

# Capstone Session - 1

## Malaria Detection Project

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# Agenda

- 1) Malaria Detection Problem Statement
- 2) Solution Steps Walkthrough
- 3) General Best Practices
- 4) Q&A

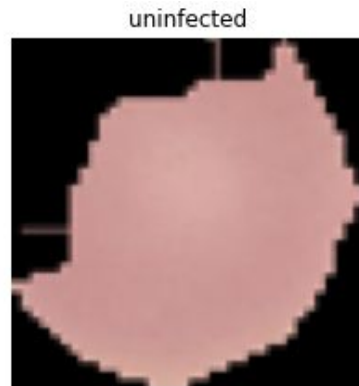
This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# Malaria Detection Dataset

- This dataset consists of colored images of red blood cells that contain parasitized and uninfected instances, where:
  - The **parasitized cells** contain the **Plasmodium parasite**
  - The **uninfected cells** are free of the Plasmodium parasites but could contain other impurities.
- The dataset contains a total of **24,958 train** and **2,600 test** images.



This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# Problem Definition and Objective

- Malaria is a contagious disease **caused by Plasmodium parasites** that are transmitted to humans through the bites of infected female Anopheles mosquitoes.
- **Almost 50% of the world's population is in danger from malaria.** There were more than **229 million malaria cases** and 400,000 malaria-related deaths reported over the world in 2019.
- **Children under 5 years of age** are the most vulnerable population group affected by malaria; in 2019 they accounted for **67% of all malaria deaths worldwide.**
- **To build a machine learning model that can help with the early and accurate detection of malaria.** It would be highly beneficial to propose a method that performs malaria detection using Deep Learning Algorithms.

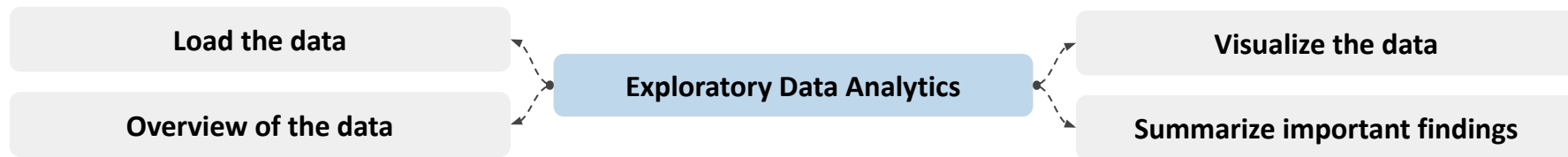


This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# Solution Step 1



Example of questions that can be answered by EDA:

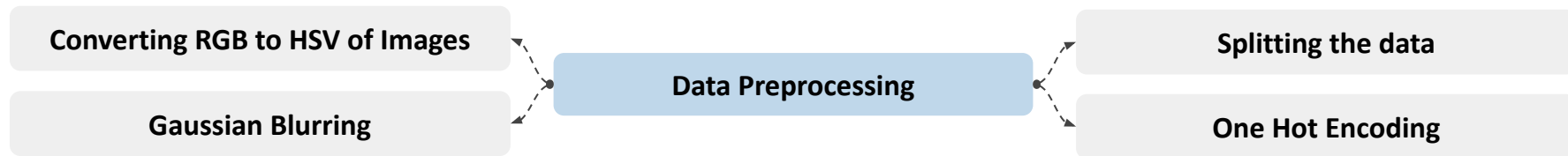
1. What is the shape of the dataset?
2. Is there imbalance in the classes of the target variable?
3. Should we normalize the images to visualize them better?
4. Are there visual distinctions between uninfected cells and parasitized cells?
5. Are there differences in the color intensity of the parasitized and uninfected cells?

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

## Solution Step 2



Example of questions that can be answered by Data Preprocessing:

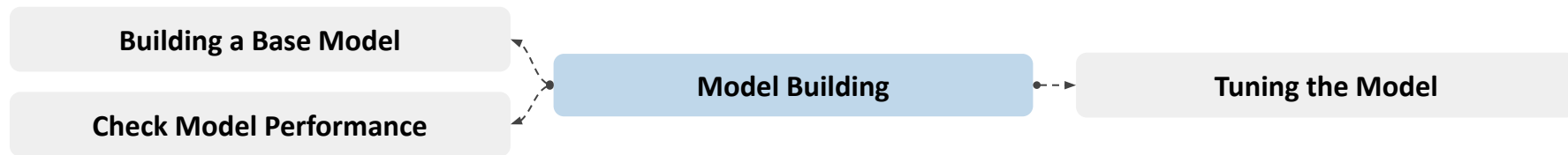
1. Are there new insights about the cells after converting the images to HSV?
2. Do you think that blurring would help us in any way?
3. Should we further split the data or the train and test folders can be simply used?
4. Do we need to encode the train and test labels?

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

## Solution Step 3



Example of questions that can be answered by Model Building:

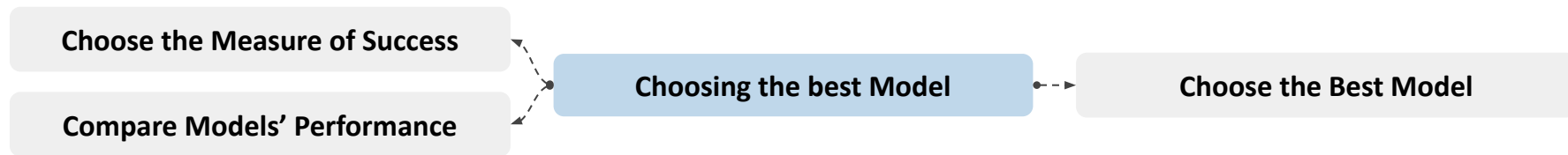
1. Can we use ANN to solve this problem?
2. Is the model overfitting the training data?
3. Can we improve the model performance by tuning different layers?
4. Would pre-trained models give better performance?
5. Should we try to train models on different types of processed images from Step 2?

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

## Solution Step 4



Example of questions that can be answered by Model Selection:

1. What is the metric (Measure of Success) for this business problem?
2. What is the architecture of the best model?

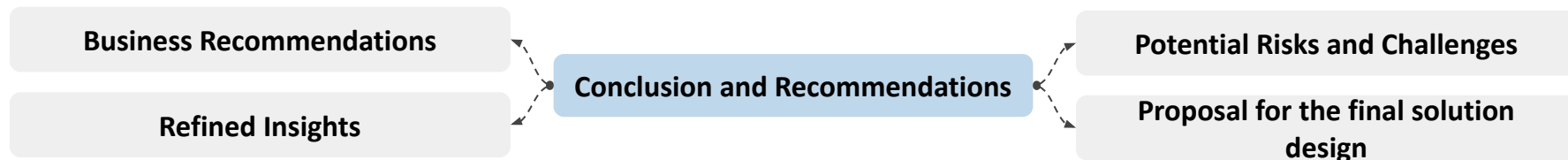
This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.



## Solution Step 5



Example of questions that can be answered by Conclusions and Recommendations:

1. What are the refined insights from EDA and model building?
2. What observations and insights can be drawn from the confusion matrix and classification report?
3. Is the model performance good enough for deployment in production?
4. What is proposal for final solution design? What are expected benefits and costs (assume numbers) of this solution design?

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# General Best Practices

Some of the best practices for submission:

- **Address all key questions in the rubric:** Make sure to read the rubric carefully and understand all the requirements. Address all the key questions asked in the rubric in your submission.
- **Provide observations and insights:** Provide observation and insight for each output, such as plots, summary statistics, missing values detection and treatment. This will help to make your work more understandable and actionable.
- **Explain your design steps:** Explain the steps you took to design your solution approach. This will help the reader to understand the overview of your solution approach and how you arrived at your final model.
- **Document the performance benchmarks:** Write down the performance benchmarks for all the Models you have trained and tested.

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# General Best Practices

- **Select the performance metric that best fits the business objectives:** Choose the performance metric that best fits the business objectives. This will help to ensure that your model is relevant and useful to the business.
- **Interpret potential benefits from the model:** Provide an interpretation of potential benefits from the model. This will help the business to determine the next steps and make informed decisions based on your work.

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

# Q&A

This file is meant for personal use by jacesca@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.