Report: Malaria Detection Using Machine Learning Techniques

**Introduction**

Malaria is a disease that virus is caused when parasites are transmitted to people by the bites of infected female Anopheles mosquitoes. Accurate and early detection of malaria is critical to the treatment and prevention (World Health Organization, 2024). Malaria diagnosis using traditional methods such as microscopy is slow and must depend on skilled personnel. The modern machine learning and deep learning approaches are making malaria detection systems automated, enabling faster and more accurate diagnosis.

In this report, we introduce a machine learning method to detect malaria from cell images. We explore two models: A baseline Convolutional Neural Network (CNN) as well as a fine tuned VGG16 model. We train both models on the Malaria Cell Images Dataset (with image both infected and uninfected cells) We aim to compare the performance of these models, but also assess whether they can effectively detect malaria. In this report, we present a machine learning-based approach to detect malaria using cell images. We explore two models: a baseline Convolutional Neural Network (CNN) and a fine-tuned VGG16 model. Both models are trained on the Malaria Cell Images Dataset, which contains images of both infected and uninfected cells. The goal is to compare the performance of these models and evaluate their effectiveness in detecting malaria.

 Dataset Overview

The Malaria Cell Images Dataset consists of 27,558 images, divided into two classes: Uninfected and Parasitized. The images are balanced in the dataset which contains 13,779 images per class. The images are stored in two subdirectories: `Parasitized` and `Uninfected`. The largest image has the size 64x64 pixels, and is a colored image (Kaggle, 2018).

First, the data is preprocessed to remove non image files (e.g. `Thumbs.db`) and that only valid images are used for training and evaluation. The images are resized to 64x64 pixels and normalized with pixel values between [0, 1]. The dataset is then split into training, validation, and test sets with the following proportions:

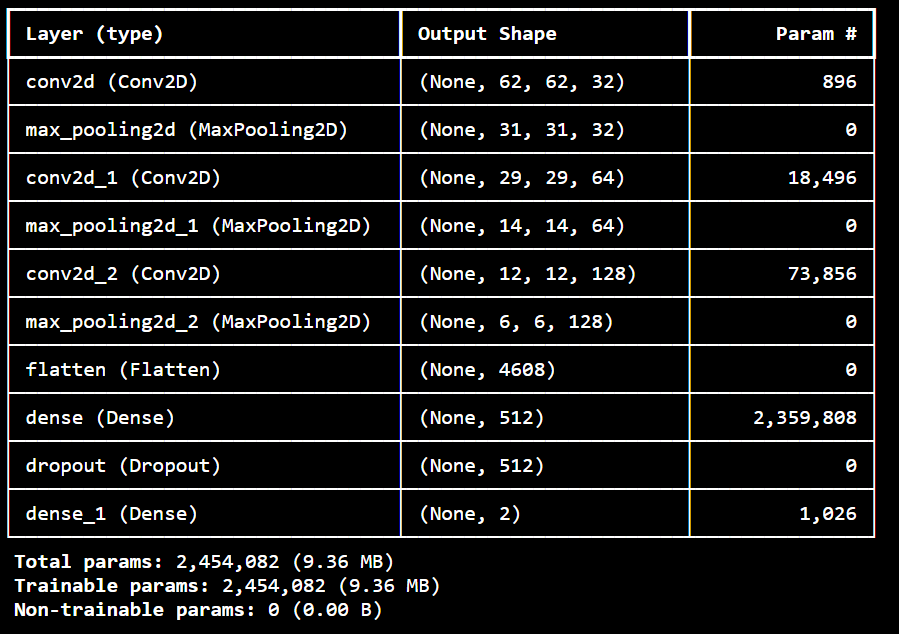
* Training set: 17,636 images (80% of the total dataset)
* Validation set: 4,410 images (10% of the total dataset)
* Test set: 5,512 images (10% of the total dataset)



 The dataset is balanced across the two classes, ensuring that the models are trained on a representative sample of the data.

**Baseline CNN Model**

 The baseline CNN model is a simple convolutional neural network designed to classify the images as either uninfected or parasitized. The model consists of three convolutional layers, each followed by a max-pooling layer, and two fully connected layers. The architecture of the model is as shown below.



 The model is compiled using the Adam optimizer, categorical cross-entropy loss, and accuracy as the evaluation metric. The model is trained for 10 epochs with a batch size of 32.

|  |  |
| --- | --- |
| Training Result | Evaluation Metrics |
| Training Accuracy: 0.9801 | Precision: 0.9486 |
| Validation Accuracy: 0.9560 | Recall: 0.9700 |
| Test Accuracy: 0.9581 | F1-Score: 0.9592 |

Confusion Matrix:

[[2568  147]

 [  84 2713]]

The baseline CNN model is shown to be quite effective, achieving a test accuracy of 95.81%. It turns out that the model correctly classifies the large majority of images, with a high recall on the parasitized class (97.00%) meaning the model is able to detect infected cells.

**Fine-Tuned VGG16 Model**

 A fine-tuned VGG16 model is a pre trained convolutional network on years of learning which is fine tuned for the malaria detection task. VGG16 is a deep neural network with 16 layers, originally pre-trained with the ImageNet dataset. It learns general features that the model is pre-trained on a big dataset of images, that can be then tuned for a particular task ( Whitfield, 2024).

To achieve a malaria detection task, we modify the VGG16 model by removing its top layers (fully connected layers) and defining custom layers to do the malaria detection. The architecture of the fine-tuned VGG16 model is as follows:

 1. VGG16 Base Model: Input shape is 64x64x3 and it’s pre-trained in VGG16 model.

2. Global Average Pooling Layer: Reduces spatial dimensions of the output

3. Dense Layer: 512 neurons, ReLU activation

4. Output Layer: softmax activation (for binary classification), 2 neurons.

Custom and base model layers are trained but, in order to avoid overfitting, only the custom layers are trained and frozen base model layers are not trained (Ying, 2019). The model is then compiled using the Adam optimizer with categorical cross entropy loss and there accuracy metric. With a batch size of 32, model is trained for 10 epochs.

|  |  |
| --- | --- |
| Training Result | Evaluation Metrics |
| Training Accuracy: 0.9379 | Precision: 0.9412 |
| Validation Accuracy: 0.9283 | Recall: 0.9099 |
| Test Accuracy: 0.9254 | F1-Score: 0.9253 |

Confusion Matrix:

[[2556  159]

 [ 252 2545]]

Our fine tuned VGG16 model achieves a test accuracy of 92.54%. However, the model performs slightly worse than the baseline CNN model in both precision and recall for both classes. Most of the infected cells (90.99% recall) can be detected by the model, however, some images that are not infected are misclassified as not infected.

**Comparison of Models**

 Further comparison between the fine tuned VGG16 model and the baseline CNN model are carried out in terms of performance on the test set. The following table summarizes the evaluation metrics for both models:

|  |  |  |
| --- | --- | --- |
| Metric | Baseline CNN | Fine-Tuned VGG16 |
| Precision | 0.9486 | 0.9412 |
| Recall | 0.97 | 0.9099 |
| F1-Score | 0.9592 | 0.9253 |

In terms of accuracy, precision, recall, and F1-score, the baseline CNN model is outperformed by the VGG16 fine-tuned model. The CNN model at the baseline however, does better at recall for the parasitized class, which indicates that it identifies infected cells better. However, this fine tuned VGG16 model has higher precision, showing that it does a better job at not creating false positives.

**Discussion**

 This study shows the performance of the baseline CNN model and fine tuned VGG16 models in the detection of malaria from cell images. Despite its simplicity, the baseline CNN model performs better than the more complex fine-tuned VGG16 model when it comes to accuracy, recall and F1-score. This means that a s simple model with less parameters may outperform the more complicated pre trained model for this particular task.

**Baseline CNN Model**

Although the baseline CNN model achieved a high accuracy (95.81%) and recall (97.00%) for the infected class we conclude that this highly accurate model is capable of identifying infected cells. In a medical context it is particularly important to not miss an infected cell which could mean that treatment is delayed and it could even cause complications (CDC, 2016). The truth is that the model also achieved a high precision (94.86%), meaning it seemed good at not making false positives – something important to avoid for unnecessary treatment of patients, and to reduce anxiety.

Being a simple model, the baseline CNN model offers itself for deployment in resource constrained environments. Being relatively simple with few parameters, it is easier to train and then deploy in comparison to models such as VGG16.

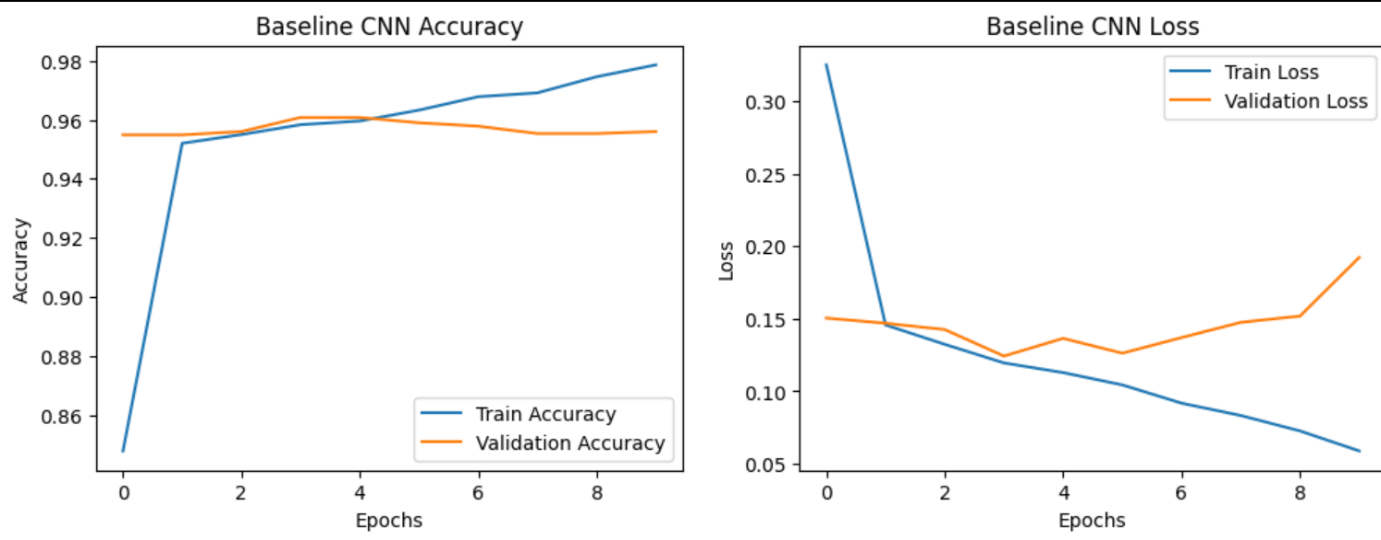
**Fine-Tuned VGG16 Model**

Although the fine tuned VGG16 model had slightly lower accuracy (92.54%) than the baseline CNN model, it still was pretty good at precision (94.12%) and recall (90.99%). The result about the model’s ability to stay away from false positives is really noteworthy because it shows that the model does not seem on very many false positives. In a medical context, false positives would cause patients to receive unnecessary treatment, or experience anxiety, and that is important.

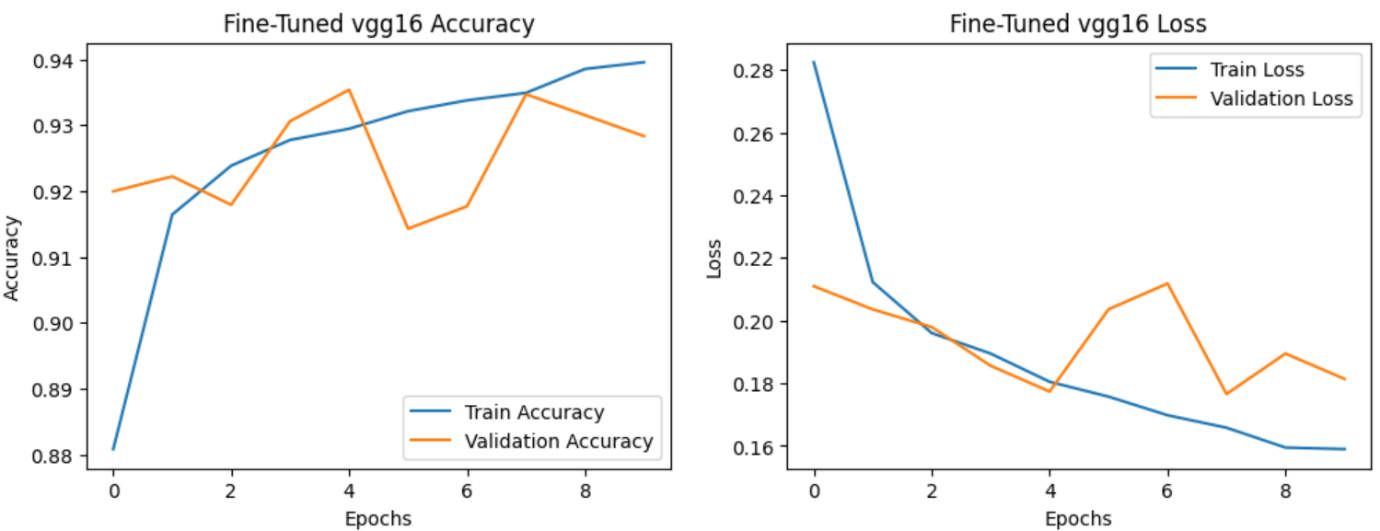
Nevertheless, the fine-tuned VGG16 model's lower recall for the parasitized class (90.99 %) is lower then the baseline CNN model recalls (97.00 %), possibly failing to capture some infected cells. Potential reason could be the complexity of the model or just that the pre trained VGG16 model wasn’t initially designed to be used to perform such task (ImageNet Image classification). For the malaria detection task, it is possible that fine tuning the model yields suboptimal performance, not fully exploiting the model’s full potential.

**Training and Validation Curves**

 Additional insights into the learning behavior of both models can be obtained from their training and validation curves. During training, the baseline CNN model has very strong performance and shows an approximately consistent performance in terms of accuracy and loss, with minimal overfitting issues. It means that the model had done well in generalizing to unseen patch.



 In contrast, the fine tunning VGG16 model, as was the case here does have a bit higher validation loss compared to the baseline CNN model suggesting that it might be overfitting the training data. This indicates that higher complexity of the model may possibly be the reason behind its lower test set performance.



  Future Work

Future work could also include examining other models that have already been trained, such as ResNet50 or InceptionV3, to see how they do compared to the baseline CNN model. Furthermore, models could be improved with a data augmentation techniques. Experimenting with data augmentation could enable the models to learn more robust features, reducing overfitting, and thus could yield better results on the test set.

Future work is also possible in the area of exploring ensemble methods, where multiple models are combined to perform better. The predictions could be reduced in variance using ensemble methods that would also help the model generalize to unseen data.

**Conclusion**

In this report, we explored two machine learning models for the detection of malaria using cell images: This is done on both a baseline CNN model and a fine tuned VGG16 model. Both the baseline CNN model, which resulted in a test accuracy of 95.81%, and the fine-tuned VGG16 model, giving a test accuracy of 92.54%, were used as performance benchmarks. In terms of accuracy, precision, recall and F1-score, the baseline CNN model outperformed the fine tuned VGG16 model.

A simple and effective baseline model for the malaria detection problem is the baseline CNN model, which achieves high accuracy and recall for classes. Although the fine tuned VGG16 model performs slightly worse, it produces high precision and recall, and is thus a viable alternative to the task.

**References**

 CDC. (2016, May 3). *CDC - DPDX - diagnostic procedures - blood specimens*. Centers for Disease Control and Prevention. https://www.cdc.gov/dpdx/diagnosticprocedures/blood/microexam.html

Kaggle. (2018, December 5). *Malaria Cell Images Dataset*. https://www.kaggle.com/datasets/iarunava/cell-images-for-detecting-malaria

Whitfield, B. (2024). *Beginners Guide to VGG16 implementation in Keras*. Built In. https://builtin.com/machine-learning/vgg16

World Health Organization. (2024). *Fact sheet about malaria*. https://www.who.int/news-room/fact-sheets/detail/malaria

Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, *1168*, 022022. https://doi.org/10.1088/1742-6596/1168/2/022022