Diagnosing Malaria: A Comparison of Transfer Learning Approaches on the NIH Malaria Dataset with Grad-CAM Visualization for Interpretable AI

Ann-Marcia Tukpah, Christoph Wippel, Elombe Calvert, Jason Xiang, Michelle Theodory

Abstract

Malaria is a lethal disease that disproportionately affects sub-Saharan Africa. Current diagnostic methods, such as microscopy, are labor-intensive and susceptible to human errors. Machine learning offers promising potential for enhancing automated medical imaging techniques in malaria diagnosis. In this research, we evaluated the efficacy of transfer learning in Convolutional Neural Networks (CNNs) compared to a baseline Random Forest model for classifying malaria-infected cells using the NIH Malaria dataset. Additionally, we investigated the influence of ImageNet weights on the VGG19 CNN model's performance. We employed Grad-CAM visualization to generate saliency representations of malaria-infected cells and assessed their interpretability. Our findings revealed that the VGG19 CNN model, when augmented with ImageNet weights, outperformed other models in accuracy. Moreover, the Grad-CAM visualization offered valuable insights into the CNN model's decision-making process by emphasizing crucial areas for classification. Our research not only establishes that deep learning can effectively enhance automated malaria diagnosis but also contributes to our understanding of disease diagnosis through its interpretability.

Introduction

Malaria is a fatal disease that is caused by Plasmodium parasite and is transmitted to humans through the bites of infected mosquitoes. It is a major health crisis as there are an estimated 241 million cases of malaria worldwide in 2020 according to the Center for Disease Control and Prevention (CDC, 2023).

Although treatable, drug resistance and poor diagnostics make it challenging to reduce mortality rates effectively. More specifically, the most common diagnostic tool is light microscopy of blood films, which entail manual counting of parasites. The manual nature of this method is not only burdensome, but is prone to error and incorrect diagnoses. There is, then, a need for a standardized and automated method which is able to correctly identify and count parasites, and thus combat the adverse effects of Malaria.

With the advancement of computer vision, machine learning methods such as Convolutional Neural Network (CNN) show potential for automated malaria diagnosis. Leveraging artificial

intelligence can speed up malaria diagnosis and particularly help regions that lack the access to healthcare professionals. Furthermore, interpretability the mechanism behind these methods can provide insights to disease pathology and facilitate better patient outcomes.

Our primary research question is to investigate which Convolutional Neural Network (CNN) provides the most accurate classification of cells infected with the malaria parasite (Plasmodium) and can the architecture be improved to increase selected metrics. We would use accuracy and area under the curve (AUC) scores as our main metrics to evaluate the performance of the model. Secondarily, we were interested in the features the model prioritizes for classification. We would examine the explainability of the above models using the Grad-CAM saliency method. We chose VGG19 as our transfer learning architecture and evaluated its performance when trained with and without ImageNet weights.

Methods

The Malaria Cell Images Dataset retrieved from kaggle.com on March 31, 2023 consists of 27,558 images of parasitized and non-parasitized red blood cells microscopy coloured images. Images were converted to a numpy array and normalized.

We loaded the data using the Kaggle download API. The images had different image-sizes, therefore we resized them to all have a uniform size of 64x64 pixels and normalized the pixel values by dividing all pixels by the max value of 255. The resulting X data array for our CNN models had a shape of (27,558, 64, 64, 3). Further, we created an array of the labels (uninfected, parasitized) and stored them in a 1-dimensional numpy array of length 27,558.

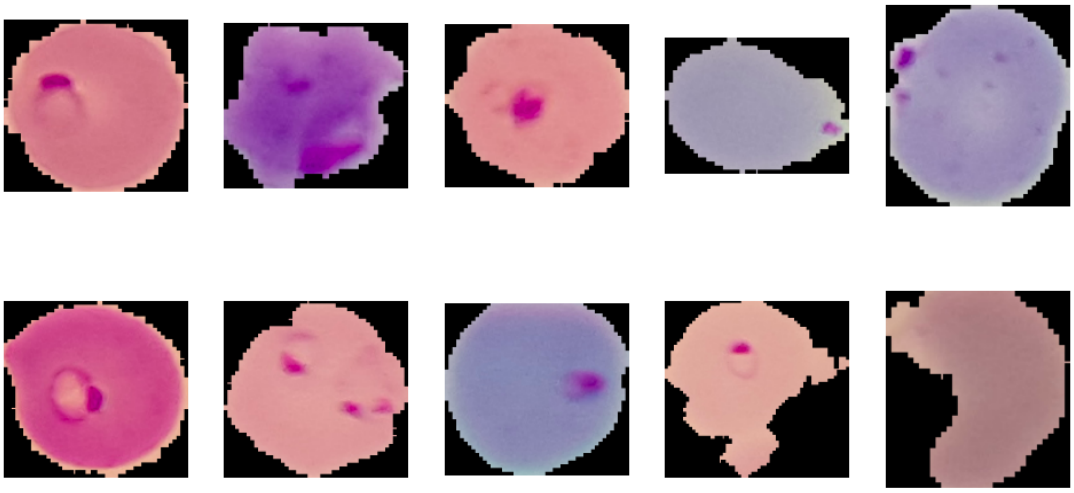
The deep learning approaches we attempted are Random Forest classification as our baseline model and Transfer Learning model using VGG19 with ImageNet weights and without.

In creating our Random Forest classification as a baseline model as a reference base model for our ML project, we split the dataset into 80% training and 20% testing. To format our data accordingly for the Random Forest model, we created flattened image sequences and stored them

as numpy arrays with 4,096 elements each. After that, we then trained a Random Forest classifier on the pre-processed data using 100 trees on the training data.

Figure 1: Example of Images in Dataset

Parasitized cells - before image processing

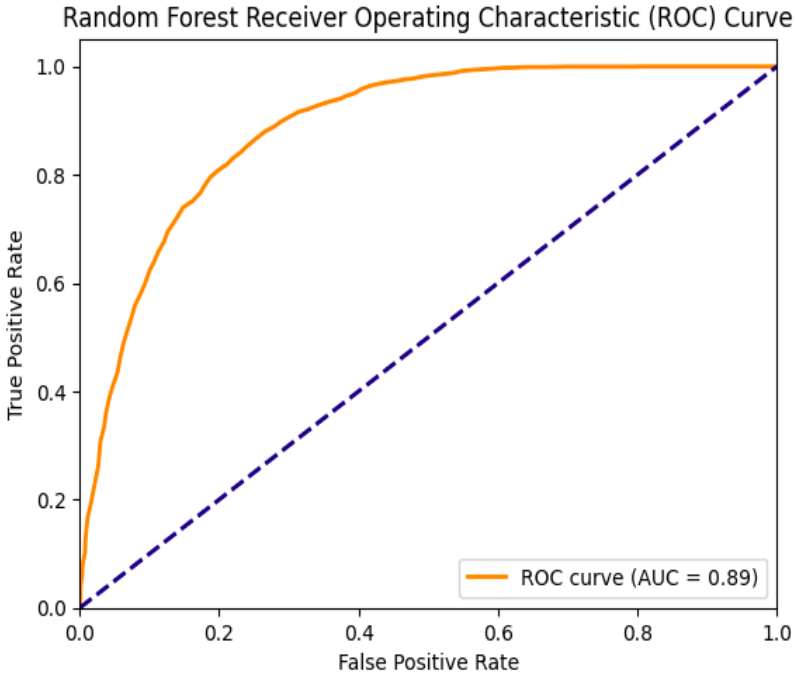


In our transfer learning, the first step taken was to import the VGG19 model with pre-trained weights on the ImageNet dataset, setting the input shape to (64,64,3). The layers of the VGG19 model are then made untrainable. A few fully connected layers are added on top of the pre-trained VGG19 layers, including a dense layer with 64 units, a dropout layer with a rate of 0.5, a batch normalization layer, a ReLU activation layer, and a flatten layer. Finally, a dense output layer with sigmoid activation is added to classify the input images as either infected or uninfected with malaria. The model is then compiled with the Adam optimizer, binary cross-entropy loss, and accuracy and AUC metrics. The model is trained for 5 epochs using a train data generator and a validation data generator. Using accuracy, as well as AUC scores, as our main evaluation metrics, we want to compare the performance with ImageNet weights and the one without ImageNet weights in VGG19.

Results

For our baseline model (Random Forest), we achieved 0.804 for accuracy and 0.887 for AUC score in the validation set (Figure 2).

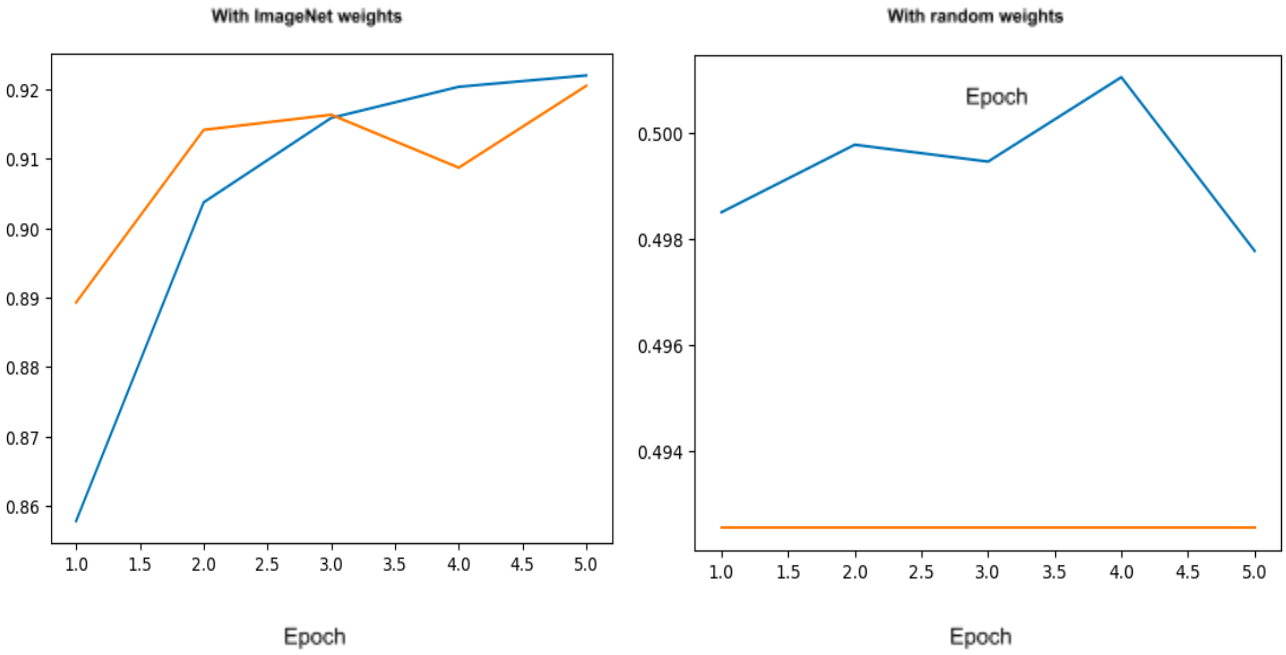
Figure 2: ROC Curve for Random Forest Model




```
Accuracy : 0.8040638606676342
AUC Score: 0.8875367057703366
```

For the VGG19 model with ImageNet weights, we achieved 0.918 for validation accuracy (orange line) and 0.921 for training accuracy (blue line) (Figure 3). We also achieved AUC scores of 0.98 and 0.97, respectively. For the VGG19 model without ImageNet weights, we achieved 0.493 for validation accuracy (orange line) and 0.498 for training accuracy (blue line) (Figure 3), and AUC scores of 0.50 and 0.56, respectively.

Figure 3: Accuracy - with Imagenet weights (left) vs dataset without ImageNet weights (right). Blue- training accuracy, orange - validation accuracy

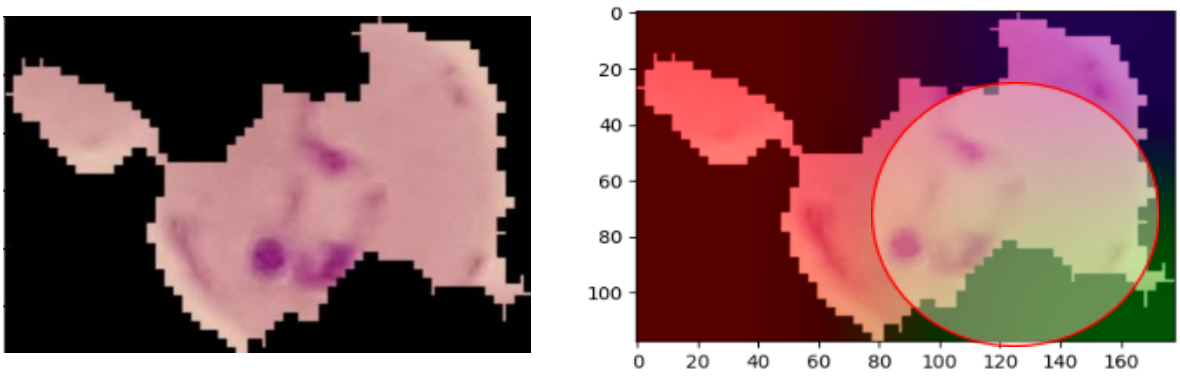


Overall, we found that the Random Forest model had 0.80 accuracy, which was better than VGG19 with random weights (without ImageNet weights). The best performing model was VGG19 using ImageNet weights, which we expected. It reaches an astounding validation accuracy of approximately 0.92, a significant improvement from 0.80 accuracy in our baseline model, as well as a final validation AUC of approximately 0.98.

In order to interpret our deep learning approach, we employed Grad-CAM to create a saliency representation, offering insight into the features the model relies on to predict cell infections. As demonstrated in the Grad-CAM figure below, it appears the model is effectively focusing on the correct areas for distinguishing infected cells from uninfected cells, as it highlights the irregular cell edges that can be indicative of infection.

Discussion

Figure 4: Grad-CAM Saliency Representation of Parasitized Image



Overall, with high validation accuracy and AUC, as well as a reasonable saliency representation, the VGG19 model trained with ImageNet weights demonstrates promising potential for classifying malaria parasite infections in cells**.** Our high AUC scores, as well as saliency maps, demonstrate that our model has appropriately learnt what features to use for prediction, and far outperforms what would be predicted by chance. Our validated VGG19 model could potentially be used in resource limited health systems that lack trained medical professionals and the equipment needed to make the diagnosis of malaria through light microscopy with manual phenotyping. This translation of AI technology into the clinic would significantly improve malaria diagnostics and ultimately improve the access to medical care needed by some of society's most vulnerable.

Although promising, our study faced a number of limitations. One challenge was the computational expense of training our transfer learning model. Despite only training the VGG19 model with five epochs, we encountered limitations in RAM and GPU on Google Colab Notebook. With additional computational resources, we could explore other hyperparameter values to optimize our model's performance, particularly for our VGG19 model that did not include ImageNet weights. Another limitation was the generalizability of our training, as the homogeneous and easily identifiable training samples may not be as applicable to more diverse real-life datasets. Future studies could benefit from training models with varied datasets from different sources to improve generalizability and robustness.

We encountered various challenges during this study. For example, when training the VGG19 model with random weights, we struggled to improve validation accuracy. Despite trying different layer freezing strategies, the model's performance stagnated even as epochs increased. To help reduce training workload, we resized images to an appropriate size without compromising performance.

For future research, we aim to enhance the generalization ability of our transfer learning model. More specifically, we will explore more diverse datasets from other sources that better reflect real-life settings through varying image quality conditions, and conduct further external validation. Furthermore, we plan to experiment with other model architectures, such as ResNet, to potentially increase the accuracy of our deep learning model.

Group Member Contribution

Ann-Marcia Tukpah: Poster writeup
Christoph Wippel: Grad-CAM analysis
Elombe Calvert: Random Forest model
Jason Xiang: Report Write-up
Michelle Theodory: Transfer Learning model

Even though each member had designated roles, everybody contributed in some way to each aspect of the project.

References

"CDC - Parasites - Malaria." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 25 Apr. 2023, https://www.cdc.gov/parasites/malaria/index.html.

"Malaria Screener." *U.S. National Library of Medicine*, National Institutes of Health, https://lhncbc.nlm.nih.gov/LHC-research/LHC-projects/image-processing/malaria-screener.html