

Malaria Detection

A Deep Learning Solution Approach

By: Nicolás Salazar B. (nicolas-salbe@github.io)

Executive Summary

Microscopy-based malaria diagnosis requires inspection by a professional microscopist who may need to handle large amounts of samples in a repetitive and error-prone procedure. A method or tool that can automate this procedure can lead to more convenient and accurate diagnosis.

Problem

Predict whether a red blood cell microscopy image comes from a malaria-infected blood sample or not.

Key takeaways

- Microscopy-based malaria diagnosis may leverage the power of Deep Learning techniques to benefit from better repeatability, accuracy, and scalability
- The proposed Deep Learning solution here shows high accuracy to differentiate parasitized from uninfected cells and yields high confidence that cells classified as uninfected are not parasitized cells
- The CNN model could result in reduced performance on unseen images processed differently than the training dataset
- Further validation of the CNN model is required to ensure trustworthy results

Solution

A 6-layer high-accuracy CNN classifier model capable of classifying red blood images as *parasitized* or *uninfected*.

Key Recommendations and Future Work

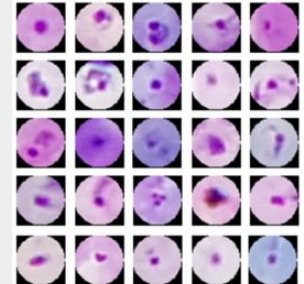
- Feedback from healthcare authorities, governments and regulating bodies is required to further develop the proposed solution
- Engineering costs and regulations need to be considered to determine the viability of the model

Problem Overview

- 247 million cases of malaria in 2021
- Prevention and treatment are a significant challenge
- Microscopy-based diagnosis is repetitive and prone to errors

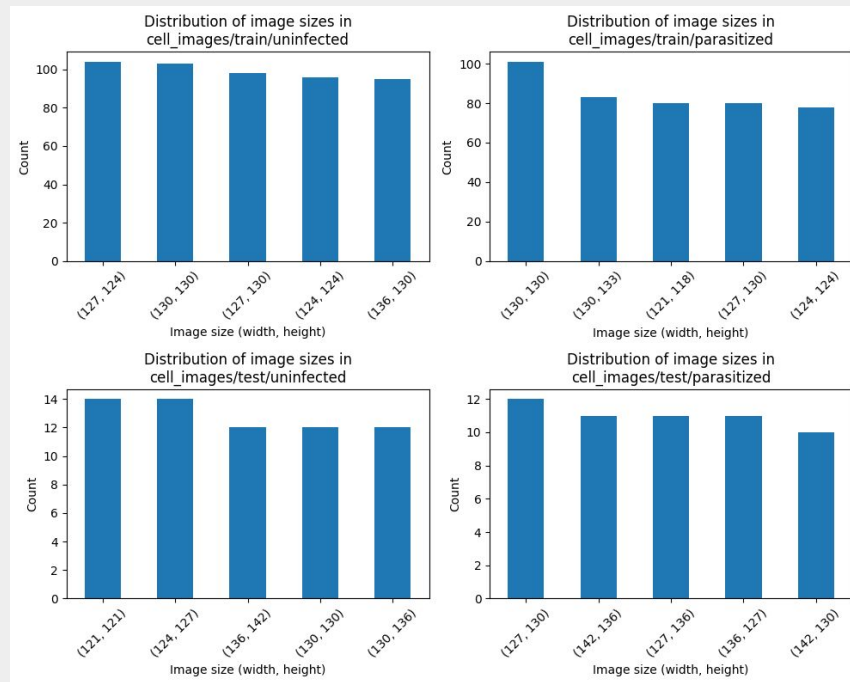


- Automation may lead to more convenient and accurate diagnosis
- A Deep Learning model may support malaria diagnosis by classifying red blood cells microscopy images as coming from a *parasitized* or an *uninfected* blood sample.



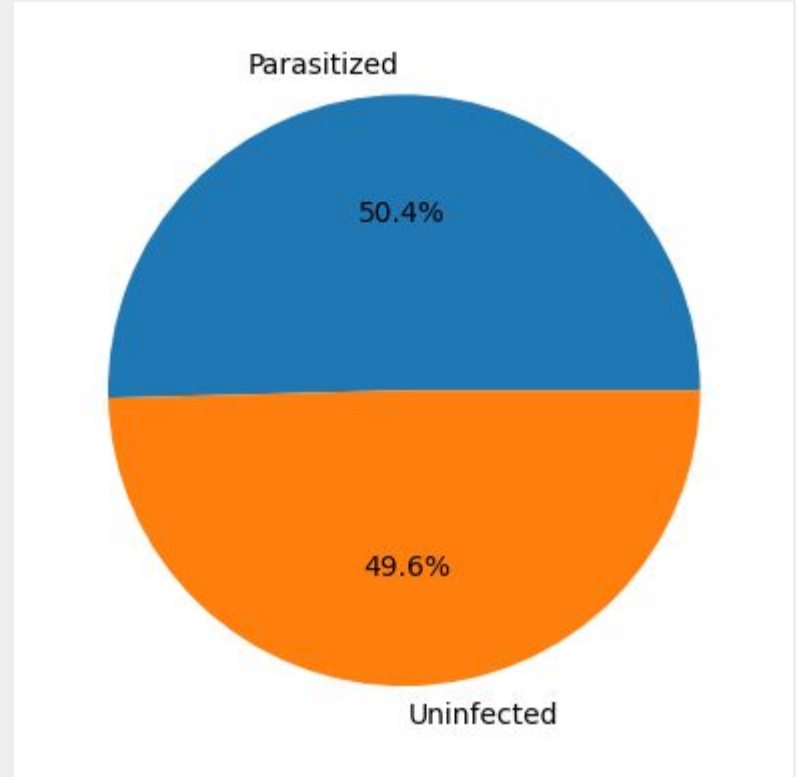
EDA

- Microscopy colored cell images dataset:
 - 24958 images for training
 - 2600 images for testing
- Images size:
 - most of the images are larger than 120x120 pixels
 - resizing to 64x64 pixels

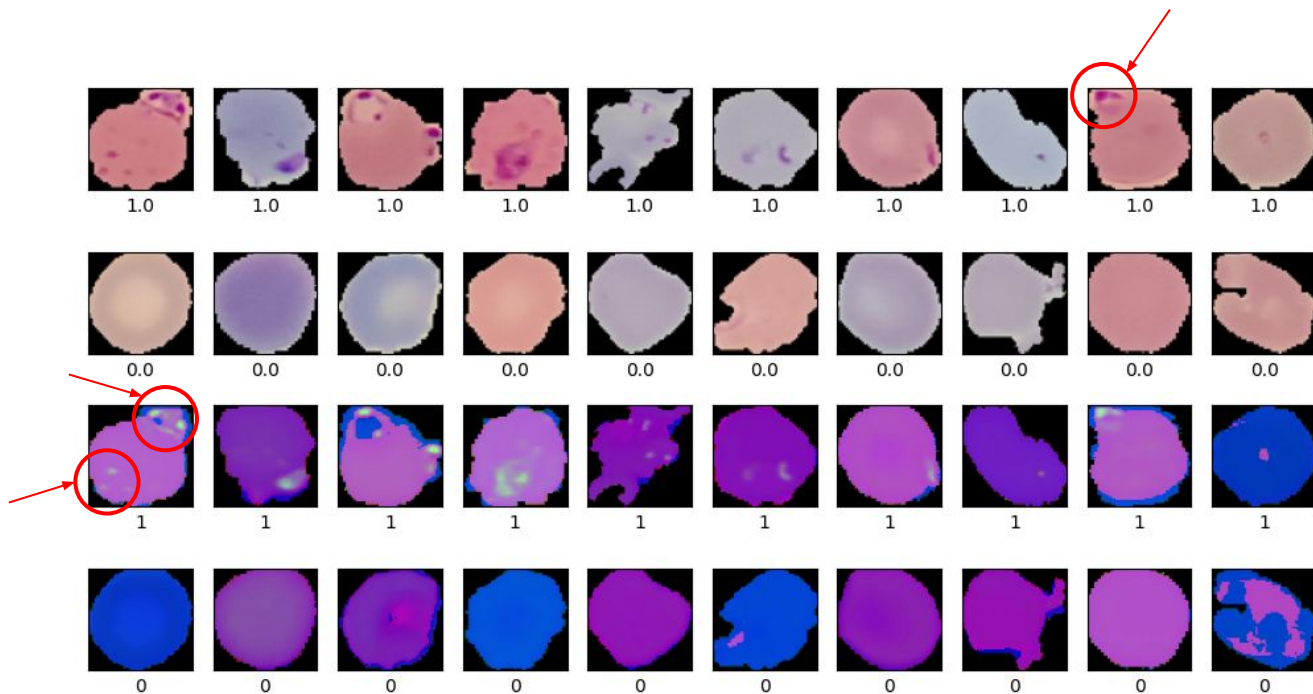


EDA

- Training dataset not perfectly balanced
 - $< \sim 1\%$ negligible



EDA

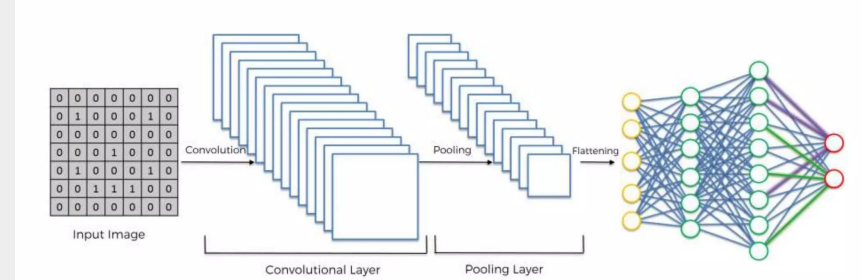
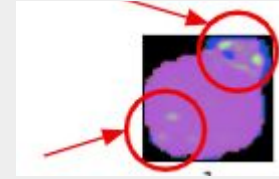
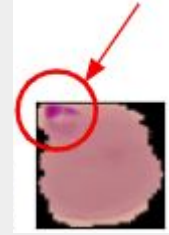


Solution Approach

- Key info is encoded in a stain pattern rather than in particular forms or locations



- Evaluate different CNN architectures and techniques that can help extracting these key features
 - Activation functions
 - Regularization methods
 - Data augmentation
 - Transfer learning



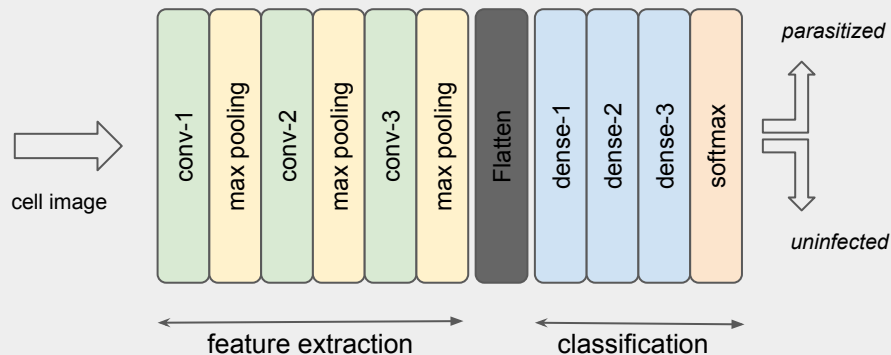
Model Comparison

	cnn_model_base	cnn_model_1	cnn_model_2	cnn_model_3	cnn_model_4	cnn_model_hsv
description	2x conv layers, 1 fc layer, relu, no regulariz...	3x conv layers, 3 fc layers, relu, dropout reg...	3x conv layers, 3 fc layers, Leaky relu, dropo...	cnn_model_1 with data augmentation	pre-trained vgg16 + cnn_model_1 fc layers	cnn_model_1 with HSV images
# trainable parameters	1053762	302178	302626	302178	76066	302178
epochs ran	18	15	23	21	29	14
training accuracy	0.9972	0.986	0.9862	0.9737	0.9474	0.9874
validation accuracy	0.9649	0.9764	0.9772	0.9808	0.9457	0.9782
test accuracy	0.9669	0.9812	0.98	0.9808	0.9385	0.985
test precision (parasitized)	0.9765	0.9692	0.972	0.9664	0.921	0.9839
test recall (parasitized)	0.9569	0.9938	0.9885	0.9962	0.9592	0.9862

Proposed Solution:

cnn_model_1

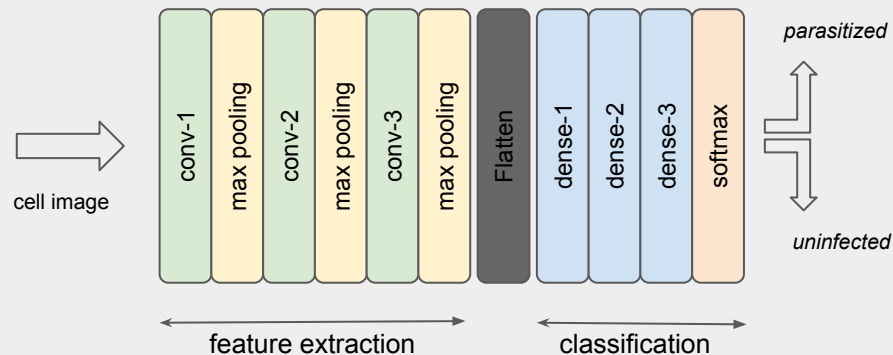
- Simple
 - Only 3 conv and 3 fc layers, max pooling, ReLU activation, and dropout regularization
- Efficient
 - Few parameters, relevant feature extraction
- High recall
 - Less than 1 person every 100 misdiagnosed when actually parasitized
- High accuracy
 - ~98% confidence that the prediction is correct



cnn_model_1	
description	3x conv layers, 3 fc layers, relu, dropout reg...
# trainable parameters	302178
epochs ran	15
training accuracy	0.986
validation accuracy	0.9764
test accuracy	0.9812
test precision (parasitized)	0.9692
test recall (parasitized)	0.9938

Proposed Solution: limitations

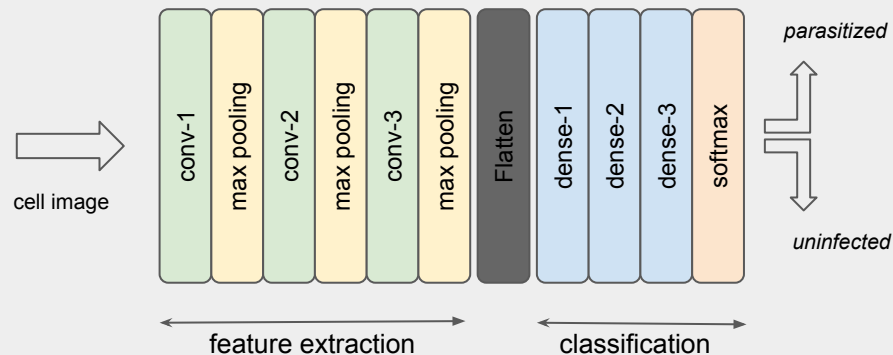
- Tests done on 2600 samples images
 - chance that the model doesn't generalize well on unseen images that differ too much from the original dataset
- Implicitly classifies the cells as of one class or the other based on probability >50%
- Better on recall than on precision for classifying parasitized cells
- Cell images should be larger or equal than 64x64
 - reduced image quality and biased result



cnn_model_1	
description	3x conv layers, 3 fc layers, relu, dropout reg...
# trainable parameters	302178
epochs ran	15
training accuracy	0.986
validation accuracy	0.9764
test accuracy	0.9812
test precision (parasitized)	0.9692
test recall (parasitized)	0.9938

Proposed Solution: improvements

- Expand training data with data augmentation
- Consider other optimizers than can avoid local minima
- Complement classification decision with other data



cnn_model_1	
description	3x conv layers, 3 fc layers, relu, dropout reg...
# trainable parameters	302178
epochs ran	15
training accuracy	0.986
validation accuracy	0.9764
test accuracy	0.9812
test precision (parasitized)	0.9692
test recall (parasitized)	0.9938

Recommendations and Future Work

- The proposed CNN solution presents good potential to benefit a microscopy-based malaria detection system with better repeatability, accuracy and scalability
- Engineering costs for integration and deployment need to be determined to better assess the viability of the CNN solution
- Healthcare authorities need to be involved to define the long-term impact that such a solution may represent in managing malaria in heavily affected areas
- New diagnosis campaigns should prioritize data collection and standardization

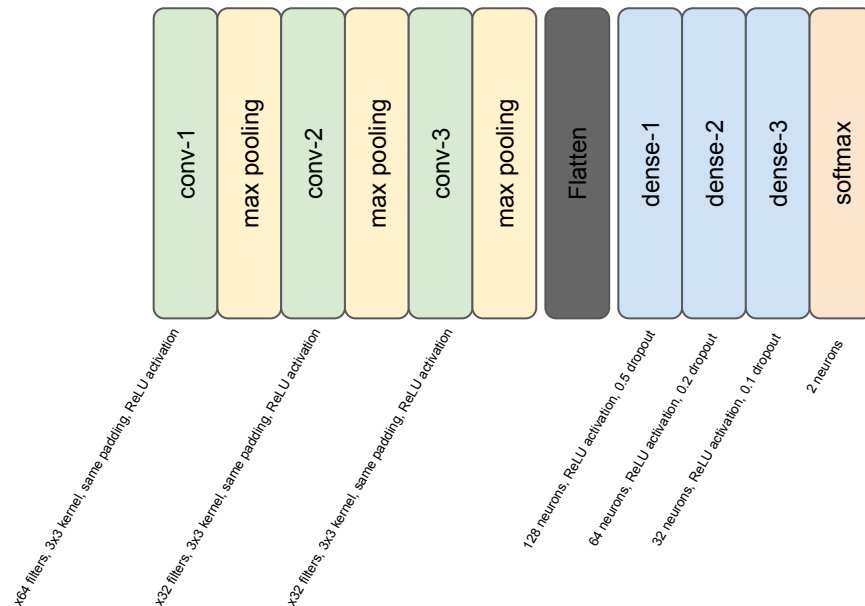
Thank you!

Sources

- <https://www.who.int/news-room/fact-sheets/detail/malaria>
- <https://www.who.int/teams/global-malaria-programme/case-management/diagnosis>
- <https://app.magicapp.org/#/guideline/LwRMXj/section/L0v9rE>
- <https://www.technologynetworks.com/analysis/articles/an-introduction-to-the-light-microscope-light-microscopy-techniques-and-applications-351924>
- <https://learnopencv.com/understanding-convolutional-neural-networks-cnn/>
- <https://www.mdpi.com/2075-4418/11/11/1994>

Appendix

Detailed *cnn_model_1* architecture



*Training performed with an Adam optimizer at a 0.001 learning rate.