

# Automated Diagnosis of Malaria Using Convolutional Neural Networks

Author: Jeyaseela Pandian V

## 1. Introduction:

Malaria remains a significant public health concern worldwide, particularly in regions with limited access to healthcare facilities. Timely and accurate diagnosis of malaria is crucial for effective treatment and disease management. In this report, we present the methodology, findings, and insights gained from the classification of malaria-infected and uninfected cells using image analysis techniques and machine learning.

## 2. Dataset Description:

The dataset comprises microscopic images of blood smear slides obtained from patients with suspected malaria infection. The images are collected at various magnifications and resolutions, capturing both infected and uninfected red blood cells. Each image is annotated with the corresponding class label (infected or uninfected). The dataset is balanced, containing an equal number of infected and uninfected cell images. The dataset is diverse, containing variations in cell morphology, staining quality, and background clutter.

## 3. Exploratory Data Analysis:

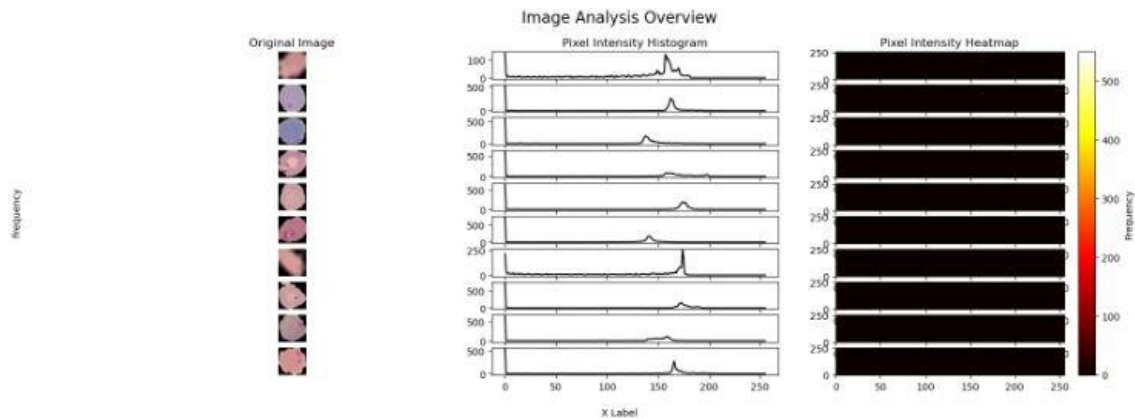
In the EDA phase, we explored the dataset to gain insights into its characteristics and distributions. Key analyses performed include:

**Data Distribution:** We visualized the distribution of infected and uninfected cell images to ensure that the dataset is balanced, which is crucial for model training.

**Image Visualization:** Sample images from the dataset were displayed to understand the visual differences between infected and uninfected cells.



**Pixel Distribution:** Histograms and heatmaps were generated to examine the distribution of pixel values in the images, identifying any patterns or variations.



#### 4. Data Preprocessing:

Before building machine learning models, we preprocessed the dataset to prepare it for training. Key preprocessing steps include:

**Image Resizing:** To standardize the input dimensions, all images are resized to a fixed size of 50x50 pixels.

**Normalization:** Pixel values are normalized to the range  $[0, 1]$  by dividing by 255, facilitating convergence during model training.

**Data Augmentation:** To enhance the dataset's diversity and prevent overfitting, we applied data augmentation techniques such as rotation, flipping, and zooming to generate additional training samples.

**Preprocessing and Highlighting:** We implement a preprocessing function that preprocesses images, classifies them using a pre-trained CNN model, and highlights potentially infected regions using adaptive thresholding and contour detection techniques.

#### 5. Model Architecture:

A Convolutional Neural Network (CNN) architecture is utilized for malaria cell classification. The model comprises multiple convolutional layers with varying filter sizes and strides, followed by max-pooling layers to downsample feature maps. Rectified Linear Unit (ReLU) activation functions introduce non-linearity, while the softmax activation function in the output layer enables binary classification (infected or uninfected).

## 6. Training and Evaluation:

**Dataset Split:** The dataset is split into training, evaluation, and test sets using a stratified split to preserve class distribution.

**Model Training:** The model is trained using the Adam optimizer with a sparse categorical cross-entropy loss function. Training is conducted for 10 epochs with a batch size of 100.

**Evaluation Metrics:** Various evaluation metrics, including accuracy, precision, recall, F1-score, and confusion matrix, are computed to assess model performance on the evaluation and test sets.

```
Accuracy: 0.9629898403483309
Confusion Matrix:
[[4017 158]
 [ 199 5272]]
Classification Report:
              precision    recall  f1-score   support

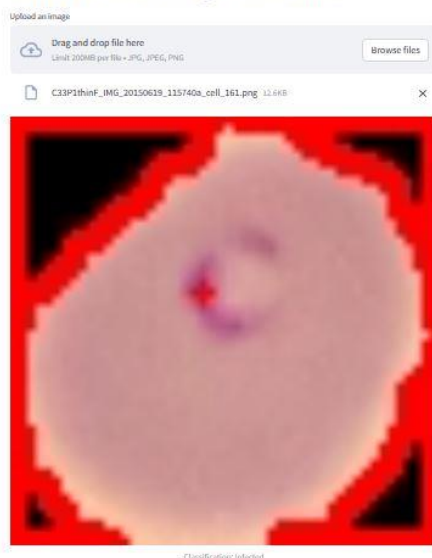
     0       0.95         0.96         0.96         4175
     1       0.97         0.96         0.97         5471

 accuracy         0.96
 macro avg        0.96
 weighted avg     0.96
```

## 7. Serialization and Deployment

Serialization involves saving the model architecture and weights to separate files. The model architecture is saved to a JSON file named "malaria\_detector\_architecture.json", while the weights are saved to an HDF5 file named "malaria\_detector\_weights.weights.h5". These files facilitate model reconstruction and deployment. Deployment is done by Streamlit library to create user-friendly web applications for uploading images and obtaining malaria diagnosis predictions.

### Malaria Cell Image Classifier



## **8. Results and Discussion:**

The trained model achieves a high accuracy of 96% on the evaluation set, demonstrating strong performance in distinguishing between infected and uninfected cells. However, there are instances where the model incorrectly predicts infected cells as uninfected, indicating potential areas for improvement. To address this, adaptive thresholding and contour detection techniques are employed to highlight potential infected regions, aiding visual interpretation and potential refinement of the classification process.

## **9. Conclusion and Future Directions**

In conclusion, our exploratory data analysis and modeling efforts demonstrate the feasibility of using machine learning techniques for malaria cell image classification. The developed CNN model, deployed using Streamlit, shows promising performance in accurately identifying malaria-infected cells, which could potentially aid healthcare professionals in early diagnosis and treatment of malaria. Further research could focus on optimizing model architecture, exploring additional preprocessing techniques, and incorporating interpretability methods to enhance model performance and interpretability.