Malaria Detection

Deep Learning

Jessica Balke 11/05/2021

Can you tell which one is parasitized?

A. 

B. 

Can you tell which one is parasitized?

A. 

B. 

Parasitized

Can you tell which one is parasitized?

A. 

B. 

Can you tell which one is parasitized?

A. 

Parasitized

B. 

Agenda

- Problem statement
- Key Questions & Insights
- Approaches
- Final Solution Design
- Implementation Roadmap
- Benefits & Costs
- Risks & Challenges

Treatable if Detected Early

Developing Countries

Young Children & Pregnant Women

229M

People with malaria cases

400K

Deaths

Why is Malaria Detection and Important Problem?

Problem Statement

How is it being detected today?

- Traditional diagnosis requires a bloodsmear and costly analysis by a lab technician.
- Time consuming process that yields inaccurate results.
- Deep learning models can yield superior results in terms of accuracy.

Key Questions

- What are the important KPI's?
  - Accuracy and Precision
- Can deep learning methods yield an acceptable accuracy and precision, without overfitting?
  - Yes!
- Is the data evenly distributed?
  - Yes - ~14K parasitized cells and 14K uninfected cells
- What is our baseline for accuracy, time, and cost of the current approach?
  - TBD

Key Insights

Parasitized

Uninfected

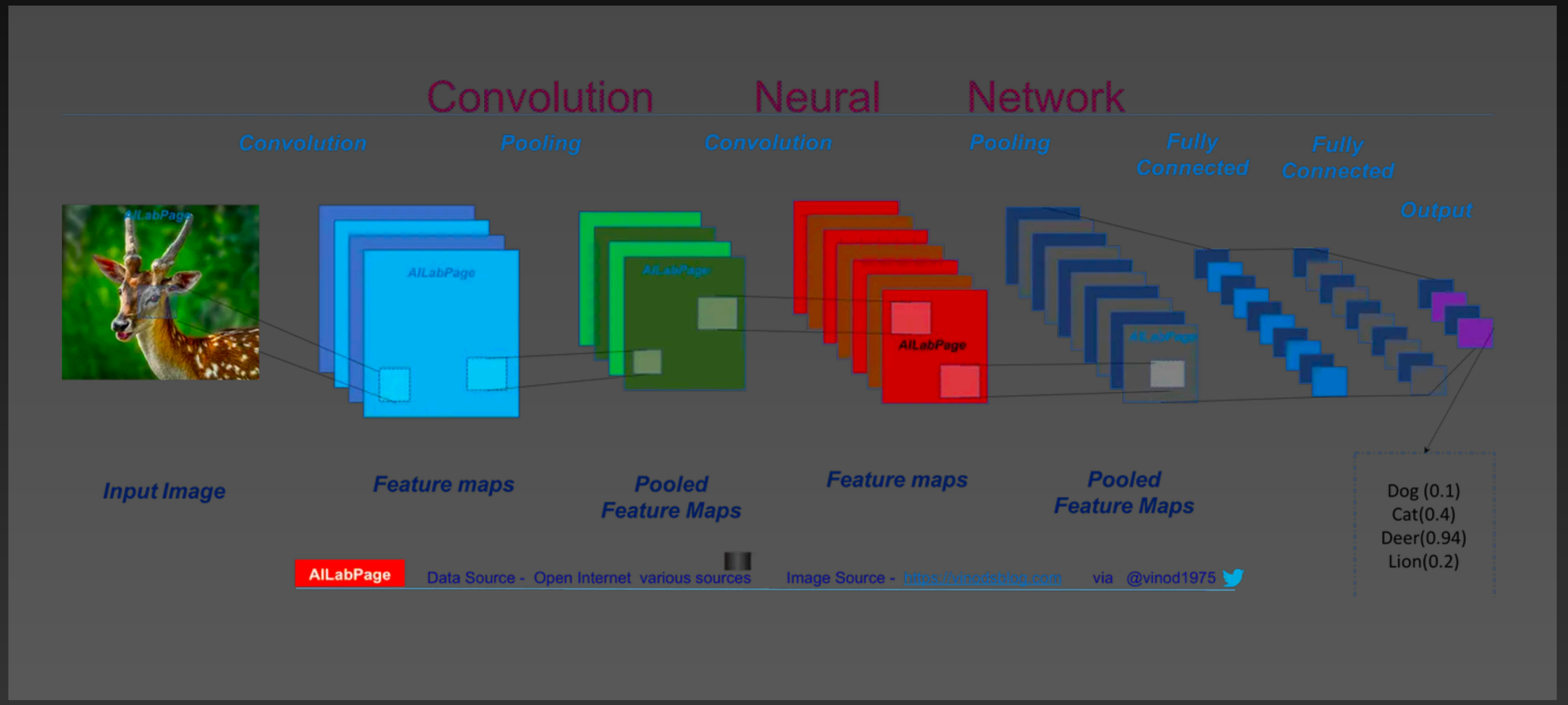
Attributes:

- Color
- Edges
- Hue
- Saturation

Approaches

- Convolutional Neural Networks (CNN)
- Transfer Learning
- Batch Normalization
- Leaky ReLU
- Data Augmentation

A Neural What?
Convolutional Neural Networks



- Consists of a convolutional layer, a pooling layer, and a fully connected layer.
- Each convolutional layer scans the image for a specific feature, whether it’s horizontal edges, vertical edges, etc.
- The essence of CNN is to reduce the image into a matrix of 1’s and 0’s that is easy to process, while retaining features good for prediction.

Approaches

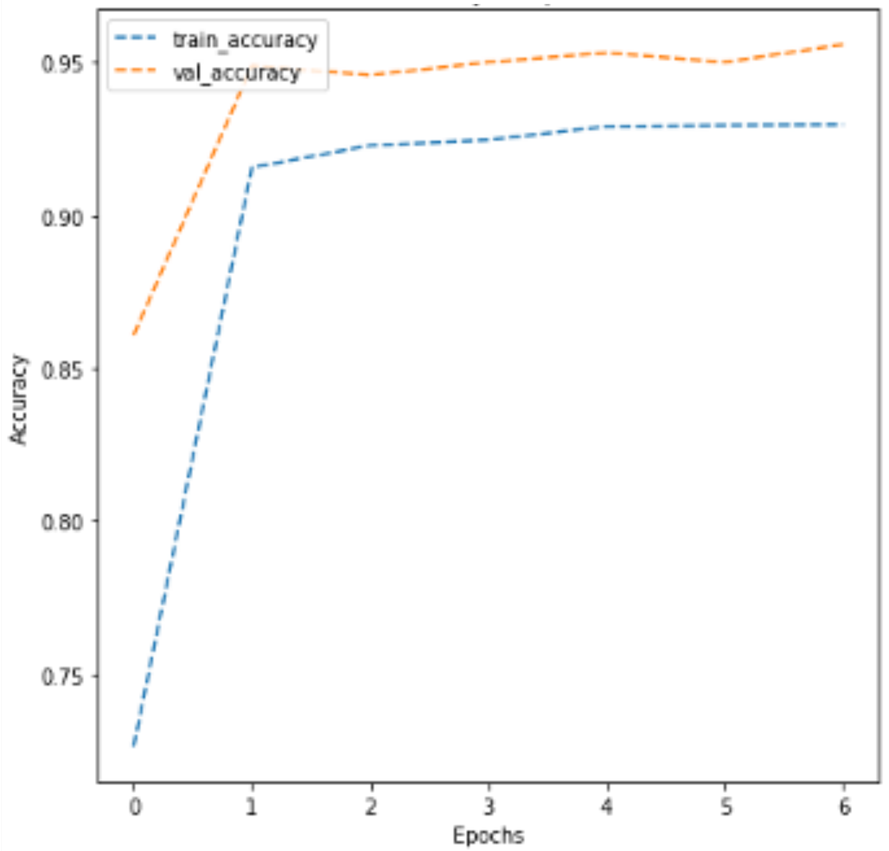
Model	Model Name	Accuracy	Precision	Recall
Baseline	Simple CNN (3 layers)	94.58%	96%	94%
Model 1	CNN Model (5 layers)	94.69%	93%	95%
Model 2	CNN Model Batch Normalization	95.19%	96%	94%
Model 3	CNN w/ Data Augmentation & Batch Norm	95.58%	97%	93%
Model 4	Transfer Learning	93.77%		

Final Solution Design

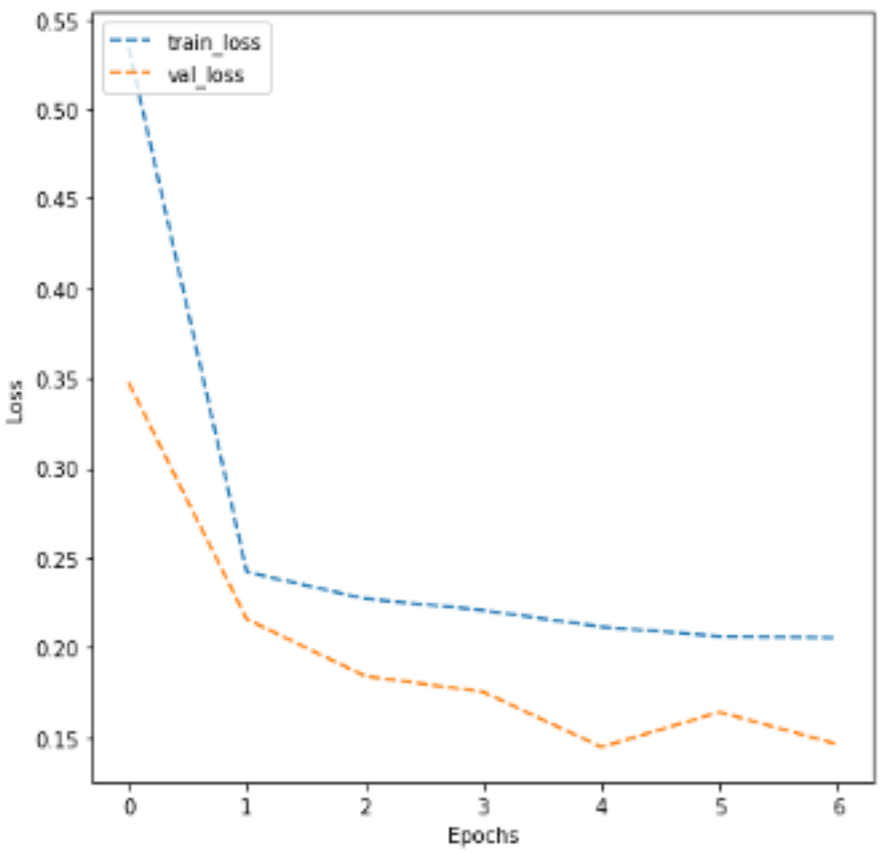
Model 3

Model 3 with Batch Normalization and Data Augmentation had highest accuracy of **95.58%** and precision of **96%**.

Accuracy



Loss



Implementation Roadmap

- Build software that contains an interface for HCP’s to upload images of cells
- They will receive a quick output of which cells were infected. It will be up to the HCP to make a diagnosis.
- Data warehousing and storage decisions
- Easy to understand and intuitive interface

Benefits and Costs

Benefits:

- Save time, money, and increase accuracy
- Malaria detection can be deployed at scale, meaning early detection, and better control of the disease.
- Save lives, improve economy, less disruptions to education, improve mental welfare and stability of nations.

Costs:

- Data warehousing, storage, and processing costs
- Software expense, hardware expenses
- Training, setup, and one-time startup costs
- Maintenance and ongoing costs
- Costs offset by savings from not paying a lab technician for manual process.

Risks and Challenges

- Deploying a solution in a developing country where infrastructure and is limited
- There may be cost restraints or hardware/software compatibility restrictions
- If solution is too computationally expensive, costly, or complex, it might not get used at all
- Important to understand what limitations are, overcome obstacles, and make sure process can fit in with constraints of system.
- Perhaps a donor or sponsor may offset some start up costs
- If complexity needs to be reduced, model will still be close to 95% accurate.