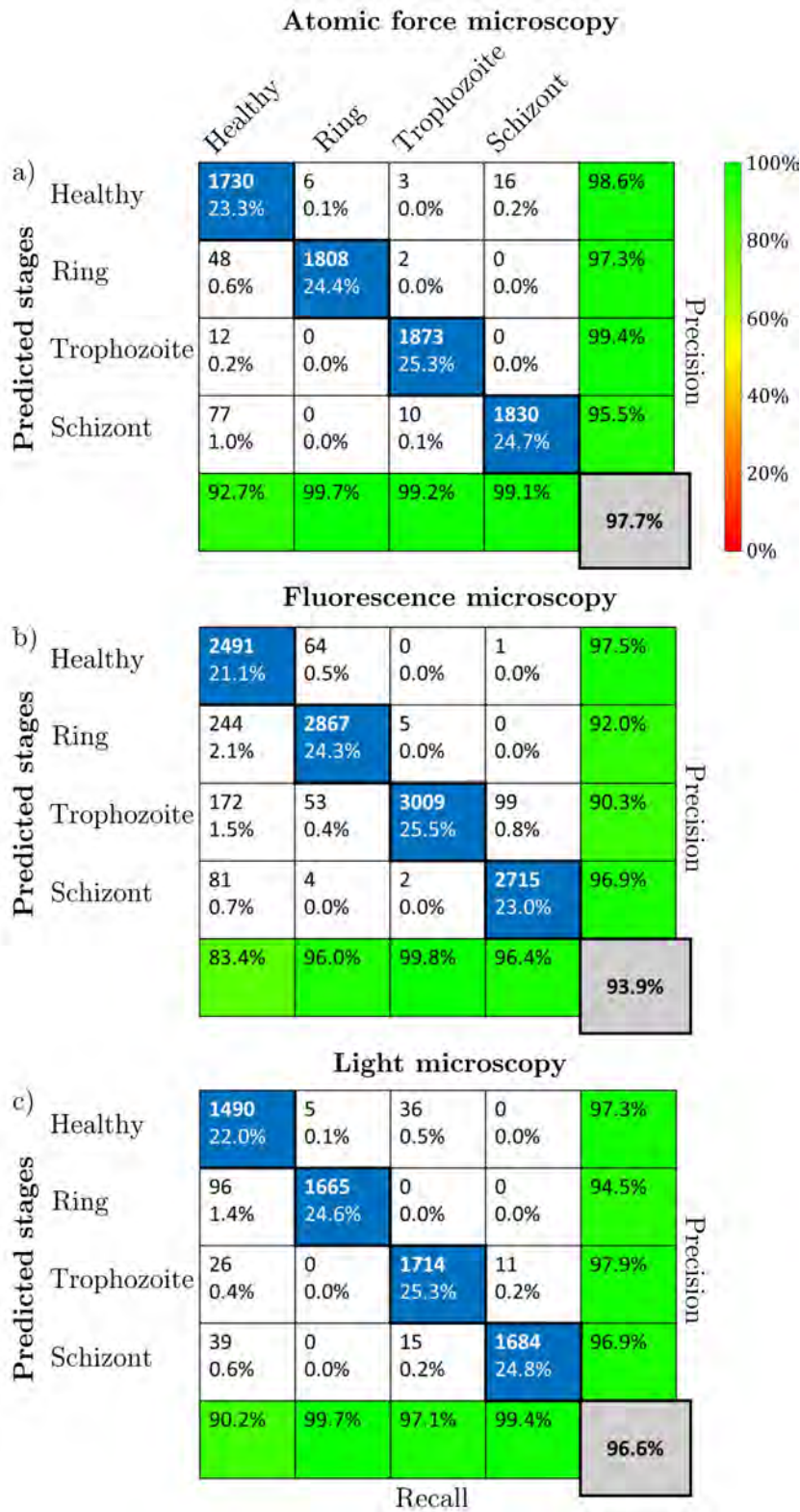


Figure A.8: **Classification performance of the NN on the parasite cut.** The orientation of the parasite cut was calculated without removing background noise.

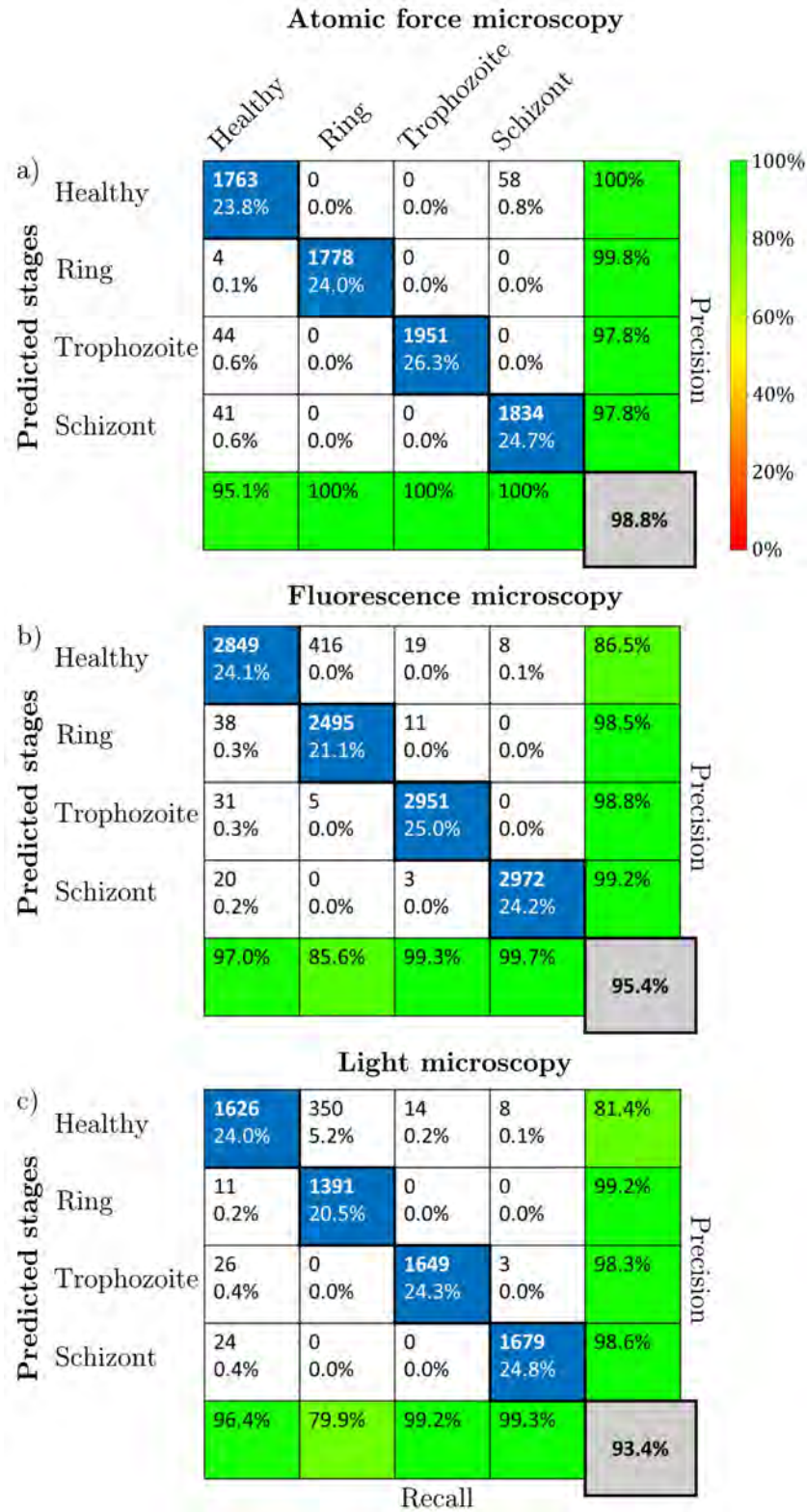


Figure A.9: Classification performance of the NN on four randomly oriented cuts.

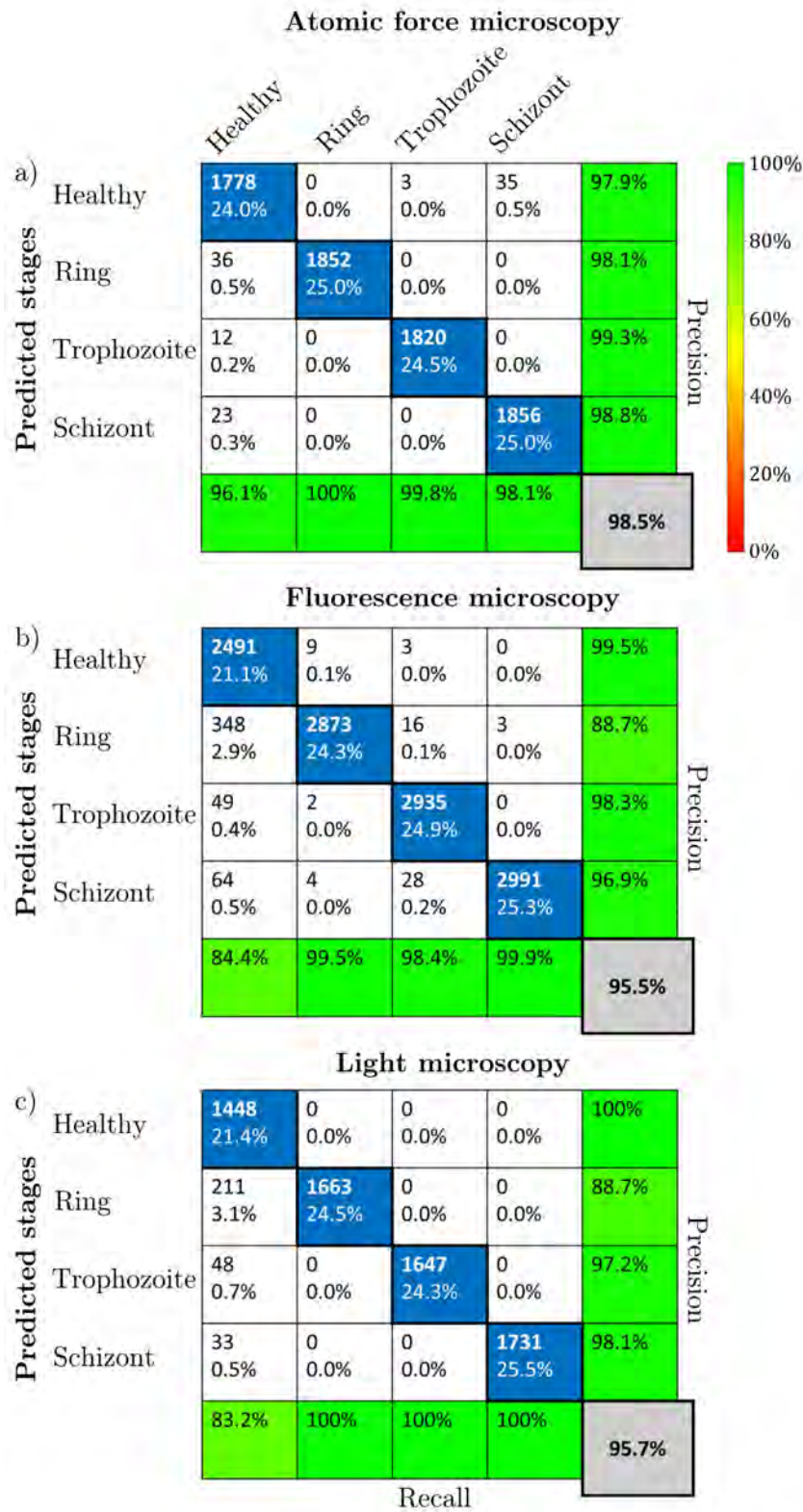


Figure A.10: Classification performance of the NN on two randomly oriented cuts.

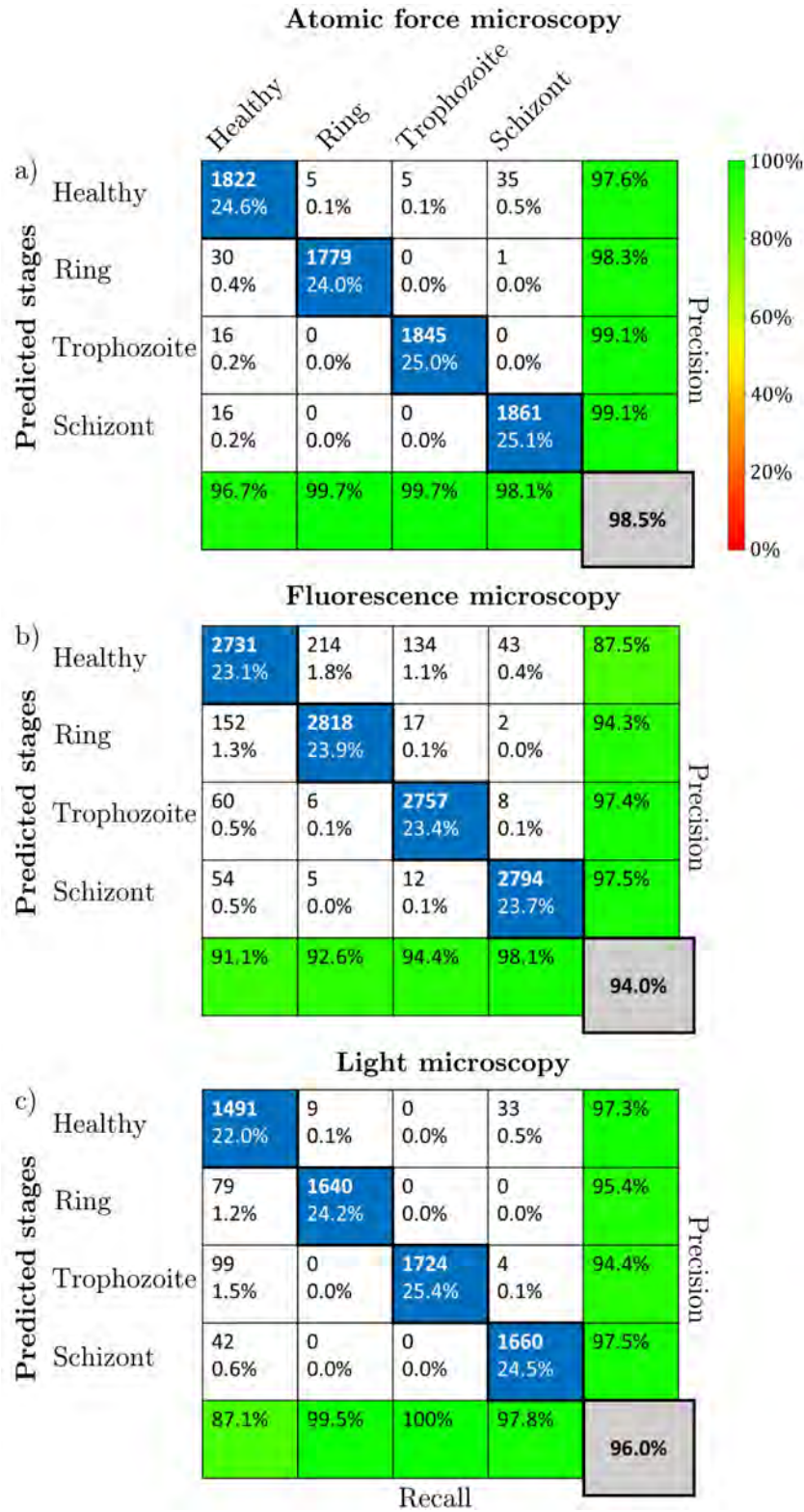


Figure A.11: Classification performance of the NN on one randomly oriented cut.

B

Documentation of the malaria stage classifier

Stage-specific-classification-of-RBCs

Release 0.0.1

Katharina Preißinger

Jul 02, 2022

CONTENTS:

- 1 Malaria stage classifier** **1**
- 2 Download** **3**
- 3 How to use the package** **5**
- 4 Module functions** **7**
 - 4.1 Module contours 7
 - 4.2 Module extractCuts 10
 - 4.3 Module NN 14
- 5 Module classes** **17**
- Index** **25**

MALARIA STAGE CLASSIFIER

This is the documentation of a Python package to classify RBCs in microscopy images. It includes:

- **a package for the stage-specific classification of RBCs (*ClassificationRBC*) with four modules:**
 - *NN.py* which initialises the neural network and trains the data
 - *classes.py* which contains classes for evaluating the properties of each RBC
 - *contours.py* which provides functions for the detection of RBCs in an image
 - *extractCuts.py* which provides functions for extracting the most characteristic profiles in the RBC

DOWNLOAD

1. Download *Code* folder which contains the code, the logo for the pop up windows, three pre-trained neural networks, sample images, and a folder for saving the evaluated data from: https://github.com/KatharinaPreissinger/Stage_specific_classification_of_RBCs
2. If you want to retrain the neural network, please download the *Datasets_for_NN* (this requires at least 200 MB of free space)
3. Install jupyter notebook on your computer <https://jupyter.org/install>
4. Read the documentation for more information about the classes and modules

HOW TO USE THE PACKAGE

1. Open the code file *Classification_of_RBCs* in jupyter
2. **Tab Settings:**
 - select the file you want to analyse
 - choose the type, in case of header lines in the text file choose the number of header lines
 - select the folder to save your output
3. **Tab Show image:**
 - displays the image
4. **Tab Threshold image:**
 - use the *Set threshold* button to show the thresholded image
 - optionally, the preset threshold value can be changed manually
5. **Tab Detect cells:**
 - the cell detection parameters can be changed manually and set back to default
6. **Tab Classify cells:**
 - *Load NN* loads the neural network
 - *Predict stages* predicts the cell stage
 - *Change predictions* provides the possibility to change the prediction by clicking on the cell
 - *Add data to NN* offers the option to add new data to the NN and retrains the NN
 - *Save predictions* saves the predictions in a text or csv file and offers the option to analyse new data

MODULE FUNCTIONS

Functions have been documented using `sphinx.ext.autodoc` and the docstring style extension `'sphinx.ext.napoleon'`:

4.1 Module contours

`ClassificationRBC.contours.al_ratio(contour_length, contour_area)`

Determines ratio of contour length and area

Parameters

- **contour_length** (*ndarray*) – 1 dimensional array with type float that contains length of contours
- **contour_area** (*ndarray*) – 1 dimensional array with type float that contains area of contours

Returns

ratio – 1 dimensional array with type float that contains ratio of contour length to area

Return type

ndarray

`ClassificationRBC.contours.draw_cont(img, contour, name, directory)`

Draws contours and returns image

Parameters

- **img** (*ndarray*) – N dimensional array with type int that contains the image to process
- **contour** (*contours*) – Contours that contains list of coordinates of detected contours
- **name** (*str*) – The name of the returned image
- **directory** (*str*) – The directory where the image is saved

Returns

new_img – N x N dimensional array that contains the selected contours

Return type

ndarray

`ClassificationRBC.contours.filter_cont(contour, index)`

Selects contours by index

Parameters

- **contour** (*contours*) – Contours that contains list of coordinates of detected contours

- **index** (*ndarray*) – 1 dimensional array with type int that contains index of selected contours

Returns

cnt – N x 2 dimensional array with type contours that contains selected contours

Return type

ndarray

`ClassificationRBC.contours.filter_df(ind, c_length, c_area, c_ratio)`

Writes new DataFrame with selected data

Parameters

- **ind** (*ndarray*) – 1 dimensional array with type int that contains the index of data selected from other DataFrame
- **c_length** (*ndarray*) – 1 dimensional array with type float that contains the length of selected contours
- **c_area** (*ndarray*) – 1 dimensional array with type float that contains the area of selected contour
- **c_ratio** (*ndarray*) – 1 dimensional array with type float that contains the ratio of contour length and area

Returns

- **df** (*DataFrame*) – DataFrame that contains properties of selected contours
- **length** (*ndarray*) – 1 dimensional array with type float that contains the length of each selected contour
- **area** (*ndarray*) – 1 dimensional array with type float that contains the area of each selected contour
- **ratio** (*ndarray*) – 1 dimensional array with type float that contains the length/area ratio of each selected contour

`ClassificationRBC.contours.filter_values(df, column, min_val, max_val)`

Filters column in DataFrame by values

Parameters

- **df** (*DataFrame*) – DataFrame that contains contour properties
- **column** (*string*) – The name of the column
- **min_val** (*float*) – The minimum value of the data
- **max_val** (*float*) – The maximum value of the data

Returns

ind – 1 dimensional array with type int that describes values, which fulfill conditions

Return type

ndarray

`ClassificationRBC.contours.find_cont(img)`

Detects contours and draws them in RGB image

Parameters

img (*ndarray*) – N x N dimensional array with type int that contains the image to process

Returns

- **canvas** (*ndarray*) – N x N dimensional array that contains the detected contours

- **contour** (*contours*) – List of coordinates of contours that were detected in the image

ClassificationRBC.contours.**find_contour_img**(*img, img_thr, directory*)

Finds contours, filters them by manually chosen values and returns filtered contours

Parameters

- **img** (*ndarray*) – N x N dimensional array with type int that contains the original image
- **img_thr** (*ndarray*) – N x N dimensional array with type int that contains the thresholded image
- **directory** (*str*) – The directory where the image is saved

Returns

c_a – N x N x 2 array with type contours that contains selected contours

Return type

ndarray

ClassificationRBC.contours.**find_extrema**(*img*)

Finds index of max and min value in image

Parameters

img (*image*) – N dimensional array with type float that contains the z_values of the data

Returns

- **z_min** (*int*) – Minimum value in array
- **z_max** (*int*) – Maximum value in array

ClassificationRBC.contours.**get_contarea**(*contour*)

Determines area that is surrounded by contour and its length

Parameters

contour (*contours*) – Contours that contains list of coordinates of detected contours

Returns

- **length** (*ndarray*) – 1 dimensional array with type float that contains length of contours
- **area** (*ndarray*) – 1 dimensional array with type float that contains area of contours

ClassificationRBC.contours.**normalise_zvalues**(*img*)

Returns normalised grayscale image

Parameters

img (*image*) – N dimensional array with type float that contains the z_values of the data

Returns

imggr – N x N dimensional array with type int that contains grayscale values of the image

Return type

ndarray

ClassificationRBC.contours.**RGB_2_bin**(*img, value*)

Converts RGB image to grayscale image

Parameters

- **img** (*image*) –
- **value** (*int*) – The value, which determines if the point is True or False

Returns

img_new – N x N dimensional array that contains a binary image

Return type

ndarray

ClassificationRBC.contours.**write_df**(*c_length, c_area, ratio*)

Writes data in pandas data frame

Parameters

- **c_length** (*ndarray*) – 1 dimensional array with type float that contains list of contour lengths
- **c_area** (*ndarray*) – 1 dimensional array with type float that contains list of contour areas
- **ratio** (*ndarray*) – 1 dimensional array with type float that contains ratio of contour length to area

Returns

df – DataFrame that contains contour length, area and ratio length/area

Return type

DataFrame

4.2 Module extractCuts

ClassificationRBC.extractCuts.**calc_asymcut**(*values, coord, ori, imgh, imgrh, imgw, imgrw, centre*)

Determines the cross-section along the most asymmetric cut through the *z_values* and shifts the origin to the geometric centre

Parameters

- **values** (*ndarray*) – N dimensional array with type float that contains the grayscale values of the original image
- **coord** (*ndarray*) – 1 dimensional array with type (float, float), (float, float) that contains the x- and y-coordinate left of a point and the x- and y-coordinate right of a point
- **ori** (*str*) – Orientation of the straight: parallel, perpendicular
- **imgh** (*int*) – The height of the image
- **imgrh** (*int*) – The resolution of the image in y-direction
- **imgw** (*int*) – The width of the image
- **imgrw** (*int*) – The resolution of the image in x-direction
- **centre** (*ndarray*) – N dimensional array with type float, float that contains the x- and y-coordinates of the geometric centre

Returns

- **xpara** (*ndarray*) – 1 dimensional array with type float that contains the x-coordinates of the asymmetric cut
- **cypara** (*ndarray*) – 1 dimensional array with type float that contains the y-coordinates of the asymmetric cut
- **zpara** (*ndarray*) – 1 dimensional array with type float that contains the z-coordinates of the asymmetric cut

- **cxperp** (*ndarray*) – 1 dimensional array with type float that contains the x-coordinates of the perpendicular asymmetric cut
- **cyperp** (*ndarray*) – 1 dimensional array with type float that contains the y-coordinates of the perpendicular asymmetric cut
- **zperp** (*ndarray*) – 1 dimensional array with type float that contains the z-coordinates of the perpendicular asymmetric cut
- **zipar** (*ndarray*) – 1 dimensional array with type float that contains the z-coordinates of the asymmetric cut
- **cut** (*ndarray*) – 1 dimensional array with type float that contains the points of the asymmetric cut in μm
- **pcut** (*ndarray*) – 1 dimensional array with type float that contains the points of the perpendicular asymmetric cut in μm
- **x2cs** (*ndarray*) – 2 dimensional array with type float that contains the x-coordinates of the asymmetric parallel and perpendicular cut
- **y2cs** (*ndarray*) – 2 dimensional array with type float that contains the y-coordinates of the asymmetric parallel and perpendicular cut
- **z2cs** (*ndarray*) – 2 dimensional array with type float that contains the z-coordinates of the asymmetric parallel and perpendicular cut
- **cut2** (*ndarray*) – 2 dimensional array with type float that contains the points of the parallel and perpendicular asymmetric cut in μm

`ClassificationRBC.extractCuts.calc_straight(circle_found, mcentre, centre, imgh, imgrh, imgw, imrw)`

Determines the straight between the centre of mass and the geometric centre and the values of the points on the straight within a range of 50 pixels (much larger than cell diameter) around the geometric centre. Calculates the perpendicular straight and the values of the points on the straight.

Parameters

- **circle_found** (*ndarray*) – 1 dimensional array with type float, float, float that contains the x- and y-coordinate of the centre and the radius of the bounding circle
- **mcentre** (*ndarray*) – N dimensional array with type float, float that contains the x- and y-coordinates of the centre of mass weighted by the corresponding z-values
- **centre** (*ndarray*) – N dimensional array with type float, float that contains the x- and y-coordinates of the geometric centre
- **imgh** (*int*) – The height of the image
- **imgrh** (*int*) – The resolution of the image in y-direction
- **imgw** (*int*) – The width of the image
- **imgrw** (*int*) – The resolution of the image in x-direction

Returns

- **xrs** (*ndarray*) – 1 dimensional array with type float that contains all x coordinates on the straight right of the geometric centre within a range of 50 pixels
- **yrs** (*ndarray*) – 1 dimensional array with type float that contains all y coordinates on the straight right of the geometric centre within a range of 50 pixels
- **xls** (*ndarray*) – 1 dimensional array with type float that contains all x coordinates on the straight left of the geometric centre within a range of 50 pixels

- **yIs** (*ndarray*) – 1 dimensional array with type float that contains all y coordinates on the straight left of the geometric centre within a range of 50 pixels
- **xprs** (*ndarray*) – 1 dimensional array with type float that contains all x coordinates on the perpendicular straight right of the geometric centre within a range of 50 pixels
- **yprs** (*ndarray*) – 1 dimensional array with type float that contains all y coordinates on the perpendicular straight right of the geometric centre within a range of 50 pixels
- **xpls** (*ndarray*) – 1 dimensional array with type float that contains all x coordinates on the perpendicular straight left of the geometric centre within a range of 50 pixels
- **ypls** (*ndarray*) – 1 dimensional array with type float that contains all y coordinates on the perpendicular straight left of the geometric centre within a range of 50 pixels

ClassificationRBC.extractCuts.**centre_of_mass**(*all_coord, img*)

Determines the centre of mass inside a rectangle weighted by the corresponding z-values

Parameters

- **all_coord** (*ndarray*) – 1 dimensional array with type float, float that contains all x- and y-coordinates inside a rectangle
- **img** (*ndarray*) – N dimensional array with type float that contains the z_values of the data

Returns

centre_mass – N x 2 array with type float, float that contains the x- and y-coordinates of the centre of mass weighted by the corresponding z-values

Return type

ndarray

ClassificationRBC.extractCuts.**find_coordborder**(*cnt_drawn, img*)

Finds points inside cell contour and returns their coordinates

Parameters

- **cnt_drawn** (*ndarray*) – 1 dimensional array with type int, int, float, float that contains the index of a rectangle with corresponding contour, the y- and x-coordinate of the contour
- **img** (*ndarray*) – N dimensional array with type float that contains the z_values of the data

Returns

img_coord – N x 3 array with type float, float, float that contains the y-coordinate, the minimum and maximum x-coordinate of the points inside a contour

Return type

ndarray

ClassificationRBC.extractCuts.**geometric_centre**(*img_coord*)

Determines the geometric centre of a contour (cell)

Parameters

img_coord (*ndarray*) – 1 dimensional array with type float, float, float that contains the y-coordinate, the minimum and maximum x-coordinate

Returns

- **centre** (*ndarray*) – N x 2 array with type float, float that contains the x- and y-coordinate of the centre of a rectangle
- **all_coord** (*ndarray*) – N x N x 2 array with type float, float that contains all x- and y-coordinates inside a rectangle

`ClassificationRBC.extractCuts.match_coord(x_straight, y_straight, cntCell)`

Finds contour points on the straight and sorts the coordinates by distance of the point from the geometric centre

Parameters

- **x_straight** (*ndarray*) – 1 dimensional array with type float that contains all x coordinates on the straight within a range of 50 pixels
- **y_straight** (*ndarray*) – 1 dimensional array with type float that contains all x coordinates on the straight within a range of 50 pixels
- **cnt_drawn** (*ndarray*) – 1 dimensional array with type int, int, float, float that contains the index of a rectangle with corresponding contour, the y- and x-coordinate of the contour

Returns

- **match_coord** (*ndarray*) – N x 3 array with type float, float, float that contains the x-coordinate, the y-coordinate and the distance of the point from the geometric centre
- **sort_coord** (*ndarray*) – N x 2 array with type float, float that contains the x- and y-coordinate of a point

`ClassificationRBC.extractCuts.points_incell(img_coord, img)`

Determines the z-value of all points inside a contour

Parameters

- **img_coord** (*ndarray*) – 1 dimensional array with type float, float, float that contains the y-coordinate, the minimum and maximum x-coordinate of the points inside a contour
- **img** (*ndarray*) – N dimensional array with type float that contains the z_values of the data

Returns

points_inside – N x N dimensional array with type float that contains all z-values inside a contour

Return type

ndarray

`ClassificationRBC.extractCuts.sort_coord(coord_rmatch, coord_lmatch)`

Sorts the matching coordinates by distance between left point and right point

Parameters

- **coord_rmatch** (*ndarray*) – 2 dimensional array with type float, float that contains the x-coordinate and the y-coordinate right of the centre
- **coord_lmatch** (*ndarray*) – 2 dimensional array with type float, float that contains the x-coordinate and the y-coordinate left of the centre

Returns

coord_sort – N x 2 dimensional array with type float, float that contains the x- and y-coordinates of the points intersecting with the straight of highest asymmetry left and right of the centre of the object

Return type

ndarray

4.3 Module NN

ClassificationRBC.NN.**format_values**(*name*, *spath*)

Formats values and writes them into array

Parameters

- **name** (*str*) – Name of the txt file
- **spath** (*str*) – Name of the directory
- **parameter** (*str*) – Name of the parameter

Returns

values – 1 dimensional array with type float that contains the content of the text file

Return type

ndarray

ClassificationRBC.NN.**format_NN_values**(*name*, *spath*)

Formats values for NN input and writes them into array

Parameters

- **name** (*str*) – Name of the txt file
- **spath** (*str*) – Name of the directory
- **parameter** (*str*) – Name of the parameter

Returns

values – 1 dimensional array with type float that contains the content of the text file

Return type

ndarray

ClassificationRBC.NN.**read_images**(*path*)

Reads text files of characteristic cuts and labels them according to their category

Parameters

path (*str*) – The name of the folder that contains the characteristic cuts

Returns

- **carr_form** (*ndarray*) – N x (N x N) dimensional array with type float that contains N times two characteristic cross-sections with 50 data points
- **label_form** (*ndarray*) – 1 dimensional array with type int that contains the labels of N cells

ClassificationRBC.NN.**train_model**(*model*, *carr_form*, *label_form*, *cut_train*, *label_train*, *directory*, *mname*, *llabel*)

Trains NN model on a combination of old and new data

Parameters

- **model** (*Sequential*) – The pre-trained neural network
- **carr_form** (*ndarray*) – N x 2 x 50 array with type float that contains N times two characteristic cross-sections with 50 data points
- **label_form** (*ndarray*) – N x 1 array with type int that contains the labels of N cells
- **directory** (*str*) – The name of the network folder
- **mname** (*str*) – The name of the new network

- **tlabel** (*tkinter Label*) – Displays the progres of the calculation

Returns

model – Sequential class that contains the trained new keras model

Return type

tf.keras.Sequential

MODULE CLASSES

Classes have been documented using `sphinx.ext.autodoc` and the docstring style extension `'sphinx.ext.napoleon'`:

class `ClassificationRBC.classes.AFMing`(*xvalues*, *yvalues*, *zvalues*, *filename*, *directory*)

A class that contains the x-, y-, z-values, filename and directory of the raw AFM file

xvalues

N dimensional array with type float that contains the x-values of the igor pro height trace data

Type

ndarray

yvalues

N dimensional array with type float that contains the y-values of the igor pro height trace data

Type

ndarray

zvalues

N dimensional array with type float that contains the z-values of the igor pro height trace data

Type

ndarray

filename

String that contains the name of the .txt file

Type

str

directory

String that contains the directory of the .txt file

Type

str

create_directory(*directory*, *name*)

Creates a new directory

Parameters

- **directory** (*str*) – The directory of the new folder
- **name** (*str*) – The name of the new folder

Returns

new_dir – The directory and name of the new folder

Return type

str

set_directory(*new_dir*)

Adds a new directory to the class

Parameters

new_dir (str) – The new directory

class ClassificationRBC.classes.**contAFMimg**(*xvalues, yvalues, zvalues, filename, directory, propAFMimg, rect_del, rect2_del, rect_exp, cnt_all, cnt_del, img_cnt, img_gr, img_rgb, coord, cdrawn*)

A class which inherits from AFMimg and contains the contour and the rectangles covering the objects in the image

rect_del

1 dimensional array with type rectangle that contains rectangles around contours

Type

ndarray

rect2_del

1 dimensional array with type int, rectangle that contains rectangles around clicked coordinates

Type

ndarray

rect_exp

1 dimensional array with type rectangle that contains the not selected expanded rectangles

Type

ndarray

cnt_all

1 dimensional array with type contours that were not selected manually

Type

ndarray

cnt_del

1 dimensional array with type contours that contains all not selected contours in the loop

Type

ndarray

img_cnt

N dimensional array with type int that contains all not selected contours

Type

image

img_gr

N dimensional array with type float that contains a greyscale image of the original image

Type

ndarray

img_rgb

N dimensional array with type float that contains img_gr as rgb image with detected contours

Type
ndarray

coord

N dimensional array with type float that contains the coordinates of points inside manually selected rectangles/contours

Type
ndarray

cdrawn

1 dimensional array with type int, int, float, float that contains the index of a rectangle with corresponding contour, the x- and y-coordinate of the contour

Type
ndarray

class ClassificationRBC.classes.**propAFMimg**(*xvalues, yvalues, zvalues, filename, directory, imgh, imgw, imgrh, imgrw, thr*)

A class which inherits from AFMimg and contains size, threshold value and the expansion factor of the rectangles covering the objects in the image

imgh

The height of the image in μm

Type
float

imgw

The width of the image in μm

Type
float

imgrh

The resolution of the yvalues

Type
int

imgrw

The resolution of the xvalues

Type
int

thr

Manually chosen threshold value for grayscale images

Type
int

class ClassificationRBC.classes.**LMimg**(*orig, values, val_enh, filename, directory*)

A class that contains the x-, y-, z-values, filename and directory of the raw LM file

orig

N dimensional array with type int that contains the colour values of the LM image

Type
ndarray

values

N dimensional array with type int that contains the greyscale values of the LM image

Type

ndarray

val_enh

N dimensional array with type int that contains the enhanced LM image

Type

ndarray

filename

String that contains the name of the .txt file

Type

str

directory

String that contains the directory of the .txt file

Type

str

create_directory(*directory, name*)

Creates a new directory

Parameters

- **directory** (*str*) – The directory of the new folder
- **name** (*str*) – The name of the new folder

Returns

new_dir – The directory and name of the new folder

Return type

str

set_directory(*new_dir*)

Adds a new directory to the class

Parameters

new_dir (*str*) – The new directory

class ClassificationRBC.classes.**contLMimg**(*orig, values, val_enh, filename, directory, propLMimg, rect_del, rect2_del, rect_exp, cnt_all, cnt_del, img_cnt, img_gr, img_rgb, coord, cdrawn*)

A class which inherits from LMimg and contains the contour and the rectangles covering the objects in the image

rect_del

1 dimensional array with type rectangle that contains rectangles around contours

Type

ndarray

rect2_del

1 dimensional array with type int, rectangle that contains rectangles around clicked coordinates

Type

ndarray

rect_exp

1 dimensional array with type rectangle that contains the not selected expanded rectangles

Type

ndarray

cnt_all

1 dimensional array with type contours that were not selected manually

Type

ndarray

cnt_del

1 dimensional array with type contours that contains all not selected contours in the loop

Type

ndarray

img_cnt

N dimensional array with type int that contains all not selected contours

Type

image

img_gr

N dimensional array with type float that contains a greyscale image of the original image

Type

ndarray

img_rgb

N dimensional array with type float that contains img_gr as rgb image with detected contours

Type

ndarray

coord

N dimensional array with type float that contains the coordinates of points inside manually selected rectangles/contours

Type

ndarray

cdrawn

1 dimensional array with type int, int, float, float that contains the index of a rectangle with corresponding contour, the x- and y-coordinate of the contour

Type

ndarray

class ClassificationRBC.classes.**propLMimg**(*orig, values, val_enh, filename, directory, imgh, imgw, imgrh, imgrw, thr*)

A class which inherits from LMimg and contains size, threshold value of the LM image

imgh

The height of the image in μm

Type

float

imgw

The width of the image in μm

Type

float

imgrh

The resolution of the yvalues

Type

int

imgrw

The resolution of the xvalues

Type

int

thr

Manually chosen threshold value for grayscale images

Type

int

class `ClassificationRBC.classes.Cell`(*img, gfimg, zcut2, label, cntCell, centre, mcentre, radius, rectChange, offset*)

A class that contains the geometric properties of a cell

img

$N \times N \times N$ dimensional array that contains the extracted cell images

Type

ndarray

gfimg

$N \times N \times N$ dimensional array that contains the gaussian filtered images

Type

ndarray

zcut2

$N \times 2$ dimensional array with type float that contains two extracted cuts

Type

ndarray

label

N dimensional array with type str that contains the labels of the images

Type

ndarray

cntCell

$N \times N \times (1, 1)$ dimensional array with type (int, int) that contains the contour of the cells

Type

ndarray

centre

N x (1, 1) dimensional array with type (float, float) that contains the centre of the cell

Type

ndarray

mcentre

N x (1, 1) dimensional array with type (float, float) that contains the centre of mass of the cell

Type

ndarray

radius

N dimensional array with type float that contains the radius of the detected circle

Type

ndarray

rectChange

N x (1, 1, 1, 1, 1, 1, 1) dimensional array with type float that contains the bounding coordinates of the bounding rectangle around a cell

Type

ndarray

offset

N x (1, 1) dimensional array with type float that contains the x and y offset of the cropped image

Type

ndarray

calc_cellCnt(*circle_found, img_hc, directory*)

Calculates the coordinates of the cell contour

Parameters

- **circle_found** (*ndarray*) – N x (1, 1, 1) dimensional array with type float that contains the centre and the radius of the detected circle
- **img_hc** (*ndarray*) – N x N dimensional array that contains the normalised zvalues

calc_geomP()

Calculates the coordinates inside the cell, its centre, and its centre of mass

extract_cut(*circle_found, imgh, imgw, imgrh, imgrw*)

Extracts cut from cell image

Parameters

- **circle_found** (*ndarray*) – N x (1, 1, 1) dimensional array with type float that contains the centre and the radius of the detected circle
- **imgh** (*float*) – The height of the image in μm
- **imgw** (*float*) – The width of the image in μm
- **imgrh** (*int*) – The resolution of the yvalues
- **imgrw** (*int*) – The resolution of the xvalues

class ClassificationRBC.classes.**ToolTip**(*widget, text*)

Class that defines a window upon hovering mouse event

hideTooltip()

Hides a window, when previously shown by hovering above an icon

showTooltip()

Shows a window, when hovering above an icon

A

AFMimg (class in *ClassificationRBC.classes*), 17
 al_ratio() (in module *ClassificationRBC.contours*), 7

C

calc_asymcut() (in module *ClassificationRBC.extractCuts*), 10
 calc_cellCnt() (*ClassificationRBC.classes.Cell* method), 23
 calc_geomPC() (*ClassificationRBC.classes.Cell* method), 23
 calc_straight() (in module *ClassificationRBC.extractCuts*), 11
 cdrawn (*ClassificationRBC.classes.contAFMimg* attribute), 19
 cdrawn (*ClassificationRBC.classes.contLMimg* attribute), 21
 Cell (class in *ClassificationRBC.classes*), 22
 centre (*ClassificationRBC.classes.Cell* attribute), 22
 centre_of_mass() (in module *ClassificationRBC.extractCuts*), 12
 cnt_all (*ClassificationRBC.classes.contAFMimg* attribute), 18
 cnt_all (*ClassificationRBC.classes.contLMimg* attribute), 21
 cnt_del (*ClassificationRBC.classes.contAFMimg* attribute), 18
 cnt_del (*ClassificationRBC.classes.contLMimg* attribute), 21
 cntCell (*ClassificationRBC.classes.Cell* attribute), 22
 contAFMimg (class in *ClassificationRBC.classes*), 18
 contLMimg (class in *ClassificationRBC.classes*), 20
 coord (*ClassificationRBC.classes.contAFMimg* attribute), 19
 coord (*ClassificationRBC.classes.contLMimg* attribute), 21
 create_directory() (*ClassificationRBC.classes.AFMimg* method), 17
 create_directory() (*ClassificationRBC.classes.LMimg* method), 20

D

directory (*ClassificationRBC.classes.AFMimg* attribute), 17
 directory (*ClassificationRBC.classes.LMimg* attribute), 20
 draw_cont() (in module *ClassificationRBC.contours*), 7

E

extract_cut() (*ClassificationRBC.classes.Cell* method), 23

F

filename (*ClassificationRBC.classes.AFMimg* attribute), 17
 filename (*ClassificationRBC.classes.LMimg* attribute), 20
 filter_cont() (in module *ClassificationRBC.contours*), 7
 filter_df() (in module *ClassificationRBC.contours*), 8
 filter_values() (in module *ClassificationRBC.contours*), 8
 find_cont() (in module *ClassificationRBC.contours*), 8
 find_contour_img() (in module *ClassificationRBC.contours*), 9
 find_coordborder() (in module *ClassificationRBC.extractCuts*), 12
 find_extrema() (in module *ClassificationRBC.contours*), 9
 format_NN_values() (in module *ClassificationRBC.NN*), 14
 format_values() (in module *ClassificationRBC.NN*), 14

G

geometric_centre() (in module *ClassificationRBC.extractCuts*), 12
 get_contarea() (in module *ClassificationRBC.contours*), 9
 gfmimg (*ClassificationRBC.classes.Cell* attribute), 22

H

hideTooltip() (*ClassificationRBC.classes.ToolTip*

- method), 23
- I
- img (ClassificationRBC.classes.Cell attribute), 22
- img_cnt (ClassificationRBC.classes.contAFMimg attribute), 18
- img_cnt (ClassificationRBC.classes.contLMimg attribute), 21
- img_gr (ClassificationRBC.classes.contAFMimg attribute), 18
- img_gr (ClassificationRBC.classes.contLMimg attribute), 21
- img_rgb (ClassificationRBC.classes.contAFMimg attribute), 18
- img_rgb (ClassificationRBC.classes.contLMimg attribute), 21
- img_h (ClassificationRBC.classes.propAFMimg attribute), 19
- img_h (ClassificationRBC.classes.propLMimg attribute), 21
- img_rh (ClassificationRBC.classes.propAFMimg attribute), 19
- img_rh (ClassificationRBC.classes.propLMimg attribute), 22
- img_rw (ClassificationRBC.classes.propAFMimg attribute), 19
- img_rw (ClassificationRBC.classes.propLMimg attribute), 22
- img_w (ClassificationRBC.classes.propAFMimg attribute), 19
- img_w (ClassificationRBC.classes.propLMimg attribute), 21
- L
- label (ClassificationRBC.classes.Cell attribute), 22
- LMimg (class in ClassificationRBC.classes), 19
- M
- match_coord() (in module ClassificationRBC.extractCuts), 12
- mcentre (ClassificationRBC.classes.Cell attribute), 23
- N
- normalise_zvalues() (in module ClassificationRBC.contours), 9
- O
- offset (ClassificationRBC.classes.Cell attribute), 23
- orig (ClassificationRBC.classes.LMimg attribute), 19
- P
- points_incell() (in module ClassificationRBC.extractCuts), 13
- propAFMimg (class in ClassificationRBC.classes), 19
- propLMimg (class in ClassificationRBC.classes), 21
- R
- radius (ClassificationRBC.classes.Cell attribute), 23
- read_images() (in module ClassificationRBC.NN), 14
- rect2_del (ClassificationRBC.classes.contAFMimg attribute), 18
- rect2_del (ClassificationRBC.classes.contLMimg attribute), 20
- rect_del (ClassificationRBC.classes.contAFMimg attribute), 18
- rect_del (ClassificationRBC.classes.contLMimg attribute), 20
- rect_exp (ClassificationRBC.classes.contAFMimg attribute), 18
- rect_exp (ClassificationRBC.classes.contLMimg attribute), 21
- rectChange (ClassificationRBC.classes.Cell attribute), 23
- RGB_2_bin() (in module ClassificationRBC.contours), 9
- S
- set_directory() (ClassificationRBC.classes.AFMimg method), 18
- set_directory() (ClassificationRBC.classes.LMimg method), 20
- showTooltip() (ClassificationRBC.classes.ToolTip method), 24
- sort_coord() (in module ClassificationRBC.extractCuts), 13
- T
- thr (ClassificationRBC.classes.propAFMimg attribute), 19
- thr (ClassificationRBC.classes.propLMimg attribute), 22
- ToolTip (class in ClassificationRBC.classes), 23
- train_model() (in module ClassificationRBC.NN), 14
- V
- val_enh (ClassificationRBC.classes.LMimg attribute), 20
- values (ClassificationRBC.classes.LMimg attribute), 19
- W
- write_df() (in module ClassificationRBC.contours), 10
- X
- xvalues (ClassificationRBC.classes.AFMimg attribute), 17
- Y
- yvalues (ClassificationRBC.classes.AFMimg attribute), 17

Z

`zcut2` (*ClassificationRBC.classes.Cell* attribute), 22

`zvalues` (*ClassificationRBC.classes.AFMimg* attribute),

17

Acronyms

1D one-dimensional. 71

2D two-dimensional. 24, 68, 76

AFM atomic force microscopy. 2, 14, 16, 17, 18, 19, 31, 32, 34, 35, 36, 37, 38, 39, 40, 41, 42, 44, 45, 46, 47, 48, 49, 50, 53, 60, 62, 63, 64, 65, 68, 71, 72, 74, 77, 79, 101, 102, 150

ANN artificial neural networks. 25

ANNs artificial neural networks. 25

CNN convolutional neural network. 13, 28, 29, 30, 68, 150

CO₂ carbon dioxide. 15, 33

DNA Deoxyribonucleic acid. 5, 10, 12

FN false negative. 30

FP false positive. 29, 30

KCl potassium chloride. 33

N₂ nitrogen. 33

NAAT nucleic acid amplification test. 10, 12

NaCl sodium chloride. 14, 15, 33

NaHCO₃ sodium bicarbonate. 33

NN neural network. 2, 3, 4, 10, 11, 12, 13, 24, 25, 28, 29, 30, 31, 37, 38, 41, 43, 62, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 84, 85, 86, 107, 108, 109, 110, 111, 112, 113, 114, 115, 150, 151, 152, 155

O₂ oxygen. 15, 33

PBS phosphate-buffered saline. 32

PCR polymerase chain reaction. 1, 4, 12

PMT photomultiplier tube. 23

RBC red blood cell. 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 24, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 67, 68, 69, 71, 72, 73, 75, 76, 77, 78, 79, 80, 82, 84, 85, 100, 104, 105, 150, 151, 153

RBCs red blood cells. 1

ReLU rectified linear unit. 27

RMOD rotating-crystal magneto-optical detection. 10, 12

RPMI 1640 Roswell Park Memorial Institute cell medium. 15, 16, 32, 33

TN true negative. 29, 30

TP true positive. 29, 30

ZnPPIX zinc protoporphyrin IX. 9, 59

List of Figures

2.1	Schematic of the malaria parasite life cycle.	6
2.2	Representative drawings of microscopy images of RBCs infected with <i>P. falciparum</i>	8
2.3	Optical absorption of haemoglobin.	9
2.4	Optical properties of malaria-infected RBCs	10
3.1	Workflow of culturing.	15
3.2	Schematic setup of AFM.	17
3.3	Tip-sample interaction.	18
3.4	Typical cantilever geometries	20
3.5	Resonance frequency in air.	21
3.6	Jablonski diagram.	22
3.7	Illumination in wide-field and confocal fluorescence microscopy	23
3.8	Components of a biological NN.	25
3.9	Schematics of a simple perceptron.	26
3.10	Schematics of backpropagation in a neural network.	28
3.11	Common activation functions.	28
3.12	Representation of a CNN.	30
3.13	Confusion matrix for a binary classification problem.	30
3.14	Confusion matrix for a four category classification.	31
4.1	Cell coverage at various RBC concentrations	35
4.2	AFM image of a healthy RBCs in liquid environment.	36
4.3	Visualisation of microscopy images with artificial colouring.	38
4.4	Binarised AFM image.	39
4.5	Graphical representation of the Hough circle method.	39
4.6	Representative images of not circular RBCs	40
4.7	Cell detection on microscopy images.	41
4.8	Images of manually classified RBCs	42
4.9	Geometric augmentations on single RBC images.	43
4.10	Illustration of radially spaced cross-sections.	44
4.11	Radial cross-sections through representative RBC images.	45
4.12	Topographical alterations of <i>P. falciparum</i> -infected erythrocytes.	47
4.13	Asymmetric height profiles of <i>P. falciparum</i> -infected RBCs.	48
4.14	Characteristic cross-sections of infected RBCs indicating the location of the food vacuole.	50

4.15	Height profiles of Giemsa-stained and unstained <i>P. falciparum</i> -infected RBCs.	51
4.16	Comparison of averaged height profiles of RBCs infected with the intra-erythrocytic stages of <i>P. falciparum</i> before and after image processing.	53
4.17	Fluorescence maps of <i>P. falciparum</i> -infected erythrocytes at various excitation wavelengths.	54
4.18	Autofluorescence maps of <i>P. falciparum</i> -infected erythrocytes.	55
4.19	Radial intensity profiles corresponding to the trophozoite- and schizont-stage parasites in figure 4.18.	56
4.20	Autofluorescence maps of hemozoin crystals.	57
4.21	Characteristic autofluorescence maps of <i>P. falciparum</i> -infected RBCs.	58
4.22	RBC images showing the emergence of autofluorescence in living parasites.	59
4.23	Light microscopy images of Giemsa-stained erythrocytes infected with <i>P. falciparum</i>	61
4.24	Statistical analysis of Giemsa-stained erythrocytes infected with <i>P. falciparum</i>	62
4.25	Artefacts in atomic force and fluorescence microscopy images due to external noise.	63
4.26	Location of the geometric and gravitational centre in simple cylinders with holes.	64
4.27	Characteristic cross-sections of a parasite.	66
4.28	Cross-sections of malaria-infected RBCs imaged with atomic force, fluorescence, and light microscopy.	67
4.29	Classification performance of the NN on microscopy images.	70
4.30	Classification of the NN on radial cuts	73
4.31	Classification performance of the NN on randomly oriented cuts.	75
4.32	Classification performance of the convolutional NN and computation time per RBC versus the amount of input data.	76
4.33	Setting of initial parameters.	79
4.34	Show image.	80
4.35	Threshold image.	81
4.36	Detect cells.	82
4.37	Cell classification.	83
A.1	Neural network architecture for the stage-specific classification of RBCs of two-dimensional input images cuts.	104
A.2	Neural network architecture for the stage-specific classification of RBCs of one-dimensional cuts.	105
A.3	Classification performance of the NN on four cuts.	107
A.4	Classification performance of the NN on the parasite plus cuts.	108
A.5	Classification performance of the NN on the parasite cut.	109
A.6	Classification performance of the NN on the four cuts.	110

A.7	Classification performance of the NN on the parasite plus cuts.	111
A.8	Classification performance of the NN on the parasite cut.	112
A.9	Classification performance of the NN on four randomly oriented cuts.	113
A.10	Classification performance of the NN on two randomly oriented cuts.	114
A.11	Classification performance of the NN on one randomly oriented cut.	115

List of Tables

3.1	Components of the culture medium.	33
4.1	Cantilever type and spring constant with corresponding success of RBC imaging.	36
4.2	Number of recorded RBCs and ratio of infected cells (parasitemia).	38
4.3	Optimum values for the sensitivity of the Hough gradient method.	40
4.4	Probability of cell detection for different imaging techniques.	41
4.5	Distribution of healthy RBCs and intra-erythrocytic stages.	42
4.6	Distribution of Giemsa-stained and unstained RBCs.	49
4.7	Characteristic values from the analysis of mean height profiles.	51
4.8	Characteristic values from the analysis of mean height profiles of RBCs before and after applying a Gaussian filter to the images.	52
4.9	Analysis of stage-specific intensity patterns obtained from fluorescence microscopy images of individual RBCs.	58
4.10	Characteristic statistical values for the intensity curves of Giemsa-stained RBCs.	61
4.11	Percentage of successful parasite detection with the parasite cut.	64
4.12	Percentage of successful parasite detection with the parasite cut.	64
4.13	Classification accuracy on a test set of RBCs.	69
4.14	Classification accuracy on the four cuts, parasite plus cuts, parasite cut.	72
4.15	Classification accuracy on the four cuts, parasite plus cuts, parasite cut.	74
4.16	Prediction accuracy of swapped parasite plus cuts ($2n$).	74
A.1	Results of the t-test with for the stage-wise comparison of stained and unstained samples (AFM).	101
A.2	Results of the t-test with for the stage-wise comparison of samples with and without background (AFM).	101
A.3	Results of the t-test with for the comparison between the stages in stained samples (AFM).	101
A.4	Results of the t-test with for the comparison between the stages in unstained samples (AFM).	101
A.5	Results of the t-test with for the comparison between the stages in samples with background (AFM).	102
A.6	Results of the t-test with for the comparison between the stages in samples without background (AFM).	102

A.7	Results of the t-test with for the stage-wise comparison in samples with and without background (FM).	102
A.8	Results of the t-test with for the comparison between the stages in samples with background (FM).	102
A.9	Results of the t-test with for the comparison between the stages in samples without background (FM).	103
A.10	Results of the t-test with for the stage-wise comparison in samples with and without background (LM)	103
A.11	Results of the t-test with for the comparison between the stages in samples with background (LM).	103
A.12	Results of the t-test with for the comparison between the stages in samples without background (LM).	103

Acknowledgements

As no PhD study is done alone, I would like to thank my colleagues, friends, and family, who made the work leading to this thesis possible. In particular:

Prof. Dr. István Kézsmárki for providing the opportunity pursue a bilateral PhD with an interesting topic in Budapest and Augsburg. As my supervisor, you guided me through the Hungarian PhD system. With your interesting ideas, you encouraged me to think outside the box.

Prof. Dr. Beáta Vértessy for acting as my co-supervisor and providing the opportunity to visit your malaria laboratory, whenever I needed. I greatly thank you for your support during my PhD and for welcoming me into your team.

Prof. Dr. János Török for helping me with all my problems during programming and for sharing your knowledge about NNs with me. I also thank you for the helpful discussions during my PhD. I really liked working with you and I am grateful that you supported me in writing my papers and thesis. Thank you for always being available when problems arose and finding creative solutions.

Prof. Dr. Miklós Kellermayer for granting me access to your lab and for your continuous interest in my work. You greatly helped me my PhD thesis with your knowledge and ideas.

Bálint Kiss for always helping me with the microscopy setup in the lab. Thank you for solving my problems with your creative ideas.

Dr. Petra Molnár, Dr. Agnes Orbán and Richard Izrael for helping me with everything connected to malaria. Thank you for great memories during my stay in Budapest. I also thank you for always answering my questions on malaria and Hungary.

Maximilian Winkler, Arthur Schulz, and Dr. Dmytro Kamenskyi for their scientific and social contributions. You made lunch and coffee breaks highlights of my work days.

Birgitta Eisenschmid and Anny Skroblied for your help with all administrative affairs.

Christine and Günther Preißinger, my parents, and Markus Preißinger, my brother, for supporting me during the challenging times in Budapest and Augsburg. I am

grateful for your patience and for reminding me of the things that matter the most.
Thank you!