



Automated characterisation of neutrophil activation phenotypes in *ex vivo* human *Candida* blood infections



Ivan Belyaev^{a,b,1}, Alessandra Marolda^{c,1}, Jan-Philipp Praetorius^{a,b}, Arjun Sarkar^{a,b}, Anna Medyukhina^{a,d}, Kerstin Hünninger^{c,e}, Oliver Kurzai^{c,e,f,*}, Marc Thilo Figge^{a,g,*}

^a Applied Systems Biology, Leibniz Institute for Natural Product Research and Infection Biology – Hans-Knöll-Institute, Jena, Germany

^b Faculty of Biological Sciences, Friedrich-Schiller-University Jena, Germany

^c Fungal Septomics, Leibniz Institute for Natural Product Research and Infection Biology – Hans-Knöll-Institute, Jena, Germany

^d Center for Bioimage Informatics, St. Jude Children's Research Hospital, Memphis, TN

^e Institute for Hygiene and Microbiology, University of Würzburg, Würzburg, Germany

^f Center for Sepsis Control and Care (CSCC), Jena University Hospital, Jena, Germany

^g Institute of Microbiology, Faculty of Biological Sciences, Friedrich-Schiller-University Jena, Germany

ARTICLE INFO

Article history:

Received 21 January 2022

Received in revised form 4 May 2022

Accepted 5 May 2022

Available online 10 May 2022

Keywords:

Candida infection

Bloodstream infection

Whole blood infection model

Image analysis

Machine learning

Diagnostic markers

ABSTRACT

Rapid identification of pathogens is required for early diagnosis and treatment of life-threatening bloodstream infections in humans. This requirement is driving the current developments of molecular diagnostic tools identifying pathogens from human whole blood after successful isolation and cultivation. An alternative approach is to determine pathogen-specific signatures from human host immune cells that have been exposed to pathogens. We hypothesise that activated immune cells, such as neutrophils, may exhibit a characteristic behaviour – for instance in terms of their speed, dynamic cell morphology – that allows (i) identifying the type of pathogen indirectly and (ii) providing information on therapeutic efficacy. In this feasibility study, we propose a method for the quantitative assessment of static and morphodynamic features of neutrophils based on label-free time-lapse imaging data. We investigate neutrophil activation phenotypes after confrontation with fungal pathogens and isolation from a human whole-blood assay. In particular, we applied a machine learning supported approach to time-lapse microscopy data from different infection scenarios and were able to distinguish between *Candida albicans* and *C. glabrata* infection scenarios with test accuracies well above 75%, and to identify pathogen-free samples with accuracy reaching 100%. These results significantly exceed the test accuracies achieved using state-of-the-art deep neural networks to classify neutrophils by their morphodynamics.

© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Candida bloodstream infections (BSI) are the most common form of invasive candidiasis and constitute the fourth leading cause of nosocomial invasive infections in Intensive Care Units (ICUs) patients in the US [1]. The study on Extended Prevalence of Infection in Intensive Care (EPIC II) revealed that the prevalence

of *Candida* BSI was 6.9 per 1000 patients with an associated mortality rate of around 43% compared to 27% caused by bacterial BSI [2]. Among hospitalised patients, *Candida* species present the most frequent isolated fungal BSI pathogens [3]. In particular, *C. albicans* and *C. glabrata* are responsible for the majority of *Candida* cases worldwide, where *C. albicans* is the predominant species with 50% of cases, while *C. glabrata* is responsible for 15–25% of invasive *Candida* infections in the US and Northern Europe [4]. These statistical data imply that methods for the fast and reliable diagnosis are urgently needed to allow for an early start of targeted treatments.

Various animal models have been used to study invasive *Candida* infections, such as fruit fly, zebrafish and mouse. In contrast, human whole-blood infection (WBI) models enable analysing host-pathogen interactions in a setting similar to *in vivo* BSI [5]. The human WBI models allowed (i) identifying virulence factors [6], (ii) analysing immune responses including time-resolved data

* Corresponding authors at: Fungal Septomics, Leibniz Institute for Natural Product Research and Infection Biology – Hans-Knöll-Institute, Jena, Germany (O. Kurzai). Applied Systems Biology, Leibniz Institute for Natural Product Research and Infection Biology – Hans-Knöll-Institute, Jena, Germany (M.T. Figge).

E-mail addresses: oliver.kurzai@leibniz-hki.de (O. Kurzai), thilo.figge@leibniz-hki.de (M.T. Figge).

¹ Authors contributed equally.

² Authors contributed equally.

on immune cell activation and pathogen status [7], and (iii) testing potential therapeutic approaches [8,9]. We have previously studied BSI by combining the human WBI model with the quantification of immune processes by virtual infection modelling [5,10]. In this context, we found that neutrophils play a central role in the defence against *C. albicans* BSI. Moreover, we have performed extensive comparative analyses for the two species *C. albicans* and *C. glabrata* and found that they are differentially recognised by neutrophils using live cell imaging combined with automated image analysis [11–14] and computational modelling [15–18]. This motivated us to study the possibility of automatically identifying the type of pathogen in BSI based on neutrophil morphological properties.

In this feasibility study, we combine the human WBI model with live cell imaging of primary neutrophils and computational analysis to extract features that allow us to detect BSI caused by *C. albicans* and *C. glabrata*. The central hypothesis is that (i) neutrophils in a human WBI model respond with morphological changes to the presence of pathogens and (ii) these changes are pathogen-specific. To advance the development of rapid and reliable diagnostic methods, we are investigating the potential of the automated characterisation of neutrophil activation phenotypes for human *Candida* BSI. From a technical point of view, this study exploits our recent developments regarding the analysis of live cell imaging data with respect to tracking of unlabelled cells over extended times [19] and segmenting cells with high accuracy for dynamical changes of their morphology (morphodynamics) [20]. Features of immune cells under different stimuli that have been previously studied include (i) changes in cell size [21], (ii) modifications of membrane topography [22,23] and (iii) variations in the migration behaviour [24]. Our study utilizes features based on the cell morphodynamics and provides a fully automated pipeline based on live cell imaging data of unlabelled primary neutrophils in order to distinguish the two scenarios of *Candida* BSI.

2. Materials and Methods

2.1. Ethics statement

This study was conducted in accordance with the Declaration of Helsinki. All protocols were approved by the Ethics Committee, University Hospital Jena (permit number: 273–12/09).

2.2. Fungal strains and culture

GFP-expressing *C. albicans* [5] and *C. glabrata* [25] strains were routinely used in all experiments. *C. albicans* and *C. glabrata* were seeded in yeast extract–peptone–dextrose medium (YPD medium: 2% D-glucose, 1% peptone, and 0.5% yeast extract, in water) and grown overnight at 30 °C and 37 °C, respectively, in a shaking incubator. Both fungal species were reseeded in fresh YPD medium, grown until they reach the mid-log phase followed by harvesting in HBSS.

2.3. Whole-blood infection model

To avoid anticoagulation and not influence complement activation, human peripheral blood from healthy donors was drawn in Hirudin S-monovettes® (Sarstedt) after informed consent. Whole-blood infection assay was performed as described previously, using an inoculum that allows rapid innate immune activation but precludes unspecific effects on adaptive immune cells [5]. In brief, HBSS (mock-infected control), *C. albicans* or *C. glabrata* were added in a final concentration of 1×10^6 fungal cells per 1 ml of whole blood and then incubated for 1 h at 37 °C on a rolling mixer. After

incubation, samples were used directly for neutrophil isolation with sequential live cells imaging of neutrophils.

2.4. Isolation of human neutrophils

Untouched neutrophils were isolated from either mock- or *Candida*-infected blood using MACSxpress Whole Blood Neutrophil Isolation Kit according to the instructions from the manufacturer (Miltenyi Biotec). Remaining erythrocytes were lysed for 5 min with ACK Lysing Buffer (Life Technologies) and purity of neutrophils was checked at flow cytometry to be >95% (see Supplementary Fig. 1). For this, neutrophils were stained with mouse anti-human CD66b antibody (BD Biosciences Cat# 561649, RRID: AB_10897169) for 20 min at 4 °C and measured with the BD FACSCanto™ II system and the BD FACSDiva™ software (both BD Biosciences). In parallel, staining with the appropriate isotype control antibody (BD Biosciences Cat# 560861, RRID: AB_10926214) was performed to ensure specificity of antibody binding. FlowJo10 software was used for analysis. Obtained neutrophils were resuspended in RPMI 1640 with 5% heat-inactivated human serum and used for live cell imaging.

2.5. Live cell imaging and Time-lapse microscopy

Live cell imaging was performed by adding 4×10^5 neutrophils isolated either from mock-, *C. albicans*- or *C. glabrata*-infected human blood in a μ -dish (ibidi) in a total volume of 2 ml RPMI 1640 containing 5% heat-inactivated human serum. 2.5 ng/ml of propidium iodide (PI, Sigma) was added into the medium to distinguish viable cells from dead ones. PI stains only nucleic acids in dying cells characterized by leak in the plasma membrane. Therefore, death of a neutrophil or a fungal cell can be identified in the video by the respective cell/fungus turning red fluorescent. Neutrophils were incubated in an environmental control chamber at 37 °C and 5% CO₂. Images were acquired every 7 s with a Zeiss LSM 780 confocal microscope, which was focused on the bottom of the dish. Cells behaviour was monitored with a 20x microscope objective (Zeiss Plan-APOCHROMAT 20x/0.8NA) using a differential interference contrast (DIC) setting with illumination by 488 nm laser. Image size was 2048 by 2048 px with the scale 0.208 μ m/px.

2.6. Segmentation and tracking of neutrophils

For cell detection and tracking we used our Algorithm for Migration and Interaction Tracking (AMIT, [13,14,19]) in its latest release of the third version [20] that is available from our GitHub repository: <https://github.com/applied-systems-biology/amit>. AMIT enables automated segmentation and tracking of label-free cells from microscopy data. In addition, it provides the possibility to eliminate track segments associated with long-lasting clusters of cells that may be indistinguishable by eye. This post-processing procedure is necessary, because the extraction of unbiased information about the morphology of individual cells inside such clusters is impossible. The maximal possible track duration is about 30 min corresponding to the 260 frames of recorded video with a frame rate of about one frame per seven seconds.

2.7. Measurement of neutrophil speed

The instantaneous cell speed was calculated from consecutive time steps in μ m/min for each cell track. These instantaneous speed values were used to compute the arithmetic mean speed value per cell. The latter was collected from all cell tracks of a video as a representative speed value distribution. In addition, for every video, the set of instantaneous cell speeds was split into two sub-

sets by a cell morphology detector. As explained in detail below, this detector distinguishes between cells with non-spreading morphology (N-morphology) and spreading morphology (S-morphology) and enables to numerically distinguish the measured neutrophil speed for these two morphology states.

2.8. Extraction of gradient-based cell features

For each frame in a video from time-lapse microscopy, we applied a morphological contrast enhancement [26] and contrast-limited adaptive histogram equalisation [27] as a pre-processing step, followed by gradient detection with the Sobel operator [28] to compute an intensity gradient magnitude map per image. Afterwards, all values were normalised to the maximal value of the gradient amplitude for a given video (see Supplementary Fig. 2) and for each previously segmented neutrophil, the value range of [10th, 80th] percentiles in the pixel intensity was used as a descriptor of cell surface roughness. This feature is also referred to as pHG-descriptor, since it is based on the percentiles of the histogram for the normalised gradient magnitudes.

2.9. Data set organisation and sampling procedures

After cell tracking and feature extraction, each video was represented by a track file consisting of a table that contains the description of each cell at every time point. The various analyses of the video data were then made based on this table.

For the evaluation of the robustness of the N-morphology detector, we performed Monte-Carlo simulations by randomly selecting 1.8×10^3 cells per iteration from every video of the mock-infected samples. These were used in the calibration set, while the video data of samples infected by *Candida* composed the test set. The number of 1.8×10^3 cells was chosen, because it corresponds approximately to the number of cells in the episodes of the first 5 min of the video with lowest cell concentration. Including data from each donor was necessary for the compensation of variations between videos regarding illumination issues that were not fully compensated during extraction of gradient-based cell features.

In the population-based analysis of snapshots, information about each video was split into two parts corresponding, respectively, to the first 42 frames (~ 5-minutes-episodes) and the following 218 frames (~ 25-minutes-episodes) of a video. Then, for the N-morphology detector calibration, we randomly selected the 1.8×10^3 cells from the 5-minutes-episodes of each video of mock-infected samples. The 25-minutes-episodes composed the test set and was used for the estimation of spreading cell fraction in each sample (see subsection 3.3) and instantaneous speed analysis (see subsection 3.4).

The morphodynamics analysis was performed on the complete videos (30 min) of *Candida*-infected samples. In these analyses of tracked cells, we compared characteristic distributions (see subsections 3.4 and 3.5). In order to reduce the influence of individual samples we used a fixed number of instances (complete cell tracks or track fragments) that were randomly selected from each video in the following way: (i) the characteristics of interest were sorted in ascending order, where multiway sorting was applied in the case of more than one descriptor; (ii) the desired number of instances was derived by generating random indexes from a uniform distribution covering the whole range of the initial vector indices. This strategy yields statistically representative sampling.

2.10. Detection of cells with N-morphology

For the detection of cells based on specific descriptors, a single-class classifier was created using the method of Data Driven Soft Independent Modelling of Class Analogy (DD SIMCA) [29,30], which is a modification [31] of the classical SIMCA [32] approach. SIMCA is a well-known tool for pattern recognition in many research and industrial applications (for example, see [33–38]). The DD SIMCA method utilises a decomposition of data by principal component analysis (PCA) [39] for a description of the target class data structure within a multicollinear feature space combined with the statistics of two distances that are used to characterise variability inside the calibration set (see Supplementary Fig. 3). The first distance refers to the position of an object (the point in a multivariate feature space which represents a real object) relative to the model (orthogonal distance, OD) and the second distance refers the displacement between the projection of the object onto the model and the centre of this model (score distance, SD). The statistics of these distances are used to establish two rules: (i) a decision rule for the detection of extreme/unusual objects, *i.e.* objects that do not follow major trends in the calibration data grasped by the principal components, and (ii) an acceptance rule for the classification of new objects. Both rules impose a comparison of the respective statistics of the distances with regard to critical values. While the classical SIMCA relies on F-statistics of OD and utilises parameters of the calibration data set (number of samples and variables) together with the number of chosen components in the PCA model for the computation of critical values, the DD SIMCA employs scaled chi-squared distributions of OD and SD for the calibration set in the estimation of critical values. Respecting the data structure makes the latter method more suitable for statistical unmixing of multivariate distributions of data.

In the present study, we applied this method in the following unsupervised way: (i) PCA was performed for the whole calibration data set comprising cells from mock-infected samples with N-morphology being the dominant form. The analysis of PCA results revealed that the first and second principle components are enough to describe more than 90% of total variation in data on mock-infected samples (see Supplementary Fig. 3). We therefore limited the number of principle components in the model to two. (ii) An outlier border was determined using the outlier significance level $\gamma = 0.01$, which specifies the probability that at least one point from the data will be erroneously considered as an outlier [30]. (iii) All data points beyond the outlier border were considered to be outliers and were removed from the calibration data set, *i.e.* we performed multidimensional distribution unmixing and obtained a representative purified population of cells with N-morphology. This filtered calibration set was then used for the final model calibration and acceptance area determination. (iv) In the classification procedure, all data points within the acceptance area were assigned to be cells with N-morphology, while all other cells were considered to have S-morphology.

All operations were done using the R package ‘mdatools’ [40].

2.11. Automated classification of infection scenarios

To automatically classify the various infection scenarios, we applied a two-step procedure: (i) each instance, *i.e.* a video frame in the population-based analysis of snapshots or a cell track in the morphodynamics analysis, was classified by a Bayesian classifier [41] after pre-calibration by a calibration set. (ii) A video was assigned to a certain infection scenario based on a majority voting by the individually classified instances.

We used the R package ‘naivebayes’ [42] to perform the Bayesian classification (in case of multiple descriptors a naïve form).

2.12. Comparison of cell characteristics for different infection scenarios

We used the multiple quantile comparison method [43]. This method utilises a combination of the Harrell-Davis quantile estimator [44] and a bootstrapping to determine the confidence interval (CI) for the difference between quantiles of the distributions to be compared. The difference in certain quantiles of any two distributions is considered to be significant in frequentist sense at confidence level α .

All computations were done with the R package ‘WRS2’ [45].

2.13. Multiple group comparison test

In cases where the data have a non-replicated complete block design we used the Quade test with post-hoc analysis [46]. This method is a generalisation of the signed paired rank test for three or more groups, where the null-hypothesis says that, apart from donor effects, the location parameter of the analysed property is the same for each infection scenario. The obtained p-values were adjusted by the Holm method [47].

All operations were done using the R package ‘PMCMRplus’ [48].

2.14. Effect size statistics

To numerically characterise a magnitude of difference between conditions and be able to compare it between different characteristics (median fraction of spreading cells per frame and average speed) we used the common language effect size (CLES), which expresses the probability that a randomly selected score from one group will be greater than a randomly sampled score from another one [49]. The values were computed with the R package ‘canprot’ [50] using empirical probability density functions.

In addition, we computed the difference between distributions via Hedges g effect size statistics [51] for paired measurements [52] and computed the standard deviations for each group individually. For this we used the R package ‘effsize’ [53].

2.15. Confidence intervals for proportions

We applied the method of Wilson’s confidence interval computation for single proportions [54,55] using the R package ‘PropCIs’ [56]. The input arguments included: (i) confidence level α (probability of type I error, set to $\alpha = 0.05$), (ii) total number Θ of entities to be examined and (iii) number of ‘successes’ θ . For the interval estimation of the fraction of cells with morphodynamics that is considered to be specific for a true infection scenario in a given sample, θ corresponds to the number of such cells among all Θ examined cells. See the following sub-section for more details.

2.16. Post-hoc analysis of errors of type II

In addition to Wilson’s confidence intervals we computed the probability for making an error of type II regarding the evaluation of two alternative hypotheses about pathogen-specific morphodynamics. The null-hypothesis corresponds to the statement that a pathogen-specific morphodynamics does not exist; therefore, fractions of cells with *C. albicans*- and *C. glabrata*-specific morphodynamics are expected to be equal 0.5 in each sample. The alternative corresponds to the hypothesis that a pathogen-specific morphodynamics does exist; therefore, observed fractions of cells with specific morphodynamics for a given infection scenario must be greater than 0.5. For samples where the detected fraction of cells specific for a given pathogen is less than 0.5, the probability of error type II for a given set of hypotheses cannot be computed.

The probability was computed using a one-sample single side test for proportions [57] as implemented in the R package ‘MKPower’ [58].

3. Results

Our results are based on time-lapse imaging data of live unlabelled neutrophils, recorded over a period of 30 min with a frame rate of about one frame per seven seconds (260 frames in total) (for details see subsection 2.5). These cells were isolated from human whole-blood infection (WBI) assays (see subsection 2.3) with either of two *Candida* species – *C. albicans* or *C. glabrata* – and were compared to neutrophils from mock-infected blood. In total we have acquired blood samples from 9 healthy donors that were each subdivided to separately study and compare the three infection scenarios.

3.1. Neutrophils exhibit morphological signatures induced by pathogen-interaction in human whole blood

Visual inspection of the video data revealed the existence of two types of dynamically appearing cell morphologies. In Fig. 1, we provide a typical example for a neutrophil that dynamically changes its morphology into the state of a spreading cell (S-morphology) and back into the morphology of a non-spreading cell (N-morphology) via a sequence of intermediate states. Thus, in a first approximation, the cell population $C(t_i)$ in video frame i can be considered as a mixed distribution of cells composed of two morphologies: $C(t_i) = S(t_i) + N(t_i)$, where $S(t_i)$ and $N(t_i)$ denote the number of cells with S- and N-morphology, respectively, at time point t_i . The fraction of cells exhibiting S-morphology is defined by.

$$\sigma(t_i) \equiv S(t_i)/C(t_i) \quad (1)$$

with $0 \leq \sigma(t_i) < 1$. In agreement with previous findings that peculiar morphological patterns are a sign of neutrophil activation [23], we observed that cells with S-morphology were only rarely present ($\sigma(t_i) \ll 1$) among neutrophils isolated from mock-infected blood (see Supplementary video set 1). In contrast, the occurrence of cells with S-morphology was observed more frequently and for a larger cell fraction after confrontation with either *C. albicans* or *C. glabrata*. This observation motivated us to design a workflow for the automated identification of neutrophil morphology and the quantitative comparison of infection scenarios by the occurrence of cells with S-morphology.

3.2. Automated classification yields highly robust predictions of neutrophil morphology

We performed the segmentation and tracking of neutrophils with our software tool AMIT [13,14], which was recently enhanced to recognise whole cell tracks [19]) and to additionally extract morphological information on dynamically changing cell shapes [20]. In particular, the distinction between S- and N-morphology of neutrophils required the identification of descriptors that are sensitive to the size and surface texture of cells and robust against varying and uneven illumination in the images as well as against inaccuracies in the cell segmentation. Considering these requirements as well as the non-rigidness of neutrophil shapes, two adequate descriptors were identified: (i) the cell’s footprint area and (ii) the intensity-gradient of segmented neutrophils (for details see subsection 2.8).

Using these descriptors, we built a one-class classifier for neutrophils with N-morphology acting as a novelty detector, *i.e.* all cells rejected by the classifier were considered to be cells with S-

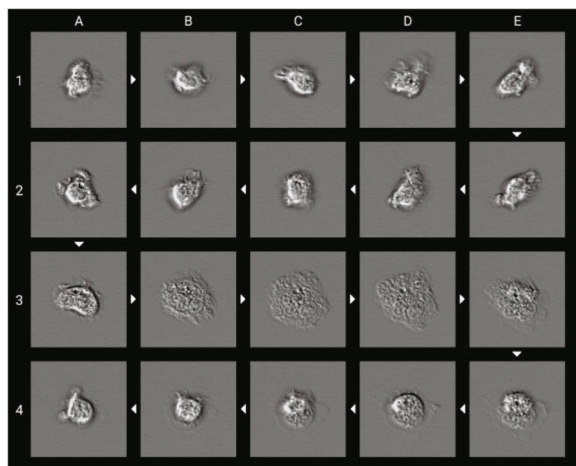


Fig. 1. Time-dependent change of a single neutrophil during 20 consecutive frames (arrows indicate the time ordering). Cells in sub-images B3–E3 can be considered as spreading cells (S-morphology).

morphology. This approach elegantly circumvents the necessity of manual distinction between S- and N-morphology for every cell, which is labour-intensive and could easily result in a bias of the classifier. Thus, we here use our observational knowledge that neutrophils with N-morphology are the predominant type within mock-infected samples. The one-class classifier corresponds to a statistical procedure for the unmixing of the morphology distribution and allows the estimation of $\sigma(t_i)$ for every frame in a video (for details see subsection 2.10).

The robustness of our classifier regarding estimation of $\sigma(t_i)$ was checked by performing Monte-Carlo simulations with 10^3 repetitions, where the N-morphology detector was calibrated using an equal number of randomly selected cells from each mock-infected sample (for details see subsection 2.9). This detector was used for cell classification in videos with neutrophils isolated from *C. albicans* and *C. glabrata* WBI. For every cell the received labels, i.e. S- or N-morphology, were recorded before counting how often each cell was assigned to be a cell with S-morphology. We then analysed the statistics of cells being associated with that class in at least one of iteration during the simulations (Fig. 2). For every infected sample, more than 80% of such cells received that label 10^3 times (Fig. 2, see also Supplementary Fig. 4). This supports the robustness of the classifier outcome with regard to providing a trustworthy estimate of $\sigma(t_i)$. Supplementary video set 2 visualises the classification results from a single iteration for various videos.

3.3. Donor variability obscures predictions based on classification by neutrophil morphology

Based on our one-class classifier, we addressed our hypothesis that the three scenarios – mock-infection, *C. albicans* infection and *C. glabrata* infection – may be distinguishable by the frequency of neutrophils occurring in the S- or N-morphology. For this analysis every video was divided into two episodes with durations of 5 and 25 min, which were used for the calibration of the N-morphology detector and for the characterisation of samples via the distributions of values $\{\sigma(t_i)\}$ (for details see subsection 2.9). Taking into account fluctuations of $\sigma(t_i)$ caused by cell migration in and out of the field of view, we focused on the median as the indicator of central tendency of the $\{\sigma(t_i)\}$ -distribution for the 25-minutes-episodes. The classification results are summarised per donor in Fig. 3a and reveal quantitative differences between the three infection scenarios (see Fig. 3b and Table 1). The statistical differences per infection scenario support our hypothesis

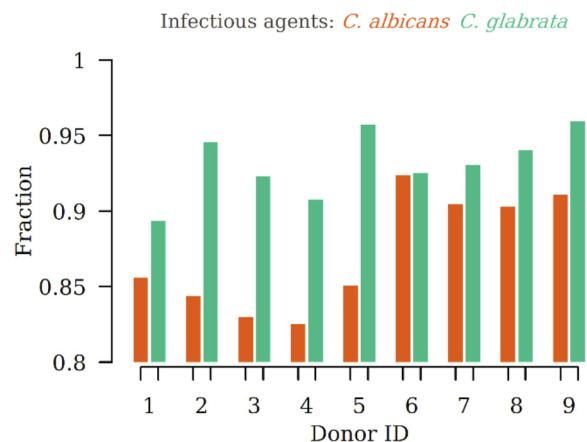


Fig. 2. Fraction of cells repeatedly identified as exhibiting S-morphology in each repetition of the Monte-Carlo simulations. Each sample includes $O(10^4)$ segmented cell images.

regarding the pathogen-specific morphological changes of neutrophils in a human WBI assay.

Next, we implemented a Bayesian classifier with majority voting to identify infection scenarios based on $\{\sigma(t_i)\}$ -distributions (for details see subsection 2.11). We performed simulations with leave-one-out cross-validation (LOOCV) [41], where we performed 9^3 iterations for the three infection scenarios with nine samples each by fixing one sample from every infection scenario as test sample and using all other samples for classifier calibration. This approach allows imitating large sample populations and measuring the classifier performance in the case of low sample numbers.

The classifier was evaluated by the observed successful classification ratio (OSCR), which equals the fraction of correctly assigned samples for a given class. As can be seen in Fig. 3c, our classification procedure recognises mock-infected samples with OSCR = 1, which confirms that the classifier can successfully distinguish infected and non-infected samples. However, as can be inferred from Fig. 3a, distinguishing between different infection scenarios is obscured by the donor variability. In fact, for *C. albicans* infection we obtained the reduced value of OSCR = 0.89, while for *C. glabrata* infection the performance dropped to OSCR = 0.67. As shown in Fig. 3b, the medians of the $\{\sigma(t_i)\}$ -distributions were not suited for achieving a clear distinction between the two infection scenarios. The LOOCV reveals that this is also true for the mock-infected samples (see Fig. 3d), as can be seen from the reduction of the certainty measure by about 11% (compare Fig. 3c and d). Nevertheless, these overall promising findings prompted us to advance our analysis from a population-based analysis of snapshots to the analysis of individual cell tracks including aspects of morphodynamics.

3.4. Neutrophil speed is inadequate for discrimination of *Candida* infection scenarios

Visual inspection of the video data gives the impression that, depending on the infection scenario with either of the two *Candida* species, the morphodynamics of neutrophils may be different (see Supplementary video set 1). In particular, neutrophils seem to (i) experience differently long episodes in the state of S-morphology and (ii) migrate slower when in the state of S-morphology compared to N-morphology. We hypothesised that a specific morphodynamics behaviour of neutrophils may be induced upon contact with a particular pathogen in human whole blood and speculated that the discrimination of infection scenarios may be improved by accounting for dynamic effects.

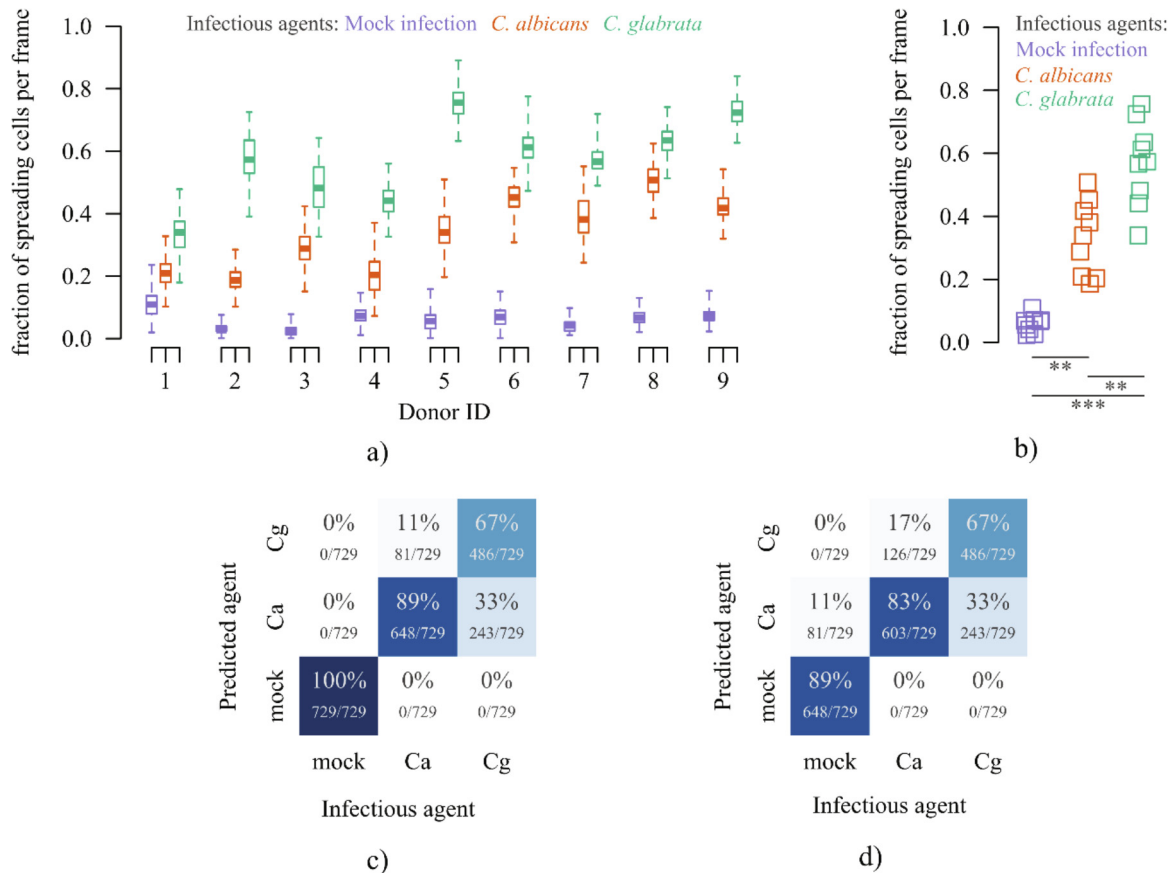


Fig. 3. a) Box diagrams for the fraction of spreading cells per video frame (260 frames in total) for each donor. b) Median value of the distributions in a) per donor. ** $p = 0.0027$, *** $p \ll 10^{-4}$ (Quade test with post-hoc analysis and p adjustment by Holm). The effect size statistics is listed in the Table 1. c) Confusion matrix for the results of the Bayesian classifications based on individual frames. Each cell of the matrix represents the ratio of proper sample classifications (numerator) for given infection scenarios over all iterations (denominator). d) Confusion matrix for the results of a sample classification based on description of whole video data.

Table 1

Comparison of effect sizes expressed via common language effect size (CLES) and Hedge's $|g|$ for median fraction of spreading cells ($CLES_{frac}, |g_{frac}|$) and for average speed per sample ($CLES_{speed}, |g_{speed}|$). Details about calculations are described in paragraph *Effect size statistics* in Materials and Methods section.

Pair for comparison	$CLES_{frac}$	$CLES_{speed}$	$ g_{frac} $	$ g_{speed} $
'mock'-' <i>C. albicans</i> '	1	0.91	3.18	1.58
'mock'-' <i>C. glabrata</i> '	1	0.94	5.47	2.13
' <i>C. albicans</i> '-' <i>C. glabrata</i> '	0.90	0.63	1.67	0.35

To quantify these observations, we first computed the average speed for each neutrophil from its individual cell track for each donor and infection scenario. However, as shown in Fig. 4a, there is no evidence for a clear pathogen-specific impact on the average neutrophil speed. On first sight, this finding may seem contradictory to a previous study where the average neutrophil speed was reported to be a suitable discrimination feature [24]. However, while that study was performed in the context of the myelodysplastic syndrome, the sample average neutrophil speed (Fig. 4b) is evidently not an adequate discrimination feature in the present context of *Candida* BSI, because this measure appears to be less pathogen-sensitive than the dynamic change in cell morphology (Fig. 3b). This was also confirmed by a quantitative comparison of effect size measures (see Table 1 and subsection 2.14) and considering our results on the morphology-based classification of infection scenarios (Fig. 3d).

Thus, while a speed-based classification of infection scenarios will not yield acceptable accuracies, we still wanted to validate

our impression from the visual inspection that there are differences in the morphodynamics of neutrophils for the two *Candida* species. To this end, neutrophils were first classified as being either in the S- or N-morphology followed by the computation of the instantaneous speed distributions for each infection scenario with the two *Candida* species. As can be seen in Fig. 5a and b, the majority of neutrophils with S-morphology are indeed statistically slower, which has also been confirmed by a comparison of the distributions using the multiple quantile comparison method [59] to compute the difference Δ between consecutive percentiles of the respective distributions for the two *Candida* infection scenarios (Fig. 5c). Another evidence for a speed difference between spreading cells and non-spreading ones is a near-monotonical decline (Spearman's $\rho = -0.74$) of the average speed per cell with increasing fraction of spreading cells (Fig. 5d). Thus, while the visual impression could be confirmed, we still have to conclude that neutrophil speed is not an adequate feature for discrimination of WBI with different *Candida* species.

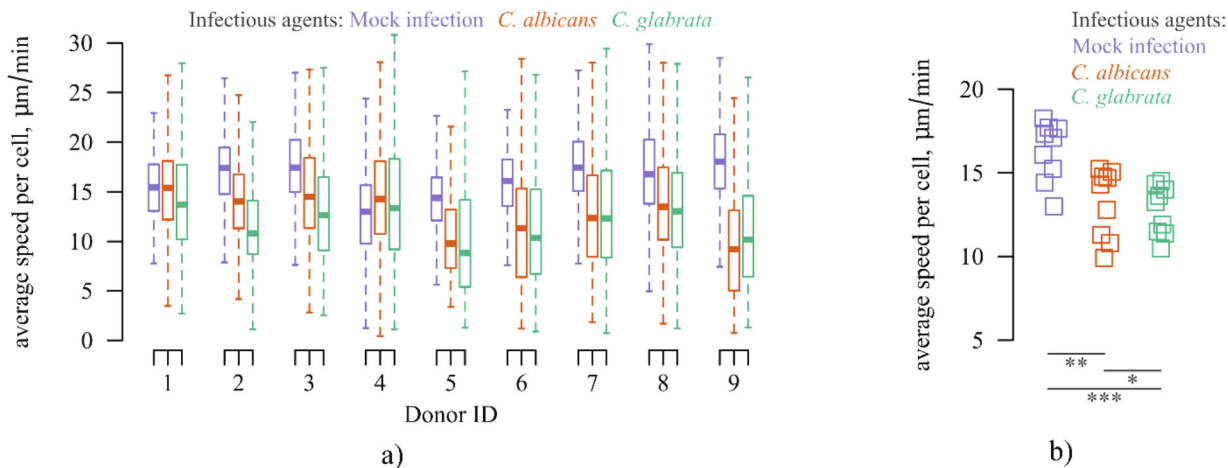


Fig. 4. Diagrams of the average speed per cell (a) and per donor (b). The number of data points per sample is $O(10^4)$, length of whiskers is not larger than 1.5 interquartile interval. For data in (b) the Quade statistical test was applied with post-hoc analysis and p adjustment by Holm: * $p = 0.1265$, ** $p = 0.0224$, *** $p = 0.0011$. The effect size statistics is listed in the Table 1.

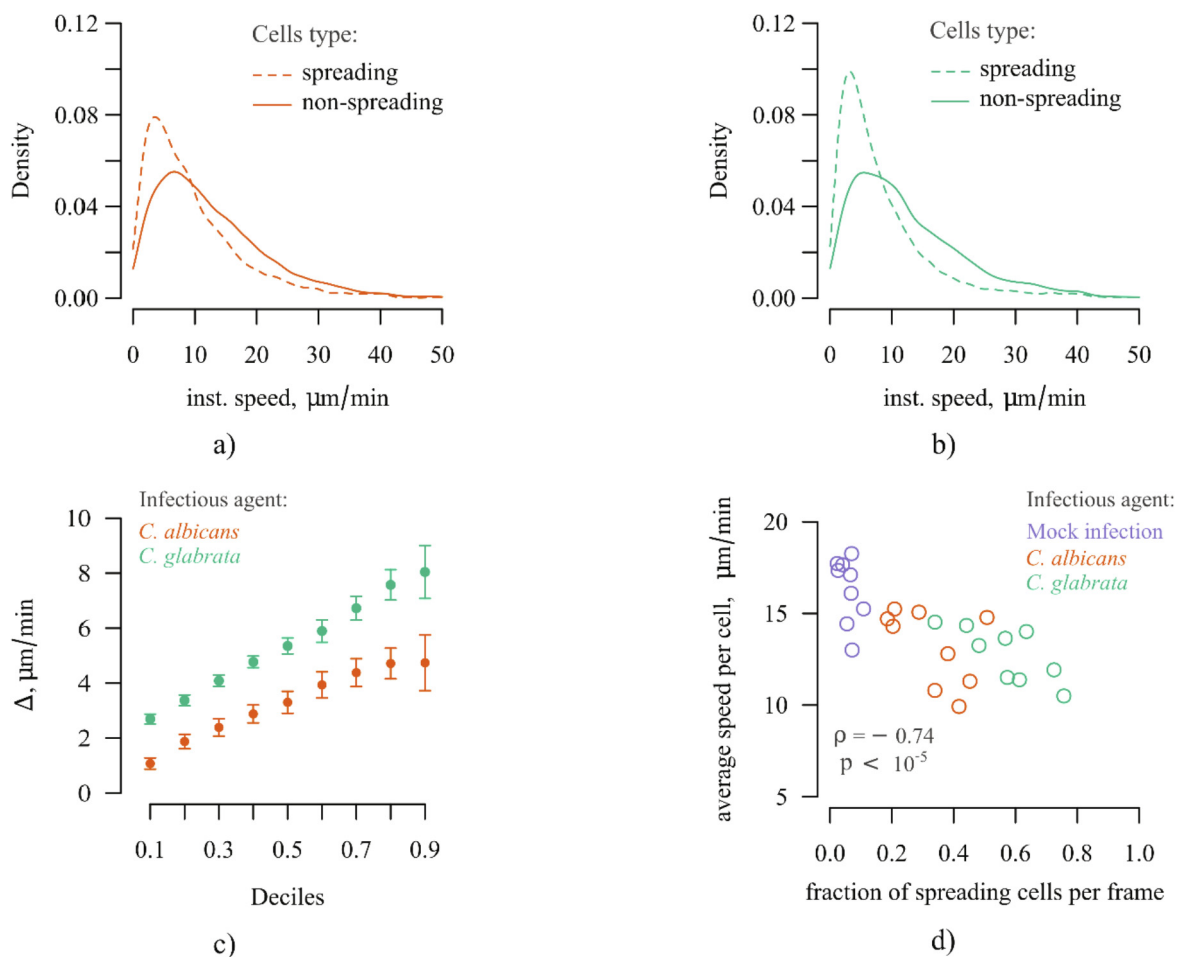


Fig. 5. Distributions of instantaneous speed for spreading and non-spreading neutrophils for a joint sample sets after confrontation with (a) *C. albicans* and (b) *C. glabrata*. Each joint sample set (represented by an individual curve) includes 9×10^3 data points composed of data from randomly selected 1×10^3 spreading cells (dashed lines) or an equal amount of non-spreading cells (solid lines) from each video. c) Shift functions presented by the difference between deciles of distributions in (a) and (b), respectively. The whiskers indicating the 0.95 bootstrap CI (for details see subsection Comparison of cell characteristics for different infection scenarios in the Materials and Methods section). d) Scatter diagram demonstrating the correlation between the median fraction of spreading cells per frame and median average speed per cell for the same sample.

3.5. Evidence for the existence of pathogen-specific morphodynamics of neutrophils

Next, we computed neutrophil morphodynamic features based on the information from the previous classification of neutrophil morphology states. Using this information every tracked cell can be characterised by a frequency of transitions to the spreading state, the total amount of time a cell exists in that state, and the duration of its longest spreading episode. Here, in order to reduce too short cell tracks and by that the noise in the data, we restricted the analysis to cells that were observed for at least 90 s (13 frames) and that switched at least once to the spreading morphology with a maximal duration of at least 28 s (4 frames). These restrictions excluded only 20% of neutrophils in samples infected by the *Candida* species (see Supplementary Fig. 5). As shown in Fig. 6, the visual impression that neutrophils tend to exhibit the S-morphology for longer episodes after confrontation with *C. glabrata* compared to the infection with *C. albicans* could be quantitatively confirmed. As can be seen in Fig. 6, the total duration of spreading episodes per track (Fig. 6b) and the maximal duration of spreading episodes per track (Fig. 6c) showed statistical differences between infection scenarios. These were considered relevant for distinguishing between infection scenarios, although these characteristics may be susceptible to donor-specific variability (see

Supplementary Fig. 6). To perform the classification task based on neutrophil morphodynamics, we used a combination of naïve Bayes classifier for individual track classification followed by a majority voting for the whole sample classification (for details see subsection 2.11). Thus, a test sample was assigned to one of the infection scenarios based on majority fraction of tracked cells being identified as *C. albicans*-specific or *C. glabrata*-specific. For classifier evaluation, we used the LOOCV per condition sampling procedure (with 9^2 iterations in total).

Using this morphodynamics-based classification we reached OSCR = 1 for the *C. albicans*-infected samples, which is higher than in the population-based analysis of snapshots. However, for the *C. glabrata*-infected samples the OSCR remains roughly the same: OSCR = 0.63. Since the maximal duration of spreading episodes per track is a characteristic that is robust against track fragmentation, which may be caused by track interruptions due to long-lasting clusters, we also used this feature alone in the Bayesian classifier. The OSCR raised to OSCR = 0.78 for *C. glabrata*, while the quantitative results for *C. albicans* remained the same (Fig. 7a). Finally, in Fig. 7b the typical detected fraction (mean value over all iterations) of cells with morphodynamics specific for the true infection scenario is shown for every donor and infection scenario. In addition, for every infection sample we performed an interval estimation (defined via Wilson’s confidence

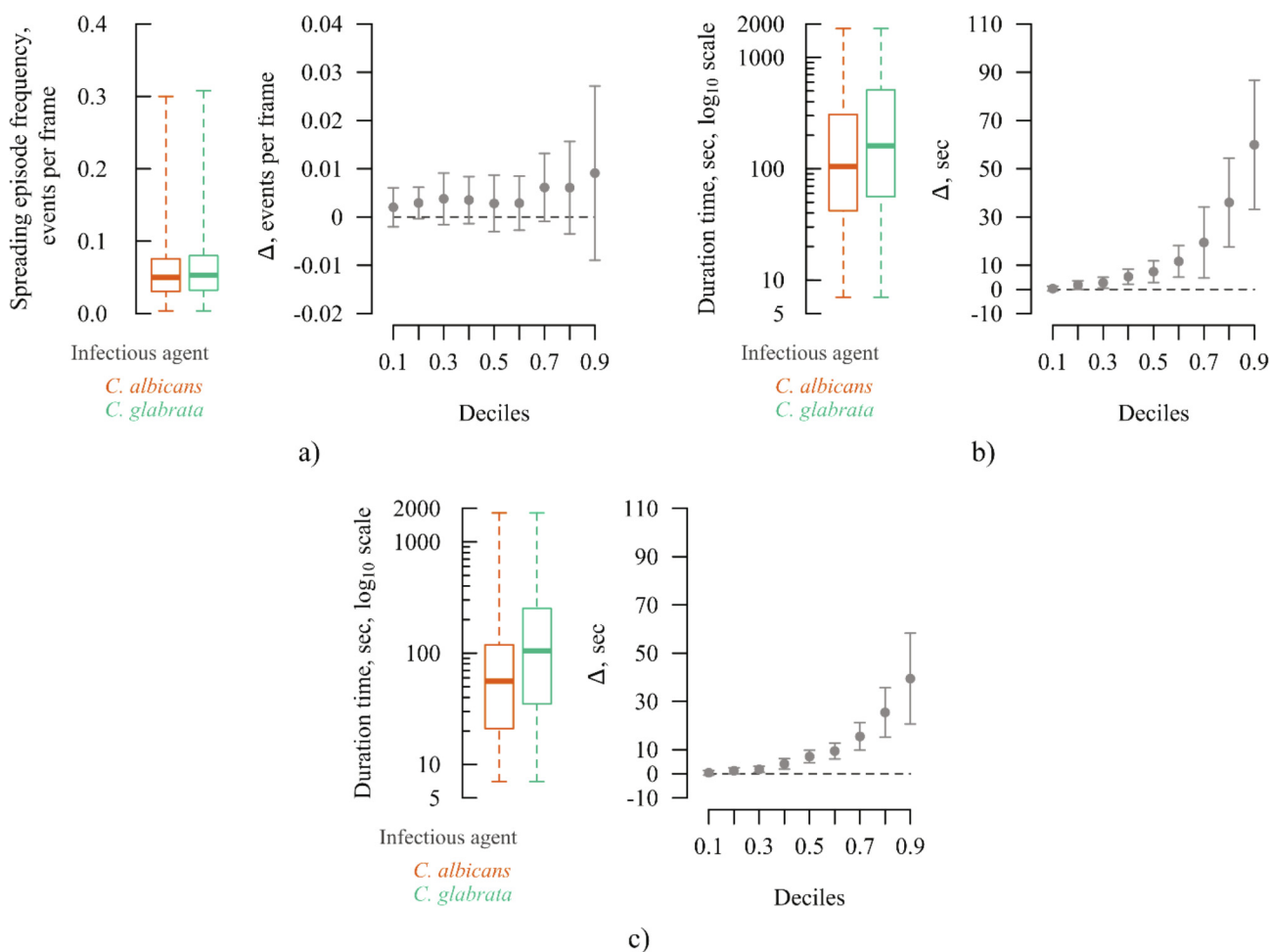


Fig. 6. Comparison of morphodynamics descriptors for joint populations of *C. albicans*- or *C. glabrata*-infected neutrophils by a box plot with whiskers indicating the whole range of values as well as decile-difference diagrams with whiskers indicating 0.95 bootstrap CI (see subsection *Comparison of cell characteristics for different infection scenarios* in the Materials and Methods section). All diagrams were built using balanced sampling (for details see subsection *Data set organisation and sampling procedures* in the Materials and Methods section). a) Distributions of the normalised number of transitions between non-spreading and spreading state. b) Distributions of the total amount of time that cells remain in state with S-morphology. c) Distributions of durations of the longest spreading episode per cell track.

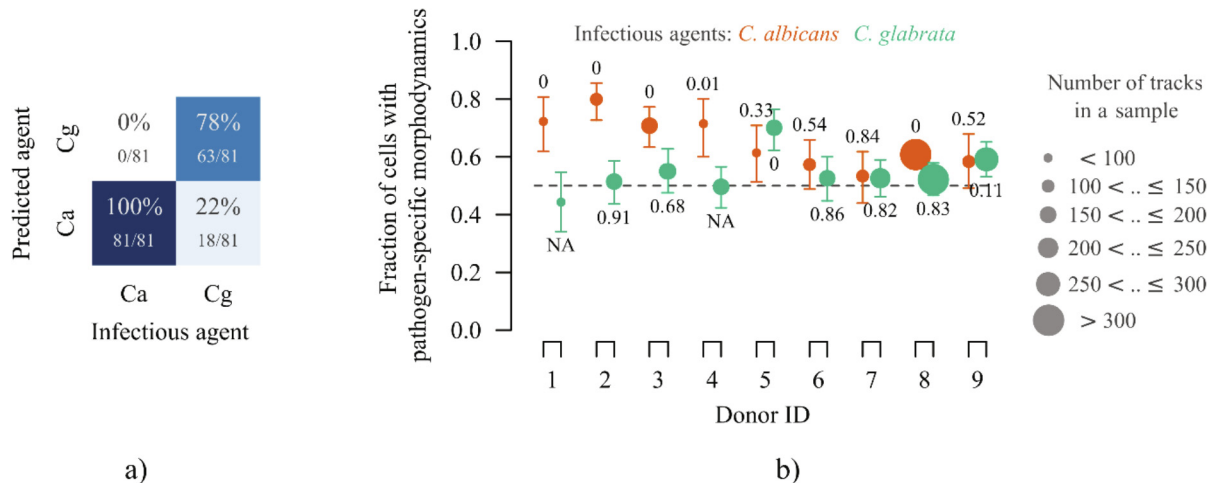


Fig. 7. a) Confusion matrix for sample classification results based on the fraction of neutrophil tracks with pathogen-specific morphodynamics. b) The typically detected fraction of cells with pathogen-specific morphodynamics in a given sample. The mean value is computed over all iterations and the whiskers indicate 0.95 CI for the detected fraction (see *Confidence intervals for proportions* in the Materials and Methods section). The number indicates a probability of the error type II for fraction of neutrophils with *C. albicans*- or *C. glabrata*-specific morphodynamics in a given sample. The symbol NA was used where the computation of this probability is not possible. For further details see subsection *Post-hoc analysis of errors of type II* in the Materials and Methods section.

interval, for details see subsection 2.15) of the fraction of cells with morphodynamics that can be considered specific for the true infection scenario. For instance, for the *C. glabrata*-infected sample from the donor with ID 2, which was represented by about 200 cells in the video, ~52% of all cells were characterised by the morphodynamics analysis to be specific for the *C. glabrata* rather than the *C. albicans* infection scenario. However, the corresponding confidence interval extends to values below 50% indicating that there is a probability that the fraction of cells in the whole blood sample with *C. glabrata*-specific morphodynamics equals that of *C. albicans*. We in fact estimated this probability to equal 91% in the case of blood donor with ID 2 (for details see subsection 2.16). In contrast, for blood donors with ID 5 and ID 9, this probability is estimated to be 0% and 11%, respectively. Moreover, for six out of the nine blood donors, the probability of having an equal number of cells showing *C. glabrata* and *C. albicans* morphodynamics for a true infection scenario with *C. albicans* is well below 35%.

Taken together, while we cannot rule out misclassifications of infection scenarios with *C. albicans* and *C. glabrata*, taking into account the morphodynamics of neutrophils does improve the classification accuracy (Fig. 7b) compared to the static analysis (Fig. 3c) from 67% to 78% for *C. glabrata* and from 89% to 100% for *C. albicans*.

3.6. White-box approach passes deep neural network challenge

Finally, we challenged our white-box approach for identifying the pathogen-specific morphodynamics of neutrophils based on the two descriptors ‘neutrophil footprint area’ and ‘intracellular intensity-gradient’. To this end, we applied state-of-the-art deep neural network technique by evaluating the classification results of a long short-term memory (LSTM) network [60] for data from the nine donors with a leave-one-out cross-validation (LOOCV) [41] (see *Blood sample classification using Deep Learning techniques* in the Supplementary materials). Overall, we achieved test accuracies (ACC) well below 65% from 7000 image sequences obtained from each class (two infection scenarios and the mock-infected samples). In particular, this LSTM-based-approach yielded only moderate values of ACC = 0.7 for mock-infected samples,

ACC = 0.63 for *C. albicans*-infected and ACC = 0.48 for *C. glabrata*-infected samples. The corresponding confusion matrix (Supplementary Fig. 8e) reveals the difficulties of the LSTM to discriminate the neutrophil morphodynamics between infection scenarios with *C. glabrata* and *C. albicans*.

4. Discussion

The application of imaging technologies is an essential component of disease diagnosis and treatment monitoring of patients with life-threatening bloodstream infections. It encompasses a wide range of tools and methods utilised to examine an organism at different levels ranging from the detection of infection foci in a whole organ by computed tomography to identifying pathogens by means of microscopy with high spatial and temporal resolution. In particular, modern label-free methods have a promising potential in the future, among which various types of spectral imaging including Raman spectroscopy, Fourier transform infrared (FTIR) spectroscopy, or matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-ToF MS) providing information about the molecular composition of individual cells. Even though the efficiency of these spectroscopy techniques has been demonstrated for fungal cultures (e.g., [61–63]), limitations for their application in the rapid identification of pathogens in human blood remain. In particular, methods based on cell cultivation require more than 24–48 h, which can lead to fatal delays in initiating pathogen-specific therapy. Moreover, for pathogen concentrations in the sample well below the observed median concentration in clinical samples, which is only 1 CFU/mL for *Candida* bloodstream infections [64], cell cultivation for pathogens may not even be successful. Therefore, since immune cells like neutrophils must have sensed the infection-causing pathogens in patient blood, these interactions are hypothesised to induce measurable changes in the readily available neutrophils that may allow for the indirect identification of pathogens. For example, as was recently shown applying Raman spectroscopy, neutrophils that were first isolated from human whole-blood and subsequently confronted *in vitro* with Gram-positive bacteria (*Staphylococcus aureus*), Gram-negative bacteria (*Escherichia coli*), and fungal pathogens (*C. albicans*) could be distinguished by their molecular

fingerprint [65]. The present study advances along these lines while optimizing various aspects: (i) we investigated activation phenotypes of isolated neutrophils after confrontation with pathogens in the human whole-blood assay to more realistically mimic pathogen detection for bloodstream infections, (ii) we focused on fungal pathogens and the distinction of two species from the *Candida* genus that require different treatment strategies, and (iii) we decided for the commonly available imaging technique of time-lapse microscopy to investigate the pathogen-specific morphodynamics of neutrophils as activation phenotype.

We developed an effective method for the automated comparative analysis of morphological and behavioural changes in neutrophils using live-cell imaging data. As a model system we used the human WBI assays with two most common fungal bloodstream pathogens – *C. albicans* and *C. glabrata*. We started with visual inspection of the acquired videos revealing that in *Candida*-infected samples neutrophils with spreading (S-) morphology appear more often, whereas in mock-infected samples neutrophils with non-spreading (N-) morphology is the dominant morphotype. Based on this observation, we constructed an N-morphology detector (one-class classifier), which was calibrated fully automatically and, therefore, free from operator errors. We could also demonstrate that the classifier outcome is weakly depending on the cells used in the calibration (Fig. 2 and Supplementary Fig. 4). Using this classifier, we were able to estimate the fraction of neutrophils with S-morphology over the whole observation period (Fig. 3). In addition, we showed that the fraction of neutrophils with S-morphology is statistically higher for infection scenarios with *C. glabrata* (Fig. 3b), suggesting the possibility for rapid differentiation between blood samples infected by the two *Candida* species (Fig. 3c and d).

Based on the classifier outcome and the tracking data, we performed an extended analysis of the behaviour of neutrophils from *Candida*-infected samples. We showed that in our experimental conditions the average neutrophil speed per sample is not a reliable marker of infection (Fig. 4), although there is a difference in neutrophil speed when comparing N- and S-morphology (Fig. 5). In contrast, regarding the morphodynamics of neutrophils, we quantitatively confirmed the observation that long-lasting spreading episodes are more often appearing for infection scenarios with *C. glabrata* than *C. albicans* (Fig. 6), which leads to the improvement of infection recognition in our WBI assays (Fig. 7a). However, we cannot assert an observation of pathogen-specific morphodynamics of neutrophils unequivocally due to sample and donor variabilities (Fig. 7b) as well as the number of blood samples. For example, as indicated by a power analysis (significance level $\alpha = 0.05$, expected power of 0.8), approximately 1000 blood donors would be required for a statistically definitive conclusion that at least 54% of neutrophils, which corresponds to the average fraction detected in our experiments, exhibit *C. glabrata*-specific morphodynamics in a *C. glabrata*-infected sample. While recruiting this large number of blood donors is clearly beyond the scope of our feasibility study, we inferred the following analysis pipeline for the best classification results: (i) calibration of the one-class classifier based on static features of neutrophils from non-infected samples, (ii) classification of samples being infected or not based on the fraction of spreading cells, (iii) including neutrophil morphodynamics to distinguish between samples from different infection scenarios.

In this study, we applied our Algorithm for Migration and Interaction Tracking (AMIT, [13,14,19]) in its latest release of the third version [20]. The performance of AMIT with regard to the automated segmentation and tracking of label-free neutrophils was previously found to outcompete established learning-based algorithms [20], such as MU_Lux-CZ [66] and SegNet [67]. Nevertheless, in the present study we checked whether deep neural

networks can improve the distinction of infection scenarios by *C. albicans* and *C. glabrata* based on a long short-term memory (LSTM) network [60], which we applied to classify the time-series of neutrophils with dynamically changing morphology. However, this black-box-approach yielded relatively moderate test accuracies (ACC) with values well below 65% compared to our white-box approach that is based on the two descriptors ‘neutrophil footprint area’ and ‘intracellular intensity-gradient’ and achieved values well above 75% for the two infection scenarios and 100% for mock-infected samples. We speculate that this may be explained by peculiarities of the LSTM network architecture, which may be unable to grasp sufficient information about aperiodic spreading events from relatively short sequences.

In further studies, instead of increasing the complexity of analysis pipelines, we consider improving the cell description by adding information about intensity and amount of neutrophil-derived trail formation [68,69] and neutrophil autofluorescence [70,71]. This could be tested after modification of the image acquisition step implying detection of transmitted light images by a high-resolution camera with a high readout speed (or global shutter) at intervals of one second or shorter. This would allow eliminating cell-movement-associated blurring effects and by that improve the accuracy in image processing with regard to cell segmentation and tracking as well as morphological information. Besides, it would pave the way for detailed analysis of dynamic transitions between the two states of N- and S-morphology, as well as performing comparative analyses of potentially different S-morphologies under various conditions. In addition, high-speed imaging would enable estimating within-donor heterogeneity, which is particularly essential regarding the neutrophil population microheterogeneity, *i.e.* existence of neutrophil sub-sets with different functions, distinct morphology as well as receptor repertoires [72–74]. Extension of this feasibility study to a larger cohort of blood donors and inclusion of BSI patients will be the next step for exploring the potential of this approach for translational research.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Leibniz ScienceCampus InfectoOptics Jena, which is financed by the funding line Strategic Networking of the Leibniz Association. Furthermore, we received financial support from the German Research Foundation (DFG) through the CRC/TR 124 FungiNet with project number 210879364 (Project B4 to MTF, Project C3 to OK). This work was supported by the Federal Ministry of Education and Research, Germany (grant number 13GW0456B) in the context of InfectoGnostics Research Campus Jena.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2022.05.007>.

References

- [1] Wisplinghoff H, Bischoff T, Tallent SM, Seifert H, Wenzel RP, Edmond MB. Nosocomial bloodstream infections in US hospitals: Analysis of 24,179 cases from a prospective nationwide surveillance study. *Clin Infect Dis* 2004;39:309–17.

- [2] Kett DH, Azoulay E, Echeverria PM, Vincent J-L. Candida bloodstream infections in intensive care units: Analysis of the extended prevalence of infection in intensive care unit study. *Crit Care Med* 2011;39:665–70.
- [3] Horn DL, Fishman JA, Steinbach WJ, Anaissie EJ, Marr KA, Olyaei AJ, et al. Presentation of the PATH Alliance® registry for prospective data collection and analysis of the epidemiology, therapy, and outcomes of invasive fungal infections. *Diagn Microbiol Infect Dis* 2007;59:407–14.
- [4] Perlroth J, Choi B, Spellberg B. Nosocomial fungal infections: Epidemiology, diagnosis, and treatment. *Med Mycol* 2007;45:321–46.
- [5] Hünninger K, Lehnert T, Bieber K, Martin R, Figge MT, Kurzai O. A Virtual Infection Model Quantifies Innate Effector Mechanisms and Candida albicans Immune Escape in Human Blood. *PLoS Comput Biol* 2014;10:e1003479.
- [6] Echenique-Rivera H, Muzzi A, Del Tordello E, Seib KL, Francois P, Rappuoli R, et al. Transcriptome Analysis of Neisseria meningitidis in Human Whole Blood and Mutagenesis Studies Identify Virulence Factors Involved in Blood Survival. *PLoS Pathog* 2011;7:e1002027.
- [7] Tena GN, Young DB, Eley B, Henderson H, Nicol MP, Levin M, et al. Failure to Control Growth of Mycobacteria in Blood from Children Infected with Human Immunodeficiency Virus and Its Relationship to T Cell Function. *J Infect Dis* 2003;187:1544–51.
- [8] Plested JS, Welsch JA, Granoff DM. Ex vivo model of meningococcal bacteremia using human blood for measuring vaccine-induced serum protective activity. *Clin Vaccine Immunol* 2009;16:785–91.
- [9] Sprong T, Brandtzaeg P, Fung M, Pharo AM, Høiby EA, Michaelsen TE, et al. Inhibition of C5a-induced inflammation with preserved C5b-9-mediated bactericidal activity in a human whole blood model of meningococcal sepsis. *Blood* 2003;102:3702–10.
- [10] Lehnert T, Timme S, Pollmächer J, Hünninger K, Kurzai O, Figge MT. Bottom-up modeling approach for the quantitative estimation of parameters in pathogen-host interactions. *Front Microbiol* 2015;6:608.
- [11] Duggan S, Essig F, Hünninger K, Mokhtari Z, Bauer L, Lehnert T, et al. Neutrophil activation by Candida glabrata but not Candida albicans promotes fungal uptake by monocytes. *Cell Microbiol* 2015;17:1259–76.
- [12] Essig F, Hünninger K, Dietrich S, Figge MT, Kurzai O. Human neutrophils dump Candida glabrata after intracellular killing. *Fungal Genet Biol* 2015;84:37–40.
- [13] Brandes S, Mokhtari Z, Essig F, Hünninger K, Kurzai O, Figge MT. Automated segmentation and tracking of non-rigid objects in time-lapse microscopy videos of polymorphonuclear neutrophils. *Med Image Anal* 2015;20:34–51.
- [14] Brandes S, Dietrich S, Hünninger K, Kurzai O, Figge MT. Migration and interaction tracking for quantitative analysis of phagocyte-pathogen confrontation assays. *Med Image Anal* 2017;36:172–83.
- [15] Lehnert T, Prauße MTE, Hünninger K, Praetorius JP, Kurzai O, Figge MT. Comparative assessment of immune evasion mechanisms in human whole-blood infection assays by a systems biology approach. *PLoS ONE* 2021;16:e0249372.
- [16] Prauße MTE, Lehnert T, Timme S, Hünninger K, Leonhardt I, Kurzai O, et al. Predictive Virtual Infection Modeling of Fungal Immune Evasion in Human Whole Blood. *Front Immunol* 2018;9:560.
- [17] Timme S, Lehnert T, Prauße MTE, Hünninger K, Leonhardt I, Kurzai O, et al. Quantitative simulations predict treatment strategies against fungal infections in virtual neutropenic patients. *Front Immunol* 2018;9:667.
- [18] Lehnert T, Leonhardt I, Timme S, Thomas-Rüddel D, Bloos F, Sponholz C, et al. Ex vivo immune profiling in patient blood enables quantification of innate immune effector functions. *Sci Reports*. 2021. 111. 2021. 11. 1–16.
- [19] Al-Zaben N, Medyukhina A, Dietrich S, Marolda A, Hünninger K, Kurzai O, et al. Automated tracking of label-free cells with enhanced recognition of whole tracks. *Sci Rep* 2019;9:3317.
- [20] Belyaev I, Praetorius JP, Medyukhina A, Figge MT. Enhanced segmentation of label-free cells for automated migration and interaction tracking. *Cytom Part A*. 2021. cyto.a.24466.
- [21] Hoffstein ST, Friedman RS, Weissmann G. Degranulation, membrane addition, and shape change during chemotactic factor-induced aggregation of human neutrophils. *J Cell Biol* 1982;95:234–41.
- [22] Wang L, Castro CE, Boyce MC. Growth strain-induced wrinkled membrane morphology of white blood cells. *Soft Matter* 2011;7:11319–24.
- [23] Jumaa MAA, Dewitt S, Hallett MB. Topographical interrogation of the living cell surface reveals its role in rapid cell shape changes during phagocytosis and spreading. *Sci Rep* 2017;7:1–11.
- [24] Schuster M, Moeller M, Bornemann L, Bessen C, Sobczak C, Schmitz S, et al. Surveillance of Myelodysplastic Syndrome via Migration Analyses of Blood Neutrophils: A Potential Prognostic Tool. *J Immunol* 2018;201:3546–57.
- [25] Seider K, Brunke S, Schild L, Jablonowski N, Wilson D, Majer O, et al. The facultative intracellular pathogen Candida glabrata subverts macrophage cytokine production and phagolysosome maturation. *J Immunol* 2011;187:3072–86.
- [26] Gonzalez RC, Woods RE. Digital Image Processing. 3rd ed. Pearson; 2007.
- [27] Zuiderveld K. Contrast Limited Adaptive Histogram Equalization. *Graph. Gems*. Elsevier. 1994. 474–85.
- [28] Sobel I. An Isotropic 3x3 Image Gradient Operator. *Present Stanford AI Proj*. 1968. 2014.
- [29] Pomerantsev AL. Acceptance areas for multivariate classification derived by projection methods. *J. Chemom*. 22. John Wiley & Sons, Ltd. 2008. 601–9.
- [30] Pomerantsev AL, Rodionova OY. Concept and role of extreme objects in PCA/SIMCA. *J Chemom* 2014;28:429–38.
- [31] Pomerantsev AL, Rodionova OY. Popular decision rules in SIMCA: Critical review. *J Chemom* 2020;34.
- [32] WOLD S, SJÖSTRÖM M. SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy. 1977. 243–82.
- [33] Tsugawa H, Tsujimoto Y, Arita M, Bamba T, Fukusaki E. GC/MS based metabolomics: development of a data mining system for metabolite identification by using soft independent modeling of class analogy (SIMCA). *BMC Bioinf* 2011;12:131.
- [34] Khanmohammadi M, Nasiri R, Ghasemi K, Samani S, Bagheri GA. Diagnosis of basal cell carcinoma by infrared spectroscopy of whole blood samples applying soft independent modeling class analogy. *J Cancer Res Clin Oncol* 2007;133:1001–10.
- [35] Masithoh RE, Yuliyanda I. NIR reflectance spectroscopy and SIMCA for classification of crops flour. *IOP Conf. Ser. Earth Environ. Sci.*, vol. 355, Institute of Physics Publishing. 2019. 012004.
- [36] Dröge JBM, Rinsma WJ, Van 'T Klooster HA, Tas AC, Van Der Greef J. An evaluation of SIMCA. Part 2 – classification of pyrolysis mass spectra of pseudomonas and serratia bacteria by pattern recognition using the SIMCA classifier. *J Chemom*. 1987. 1. 231–41.
- [37] Mazivila SJ, Páscoa RNMJ, Castro RC, Ribeiro DSM, Santos JLM. Detection of melamine and sucrose as adulterants in milk powder using near-infrared spectroscopy with DD-SIMCA as one-class classifier and MCR-ALS as a means to provide pure profiles of milk and of both adulterants with forensic evidence: A short communication. *Talanta*. 2020. 216. 120937.
- [38] Neves MDG, Poppi RJ. Authentication and identification of adulterants in virgin coconut oil using ATR/FTIR in tandem with DD-SIMCA one class modeling. *Talanta* 2020;219:121338.
- [39] Pearson K. LIII. On lines and planes of closest fit to systems of points in space. *London, Edinburgh, Dublin Philos Mag J Sci* 1901;2:559–72.
- [40] Kucheryavskiy S. mdatools – R package for chemometrics. *Chemom Intell Lab Syst* 2020;198:103937.
- [41] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. New York, NY: Springer, New York; 2009.
- [42] Michal Majka. naivebayes: High Performance Implementation of the Naive Bayes Algorithm in R. 2019.
- [43] Wilcox RR, Erceg-Hurn DM. Comparing two dependent groups via quantiles. *J Appl Stat* 2012;39:2655–64.
- [44] Harrell FE, Davis CE. A new distribution-free quantile estimator. *Biometrika* 1982;69:635–40.
- [45] Mair P, Wilcox R. Robust statistical methods in R using the WRS2 package. *Behav Res Methods*. 2019. 522. 2019. 52. 464–88.
- [46] Quade D. Using weighted rankings in the analysis of complete blocks with additive block effects. *J Am Stat Assoc* 1979;74:680–3.
- [47] Holm S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand J Stat* 1979;6:65–70.
- [48] Pohlert P. PMCMRplus: Calculate Pairwise Multiple Comparisons of Mean Rank Sums Extended. R package version 1.4.4. 2020.
- [49] McGraw KO, Wong SP. A Common Language Effect Size Statistic. *Psychol Bull* 1992;111:361–5.
- [50] Dick J, Bolker B. canprot: Compositional Analysis of Differentially Expressed Proteins in Cancer. 2020.
- [51] Hedges LV, Olkin I. Statistical Methods for Meta-Analysis. Orlando: Academic Press Inc (Verlag); 1985.
- [52] Gibbons RD, Hedeker DR, Davis JM. Estimation of Effect Size From a Series of Experiments Involving Paired Comparisons. *J Educ Stat* 1993;18:271–9.
- [53] Torchiano M. Effsize – A package for efficient effect size computation. 2016.
- [54] Wilson EB. Probable Inference, the Law of Succession, and Statistical Inference. *J Am Stat Assoc* 1927;22:209–12.
- [55] Agresti A, Coull BA. Approximate is better than “Exact” for interval estimation of binomial proportions. *Am Stat* 1998;52:119–26.
- [56] PropCIs SR. Various confidence interval methods for proportions. R Package Version 2018:03.
- [57] Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions. John Wiley & Sons, Inc. 2003.
- [58] Kohl M. MKpower: Power Analysis and Sample Size Calculation. 2020.
- [59] Wilcox RR. Comparing Two Independent Groups Via Multiple Quantiles. *Stat* 1995;44:91–9.
- [60] Hochreiter S, Schmidhuber J. Long Short-Term Memory. *Neural Comput* 1997;9:1735–80.
- [61] Witkowska E, Jagielski T, Kamińska A, Kowalska A, Hryniewicz-Gwóźdź A, Waluk J. Detection and identification of human fungal pathogens using surface-enhanced Raman spectroscopy and principal component analysis. *Anal Methods* 2016;8:8427–34.
- [62] Lam LMT, Dufresne PJ, Longtin J, Sedman J, Ismail AA. Reagent-free identification of clinical yeasts by use of attenuated total reflectance fourier transform infrared spectroscopy. *J Clin Microbiol* 2019;57.
- [63] Normand A-C, Gabriel F, Riat A, Cassagne C, Bourgeois N, Huguenin A, et al. Optimization of MALDI-ToF mass spectrometry for yeast identification: a multicenter study. *Med Mycol* 2020;58:639–49.
- [64] Pfeiffer CD, Samsa GP, Schell WA, Reller LB, Perfect JR, Alexander BD. Quantitation of Candida CFU in initial positive blood cultures. *J Clin Microbiol* 2011;49:2879–83.
- [65] Arend N, Pittner A, Ramoji A, Mondol AS, Dahms M, Rieger J, et al. Detection and Differentiation of Bacterial and Fungal Infection of Neutrophils from Peripheral Blood Using Raman Spectroscopy. *Anal Chem* 2020;92:10560–8.
- [66] Lux F, Matula P. Cell Segmentation by Combining Marker-Controlled Watershed and Deep Learning. 2020.

- [67] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans Pattern Anal Mach Intell* 2017;39:2481–95.
- [68] Youn Y-j, Shrestha S, Lee Y-B, Kim J-K, Lee JH, Hur K, et al. Neutrophil-derived trail is a proinflammatory subtype of neutrophil-derived extracellular vesicles. *Theranostics* 2021;11:2770–87.
- [69] Marki A, Buscher K, Lorenzini C, Meyer M, Saigusa R, Fan Z, et al. Elongated neutrophil-derived structures are blood-borne microparticles formed by rolling neutrophils during sepsis. *J Exp Med* 2021;218.
- [70] Monici M, Pratesi R, Bernabei PA, Caporale R, Ferrini PR, Croce AC, et al. Natural fluorescence of white blood cells: spectroscopic and imaging study. *J Photochem Photobiol B Biol* 1995;30:29–37.
- [71] Monsel A, Lécart S, Roquilly A, Broquet A, Jacqueline C, Mirault T, et al. Analysis of autofluorescence in polymorphonuclear neutrophils: A new tool for early infection diagnosis. *PLoS ONE* 2014;9:e92564.
- [72] Beyrau M, Bodkin JV, Nourshargh S. Neutrophil heterogeneity in health and disease: a revitalized avenue in inflammation and immunity. *Open Biol* 2012;2:120134.
- [73] Silvestre-Roig C, Hidalgo A, Soehnlein O. Neutrophil heterogeneity: Implications for homeostasis and pathogenesis. *Blood* 2016;127:2173–81.
- [74] Hong CW. Current understanding in neutrophil differentiation and heterogeneity. *Immune Netw* 2017;17:298–306.