Author: Hermela Shimelis

Table of Contents

Malaria detection

Final report

Executive summary

This project proposes the Convolutional Neural Network (CNN) model to aid in the automation of malaria detection and diagnosis in a healthcare setting. Automation of malaria detection will improve diagnostic accuracy by reducing variability due to human error, aid in reducing manual labor, and provide opportunities for the broad availability of tests. This model has the potential to be used in clinical settings because its sensitivity and specificity exceed those achieved by manual malaria detection in most low socio-economic regions. A successful CNN model can be deployed as an application loaded on a computer or smartphone that can be used to classify images of microscopic blood smear slides as either parasitized or uninfected.

Problem summary

The gold standard for malaria detection is through a microscopic evaluation of blood smears on a microscope slide. However, the process is labor intensive, the accuracy of the test heavily depends on the person evaluating the slide and there are limited testing sites due to the lack of trained personnel. Because the disease is treatable if detected early, a timely malaria diagnostic test is required if patients are suspected of being infected, increasing the demand for the test.

Proposed solution

In order to overcome the malaria diagnosis bottleneck, we propose the automation of malaria detection, using a deep learning algorithm with high sensitivity and specificity for classifying images of microscopic blood smear slides as parasitized or uninfected.

Solution design

Automation of malaria detection involves 1) blood smear slides imaging of a single red blood cell 2) Image classification using CNN model 3) quantification of parasitized cells 4) diagnosis. CNN is a well-known deep learning method that has been successfully used for image processing and classification, and thus will be used to solve the current image classification problem.

Train and test images:
- Train data: 24,958 images (12,582 parasitized and 12,373 uninfected)
- Test data: 2,600 images (1,300 parasitized and 1,300 uninfected)

Performance metrics: The most important performance metric for the current problem is parasitized cell recall. The objective is to minimize the misclassification of a parasitized cell. A model with the highest sensitivity and specificity is favored.

Exploratory analysis: Image visualization

An example of images used for training the CNN model

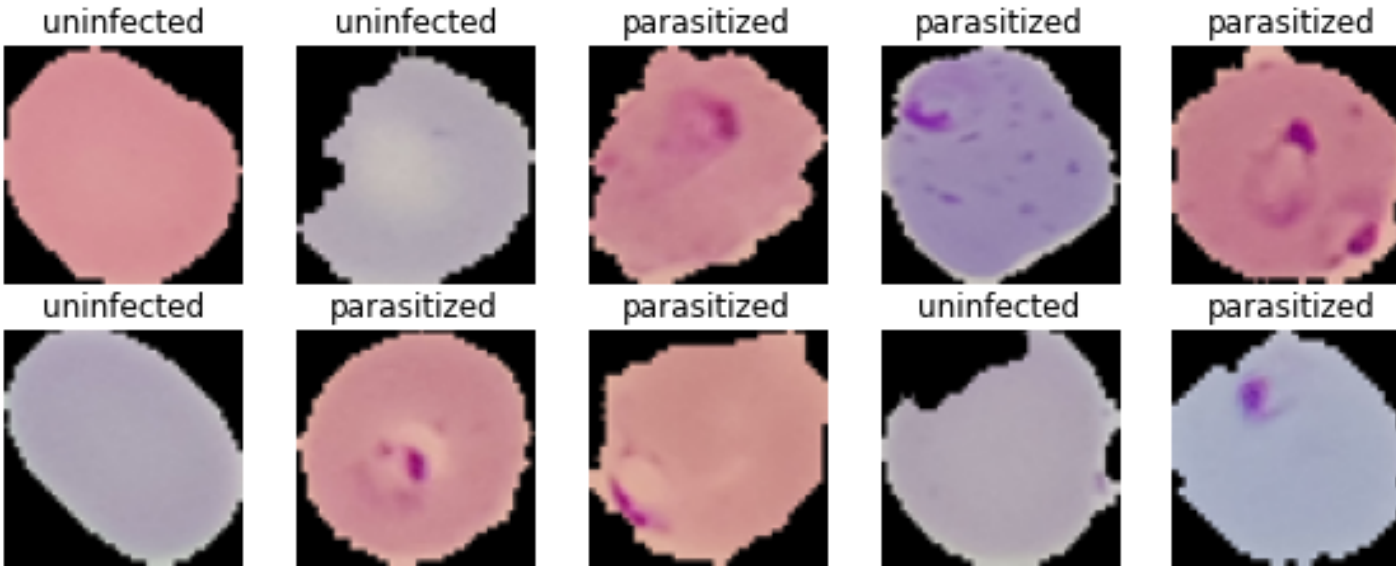

Figure 1: Uninfected and parasitized cell Images

Key takeaways:

- Because the difference between parasitized and uninfected images is small, extracting the high-intensity pixels in the positive images from the background is important, hence max-pooling layers are important to include in the CNN model.

- Generally, augmentation helps to minimize the false negatives by training the model with images with various transformations. Since the high-intensity pixels could be in a different area of the cell, augmented images will be used for model training.

- Because the small coloration difference between parasitized and uninfected cells is the main difference between the two groups, extracting the settled color differences from the background is important for the neural network. HSV is better for representing color than RGB.

Model 1: Base model

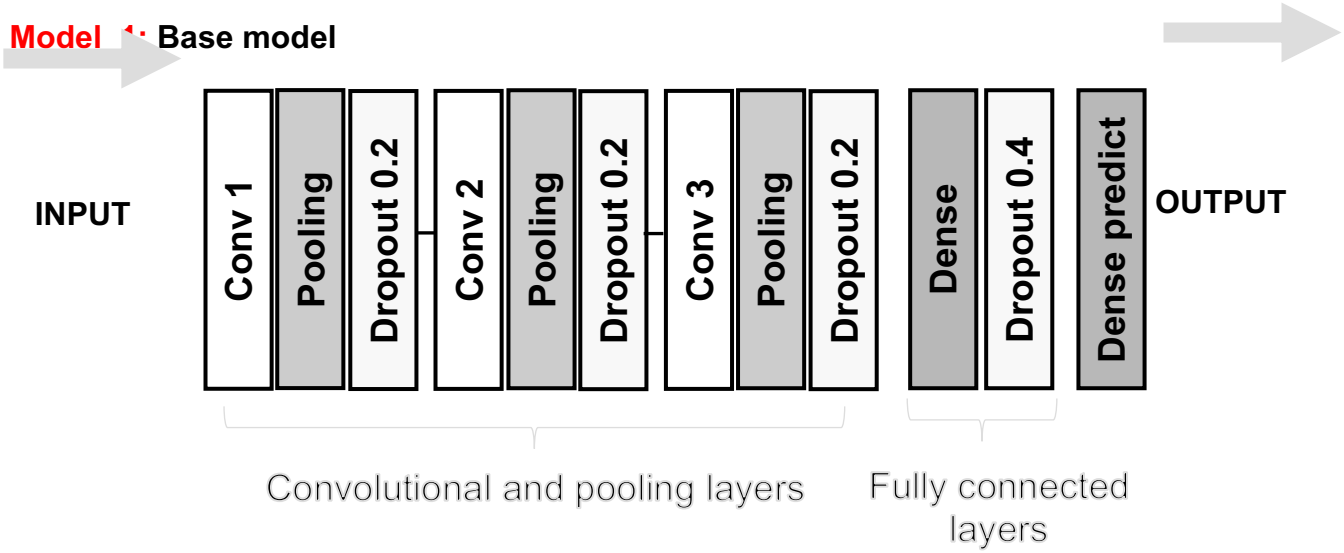

Figure 2: Model_1 architecture

Trainable params: 1,058,786

Table 1: Model_1 performance metrics

	Recall (%)	Precession (%)
Parasitized	97.4	99.0
Uninfected	99.0	97.4

Accuracy (%)	F1-score (%)
97.4	99.0

Takeaway: Because the model is not overfitting, the parasitized cell recall (97.4%) can be improved by reducing regularization and increasing convolutional layers.

Model 2: Tuned convolutional layers and regularization

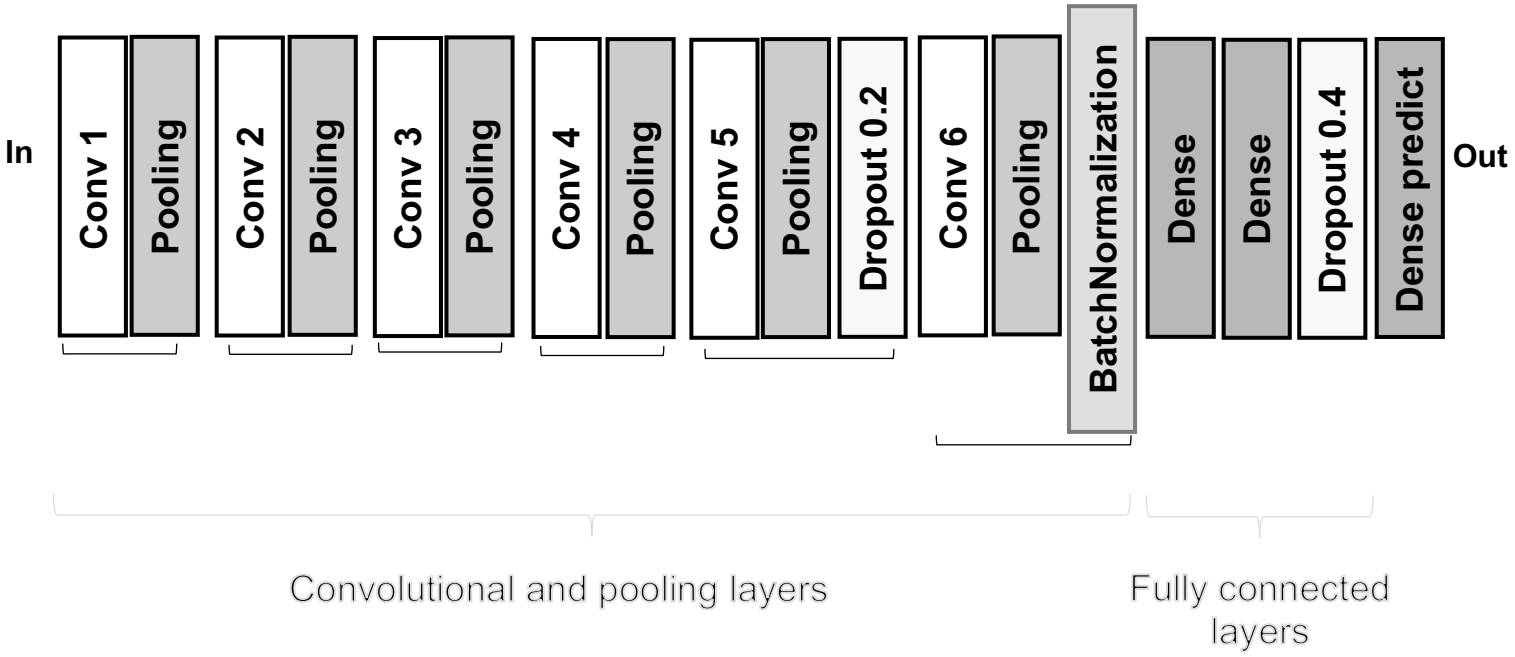

Figure 3: Model_2 architecture

Trainable params: 272,394

Table 2: Model_2 performance metrics

	Recall (%)	Precession (%)
Parasitized	98.7	98.9

Uninfected	98.9	98.7

Accuracy (%)	F1-score (%)
98.8	98.8

Takeaway: The recall on the parasitized image improved from 97.4% in Model_1 to 98.7 in Model_2 although the number of trainable parameters decreased about 4-fold compared to Model_1.

Model_3: Model 2 neural network trained with augmented images

Table 3: Image augmentation table

Augmentation type	Parameters
Horizontal flip	True
Zoom_range	0.05
Rotation_range	20
Width_shift_range	0.05
Height_shift_range	0.05
Shear_range	0.05

Table 4: Model_3 performance metrics

	Recall (%)	Precession (%)
Parasitized	97.8	99.3
Uninfected	99.3	97.8

Accuracy (%)	F1-score (%)
98.8	98.8

Takeaway: The recall on the parasitized image was not improved by image augmentation.

Model_4: Model 2 neural network trained with HSV color space images

Images in HSV color space

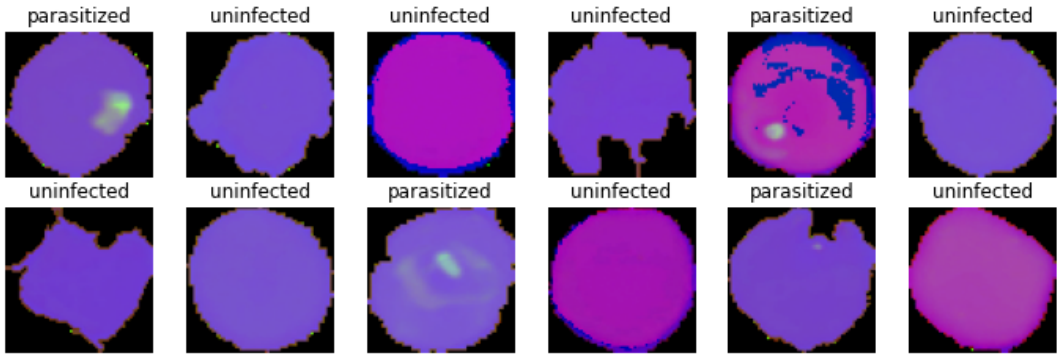

Figure 4: Uninfected and parasitized cell Images in HSV color space

Table 5: Model_4 performance metrics

	Recall (%)	Precession (%)

Parasitized	98.8	97.7
Uninfected	97.7	98.8

Accuracy (%)	F1-score (%)
98.3	98.3

Takeaway: The recall on the parasitized test images was not improved much by using HSV color space (98.8% in Model_4 and 98.7 in Model_2), while the specificity was reduced from 98.9% in model_2 to 97.7% in model_4. Therefore, using HSV color space didn't improve model performance.

Model_5: Based on transfer learning from VGG16 pretrained model

VGG16 pre-trained weights from the first 5 convolutional layers (blocks 1-5) will be used to build Model 3_vgg. Three dense layers will be added to the convolution blocks that will be frozen.

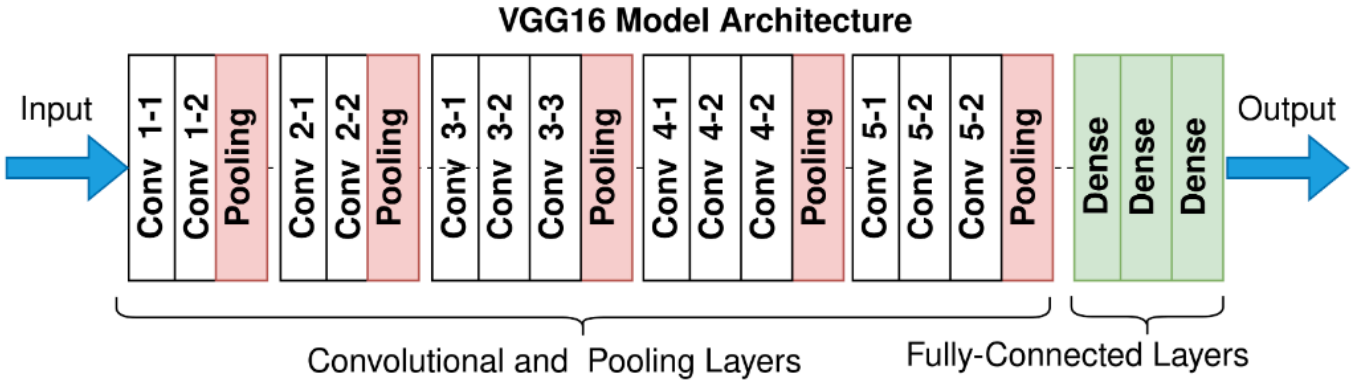

Figure 5: VGG16 blocks 1-5 and 3 dense layers

Trainable params: 14,714,688

Table 5: Model_5 performance metrics

	Recall (%)	Precession (%)
Parasitized	95.7	93.3
Uninfected	93.2	95.6

Accuracy (%)	F1-score (%)
94.4	94.4

Takeaway: Despite having a large number of parameters to train, this model did not outperform Model 2, which had far fewer parameters.

Model performance comparison

Table 6: Model performance on test data

	Model 1	Model 2	Model 3	Model 4	Model 5
Sensitivity (%)	97.4	99.2	97.8	98.8	94.0
Specificity (%)	99.0	98.3	99.3	97.7	95.9

Accuracy (%)	98.2	98.7	98.5	98.3	95.0

Final proposed model

Model 2 performs the best of all the models on the test data, with a sensitivity of 99.2%, a specificity of 98.3%, and overall accuracy of 98.7%. Because it would be detrimental to misclassify a parasitized cell image, a model with the highest sensitivity, and reasonably high specificity (>98%) is selected. The model generalizability is good across all models as model accuracy on both validation and test data is comparable.

Recommendations for implementation

The selected deep learning model achieved very high sensitivity (99.2%) and specificity (98.3%) for the classification of red blood cell images as parasitized or uninfected. Because the model performance is within the range of accuracy required for clinical lab tests, the proposed model has the potential to be further considered for implementation in malaria detection automation. The proposed model can be deployed as an application loaded on a computer or smartphone that can be used to detect malaria from images of microscopic blood smear slides.

The automation of malaria detection using the proposed deep learning model provides solutions to the malaria diagnosis bottleneck, however, there are some limitations to the automation of the process. Both benefits and limitations are discussed below.

Benefits of malaria detection automation using the proposed deep learning model:

- Quick and accurate malaria diagnosis
- scalable as it reduces the strain of the need for a human expert
- Reduce diagnosis accuracy variability due to the level of human expertise
- High sensitivity and specificity
- The selected model has a small number of trainable parameters, so it would not require a high computational capacity to deploy. This is an important factor in low socioeconomic regions where high-performance computing may not be available.

Challenges and Limitations of malaria detection automation using the proposed deep learning model:

- The current solution provides only partial automation of malaria diagnosis as the need for trained personnel to obtain blood smears on a microscopic slide and obtain single-cell images.
- There may be variability in the quality of blood smears and images, which may lead to the degradation of the algorithm's performance if the slide preparation has unfamiliar characteristics.
- The effectiveness of the current deep learning model in the detection of parasites of different stages has not been tested.
- There are 6 different strains of plasmodium parasites that cause malaria and the representation of each type in the train and test images is not known, therefore model generalizability across different strains is not known

Recommendation and key actionable items before deployment:

- Train the model using samples collected from different clinics to familiarize the model with different sample sets.
- The deep learning model may need to be retrained using clinic-specific samples
- The sample collection and imaging may need to be standardized to reduce inaccurate prediction due to variability in blood smear image preparations. For example, provide a camera for imaging the blood smear slides along with the malaria detection software.
- The automated malaria detection system requires parasitized cell counting capability to quantify the total malaria parasite per microliter of blood and this capability has not been included in the current classification model.

Automation of malaria detection using the current deep learning model provides a solution to the malaria detection bottleneck as it helps with the wide availability of tests, reduce the need for trained microscopist, and provides a highly sensitive and specific malaria detection. Automation of malaria detection is likely cost-effective since the deep learning algorithm can be installed as software on a desktop equipped with a camera.