Malaria Diagnostics Using Deep Learning

MIT ASDP Capstone
June 12, 2022
Michael Kasper

Problem Definition - Malaria

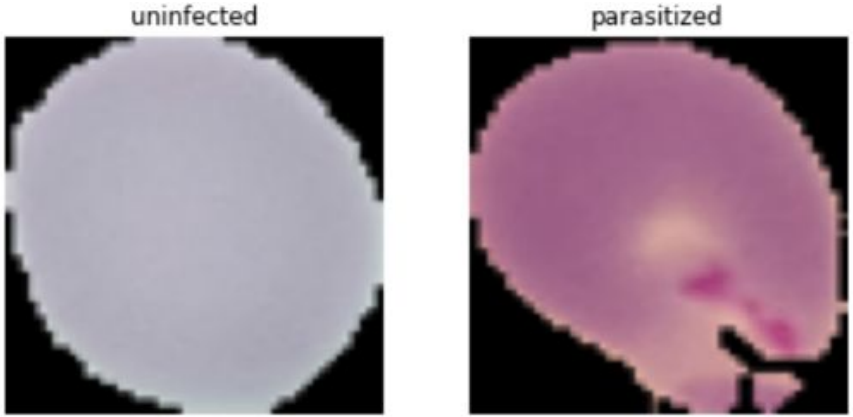
- Over 229 million cases in 2019

- 400,000 malaria-related deaths worldwide in 2019

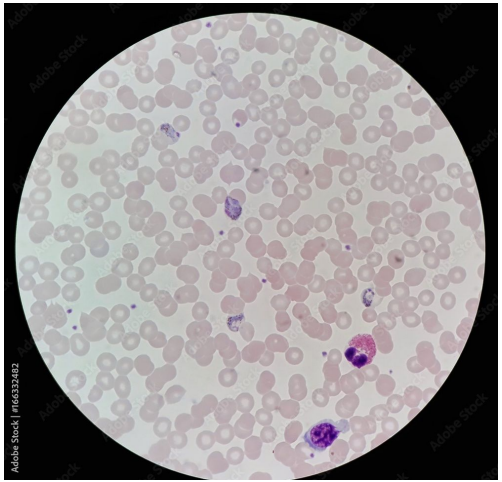
- 67% of these deaths were children under the age of 5.

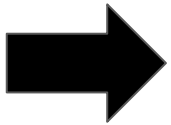
- Gold standard of malaria requires an experienced laboratory technician.

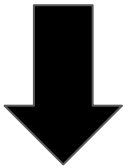
- Early diagnosis of malaria reduces disease transmission and prevents deaths.



Problem to solve



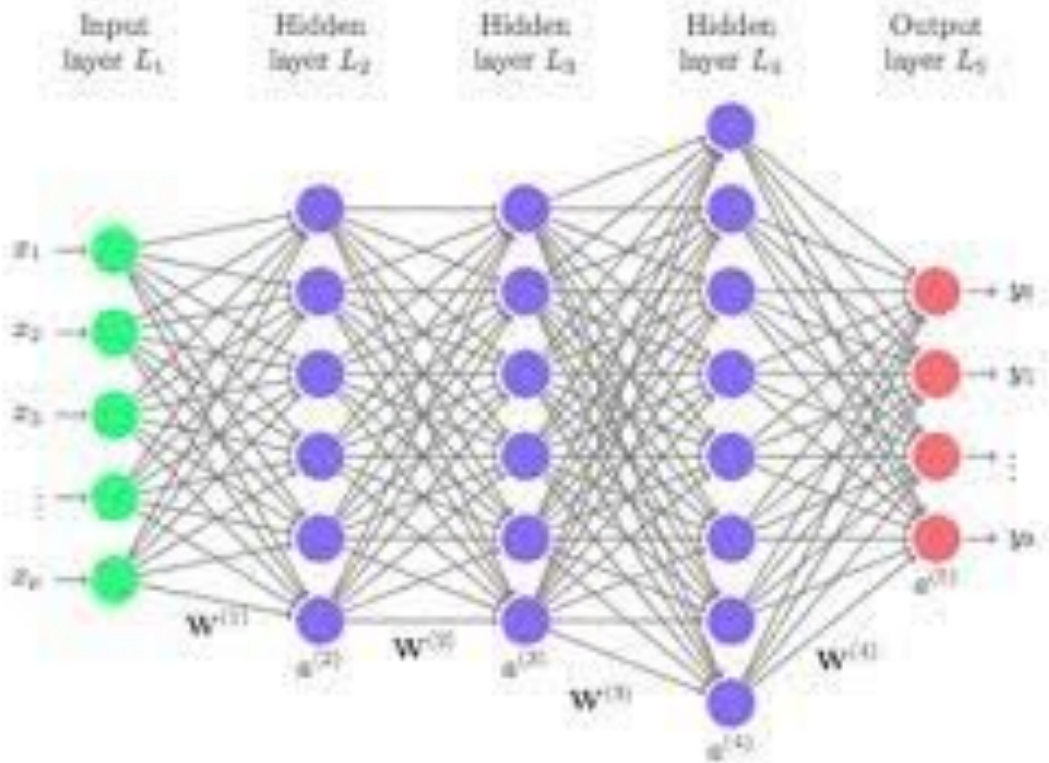




- Can we use a deep learning model to increase diagnostic accuracy?

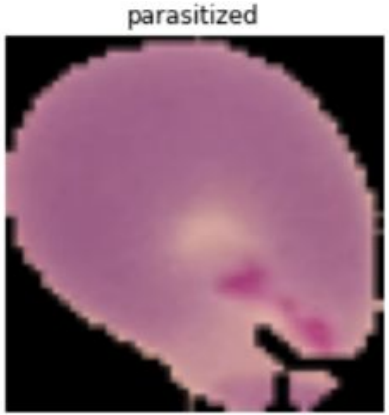
Accurate malaria diagnosis?

Convolutional Neural Networks



Solution Summary - Data Description and EDA

- Images are of single red blood cells taken from microscopy of blood smears. There is a training set of 24,985 images and testing set of 2,600 images.
- Data is balanced between uninfected and parasitized.
- Gaussian blurring to reduce noise was evaluated but not used in final model.
- Images were converted into arrays and normalized before running models.



Solution Summary

Model 1: 3 convolutional layers with 32 filters. A simple model.

Model 2: Slightly more complicated model with 2 more convolutional layers.

Model 3: Data normalization between layers.

Model 4: Model using augmented data.

Model 5: Transfer learning model from a pretrained VGG16 CNN.

Insights from Accuracy Graph - First Model

a.

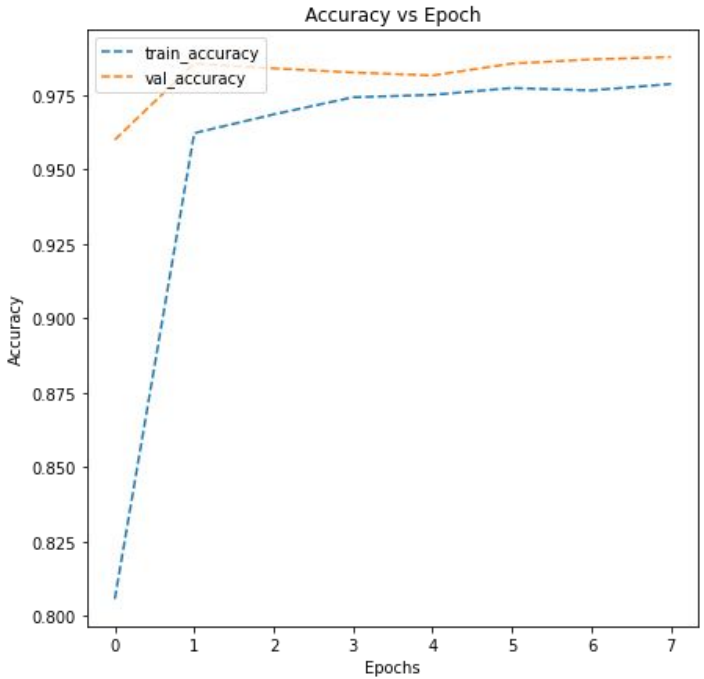


Figure 3: a. Model 1 accuracy vs. epoch graph shows that the model is fitting well. Training and validation accuracy are both near 0.98 after 7 epochs.

Model Accuracy = 98.42%

Bibliography and Acknowledgements

- Thank you to Great Learning for fantastic resources.
- Special thanks to Grant Haskins and Misbah Khan for mentorship.
- https://www.mmv.org/malaria-medicines/malaria-facts-figures?gclid=CjwKCAjw14uVBhBEEiwAaufYxxgNl2f1L54dkc3E2PgE1noJVsxD5Ukqqr0VA2OeT9X8jkCJTpgPXBoCSnMQAvD_BwE
- https://www.who.int/en/news-room/fact-sheets/detail/malaria

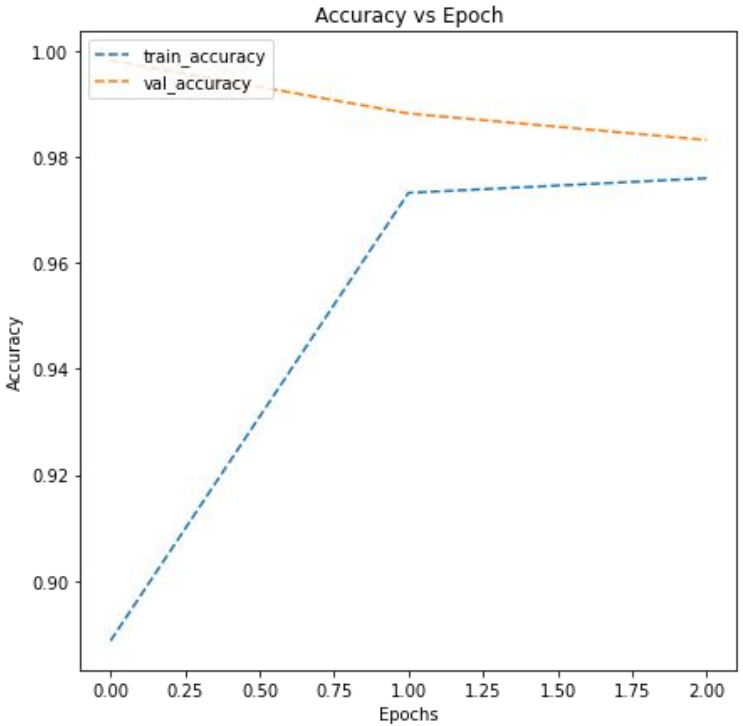
Images from following sources:

- https://www.nih.gov/news-events/nih-research-matters/how-mosquitoes-detect-people
- https://bdtechtalks.com/2019/02/15/what-is-deep-learning-neural-networks/
- https://www.flinnsci.com/flinn-basic-microscope/ms1130/
- https://stock.adobe.com/images/under-100x-light-microscope-human-parasite-on-thin-film-of-blood-smear-with-plasmodium-vivax-form-malaria-infected-red-blood-cells-selective-focus/166332482
- https://towardsdatascience.com/its-deep-learning-times-a-new-frontier-of-data-a1e9ef9fe9a8

Appendix - Best Solution Model Architecture

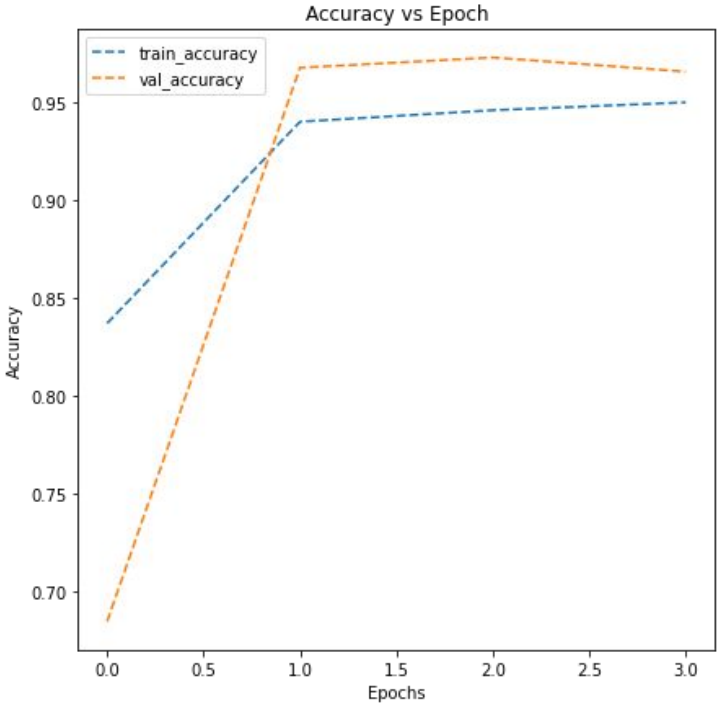
```
model=Sequential()
model.add(Conv2D(filters=32,kernel_size=2,padding="same",activation="relu",input_shape=(64,64,3)))
model.add(MaxPooling2D(pool_size=2))
model.add(Dropout(0.2))
model.add(Conv2D(filters=32,kernel_size=2,padding="same",activation="relu"))
model.add(MaxPooling2D(pool_size=2))
model.add(Dropout(0.2))
model.add(Conv2D(filters=32,kernel_size=2,padding="same",activation="relu"))
model.add(MaxPooling2D(pool_size=2))
model.add(Dropout(0.2))
model.add(Flatten())
model.add(Dense(512,activation="relu"))
model.add(Dropout(0.4))
model.add(Dense(2,activation="softmax"))#2 represent output layer neurons
model.summary()
```

Appendix - Accuracy Graph model 3



This is the accuracy vs epoch from the model using batch normalization. We can see that this model is overfitting. There are only 2 epochs in this model fit because of callbacks code line so when validation accuracy continues to decrease, the model stops fitting.

Appendix - Accuracy Graph Model 4



This is the model from data augmentation. This might be a bit underfit. The accuracies are not as high as in the original mode.