

CS 395 Selected Topics in CS-1

Research Project

Report Submitted for Fulfillment of the Requirements and ILO's for Selected Topics
in CS-1 course for Fall 2021

Team No. 24

	ID	Name	Grade
1.	201900505	عمر حامد عبدالنبي سند	
2.	20180557	محمود رمضان احمد عيده	
3.	20180541	محمد مصطفى عبدالعال فرج	
4.	201900768	محمود سعد محمد بكر	
5.	201900382	صلاح بكر فاروق عبدالله	
6.	201900397	عبدالرحمن ابراهيم عبدالرحمن محمد	
7.	201900150	اسماء محمود عبدالجبار محمود	

Delivered to:

Dr. Wessam El-Behaidy

Eng. Islam Gamal

Eng. Muhammed Kamal

1. Project Introduction

General information on Dataset

Numerical Dataset:

Name: heart_failure_clinical_records_dataset

URL:<https://www.kaggle.com/anderwmvd/heart-failure-clinical-data>

Number of class (features):12

Type of label	Name
features	Age : age person
features	Anaemia :decrease of blood cells or hemoglobin
features	creatinine_phosphokinase :level of cpk enzyme in the blood
features	Diabetes : if the patient has diabetes
features	ejection_fraction: precentage of blood leaving the heart at each contraction
features	high_blood_pressure : if the patient has hypertension
features	Platelets : plateles in the blood
features	serum_creatinine : level of creatine in the blood
features	serum_sodium : level of serum sodium in the blood
features	Sex : woman or man
features	Smoking
features	time
Target	death event : true or not

Number of sample	
training	293 sample
validation	60 sample
testing	60 sample
Total sample	299 sample

Image Dataset:

Name: kagglecatsanddogs_3367a

URL : [https://www.kaggle.com/ karakaggle/kaggle-cat-vs-dog-dataset](https://www.kaggle.com/karakaggle/kaggle-cat-vs-dog-dataset)

Number of class (features): 9216

Number of sample	
training	19930 sample
Validation training	19930 sample
testing	4983 sample
validation test	4983

Model_1: support vector machine for dataset numerical

1- Features extraction

The number of features is 12 features

The number of sample and features for training data: (239, 12)

The number of sample and features for testing data: (60, 12)

The number of sample and features for validation data: (60,)

2- Preprocessing data

We make standard scaling for the features

3- Cross validation

We use the library KFold and cross_val_score to make the cross validation

Number of fold = 10 splits

The ratio of training scores: 0.679 (mean)

The standard derivation: 0.089 (std)

4- Grid search

we use the Grid Search method to choose the best parameters for the model

the best parameters for model support vector machine :

C : 1000 ()

Gamma: 0.0001 ()

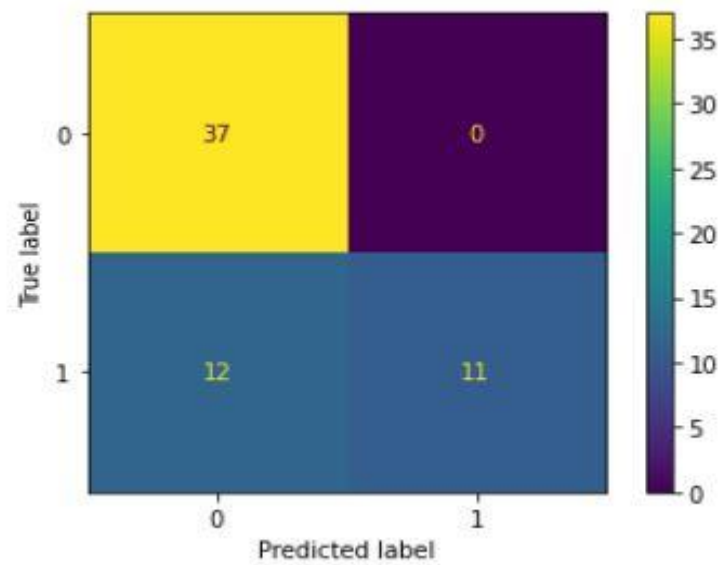
Kernel : rbf ()

The best score in grid is 0.83677

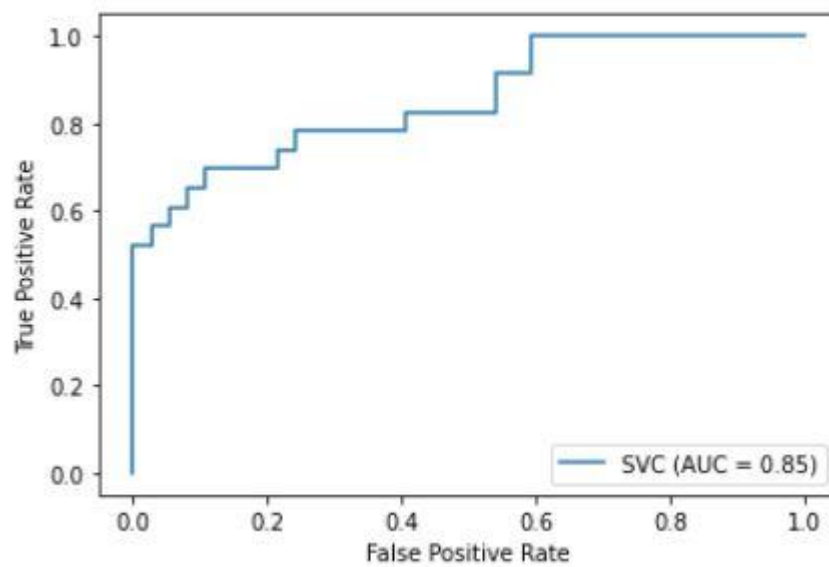
5- The classification report

precision	recall	f1-score	support	
0	0.76	1.00	0.86	37
1	1.00	0.48	0.65	23
accuracy			0.80	60
macro avg	0.88	0.74	0.75	60
weighted avg	0.85	0.80	0.78	60

6- confusion matrix



7- ROC curve



Model_2: artificial neural network for dataset numerical

1- Features extraction

The number of features is 12 features

The number of sample and features for training data: (239, 12)

The number of sample and features for testing data: (60, 12)

The number of sample and features for validation data: (60,)

2- Preprocessing data

We make standard scaling for the features

3- Cross validation

We use the library KFold and cross_val_score to make the cross validation

Number of fold = 10 splits

4- Grid search

we use the Grid Search method to choose the best parameters for the model

The best parameters for model support vector machine :

Batch size :10

Dropout rate: 0

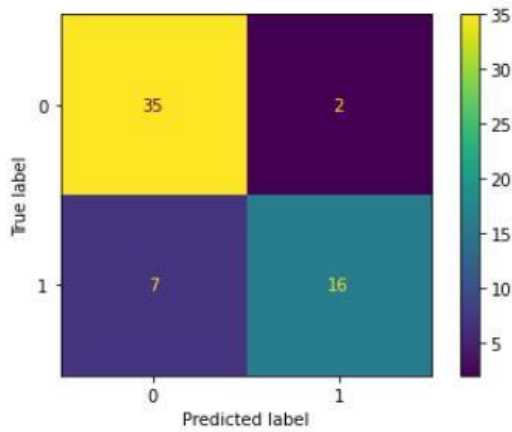
Epochs:10

Learning rate: 0.001

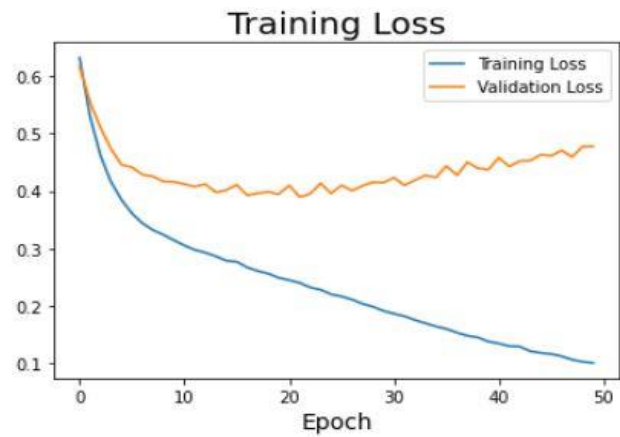
5- The classification report

	precision	recall	f1-score	support	
	0	0.83	0.95	0.89	37
	1	0.89	0.70	0.78	23
accuracy				0.85	60
macro avg		0.86	0.82	0.83	60
weighted avg		0.85	0.85	0.85	60

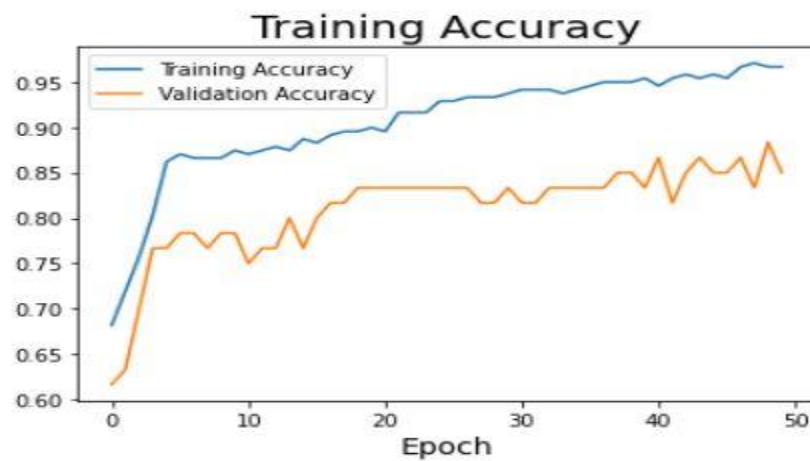
6- confusion matrix



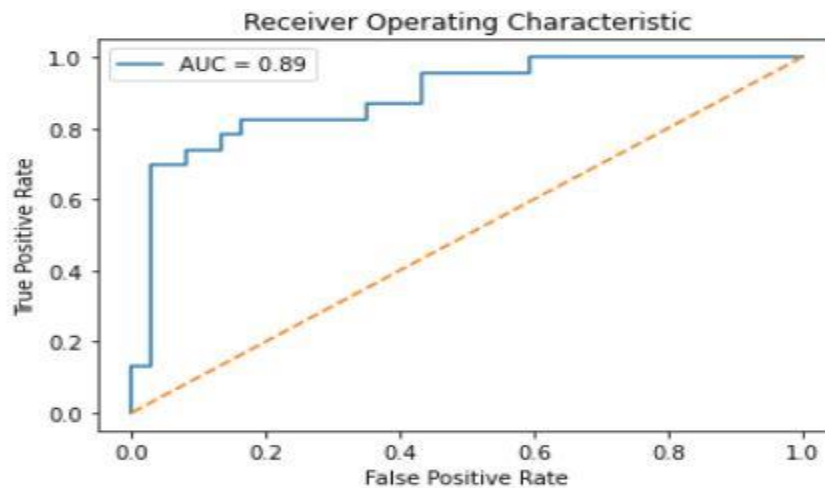
7- loss curve



8- training curve



9-(ROC –AUC) curve.



Model_3 Artificial neural network for dataset image

1- Features extraction

the number of features extraction :9216
the number of sample for training: 19930
the number of sample for testing 4983
the number of sample for validation training: 19930
the number of sample for validation test: 4983

2- Preprocessing data

- 1- we read image form the dataset
- 2- convert image from RGB to grayscale
- 3- convert image from grayscale to Histograms features
- 4- flatten the image that is make histograms

3- Cross validation

We use the library KFold and cross_val_score to make the cross validation

Number of fold = 2 splits

4- Grid search

we use the Grid Search method to choose the best parameters for the model

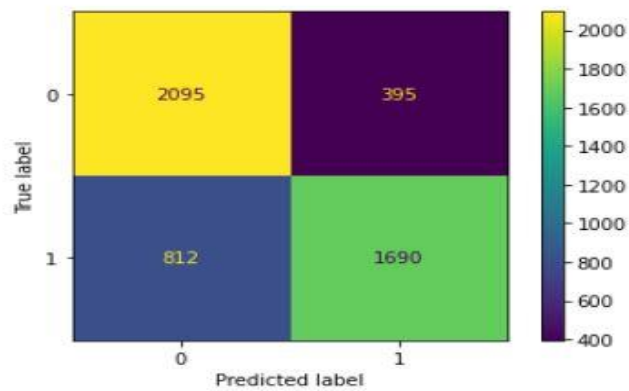
The best parameters for model support vector machine :

Batch size : 20 Dropout rate: 0
Epochs: 20 Learning rate: 0.001

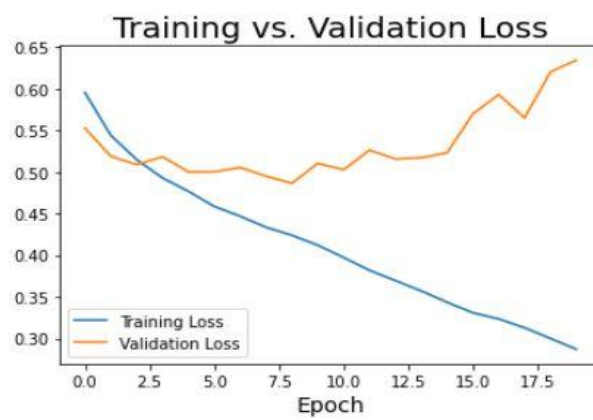
5- The classification report

precision	recall	f1-score	support	
0	0.73	0.80	0.77	2477
1	0.78	0.71	0.75	2506
accuracy			0.76	4983
macro avg	0.76	0.76	0.76	4983
weighted avg	0.76	0.76	0.76	4983

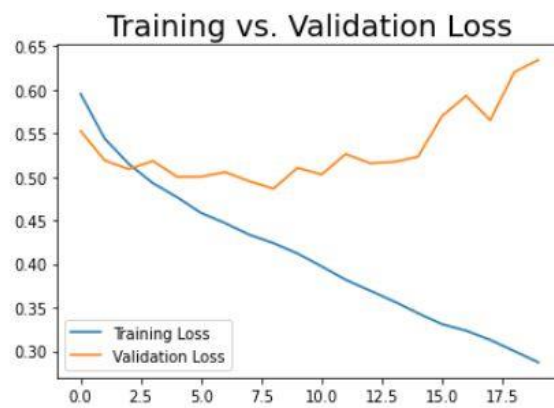
6- confusion matrix



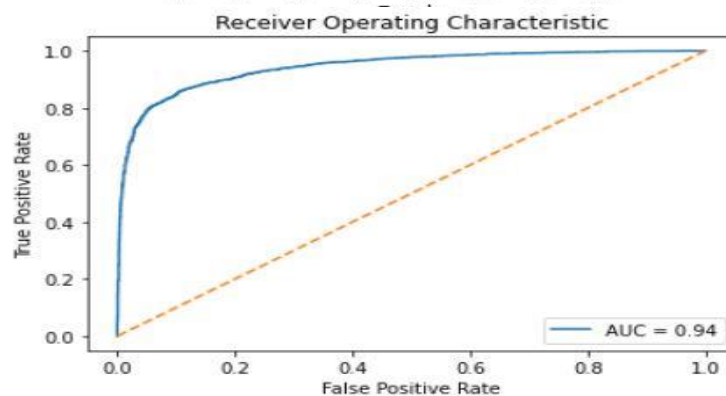
7- loss curve



8- learning curve



9 –ROC curve



Model_4 Artificial neural network for dataset image

1- Features extraction

the number of features extraction :9216
the number of sample for training: 19930
the number of sample for testing 4983
the number of sample for validation training: 19930
the number of sample for validation test: 4983

2- Preprocessing data

- 1- we read image form the dataset
- 2- convert image from RGB to grayscale
- 3- convert image from grayscale to Histograms features
- 4- flatten the image that is make histograms
- 3- Standard scalar and PCA

We used standard scalar to scale features between small value

Then used the PCA to downsize the number of features for image

3- Cross validation

We use the library KFold and cross_val_score to make the cross validation

Number of fold = 4 splits

Accuracy: 0.706 (0.010)

4- Grid search

we use the Grid Search method to choose the best parameters for the model

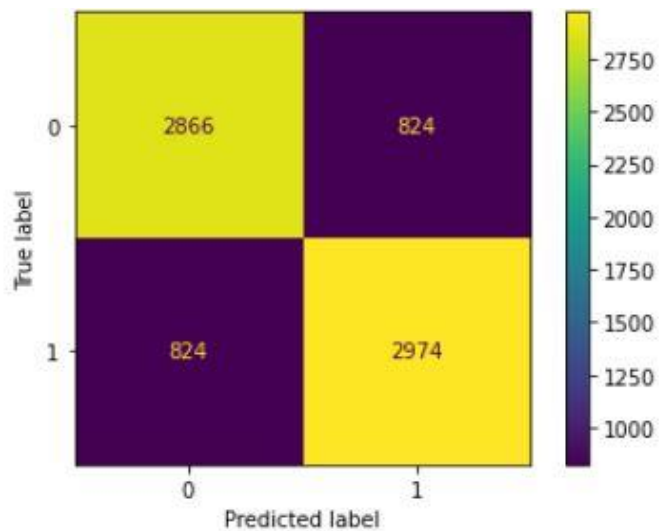
The best parameters for model support vector machine :

C': 1, 'gamma': 0.001, 'kernel': 'rbf'
Best score 0.71

5- The classification report

	Precision	recall	f1-score	support
0	0.78	0.78	0.78	3690
1	0.78	0.78	0.78	3798
Accuracy			0.78	7488
Macro avg	0.78	0.78	0.78	7488
Weighted avg	0.78	0.78	0.78	7488

6- Confusion matrix



6- Roc curve

