Composition Counts: A Machine Learning View on Immunothrombosis using Quantitative Phase Imaging

David Fresacher[*1] DAVID.FRESACHER@TUM.DE
Stefan Röhrl[*1] STEFAN.ROEHRL@TUM.DE
Christian Klenk[2] CHRISTIAN.KLENK@TUM.DE
Johanna Erber[3] JOHANNA.ERBER@MRI.TUM.DE
Hedwig Irl[3] HEDWIG.IRL@MRI.TUM.DE
Dominik Heim[2] DOMINIK.HEIM@TUM.DE
Manuel Lengl[1] M.LENGL@TUM.DE
Simon Schumann[1] SIMON.SCHUMANN@TUM.DE
Martin Knopp[1,2] MARTIN.KNOPP@TUM.DE
Martin Schlegel[3] MARTIN.SCHLEGEL@TUM.DE
Sebastian Rasch[3] SEBASTIAN.RASCH@TUM.DE
Oliver Hayden[2] OLIVER.HAYDEN@TUM.DE
Klaus Diepold[1] KLDI@TUM.DE

[1] *Chair of Data Processing,*
[2] *Heinz-Nixdorf-Chair of Biomedical Electronics, and*
[3] *University Hospital rechts der Isar*
Technical University of Munich, Germany
[*] *These two authors contributed equally to this work*

Abstract

Thrombotic complications are a leading cause of death worldwide, often triggered by inflammatory conditions such as sepsis and COVID-19, due to a close relationship between inflammation and hemostasis known as immunothrombosis. Platelet activation and leukocyte-platelet aggregation play key roles in microthrombotic events, yet there are no routine diagnostic predictive biomarkers based on these factors. This work presents a novel processing pipeline using label-free Quantitative Phase Imaging (QPI) for the detection and quantitative analysis of blood cell aggregates without sample preparation. For evaluation, we use different test scenarios and measure performance at different stages of the pipeline to gain a better understanding of the critical points. We show that, among other classical and machine learning techniques, the Mask R-CNN approach achieves the best results for detection, segmentation, and classification of cell aggregates. The method successfully identifies aggregate levels in whole blood samples and shows elevated levels in >90% of patients with COVID-19 or sepsis compared to healthy reference samples, indicating the potential of platelet and leukocyte-platelet aggregates as biomarkers for thrombotic diseases.

1. Introduction

Motivation Thrombotic conditions are considered the leading cause of mortality worldwide and the number of patients is steadily increasing, especially in developing and first world countries (Wendelboe and Raskob, 2016). Different types of thrombosis include arte-

rial thrombosis (e.g. in the form of coronary heart disease or ischemic stroke) and venous thrombosis (e.g. in the form of deep vein thrombosis or pulmonary embolism). As thrombotic events are closely related to coagulation (blood clotting) and hemostasis in general, one of the key players are platelets (thrombocytes). Their dysfunction can have serious consequences. Thrombocyte hyperreactivity can lead to venous or arterial thrombosis and subsequently to pulmonary embolism, myocardial infarction, and stroke (Engelmann and Massberg, 2013; Nicolai et al., 2020).

Until recently, hemostasis and inflammation were thought to be completely separate physiological processes. However, recent research has shown that these two processes are intimately linked. This close relationship between coagulation and inflammation is called **immunothrombosis** (Engelmann and Massberg, 2013), which is based on the interaction of immune cells and thrombosis-related molecules. Immunothrombosis is an important defense mechanism to prevent the systemic spread of pathogens through the bloodstream by facilitating the recognition, containment, and destruction of pathogens (Stark and Massberg, 2021). However, uncontrolled immunothrombosis leads to a general risk of blood clotting, promoting the formation of microthrombi and, in the worst case, organ failure (Engelmann and Massberg, 2013).

The most recent and prominent example of an uncontrolled inflammatory response associated with thrombotic risk is COVID-19. While in most cases this infection is asymptomatic or accompanied by mild flu-like symptoms, in severe cases pulmonary complications associated with a systemic inflammatory response can occur, with potentially fatal consequences. Many recent publications indicate the occurrence of immunothrombosis with micro- and macrovascular thrombi (Nicolai et al., 2020; Schulte-Schrepping et al., 2020; Nishikawa et al., 2021; Zuo et al., 2021). Another example of the emergence of immunothrombosis is sepsis, where an initially appropriate and targeted immune response becomes generalized and harmful hyperactivation leading to organ failure (Hotchkiss et al., 2016). The appearance of activated platelets and leukocyte-platelet aggregates plays an important role in this process (Assinger et al., 2019). Due to its acute pathology, an immediate medical response is required, showing the necessity of early diagnosis (Rhodes et al., 2017).

Despite the emerging demand, no diagnostic predictive biomarker is available for routine economic diagnosis due to the highly complex pre-analytics and sample preparation required (with typically expensive antibody-based activation markers) as well as the short lifetime of cell aggregates (Finsterbusch et al., 2018). However, with the use of QPI, label-free analysis of blood cells and their aggregates becomes feasible, possibly even in a point-of-care application (Nguyen et al., 2022).

Problem statement While cell detection and classification have already been demonstrated for phase images of blood cells obtained with QPI (Ugele et al., 2018b,a; Paidi et al., 2021), the analysis of cell aggregates has proven to be more difficult due to their complex morphology, small details and short lifetime (Finsterbusch et al., 2018). In addition, their rare occurrence usually requires extensive sample preparation (Nishikawa et al., 2021).

Therefore, in this work, we design and test a data processing system that allows for the analysis of phase images of whole blood samples (obtained by QPI) for the size, number and composition of platelet and leukocyte-platelet aggregates. In addition, we evaluate relationships and correlations of these aggregate data with disease and infection using clin-

ical samples from patients with COVID-19 and sepsis. The concept is the implementation and evaluation of a three-step pipeline for the quantitative analysis of aggregates and their components. The first step is the **detection and separation** of aggregates, specifically platelet aggregates and leukocyte-platelet aggregates, in whole blood samples. The second step is to **evaluate** the detected aggregates. This includes assessing the number of cells in an aggregate and the specific type of each cell. The last step is the integration of all the previous results and the search for **correlations with immunothrombotic diseases**. Specifically, we are analyzing sepsis and COVID-19 in comparison to healthy individuals using multiple samples from 27 subjects.

Generalizable Insights about Machine Learning in the Context of Healthcare

In our work, we present improved approaches to better understand the effects of immunothrombosis and to generate detailed information about the composition of volatile microthrombotic events. This is done under more demanding conditions because, unlike previous methods, we work label-free and with whole blood, which minimizes sample preparation. Here, we can show that our proposed machine learning pipeline is more robust to these conditions and generalizes better than the state of the art. To this aim, we not only evaluate the end-to-end performance, but also measure meaningful metrics at different points within the pipeline to better assess the behavior of the algorithms. In addition, we are introducing test scenarios to incrementally approach real-world conditions and exemplary clinical use cases in order to identify the factors that cause problems for the algorithms. We hope to lay the groundwork for using the QPI platform technology to analyze blood cell aggregates as biomarkers for predictive and individual diagnostics in subsequent clinical studies. Finally, we provide best practices for expansion into new applications that have already been shown to be related to immunothrombosis, such as hemophilia (Riedl et al., 2017), anticoagulation therapies (Lazaridis et al., 2022) or cardiovascular diseases in general (Furman et al., 2001; Allen et al., 2019).

2. Background and Related Work

Before proceeding to our proposed approach, we will look at the state of the art in observing the biomedical effects we are interested in. We will also give insights into the QPI technology and its combination with machine learning.

2.1. Medical Relevance

While the coagulation process was first discovered more than 100 years ago, in recent years coagulopathy, thrombocytopathy, and immunothrombosis have attracted increasing interest in the scientific community due to the discovery of the important role that coagulation plays in the development of cardiovascular diseases (Bhatt and Topol, 2003). Since then, a great deal of research has been conducted in this area. The role of thrombosis as an independent process of innate immunity was investigated by Engelmann and Massberg (2013), leading to the introduction of the term immunothrombosis. Successively, several researchers have shown the intricate relationship between hemostasis and inflammation (Stark and Massberg, 2021; Reyes et al., 2020; van der Poll et al., 2017).

Due to the recent emergence of a new variant of coronaviruses causing COVID-19, which has evolved into a worldwide pandemic, a great deal of research has been initiated targeting the thrombotic features of this disease. Nicolai et al. (2020) provided evidence for the involvement of immunothrombosis, while Zuo et al. (2021) discussed the process behind the formation of microthrombi, and Schulte-Schrepping et al. (2020) provided detailed insights into the systemic immune response. Most notable is the work of Nishikawa et al. (2021), who were able to show a direct link between aggregates and disease severity. Unfortunately, the analysis of the aggregates remains superficial and is limited to estimating the area of the aggregates and a fixed conversion factor for the number of platelets. No individual analysis of the aggregate components is performed, as proposed by Klenk et al. (2023). Moreover, their method requires a laborious sample preparation of up to eight hours, which, as shown, denies access to most volatile microthrombotic events (Finsterbusch et al., 2018).

Although not a new topic, sepsis has recently gained importance in scientific research. Among others Levi et al. (2013) discussed thromboembolic disease, thrombophilia, and coagulopathy in septic patients, and Assinger et al. (2019) examined the contribution of platelets to sepsis severity and outcome.

2.2. Technical Background

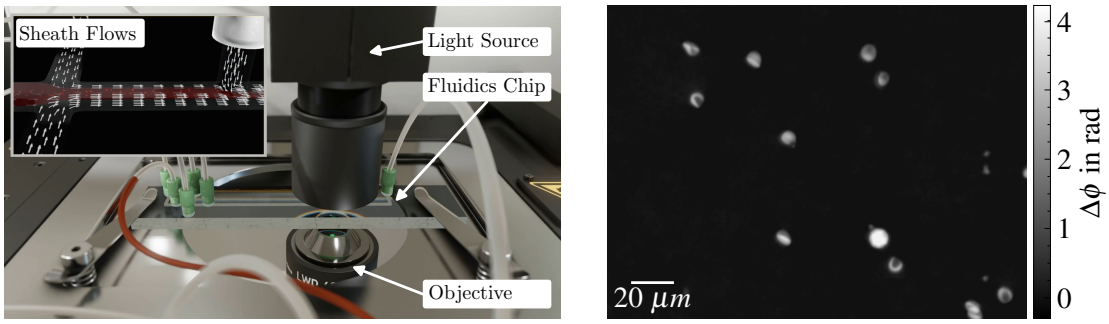
A QPI microscope uses the principle of interference to measure not only the transmission of light, but also its phase shift $\Delta\phi$, and thus to infer the optical density of cellular structures. Recently, QPI has gained relevance through its combination with machine learning, transforming cytometry into a computer vision problem (Jo et al., 2018). As a new platform technology it solves the problem of low contrast associated with typical brightfield microscopy, caused by the transparent nature of most cells. Traditionally, this would require time-consuming sample preparation, staining or genetic fluorescent labeling of cells, which can directly affect cell morphology (Barcia, 2007; Sahoo, 2012; Klenk et al., 2019).

For this project, we utilize an *off-axis diffraction phase microscope* by *Ovizio Imaging Systems* as shown in Figure 1(*a*). In combination with a microfluidics channel, it allows label-free cell imaging of unprocessed blood cells in suspension under near *in-vivo* conditions. A 528 nm *Super-LED Köhler illumination* provides the light source, shining on a 50 µm × 500 µm polymethyl methacrylate (PMMA) microfluidics channel that uses sheath flows to focus the sample stream within the depth of field of the 40× objective. The integrated optical setup then projects the interference patterns onto a camera sensor that captures the cells at 105 frames per second. For more information on the setup used, see Dubois and Yourassowsky (2008) and Ugele et al. (2018b).

The resulting interference patterns (hologram images) contain the intensity and phase information, that can be extracted by a reconstruction algorithm (Schnars and Jüptner, 1994). In this work, we use only phase images, as shown in Figure 1(*b*), because they contain most of the information about the internal structure and morphology of the observed cells.

2.3. Quantitative Phase Imaging and Machine Learning

Machine Learning methods recently entered the field of quantitative phase imaging. Besides their application in the phase reconstruction process itself (Jo et al., 2018; Paine and Fienup, 2018), the strength of these techniques, especially the Convolutional Neural



(*a*) Microscope components (*b*) A raw phase image

Figure 1: Optical setup and the resulting image

Network (CNN), lies in the segmentation and enhancement of the images as well as the differentiation of individual cells. The U-Net (Ronneberger et al., 2015; Zhang et al., 2018; Midtvedt et al., 2021) and the Mask R-CNN (He et al., 2017; Kutscher et al., 2021) show promising results for identifying and segmenting blood and tissue cells. The greatest opportunities for these technologies lie in the combination of the label-free holography and the beyond-human classification power of current neural network architectures. Bacteria (Jo et al., 2015) or leukocytes (Ozaki et al., 2019) are analyzed and classified based on their sub-cellular structures. In oncology, leukemia and the detection of its sub-types can be addressed using dimensionality reduction techniques and morphological features (Ugele et al., 2018b; Paidi et al., 2021). Nevertheless, the automated filtering and classification of cells in a high-throughput scenario like ours remains. As the phase representation of most cell types is unfamiliar to biological and medical experts, the generation of a ground truth needed for supervised learning is laborious, if not impossible (Filby, 2016; Ugele, 2019). Another obstacle is the work with whole blood samples which would provide the most convenient and simple clinical workflow without intensive and time consuming sample preparation. Reaching for rare events demands a reliable chain of filtering and outlier detection techniques, since otherwise feature extraction, dimensionality reduction, neural networks as well as classical discriminators are prone to failure (Röhrl et al., 2019).

3. Methods

In this section, we introduce the algorithms used in different steps of our proposed cell aggregate analysis pipeline. Also the metrics for their evaluation are presented.

3.1. Segmentation

The first step in the analysis of blood cell aggregates is segmentation, which allows for the identification of aggregates and their components. Numerous methods have been developed and evaluated for use in biomedical imaging, as segmentation is an essential practice. In this work, we focus on three methods, one classical segmentation method, namely watershed segmentation, and two machine learning methods, U-NET and Mask R-CNN, all of which have previously been successfully used in biomedical segmentation tasks.

Watershed Watershed segmentation is a classical method based on region growing (Beucher and Lantuéjoul, 1979; Vincent and Soille, 1991). It starts from a seed point and iteratively adds neighboring pixels in a similar way to how water floods a region. Watershed is relatively simple and fast, especially compared to machine learning based segmentation methods, yet provides good segmentation results for biomedical purposes (Ng et al., 2006).

U-NET The U-NET is a type of CNN specifically designed for semantic segmentation of biomedical images (Ronneberger et al., 2015). It is a derivative of the fully convolution network (FCN) (Shelhamer et al., 2017) designed to work with very few training images. The U-NET takes its name from its symmetrical U-shape, which consists of a contracting path and an expanding path. The contracting path uses a typical CNN architecture consisting of convolutional, rectified linear unit (ReLU), and max-pooling layers. Each step of the expansive path uses upsampling and convolution while concatenating higher resolution feature maps from the contractive path with the upsampled features. While the U-NET was originally designed for semantic segmentation only, the use of a *boundary loss* function allows its adoption for instance segmentation. In this work, an *Adam* optimizer (Kingma and Ba, 2015) was combined with a compound loss function of *cube loss* (Wang and Chung, 2018) and *boundary loss* (Kervadec et al., 2019).

Mask R-CNN The Mask R-CNN is designed for instance segmentation (He et al., 2017). It performs both object detection and object mask computation simultaneously. The Mask R-CNN is based on the Faster R-CNN (Ren et al., 2017), a region-based CNN. In this work, a ResNet50 (He et al., 2016) is used as the backbone. For the training process, a *stochastic gradient descent* optimizer with momentum was combined with a compound loss function of *classification loss*, *bounding-box loss*, and *mask loss* as defined by Ren et al. (2017) and He et al. (2017).

3.2. Classification

Unless classification has already been performed during the segmentation, the second step in our pipeline is to classify segmented cell images. This classification task considers three classes of cells relevant to blood analysis, the coagulation system, and possible diseases (erythrocytes, leukocytes, and platelets), as described in Section 2.1.

Gating Gating is a popular method in biology and medicine for manually dividing a set of cells into distinct clusters or populations (Staats et al., 2019). It typically relies on the use of software to apply a set of manually drawn gates that select regions in a 2D graphical representation of the data. This technique is most commonly used to analyze flow cytometry data. The advantages are its simplicity and explainability, since the gates are generally based on expert knowledge of the cell characteristics. This explains its widespread use in biology and medicine. However, gating shows limited suitability for high-dimensional data and is typically based on manual subjective decisions leading to high inter-observer variability (Staats et al., 2019).

Morphological features In order to successfully apply manual gating techniques to images, features must first be extracted from the images. As suggested by a Ugele et al. (2018b) or Paidi et al. (2021), a set of hand-crafted morphological features based on cell

size, shape, and texture is calculated for each individual cell. The outer contour line forms the basis for features describing size and shape. Texture features are computed from the gray-level co-occurrence matrix (GLCM), which represents the distribution of co-occurring pixel values in an image and is commonly used for texture analysis in image processing (Haralick et al., 1973). These features are highly intuitive and explainable, providing excellent interpretability for expert gating. To classify the elements of cell aggregates, we only use features that are robust to changes in the shape and contour of the cell (like *homogeneity* or *optical height*), since others can experience shifting due to aggregation.

Random forest Random forest is an ensemble classification method based on decision trees proposed by Breiman (2001). A decision tree is a machine learning model that combines a series of decisions based on variable values. For random forest classification, a large number of decision trees are automatically constructed based on different fractions of the given data set. For classification of unknown samples, the average result of all trained decision trees is used. This concept reduces overfitting very effectively and works well for more complex classification tasks in high-dimensional feature spaces. Nguyen et al. (2017) successfully used a combination of morphological features and random forest classification for the grading of prostate cancer.

3.3. Evaluation Metrics

To assess the segmentation and classification quality of the proposed methods, evaluation metrics are needed. For this purpose, the Intersection over Union (IoU) is used, as usual for the evaluation of segmentation and object detection methods, and combined with the metrics precision, recall, and F_1-score.

Segmentation performance The **IoU**, or Jaccard index (Jaccard, 1912), is a popular evaluation metric used in instance segmentation and object detection. It is a measure of the similarity between two shapes, in the case of instance segmentation, the predicted region $\hat{A}$ and the ground-truth A

$$\text{IoU} = \frac{\text{area of overlap}}{\text{area of union}} = \frac{A \cap \hat{A}}{A \cup \hat{A}} \in [0, 1]\,. \tag{1}$$

The IoU is invariant to scale and therefore a very powerful metric for the evaluation of segmentation algorithms.

Detection and classification performance Since the IoU alone is only partially useful for evaluating a real-world application, a minimum IoU threshold is typically defined for an instance (or object) to be considered as correctly recognized, and evaluation metrics such as precision or recall are used. In this work, we use an IoU threshold of 0.4, since overly detailed localization and masking of the cells is not necessarily needed, while correct detection of the cell amounts and types is more important. We use the definitions given by Powers (2011) for **precision** and **recall**

$$\text{Precision} = \frac{T_p}{T_p + F_p}, \quad \text{Recall} = \frac{T_p}{T_p + F_n} \qquad \text{with} \quad \begin{array}{l} T_p : \text{true positives} \\ F_p : \text{false positives} \\ F_n : \text{false negatives} \end{array} \tag{2}$$

and the resulting harmonic mean as the so called **F_1-Score**

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \,. \tag{3}$$

Aggregate analysis To better evaluate the segmentation performance for the specific task of detecting and counting single cells and cell aggregates, two custom metrics are used: The **aggregate composition** score evaluates whether all parts of the analyzed aggregate are correctly detected quantitatively as defined by

$$\text{AC} = \frac{1}{K}\sum_{i=1}^{K} ac_i \quad \text{with} \quad ac_i = \begin{cases} 1 & \text{if } \hat{n}_{class} = n_{class}\ \forall class \in \{\text{ery}, \text{leuko}, \text{thrombo}\} \\ 0 & \text{else} \end{cases} \tag{4}$$

where n_{class} is the number of elements of class *class* in the image patch and K is the number of images patches in the dataset. The **event type** score

$$\text{ET} = \frac{1}{K}\sum_{i=1}^{K} et_i \quad \text{with} \quad et_i = \begin{cases} 1 & \text{if } \hat{t}_i = t_i \\ 0 & \text{else} \end{cases} \tag{5}$$

and t as the type of aggregate or single cell $t \in \{$*single erythrocyte*, *multiple erythrocytes*, *single platelet*, *platelet aggregate*, *single leukocyte*, *leukocyte-platelet aggregate*$\}$ assesses if the type of aggregate or cell is correctly detected (qualitatively).

Regression analysis To evaluate possible correlations in our experiments concerning mixing ratios or activation, we employ the following models and methods: For **linear relations** we use a simple linear mapping of an independent variable x on a dependent variable y

$$\hat{y} = ax + b \qquad \text{minimizing} \quad L(x) = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{6}$$

where y is the real world observation and $\hat{y}$ is the prediction of the model. For **nonlinear relations** we chose an exponential model

$$\hat{y} = a \log(x) + b \tag{7}$$

which can be fitted by an iterative estimation algorithm. Therefore, we use the *Levenberg–Marquardt algorithm* (Levenberg, 1944). To evaluate the fit of the regression models we apply the Normalized Root-Mean-Square Error (NRMSE)

$$\text{NRMSE} = \frac{\sqrt{\text{MSE}}}{y_{max} - y_{min}} \qquad \text{with} \quad \text{MSE} = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{8}$$

where N is the number of observations, which allows us to compare data and models of different scales (James et al., 2013). The coefficient of determination

$$R^2 = \frac{\text{Var}(\hat{y})}{\text{Var}(y)} = 1 - \frac{\text{Var}(e)}{\text{Var}(y)} \qquad \text{with} \quad e = y - \hat{y} \tag{9}$$

provides us with a measure of the quality for the respective fit, by comparing the variance of the observed data $\text{Var}(y)$ to the variance explained by the model $\text{Var}(\hat{y})$ (Devore, 2015).

3.4. Experimental Setup

In order to achieve a quantitative evaluation of aggregates and their components and to enable pathological analysis of clinical samples, four processing pipelines have been developed. The first two approaches represent rather simple computer vision-based methods, consisting of a combination of watershed segmentation and different classification algorithms, namely expert gating and random forest. They require four processing steps: segmentation, feature extraction, classification, and a final analysis of the results. The third approach uses a U-NET and the fourth approach is based on a Mask R-CNN. These two approaches require only two processing steps because both U-NET and Mask R-CNN are capable of both segmentation and classification.

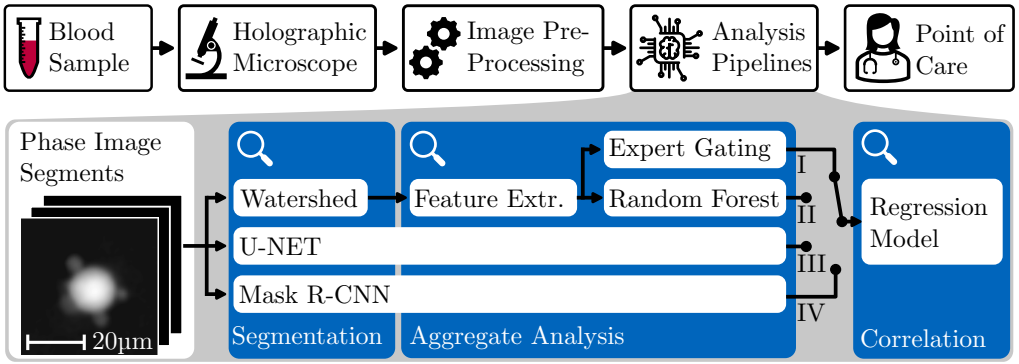


Figure 2: While the overall workflow is kept close to existing assays, we investigate the performances of the aggregate analysis pipelines I-IV.

4. Data Set

For our experiments and trainings, we use several data sets to better assess the strengths and weaknesses of the evaluated approaches and to gradually increase the level of difficulty.

4.1. BBBC038

To obtain a baseline evaluation of the segmentation methods, the publicly available BBBC038 data set (Caicedo et al., 2019) is used. It contains a variety of two-dimensional light microscopy images of stained nuclei as displayed in Figure 5 in the Appendix. For this experiment, the BBBC038 data set is divided into three parts, a training, test, and validation set (60:20:20). The training set is used to train both U-NET and Mask R-CNN, the validation set is used for hyperparameter tuning, and all three methods are evaluated on the test set.

4.2. Expert Labeled Data Set

For our particular use case we need a more accurate assessment of the segmentation. Therefore, we labeled a data set of 100 images of blood cells captured by our QPI microscope. It consists of 50% single and multiple erythrocytes and 50% platelets and leukocyte-platelet aggregates. The images were manually masked by biomedical personnel using a brush tool, resulting in a very accurate segmentation. Respective examples are shown in Figure 7 in the Appendix.

4.3. Synthetic Aggregate Data Set

The performance and reliability of a neural network is highly dependent on the quality, quantity, and selection of the training data. Acquiring ground-truth data using an unsupervised or self-supervised method, as is sometimes done when no ground-truth is available, does not solve this problem, because the trained network will never be able to outperform the quality of the training data. Alternatively, especially for the Mask R-CNN, training with only single cell images would produce adequate results, but the shape and contour of the cell images and masks will change slightly when the cells are part of aggregates, which will degrade the performance of the network, as analyzed in section 5.1. As labeling by experts is costly and time consuming we chose to generate a synthetic data set like Prastawa et al. (2005) or Gupta et al. (2016) by stitching together multiple single cell images to form cell aggregates. Its generation procedure and example images can be found in Appendix C.1.

4.4. Clinical Samples

Activated platelets For this data set, platelets are extracted from whole blood using two centrifugation steps to first extract platelet-rich plasma and then concentrate it to a pellet, which is then resuspended in a buffer solution (Bernlochner et al., 2021). After extraction, the platelets are artificially activated with the platelet activator thrombin receptor activating peptide (TRAP). Activation causes the platelets to form volatile aggregates that disintegrate over time. As shown experimentally by Michelson et al. (2001), based on measurements of platelet surface P-selectin and the occurrence of monocyte-platelet and neutrophil-platelet aggregates in whole blood, a peak of aggregation is expected after a few minutes, followed by a decline until normal levels are restored 60-120 minutes after the addition of TRAP (Michelson et al., 2001).

Five series of measurements are performed, to assess platelet aggregation levels at 7.5, 15, 30, 60, 90, and 120 minutes after application of 10 µM TRAP. Three measurements are taken at each time step, each containing approximately 5,000 platelets.

Activated platelets spiked in whole blood Aiming for data more closely related to whole blood, and to test the robust detection of aggregates as a tiny minority of events in the sample stream, we created another data set. As before, pure platelet samples are extracted from whole blood and then activated with TRAP. The activated platelets are then mixed with whole blood samples at various mixing ratios [0%, 10%, 30%, 50%, 70%, 90%, 100%]. Three samples, each containing approximately 40,000 cells, are measured for each mixing ratio. These mixing ratios should be clearly observable and the amount of aggregates detected should be dependent on the amount of activated platelets added.

Activated whole blood samples To match the conditions of Michelson et al. (2001), this data set consists of whole blood samples collected and activated by either adenosine diphosphate (ADP) or TRAP. While TRAP is a synthetic peptide, ADP is a nucleotide that binds to three specific platelet membrane receptors, triggering platelet aggregation and shape change (Murugappa and Kunapuli, 2006). For comparison, three types of samples were analyzed: untreated whole blood, whole blood activated by adding 10 µM ADP, and

whole blood activated by adding 10 µM TRAP. Each sample was captured six times, with each capture containing approximately 40,000 cells.

Healthy reference Reference samples are collected from seven healthy donors, both male and female, between the ages of 29 and 67, with no history of disease. Blood samples are diluted 1:100 in a measurement buffer consisting of phosphate buffered saline (PBS) and polyethylene oxide analogous to Klenk et al. (2023) and analyzed immediately. From each sample, we record three measurements of approximately 40,000 cells each.

Sepsis For an immunothrombotic disease associated with blood cell aggregation, we collect samples from seven Intensive Care Unit (ICU) patients diagnosed with sepsis, both male and female, between the ages of 45 and 80, at multi-day intervals, typically three samples per patient. The time between blood draw and sample analysis is less than 30 minutes, which is critical for accurate assessment of aggregation. Samples are carefully transported to the nearby prototype to ensure minimal mechanical disturbance. Three measurements of 7,500 images each are taken from each sample for analysis. A single measurement typically contains approximately 30,000 cells.

COVID-19 For COVID-19, we collect samples from thirteen ICU patients (both male and female) diagnosed with PCR-confirmed wild-type SARS-CoV-2 infection between the ages of 51 and 91 at multi-day intervals, typically five samples per patient. As before, less than 30 minutes elapse between blood collection and sample analysis. From each sample, three measurements of 7,500 images each are recorded for analysis, with each measurement typically containing approximately 30,000 cells.

5. Results

The experimental results are organized in three sections. The first two sections will preselect the most appropriate aggregate analysis pipeline, which is then used to process the clinical samples in the last section. Therefore, we employ the proposed measurement points marked in blue in Figure 2 and the corresponding evaluation metrics from Section 3.3.

5.1. Segmentation

BBBC038 data set Testing the algorithms on the BBBC038 data set provides a first trend for their segmentation performance. Although this data set does not contain blood cell aggregates, it is very diverse and challenging due to its complex structures. Our observations are printed in Table 1(*a*). While watershed segmentation achieves only mediocre results, U-NET and Mask R-CNN achieve reasonably good results, with the Mask R-CNN showing the best overall recall and precision. Figure 8 in the Appendix shows the according visualizations.

Expert labeled data set To see if these trends continue, we switch to the expert labeled data set, which is closer related to our real-world applications. Table 1(*b*) demonstrates again the superiority of the Mask R-CNN (see also Appendix Figure 9). However, the watershed algorithm achieves very close results. Interestingly, when comparing the performance of the differently trained Mask R-CNNs, a Mask R-CNN trained on the BBBC038 data set already shows quite good results. As expected, segmentation using the synthetic

data set for training outperforms training using only single cell images due to the changing morphology of cells that are part of aggregates. For this application, the U-NET achieves the worst results, which is most likely due to the fact that it is not perfectly suited for instance segmentation, especially of small, slightly overlapping cells.

Table 1: Segmentation quality of used methods on the test data

(a) BBBC038 data set

	∅IoU	Recall	Precision	F_1
Watershed	0.594	0.747	0.641	0.690
U-NET	0.584	0.760	0.831	0.794
Mask R-CNN	**0.758**	**0.909**	**0.861**	**0.884**

[1]trained on BBBC038, [2]trained on single cell images, [3]trained on synthetic aggregate data set

(b) Expert labeled data set

	∅IoU	Recall	Precision	F_1
Watershed	0.727	0.928	0.931	0.930
U-NET[3]	0.619	0.801	0.914	0.854
Mask R-CNN[1]	0.716	0.878	0.821	0.849
Mask R-CNN[2]	0.583	0.831	0.852	0.841
Mask R-CNN[3]	**0.741**	**0.931**	**0.956**	**0.943**

5.2. Classification

As before, the expert labeled data set of 100 images of blood cells, here including the cell labels, is used to evaluate the performance of aggregate detection combining segmentation and classification. Again, the segmentation quality is evaluated based on the metrics described in Section 3.3, adding the correctness of the predicted classes as a requirement for accepted detected instances. The evaluation results are shown in Table 2. Similar to the previous results, Mask R-CNN shows the best performance with a slight decrease in both precision and recall due to classification inaccuracies. For watershed-based methods, random forest classification shows significantly better results than expert gating. The U-NET achieves slightly better results, but worse than the Mask R-CNN. This also shows in the scores for aggregate composition (AC) and event type (ET). The Mask R-CNN qualifies as an excellent aggregate detector having an ET score of 0.970. The U-Net is also quite suitable, while the watershed-based pipelines are too coarse to detect all aggregates or mix up the contained classes. Therefore, we will use the Mask R-CNN for the following experiments.

Table 2: Segmentation and classification performance on the expert labeled dataset

	∅IoU	Recall	Precision	F_1	AC	ET
Watershed + Gating	0.539	0.672	0.658	0.665	0.510	0.650
Watershed + Random Forest	0.630	0.790	0.773	0.782	0.580	0.730
U-NET	0.596	0.810	0.826	0.818	0.660	0.930
Mask R-CNN	**0.676**	**0.917**	**0.912**	**0.915**	**0.780**	**0.970**

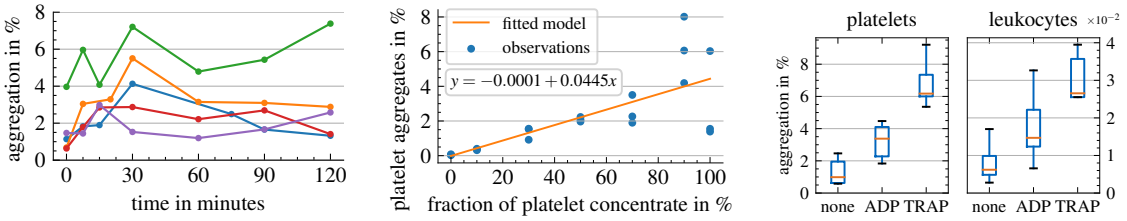
5.3. Clinical Samples

Activated platelets Analysis of the five time-series measurements of activated platelets using our Mask R-CNN pipeline results in slightly different curves, as drawn in 3(a). However, there is a clear trend that shows a sharp increase in platelet aggregation from the beginning, peaking between 15 and 30 minutes. Thereafter, the aggregates begin to break down. These observations are roughly in line with expectations based on previous research. Only the activation seems to be a bit slower, reaching a maximum activation of only 3%-7%. This is probably caused by the fact that these experiments were performed with extracted platelets as opposed to whole blood in the case of Michelson et al. (2001). Platelet activation

and aggregation is a complex process based on the coagulation cascade, and platelet-only samples lack many of the coagulants that normally promote platelet aggregation in whole blood.

Activated platelets spiked in whole blood To evaluate the detected platelet concentrations in the samples with different mixing ratios, we use linear regression. The fitted model $y = 0.0288 + 0.932x$ gives an almost perfect fit with a high coefficient of determination of $R^2 = 0.997$ and a NRMSE $= 0.018$, as shown in Figure 10 in the Appendix. These observations fit the expectation, as evidenced by the intercept, suggesting about 3% platelets in the whole blood sample, which is reasonable since it is in the typical range of 2.5%-8% (Bain, 2017). For the aggregation analysis, we also use linear regression. The fitted model can be seen in Figure 3(*b*) and features $R^2 = 0.552$ and NRMSE $= 0.18$.

Activated whole blood samples Analysis of the levels of platelet aggregates detected, as depicted in Figure 3(*c*), shows a clearly visible effect. In untreated whole blood samples almost no aggregates are observed, whereas in ADP activated samples 2 % to 4 % and TRAP activated samples 5.5 % to 8 % platelet aggregates are detected. Similarly, almost no leukocyte-platelet aggregates are observed in untreated blood samples. ADP activated samples showed 0.15 % to 0.25 % and TRAP activated samples 0.25 % to 0.35 % leukocyte-platelet aggregates. This difference between ADP- and TRAP-induced platelet aggregation is consistent with previous research by Olivier et al. (2016), where application of TRAP showed approximately 2.5 times higher aggregation than application of ADP.



(*a*) Activation time series (*b*) Aggregates in whole blood (*c*) Activated whole blood

Figure 3: Detection of platelet aggregation induced by ADP or TRAP activation

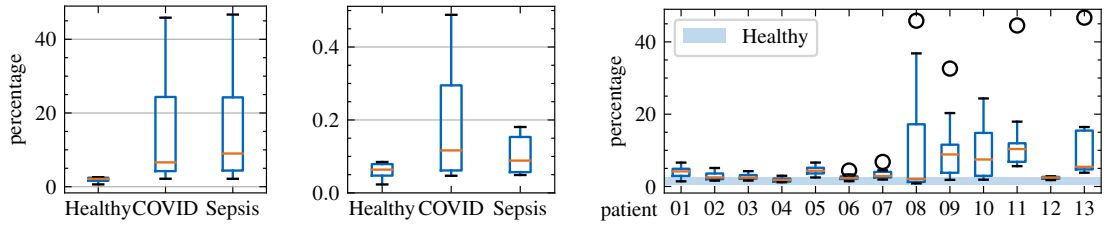
Healthy reference For the reference samples, platelet aggregates are in the range 0.5 % to 2.5 % and leukocyte-platelet aggregates are in the range 0.01 % to 0.08 %, as shown in Figure 11 in the Appendix. These results are generally consistent with previous studies by Leytin et al. (2000) and Gerrits et al. (2016), which reported (1.02 ± 0.49) % and 0.001 % to 0.03 % respectively. For platelet aggregates, mostly 2-cell aggregates are observed and very few 3-cell or 4-cell aggregates, similar to the previous activation experiments. Almost all leukocyte-platelet aggregates contain only one leukocyte and mostly one (or sometimes two) platelets. This behavior is in line with expectations, as only minimal and very small aggregates are expected in healthy whole blood samples, since larger aggregates already pose a significant health risk, as described in Section 2.1.

Sepsis In our exemplary sepsis cohort, we observe both elevated platelet aggregation levels, as shown in Figure 4(*a*), and elevated leukocyte-platelet aggregation levels, as shown

in Figure 4(*b*). Three out of seven patients (01, 02, and 04) have severely elevated levels of **platelet aggregates**, while only one is completely within the healthy reference range, as shown in Figure 12(*a*) in the Appendix. For **leukocyte-platelet aggregates**, all but one patient feature increased levels at least once, as plotted in Figure 12(*b*) in the Appendix. In addition, patients with higher aggregate levels also show a shift in aggregate size distribution with comparatively more larger aggregates, as shown in Appendix G.

COVID-19 For the COVID-19 patients, we record a similar picture with both elevated platelet aggregation levels and elevated leukocyte platelet aggregation levels, as shown in Figure 4(*a*) and 4(*b*). The effect is even more remarkable as 12 out of 13 patients have increased levels of **platelet aggregates** compared to the healthy donors. On closer inspection of Figure 4(*c*), some patients show extremely elevated levels, specifically patients 08, 09, 10, 11, and 13. Consistent with these observations, three of these four patients had a collapse of their clinical condition during observation (e.g., lung failure). In contrast, patient 04, who was transferred to the general ward during observation, shows very low platelet aggregation levels that remain within the reference range.

Looking at **leukocyte-platelet aggregates**, 11 of the 13 patients demonstrate elevated levels at least once, as shown in Figure 13(*b*) in the Appendix. For more severe cases (especially patients 08-11 and 13) this effect is clearly an indicator, but in milder cases leukocyte-platelet aggregate levels do not show a substantial shift. Similar to the sepsis experiments, samples with higher levels of aggregation also show comparatively more larger aggregates, both platelet aggregates and leukocyte-platelet aggregates, as shown in Appendix G.



(*a*) P aggregates (*b*) LP aggregates (*c*) P aggregates in COVID-19 patients

Figure 4: Platelet (P) and leukocyte-platelet (LP) aggregates for COVID-19 and sepsis

6. Discussion and Conclusion

In this work, we present a novel processing pipeline for the detection and quantitative analysis of blood cell aggregates and their components. Using QPI, this approach allows the assessment of platelet aggregation and microthrombus formation in label-free whole blood samples without the need for sample preparation. Specific detection of each component of an aggregate allows evaluation of the size and number of platelet and leukocyte-platelet aggregates.

In various test scenarios, we compared four different approaches using established and custom metrics at different stages of the pipelines. The first approach is a non-machine

learning method consisting of a combination of watershed segmentation and two different classification methods. Watershed segmentation showed great potential for segmenting blood cell aggregates, but the evaluated classification method was not able to reliably distinguish leukocytes from erythrocytes. The extension with a data-driven random forest classifier as a second approach did not lead to the desired improvements. The third approach uses a U-NET, which we adapted for instance segmentation by using a boundary loss function, which showed decent results. The fourth approach is based on a Mask R-CNN trained on an artificially created synthetic aggregate data set. This approach showed the best results, with a precision of 0.956 and a recall of 0.931 on an expert-labeled test set, and most importantly, it yielded the correct category of cell or cell aggregate in 97% of the cases.

The Mask R-CNN processing pipeline was then evaluated on defined medical samples comprising activated platelets, activated platelets spiked in whole blood, and activated whole blood. These experiments demonstrated very reliable detection of platelet aggregates, but showed some limitations for leukocyte-platelet aggregates due to low statistical power.

Finally, we evaluated the quality of this method as a diagnostic predictive biomarker for immunothrombotic diseases by analyzing samples from patients with COVID-19 and sepsis. In both diseases, 90% of patients had aggregate levels above the healthy reference interval, with all severe patients having substantially higher aggregate levels (5-10 times higher than reference samples). In addition, these samples with particularly high aggregate levels also had consistently higher amounts of larger aggregates. In conclusion, the analysis of these clinical samples demonstrated the effectiveness of the proposed method and the potential of using the occurrence of platelet and leukocyte-platelet aggregates as biomarkers for the presence and severity of immunothrombotic diseases.

Limitations Due to the difficulty in obtaining blood samples from COVID-19 and sepsis patients during the pandemic, only a limited number of clinical patients were analyzed in this work, which does not allow a concrete diagnostic and therapeutic assessment of the occurrence of aggregates in the studied diseases. For a higher statistical power and a more precise assertion of the severity of both sepsis and COVID-19 (Rampotas and Pavord, 2021), a larger clinical study needs to be performed. We also need to collect more medical parameters as described by Poudel et al. (2021) or Gorog et al. (2022) to correlate our new biomarker with established biomarkers to prove advantages or discrepancies. In addition, we observed a significant decrease in aggregate levels between multiple acquisitions of the same sample, demonstrating the short lifespan of blood cell aggregates and confirming the need for immediate sample analysis, ideally in a point-of-care environment. The choice of anticoagulant also plays an important role and needs to be evaluated in future studies (Klenk et al., 2023). Finally, the new methodology needs to be proven in clinical applications before this platform technology can add value at the point of care. This is not least due to the acceptance of black box models that still needs to be built up, which can be achieved through long-term successful use. However, we hope that our methodology will open the way for further applications that can benefit from the detailed analysis of immunothrombosis (Engelmann and Massberg, 2013; Wendelboe and Raskob, 2016; Nicolai et al., 2020; Stark and Massberg, 2021).

References

Nicole Allen, Tessa J. Barrett, Yu Guo, Michael Nardi, Bhama Ramkhelawon, Caron B. Rockman, Judith S. Hochman, and Jeffrey S. Berger. Circulating monocyte-platelet aggregates are a robust marker of platelet activity in cardiovascular disease. *Atherosclerosis*, 282:11–18, 2019.

Alice Assinger, Waltraud C. Schrottmaier, Manuel Salzmann, and Julie Rayes. Platelets in sepsis: An update on experimental models and clinical data. *Frontiers in Immunology*, 10:1687, 2019.

Barbara J. Bain. *A Beginner's Guide to Blood Cells*. John Wiley & Sons Ltd, 3rd edition, 2017.

Juan José Barcia. The giemsa stain: Its history and applications. *International Journal of Surgical Pathology*, 15(3):292–296, 2007.

Isabell Bernlochner, Melissa Klug, Ditya Larasati, Moritz Von Scheidt, Donato Santovito, Michael Hristov, Christian Weber, Karl-Ludwig Laugwitz, and Dario Bongiovanni. Sorting and magnetic-based isolation of reticulated platelets from peripheral blood. *Platelets*, 32(1):113–119, 2021.

Serge Beucher and Christian Lantuéjoul. Use of watersheds in contour detection. In *International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation*, volume 132, 1979.

Deepak L. Bhatt and Eric J. Topol. Scientific and therapeutic advances in antiplatelet therapy. *Nature Reviews Drug Discovery*, 2(1):15–28, 2003.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

Wilhelm Burger and Mark J. Burge. *Morphological Filters*, pages 181–208. Springer London, London, 2016.

Juan C. Caicedo, Allen Goodman, Kyle W. Karhohs, Beth A. Cimini, Jeanelle Ackerman, Marzieh Haghighi, CherKeng Heng, Tim Becker, Minh Doan, Claire McQuin, Mohammad Rohban, Shantanu Singh, and Anne E. Carpenter. Nucleus segmentation across imaging experiments: the 2018 data science bowl. *Nature Methods*, 16(12):1247–1253, 2019.

Jay L. Devore. *Probability and Statistics for Engineering and the Sciences*. Cengage Learning, 2015.

Frank Dubois and Catherine Yourassowsky. US patent 7,362.449 b2, 2008.

Bernd Engelmann and Steffen Massberg. Thrombosis as an intravascular effector of innate immunity. *Nature Reviews. Immunology*, 13(1):34–45, 2013.

Andrew Filby. Sample preparation for flow cytometry benefits from some lateral thinking. *Cytometry Part A*, 89(12):1054–1056, 2016.

Michaela Finsterbusch, Waltraud C. Schrottmaier, Julia B. Kral-Pointner, Manuel Salzmann, and Alice Assinger. Measuring and interpreting platelet-leukocyte aggregates. *Platelets*, 29(7):677–685, 2018.

Mark I. Furman, Marc R. Barnard, Lori A. Krueger, Marsha L. Fox, Elizabeth A. Shilale, Darleen M. Lessard, Peter Marchese, A.L Frelinger, Robert J. Goldberg, and Alan D. Michelson. Circulating monocyte-platelet aggregates are an early marker of acute myocardial infarction. *Journal of the American College of Cardiology*, 38(4):1002–1006, 2001.

Anja J. Gerrits, Andrew L. Frelinger, and Alan D. Michelson. Whole blood analysis of leukocyte-platelet aggregates. *Current Protocols in Cytometry*, 78:6.15.1–6.15.10, 2016.

Diana A Gorog, Robert F Storey, Paul A Gurbel, Udaya S Tantry, Jeffrey S Berger, Mark Y Chan, Daniel Duerschmied, Susan S Smyth, William A E Parker, Ramzi A Ajjan, Gemma Vilahur, Lina Badimon, Jurrien M Ten Berg, Hugo Ten Cate, Flora Peyvandi, Taia T Wang, and Richard C Becker. Current and novel biomarkers of thrombotic risk in covid-19: a consensus statement from the international covid-19 thrombosis biomarkers colloquium. *Nature reviews. Cardiology*, 19(7):475—495, 2022.

Ankush Gupta, Andrea Vedaldi, and Andrew Zisserman. Synthetic data for text localisation in natural images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2315–2324, 2016.

Robert M. Haralick, Its'hak Dinstein, and K. Shanmugam. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(6):610–621, 1973.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2961–2969, 2017.

Richard S. Hotchkiss, Lyle L. Moldawer, Steven M. Opal, Konrad Reinhart, Isaiah R. Turnbull, and Jean-Louis Vincent. Sepsis and septic shock. *Nature reviews. Disease primers*, 2:16045, 2016.

Paul Jaccard. The distribution of the flora in the alpine zone. *New Phytologist*, 11(2):37–50, 1912.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. Springer-Verlag, 2013.

Young Ju Jo, Jae Hwang Jung, Min hyeok Kim, Hyun Joo Park, Suk-Jo Kang, and Yong Keun Park. Label-free identification of individual bacteria using fourier transform light scattering. *Optics Express*, 23(12):15792–15805, 2015.

Young Ju Jo, Hyungjoo Cho, Sang Yun Lee, Gunho Choi, Geon Kim, Hyun Seok Min, and Yong Keun Park. Quantitative phase imaging and artificial intelligence: A review. *IEEE Journal of Selected Topics in Quantum Electronics*, 25(1):1–14, 2018.

Hoel Kervadec, Jihene Bouchtiba, Christian Desrosiers, Eric Granger, Jose Dolz, and Ismail Ben Ayed. Boundary loss for highly unbalanced segmentation. In *International Conference on Medical Imaging with Deep Learning*, pages 285–296. PMLR, 2019.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

Christian Klenk, Dominik Heim, Matthias Ugele, and Oliver Hayden. Impact of sample preparation on holographic imaging of leukocytes. *Optical Engineering*, 59(10):102403, 2019.

Christian Klenk, David Fresacher, Stefan Röhrl, Dominik Heim, Manuel Lengl, Simon Schumann, Martin Knopp, Klaus Diepold, Stefan Holdenrieder, and Oliver Hayden. Measurement of platelet aggregation in ageing samples and after in-vitro activation. In *Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 2: BIOIMAGING*, pages 57–65. INSTICC, SciTePress, 2023.

Sotiris B. Kotsiantis, Dimitris Kanellopoulos, and Panagiotis E. Pintelas. Data preprocessing for supervised leaning. *International Journal of Computer and Information Engineering*, 1(12):4104–4109, 2007.

Tobias Kutscher, Kai Eder, Anne Marzi, Álvaro Barroso, Jürgen Schnekenburger, and Björn Kemper. Cell Detection and Segmentation in Quantitative Digital Holographic Phase Contrast Images Utilizing a Mask Region-based Convolutional Neural Network. In *OSA Optical Sensors and Sensing Congress*, page JTu5A.23. Optica Publishing Group, 2021.

Dovena Lazaridis, Simon Leung, Lisa Kohler, Carla Hawkins Smith, Margaretta L Kearson, and Nathaniel Eraikhuemen. The impact of anticoagulation on covid-19 (sars cov-2) patient outcomes: a systematic review. *Journal of Pharmacy Practice*, 35(6):1000–1006, 2022.

Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164–168, 1944.

Marcel Levi, Marcus Schultz, and Tom van der Poll. Sepsis and thrombosis. *Seminars in Thrombosis and Hemostasis*, 39(5):559–566, 2013.

Valery Leytin, Meera Mody, John W Semple, Bernadette Garvey, and John Freedman. Flow cytometric parameters for characterizing platelet activation by measuring p-selectin (CD62) expression: theoretical consideration and evaluation in thrombin-treated platelet populations. *Biochemical and Biophysical Research Communications*, 269(1):85–90, 2000.

Alan D. Michelson, Marc R. Barnard, Lori A. Krueger, C. Robert Valeri, and Mark I. Furman. Circulating monocyte-platelet aggregates are a more sensitive marker of in vivo platelet activation than platelet surface p-selectin. *Circulation*, 104(13):1533–1537, 2001.

Benjamin Midtvedt, Saga Helgadottir, Aykut Argun, Jesús Pineda, Daniel Midtvedt, and Giovanni Volpe. Quantitative digital microscopy with deep learning. *Applied Physics Reviews*, 8(1):011310, 2021.

Swaminathan Murugappa and Satya P. Kunapuli. The role of ADP receptors in platelet function. *Frontiers in Bioscience: A Journal and Virtual Library*, 11:1977–1986, 2006.

H.P. Ng, S.H. Ong, K.W.C. Foong, P.S. Goh, and W.L. Nowinski. Medical image segmentation using k-means clustering and improved watershed algorithm. In *2006 IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 61–65, 2006.

Tan Huu Nguyen, Shamira Sridharan, Virgilia Macias, Andre Kajdacsy-Balla, Jonathan Melamed, Minh N. Do, and Gabriel Popescu. Automatic Gleason grading of prostate cancer using quantitative phase imaging and machine learning. *Journal of Biomedical Optics*, 22(3):036015, 2017.

Thang L. Nguyen, Soorya Pradeep, Robert L. Judson-Torres, Jason Reed, Michael A. Teitell, and Thomas A. Zangle. Quantitative phase imaging: Recent advances and expanding potential in biomedicine. *American Chemical Society Nano*, 16(8):11516–11544, 2022.

Leo Nicolai, Alexander Leunig, Sophia Brambs, Rainer Kaiser, Tobias Weinberger, Michael Weigand, Maximilian Muenchhoff, Johannes C. Hellmuth, Stephan Ledderose, Heiko Schulz, Clemens Scherer, Martina Rudelius, Michael Zoller, Dominik Höchter, Oliver Keppler, Daniel Teupser, Bernhard Zwißler, Michael von Bergwelt-Baildon, Stefan Kääb, Steffen Massberg, Kami Pekayvaz, and Konstantin Stark. Immunothrombotic dysregulation in COVID-19 pneumonia is associated with respiratory failure and coagulopathy. *Circulation*, 142(12):1176–1189, 2020.

Masako Nishikawa, Hiroshi Kanno, Yuqi Zhou, Ting-Hui Xiao, Takuma Suzuki, Yuma Ibayashi, Jeffrey Harmon, Shigekazu Takizawa, Kotaro Hiramatsu, Nao Nitta, et al. Massive image-based single-cell profiling reveals high levels of circulating platelet aggregates in patients with covid-19. *Nature Communications*, 12(1):1–12, 2021.

Christoph B. Olivier, Melanie Meyer, Hans Bauer, Katharina Schnabel, Patrick Weik, Qian Zhou, Christoph Bode, Martin Moser, and Philipp Diehl. The ratio of ADP- to TRAP-induced platelet aggregation quantifies p2y12-dependent platelet inhibition independently of the platelet count. *PloS One*, 11(2):e0149053, 2016.

Yusuke Ozaki, Hidenao Yamada, Hirotoshi Kikuchi, Amane Hirotsu, Tomohiro Murakami, Tomohiro Matsumoto, Toshiki Kawabata, Yoshihiro Hiramatsu, Kinji Kamiya, Toyohiko Yamauchi, Kentaro Goto, Yukio Ueda, Shigetoshi Okazaki, Masatoshi Kitagawa, Hiroya Takeuchi, and Hiroyuki Konno. Label-free classification of cells based on supervised machine learning of subcellular structures. *PLOS ONE*, 14:1–20, 2019.

Santosh Kumar Paidi, Piyush Raj, Rosalie Bordett, Chi Zhang, Sukrut H. Karandikar, Rishikesh Pandey, and Ishan Barman. Raman and quantitative phase imaging allow morpho-molecular recognition of malignancy and stages of B-cell acute lymphoblastic leukemia. *Biosensors and Bioelectronics*, 190:113403, 2021.

Scott W. Paine and James R. Fienup. Machine learning for improved image-based wavefront sensing. *Optics Letters*, 43(6):1235–1238, 2018.

Ayusha Poudel, Yashasa Poudel, Anurag Adhikari, Barun Babu Aryal, Debika Dangol, Tamanna Bajracharya, Anil Maharjan, and Rakshya Gautam. D-dimer as a biomarker for assessment of covid-19 prognosis: D-dimer levels on admission and its role in predicting disease outcome in hospitalized patients with covid-19. *PloS ONE*, 16(8):e0256744, 2021.

David M. W. Powers. Evaluation: from precision, recall and f-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1):37–63, 2011.

Marcel Prastawa, Elizabeth Bullitt, and Guido Gerig. Synthetic ground truth for validation of brain tumor MRI segmentation. In James S. Duncan and Guido Gerig, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005*, Lecture Notes in Computer Science, pages 26–33. Springer, 2005.

Alexandros Rampotas and Sue Pavord. Platelet aggregates, a marker of severe covid-19 disease. *Journal of Clinical Pathology*, 74(11):750–751, 2021.

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, 2017.

Miguel Reyes, Michael R. Filbin, Roby P. Bhattacharyya, Kianna Billman, Thomas Eisenhaure, Deborah T. Hung, Bruce D. Levy, Rebecca M. Baron, Paul C. Blainey, Marcia B. Goldberg, and Nir Hacohen. An immune-cell signature of bacterial sepsis. *Nature Medicine*, 26(3):333–340, 2020.

Andrew Rhodes, Laura E. Evans, Waleed Alhazzani, Mitchell M. Levy, Massimo Antonelli, Ricard Ferrer, Anand Kumar, Jonathan E. Sevransky, Charles L. Sprung, Mark E. Nunnally, Bram Rochwerg, Gordon D. Rubenfeld, Derek C. Angus, Djillali Annane, Richard J. Beale, Geoffrey J. Bellinghan, Gordon R. Bernard, Jean-Daniel Chiche, Craig Coopersmith, Daniel P. De Backer, Craig J. French, Seitaro Fujishima, Herwig Gerlach, Jorge Luis Hidalgo, Steven M. Hollenberg, Alan E. Jones, Dilip R. Karnad, Ruth M. Kleinpell, Younsuk Koh, Thiago Costa Lisboa, Flavia R. Machado, John J. Marini, John C. Marshall, John E. Mazuski, Lauralyn A. McIntyre, Anthony S. McLean, Sangeeta Mehta, Rui P. Moreno, John Myburgh, Paolo Navalesi, Osamu Nishida, Tiffany M. Osborn, Anders Perner, Colleen M. Plunkett, Marco Ranieri, Christa A. Schorr, Maureen A. Seckel, Christopher W. Seymour, Lisa Shieh, Khalid A. Shukri, Steven Q. Simpson, Mervyn Singer, B. Taylor Thompson, Sean R. Townsend, Thomas Van der Poll, Jean-Louis Vincent, W. Joost Wiersinga, Janice L. Zimmerman, and R. Phillip Dellinger. Surviving sepsis campaign: International guidelines for management of sepsis and septic shock: 2016. *Intensive Care Medicine*, 43(3):304–377, 2017.

Julia Riedl, Cihan Ay, and Ingrid Pabinger. Platelets and hemophilia: A review of the literature. *Thrombosis Research*, 155:131–139, 2017.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Lecture Notes in Computer Science, pages 234–241. Springer International Publishing, 2015.

Stefan Röhrl, Matthias Ugele, Christian Klenk, Dominik Heim, Oliver Hayden, and Klaus Diepold. Autoencoder features for differentiation of leukocytes based on digital holographic microscopy (DHM). In *Computer Aided Systems Theory - EUROCAST 2019*. Springer, 2019.

Harekrushna Sahoo. Fluorescent labeling techniques in biomolecules: A flashback. *RSC Advances*, 2(18):7017–7029, 2012.

Ulf Schnars and Werner Jüptner. Direct recording of holograms by a CCD target and numerical reconstruction. *Applied optics*, 33(2):179–181, 1994.

Jonas Schulte-Schrepping, Nico Reusch, Daniela Paclik, Kevin Baßler, Stephan Schlickeiser, Bowen Zhang, Benjamin Krämer, Tobias Krammer, Sophia Brumhard, Lorenzo Bonaguro, Elena De Domenico, Daniel Wendisch, Martin Grasshoff, Theodore S. Kapellos, Michael Beckstette, Tal Pecht, Adem Saglam, Oliver Dietrich, Henrik E. Mei, Axel R. Schulz, Claudia Conrad, Désirée Kunkel, Ehsan Vafadarnejad, Cheng-Jian Xu, Arik Horne, Miriam Herbert, Anna Drews, Charlotte Thibeault, Moritz Pfeiffer, Stefan Hippenstiel, Andreas Hocke, Holger Müller-Redetzky, Katrin-Moira Heim, Felix Machleidt, Alexander Uhrig, Laure Bosquillon de Jarcy, Linda Jürgens, Miriam Stegemann, Christoph R. Glösenkamp, Hans-Dieter Volk, Christine Goffinet, Markus Landthaler, Emanuel Wyler, Philipp Georg, Maria Schneider, Chantip Dang-Heine, Nick Neuwinger, Kai Kappert, Rudolf Tauber, Victor Corman, Jan Raabe, Kim Melanie Kaiser, Michael To Vinh, Gereon Rieke, Christian Meisel, Thomas Ulas, Matthias Becker, Robert Geffers, Martin Witzenrath, Christian Drosten, Norbert Suttorp, Christof von Kalle, Florian Kurth, Kristian Händler, Joachim L. Schultze, Anna C. Aschenbrenner, Yang Li, Jacob Nattermann, Birgit Sawitzki, Antoine-Emmanuel Saliba, Leif Erik Sander, and Deutsche COVID-19 OMICS Initiative (DeCOI). Severe COVID-19 is marked by a dysregulated myeloid cell compartment. *Cell*, 182(6):1419–1440.e23, 2020.

Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651, 2017.

Dalwinder Singh and Birmohan Singh. Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97:105524, 2020.

Janet Staats, Anagha Divekar, J Philip McCoy, and Holden T Maecker. Guidelines for gating flow cytometry data for immunological assays. *Immunophenotyping: Methods and Protocols*, pages 81–104, 2019.

Konstantin Stark and Steffen Massberg. Interplay between inflammation and thrombosis in cardiovascular pathology. *Nature Reviews Cardiology*, 18(9):666–682, 2021.

Satoshi Suzuki and Keiichi Abe. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics and Image Processing*, 30(1):32–46, 1985.

Matthias Ugele. *High-throughput hematology analysis with digital holographic microscopy*. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 2019.

Matthias Ugele, Markus Weniger, Maria Leidenberger, Yiwei Huang, Michael Bassler, Oliver Friedrich, Barbara Kappes, Oliver Hayden, and Lukas Richter. Label-free, high-throughput detection of p. falciparum infection in sphered erythrocytes with digital holographic microscopy. *Lab on a Chip*, 18(12):1704–1712, 2018a.

Matthias Ugele, Markus Weniger, Manfred Stanzel, Michael Bassler, Stefan W. Krause, Oliver Friedrich, Oliver Hayden, and Lukas Richter. Label-free high-throughput leukemia detection by holographic microscopy. *Advanced Science*, 5(12), 2018b.

Tom van der Poll, Frank L. van de Veerdonk, Brendon P. Scicluna, and Mihai G. Netea. The immunopathology of sepsis and potential therapeutic targets. *Nature Reviews. Immunology*, 17(7):407–420, 2017.

Luc Vincent and Pierre Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598, 1991.

Pei Wang and Albert C. S. Chung. Focal dice loss and image dilation for brain tumor segmentation. In Danail Stoyanov, Zeike Taylor, Gustavo Carneiro, Tanveer Syeda-Mahmood, Anne Martel, Lena Maier-Hein, João Manuel R.S. Tavares, Andrew Bradley, João Paulo Papa, Vasileios Belagiannis, Jacinto C. Nascimento, Zhi Lu, Sailesh Conjeti, Mehdi Moradi, Hayit Greenspan, and Anant Madabhushi, editors, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Lecture Notes in Computer Science, pages 119–127. Springer International Publishing, 2018.

Aaron M. Wendelboe and Gary E. Raskob. Global burden of thrombosis: Epidemiologic aspects. *Circulation Research*, 118(9):1340–1347, 2016.

Gong Zhang, Tian Guan, Zhiyuan Shen, Xiangnan Wang, Tao Hu, Delai Wang, Yonghong He, and Ni Xie. Fast phase retrieval in off-axis digital holographic microscopy through deep learning. *Optics Express*, 26(15):19388–19405, 2018.

Yu Zuo, Yogendra Kanthi, Jason S. Knight, and Alfred H. J. Kim. The interplay between neutrophils, complement, and microthrombi in COVID-19. *Best Practice & Research. Clinical Rheumatology*, 35(1):101661, 2021.

Appendix A. Preprocessing

Preprocessing is an essential requirement in achieving good segmentation and classification results. The QPI setup provides 512 px by 384 px phase images containing multiple cells as displayed in Figure 1(*b*). These must be prepared to obtain usable images containing only a single cell or cell aggregate, while keeping cell aggregates intact. The operations required to achieve this are discussed below.

A.1. Background Subtraction

Disturbing artifacts and background noise can be removed by calculating the median of 100 images and subtracting it from each frame. This can be done as the imaging setup and the channel is assumed static.

A.2. Cell Detection

The detection of cells in the acquired images is done by thresholding and contour finding. First, binary thresholding is applied to the phase images. From the resulting binary images, contours are extracted based on the algorithm of Suzuki and Abe (1985). The extracted contours are filtered according to a minimum contour area and each cell or cell aggregate (represented by a contour) was then saved as an image snippet of 100 px by 100 px for further processing.

A.3. Masking

Individual cell masking is used to remove unwanted noise from fluid, particles, and other cells. This is done by first thresholding to remove any residual noise caused by the microfluidics channel. To improve the resulting mask, other cells or particles in the image are removed from the mask using the previously calculated contour, and any holes in the mask are filled using morphological dilation and erosion (Burger and Burge, 2016).

A.4. Normalization

Normalization is an essential preprocessing step in any machine learning application, especially when using neural networks. It transforms the feature or image values into a common range. Typical methods are either mean and standard deviation based (like z-score normalization) or minimum-maximum based (Singh and Singh, 2020) (Kotsiantis et al., 2007). For this work, the images were first clipped to limit the value range, as the images resulting from the holographic microscope (theoretically) have an unlimited value range. A minimum clipping value of 0.2 (due to the background) and a maximum clipping value of 4 were used, which showed good results for a combination of platelets, erythrocytes, and leukocytes (with a fully used value range and minimal clipping of cells). *Min-Max normalization* was then applied to transform the image values into the $[0, 1]$ interval suitable for neural networks.

Appendix B. Hyperparameter Optimization and Training

Hyperparameters are all configuration parameters of a neural network that can be set by the user. They directly control the behavior of the network during training and have a dominant impact on model performance. Hyperparameters control the network's architecture, regularization, and most importantly optimization. Tuning of hyperparameters, called hyperparameter optimization, is therefore needed to be able to exploit the full potential of a neural network. Since manual tuning is tedious and inefficient, automatic optimization is widely used.

The simplest optimization methods are grid search, which traverses the search space on a grid in an ordered fashion, and random search, which tries random combinations of hyperparameters. More advanced methods use Gaussian processes and early stopping.

In this work, a combination of the tree-structured parzen estimator (TPE) and the asynchronous successive halving approach (ASHA) is used. TPE is a sequential model based optimization approach. By describing the search space with a graph-structured generative process, a model of the relation between hyperparameters and measured performance of the neural network can be created. This model is then successively optimized by sequentially constructing models to approximate the performance of hyperparameters based on previous measurements and subsequently proposing new hyperparameter combinations. ASHA uses aggressive early stopping of bad performing training steps to allocate more time and computing power to more promising configurations. The combination of these two methods makes very efficient hyperparameter optimization possible.

U-Net For the U-NET, the parameters α, β_1, β_2 and ϵ were optimized based on the search space defined in Table 3. The best choices were $\alpha = 3 \times 10^{-4}$, $\beta_1 = 0.81$, $\beta_2 = 0.994$, and $\epsilon = 4 \times 10^{-8}$.

Parameter	Search space	Choice
learning rate α	loguniform$(1 \times 10^{-5}, 1 \times 10^{-2})$	3×10^{-4}
exp decay rate β_1	uniform$(0, 0.9)$	0.81
exp decay rate β_2	uniform$(0.9, 0.999)$	0.994
numerical stability parameter ϵ	loguniform$(1 \times 10^{-8}, 1)$	4×10^{-8}

Table 3: Search space and chosen value of the hyperparameter optimization of the U-NET

Mask R-CNN For the Mask R-CNN, the parameters learning rate, momentum and weight decay were optimized using the search space defined in Table 4. A learning rate of 1×10^{-3}, a momentum of 0.97, and a weight decay of 5×10^{-4} yielded the best results.

Parameter	Search space	Choice
learning rate lr	loguniform$(1 \times 10^{-5}, 1 \times 10^{-2})$	1×10^{-3}
momentum γ	uniform$(0.9, 0.999)$	0.97
weight decay wd	loguniform$(1 \times 10^{-4}, 1 \times 10^{-1})$	5×10^{-4}

Table 4: Search space and chosen value of the hyperparameter optimization of the Mask R-CNN

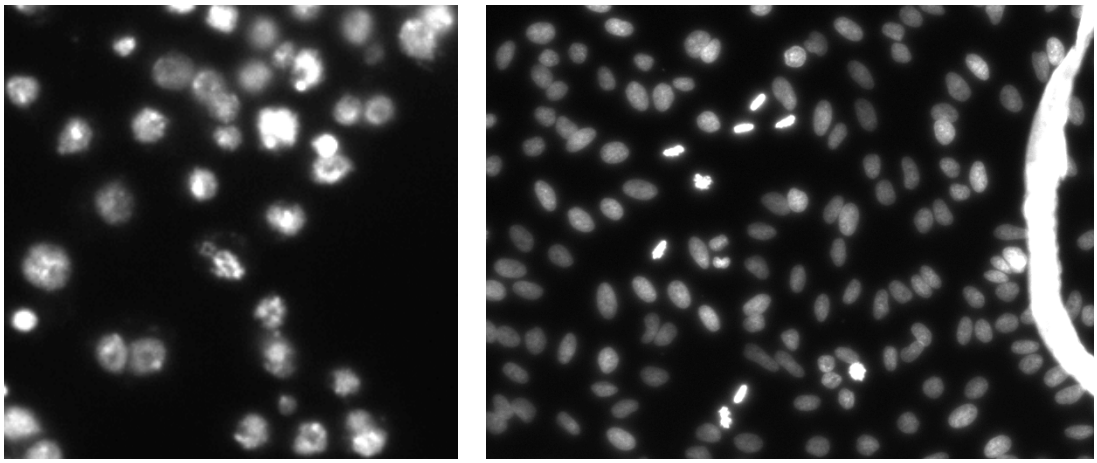
Appendix C. Visualizations

Since our work is very visual, we do not want to deprive readers of the corresponding images and segmentations.

C.1. Data Sets

This section contains exemplary images for the employed data sets.

BBBC038 The BBBC038 data set contains a variety of two-dimensional light microscopy images of stained nuclei. Two examplariy images are displayed in Figure 5.

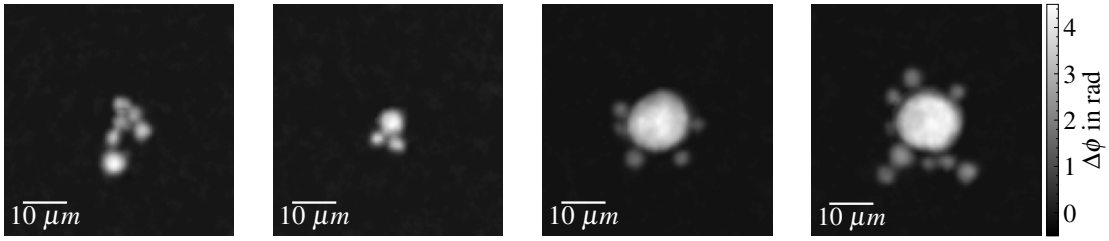


(a) Example image 1 (b) Example image 2

Figure 5: BBBC038 by Caicedo et al. (2019)

Synthetic data set To control the composition of Aggregates, we created a synthetic data set of aggregates, by stitching together multiple single cell images to form cell aggregates. This is based on pure blood cell populations (platelets, erythrocytes, leukocytes) extracted from whole blood by differential centrifugation and density gradient centrifugation. Single cell images and corresponding masks were extracted using threshold segmentation and manual filtering to remove cell duplicates, out-of-focus cells, and flow channel artifacts. The synthetic aggregate images are then iteratively assembled by randomly placing them side by side based on their contour. Since this is done simultaneously for the mask of the

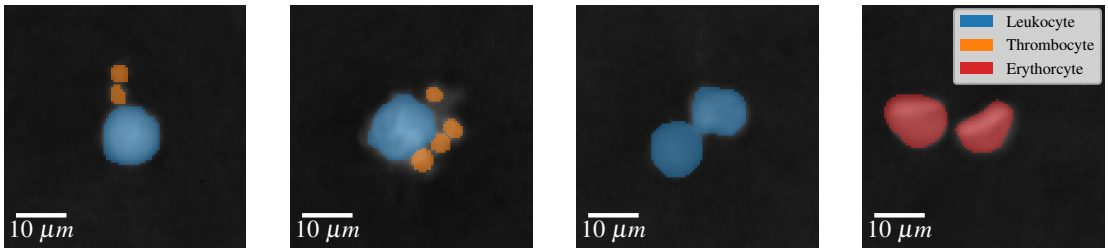
cell images, the ground truth label mask needed for training is created. The results are extremely close to real blood cell aggregates, as visualized in Figure 6.



(*a*) P Aggregate (*b*) P Aggregate (*c*) LP Aggregate (*d*) LP Aggregate

Figure 6: Synthetic platelet (P) and leukocyte-platelet (LP) aggregates

Expert labeled data set We asked a team of biomedical researchers and experts on QPI blood cell analysis to label a data set of 100 images. It consists of 50% single and multiple erythrocytes and 50% platelets and leukocyte-platelet aggregates. The images were manually masked by using a brush tool, resulting in a very accurate segmentation. The according examples are shown in Figure 7.



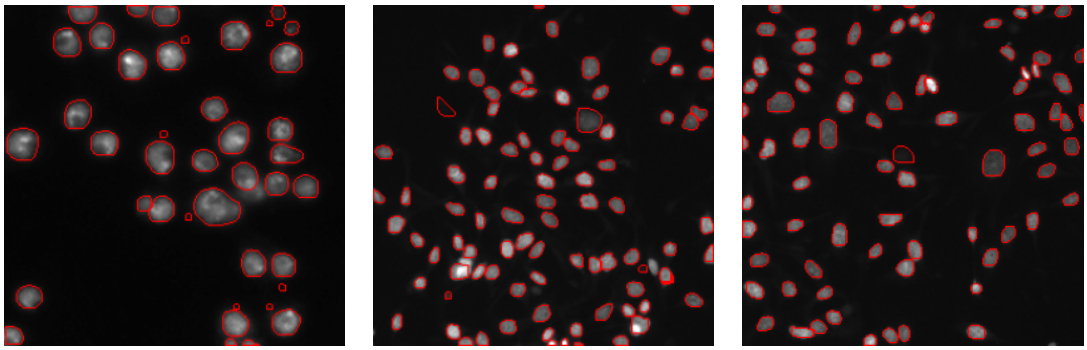
(*a*) LP Aggregate (*b*) LP Aggregate (*c*) Leukocytes (*d*) Erythrocytes

Figure 7: Expert labeled data set

C.2. Segmentation

This section contains sample images generated by our analysis pipeline. Here, the segmentation performance of the Mask R-CNN approach can be observed on the retrieved images patches.

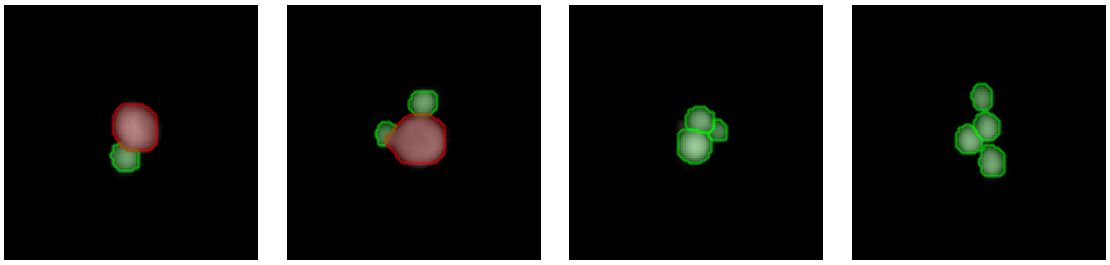
BBBC038 The complexity of the BBBC038 data set is a good benchmark to test the segmentation capabilities of the examined approaches. The Mask R-CNN does an excellent job of detecting a wide range of small and large cells while achieving a high IoU. The contours of the recognized cells are drawn red in Figure 8.



(*a*) Example 1 (*b*) Example 2 (*c*) Example 3

Figure 8: Segmentation examples of the Mask R-CNN on the BBBC038 data set

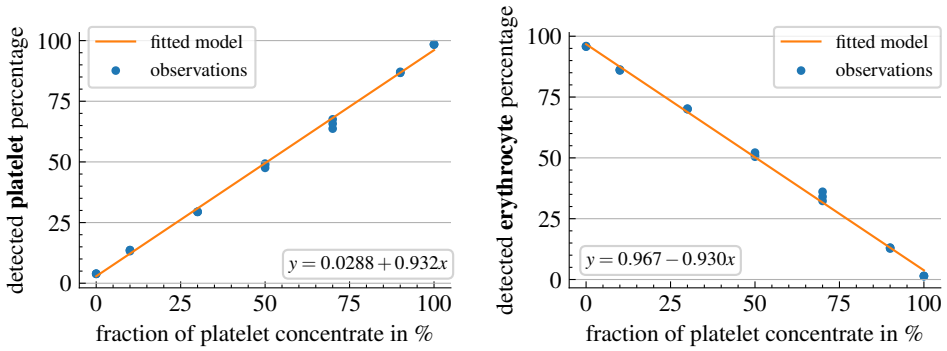
Expert labeled data set Closer to our actual use case, the segmentation of the expert labeled data set puts the focus on the detection of the individual aggregate components. Also for this challenging task, the Mask R-CNN shows a good performance in all quality measures (see Section 5.1). The according visualizations using red and green contours for each detected component can be seen in Figure 9.



(*a*) LP Aggregate (*b*) LP Aggregate (*c*) P Aggregate (*d*) P Aggregate

Figure 9: Segmentation examples of Mask R-CNN on the expert labeled data set: The color of the contour represents the predicted type. Leukocytes (L) are drawn in red and platelets (P) in green.

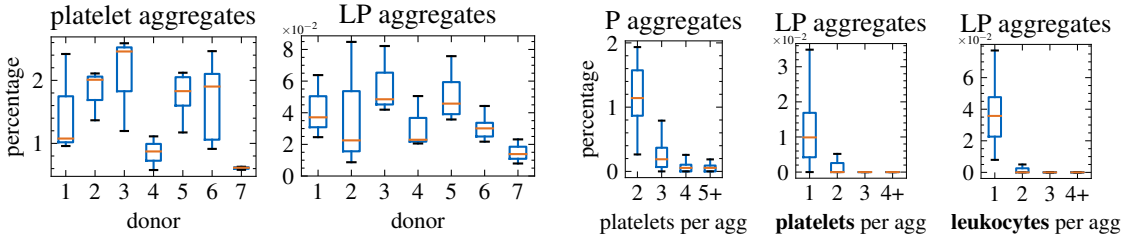
Appendix D. Activated platelets spiked in whole blood



(*a*) Detected platelet percentage (*b*) Detected erythrocyte percentage

Figure 10: Activated platelets spiked in whole blood: The dots show the observations and the curve represents the fitted function.

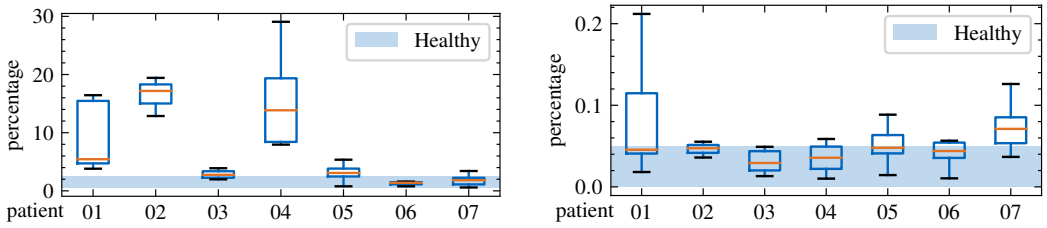
Appendix E. Healthy Reference



(*a*) Platelet (P) and leukocyte-platelet (LP) aggregate occurrence (*b*) Aggregate composition of platelet (P) and leukocyte-platelet (LP) aggregates

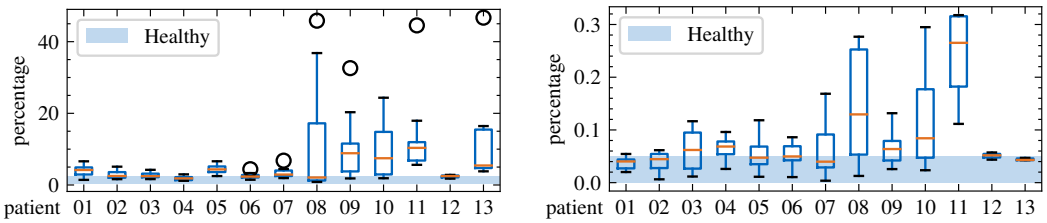
Figure 11: Aggregates in samples from healthy donors

Appendix F. Sepsis and COVID-19 in Detail



(*a*) Platelet aggregates (*b*) Leukocyte-platelet aggregates

Figure 12: Aggregate occurrence in samples from patients with sepsis



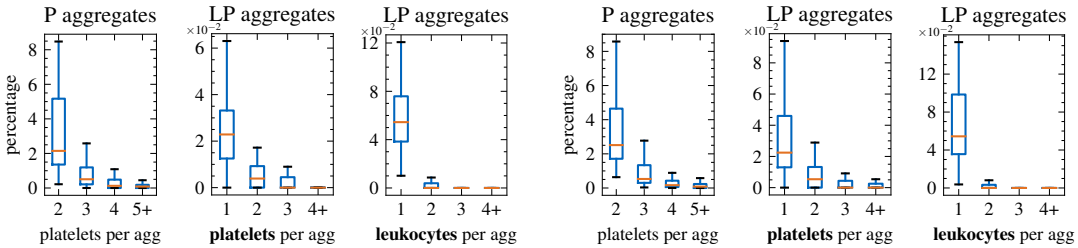
(*a*) Platelet aggregates (*b*) Leukocyte-platelet aggregates

Figure 13: Aggregate occurrence in samples from patients with COVID-19

Appendix G. Aggregate Composition

In order to gain a better understanding of the characteristics of aggregates, we analyzed the size distribution of aggregates of both the sepsis and COVID-19 samples. For the assessment of platelet aggregate size, the samples are divided into two categories, samples with lower platelet aggregate levels and samples with higher platelet aggregate levels, separated by a threshold of 10%. For the sepsis cohort, patients with sepsis with fewer observed aggregates show a similar distribution of aggregate size to healthy donors, while samples with higher aggregate levels also show comparatively more larger aggregates, as shown in Figure 14(*a*). The same is observable for COVID-19 patients, as shown in Figure 14(*b*).

To analyze the amount of platelets in leukocyte-platelet aggregates a threshold of 0.2% was chosen to divide the samples into two categories of lower and higher observed leukocyte-platelet aggregates. Analysis of the amount of platelets in leukocyte-platelet aggregates shows similar results to those of platelet aggregates for sepsis patients, as shown in Figure 14(*a*), as well as COVID-19 patients, as shown in Figure 14(*b*). Patients with more observed aggregates also showed comparatively more larger aggregates. However, the analysis of leukocyte amounts in leukocyte-platelet aggregates showed slightly different results, as shown in Figure 14(*a*) and 14(*b*). Both in healthy donors, in patients with fewer observed leukocyte-platelet aggregates, and in patients with more leukocyte-platelet aggregates almost only aggregates containing a single leukocyte are observed, all of these featuring a similar distribution, just with different levels in general.



(*a*) In sepsis samples (*b*) In COVID-19 samples

Figure 14: Aggregate composition of platelet (P) and leukocyte-platelet (LP) aggregates