ASSESSING MALARIA BLOOD SAMPLES USING MACHINE LEARNING

KIERNAN HARDING

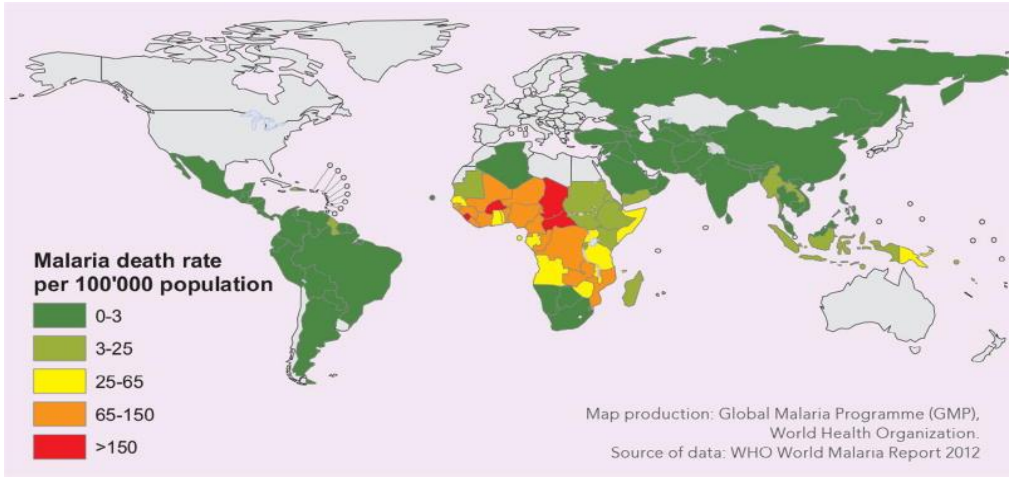
MaI.ARIA

IDENTIFICATION MADE SIMPLE

University of Essex

WHY?

- Malaria has a major impact on global health
- 2018: estimated 228 million cases worldwide & 405,000 deaths [2]
- Only 6 countries account for more than 50% of all cases [2]
- Nigeria accounts for 25% of all cases [2] and has 34% of the world average GDP (per Capita 2017) [8]
- Microscopists are required for testing and this can lead to many problems…

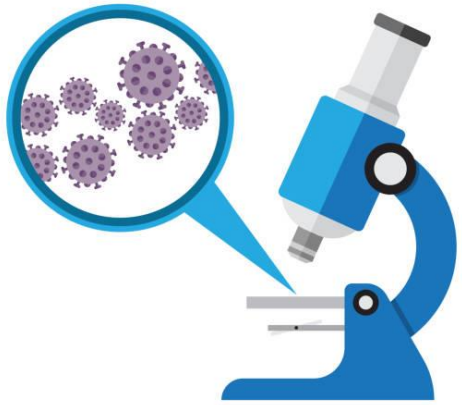


[1]

MAL.ARIA
IDENTIFICATION MADE SIMPLE

WHY USE AI?

- Yearly hundreds of millions of blood films examined by a trained microscopist
- This involves manual counting of parasites in red blood cells = timely
- Microscopists in poorer areas may have low quality training
  - Low-resource areas and poor-quality control setting
- Lead to incorrect diagnosis
  - False positives and False negatives
- Standardise using AI
- Faster testing
- Reduced workload
- More reliable (with an accurate model)



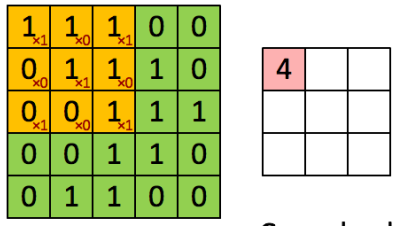
MaI.ARIA
IDENTIFICATION MADE SIMPLE

RESEARCH TOPICS

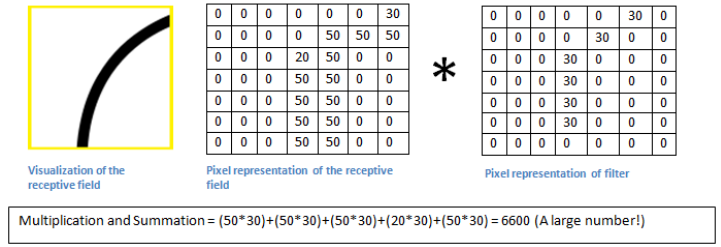
- Effects of malaria on world health and why the testing process would benefit from AI
- Using deep learning for image classifiers: specifically convolutional neural networks
- How to use TensorFlow and Keras to create a convolutional neural network (dogs and cats tutorial)
- How to implement a CNN model into a website (using python with an interactive website)

MAL.ARIA
IDENTIFICATION MADE SIMPLE

WHAT IS A CNN? (ML)

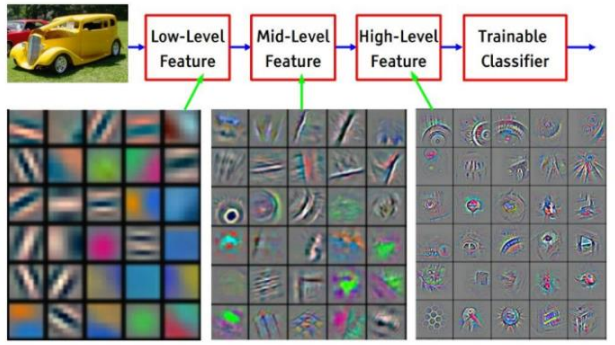


Convolutional Layer [10]

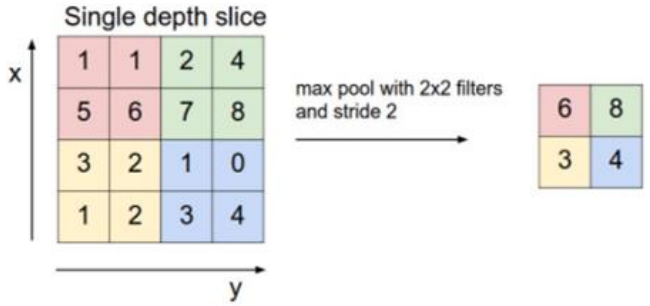


Applying filters to an image [5]

- Short for 'convolutional neural network'
- Class of deep learning neural networks
- Learns similarly to a child – inspired by the human brain
- Trained through backpropagation:
  - Forward pass – pass input data through network as normal as caches values
  - Backwards pass – go back through network and alter loss (similar to telling a child they're incorrect, which they then learn from)
- Different types of layers:
  - Convolutional – identifies features in images, such as straight lines and curves, aids learning process (becoming less abstract over time)
    - Activation (e.g. ReLU) – aids the network to learn complex patterns in the data – increase non-linearity (ReLU: returns 0 if negative and value x if positive)
  - Pooling (e.g. max pooling) – reduces sample size and thus speeds up processing
  - Fully connected – essentially the output layer (provides probabilistic values)



Convolved feature maps [9]

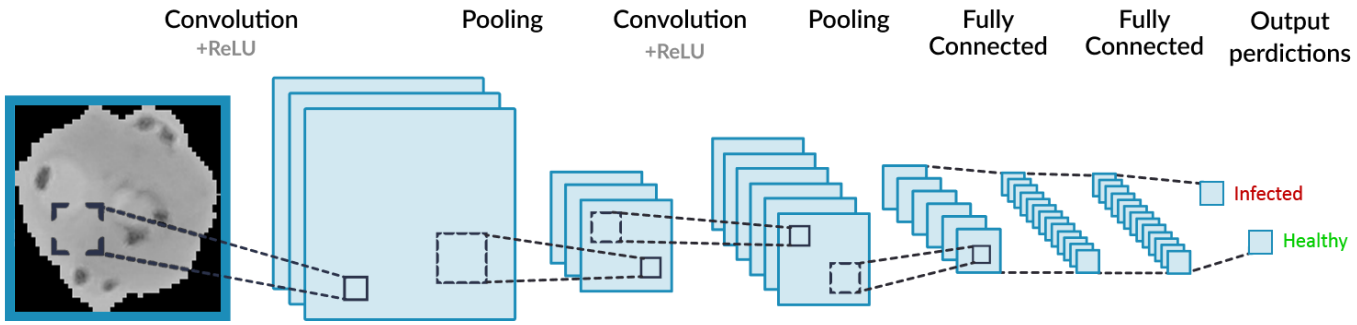


Max pooling [7]

MAL.ARIA
IDENTIFICATION MADE SIMPLE

CREATING MY CNN MODEL

- Splitting the dataset into training and testing (80/20)
- Preparing the dataset for training – resizing, greyscale…
- Created a convolutional neural network in python using TensorFlow and Keras
  - 3 main layers and out output layer (the fully connected layer)
  - These layers include: convolutional, activation (ReLU), pooling and fully connected
- Created a script and tested the models prediction accuracy: 71.4% (3,936/5,512) – overfitting
  - Incredibly bias towards infected samples: parasitized = 91.1%, healthy = 51.7%
- Created a windows batch file which automatically installs the required libraries for running any project related code



MAL.ARIA
IDENTIFICATION MADE SIMPLE

UI – WEB APPLICATION

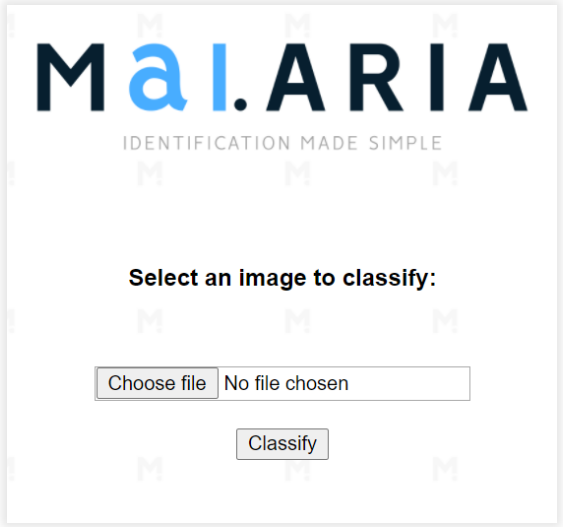
- Use of web development to increase the applications accessibility in poorer areas
- Microscope, camera, ideally an internet connection
- No specific requirements (e.g. android application = limits reach)
- Uses the Flask library to connect the python script to a website
- Simple as possible
- Takes an image input of a blood sample
- Classifies and outputs whether the sample is infected

MaI.ARIA
IDENTIFICATION MADE SIMPLE

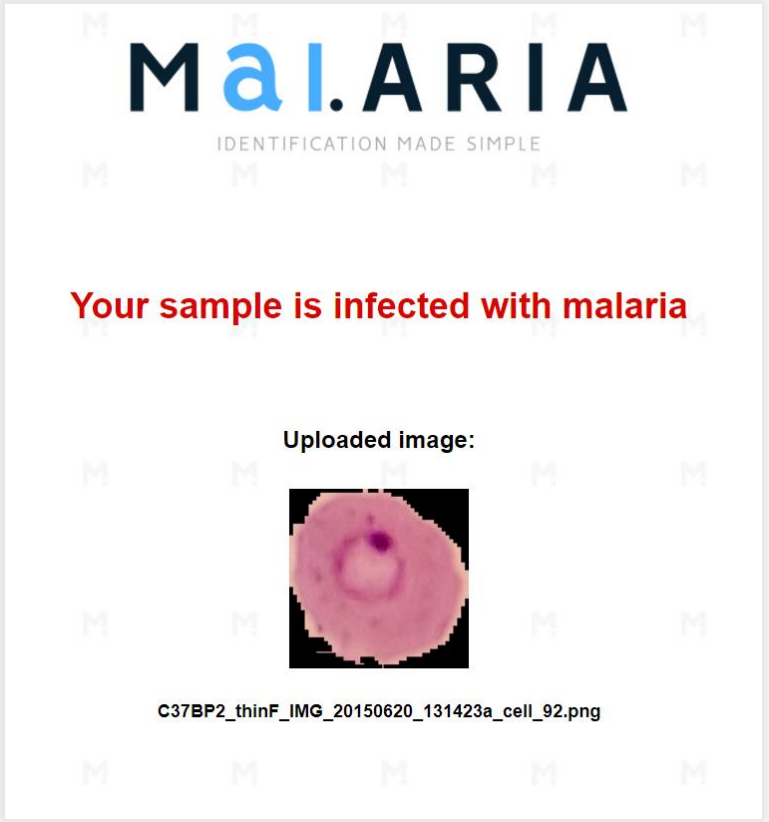
LIVE DEMO

- A web application that includes a back-end machine learning classifying model
- Examples of infected blood cells being classified
- Examples of healthy blood cells being classified

Back-up example:



>>



MAL.ARIA
IDENTIFICATION MADE SIMPLE

MAIN ISSUES ENCOUNTERED

- Installation of the required local environment – e.g. libraries such as TensorFlow
- Issues with array types between libraries – Occurred whilst creating a working convolutional neural network using tutorials (using the Kaggle dogs and cats dataset)
- Trying to save, use and classify the input image whilst designing the web application

MAI.ARIA
IDENTIFICATION MADE SIMPLE

MAIN INTERIM GOALS – HOW DID IT GO?

What was adjusted and achieved?

Specific MVP milestones	Estimated start/end dates
Create and train a convolutional neural network using python that identifies whether a red blood cell is infected.	19/10/20 – 16/11/20
Test the CNN and adjust accordingly to increase accuracy and remove unfavorable features such as overfitting.	16/11/20 – 23/11/20
Connect the algorithm to a local host web page. At this point it should take an image input and return whether the sample is infected.	23/11/20 – MVP Presentation

MaI.ARIA
IDENTIFICATION MADE SIMPLE

TIMELINE OF FUTURE OBJECTIVES

Improve my convolutional neural network model through research and structured testing *(e.g. add colour)*	Implement bootstrap into my website to improve usability across all devices	Improve the visual styling of the user interface	Slack for any issues encountered before the submissions
8/2/21	22/2/21	8/3/21	17/3/21

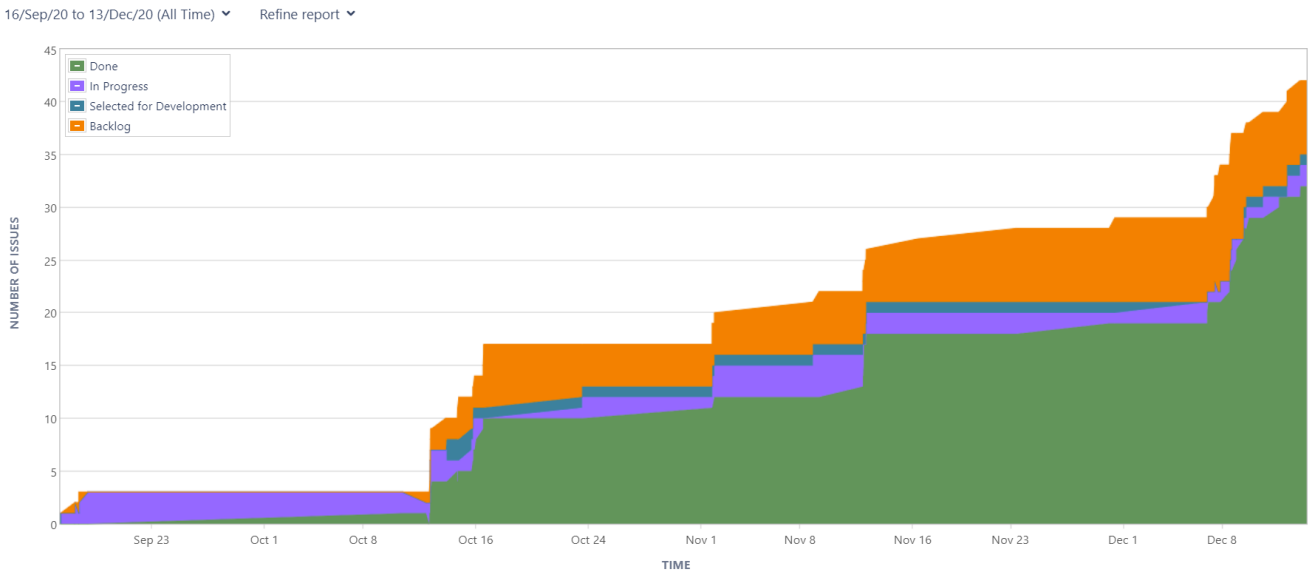
Now → 17/3/21 – Open Day Submissions

MalARIA
IDENTIFICATION MADE SIMPLE

USE OF JIRA

- Consistent use of Jira
- Use of tasks, sub-tasks comments and releases to document updates and plan future tasks

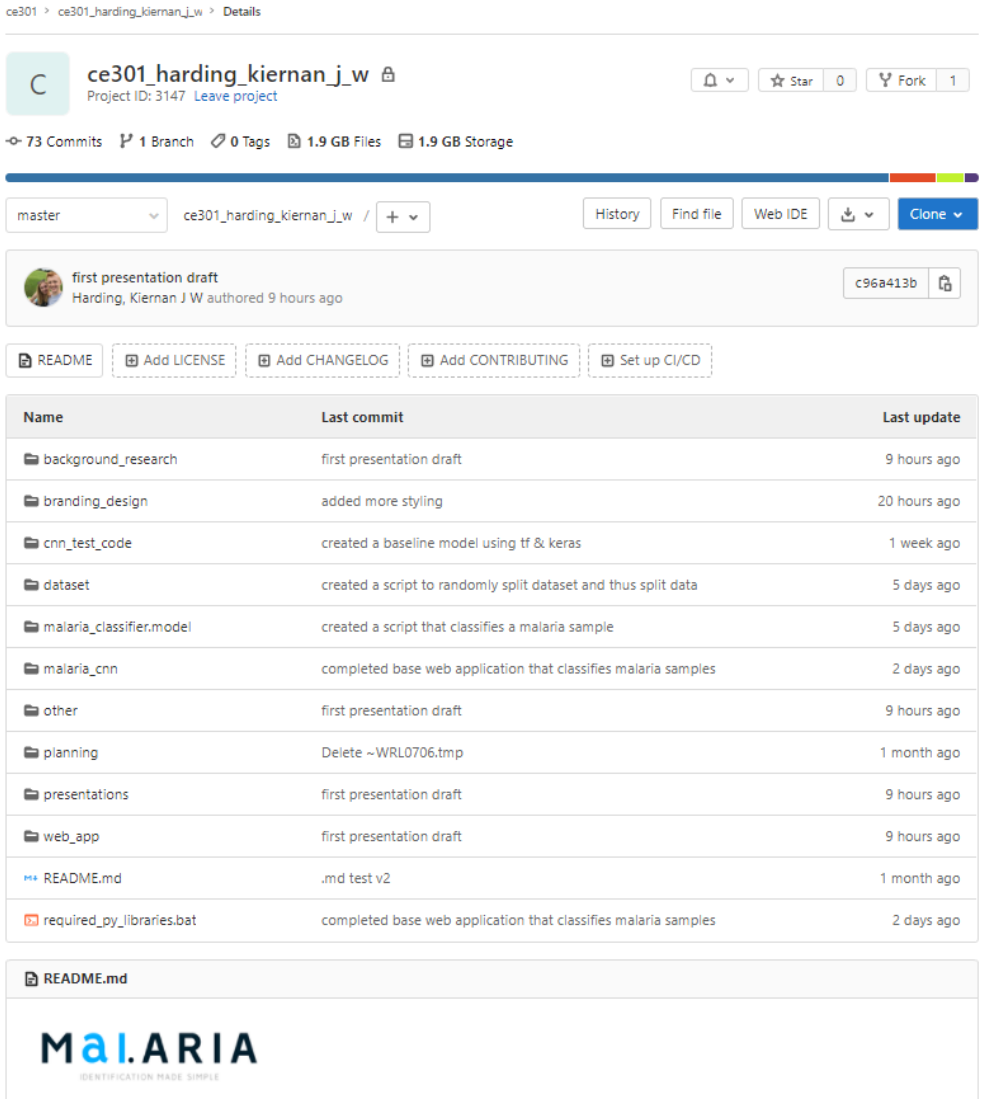
Cumulative flow diagram since the start of September



MAL.ARIA
IDENTIFICATION MADE SIMPLE

USE OF GIT

- Consistent use of Git
- Regular commits to Git for:
  - Significant changes
  - Keep important work backed-up!
  - Small required alterations: deletions
- Short, meaningful commit messages



ce301 > ce301_harding_kiernan_j_w > Details

C **ce301_harding_kiernan_j_w**
Project ID: 3147 Leave project

Star 0 Fork 1

73 Commits | 1 Branch | 0 Tags | 1.9 GB Files | 1.9 GB Storage

master | ce301_harding_kiernan_j_w / | History | Find file | Web IDE | Clone

first presentation draft
Harding, Kiernan J W authored 9 hours ago — c96a413b

README | Add LICENSE | Add CHANGELOG | Add CONTRIBUTING | Set up CI/CD

Name	Last commit	Last update
background_research	first presentation draft	9 hours ago
branding_design	added more styling	20 hours ago
cnn_test_code	created a baseline model using tf & keras	1 week ago
dataset	created a script to randomly split dataset and thus split data	5 days ago
malaria_classifier.model	created a script that classifies a malaria sample	5 days ago
malaria_cnn	completed base web application that classifies malaria samples	2 days ago
other	first presentation draft	9 hours ago
planning	Delete ~WRL0706.tmp	1 month ago
presentations	first presentation draft	9 hours ago
web_app	first presentation draft	9 hours ago
README.md	.md test v2	1 month ago
required_py_libraries.bat	completed base web application that classifies malaria samples	2 days ago

README.md

MAL.ARIA
IDENTIFICATION MADE SIMPLE

Snapshot of Git 14/12/20

MAL.ARIA
IDENTIFICATION MADE SIMPLE

THANK YOU, I HOPE YOU ENJOYED.

ANY QUESTIONS?

MAL.ARIA
IDENTIFICATION MADE SIMPLE

REFERENCES

- [1] M. Poostchi, K. Silamut, R. J. Maude, S. Jaeger, G. Thoma. (2018). Image analysis and machine learning for detecting malaria [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5840030/
- [2] WHO. (2019, Dec. 4). The "World malaria report 2019" at a glance [Online]. Available: https://www.who.int/news-room/feature-stories/detail/world-malaria-report-2019
- [3] S. Rajaraman, S. K. Antani, M. Poostchi, K. Silamut, Md. A. Hossain, R. J. Maude, S. Jaeger, G. R. Thoma1. (2018, Apr. 16). Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images [Online]. Available: https://peerj.com/articles/4568/
- [4] Expert System Team. (2020, May. 6). What is Machine Learning? A definition [Online]. Available: https://expertsystem.com/machine-learning-definition/
- [5] A. Deshpande. (2016, Jul. 20). A Beginner's Guide to Understanding Convolutional Neural Networks [Online]. Available: https://adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural- Networks/
- [6] H. Pokharna. (2016, Jul. 28). The best explanation of Convolutional Neural Networks on the Internet [Online]. Available: https://medium.com/technologymadeeasy/the-best-explanation-of-convolutional-neural-networks-on-the-internet-fbb8b1ad5df8
- [7] Stanford University (n.d.). CS231n Convolutional Neural Networks for Visual Recognition [Online]. Available: https://cs231n.github.io/convolutional-networks/
- [8] Worldometer (2017). GDP per Capita [Online]. Available: https://www.worldometers.info/gdp/gdp-per-capita/
- [9] Rohith Gandhi (May. 2018). An introduction to CNN and code [Online]. Available: https://towardsdatascience.com/build-your-own-convolution-neural-network-in-5-mins-4217c2cf964f
- [10] Sumit Saha (Dec. 2018). A Comprehensive Guide to Convolutional Neural Networks [Online]. Available: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

MAL.ARIA
IDENTIFICATION MADE SIMPLE