

# Detecting Malaria in Cell Images Using CNNs and Tensorflow 2.0

**Course Project for: B55.1 WT Ausgewählte Kapitel sozialer Webtechnologien**

Lukas Wagner s0556753, Laila Westphal s0556469

## INTRODUCTION

The general topic of this project is to use Artificial Intelligence (AI), namely Convolutional Neural Networks (CNN) to detect malaria in blood cell images.

Being a life-threatening disease, malaria is caused by Plasmodium parasites being transmitted to humans by the bites of female Anopheles mosquitoes [1]. There are five parasite species responsible for malaria in humans. Two of them represent the biggest threat:

- P. falciparum, responsible for 99.7% of the estimated malaria cases in african countries.
- P. vivax, responsible for about 74.1% of the malaria cases in the Americas. The World Malaria Report 2018 states that amongst the areas where malaria cases increased by more than 20% is the WHO Region of the Americas [2]. 84% of that increase is due to malaria cases reported in Venezuela.

Even though malaria can be cured and prevented, according to the World Health Organisation (WHO), almost half of the world's population was at risk of it in 2017.

An article of the Korean Journal of Parasitology referenced by the National Center for Biotechnology Information (NCBI) written by N. Tangpukdee et al. states that the microscopic diagnosis of malaria requires a well trained microscopist [3]. In regions where malaria isn't endemic any longer, its diagnosis can be difficult because clinicians might not consider it as a possible cause. Microscopists might also fail to detect it, as they are not familiar with malaria and would possibly not recognize the parasites.

An important aspect of the proposed topic is that it is a great example of how AI can be used to save human lives. A CNN can be trained to be pretty proficient at detecting the disease. Once trained it can be used easily without producing a lot of costs and might therefore help to facilitate malaria diagnosis significantly.

## PROJECT OBJECTIVE

The aim of this project is to train a CNN to detect cells that contain malaria. Therefore a dataset from Kaggle, originally from the official NIH (US National Library of Medicine) Website<sup>1</sup> is being used. The dataset contains images divided into two categories: Parasitized and Uninfected.

The topic of detecting malaria in cell images fits into the context of machine learning and neural networks because it relies on image recognition. The foundation of this task is building and training a CNN, which requires a good understanding of the underlying

---

<sup>1</sup> <https://ceb.nlm.nih.gov/repositories/malaria-datasets/>

algorithms and mathematical functions. Due to the relatively small dataset, data augmentation is needed to produce good accuracy results.

Research about image classification and dataset sizes showed that bigger datasets are crucial for training CNNs in order to get stable results. As the malaria dataset is relatively small, data augmentation will be the first important step in our implementation. We anticipate to show the difference it makes by training the CNN with both, the original and the augmented dataset separately and compare the results.

At the end of the project, there should be an easy-to-reproduce jupyter notebook containing a brief discussion of the existing kernels and an implementation with our own approach. The main learning objectives we will focus on in the project are:

1. Getting familiar with TensorFlow 2.0
2. Deepen understanding of CNNs
3. Develop competencies and the necessary understanding for data augmentation

## METHODOLOGY

There are a lot of Kernels using this dataset on Kaggle. Some of them will be used as inspiration.<sup>2 3 4</sup> The architectures, score- and loss functions used by those kernels will be used as a guideline in developing our own approach. PyTorch and TensorFlow 2.0 have been considered for implementation. As TensorFlow has recently been updated to 2.0, we have decided to use it due to curiosity.

The Malaria dataset contains 27.558 images divided into two categories/folders: Infected and Uninfected. 13779 images of parasitized blood cells and an equal number of uninfected blood cells is available. All images are in PNG format. In order to enlarge the dataset we will use data augmentation techniques like rotation and flipping.

Keras has the `ImageDataGenerator` class<sup>5</sup> and TensorFlow has the `tflearn.data_augmentation.ImageAugmentation` class<sup>6</sup> to apply data augmentation techniques. The rotation technique will be applied by methods like `add_random90degrees_rotation()`, `add_random_rotation()` etc. Likewise, the flipping will be implemented by `add_random_flip_leftright()`, `add_random_flip_updown()`, etc. To process the images in batches they must be of fixed sizes, therefore, the method `tf.image.resize_images()` or a similar method will be used. The building and training of the CNN are based on the framework TensorFlow at version 2.0 and the high-level API Keras. After importing the TensorFlow implementation of Keras `tf.keras` the CNN can be put together with layers and models. Multiple layers will be stacked up to form a model. Keras provides methods to build layers including the activation function, the number of nodes and regularization schemes to prevent overfitting. The next task is to train the CNN, for which the method `tf.keras.Model.compile()` can be used. It configures the learning process of the model by setting the optimizer and the loss function. The optimizer is a specification of the training procedure. As an evaluation strategy, the loss function will provide feedback regarding the efficiency of the current structure and configuration of the CNN.

---

<sup>2</sup> <https://www.kaggle.com/ingbiodanielh/malaria-detection-with-fastai-v1>

<sup>3</sup> <https://www.kaggle.com/kushal1996/detecting-malaria-cnn>

<sup>4</sup> <https://www.kaggle.com/sid321axn/regularization-techniques-in-deep-learning>

<sup>5</sup> <https://keras.io/preprocessing/image/>

<sup>6</sup> [http://tflearn.org/data\\_augmentation/](http://tflearn.org/data_augmentation/)

## THEORETICAL BACKGROUND

Automatic identification of infected cells in general [4] and malaria infected cells in particular [5] has been studied by various research groups. CNNs are known to produce high accuracy results for image recognition problems and require little input from human experts besides labelled image data. According to Saurabh Yadav (Medium), CNNs are *"The most successful type of models for image analysis till date"* [6]. They consist of many different layers, the most important ones being convolutional and pooling layers [7]. In the convolutional layers, a kernel (or filter) is used to convolve the image. This process can be repeated several times and then pooling is applied, which leads to the recognition of some features, like edges in the image. The convolutional layers and the pooling layer are usually followed by a fully connected layer (FC) at the end.

## TIME MANAGEMENT AND TEAM ORGANISATION

Week	Class	Milestones	Lukas	Laila
13.05-19.05	Canceled	Research	x	x
20.05-26.05		TensorFlow, study kernels	Study Kernel a, compare results	Study Kernel b, compare results
27.05-02.06		Data augmentation, start implementation	Rotation technique	Flipping
03.06-09.06		Decide on functions, training CNN	x	x
10.06-16.06		Adapt, experiment	x	x
17.06-23.06		Hurry up	x	x
24.06-30.06	Project Work	Prepare presentation	x	x
01.07-07.07	Presentation		x	x

## REFERENCES

[1] WHO (2019, March 27). *Malaria*. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/malaria>

[2] WHO (2018, November). *World Malaria Report 2018*. Retrieved from <https://www.who.int/malaria/publications/world-malaria-report-2018/report/en/>

[3] Tangpukdee, N., Duangdee, C., Wilairatana, P., & Krudsood, S. (2009). Malaria diagnosis: a brief review. *The Korean journal of parasitology*.

[4] Hirimutugoda, Y. M., & Wijayarathna, G. (2010). Image analysis system for detection of red cell disorders using artificial neural networks. *Sri Lanka Journal of Bio-Medical Informatics*.

[5] Dong, Y., Jiang, Z., Shen, H., Pan, W. D., Williams, L. A., Reddy, V. V., & 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.

[6] Saurabh Yadav (2018, October 16). Brief Intro to Medical Image Analysis and Deep Learning. Retrieved from <https://medium.com/@saurabh.yadav919/brief-intro-of-medical-image-analysis-and-deep-learning-810df940d2f7>

[7] Harsh Pokharna (2016, July 28). The best explanation of Convolutional Neural Networks on the Internet!. Retrieved from <https://medium.com/technologymadeeasy/the-best-explanation-of-convolutional-neural-networks-on-the-internet-fbb8b1ad5df8>