Numerical dataset

General Information on Dataset:

- **Name:** Life Expectancy Data
  https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who/data

- **Number of columns: 22**
  The total number of samples in the dataset: 2938
  and the size of each:
  Linear Regression → Training Samples: 80% & **Testing Samples:** 20%
  KNN → Training Samples: 80% & **Testing Samples:** 20%

Implementation details:

Linear Regression

```
Mean Squared Error: 3.435781597067537
Mean Absolute Error: 1.1853777135523018
R^2 Score: 0.9603420024762042
```

KNN

```
Mean Absolute Error: 1.1110373476913207
Mean Squared Error: 3.841911654599326
R-squared: 0.9569447569254192
```

Image dataset

General Information on Dataset:

- **Name:** Cell Images for Detecting Malaria
  https://www.kaggle.com/datasets/iarunava/cell-images-for-detecting-malaria

- **Number of classes:** 2
  & their labels: Parasitized & Uninfected
- **The total number of samples in dataset:** 27,558 images
  and the size of each: 64*64
- **The number of samples used in training:**
  Logistic Regression → Training Samples: 75% & **Testing Samples:** 25%
  KNN → Training Samples: 70% & **Testing Samples:** 30%

Implementation details:

Logistic Regression

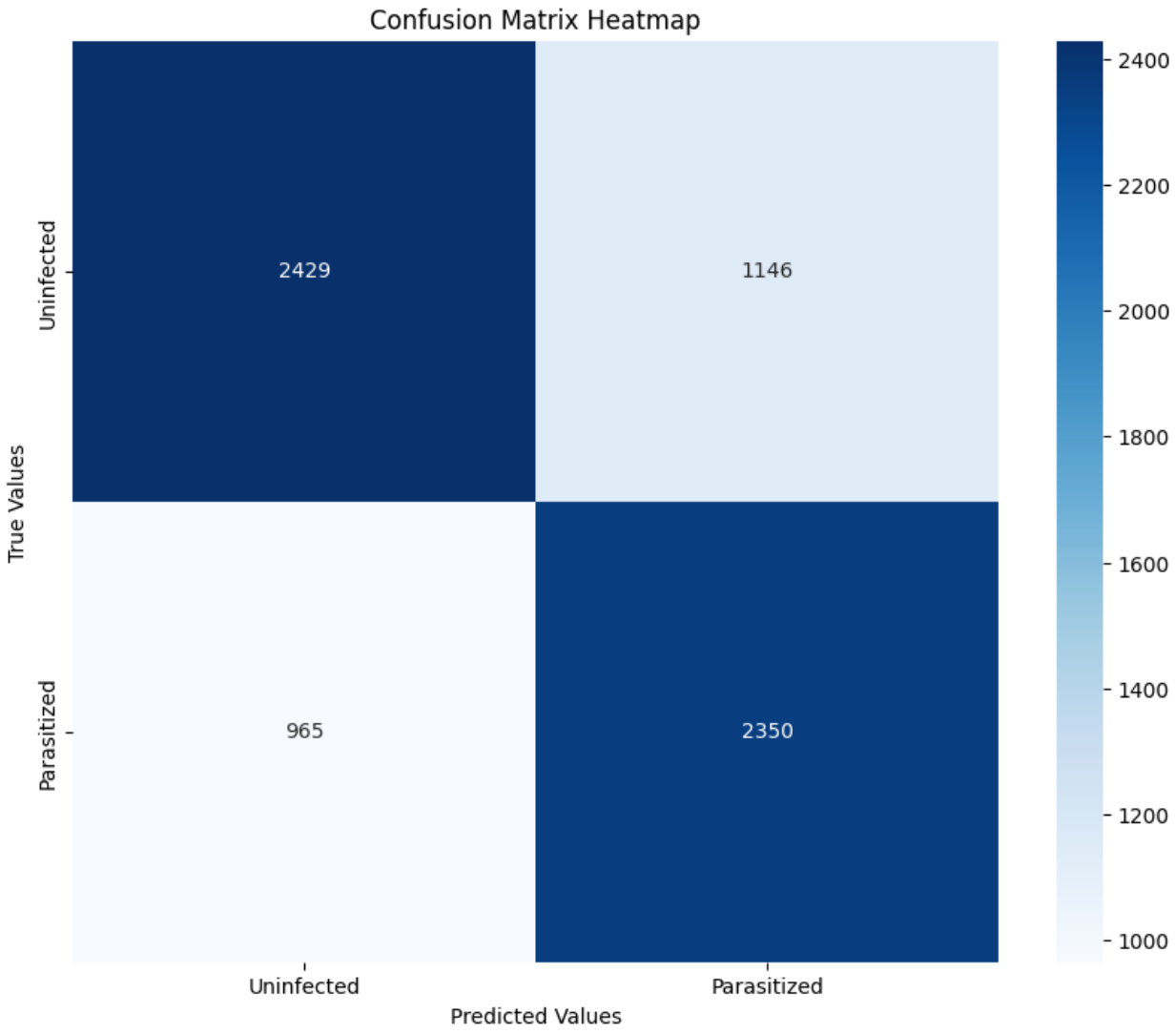
Accuracy

```
Logistic Regression accuracy is :  0.6936139332365747
```

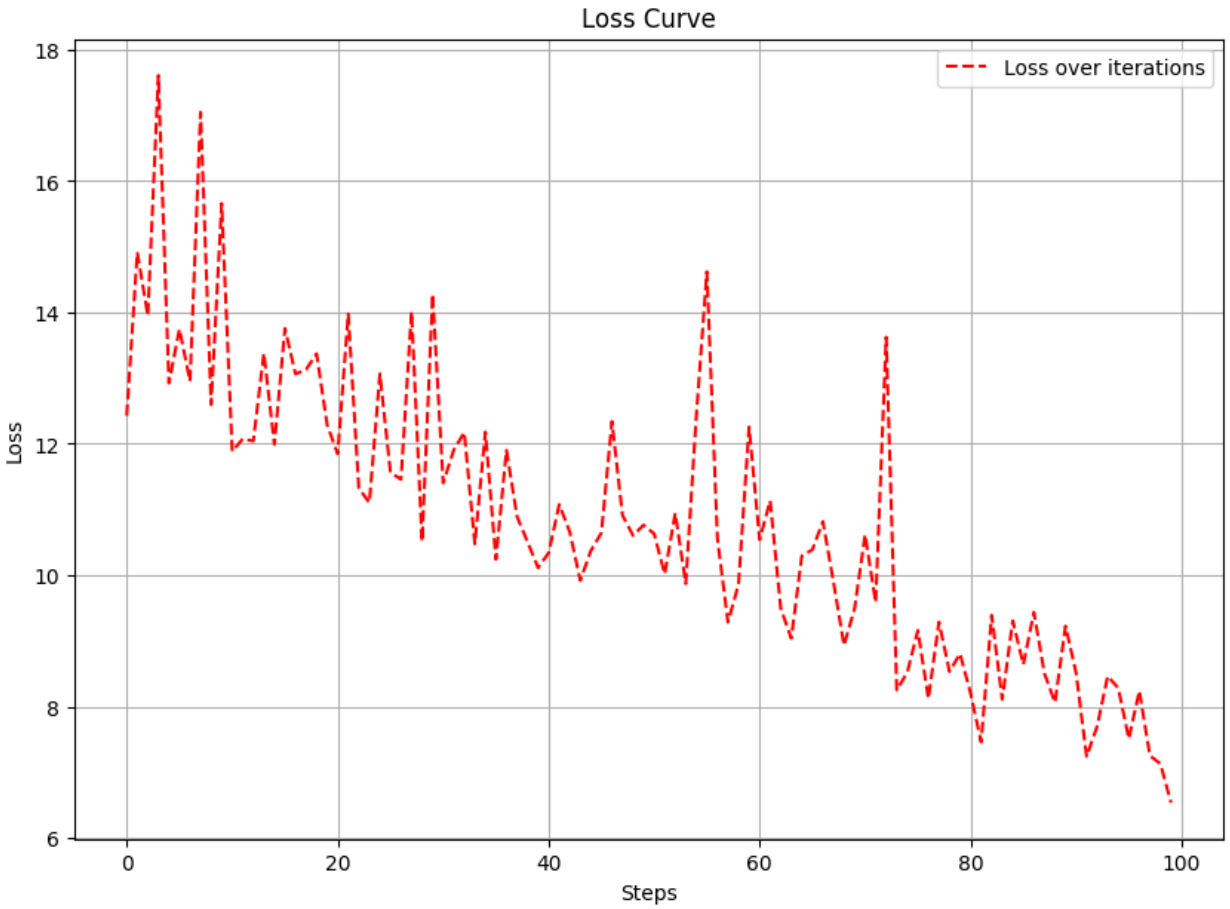
Precision & recall

	precision	recall
False	0.679441	0.715675
True	0.708899	0.672197
accuracy	0.693614	0.693614

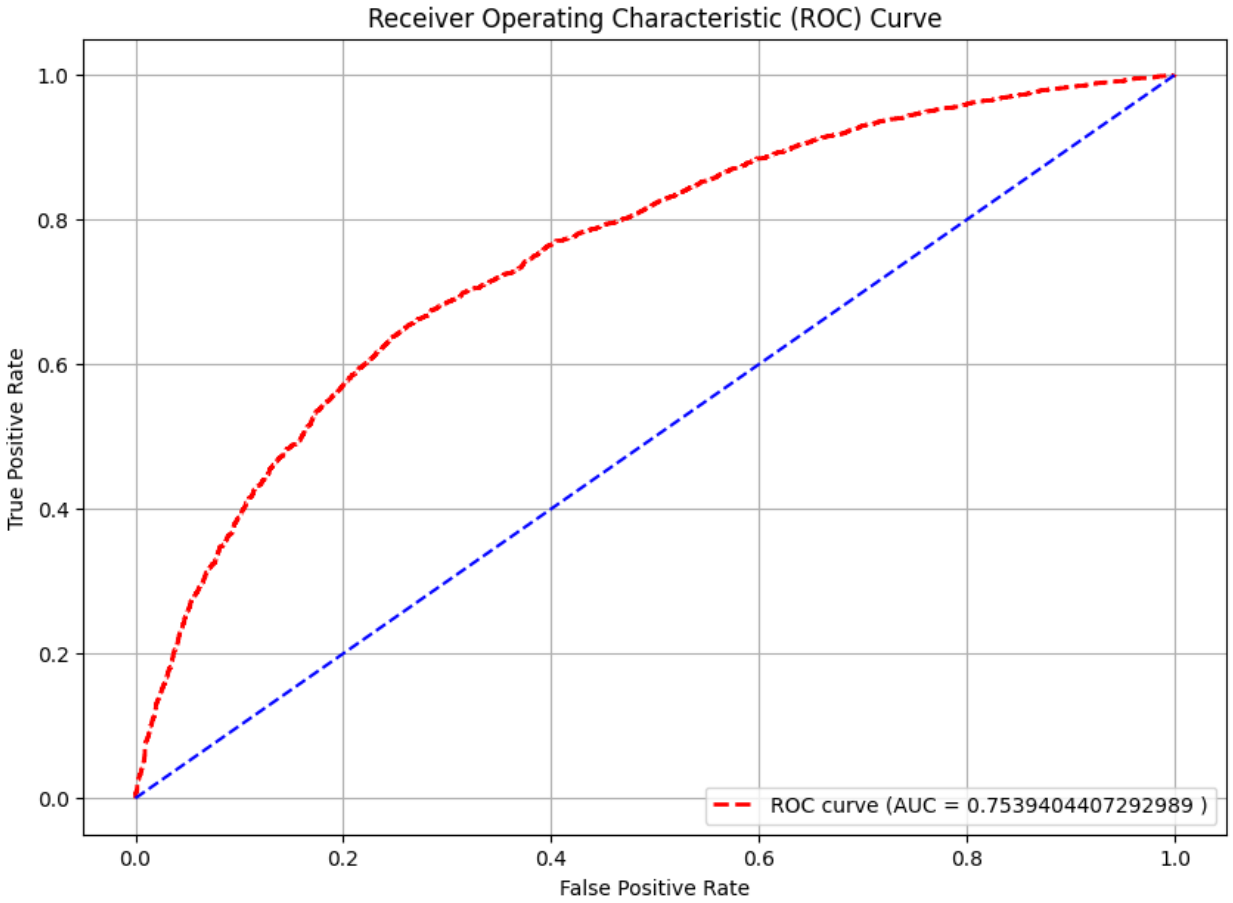
Logistic Regression



Logistic Regression



Logistic Regression



KNN

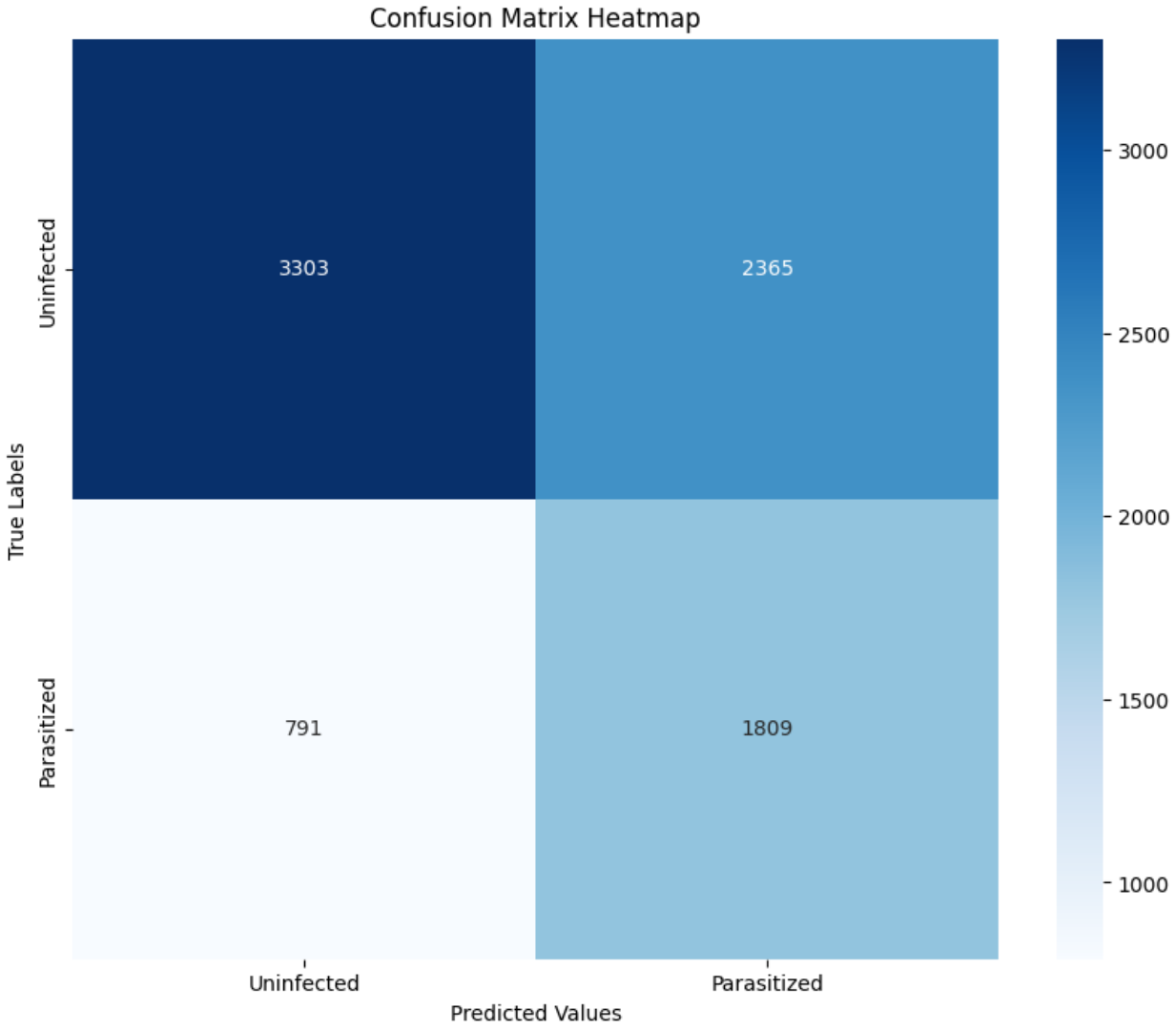
Accuracy

```
Accuracy is : 0.6182873730043541
```

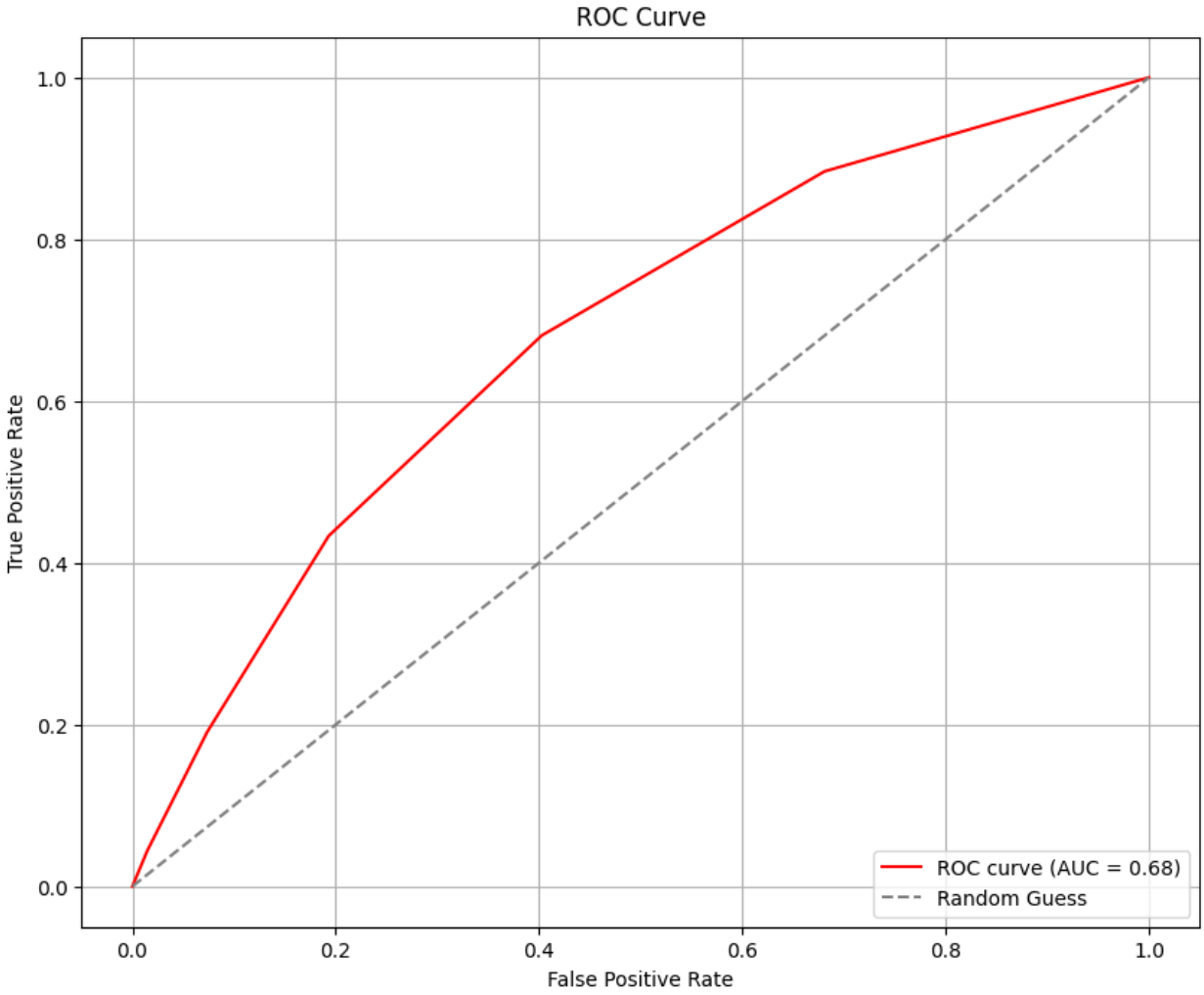
Precision & recall

	precision	recall
False	0.582745	0.806790
True	0.695769	0.433397
accuracy	0.618287	0.618287

KNN


Confusion Matrix Heatmap
Uninfected
Parasitized
True Labels
3303
2365
791
1809
Uninfected
Parasitized
Predicted Values
3000
2500
2000
1500
1000

KNN


ROC Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (AUC = 0.68)
Random Guess