

# Malaria Detection Using Blood Smear Images

Thosar, Sharayu Shekhar, [thosar.sh@northeastern.edu](mailto:thosar.sh@northeastern.edu)

Ruifeng, Song, [song.ruif@northeastern.edu](mailto:song.ruif@northeastern.edu)

Kumar, Priyanka Senthil, [senthilkumar.pri@northeastern.edu](mailto:senthilkumar.pri@northeastern.edu)

## Objective

This project aims to develop a deep learning model for the classification of malaria-infected and uninfected cells from blood smear images. Utilizing a dataset of 27,558 images sourced from the NIH Malaria Datasets, the goal is to build a robust model capable of accurately distinguishing between two classes: uninfected cells (healthy red blood cells) and parasitized cells (red blood cells infected with Plasmodium parasites). The project seeks to enhance diagnostic accuracy in malaria detection through advanced image classification techniques.

## Current State of Art

The survey of related published works highlights significant advancements in deep learning techniques for image classification, especially in medical applications such as malaria detection. Recent studies focus on developing efficient model architectures, including interleaved grouped filters and snapshot ensembling, which streamline computational complexity while preserving accuracy. Additionally, novel methodologies for dynamic model development are enhancing scalability and performance in multitask learning scenarios. These innovations represent a notable progression towards more efficient, interpretable, and scalable AI solutions for medical image analysis and beyond.

## Approach

1. **Data Acquisition and Exploration** - Explore the dataset to understand its structure, image dimensions, and distribution of classes.
2. **Data Preprocessing** - Preprocess the images by standardizing dimensions, normalizing pixel values, and potentially applying data augmentation techniques (e.g., rotation, flipping) to enhance model robustness.
3. **Model Selection** - Choose a suitable CNN architecture known for image classification tasks.
4. **Model Development** - Implement the CNN architecture using TensorFlow and design the network layers for feature extraction and classification.
5. **Model Training** - Train the CNN model on the preprocessed dataset using a training-validation split.
6. **Model Evaluation** - Evaluate the trained model on a separate test set to assess its performance metrics, including accuracy, precision, recall, and F1-score.
7. **Optimization and Fine-tuning** - Fine-tune the model by adjusting hyperparameters and exploring different optimization strategies.
8. **Documentation and Reporting** - Document the entire process, and provide clear explanations of decisions made and insights gained throughout the project.
9. **Ethical Considerations** - Adhere to ethical guidelines regarding the use of data.

## Dataset to be used

[Malaria dataset in TensorFlow Package](#) which contain 27,558 cell images with the total size of 317.62 MiB.

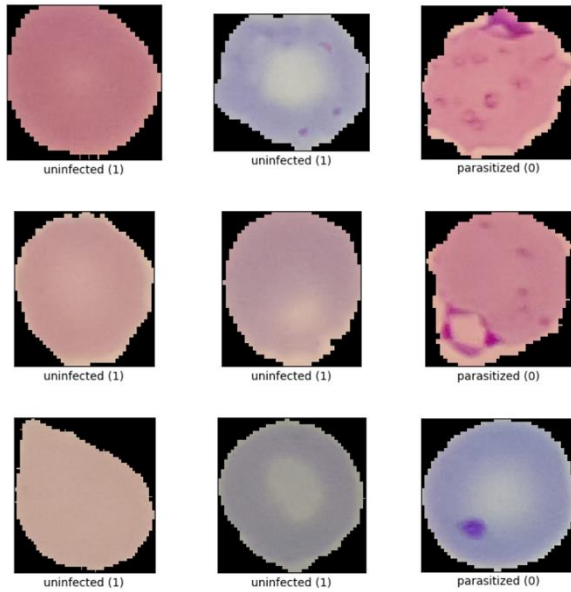


Figure 1. Data example image

## Progress Timeline

To be determined.

## Deliverables

Tech flow and relevant code to develop a deep learning model to predict whether a given cell is affected by Malaria or not.