Capstone Project: Deep Learning

Identification of Plasmodium spp. from Thin Blood Smears using Deep Learning.

LaKeta L Kemp
Applied Data Science Program
Massachusetts Institute of Technology, Great Learning
December 13, 2022

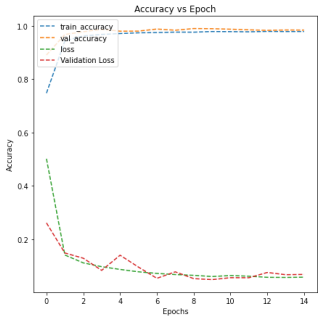
Problem Definition

- Malaria is a parasite transmitted by mosquitos.
- There are 241 million infections and 627,000 deaths annually.
- Thin blood smears read by trained and certified Parasitologist with specialized training.
- Subject matter expertise is a small group of individuals.
- A computer vision model can be trained to identify Malaria infections.

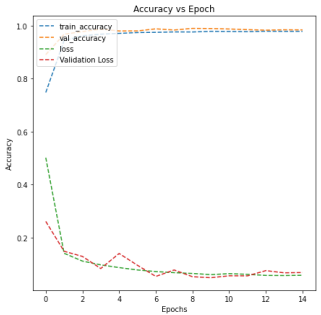
Solutions Design

- Differential microscopy allows identification of key features.
- Two dimensional convolutional layers with
  - Dense layers, pooling, normalization, regularization, reshaping
- Model Evaluation Metric minimizes validation loss.
- Model Performance Metric is the ratio of recall and precision to maximize overall performance.
- Reduce false reporting to the minimum.

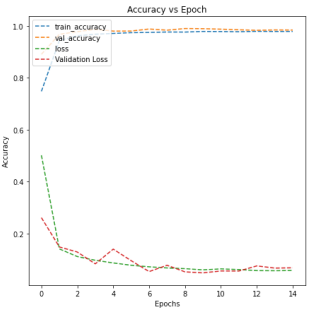
Model Performance



Base Model



Model 2



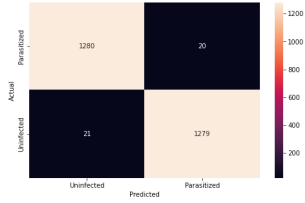
VGG16 Model

Model Evaluation

```
82/82 [==============================] - 3s 40ms/step
              precision    recall  f1-score   support

           0       0.98      0.98      0.98      1300
           1       0.98      0.98      0.98      1300

    accuracy                           0.98      2600
   macro avg       0.98      0.98      0.98      2600
weighted avg       0.98      0.98      0.98      2600
```

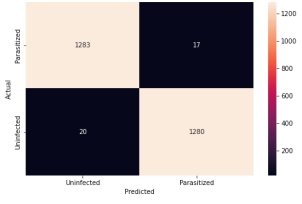


Base Model

```
82/82 [==============================] - 5s 63ms/step
              precision    recall  f1-score   support

           0       0.98      0.99      0.99      1300
           1       0.99      0.98      0.99      1300

    accuracy                           0.99      2600
   macro avg       0.99      0.99      0.99      2600
weighted avg       0.99      0.99      0.99      2600
```

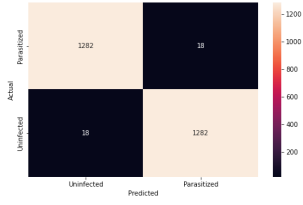


Model 2

```
82/82 [==============================] - 3s 41ms/step
              precision    recall  f1-score   support

           0       0.99      0.99      0.99      1300
           1       0.99      0.99      0.99      1300

    accuracy                           0.99      2600
   macro avg       0.99      0.99      0.99      2600
weighted avg       0.99      0.99      0.99      2600
```



VGG16 Model

Key Observations & Insights

- VGG16 performs better than Model 2.
- False reporting has serious health outcomes, minimizing false reporting is a key performance indicator for the model.
  - False Negatives: serious negative health outcomes including death without treatment.
  - False Positives: negative health outcomes due to toxicity of treatment.
- The VGG16 matches the current accuracy and precision for light microscopy.

Recommendations

- The VGG16 model is recommended to identify Malaria in blood smear images.
  - Positive Predictive Value (Precision): 99%
  - Correct Identification (Accuracy): 99%
  - Predictive Performance (Recall): 99%
  - Ratio of False Identifications (F1 Score): 99%
  - False Negatives: <1%
  - False Positives: <1%

Works Cited

1. *Giemsa Stain.* Clinical Microbiology Procedures Handbook, 3 Volume Set, 4th Edition Amy L. Leber. doi:10.1128/9781555818814.ch.9.8.1
2. *Detection of Blood Parasites.* Clinical Microbiology Procedures Handbook, 3 Volume Set, 4th Edition Amy L. Leber. doi:10.1128/9781555818814.ch.9.8.1
3. Pillow Python Imaging Library. PIL Fork Documentation.
4. OpenCV. OpenCV: Smoothing Images API Documentation.
5. WHO Guidelines for malaria, 25 November 2022. Geneva: World Health Organization; 2022 (WHO/UCN/GMP/2022.01 Rev.3). License: CC BY-NC-SA 3.0 IGO.
6. Center for Disease Control and Prevention Division of Parasitic Diseases and Malaria (DPDM). DPDx - Laboratory Identification of Parasites of Public Health Concern
7. A. A. Alonso-Ramírez et al., *Classifying Parasitized and Uninfected Malaria Red Blood Cells Using Convolutional-Recurrent Neural Networks.* IEEE Access, vol. 10, pp. 97348-97359, 2022, doi: 10.1109/ACCESS.2022.3206266.
8. Jdey, I., Hcini, G., & Ltifi, H. (2022). Deep learning and machine learning for Malaria detection: overview, challenges, and future directions. arXiv preprint arXiv:2209.13292.
9. *How to Configure Image Data Augmentation in Keras.* Deep Learning for Computer Vision. Brownlee J. https://machinelearningmastery.com/how-to-configure-image-data-augmentation-when-training-deep-learning-neural-networks/
10. *Understanding Data Augmentation | What is Data Augmentation & how it works?* Great Learning Blog AI and Machine Learning. Arun K. https://www.mygreatlearning.com/blog/understanding-data-augmentation/
11. *Everything you need to know about VGG16.* Medium.com Blog. Rohini G. September 23, 2021. Everything you need to know about VGG16.
12. *Step by step VGG16 implementation in Keras for beginners.* Towards Data Science Blog. Thakur, Rohit. August 6, 2019. Step by step VGG16 implementation in Keras for beginners.
13. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. Simonyan, K. and Zisserman, A. https://arxiv.org/abs/1409.1556. September 4, 2014