

{PROJECT Description}

ML-Project



|  |  |
| --- | --- |
| **NAME** | **ID** |
| **Shahd ahmed** | **20210464** |
| **Hager sayed** | **20211023** |
| **Slama taha** | **20210414** |
| **Soliman Khalil** | **20210420** |
| **Huda maher** | **20211031** |
| **Omar abdallah** | **20210609** |

Numerical dataset:

A-General information

Name: wine dataset

number of samples used in training, validation, and testing*:*(1087,217,272)

20% testing,80%traning

B-Implementation details:

No feature extraction: because our dataset just numbers

Cross-validation:

A screenshot of a computer

Description automatically generated

C-result details:

Loss curve for

linear regression:

A screenshot of a computer

Description automatically generated

Accuracy:40%

Confusion matrix: no for linear regression because it is continuous values and *confusion matrix used in categorical data*

*Roc curve: use for logistic regression*

Knn:

A screen shot of a computer

Description automatically generated

Accuracy:63%

Confusion matrix:

A screenshot of a computer

Description automatically generated

Image dataset:

(logistic regression)

A - ***General Information on dataset***

Image dataset:

- Name: Malaria cell dataset.

- Number of classes = 2

- The labels are Parasitized and Uninfected

- the total number of images = 27558

- the size of each is 18.0 KB

* - Training Set: 80% × 27,558= 22,046 images
* - Testing Set: 10% ×27,558= 2,756 images
* - Validation Set: 10% ×27,558= 2,756 images

B - ***Implementation details:***

* One feature ( HOG )
* The dimension are 2 hog\_features hog\_image
* in the cross-validation the number of folds are 5 and the ratio of training data to validation data is 4:1 as the same ratio in initial train-test split
* 'log\_loss,' which includes a regularization term. The strength of the regularization can be adjusted with the alpha parameter.
* The patch size …updates the model parameters after processing each individual sample The effective batch size is 1.
* The number of epochs is 100 iterations

C - ***Results details***

* Loss curve

A graph with a blue line

Description automatically generated

* Accuracy

A screenshot of a computer

Description automatically generated

* *confusion matrix*

A blue squares with white text

Description automatically generated

(k-means)

A- Name: Malaria cell dataset.

the total of image in daatset=27550 image but Num of used in train and test totally =2000

Use in train=70% ,use in test=30%.

Size of images(224,224)

B-Implementation details:

1\_ resize images:

A screenshot of a computer program

Description automatically generated A screenshot of a computer

Description automatically generated

2\_ feature extraction:

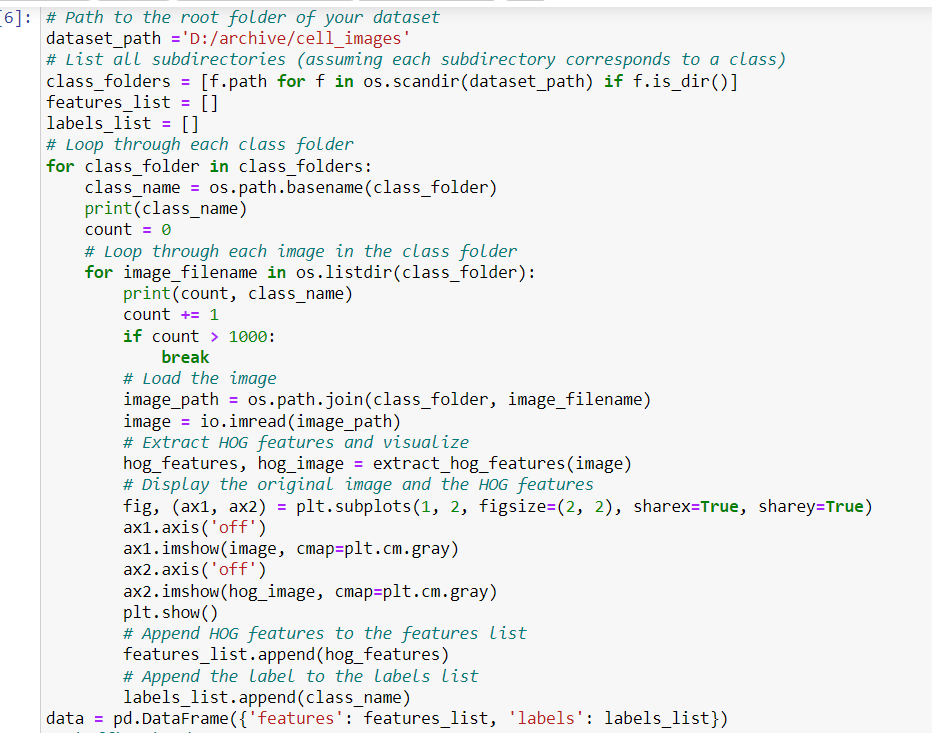
Num of feature :2000.

Name of feature is uninfected and Parasitized.

The dimension is (224,224).

The code of feature extraction:

A screenshot of a computer program

Description automatically generated 

Output:

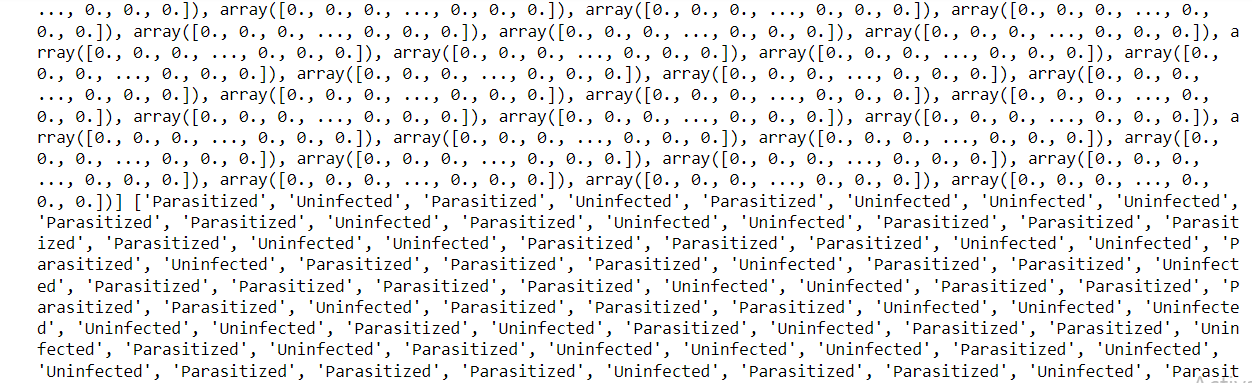
A screenshot of a computer screen

Description automatically generated

3\_ shuffled data :

A screen shot of a computer code

Description automatically generated



3\_ Convert lists to NumPy arrays:

A screenshot of a computer program

Description automatically generated

accuracy before k\_mean algorithm is: 0.97. A screenshot of a computer

Description automatically generated

the k\_mean algorithm:

A screenshot of a computer code

Description automatically generated

A screenshot of a computer screen

Description automatically generated

K \_ mean algorithm accuracy :

