Copenhagen Business School

Master of Science in Business Administration & Data Science
Data Mining, Machine Learning, and Deep Learning

Supervisors: Somnath Mazumdar

Final Project

Enhancing Malaria Diagnosis: Comparative Analysis of CNN and SVM Models for *P. vivax* Cell Detection

Dataset and Code

https://drive.google.com/drive/folders/1VTQhyxTC2UBCjQpL3E0wg4uV4JhX8GfV?usp=sharing

Authors	Konstantin Kleffke, Linda Schombach & Mira Metzger
Student IDs	158284, 158305 and 159107
Pages	15
Characters	33.758
Submission	18th of May 2023

Abstract

This research paper focuses on the detection of *P. vivax* malaria-infected cells. It aims to close the existing research gap of the *P. vivax* parasite and decrease cost and labor necessities for malaria identification. Using the Broad Bioimage Benchmark Collection (BBBC) dataset of over 85.000 blood smear images, different classification models were developed and compared. As hypothesized, the Support Vector Machines (SVM) model was outperformed by the two Convolutional Neural Networks (CNN) models. The highest accuracy was demonstrated by CNN1 with 0.98. Results support the possibility of practically implement the models into automated microscopy systems or mobile applications. This could lead to real-time diagnosis, malaria control strategies, and better resource allocation. Especially for low resource countries this could have significant positive effects. However, ethical implications must be considered. In conclusion, the study highlights the effectiveness of CNN models for binary classifying cells as malaria infected or uninfected. Further studies are recommended include different malaria parasites, potentially enabling to distinguish between parasites and to be transferable to new emerging parasites as well.

Keywords: Malaria Cell Detection, P. vivax Parasite, Machine Learning, Deep Learning, Image Classification, Convolutional Neural Networks, Support Vector Machines

Table of Content

Table of Figures

Table of Tables

1. Introduction

Malaria is a highly dangerous disease that is transmitted through bits from *Plasmodium* parasites. In 2021, 247 million malaria cases and 619.000 deaths were reported in 84 malaria endemic countries. Malaria occurs primarily in developing and underdeveloped countries (95% of global cases are in Africa) and is highly dangerous for children (76% of total malaria deaths in children aged under 5 years). Though deaths are steadily declining (897.000 deaths in 2000), several challenges raise concerns such as disruptions in treatment and detection caused by the Covid-19 pandemic and increasing resistance of insecticide-treated nets. Furthermore, new parasites seem to not be detectable by widely used diagnostic tests, which is why additional detection methods are necessary. (*World Malaria Report*, 2022)

In total, seven species of *Plasmodium* parasites affect human of which *P. falciparum* and *P. vivax* accumulated cause 99% of all malaria cases. While *P. falciparum* is the deadliest, *P. vivax* is the geographically most diversified parasite. Consequently, previous research focused on the former. However, studies revealed that *P. vivax* can lead to severe outcomes as well and emphasized the need to close the research gap for this parasite. Further, while a vaccine for *P. falciparum* malaria already exists, there are only few vaccines for *P. vivax* under development. (Veiga et al., 2023; *World Malaria Report*, 2022)

In absence of vaccines, a rapid diagnosis and treatment is the best method to prevent severe outcomes and deaths, which illustrates the importance of research of accurate malaria detection methods (Landier et al., 2016). Existing laboratory tests such as PCR, are often very costly and infeasible due to a lack of equipment and trained health personnel. (CDC, 2019) Also, successfully identifying Malaria in red blood cells by manual microscopy strongly relies on the expert knowledge of the health personnel. Consequently, an accurate detection represents a significant challenge especially in low-resource areas.
To close this knowledge gap and to facilitate an early detection also of *P. vivax* infections, machine learning techniques could be leveraged as a cost-efficient way. (Meng et al., 2022; Zhao et al., 2020) Thus, this paper aims to develop and compare different models that classify blood smear images with cells as *P. vivax* infected or uninfected. Therefore, the Broad Bioimage Benchmark Collection (BBBC) dataset of over 85.000 blood smear images will be used. Based on results of prior research, Support Vector Machines (SVMs) model as well as two Convolutional Neural Networks (CNN) models were developed and compared. All models were trained and tested on the same dataset, ensuring a fair comparison.

2. Related Work

There is a vast amount of research on Malaria cell detection as well as classification using various datasets. As stated previously, the focus has been primarily on *P. falciparum* cells (Gopakumar et al., 2018; Rahman et al., 2019; Rajaraman et al., 2018, 2019). Zhao et al. (2020) tested a *P. falciparum* trained model on a *P. vivax* dataset. The results indicate that the model trained on only P. *falciparum* is incapable of predicting infections caused by P. *vivax* as the model is overtrained for the specific task. This finding underscores the importance of training separate models for each parasite to ensure accurate predictions.

Delas Peñas et al. (2017) trained a CNN model for detection of malaria cells in a blood smear (92.4% accuracy). However, when distinguishing between *P. falciparum* and *P. vivax* reached only an accuracy of 87.9%. Hung and Carpenter (2017) used the faster region-based CNN model developed for cell segmentation also to classify those segmented cells. However, a non-binary classification was unsuccessful (accuracy 59%). For this task Vijayalakshmi & Rajesh (2019) proposed a transfer learning approach by combining a Visual Geometry Group (VGG) 19 and a SVM models. Results show an 93.1% classification accuracy including detection.

Considering classification technique's, Das et al. (2013) compared SVM with a Bayesian approach to detect malaria infections. In more recent years, deep learning architectures have become the primary method for malaria diagnosis. Especially CNNs have proven successful. Lian et al. (2016) developed a 16-layer CNN model which has a 97.37% accuracy in a binary classification between infected and uninfected cells. For the same task, Dong et al. (2017) evaluated three types of CNNs: LeNet, AlexNet and GoogLeNet. All of them reached an accuracy of over 95% and thereby outperformed the SVM model (92% accuracy). Narayanan et al. (2019) obtained accuracies of over 96% using deep neural networks (GoogLeNet and ResNet). Similarly, Rajaraman et al. (2019) found that the VGG19 model outperformed several other techniques including ensemble models, causing an accuracy of 99.32%.

Extending the binary classification, Li et al. (2021) and Meng et al. (2022) developed models for multi-stage malaria detection – both using the BBBC dataset. The former proposed the DTGCN framework consisting of CNN feature learning, source transfer graph building, and unsupervised graph convolutional network (UGCN). This model outperformed several other tested models such as VeggNet, GoogLeNet and ResNet models, yielding over 99% accuracy in detection different malaria stages. The latter followed a Neighbor Correlated Graph Convolutional Network (NCGCN) approach which combines CNN feature learning, neighbor

correlation mining (using K-Nearest Neighbor (KNN) and ϵ-radius graph building algorithms), and graph representation modules (using graph convolutional network (GCN)). In comparison to other models, the NCGCN reached a better outcome with an accuracy of over 94%.

3. Conceptual Framework

The primary objective of this study is to construct a predictive model which can distinguishes between *P. vivax* infected and uninfected cells, based on microscopic images. Thus, the main contribution of this paper is to support the accurate detection of infections caused by *P. vivax* which has been neglected in prior research. Further, this study is the first to apply binary classification techniques to this specific dataset to the best of our knowledge. From a clinical standpoint, the initial critical information needed is the detection of an infection rather than the exact type. Given the task's simplicity compared to multi-label classification, improved model performance is anticipated.

Aligning with previous malaria research findings, it is hypothesized that deep learning will outperform machine learning models. Specifically, this study develops and compares the benchmarking SVM with two CNNs models. Distinct architectures were employed to explore the effect of model complexity on predictive accuracy. The selection of these models was driven by their respective strengths in cell classification (Shifat-E-Rabbi et al., 2020) and proven track records in similar image classification tasks as describe above.

SVM, a traditional machine learning algorithm, has been lauded for its ability to handle high-dimensional data and its robustness against overfitting (Cervantes et al., 2020; Porcello, 2019). Its underlying principle, which involves finding the optimal hyperplane to separate different classes, makes it particularly effective for binary classification tasks such as the task at hand (Cervantes et al., 2020).

CNN, a type of deep learning model, is designed specifically for processing grid-like data such as images. As in the case of images each pixel represents a data point, training a regular deep neural network is highly computational expensive. CNNs can overcome this challenge and are therefore widely used for image classification tasks. Convolutional layers are only partially connected with the previous layer which reduces the total number of connections needed. (Géron, 2019) Its ability to automatically learn and identify intricate patterns and structures in the data makes it a strong contender for complex image classification tasks (Sultana et al., 2018). Additionally, CNNs can effectively learn hierarchical representations (Sharma et al.,

2018; Szegedy et al., 2015), which are highly desirable for identifying of *P. vivax* infection involves recognizing patterns at various levels of complexity in the blood smear images.

4. Methodology

4.1. Dataset Description

The public available image dataset provided by the Broad Bioimage Benchmark Collection (BBBC) website (https://bbbc.broadinstitute.org/BBBC041) was used. (Hung & Carpenter, 2017) It holds 1.228 images that were manually created from *P. vivax* parasite infected humans in Manaus, Brazil, and Thailand. It includes over 80.000 cells. For each of those cells bounding box coordinates and manually created labels were provided. The labels are either uninfected (including Red Blood Cell and leukocytes class), infected (including ametocytes, rings, trophozoites, and schizonts) or difficult if not identifiable by the experts.

4.2. Data Pre-Processing

To begin with, the data is preprocessed as visualized in Figure 1. First, two datasets from the BBBC were loaded. It must be mentioned that the original data download was already divided into training and test set. However, as the criteria for the division were unclear, a random division was preferred. Thus, in the second step both datasets were merged. Several pre-processing steps followed which will be explained in more detail in the following.

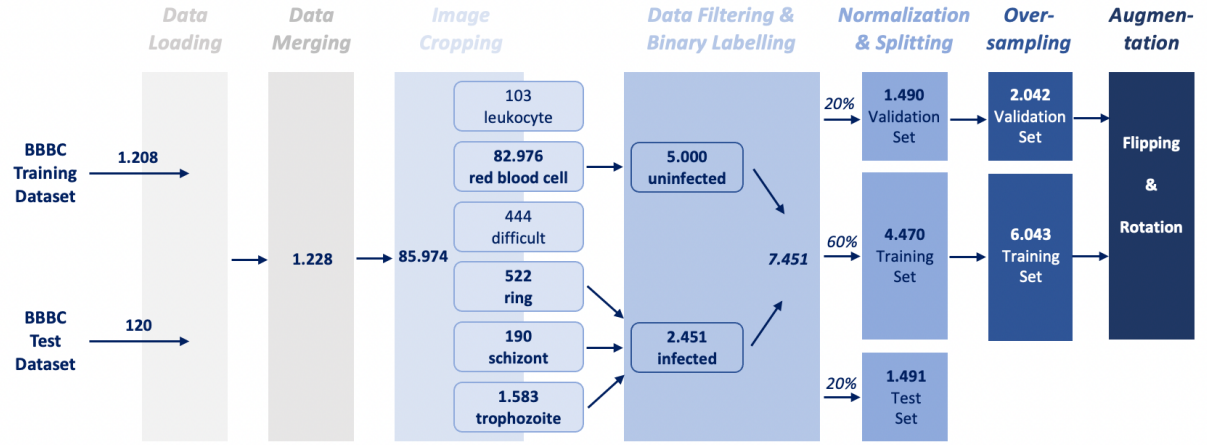


Figure 1: Pre-processing Pipeline

4.2.1. Image Cropping

The dataset comprises a collection of images, each encompassing multiple cells. To develop a model which is specifically designed to distinguish between uninfected and infected cells, it was necessary to extract each individual cells. The dataset supplied bounding box coordinates, which served as the basis for cropping the images in this case. To execute this task, a script

was developed which iterates through the dataframe containing these coordinates. Each corresponding image was selected, and the corner points were utilized to crop the cell. Subsequently, the images were saved for further use. One original image and an excerpt of some of the resulting cropped images are presented in Figure 2.

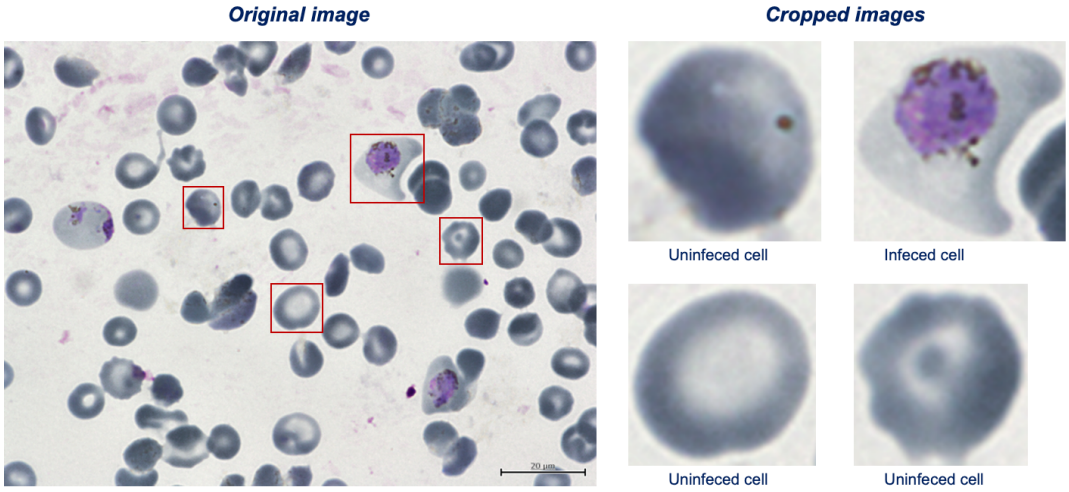


Figure 2: Original & Cropped Images

4.2.2. Data Filtering, Normalization & Splitting

Next, cell images labelled as difficult as well as leukocyte were removed. This leads to a dataset consisting of 85.042 images. However, the classes are highly imbalanced with 82.976 uninfected and 2.451 infected cells. To handle this issue, various pre-processing techniques were applied. Following Li et al. (2021) only 5.000 uninfected cell images were randomly selected to reduce the majority class and reduce computation time. Also, the assessment of duplicates revealed that no duplicates exist. Second, oversampling was applied which will be explained in detail below.

As visible in Figure 2, the boundary boxes had different sizes and thus, led to varying height and width of the images. Figure 3 shows the distribution of the images by height and width for images labeled as infected and uninfected. Commonly used sizes for images are 64x64, 128x128 or 224x224 pixel. The plots

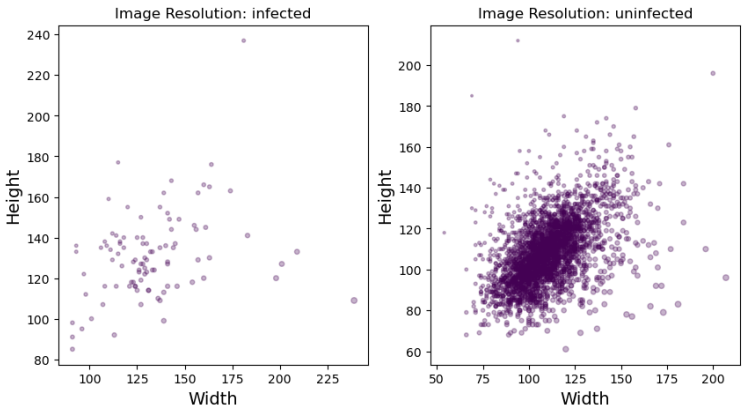


Figure 3: Distribution of Images by Resolution

show that most images are centered around 128x128 pixel. As a smaller or larger size would cause loss of information and detail, all images were resized into 128x128 pixel for consistency. Also, this size has been proven successful by previous research using the same dataset (Li et al., 2021).

In addition, all pixel values were scaled between 0 and 1. After normalizing the data, the image set was randomly divided into a training (60%), validation (20%) and test (20%) set. As mentioned earlier the original division into training and test set was not kept for this project.

4.2.3. Oversampling & Data Augmentation

Besides undersampling the majority class, oversampling is used to further balance the distribution of the two classes and reduce bias. In the case of malaria images, infected cells are more rare than uninfected cells and therefore underrepresented in the dataset. However, classifying the infected cells correctly and as accurate as possible is the main objective of the classification task. To overcome this issue, the Adaptive Synthetic Sampling (ADASYN) method is used on the training and validation set. ADASYN generates synthetic samples for the minority class by considering the density distribution of the feature space. It generates new datapoints in areas where the minority class is underrepresented and focus on samples that are more difficult to learn. (He et al., 2008) As result, the validation set holds 2.042 and the training set 6.043 images with an approximately even distribution of un- and infected classes.

Data augmentation is a common technique in machine learning, especially in deep learning, that enhances model performance by increasing the amount of training data available. It involves creating new samples by applying transformations to existing data. (Perez & Wang, 2017; Shorten & Khoshgoftaar, 2019). The study only employed geometric transformations, including rotation, and flipping as examined by Shifat-E-Rabbi et al. (2020) in the context of cell images. Unlike directional-dependent objects such as images of dogs, cells can be viewed from any orientation.

Thus, horizontal and vertical flipping proved viable as these transformations preserved the integrity of the class labels (Shifat-E-Rabbi et al., 2020). However, scaling and cropping were not employed due to the risk of losing important details in the cell images. The proximity of the boundary boxes around individual cells makes these transformations potentially detrimental, as they could diminish or eliminate the informational value of the image.

The *ImageDataGenerator* from TensorFlow was used for an efficient application of these transformations, eliminating the requirement to store all augmented images.

4.3. Modelling Framework

4.3.1. Convolutional Neural Networks

As described above, CNN is a commonly used image classification algorithm due to reducing the computational complexity. The images used in this project contain 128 x 128 pixels with three RGB color channels which makes 49.152 total data points. If all pixels and RGB channels are used as an input for a neuron in the first layer, over 2.4 million connections are needed to build the neural network. Thus, the partial connection between convolutional layers is very helpful for this study. (Géron, 2019)

In this project, two CNN models are developed. The first one was built from scratch and the second one adapted from a similar classification task. Both models use *Adam* as an optimizer with a learning rate of 0.001. *Adam* was chosen as it outperforms other optimizers and results in high performance in a similar study conducted by Kumar et al. (2020) comparing different optimizers for malaria detection. During training, the binary cross entropy loss is minimized and evaluated with the accuracy metric. To prevent overfitting, early stopping is implemented that terminates the training process when the validation error increase which is an indication for overfitting (Géron, 2019). Overfitting occurs when the performance is high on the training data but low on classifying unseen instances in the validation set (Géron, 2019).

The first CNN model was built by starting with a simple model architecture with one hidden layer. Deep neural networks with more hidden layers tend to reach better performance, converge faster, and generalizes well on unseen data (Schilling et al., 2018). With hidden layers, the number of neurons towards the output layer are reduced. This hierarchical architecture is similar to the way real world data is structured, allowing the model to learn low-level structures at the layers close to the input layer and high-level structures towards the output layer. (Géron, 2019) For this reason, more hidden layers were added iteratively while controlling for overfitting.

According to Géron (2019), observing the learning curves reveals a pattern of overfitting when the training loss consistently decreases, but the validation loss begins to rise or remains stagnant. Additionally, the gap between the training and validation curve should be small indicating that the model can generalize well on unseen data (Géron, 2019). The first CNN model consists of five convolutional layers, five max pooling layers and two dense layers as can be seen in Figure 4. Two dropout layers are added to prevent overfitting. Dropout is a regularization technique that randomly selects a fraction of the input neurons, in this case

pixels, which are not considered in the current epoch. The final model has approximately 1.2 million trainable parameters.

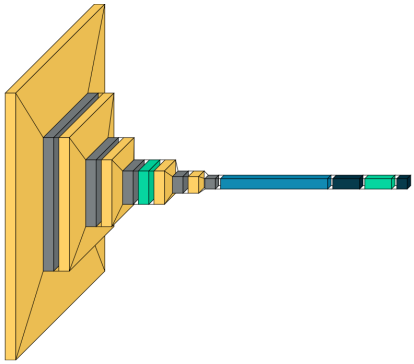


Figure 4: Architecture CNN Model 1

Figure 5: Architecture CNN Model 2

The second CNN model is adapted from Liang et al. (2016). As mentioned in section two, they reached 97.37% accuracy with a 16-layer CNN for malaria infection detection. In their study, the input images were of size 44 x 44 x 3. However, since the task at hand involves images with dimensions of 128 x 128 x 3, slight modifications were made to the model architecture to accommodate for the larger input size. Specifically, the filter size of the second convolutional layer was increased. After training the model with the cell images, performance was very low, indicating that the model is unable to learn the features of the images. Although this model architecture worked well for a similar classification task, the model architecture needs to be adjusted to accurately learn the features of the new images. To enhance the model's performance, two modifications were made. First, the size of the first max-pooling kernel was reduced. Second, the model was simplified by removing convolutional layers that immediately followed another convolutional layer. With these adjustments, the second CNN model has approximately 6.7 million trainable parameters. Figure 5 shows the architecture of the second CNN model, that consists of four convolutional layers, one max pooling and one average pooling layer as well as two dense layers.

4.3.2. Support Vector Machine

The performance of the two CNN models is compared with a SVM model. For the SVM model, the images are not augmented with the *ImageDataGenerator* as described in section 4.2.3 as the SVM serves as a baseline model. Before training the SVM model with the preprocessed images, Principal Component Analysis (PCA) is applied to reduce the dimensionality of the dataset while preserving 99% of the variance. According to Chandra & Bedi (2021), extracting informative features is essential to obtain high accuracy on a classification task. Furthermore, applying PCA in pre-processing step is essential to reduce computation time needed to train

high-dimensional features (Géron, 2019). The idea behind PCA is that not all pixels are needed to correctly classify an image. For instance, only certain areas within the cell are relevant to determine whether a cell is malaria infected. By identifying the principal components – the parts that contain the pixels with most of the information - the image dimensions are reduced to 460 instead of 49.152 pixels. (Géron, 2019)

SVM can be used for both linear and nonlinear data. To classify non-linearly separable data, kernel functions are used to transform the images into a high-dimensional space and consequently find the optimal decision boundaries in this new high-dimensional space. Various types of kernels are used such as linear, polynomial, or the radial basis function (RBF). The choice of the optimal kernel for a given classification task depends on the data and is therefore found by performing grid search. (Chandra & Bedi, 2021)

The right choice of hyperparameters in the SVM model determine how well the model performs for the given classification task. The hyperparameter C controls for the balance between maximizing the margin and minimizing the training errors. For instance, a small C value results in a wide “street” separating the two classes but in datapoints will be within the boundary. By tuning the hyperparameter C, the model can be regularized to control for the complexity of the model and prevent overfitting which allows the model to better generalize on unseen data. Another relevant hyperparameter that can prevent overfitting in the SVM model is gamma which influences how the decision boundary adapts to the different data points. A high gamma adjusts the decision boundaries to the individual data points which could lead to overfitting and a bad capability to generalize. (Géron, 2019)
The optimal values for C and gamma as well as choice of the kernel largely depend on the dataset. To find the optimal hyperparameters, a threefold grid search cross validation is performed. This search refers to the technique that evaluates different combinations of parameters based on cross validation. Thereby, the model is fitted on three randomly splitted subsets of the dataset and evaluated on the validations set. The optimal parameters obtained from the grid search are RBF kernel, C = 100 and gamma = 0.0001. The relatively high value for C and the low value for gamma indicate a stricter regularization of the model (Géron, 2019).

5. Results

5.1. Comparison of Models

The results (Table 1) indicate that CNN1 exhibits the highest performance across all metrics. It achieved the highest accuracy of 0.9826, indicating a very high proportion of true results

(both true positives and true negatives) among the total number of cases examined. It also has the highest precision, recall, and F1-score averages of 0.98, which demonstrate its balanced performance in terms of both precision (correct positive predictions out of total predicted positives) and recall (correct positive predictions out of total actual positives). The latter is crucial for identifying every malaria infection and thereby selected as the most important comparison metric. Despite potential false positives, early detection and treatment lessen the severity of the illness (Landier et al., 2016).

The CNN2 is the second-best performing model, with an accuracy of 0.9772 and F1-score of 0.97. It also maintains a high level of precision and recall, making it a strong model despite being slightly outperformed by CNN1.
Meanwhile, the SVM as baseline model has the lowest metrics. While it maintains a decent absolute accuracy of 0.9269 and F1 score of 0.92, in comparison it falls behind the CNN models in every category.

Model	Accuracy	Precision	Recall	F1-Score
SVM	0.9269	0.92	0.91	0.92
CNN1	0.9826	0.98	0.98	0.98
CNN2	0.9772	0.97	0.98	0.97

Table 1: Model Results

The confusion matrices in Figure 6 substantiate the earlier metrics. SVM has the most misclassifications, reflecting its lower performance. CNN1 exhibits the best performance with the least misclassifications, aligning with its top scores across all metrics. CNN2 is a close second, demonstrating a good but slightly lesser performance than CNN1. Most notably, CNN1s misclassified only 7 truly labeled infected cells (high recall) which is the most important evaluation metrics due to the high implications of wrongly identifying infected cells. At the same time, the misclassified uninfected cells sum up to 19. Thus, though the precision is slightly lower than the recall, it is still the lowest of all models.

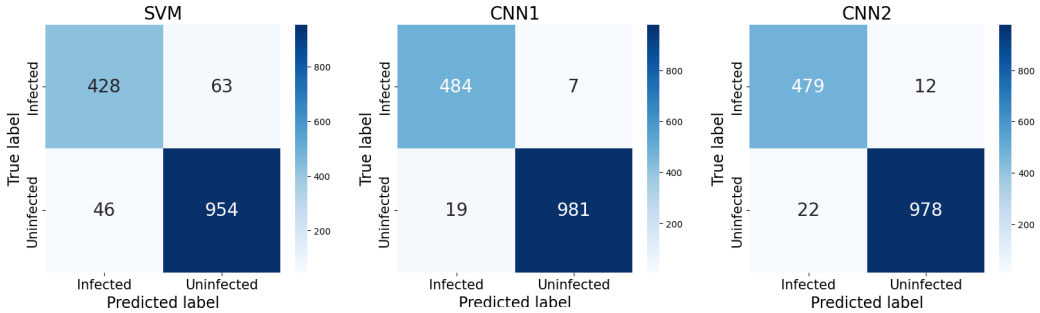


Figure 5: Confusion matrices

5.2. Run time and complexity comparison

Runtime analysis is vital in model evaluation as a quicker model with slightly less accuracy might be more desirable. Figure 7 visualizes the models' runtimes, showing negligible differences in training time.

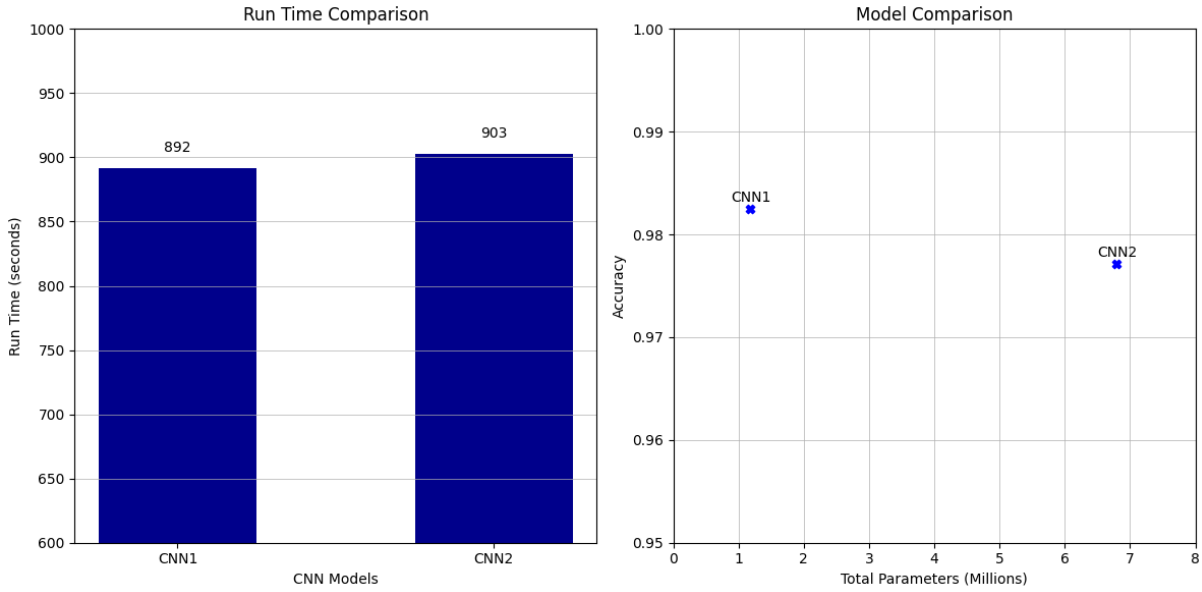


Figure 6: Run Time & Model Comparison

The scatter plot in Figure 6 reveals that, despite having fewer trainable parameters, the self-built CNN1 outperforms CNN2. All models were trained on a Google Colab Pro+ machine with 25,5 GB RAM and a Tesla T4 GPU, a processor designed to for deep learning training. The exclusion of the SVM model from the run time comparison arose from the fact that the CNN models had the advantage of utilizing the GPU resources provided by Google Colab, whereas the SVM model was trained without using additional GPUs. Thus, a consistent baseline for comparison was not established.

5.3. Training Performance

Examining the loss and accuracy plots of the CNN models (Figure 8 and 9) training process reveals several insights into the model's performance and learning efficiency. The CNN 1 model (Figure 8) was initially trained for 50 epochs, employing the "early stopping" callback which halted the learning process at epoch 30. Throughout the training period, the loss graph exhibits a consistent downward trend, although it eventually reaches a plateau. Notably, the validation curve displays occasional spikes, which can be attribute to the utilization of a batch size of 64. It is worth mentioning that a smaller batch size introduces more variability in the

validation curve, as smaller batches are less likely to provide a comprehensive representation of the entire dataset (Radiuk, 2017). Despite these intermittent spikes, the validation curve generally closely mirrors the progression of the training curve.

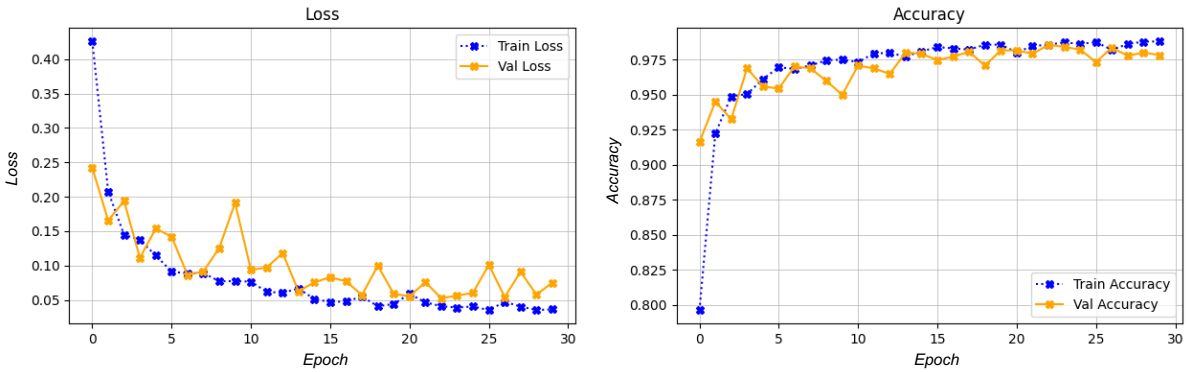


Figure 7: Loss & Accuracy of CNN1

During the training phase, the loss graph of the second CNN model consistently shows a decline, although it eventually stabilizes (Figure 9). The validation curve, however, demonstrates some sharp rises in loss after the 10th epoch in the validation set.

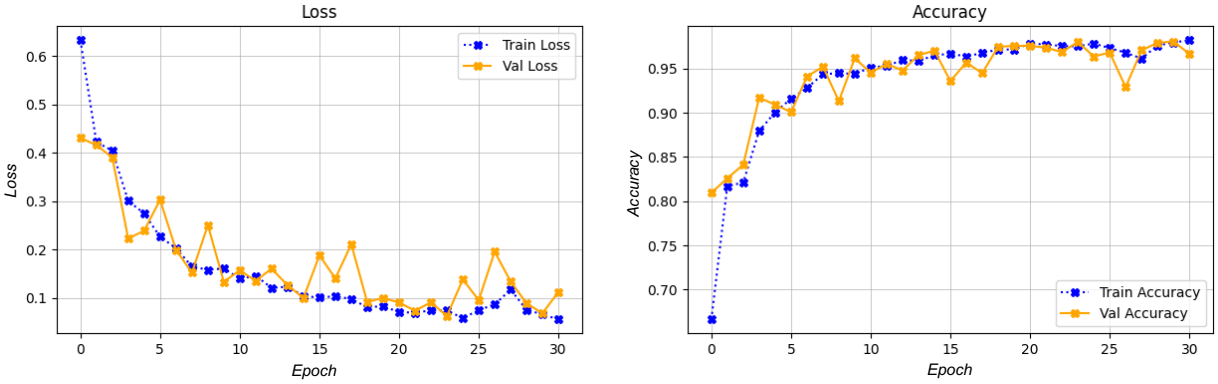


Figure 8: Loss & Accuracy CNN2

For both models, neither underfitting nor overfitting seems to be an issue, as both models exhibit a consistent and promising performance on both the training and validation data. This is indicative of a well-trained model that should prove robust and reliable in its predictions. (Dietterich, 1995)

6. Discussion

6.1. Comparison of Results

The first CNN model recorded an outstanding accuracy of 98.6%. CNN2, also demonstrated excellent performance, with an accuracy score of 97.7%. The SVM model showed a lower accuracy of 92.6%, thus reinforcing the superiority of CNN models in this context. These results

are in line with previous research, emphasizing the advantage of CNN models for binary classification tasks, especially in differentiating malaria-infected cells from uninfected ones. Also, the conducted study validates that binary classification improves overall accuracy, outperforming multilabel classification models using the same BBBC dataset like Li et al.'s CNN model (98.3%) and Meng et al.'s (94.17%).

Both CNN models were evaluated without image augmentation, considering findings from Shifat-E-Rabbi et al. (2020) that suggested worse results with augmented images for cell classification. This test showed clear overfitting, which confirms that image augmentation is necessary in this case.

6.2. Error Analysis

To assess the model's predictions, the types of errors are analyzed. The model misclassifies some images as shown in exemplary in Figure 10 below. A predicted probability close to 0 means the image is classified as infected and close to 1 as uninfected. By visually inspecting the cell images, it cannot be determined with certainty whether the model made an actual error or if the labels are wrongly classified. As an example, the third image which the model classifies as infected, while the true label indicates it is uninfected. One potential explanation for this error could be that it seems that some cells are overlapping. To evaluate these errors, domain experts would need to reevaluate the images and potentially relabel the images.

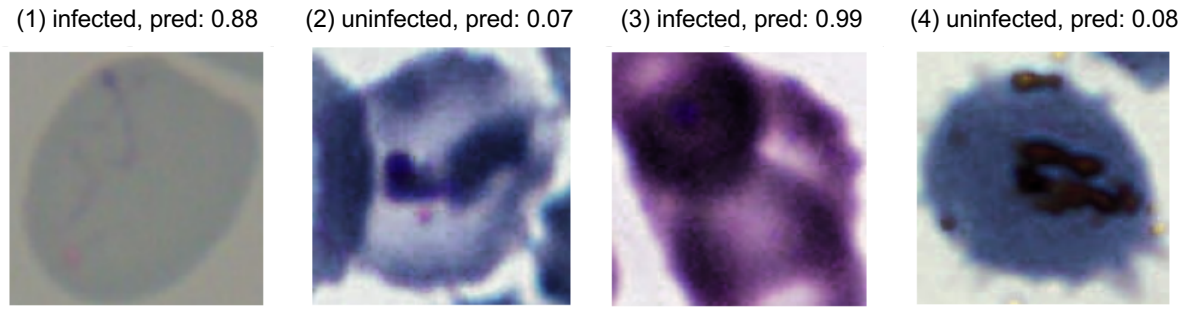


Figure 9: Misclassified Images

6.3. Practical & Ethical Implications

Considering practical implications, the developed models could be integrated into automated microscopy systems or mobile applications. This would enable real-time diagnosis of malaria infections. Further, such technologies could reduce the burden on skilled laboratory personnel. Automation can increase efficiency, allowing healthcare professionals to focus on other tasks which is especially helpful in low resource countries that are severely affected by malaria. On a higher level, those models could monitor malaria developments on a broader level: disease

trends can be tracked, and hotspots identified. For instance, this information can guide resource allocation, enhancing the overall malaria control strategy.

Several ethical implications must be addressed. Firstly, sensitive data was used, including patient images and health information. Thus, it is necessary to ensure proper anonymization and protection of patient privacy throughout the data collection, storage, and analysis processes. Regulatory oversight is crucial to ensure compliance with ethical guidelines and safety standards. Second, equal access to technologies leveraging those deep learning models should be aimed, particularly in resource-limited settings heavily burdened by malaria. A focus must lie on affordability, availability, and accessibility of diagnostic tools from a social justice point of view. Lastly, while those deep learning models could be a valuable support for healthcare professional, it must be emphasized that they should not replace them. Human experts should be introduced to the tools, how to use them and what their limitations are. For patience it is essential to have an educated and trained medic.

6.4. Limitations

The project demonstrates several limitations that warrant further exploration. Firstly, the binary classifier was designed solely distinguishing infected from uninfected malaria cells. Thus, additional factors such as cell type or stages of infection were not considered. This binary approach may limit the model's utility in broader applications. For instance, understanding variations in infection patterns among different cell types or infection stages could be invaluable for medical research and treatment development.

Secondly, the model's performance could be overestimated due to the homogeneous nature of the data used, which comprises only already segmented and cropped malaria cells. Its ability to generalize to other cell types or conditions remains untested, and there's no guarantee of comparable performance when applied to different images. To enhance the model's versatility, transfer learning techniques could be employed to adapt it to other cell classification tasks.

Further, the dataset holds 444 images that were labeled by the experts as “difficult”. Thus, especially those cases would be interesting to be labeled by machine learning as it could detect patterns medical personal cannot. For instance, an unsupervised algorithm could be trained, maybe revolutionizing malaria identification.

Lastly, it was opted to use RGB color data, despite the computational advantages of grayscale image data. This choice assumed that color information was vital in distinguishing infected cells. However, the potential performance trade-offs of using grayscale instead of RGB data could be further investigates. This could lead to efficient models with comparable accuracy.

7. Conclusion & Future Work

This study aimed to contribute an accurate identification of *P. vivax* malaria cases. It was an attempt to close the existing research gap regarding this parasite and support a cost- and highly trained labor-efficient malaria recognition. Therefore, an SVM and two CNN models were trained and tested on the BBBC dataset. Results show that the first CNN model performed the best, resulting in an accuracy of over 0.982. In comparison, the second CNN model yield an accuracy of 0.977. The by other researchers suggested higher performance of deep learning models compare to machine learning models (SVM with accuracy of 0.926) was replicated. In conclusion, the present study provides evidence that CNN models are best in binary (infected and uninfected) classification of cells regarding malaria.

Based on those results, future work could extend the classification models to distinguish between *P. vivax* and *P. falciparum*. In the long run, other less prevalent species can be an interesting. This can contribute to more targeted treatment strategies and enhance our understanding of different parasite behavior. Also, the detection of malaria infections caused by new parasites could increase. Those models could be implemented in tools easily accessible in countries that have malaria cases.

References

CDC. (2019, January 28). *Centers for Disease Control and Prevention (CDC)—Malaria Worldwide—How Can Malaria Cases and Deaths Be Reduced? - Diagnosis and Treatment*. https://www.cdc.gov/malaria/malaria_worldwide/reduction/dx_tx.html

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Chandra, M. A., & Bedi, S. S. (2021). Survey on SVM and their application in image classification. *International Journal of Information Technology*, *13*(5), 1–11. https://doi.org/10.1007/s41870-017-0080-1

Das, D. K., Ghosh, M., Pal, M., Maiti, A. K., & Chakraborty, C. (2013). Machine learning approach for automated screening of malaria parasite using light microscopic images. *Micron*, *45*, 97–106. https://doi.org/10.1016/j.micron.2012.11.002

Delas Peñas, K. E., Rivera, P. T., & Naval, P. C. (2017). Malaria Parasite Detection and Species Identification on Thin Blood Smears Using a Convolutional Neural Network. *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 1–6. https://doi.org/10.1109/CHASE.2017.51

Dietterich, T. (1995). Overfitting and undercomputing in machine learning. *ACM Computing Surveys*, *27*(3), 326–327. https://doi.org/10.1145/212094.212114

Dong, Y., Jiang, Z., Shen, H., David Pan, W., Williams, L. A., Reddy, V. V. B., Benjamin, W. H., & Bryan, A. W. (2017). Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 101–104. https://doi.org/10.1109/BHI.2017.7897215

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (Second edition). O'Reilly Media, Inc.

Gopakumar, G. P., Swetha, M., Sai Siva, G., & Sai Subrahmanyam, G. R. K. (2018). Convolutional neural network-based malaria diagnosis from focus stack of blood

smear images acquired using custom-built slide scanner. *Journal of Biophotonics*, *11*(3), e201700003. https://doi.org/10.1002/jbio.201700003

Haibo He, Yang Bai, Garcia, E. A., & Shutao Li. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 1322–1328. https://doi.org/10.1109/IJCNN.2008.4633969

Hung, J., & Carpenter, A. (2017). Applying Faster R-CNN for Object Detection on Malaria Images. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 808–813. https://doi.org/10.1109/CVPRW.2017.112

Krämer, J., Senellart, P., & de Streel, A. (2020). *Making data portability more effective for the digital economy: Economic implications and regulatory challenges*. Centre on Regulation in Europe asbl (CERRE).

Kumar, A., Sarkar, S., & Pradhan, C. (2020). Malaria Disease Detection Using CNN Technique with SGD, RMSprop and ADAM Optimizers. In S. Dash, B. R. Acharya, M. Mittal, A. Abraham, & A. Kelemen (Eds.), *Deep Learning Techniques for Biomedical and Health Informatics* (Vol. 68, pp. 211–230). Springer International Publishing. https://doi.org/10.1007/978-3-030-33966-1_11

Landier, J., Parker, D. M., Thu, A. M., Carrara, V. I., Lwin, K. M., Bonnington, C. A., Pukrittayakamee, S., Delmas, G., & Nosten, F. H. (2016). The role of early detection and treatment in malaria elimination. *Malaria Journal*, *15*(1), 363. https://doi.org/10.1186/s12936-016-1399-y

Li, S., Du, Z., Meng, X., & Zhang, Y. (2021). Multi-stage malaria parasite recognition by deep learning. *GigaScience*, *10*(6), giab040. https://doi.org/10.1093/gigascience/giab040

Liang, Z., Powell, A., Ersoy, I., Poostchi, M., Silamut, K., Palaniappan, K., Guo, P., Hossain, M. A., Sameer, A., Maude, R. J., Huang, J. X., Jaeger, S., & Thoma, G. (2016). CNN-based image analysis for malaria diagnosis. *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 493–496. https://doi.org/10.1109/BIBM.2016.7822567

Meng, X., Ha, Y., & Tian, J. (2022). Neighbor Correlated Graph Convolutional Network for multi-stage malaria parasite recognition. *Multimedia Tools and Applications*, *81*(8), 11393–11414. https://doi.org/10.1007/s11042-022-12098-6

Narayanan, B. N., De Silva, M. S., Hardie, R. C., Kueterman, N. K., & Ali, R. (2019). *Understanding Deep Neural Network Predictions for Medical Imaging Applications* (arXiv:1912.09621). arXiv. http://arxiv.org/abs/1912.09621

Perez, L., & Wang, J. (2017). *The Effectiveness of Data Augmentation in Image Classification using Deep Learning* (arXiv:1712.04621). arXiv. http://arxiv.org/abs/1712.04621

Porcello, J. C. (2019). Designing and Implementing SVMs for High-Dimensional Knowledge Discovery Using FPGAs. *2019 IEEE Aerospace Conference*, 1–8. https://doi.org/10.1109/AERO.2019.8741916

Radiuk, P. M. (2017). Impact of Training Set Batch Size on the Performance of Convolutional Neural Networks for Diverse Datasets. *Information Technology and Management Science*, *20*(1). https://doi.org/10.1515/itms-2017-0003

Rahman, A., Zunair, H., Rahman, M. S., Yuki, J. Q., Biswas, S., Alam, M. A., Alam, N. B., & Mahdy, M. R. C. (2019). *Improving Malaria Parasite Detection from Red Blood Cell using Deep Convolutional Neural Networks* (arXiv:1907.10418). arXiv. https://doi.org/10.48550/arXiv.1907.10418

Rajaraman, S., Antani, S. K., Poostchi, M., Silamut, K., Hossain, M. A., Maude, R. J., Jaeger, S., & Thoma, G. R. (2018). Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ*, *6*, e4568. https://doi.org/10.7717/peerj.4568

Rajaraman, S., Jaeger, S., & Antani, S. K. (2019). Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images. *PeerJ*, *7*, e6977. https://doi.org/10.7717/peerj.6977

Schilling, A., Metzner, C., Rietsch, J., Gerum, R., Schulze, H., & Krauss, P. (2018). *How deep is deep enough? -- Quantifying class separability in the hidden layers of deep neural networks*. https://doi.org/10.48550/ARXIV.1811.01753

Sharma, N., Jain, V., & Mishra, A. (2018). An Analysis Of Convolutional Neural Networks For Image Classification. *Procedia Computer Science*, *132*, 377–384. https://doi.org/10.1016/j.procs.2018.05.198

Shifat-E-Rabbi, M., Yin, X., Fitzgerald, C. E., & Rohde, G. K. (2020). Cell Image Classification: A Comparative Overview. *Cytometry Part A*, *97*(4), 347–362. https://doi.org/10.1002/cyto.a.23984

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*(1), 60. https://doi.org/10.1186/s40537-019-0197-0

Sultana, F., Sufian, A., & Dutta, P. (2018). Advancements in Image Classification using Convolutional Neural Network. *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, 122–129. https://doi.org/10.1109/ICRCICN.2018.8718718

Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Veiga, G. T. S. da, Moriggi, M. R., Vettorazzi, J. F., Müller-Santos, M., & Albrecht, L. (2023). Plasmodium vivax vaccine: What is the best way to go? *Frontiers in Immunology*, *13*. https://www.frontiersin.org/articles/10.3389/fimmu.2022.910236

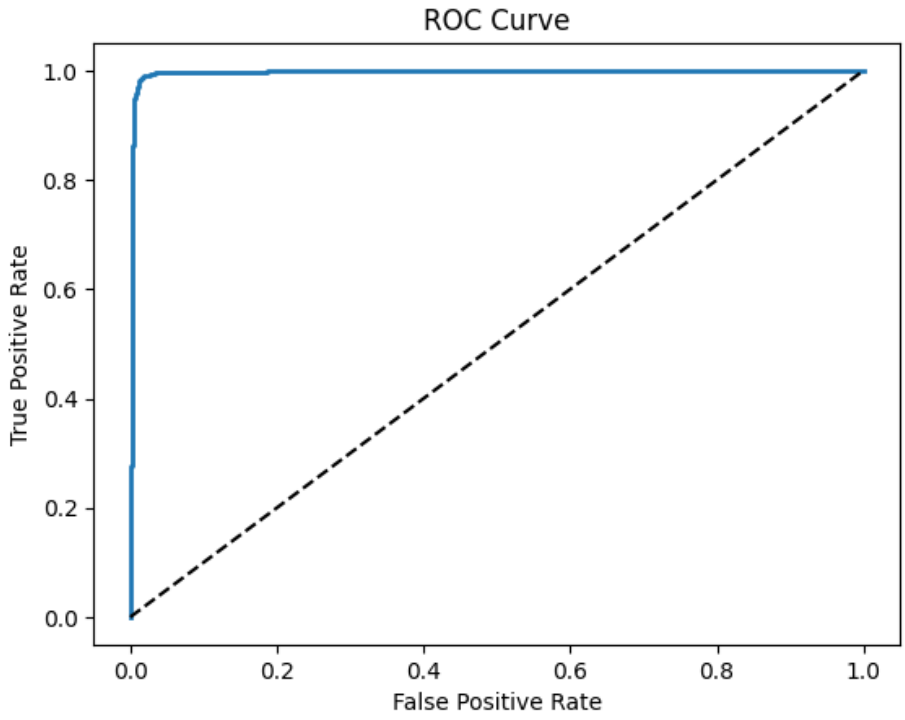
Vijayalakshmi & Rajesh Kanna. (2019). Deep learning approach to detect malaria from microscopic images. *Multimedia Tools and Applications*, *79*(21), 15297–15317. https://doi.org/10.1007/s11042-019-7162-y

World Malaria Report. (2022). World Health Organization. https://books.google.dk/books?hl=de&lr=&id=ST-hEAAAQBAJ&oi=fnd&pg=PR6&dq=world+malaria+report&ots=YYAVUcTkrm&sig=tzXlo1pGDNHTZXfu-4c2skzqmc0&redir_esc=y#v=onepage&q=world%20malaria%20report&f=false

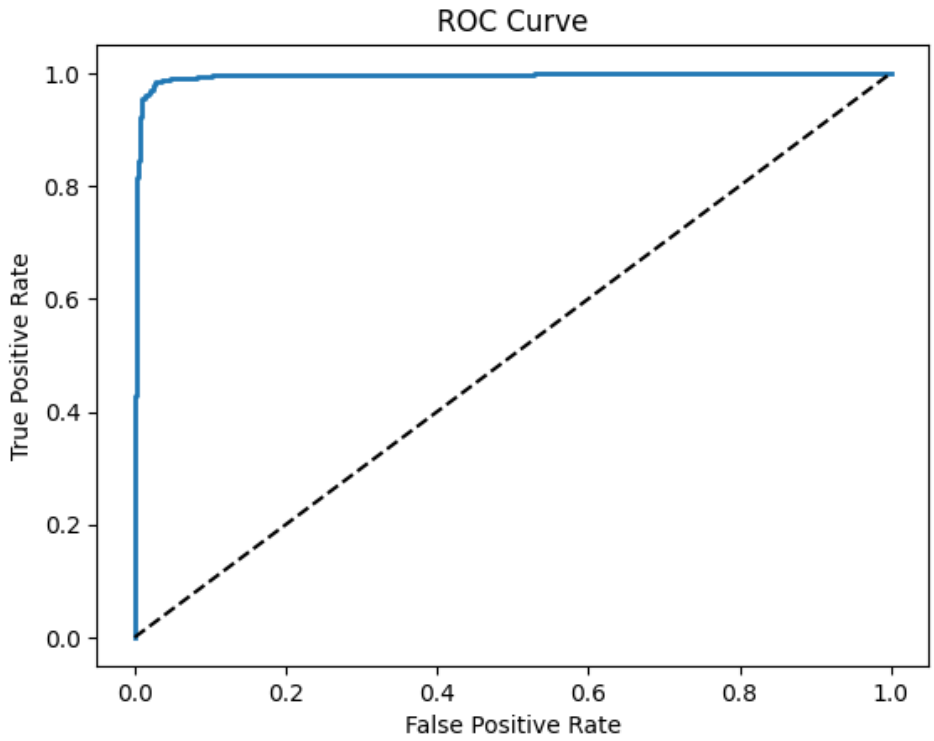
Zhao, O. S., Kolluri, N., Anand, A., Chu, N., Bhavaraju, R., Ojha, A., Tiku, S., Nguyen, D., Chen, R., Morales, A., Valliappan, D., Patel, J. P., & Nguyen, K. (2020). Convolutional neural networks to automate the screening of malaria in low-resource countries. *PeerJ*, *8*, e9674. https://doi.org/10.7717/peerj.9674

Appendix

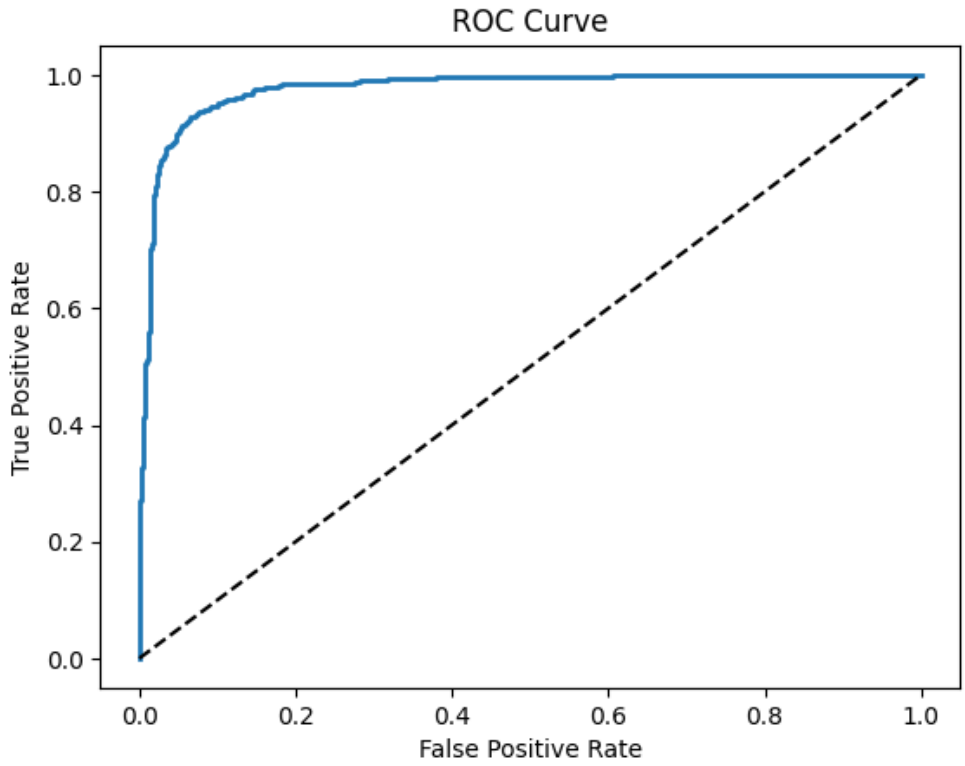
CNN1 – ROC Curve



CNN2 – ROC Curve



SVM – ROC Curve



Performance metrices divided by label

Label	Model	Precision	Recall	F1-Score
Infected	SVM	0.90	0.87	0.89
Not infected	SVM	0.94	0.95	0.95
Infected	CNN1	0.96	0.99	0.97
Not infected	CNN1	0.99	0.98	0.99
Infected	CNN2	0.96	0.98	0.97
Not infected	CNN2	0.99	0.98	0.98