

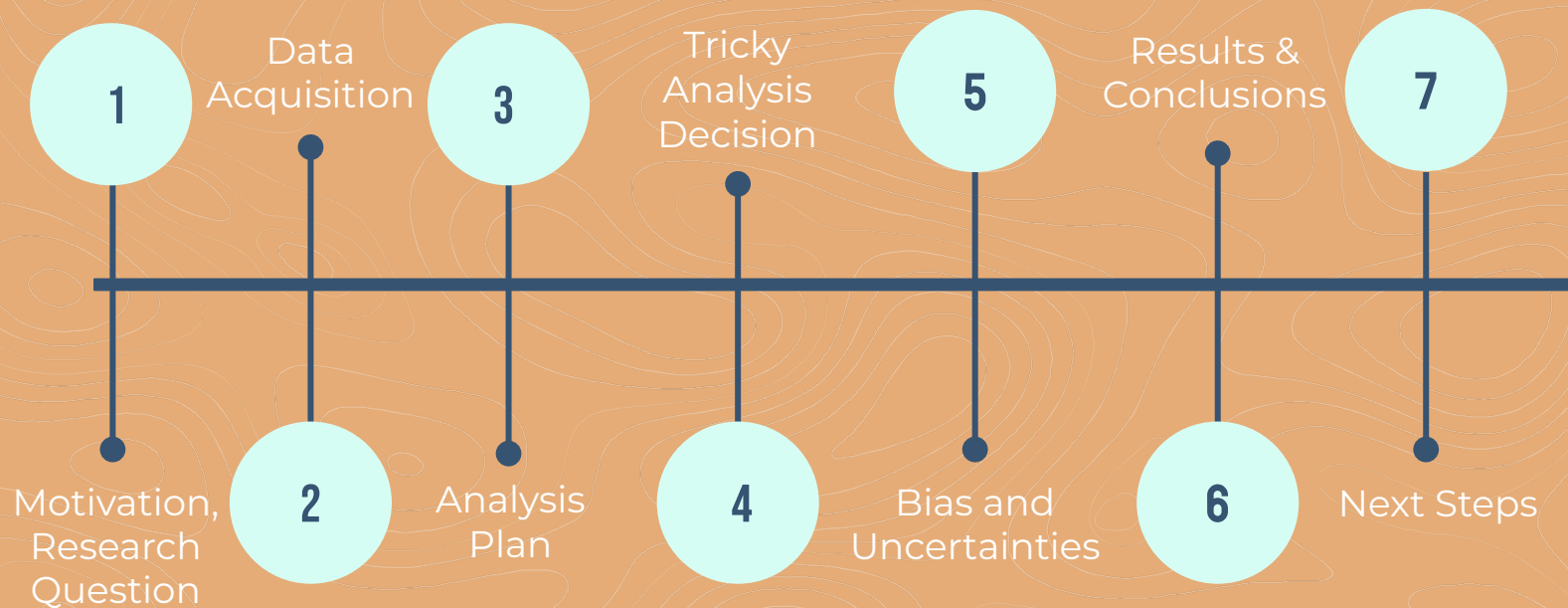
AUTOMATED IDENTIFICATION OF MALARIA

Group 7: Alex Smithgall (Leader), Andre
Chuabio, Sam Rea

DS 4002 4/23/24

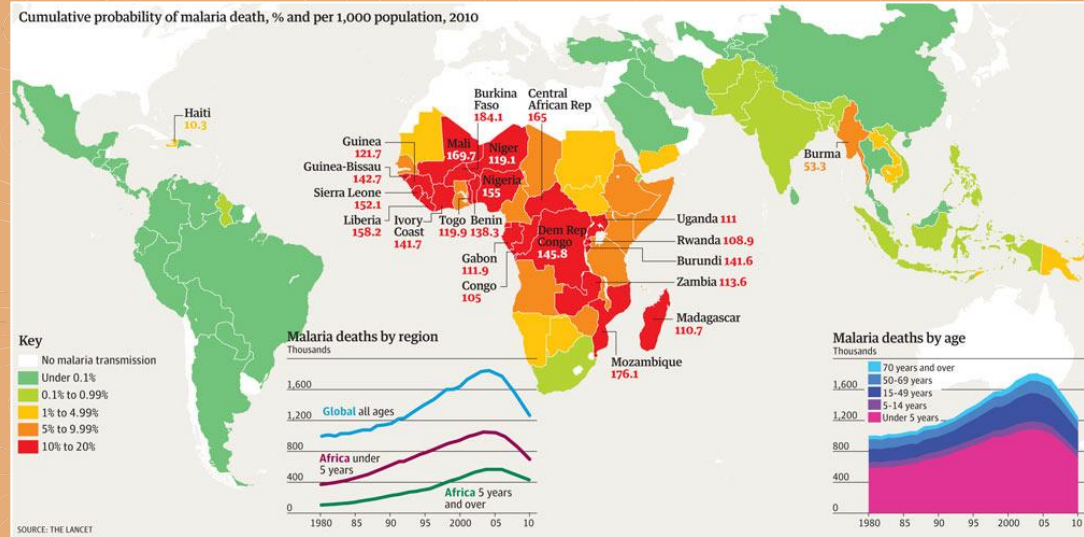


TABLE OF CONTENTS



BACKGROUND

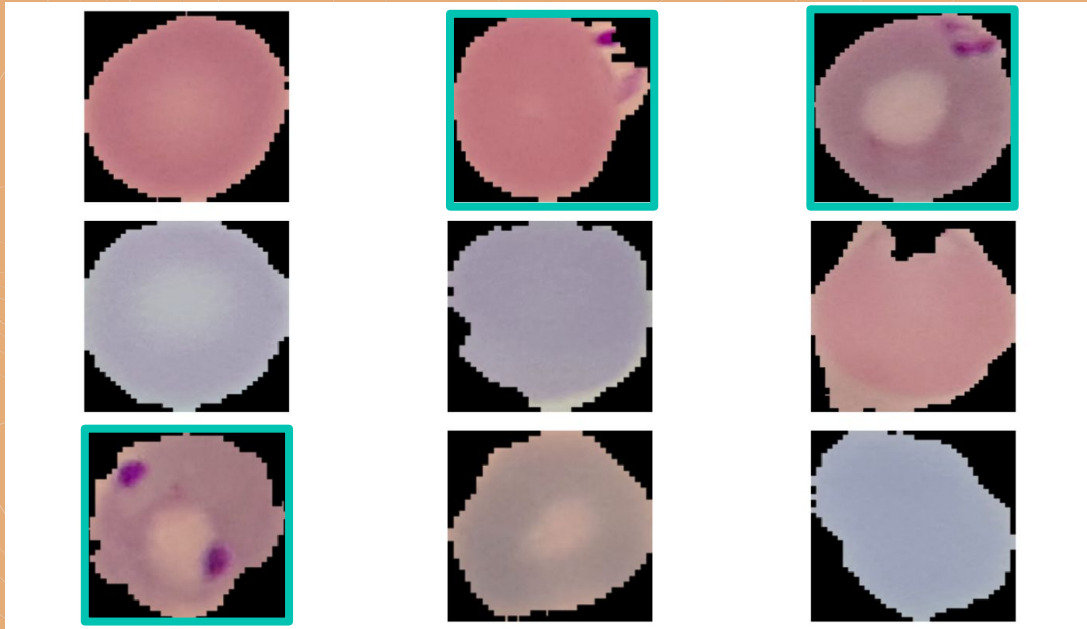
- 94% of Malaria cases, 95% of deaths in Sub-Saharan Africa
 - “Getting treatment early for mild malaria can stop the infection from becoming severe”
- Identifiable through Microscopy - reading a blood smear, hard to do in rural areas
- Africa: many healthcare workers with limited training, RDTs potentially unreliable



OUTLINE

Research Question	What level of specificity can we achieve in classifying malaria cells using deep learning algorithms?
Hypothesis	We will be able to identify cells infected with malaria with 95% accuracy
Approach	We will use TensorFlow and the Keras package in R to train a neural network image classification model, which will determine if cells are infected or not

DATA ACQUISITION



27,558 images, 2988x5312 JPEGs: 50/50 Parasitized/Uninfected
Official source is NIH but found tidier version on Kaggle (organized by subfolder)

ANALYSIS PLAN



Establish Dataset:
Load in Malaria image
data, split into train/tes.



Data Wrangling:
Preprocess images, scale
resolution, standardize



Modeling:
Run Keras Image
Classification with **25**
Epochs



Evaluation:
Evaluate on accuracy
>0.95

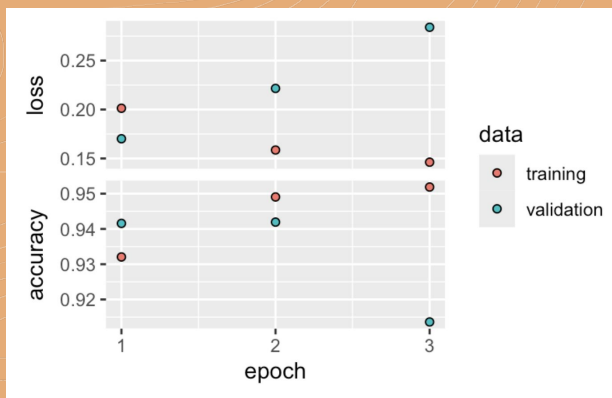
TRICKY ANALYSIS DECISION

ISSUE

- Balancing between image resolution and time to run model
- How many Epochs to run the model

SOLUTION

- Compromised at 256x256 pixels (blurrier)
- Ran 25 epochs to reduce variability



BIAS AND UNCERTAINTY EVALUATION

CELL SELECTION

- Quality of cell images dependant on the preparer
- Possible false negatives due to low Malaria incidence

MODEL OVERFITTING

- Volatility in validation accuracy - could increase validation split
- Used dropout layer to reduce overfitting

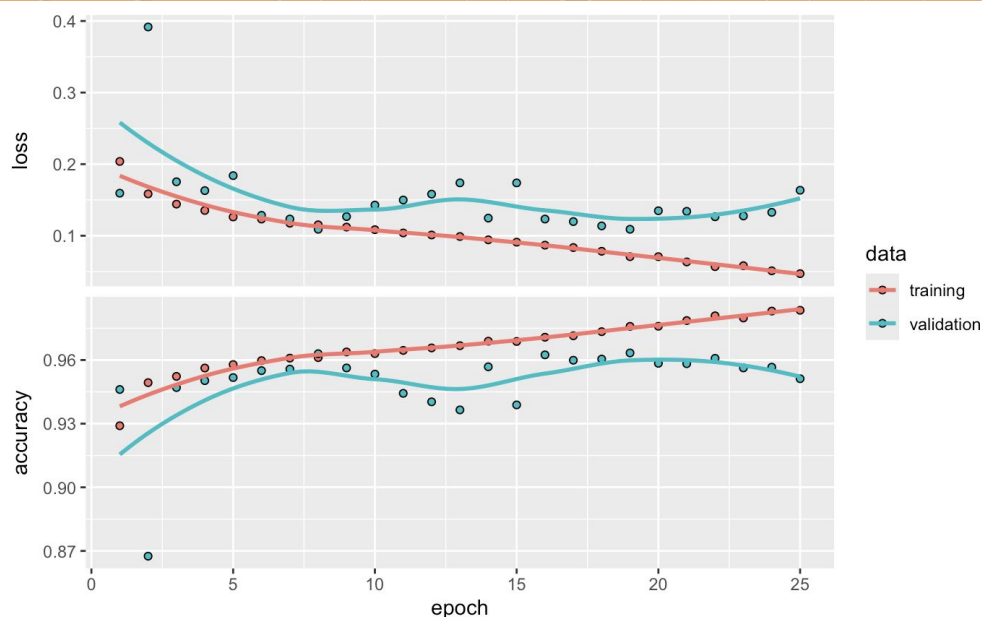


RESULTS

ACCURACY: 0.9834 (TRAINING), ~0.95 (TESTING)

Epoch 1/25
689/689 [=====] - 9893s 14s/step - loss: 0.2038 - accuracy: 0.9290
l_accuracy: 0.9461
Epoch 2/25
689/689 [=====] - 9896s 14s/step - loss: 0.1584 - accuracy: 0.9493
l_accuracy: 0.8675
Epoch 3/25
689/689 [=====] - 9809s 14s/step - loss: 0.1442 - accuracy: 0.9523
l_accuracy: 0.9470
Epoch 4/25
689/689 [=====] - 9516s 14s/step - loss: 0.1353 - accuracy: 0.9562
l_accuracy: 0.9503
Epoch 5/25
689/689 [=====] - 9524s 14s/step - loss: 0.1262 - accuracy: 0.9579
l_accuracy: 0.9517

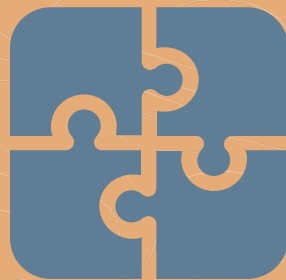
Epoch 21/25
689/689 [=====] - 9528s 14s/step - loss: 0.0634 - accuracy: 0.9786
val_loss: 0.1341 - val_accuracy: 0.9583
Epoch 22/25
689/689 [=====] - 9406s 14s/step - loss: 0.0568 - accuracy: 0.9809
val_loss: 0.1265 - val_accuracy: 0.9608
Epoch 23/25
689/689 [=====] - 9637s 14s/step - loss: 0.0581 - accuracy: 0.9798
val_loss: 0.1276 - val_accuracy: 0.9563
Epoch 24/25
689/689 [=====] - 9750s 14s/step - loss: 0.0510 - accuracy: 0.9831
val_loss: 0.1327 - val_accuracy: 0.9566
Epoch 25/25
689/689 [=====] - 9775s 14s/step - loss: 0.0470 - accuracy: 0.9834
val_loss: 0.1635 - val_accuracy: 0.9512



NEXT STEPS



**CONDUCT REAL-LIFE
SAMPLES**



REDUCE OVERFITTING

(Gather more data,
run more epochs,
experiment with
resolution scaling)



**LOOK INTO MORE
IN-DEPTH
CLASSIFICATION**

REFERENCES

- [1] "NLM - Malaria Data." [Online]. Available: <https://lhncbc.nlm.nih.gov/LHC-research/LHC-projects/image-processing/malaria-datasheet.html>
- [2] "TensorFlow for R - Image classification from scratch." Accessed: Apr. 12, 2024. [Online]. Available: https://tensorflow.rstudio.com/examples/image_classification_from_scrat
- [3] "Fact sheet about malaria." Accessed: Apr. 21, 2024. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/malaria>
- [4] P. J. Rosenthal, "How Do We Best Diagnose Malaria in Africa?," *Am J Trop Med Hyg*, vol. 86, no. 2, pp. 192-193, Feb. 2012, doi: 10.4269/ajtmh.2012.11-0619.
- [5] J. B. Parr *et al.*, "Analysis of false-negative rapid diagnostic tests for symptomatic malaria in the Democratic Republic of the Congo," *Sci Rep*, vol. 11, no. 1, p. 6495, Mar. 2021, doi: 10.1038/s41598-021-85913-z

Project GitHub: <https://github.com/samrea5/Project-3-Identifying-Malaria-Infections-with-Image-Classification/tree/main>.

