



**Cairo University**  
**Faculty of Computers and**  
**Artificial Intelligence**



**AI322 – Supervised learning**

**Under the supervision of Dr. Khaled Mostafa**

# ***Assignment 3***

<i>Name</i>	<i>ID</i>
<b>Marwan Mohamed Abd El-monem</b>	<b>20190513</b>
<b><i>Youssef Hesham Mohamed</i></b>	<b><i>20190648</i></b>

# REPORT FOR MNIST DATASET USING CNN

It's required to implement a machine learning algorithm for classification of the handwritten digits (MNIST) using CNN

## preprocessing phase:

60,000 samples are used from the dataset for training and 10,000 for testing and since its saved in a 1-dimentional array of size 784 so we will reshape it in form of 2-dimentional array with size of 28\*28 to be easily manipulated and visualized, we label it by creating a one-hot vector which contains 10 values (0-9 classes) where only one index is valued 1 which is the correct class

## processing phase:

### ANALYZING MODEL\_1: (CHANGING EPOCHS)

- Final accuracy: 96.46% and first 5 epoch (74.46%, 92.06%, 93.69%, 94.51%, 95.09%)
- Total parameters: 175,946
- epochs=10
- batch size=32
- learning rate=0.01
- The average time to train: 10~11s
- The average time to test :2s
- loss = 0.1162
- In this model we use 2 CNN layers which has 16 kernel in first layer , 32 in second and 2 FC layers with activation function (relu) in all layers and (softmax) for last layer
- SGD optimizer

---

#### ANALYZING MODEL\_2:

- Final accuracy: 96.86% and first 5 epoch (73.99%, 91.98%, 93.59%, 94.43%, 94.99%)
- Total parameters: 175,946
- epochs=12
- batch size=32
- learning rate=0.01
- The average time to train: 10~11s
- The average time to test :2s
- loss = 0.1053
- In this model we use 2 CNN layers (kernel same as previous) and 2 FC layers with activation function (relu) in all layers and (softmax) for last layer
- SGD optimizer

---

#### ANALYZING MODEL\_3:

- Final accuracy: 97.44% and first 5 epoch (70.12%, 92.12%, 93.75%, 94.66%, 95.12%)
- Total parameters: 175,946
- epochs=15
- batch size=32
- learning rate=0.01
- The average time to train: 10~11s
- The average time to test :2s
- loss = 0.0813

- In this model we use 2 CNN layers (kernel same as previous) and 2 FC layers with activation function (relu) in all layers and (softmax) for last layer
  - SGD optimizer
  - As we see in first three models (1,2,3) with changing the number of epochs it effects on the accuracy of the model and on model loss, so we will use 15 epochs in coming models.
- 

#### ANALYZING MODEL\_4: (CHANGING LEARNING RATE)

- Final accuracy: 98.15% and first 5 epoch (87.36%, 94.82%, 96.69%, 97.50%, 98.02%)
  - Total parameters: 175,946
  - epochs=15
  - batch size=32
  - learning rate=0.05
  - The average time to train: 10~11s
  - The average time to test :2s
  - loss = 0.0800
  - In this model we use 2 CNN layers (kernel same as previous) and 2 FC layers with activation function (relu) in all layers and (softmax) for last layer
  - SGD optimizer
- 

#### ANALYZING MODEL\_5:

- Final accuracy: 98.14% and first 5 epoch (88.20%, 95.36%, 97.13%, 98.44%, 98.77%)
- Total parameters: 175,946
- epochs=15
- batch size=32
- learning rate= 0.08

- The average time to train: 10~11s
  - The average time to test :2s
  - loss = 0.0787
  - In this model we use 2 CNN layers (kernel same as previous) and 2 FC layers with activation function (relu) in all layers and (softmax) for last layer
  - SGD optimizer
- 

#### ANALYZING MODEL\_6:

- Final accuracy: 98.17% and first 5 epoch (82.42%, 93.52%, 94.85%, 95.79%, 96.50%)
  - Total parameters: 175,946
  - epochs=15
  - batch size=32
  - learning rate= 0.03
  - The average time to train: 10~11s
  - The average time to test :2s
  - loss = 0.0691
  - In this model we use 2 CNN layers (kernel same as previous) and 2 FC layers with activation function (relu) in all layers and (softmax) for last layer
  - SGD optimizer
  - As we see in three models (4,5,6) with changing the learning rate it effects on the accuracy of the model and on model loss, so we will use 0.03 in coming models.
- 

#### ANALYZING MODEL\_7: (ADDING/REMOVING LAYERS)

- Final accuracy: 97.42% and first 5 epoch (97.33%, 97.51%, 97.61%, 97.78%, 97.87%)
- Total param s: 43,530

- epochs=15
  - batch size=32
  - learning rate= 0.03
  - The average time to train: 8s
  - The average time to test: 1s
  - loss = 0.0906
  - In this model we use 1 CNN layer which has 16 kernel and 1 FC layers with activation function (relu) in all layers and (softmax) for last layer
  - SGD optimizer
- 

#### ANALYZING MODEL\_8:

- Final accuracy: 97.34% and first 5 epoch (82.70%, 92.52%, 94.25%, 95.21%, 95.84%)
  - Total parameters: 173,946
  - epochs=15
  - batch size=32
  - learning rate= 0.03
  - The average time to train: 7~8s
  - The average time to test :1s
  - loss = 0.0871
  - In this model we use 1 CNN layer which has 16 kernel and 1 FC layers with activation function (relu) in all layers and (softmax) for last layer
  - SGD optimizer
- 

#### ANALYZING MODEL\_9:

- Final accuracy: 98.13% and first 5 epoch (85.61%, 93.83%, 95.14%, 96.03%, 96.72%)

- Total parameters: 175,578
  - epochs=15
  - batch size=32
  - learning rate= 0.03
  - The average time to train: 10~11s
  - The average time to test: 2s
  - loss = 0.0582
  - In this model we use 2 CNN layers which has 16 kernel in first layer , 32 in second and 1 FC layers with activation function (relu) in all layers and (softmax) for last layer
  - SGD optimizer
- 

#### ANALYZING MODEL\_10:

- Final accuracy: 98.45% and first 5 epoch (84.02%, 96.14 %, 97.33%, 97.96%, 98.28%)
- Total parameters: 160,346
- epochs=15
- batch size=32
- learning rate= 0.03
- The average time to train: 13~14s
- The average time to test :2s
- loss = 0.0534
- In this model we use 3 CNN layers which has 16 kernel in first ,32 in second 64 in third and 2 FC layers with activation function (relu) in all layers and (softmax) for last layer
- SGD optimizer

---

#### ANALYZING MODEL\_11:

- Final accuracy: 98.49% and first 5 epoch (82.07%, 96.03%, 97.16%, 97.72%, 98.09%)
- Total parameters: 160,714
- epochs=15
- batch size=32
- learning rate= 0.03
- The average time to train: 14s
- The average time to test :2s
- loss = 0.0572
- In this model we use 3 CNN layers (kernel same as previous) and 3 FC layers with activation function (relu) in all layers and (softmax) for last layer
- SGD optimizer
- As we see in five models (7,8,9,10,11) with changing the numbers of layers it not effects on the accuracy of the model, model loss only it also effects on the number of parameters which refers to the amount of memory we use, so we will use 3 CNN layers and 2 FC layers in coming models.

---

#### ANALYZING MODEL\_12: (BATCH SIZE)

- Final accuracy: 98.79% and first 5 epoch (87.87%, 97.00 %, 97.87%, 98.32%, 98.68%)
- Total parameters: 160,346
- epochs=15
- batch size=20
- learning rate= 0.03



- The average time to train 18~19s
  - The average time to test :2s
  - loss = 0.0620
  - In this model we use 3 CNN layers (kernel same as previous) and 2 FC layers with activation function (relu) in all layers and (softmax) for last layer
  - SGD optimizer
- 

#### ANALYZING MODEL\_13:

- Final accuracy: 97.44% and first 5 epoch (75.51%, 94.57%, 96.53%, 97.25%, 97.70%)
  - Total parameters: 160,346
  - epochs=15
  - batch size=64
  - learning rate= 0.03
  - The average time to train 8~9s
  - The average time to test :2s
  - loss = 0.0824
  - In this model we use 3 CNN layers (kernel same as previous) and 2 FC layers with activation function (relu) in all layers and (softmax) for last layer
  - SGD optimizer
  - As we see in tow models (12,13) with changing the batch size it not effects on the accuracy of the model, model loss only it also effects on the average time of training but we chose to focus on the accuracy, so we will use 20 batch size in coming models.
- 

#### ANALYZING MODEL\_14: (ACTIVATION FUNCTION)

- Final accuracy: 11.35% and first 5 epoch (10.59%, 10.53%, 10.74%, 10.90%, 10.87%)
  - Total parameters: 160,346
  - epochs=15
  - batch size=20
  - learning rate= 0.03
  - The average time to train 17~18s
  - The average time to test :2s
  - loss = 2.3012
  - In this model we use 3 CNN layers (kernel same as previous) and 2 FC layers with activation function (sigmoid) in all layers and (softmax) for last layer
  - SGD optimizer
- 

#### ANALYZING MODEL\_15:

- Final accuracy: 98.86% and first 5 epoch (90.60%, 97.20%, 98.21%, 98.66%, 98.95%)
- Total parameters: 160,346
- epochs=15
- batch size=20
- learning rate= 0.03
- The average time to train: 17~18s
- The average time to test :2s
- loss = 0.0379
- In this model we use 3 CNN layers (kernel same as previous) and 2 FC layers with activation function (tanh) in all layers and (softmax) for last layer

- SGD optimizer
- 

#### ANALYZING MODEL\_16:

- Final accuracy: 98.85% and first 5 epoch (87.32%, 95.65%, 97%, 97.72%, 98.21%)
  - Total parameters: 160,346
  - epochs=15
  - batch size=20
  - learning rate= 0.03
  - The average time to train 18~19s
  - The average time to test :2s
  - loss = 0.0381
  - In this model we use 3 CNN layers (kernel same as previous) and 2 FC layers with activation function (softsign) in all layers and (softmax) for last layer
  - SGD optimizer
  
  - As we see in three models (14,15,16) with changing the activation function it effects on the accuracy of the model, model loss, so we will use tanh function in coming models.
- 

#### ANALYZING MODEL\_17: (OPTIMIZERS)

- Final accuracy: 98.51% and first 5 epoch (95.56%, 98.19%, 98.57%, 98.82%, 98.99%)
- Total parameters: 160,346
- epochs=15
- batch size=20

- The average time to train 19~20s
  - The average time to test :2s
  - loss = 0.0553
  - In this model we use 3 CNN layers (kernel same as previous) and 2 FC layers with activation function (tanh) in all layers and (softmax) for last layer
  - Adam optimizer
- 

#### ANALYZING MODEL\_18:

- Final accuracy: 98.76% and first 5 epoch (95.74%, 98.32%, 98.83%, 99.15%, 99.30%)
  - Total parameters: 160,346
  - epochs=15
  - batch size=20
  - learning rate= 0.03
  - The average time to train: 23s
  - The average time to test :2s
  - loss = 0.0796
  - In this model we use 3 CNN layers (kernel same as previous) and 2 FC layers with activation function (tanh) in all layers and (softmax) for last layer
  - RMSprop optimizer
- 
- As we see in tow models (17,18) with changing the optimizer it not effects on the accuracy of the model, model loss only it also effects on the average time of training, so we will use SGD optimizer (model 15) in coming models.
-

#### ANALYZING MODEL\_19: (DROPOUT LOCATION)

- Final accuracy: 98.66% and first 5 epoch (87.92%, 95.29%, 96.61%, 97.12%, 97.64%)
  - Total parameters: 160,346
  - epochs=15
  - batch size=20
  - learning rate= 0.03
  - The average time to train: 19s
  - The average time to test :2s
  - loss = 0.0444
  - In this model we use 3 CNN layers (kernel same as previous) and 2 FC layers with activation function (tanh) in all layers and (softmax) for last layer
  - SGD optimizer
  - We put the drop out after the first FC layer
  - Dropout rate: 50%
- 

#### ANALYZING MODEL\_20:

- Final accuracy: 98.83% and first 5 epoch (88.26%, 96.09%, 97.34%, 97.91%, 98.30%)
- Total parameters: 160,346
- epochs=15
- batch size=20
- learning rate= 0.03
- The average time to train: 19s
- The average time to test :2s

- loss = 0.0433
  - In this model we use 3 CNN layers (kernel same as previous) and 2 FC layers with activation function (tanh) in all layers and (softmax) for last layer
  - SGD optimizer
  - We put the drop out after the second FC layer
  - Dropout rate: 50%
  - As we see in tow models (19,20) with changing the Dropout location it effects on the accuracy of the model, model loss, so we will use after second layer in coming models.
- 

#### ANALYZING MODEL\_21: (DROPOUT RATE)

- Final accuracy: 98.98% and first 5 epoch (98.10%, 98.60%, 98.76%, 99.02%, 99.15%)
  - Total parameters: 160,346
  - epochs=15
  - batch size=20
  - learning rate= 0.03
  - The average time to train: 19s
  - The average time to test :2s
  - loss = 0.0349
  - In this model we use 3 CNN layers (kernel same as previous) and 2 FC layers with activation function (tanh) in all layers and (softmax) for last layer
  - SGD optimizer
  - We put the drop out after the second FC layer
  - Dropout rate: 25%
-

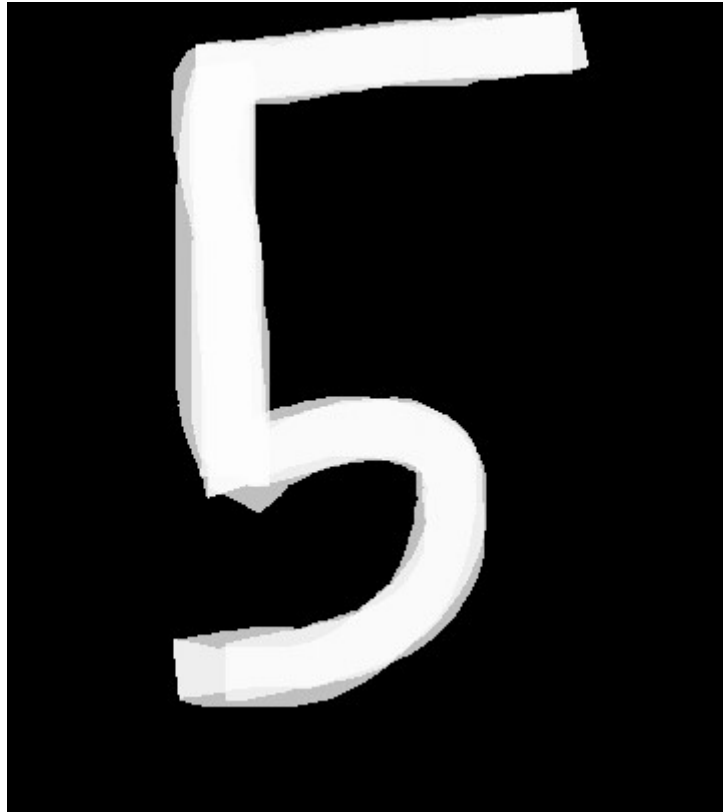
#### ANALYZING MODEL\_22:

- Final accuracy: 98.59% and first 5 epoch (84.04%, 93.62%, 95.54%, 96.11%, 96.54%)
- Total parameters: 160,346
- epochs=15
- batch size=20
- learning rate= 0.03
- The average time to train: 19s
- The average time to test :2s
- loss =0.0648
- In this model we use 3 CNN layers (kernel same as previous) and 2 FC layers with activation function (tanh) in all layers and (softmax) for last layer
- SGD optimizer
- We put the drop out after the second FC layer
- Dropout rate: 70%
- As we see in two models (21,22) with changing the Dropout rate it effects on the accuracy of the model, model loss, so we will use 25%.

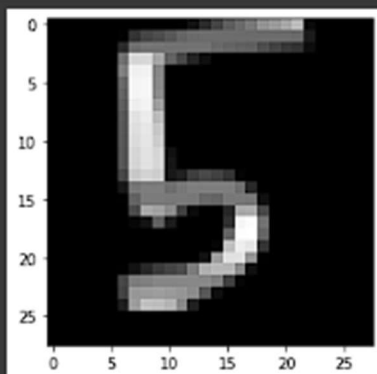
#### CONCLUSION

SO, WE CAN DEDUCE THAT MODEL 21 IS THE BEST MODEL IN ASPECTS OF ACCURACY AND LOSS , WHERE THE ACCURACY IS CHANGEABLE DUE TO SHUFFLING THE DATA

- WE USED A USER INPUT DATA USING 'PAINT' TO TEST THE MODEL WHERE THE PROGRAMME CHOSE THE BEST MODEL TO PROCESS UPON THIS PHOTO



```
[ ] test_img = tf.keras.utils.normalize(test_img,axis = 1)
    plot_img(test_img)
```



```
[ ] test_img = reshape_img(np.array(test_img))
    test_img.shape
```

```
(1, 28, 28, 1)
```

```
[ ] p = model_21.predict(test_img)
    print(np.argmax(p))
```

```
5
```