



Research Article

Machine learning for predicting Plasmodium liver stage development in vitro using microscopy imaging

Corin F. Otesteanu^{a,*}, Reto Caldelari^b, Volker Heussler^b, Raphael Sznitman^{a,*}^a Artificial Intelligence in Medicine group, University of Bern, Switzerland^b Institute of Cell Biology, University of Bern, Switzerland

ARTICLE INFO

Dataset link: <https://doi.org/10.5281/zenodo.10909406>Dataset link: <https://github.com/ml-malaria-prediction/ml-malaria-prediction>

Keywords:

Malaria
Microscopy imaging
Neural networks
Deep learning

ABSTRACT

Malaria, a significant global health challenge, is caused by *Plasmodium* parasites. The *Plasmodium* liver stage plays a pivotal role in the establishment of the infection. This study focuses on the liver stage development of the model organism *Plasmodium berghei*, employing fluorescent microscopy imaging and convolutional neural networks (CNNs) for analysis. Convolutional neural networks have been recently proposed as a viable option for tasks such as malaria detection, prediction of host-pathogen interactions, or drug discovery. Our research aimed to predict the transition of *Plasmodium*-infected liver cells to the merozoite stage, a key development phase, 15 hours in advance. We collected and analyzed hourly imaging data over a span of at least 38 hours from 400 sequences, encompassing 502 parasites. Our method was compared to human annotations to validate its efficacy. Performance metrics, including the area under the receiver operating characteristic curve (AUC), sensitivity, and specificity, were evaluated on an independent test dataset. The outcomes revealed an AUC of 0.873, a sensitivity of 84.6%, and a specificity of 83.3%, underscoring the potential of our CNN-based framework to predict liver stage development of *P. berghei*. These findings not only demonstrate the feasibility of our methodology but also could potentially contribute to the broader understanding of parasite biology.

1. Introduction

Malaria is a dangerous infectious disease caused by a variety of *Plasmodium* species that are transmitted to people through the bites of infected female *Anopheles* mosquitoes. In 2021, nearly half of the world's population was at risk of malaria. That year, there were an estimated 247 million cases of malaria worldwide and the estimated number of malaria deaths stood at 619,000, according to the World Health Organization report from 2021. Studying *Plasmodium* parasite behavior [1], the epidemiology [2], diagnosis [3], treatment and risk factors [4,5] associated with malaria is crucial for effective control and prevention strategies. The progress in microscopy imaging has enabled advanced imaging of the *Plasmodium falciparum* erythrocytic cycle and the *P. berghei* pre-erythrocytic stages [6]. Tracking the development of infected cells is an essential step in understanding parasitic cell biology and host-pathogen interactions [7]. Traditionally, such procedures rely on manual expert evaluation of cell morphology in microscopy images and, therefore, can suffer from subjectivity regarding cell features and limited throughput [8], while an automated system based on tracking cell features over time could mitigate these issues.

Artificial intelligence using machine learning has been effectively used for various tasks in malaria research: *prediction*: predicting human-parasite protein associations using network topological profiles and machine learning [9], parasite protein phosphosite prediction using machine-learning [10]; *detection*: detecting malaria parasites using light microscopy and convolutional neural networks (CNNs) [11,12], using patient information from parasite case reports, and machine learning techniques [13], using patient symptoms and demographic features together with machine learning models [14], using protein sequences and natural language processing techniques [15], or using bead-based antigen detection assay in conjunction with decision trees [16], and *drug discovery* [17].

Most recent studies investigating the *Plasmodium* parasite used neural networks or an ensemble of neural networks [11]. Rajaraman et al. [18] used a pre-trained ResNet CNN model, while Hemachandran et al. used a pre-trained MobileNetV2 [19]. Masud et al. [20] proposed a custom CNN with four convolutional blocks and two fully connected layers, Cho et al. used a custom 3 layer neural network [21] and Alok et al. [22] proposed a custom CNN with 4 convolutional blocks and two fully connected layers for classification of malaria-infected or healthy

* Corresponding authors.

E-mail addresses: corin.osteanul@unibe.ch (C.F. Otesteanu), raphael.sznitman@unibe.ch (R. Sznitman).

blood cells. A hybrid model with Bayesian optimization and Deep Residual CNN features was used for classification of malarial cell images by Di Ker et al. [23]. Single cells in thin blood smear slides were classified as infected or uninfected using a custom 16-layer CNN by Liang et al. [24]. A study from Huang et al. [25] used the Faster R-CNN to distinguish malaria-infected and healthy blood cells, while Fuhat et al. [26] used a custom 3-layer autoencoder. For the same purpose, a watershed algorithm was used together with a multi-layer perception (MLP) on predetermined features by Manning et al. [27], while Dey et al. [28] proposed a ResNet 152 model integrated with a Deep Greedy Network for training. De Souza et al. [11] proposed using the HSV color components as inputs to an MLP for pixel-wise segmentation together with a custom 34-layer CNN for the malaria classification tasks in thick blood smear images. Arshad et al. [29] proposed a U-net CNN together with a ResNet for cell localization and classification of the *Plasmodium* life-cycle stage in 4 classes. Similarly, Yang et al. [30], developed a framework to segment the *Plasmodium* parasite from thin film images and to classify its species among four dominant classes using a U-net together with a custom CNN with 4 convolutional blocks and two fully connected layers.

Upon the bite of an infected *Anopheles* mosquito *Plasmodium* sporozoites migrate from the skin to the liver, where they establish their initial site of replication within the host [31]. Therefore, understanding the liver stage development of the parasite is of great importance [32]. Upon rupture of the parasitophorous vacuole membrane, liver-stage merozoites were shown to destabilize and substantially alter the protein composition of the host cell membrane and induce separation of the host cell actin cytoskeleton from it [33]. Moreover, Burda et al. [34] tracked the development of the *Plasmodium* parasite plasma membrane, revealing the formation of membrane contact sites with the endoplasmic reticulum, which may facilitate lipid delivery to support the expansion of the plasma membrane during the parasite's life cycle. De Niz et al. have shown that the parasite fragments the host cell Golgi into miniaturized stacks, to optimize its own intracellular development [35]. Niklaus et al. [36], characterized lysosomal interactions with the parasitophorous vacuole and provided insights into cellular details of intracellular killing and lysosomal elimination of *Plasmodium* parasites independent of immune system cells.

In this study, we delve into the liver stage development of the *P. berghei* parasite and examine the host-parasite interactions using HeLa cells. HeLa cells, derived from a human cervical cancer cell line [37], have been extensively utilized for *P. berghei* infections, as demonstrated in our recent publications [38–40]. This research is primarily focused on in vitro analysis and does not extend to in vivo applications. The reason is that traditional in vivo methodologies, such as intravital microscopy, are inherently invasive and applicable only for short durations, whereas bioluminescent imaging, though less invasive, does not provide detailed information on individual parasites. This study pioneers the integration of fluorescent microscopy with artificial intelligence to meticulously track and predict the developmental milestones of *Plasmodium* liver stage development.

2. Material and methods

2.1. Animal ethics statement

Animal studies were carried out under the approval of the Animal Research Ethics Committee of Canton Bern, Switzerland (License Number: BE86/19). Balb/c mice were purchased from Janvier Labs (Saint Berthevin, France) or bred in-house and used from 8–10 weeks of age.

2.2. Imaging

3D image acquisitions (10 μm) of infected cells were acquired hourly, starting at around 24 hours post-infection (hpi), using a fully automated Nikon CSU-W1 spinning disc (4000 rpm) confocal microscope

set at 37 °C and 5% CO₂ (Omicron Laser 561 nm, 100 ms exposures with 2x2 binning and 16-bit depth on Photometrics Prime BSI CMOS camera, Plan Apo λ 40 \times (NA 0.95) objective). The resulting images were $N_x \times N_y \times N_z$. Specifically, $N_x = 1024$, $N_y = 1024$, and $N_z = 6$, correspond to the image's height, width, and depth, respectively, in pixels. At each acquisition, a 2D brightfield widefield image was acquired with an image size of $N_x \times N_y$ pixels.

2.3. Parasite infections

Blood stabilates of *P. berghei* constitutively expressing mCherry [41] were used to infect mice. Upon sufficient sexual forms of the parasites (gametocytes), mice were anesthetized using Ketamin/Xylazine. Female *Anopheles stephensi* mosquitoes were allowed to feed on the anesthetized mice for 30 minutes. Mosquitoes were kept at 20.5 °C and > 80% relative humidity and fed daily with 8% fructose-containing 0.2% para-aminobenzoic acid. From day 18 to 26 post feeding salivary glands of infected mosquitoes were dissected and used for liver-stage infections. Sporozoites were released using a pellet pestle mixer. Around 20,000 Sporozoites released from salivary glands were used to infect confluent HeLa cells (40,000 cells/96well in MEM10%FCS, PenStrep seeded the day before). Two hours post-infection the cells were detached with accutase (Innovative Cell Technologies) and reseeded onto 24 well-glass bottom well plates and incubated at 37 °C/ 5% CO₂. Medium was replaced the next day (2 ml).

2.4. Data annotation

To help with the data annotation process, a relevance filter was implemented to filter out cells with living parasites in the last acquisition hours. Binary thresholding based on (expected) signal-to-noise ratio was used, with noise determined from empty images. In the post filtering images, the cell structure was visually inspected for manual annotation. Parasite shape, elevation, and merozoite formation were the assessed visual cues.

2.5. Parasite segmentation and deep learning classification model

Throughout this paper, we use the term *capture position* to denote the same 3D volume location acquired at every time point (hour) during the acquisition process. Therefore, each capture position may contain more than one infected cell, as it was not possible to control the number of infected cells at a certain location. Every capture position had a size of $N_x \times N_y \times N_z \times N_t$, where N_t is the number of acquired volumes. Adaptive binary thresholding was used, and small groups of pixels (<40) equivalent to less than 4.43 μm were filtered out and considered noise. Peak detection was used for single cell localization and k-means clustering for cell boundary delineation. Parasite features such as position, volume, signal intensity, and bright elements inside the infected cell - a proxy for merozoites, and others, were calculated. Features were weighted according to their consistency for each parasite in consecutive time volumes. Using their position (x,y,z) and parasite volume, parasites were tracked across volumes with the Hungarian Algorithm for Linear Assignment Problem [42]. Throughout this paper, we use the term *successfully developing* to denote parasites that reach the merozoite stage of development and form detached cells by the last acquired volume. The term *non-developing* was used to denote parasites that died and lost fluorescent signal - or those that did not reach the merozoite stage.

Time series binary classification was used to distinguish between successfully developing and non-developing parasites from every time-volume. For this, long short-term memory recurring neural networks were used, capable of learning long-term dependencies in time series sequence prediction problems. To handle data imbalances caused by a different number of images for each training class, the minority class was oversampled to that of the majority class during the training process.

To achieve this, random oversampling was used by randomly selecting samples from the minority class (with replacement) and adding them to the training dataset. To limit neural network overfitting, a dropout layer of 5% was used after the LSTM layer which consisted in 400 hidden units. Afterward, a fully connected layer was followed by a softmax activation function, which would output the score our model gives for a particular input image pertaining to a class. To train the model, stochastic gradient descent [43] was used to optimize the binary cross entropy loss function (1).

$$\text{crossentropy} = -(y \log(p) + (1 - y) \log(1 - p)), \quad (1)$$

where y is the binary indicator (0 or 1) of the class label, and p is the predicted probability.

2.6. Evaluation metrics and statistical analysis

To evaluate the performance of our model, we compute various measures, including the area under the receiver-operating-characteristic curve (AUC), sensitivity = $\frac{TP}{TP+FN}$, specificity = $\frac{TN}{FP+TN}$, and weighted accuracy. Here, TP represents the true positives, TN the true negative, FP the false positives, and FN the false negatives. The weighted accuracy is defined as the sum of the proportion of correct classifications performed by our model in each class, weighted by the number of images in each class.

$$\text{acc}_w = \frac{1}{n} \sum_{c=1}^n \sum_{o=1}^{n_c} \frac{y_{o,c}}{n_c}, \quad (2)$$

where n_c is the number of images in class c and $y_{o,c}$ is the binary indicator (0 or 1) if class label c is the correct classification for the observed image o . The Sorensen–Dice coefficient was used to estimate the similarity between the ground truth segmented data and the results of the automatized segmentation.

$$\text{Dice}(A, B) = \frac{2|A \cdot B|}{|A| + |B|}, \quad (3)$$

where A and B are binary vectors with positive values for elements inside the group and 0 otherwise, and signify the ground truth and segmentation result, respectively. The Permutation Importance method [44] was used to measure the influence that the features of our data (i.e., parasite properties) have on the predictions of the machine learning model. It consists of shuffling the values of each feature and calculating the decrease in the model's performance (4).

$$i_j = s - \frac{1}{K} \sum_{k=1}^K s_{k,j}, \quad (4)$$

where i_j is the importance of feature f_j , K is the number of times the shuffle-evaluation process is repeated, and $s_{k,j}$, the performance of the model on the shuffled feature j for repetition k . The kernel density estimation, a non-parametric approach for estimating the probability density function of a population, was used to model the probability density function for each parasite parameter across the positive and negative classes. The Hellinger distance, a type of f-divergence function that measures the difference between two probability distributions was used to quantify the similarity between positive and negative class parasite parameters.

$$H(P, Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^k (\sqrt{p_i} - \sqrt{q_i})^2}, \quad (5)$$

where P and Q represent the probability distribution of the positive and negative class parasite features.

Table 1
Dataset distribution overview.

	Positive samples	Negative samples	Total
Training sets	49 - 50	200	249 - 250
Training sets [%]	19.7 - 20	80 - 80.3	-
Validation sets	12 - 13	50	62 - 63
Validation sets [%]	19.4 - 20.6	79.4 - 80.6	-
Total train_val	62	250	312
Total train_val [%]	19.9	80.1	-
Test set	61	129	190
Test set [%]	32.1	67.9	-
Total sets	123	379	502
Total sets [%]	24.5	75.5	-

3. Results

3.1. Dataset properties

We collected data from 400 capture positions yielding 502 infected cells. Among these, 123 cells harbored developing parasites, while the remaining 379 did not exhibit progression to the final developmental stage. These images were used to train, validate, and test the performance of the deep learning system. The system was trained and validated on 250 capture positions consisting of 312 cells. An additional set of 150 capture positions collected at the same research facility but in a separate experiment and at a separate time consisting of 190 cells was used as a testing set. An overview of training, validation, and (holdout) test dataset population is summarized in Table 1. The infection rate was consistent between the two datasets.

The typical progression of parasite development, illustrated in Fig. 1A, is depicted through z-projections of fluorescent signals captured at key post-infection intervals. Initially, the parasite's growth is predominantly two-dimensional, as shown in the first row, followed by an intensification of the fluorescent signal. By 55 hours post-infection (hpi), segmentations indicative of the cytomere stage become evident, leading to further segmentation into merozoites by 65 hpi, marking the final merozoite stage with dispersed merozoites within the host cell cytoplasm. Contrastingly, Figs. 1B-D present capture positions where parasites fail to progress to this final stage, highlighting the variability in development outcomes. Each panel in Fig. 1 represents the development of a single parasite over time. Further elaboration on these observations, including supplementary fluorescence images juxtaposed with brightfield images, is provided in Fig. S1, showcasing a range of infected cell conditions.

The process of manual annotation and prediction was focused on the 55 hpi mark, a critical juncture expected to showcase the cytomere stage. The absence of segmentation at this stage signals developmental anomalies, serving as a basis for evaluating our deep learning model. The Dice score metrics was used to gauge the similarity between the manual segmentation, used as ground truth, and those generated by the automatized system. A subset of 600 images, equally divided between positive and negative classes, was manually segmented to generate ground-truth examples. The mean computed Dice score for the positive class was 0.86 (0.054 standard deviation) and 0.89 (0.075 s.d.) for the negative class. Using the segmented data, various parasite features were tracked over time: volume, convex volume, mean and maximum signal intensity, position, elevation, distance moved, principal axes lengths of the parasite ellipsoid, Euler angles, solidity (computed as ratio of volume to convex volume), number of bright elements inside the parasite-infected cell, equivalent diameter ($\sqrt[3]{\frac{6 \cdot \text{Volume}}{\pi}}$), ratio of width to height and ratio of volume to equivalent diameter.

Fig. 2 delineates the features most indicative of parasite development by 70 hpi, highlighting significant differences between successful and unsuccessful developmental outcomes. The mean feature values at every time point across the whole dataset are shown with a blue line (positive classes) and red line (negative classes), while the whole range

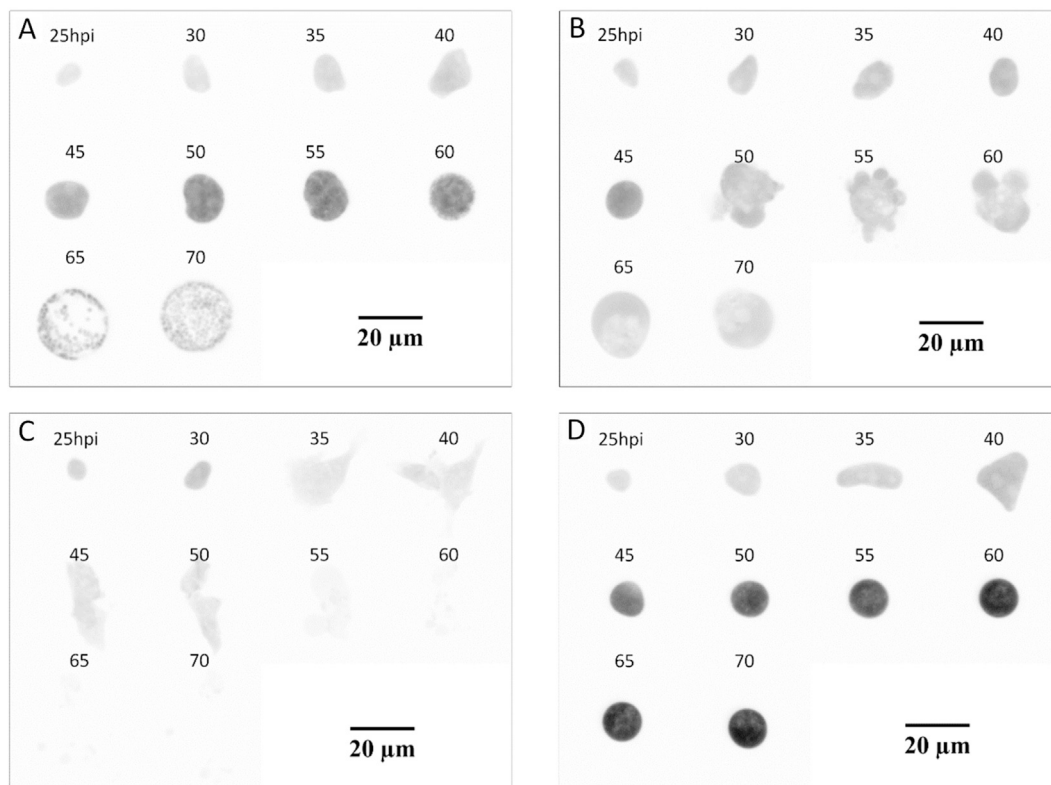


Fig. 1. Development of *P. berghei* parasites. Individual parasites were followed by time-lapse fluorescent microscopy. A) Normal development and (B–D) defective development are shown as still images for the indicated timepoints. In B) the parasite has lost its integrity at 50 hpi, in C) already at 35 hpi and in D) the parasite did develop normally till 45 hpi but then was not progressing.

of values is depicted with the corresponding lighter-shaded color. The overlapping values between classes are depicted with dark red. The figures illustrate that features related to size (volume, equivalent diameter, ratio of volume to equivalent diameter) are significantly larger in the positive class compared to the negative class. The mean signal intensity captured by the fluorescence microscope tends to increase in time at a higher rate in the positive class compared to the negative class, but then it decreases at 55–60 hours post-infection in the positive cases. Mean parasite elevation is similar in both classes but tends higher after 65 hours post-infection in the positive class. The number of bright elements inside a parasite-infected cell - a proxy for merozoites, is observed to increase at a faster rate in the positive class, reaching a 6-fold difference between classes at 70 hours post-infection. The mean solidity parameter is observed as static in time. However, after 60 hours post-infection it decreases in the positive class. The mean distance traveled, and the ratio of parasite width to length are similar in both classes. Notably, features like volume (and size related features) and the presence of bright internal elements, suggestive of merozoite formation, were markedly distinct in successfully developing parasites. The other features are shown to have a higher overlap between classes. This comprehensive tracking of parasitic features across the developmental timeline provides crucial insights into the dynamics of *P. berghei* infection *in vitro*, offering a robust framework for predicting developmental outcomes based on early-stage observations.

3.2. Training characteristics

With these measured features, time series binary classification was used to train a classifier to distinguish between successfully developing and non-developing parasites at every time point. To reduce the risk of selection bias, the training and evaluation procedure was conducted using a five-fold cross-validation procedure. This involved running the experiment five times, each using a different training and validation set.

Here 80% of the data was used for training, 20% was used for validation (cf. Table 1). For every fold, training accuracy and loss plateaued and converged after 500 to 600 epochs. Our performance metrics are reported as means and standard deviations across the five test folds corresponding to each training cross-validation cycle. We report the performance metrics on the holdout test dataset by applying each of the five previously derived machine learning models to this dataset. This comprehensive validation process ensures the reliability and generalizability of our classifier in distinguishing between the developmental outcomes of *P. berghei* parasites.

3.3. Classification performance in the validation dataset

Our analysis focused on evaluating the classifier's performance at two critical junctures: as a classifier at the end of the data collection period, approximately 70 hours post-infection (hpi), and as a predictive tool at 55 hpi, forecasting the development outcome of infected cells by 70 hpi. In the validation dataset, our system discriminated successfully developing parasites from the non-successful ones (cf. Table 2) with an AUC of 0.933 (s.d. 0.0027), a sensitivity of 95.0% (s.d. 7.5), specificity of 82.8% (s.d. 7.6), and accuracy of 88.9% (s.d. 4). At 55 hours post-infection the performance metrics were an AUC of 0.896 (s.d. 0.029), a sensitivity of 90.4% (s.d. 6.7) specificity of 82.0% (s.d. 6.2) and accuracy of 86.2% (s.d. 2.1). The prediction accuracy plot from Fig. 3A rises from a mean of 50% at 32 hours post-infection to 89% at 70 hours post-infection. The receiver operator curves are plotted for the 70 and 55 hours post-infection time points for each of the five folds in the validation dataset (Fig. 3B–C). To contextualize these results, two human annotators performed the task of parasite development prediction at 55 hours post-infection on a subset of 44 capture positions comprising of 78 infected cells. While all capture positions contain infected cells, by 55 hpi some parasites present in the early acquisition stages might have died or left the field of view. Predictions were made using the en-

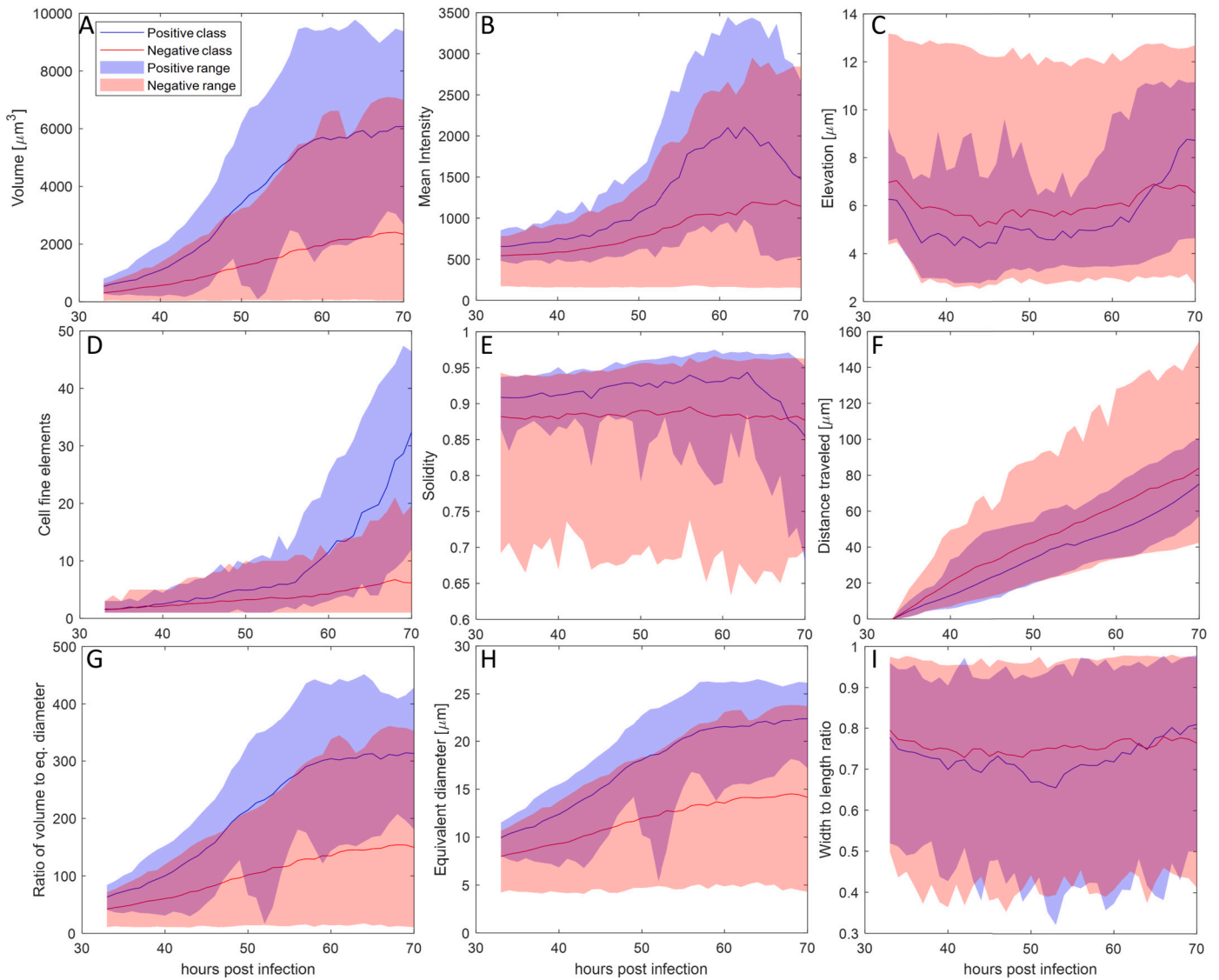


Fig. 2. Various parasite features tracked over time illustrating the parasite development: A) volume of parasite-infected cell, B) mean fluorescent signal intensity, C) infected cell elevation, D) number of bright elements inside a parasite-infected cell as proxy for merozoites, E) solidity - computed as ratio of volume to convex volume, F) distance traveled, G) ratio of volume to equivalent diameter, H) equivalent diameter, I) ratio of cell width to height. Mean values are shown with a blue line (positive) and red line (negative). The shaded areas represent the range of the features and overlapping ranges are marked with dark red.

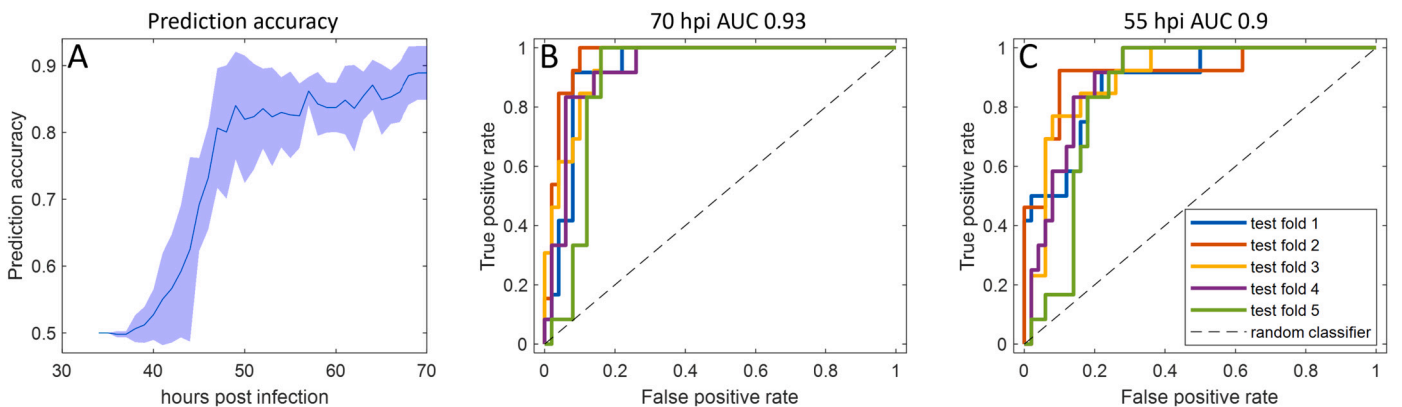


Fig. 3. A) Prediction accuracy of parasite development at increments of 1 hour until end of acquisition. B) and C) receiver operator characteristic for the 5 cross validation folds for predicting *P. berghei* infected cell development in the validation dataset at 70 and 55 hours post-infection.

Table 2
Performance metrics for validation dataset.

	Balanced Accuracy	Sensitivity	Specificity	AUC
Primary validation dataset				
70 hours post-infection	88.9 ± 4	95 ± 7.5	82.8 ± 7.6	0.933 ± 0.027
55 hours post-infection	86.2 ± 2.1	90.4 ± 6.7	82 ± 6.2	0.896 ± 0.029
Test dataset				
70 hours post-infection	83.9 ± 1.7	84.6 ± 3.0	83.3 ± 2.8	0.873 ± 0.009
55 hours post-infection	75.6 ± 1.0	68.2 ± 3.6	82.6 ± 1.9	0.825 ± 0.022

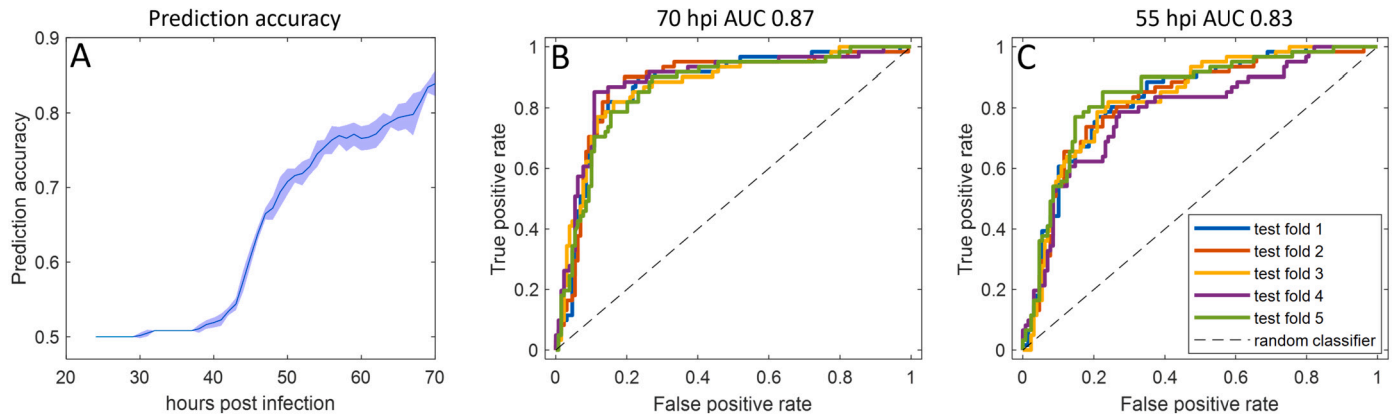


Fig. 4. A) Prediction accuracy of parasite development at increments of 1 hour until end of acquisition. B) and C) receiver operator characteristic for predicting *P. berghei* infected cell development in the test dataset at 70 and 55 hours post-infection.

tire time-volume sequence up to 55 hours post-infection. An accuracy of 79.3% was achieved by the first annotator and 72.5% by the second annotator.

3.4. Classification performance in the test dataset

In the test dataset the performance metrics at 70 and 55 hours post-infection were (cf. Table 2): an AUC of 0.873 (s.d. 0.009), respectively 0.825 (s.d. 0.022), a sensitivity of 84.6% (s.d. 3), respectively 68.2% (s.d. 3.6), a specificity of 83.3% (s.d. 2.8), respectively 82.6% (s.d. 1.9) and an accuracy of 83.9% (s.d. 1.7), respectively 75.6% (s.d. 1.0). The prediction accuracy plot from Fig. 4A rises from a mean of 50% at 24 hours post-infection to 83.3% at 70 hours post-infection. The receiver operator curves are plotted for both the 70 and 55 hours post-infection time points on the test dataset for each model resulting from the cross-validation process (Fig. 4 B-C). The same two human annotators performed the task of parasite development prediction at 55 hours post-infection, achieving an accuracy of 74.2%, respectively 77.6%.

3.5. Feature importance

We used permutation importance to evaluate our model's reliance on each parasite feature. For each feature $j \in \{1, \dots, p\}$ in the dataset X , a new feature matrix X_j was generated where values of feature j were permuted across the dataset X . The feature importance is reported as the prediction accuracy difference of each model on the initial dataset X and on the feature permuted dataset X_j . This process was repeated ten times, generating a different permutation sequence for feature j and each of the five models from the cross-validation process. Fig. 5 illustrates the computed feature importance for the validation and test dataset at 70 hours post-infection. The most important features in both cases were the volume, convex envelope, and maximum signal intensity, with the other features having an importance of less than 0.5%.

To study the effect of feature number, we trained new models on a subset of the 17 features. The subset was chosen by random sampling a number $N_{features}$ from the features that were used to train the model. The 4 most important features (volume, convex envelope, mean and

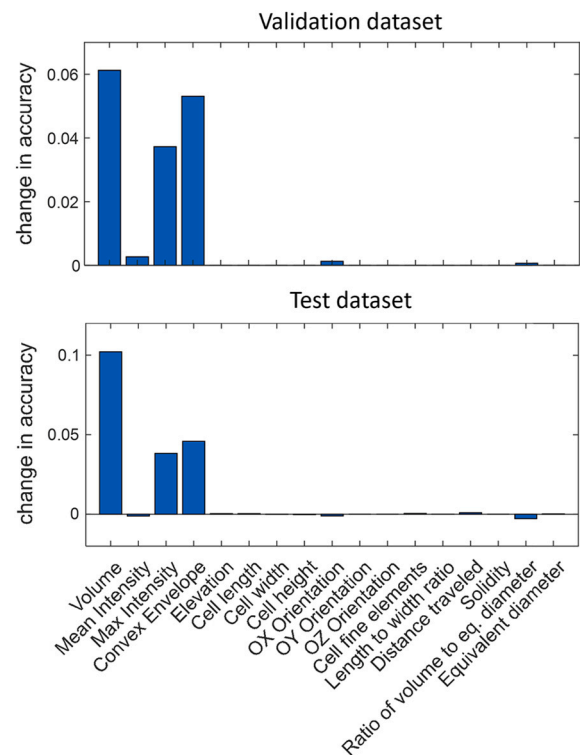


Fig. 5. Feature importance computed using the permutation importance algorithm on the validation (top) and test dataset (bottom) for each of the 17 parasite features.

maximum signal intensity) were always used. This was repeated ten times to account for stochasticity in training and random feature selection. The number of features $N_{features}$ was swept from 4 to 17. We did not identify any significant trend in the evaluation metrics at 70 and 55 hpi, except for a larger standard deviation when $N_{features}$ is 4 (Fig. 6).

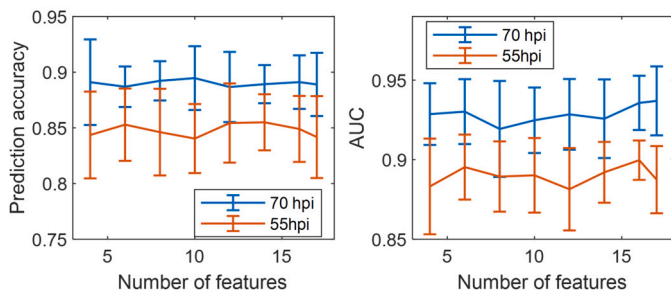


Fig. 6. Accuracy (left) and AUC (right) for predicting parasite development at 70 and 55 hours post infection as a function of the number of features used for training.

4. Discussion

The initial stage of *Plasmodium* infection occurs in the liver, where tens of thousands of blood-infectious merozoites are generated in each infected hepatocyte. The entire liver stage is typically asymptomatic; however, the emergence of parasites into the bloodstream marks the onset of symptomatic malaria, such as recurring fever. A deeper understanding of this liver stage is crucial for developing targeted treatments. Our study embarked on exploring the potential of an automated system to predict the developmental outcomes of the *P. berghei* parasite during this critical stage. Our first goal was to develop a method for parasite segmentation and tracking to extract physical parameters for characterizing parasite development. Our second goal was to evaluate the performance of the proposed convolutional neural network-based algorithm on predicting parasite development at 55 hours post-infection. This time-point was chosen based on the expected onset of the cytomere stage of the parasite [41]. Our principal discovery was the system's capability to anticipate which parasites would progress to a stage where merozoites freely navigate within host cells.

In our test set, we achieved a balanced accuracy of 75.6%, sensitivity of 68.2%, and specificity of 82.6% for predicting the final development of *Plasmodium* parasite development at 55 hours post-infection. The results are similar to those of human annotators of 75.9% accuracy and indicate the feasibility of the proposed method. To our knowledge, this marks the first instance of employing fluorescent microscopy images alongside machine learning techniques for predicting, rather than merely classifying, parasite development. Such predictive capabilities could significantly benefit in vitro studies by identifying parasites unlikely to complete their developmental journey. Automating these predictions not only alleviates the manual annotation workload but also offers clearer, more interpretable results with reduced computational demands.

Moreover, we hypothesize that tracking the evolution of various cell features, such as volume, elevation, shape, and others, might yield additional insights, especially in drug discovery contexts. Fig. 2 shows that features related to parasite size are larger in the positive class. It is known that quite a substantial number of parasites are eliminated during liver stage development [36]. Normal growing parasites increase in mass, which can be seen by the larger sizes and increased fluorescence intensity. However, there was no difference between the positive and negative classes concerning the ratio of cell length to width. The fluorescent signal decreases in intensity from 60 hours post-infection. This can be explained by the rupture of the parasitophorous membrane and liberation of merozoites into the host cell cytoplasm and therefore, dispersion of the fluorescent signal. The same explanation holds for the solidity parameter, which is observed to decrease after 60 hours post-infection in the positive class. Merozoites get liberated and move freely in the host cell cytoplasm. The infected host cells undergo dramatic cytoskeletal changes at late stage parasite development. Virtually all actin cytoskeleton retracts from the host cell plasma membrane, which in vitro leads to contact loss from cell culture vessel/surface [33]. The detached cell resembles a spherical shape and should be seen at higher elevations. However, cell elevation was similar before 65 hours post-infection in both classes, we hypothesize this is because cells undergo division two to four times during the imaging period, which causes temporary cell elevation. Also, if not only the parasite dies but also the host cell, the dead cell also loses contact with the culture vessel and elevates. While the distance traveled did not seem to be an important parameter in these experiments, it could provide additional insights when conducting future studies, for example, for drug therapies that might affect infected cell motility [45]. Interestingly, we observed in $\approx 5\%$ of the cases in our experiments that a cell division after 50 hours post-infection causes a several hour delay in the development of the parasite.

In both validation and test dataset we observe a consistent increase in prediction accuracy up to ≈ 45 hours post-infection, followed by a slower increase till ≈ 65 . The prediction accuracy is experiencing a

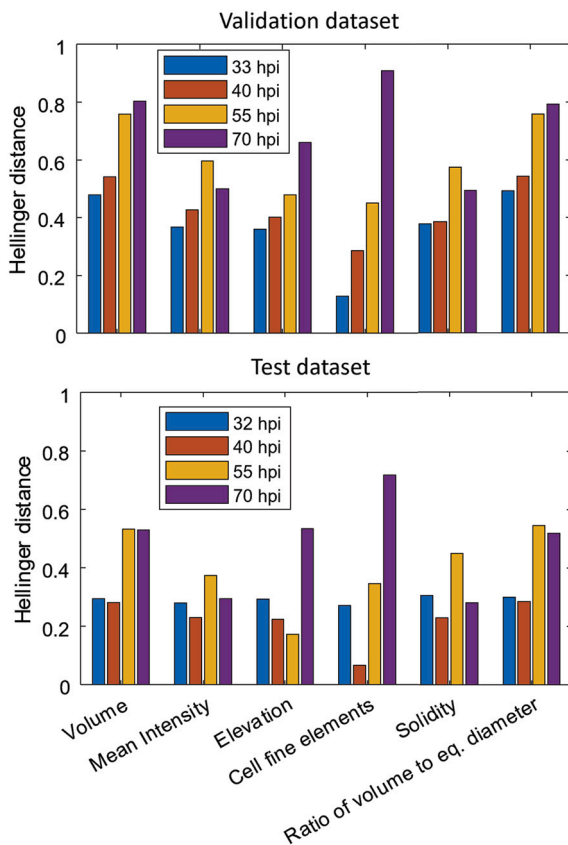


Fig. 7. Distance between features (volume, mean fluorescence signal intensity, parasite elevation, number of bright elements inside cell as proxy for merozoites, cell solidity and ratio of volume to eq. diameter) of positive and negative classes at 32-33, 40, 55, and 70 hours post infection computed as Hellinger distance for validation and test datasets.

Another way to visualize the importance of features is to compute the distance between the feature values (seen as a probability distribution) of the positive and negative classes. While the permutation importance was computed after the model was trained, this method is computed on the data itself without requiring training. Fig. 7 illustrates the Hellinger distance at 32, 40, 55, and 70 hours post-infection for the following features: volume, mean intensity, elevation, cell fine elements, solidity, and the ratio of volume to equivalent diameter. At 70 hours post-infection the features offering the largest distance between classes were the number of bright elements inside the host cell, then volume, and the ratio of volume to the equivalent diameter of the parasite.

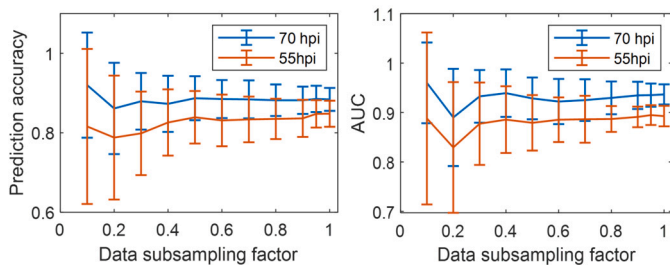


Fig. 8. Accuracy (left) and AUC (right) for predicting parasite development at 70 and 55 hours post infection as a function of training data subsampling factor.

qualitatively sharper increase again until the end of the imaging period. We assume this last increase is because, at this time, merozoites appear, and detached cells start to form. Therefore, different parasite features might be more important in each development period. We observe the standard deviation of prediction accuracies to be lower in the test dataset. We believe this is because, during the five-fold cross-validation process, the models are evaluated on the different data splits. In the test dataset case, all the models are evaluated on the same test dataset. This also explains the coarser look of the receiver operator characteristic curves in the validation dataset. In the cross-validation process, the validation dataset was split into five, leading to 62–63 capture positions in each validation set, compared to 190 in the test dataset.

We used permutation importance to evaluate our model’s reliance on each cell parameter. The assumption is that if a feature is significant, shuffling its values will lead to a significant drop in the model’s performance. It does not require model retraining and can capture variable interactions. Our results in Fig. 5 show a good agreement between validation and test data feature importance, indicating the generalizability of our model. Interestingly, the number of bright elements inside a parasite-infected cell was not identified as an essential feature. While shown to be effective [44], a drawback of this method is that the importance of correlated features may be overestimated [46] and that the feature importance accuracy is highly dependent on the model’s accuracy.

Fig. 7 illustrates the Hellinger distance between features at 32, 40, 55, and 70 hours post-infection. As expected, in the early to late schizont phase (32,40 hours post-infection), the feature distance between classes is very small (< 0.1), resulting in poor predictions. We assume this is because up to the point of ≈ 55 –65 hours post-infection, the hallmarks of a successfully developing parasite are not yet visible.

We compared the results of our method to those of a logistic regression based classifier. For training, the dataset $X_{j,t}$, with features $j \in \{1, \dots, p\}$ and time-points $t \in \{1, \dots, p\}$ was flattened into a training set Xr_{jt} , with $J_{reg} \in \{1, \dots, p \times t\}$. The regression-based classifier resulted in an accuracy of 81.2 (s.d. 11.5) and AUC of 0.883 (s.d. 0.1), at 70 hpi and an accuracy of 78.5 (s.d. 11.3) and AUC of 0.832 (s.d. 0.126) at 55 hpi for the validation set. We assume the proposed LSTM model performed better because it can handle sequential data and learn relationships between past and current data points.

The performance of our model trained using oversampling of the minority class was also compared with a model trained without data balancing. The model without oversampling resulted in lower accuracy metrics, with an accuracy of 81.3 (s.d. 8.9) and AUC of 0.930 (s.d. 0.025), at 70 hpi and an accuracy of 64.5 (s.d. 8.3) and AUC of 0.874 (s.d. 0.027) at 55 hpi for the validation set.

To study the importance of dataset size, new models were trained on a subset of the data with a subsampling factor ranging from 0.1 to 1. This was repeated five times to account for stochasticity in training and random data selection. Larger standard deviations were observed starting from a subsampling factor of 0.95, with twice higher standard deviations after a subsampling of 0.5. Moreover, a decrease in AUC and accuracy can be observed after the 0.5 subsampling factor (Fig. 8).

Despite these advancements, our study is not without limitations. First, the infection process was not controlled leading to an uneven distribution of infected cells across the petri dish, which can create high-confluency cell areas. These were not included in our study. The challenges of ensuring a balanced dataset, given the uncontrolled infection rates and the variable outcomes of parasite development, were addressed through a five-fold cross-validation approach and random sampling which ensured similar class ratios for each training-validation split. Moreover, to ensure that the training, validation, and testing of our neural network model were not affected by variability in the number of cases in each class, we used oversampling of the minority classes to the number of images in the majority class during training. When reporting the results, we used a weighted accuracy instead of standard accuracy as an evaluation metric, to further ensure our results are not biased by the number of images in each class. Secondly, compared to a few other studies [16,47,30] which were collected from diverse data sources, our data was collected in a single center, the University of Bern using the Nikon microscope. Therefore further investigation is required to determine if these results are generalizable to other acquisition devices or specimens prepared in a different way. Thirdly, it is important to note that our model classifies the data only of the fluorescent signal of mCherryexpressing parasites, without using the brightfield microscopy image data. Moreover, although we acquired z-stack fluorescent data at every hour with very low laser energy, we cannot exclude phototoxicity affecting the parasite development [48]. However, this effect could be lessened in future studies by using the proposed approach of predicting parasite development and therefore stopping the experiments early.

5. Conclusion

In this study, we explored the integration of fluorescent microscopy imaging and neural network based algorithms to predict the development of *P. berghei* during its liver stage. Our automated approach for cell segmentation and tracking has proven effective, successfully identifying developing parasites with an accuracy of 87.3%. Crucially, our method achieved a 75.6% prediction accuracy at 55 hours post-infection, paralleling the proficiency of human annotators. These findings affirm the potential of our approach to enhance the study of *Plasmodium* liver stage development, offering a promising tool for malaria research with implications for both understanding the disease and developing new treatments.

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.

Data availability

The datasets used during the current study are available on Zenodo.org under the following DOI <https://doi.org/10.5281/zenodo.10909406>. The neural network-based model and scripts used in the context of this paper are available on GitHub (<https://github.com/ml-malaria-prediction/ml-malaria-prediction>.)

Acknowledgement

This work was supported by the Swiss National Science Foundation to VH (grant number CRSII5_198543).

Appendix A. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.csbj.2024.04.029>.

References

- [1] Howick VM, Russell AJC, Andrews T, Heaton H, Reid AJ, Natarajan K, et al. The malaria cell atlas: single parasite transcriptomes across the complete plasmodium life cycle. *Science* Aug. 2019;365(6455).
- [2] Bousema T, Drakeley C. Epidemiology and infectivity of plasmodium falciparum and plasmodium vivax gametocytes in relation to malaria control and elimination. *Clin Microbiol Rev* Apr. 2011;24(2):377–410.
- [3] Gitta B, Killian N. Diagnosis of malaria parasites plasmodium spp. in endemic areas: current strategies for an ancient disease. *BioEssays* Jan. 2020;42(1):e1900138.
- [4] Shamriz S, Ofoghi H, Moazami N. Effect of linker length and residues on the structure and stability of a fusion protein with malaria vaccine application. *Comput Biol Med* Sep. 2016;76:24–9.
- [5] Ferruz N, Heinzinger M, Akdel M, Goncarenco A, Naef L, Dallago C. From sequence to function through structure: deep learning for protein design. *Comput Struct Biotechnol J* 2023;21:238–50.
- [6] De Niz M, Burda P-C, Kaiser G, Del Portillo HA, Spielmann T, Frischknecht F, et al. Progress in imaging methods: insights gained into plasmodium biology. *Nat Rev Microbiol* Jan. 2017;15(1):37–54.
- [7] De Niz M, Heussler VT. Rodent malaria models: insights into human disease and parasite biology. *Curr Opin Microbiol* Dec. 2018;46:93–101.
- [8] Bailey JW, Williams J, Bain BJ, Parker-Williams J, Chiodini PL, General Haematology Task Force of the British Committee for Standards in Haematology. Guideline: the laboratory diagnosis of malaria. *Br J Haematol* Dec. 2013;163(5):573–80.
- [9] Suratane A, Buaboocha T, Plaimas K. Prediction of Human-Plasmodium vivax protein associations from heterogeneous network structures based on Machine-Learning approach. *Bioinform Biol Insights* Jun. 2021;15:11779322211013350.
- [10] Gupta P, Venkadesan S, Mohanty D. Pf-phospho: a machine learning-based phosphorylation sites prediction tool for plasmodium proteins. *Brief Bioinform* Jul. 2022;23(4).
- [11] de Souza Oliveira A, Guimarães Fernandes Costa M, das Graças Vale Barbosa M, Ferreira Fernandes Costa Filho C. A new approach for malaria diagnosis in thick blood smear images. *Biomed Signal Process Control* Sep. 2022;78:103931.
- [12] Zhang C, Jiang H, Jiang H, Xi H, Chen B, Liu Y, et al. Deep learning for microscopic examination of protozoan parasites. *Comput Struct Biotechnol J* Feb. 2022;20:1036–43.
- [13] Agany DDM, Pietri JE, Gnimpieba EZ. Assessment of vector-host-pathogen relationships using data mining and machine learning. *Comput Struct Biotechnol J* Jun. 2020;18:1704–21.
- [14] Mariki M, Mkoba E, Mduma N. Combining clinical symptoms and patient features for malaria diagnosis: machine learning approach. *Appl Artif Intell* Dec. 2022;36(1):2031826.
- [15] Hayat M, Tahir M, Alarfaj FK, Alturki R, Gazzawe F. NLP-BCH-Ens: NLP-based intelligent computational model for discrimination of malaria parasite. *Comput Biol Med* Oct. 2022;149:105962.
- [16] Schmedes SE, Dimbu RP, Steinhardt L, Lemoine JF, Chang MA, Plucinski M, et al. Predicting plasmodium falciparum infection status in blood using a multiplexed bead-based antigen detection assay and machine learning approaches. *PLoS ONE* Sep. 2022;17(9):e0275096.
- [17] Oguike OE, Ugwuishiwu CH, Asogwa CN, Nnadi CO, Obonga WO, Attama AA. Systematic review on the application of machine learning to quantitative structure-activity relationship modeling against plasmodium falciparum. *Mol Divers* Dec. 2022;26(6):3447–62.
- [18] Rajaraman S, Antani SK, Poostchi M, Silamut K, Hossain MA, Maude RJ, et al. Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ* Apr. 2018;6:e4568.
- [19] Hemachandran K, Alasiry A, Marzougui M, Ganie SM, Pise AA, Alouane MT-H, et al. Performance analysis of deep learning algorithms in diagnosis of malaria disease. *Diagnostics* Feb. 2023;13(3).
- [20] Masud M, Alhuymani H, Alshamrani SS, Cheikhrouhou O, Ibrahim S, Muhammad G, et al. Leveraging deep learning techniques for malaria parasite detection using mobile application. In: *Proc. int. wirel. commun. mob. comput. conf.*, vol. 2020. Jul. 2020.
- [21] Cho YS, Hong PC. Applying machine learning to healthcare operations management: CNN-based model for malaria diagnosis. *Healthcare* Jun. 2023;11(12).
- [22] Alok N, Krishan K, Chauhan P. Deep learning-based image classifier for malaria cell detection; Apr. 2021. p. 187–97.
- [23] Di Ker A. An efficient model of residual based convolutional neural network with Bayesian optimization for the classification of malarial cell images. *Comput Biol Med* Sep. 2022;148:105635.
- [24] Liang Z, Powell A, Ersoy I, Poostchi M, Silamut K, Palaniappan K, et al. CNN-based image analysis for malaria diagnosis. In: *2016 IEEE international conference on bioinformatics and biomedicine (BIBM)*; Dec. 2016. p. 493–6.
- [25] Hung J, Lopes SCP, Nery OA, Nosten F, Ferreira MU, Duraisingh MT, et al. Applying faster R-CNN for object detection on malaria images. *Conf Comput Vis Pattern Recognit Workshops* Jul. 2017;2017:808–13.
- [26] Fuhad KMF, Tuba JF, Sarker MRA, Momen S, Mohammed N, Rahman T. Deep learning based automatic malaria parasite detection from blood smear and its smartphone based application. *Diagnostics* May 2020;10(5).
- [27] Manning K, Zhai X, Yu W. Image analysis and machine learning-based malaria assessment system. *Digit Commun Netw* Apr. 2022;8(2):132–42.
- [28] Dey S, Nath P, Biswas S, Nath S, Ganguly A. Malaria detection through digital microscopic imaging using deep greedy network with transfer learning. *J Med Imag Sep.* 2021;8(5):054502.
- [29] Arshad QA, Ali M, Hassan S-U, Chen C, Imran A, Rasul G, et al. A dataset and benchmark for malaria life-cycle classification in thin blood smear images. *Neural Comput Appl* Nov. 2021.
- [30] Yang Z, Benhabiles H, Hammoudi K, Windal F, He R, Collard D. A generalized deep learning-based framework for assistance to the human malaria diagnosis from microscopic images. *Neural Comput Appl* Sep. 2022;34(17):223–14238.
- [31] Singh AP, Buscaglia CA, Wang Q, Levay A, Nussenzweig DR, Walker JR, et al. Plasmodium circumsporozoite protein promotes the development of the liver stages of the parasite. *Cell* Nov. 2007;131(3):492–504.
- [32] Prudêncio M, Rodríguez A, Mota MM. The silent path to thousands of merozoites: the plasmodium liver stage. *Nat Rev Microbiol* Nov. 2006;4(11):849–56.
- [33] Burda P-C, Caldelari R, Heussler VT. Manipulation of the host cell membrane during plasmodium liver stage egress. *mBio* Apr. 2017;8(2).
- [34] Burda P-C, Schaffner M, Kaiser G, Roques M, Zuber B, Heussler VT. A plasmodium plasma membrane reporter reveals membrane dynamics by live-cell microscopy. *Sci Rep* Aug. 2017;7(1):9740.
- [35] De Niz M, Caldelari R, Kaiser G, Zuber B, Heo WD, Heussler VT, et al. Hijacking of the host cell Golgi by plasmodium berghei liver stage parasites. *J Cell Sci* May 2021;134(10).
- [36] Niklaus L, Agop-Nersesian C, Schmuckli-Maurer J, Wacker R, Grünig V, Heussler VT. Deciphering host lysosome-mediated elimination of plasmodium berghei liver stage parasites. *Sci Rep* May 2019;9(1):7967.
- [37] Masters JR. HeLa cells 50 years on: the good, the bad and the ugly. *Nat Rev Cancer* Apr. 2002;2(4):315–9.
- [38] Beyeler R, Jordan M, Dorner L, He B, Cyrklaff M, Roques M, et al. Putative prefolin complex subunit 5 of plasmodium berghei is crucial for microtubule formation and parasite development in the mosquito. *Mol Microbiol* Mar. 2024;121(3):481–96.
- [39] Bindschedler A, Schmuckli-Maurer J, Wacker R, et al. Plasmodium berghei-mediated NRF2 activation in infected hepatocytes enhances parasite survival. *Cellular* 2022.
- [40] Atchou K, Berger BM, Heussler V, Ochsenreiter T. Pre-gelation staining expansion microscopy for visualisation of the plasmodium liver stage. *J Cell Sci* Nov. 2023;136(22).
- [41] Burda P-C, Roelli MA, Schaffner M, Khan SM, Janse CJ, Heussler VT. A plasmodium phospholipase is involved in disruption of the liver stage parasitophorous vacuole membrane. *PLoS Pathog* Mar. 2015;11(3):e1004760.
- [42] Kuhn HW. The Hungarian method for the assignment problem. *Nav Res Logist Q* Mar. 1955;2(1–2):83–97.
- [43] Amari S-I. Backpropagation and stochastic gradient descent method. *Neurocomputing* Jun. 1993;5(4):185–96.
- [44] Fisher A, Rudin C, Dominici F. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. *J Mach Learn Res* 2019;20.
- [45] Chen S, Shenk T, Nogalski MT. P2Y2 purinergic receptor modulates virus yield, calcium homeostasis, and cell motility in human cytomegalovirus-infected cells. *Proc Natl Acad Sci USA* Sep. 2019;116(38):971–18982.
- [46] Strobl C, Boulesteix A-L, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. *BMC Bioinformatics* Jul. 2008;9:307.
- [47] Lee YW, Choi JW, Shin E-H. Machine learning model for predicting malaria using clinical information. *Comput Biol Med* Feb. 2021;129:104151.
- [48] Yahya S, Jordan S, Smith HX, Gaboriau DCA, Famodimu MT, Dahalan FA, et al. Live-cell fluorescence imaging of microgametogenesis in the human malaria parasite plasmodium falciparum. *PLoS Pathog* Feb. 2022;18(2):e1010276.