# Malaria Cell classification
## Architectural Decisions Document

### 1.1 Data Source

The data source for the project is taken from the official NIH website's data repository. It contains segmented cells from the thin blood smear slide images from the Malaria Screener research activity. Data set consist of 27,558 images with 13779 images of infected images and 13779 images of uninfected images.

#### 1.1.1 Technology Choice

IBM Watson Studio jupyter Notebook was used to get the data set directly from NIH website.

#### 1.1.2 Justification

!wget method directly connects to the NIH data repository and download the data set to the current IBM Watson studio working environment.

### 1.2 Enterprise Data

No enterprise data

#### 1.2.1 Technology Choice

#### 1.2.2 Justification

### 1.3 Streaming analytics

No Streaming analytics

#### 1.3.1 Technology Choice

#### 1.3.2 Justification

### 1.4 Data Integration

#### 1.4.1 Technology Choice

IBM Watson Studio jupyter Notebook was used for ETL process

#### 1.4.2 Justification

IBM Watson Studio Notebook can use to extract data from a data source and transform data to acceptable formats (for example keras model acceptable formats)

### 1.5 Data Repository

#### 1.5.1 Technology Choice

IBM Object store

### 1.5.2 Justification

Because IBM cloud object store is a cloud offering no administrative skills required. Data can be manipulated through Apache Spark, SparkSQL.

## 1.6 Discovery and Exploration

### 1.6.1 Technology Choice

IBM Watson Studio jupyter Notebook

### 1.6.2 Justification

Data can manipulate and visualize using IBM Watson Studio jupyter Notebook via python libraries such as pyspark, pandas, Matplotlib, opencv and numpy

## 1.7 Actionable Insights

### 1.7.1 Technology Choice

IBM Watson Studio jupyter Notebook

### 1.7.2 Justification

Keras can use to develop deep learning models for the image classification while scikit-learn can be used to develop non deep learning models to compare the model performance between deep learning and non-deep learning models.

## 1.8 Applications / Data Products

### 1.8.1 Technology Choice

IBM Watson Studio jupyter Notebook

### 1.8.2 Justification

Final product will be delivered as a jupyter Notebook. Also model weight can be save and load as HDF5 format and model structure can be saved as JASON or YAML.

## 1.9 Security, Information Governance and Systems Management

### 1.9.1 Technology Choice

IBM Watson Studio

### 1.9.2 Justification

IBM Cloud I Services provides global points of presence which includes DNS, global load balancer, distributed denial of service protection, web application firewall and transport layer security.