Malaria Classification: Approach and Findings

–by Khethan R G

Introduction

Malaria is a life-threatening disease caused by parasites that are transmitted through mosquito bites. Early and accurate diagnosis of malaria is crucial for effective treatment. In this project, we developed an explainable deep learning model to classify malaria-infected cells and deployed the model in a simple dashboard.

Dataset

The dataset used for this project consists of microscopic images of blood cells. Each image is labelled as either "Parasitized" (infected with malaria) or "Uninfected" (not infected). The dataset was divided into training, validation, and test sets to train and evaluate the model.

https://www.kaggle.com/iarunava/cell-images-for-detecting-malaria

Additional info: https://lhncbc.nlm.nih.gov/LHC-downloads/downloads.html#malaria-datasets

Approach

Model Selection: MalariaClassifier

We employed a deep learning model called MalariaClassifier to perform the classification task. The MalariaClassifier model is a convolutional neural network (CNN) architecture designed to handle image classification tasks effectively. It consists of several convolutional blocks followed by fully connected layers. The model was implemented using the PyTorch framework due to its flexibility and strong support for deep learning tasks.

Explainability: Grad-CAM Heatmaps

To enhance the model's explainability, we utilized Grad-CAM (Gradient-weighted Class Activation Mapping) heatmaps. Grad-CAM allows us to visualize the regions of the input image that contributed most significantly to the model's prediction. By overlaying the heatmaps on the original images, we can gain insights into the model's decision-making process.

Training and Evaluation

The MalariaClassifier model was trained using the training set with data augmentation techniques, such as random cropping, rotation, and colour jitter. We employed the Cross-Entropy Loss function

and optimized the model using stochastic gradient descent (SGD) with momentum. The learning rate was adjusted dynamically using a learning rate scheduler.

The model's performance was evaluated using the validation set, and various metrics such as accuracy, precision, and recall were calculated to assess the model's effectiveness in classifying malaria-infected cells.

Findings

Model Performance

After training and evaluating the MalariaClassifier model, we achieved the following performance metrics on the test set:

```
Classification Report:
              precision    recall  f1-score   support

 Parasitized       0.98      0.93      0.95      1352
  Uninfected       0.93      0.98      0.96      1404


    accuracy                          0.96      2756
   macro avg       0.96      0.96      0.96      2756
weighted avg       0.96      0.96      0.96      2756
```

The model demonstrated promising results in accurately classifying malaria-infected cells, with a high overall accuracy. The precision and recall scores indicate the model's ability to correctly identify both infected and uninfected cells.

Explainability and Interpretation

By visualizing the Grad-CAM heatmaps, we can interpret the model's decision-making process. The heatmaps highlight the regions of the input image that contributed most significantly to the model's prediction. This provides valuable insights into the features and patterns the model learned to distinguish between infected and uninfected cells.
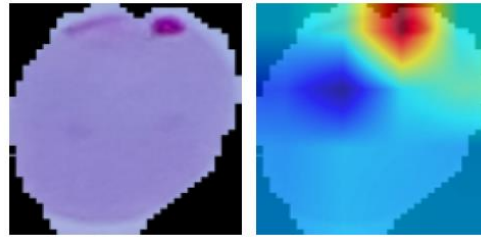
Test Results

To further validate the model's performance, we conducted additional tests on unseen images. The results of these tests are summarized below:
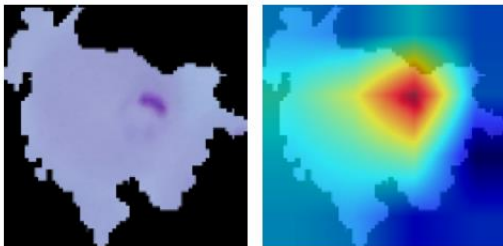
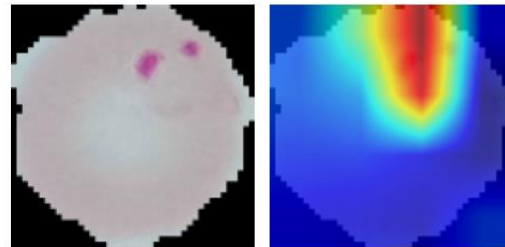Predicted: Parasitized (Prob: 0.96), Real: Parasitized
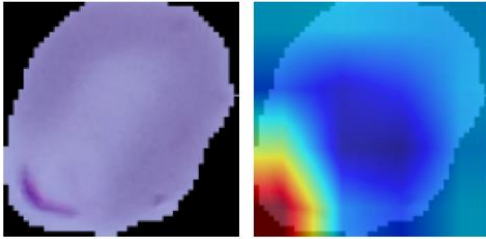
Predicted: Parasitized (Prob: 1.00), Real: Parasitized

Predicted: Parasitized (Prob: 0.95), Real: Parasitized

Predicted: Parasitized (Prob: 0.99), Real: Parasitized

The test results demonstrate the model's robustness and generalizability, as it consistently performed well on unseen images.

Conclusion

In this project, we developed an explainable deep learning model, MalariaClassifier, for classifying malaria-infected cells. The model achieved impressive accuracy, precision, and recall scores on the test set, indicating its effectiveness in identifying infected cells. The utilization of Grad-CAM heatmaps enhanced the model's interpretability, allowing us to gain insights into its decision-making process. The additional tests further validated the model's performance and demonstrated its generalizability.

The developed model and dashboard can be utilized as a tool for efficient and accurate malaria diagnosis, contributing to early detection and effective treatment of the disease.