Sign Language Classification

Hand Gesture Recognition Task on MNIST images

Michele Lotto 875922

Ca'Foscari University - DAIS

January 6, 2024

Ca' Foscari
University
of Venice

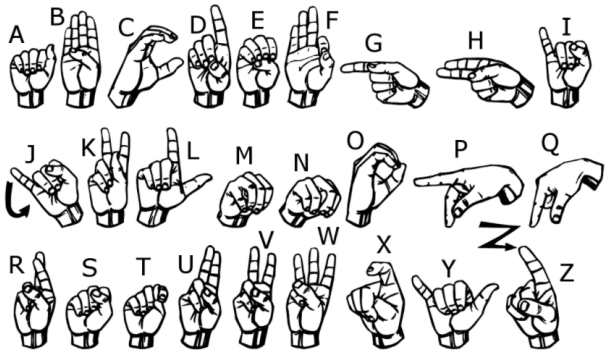
Contents

1. Introduction
2. MNIST image dataset
3. Base architectures
   - LeNet5
   - Classifier 2
   - Classifier 3
4. Training procedure
   - Data augmentation
5. Results
   - LeNet5 results
   - Classifier 2 results
   - Classifier 3 results
6. References

Introduction

Project Goal

Create a robust classifier to recognize **American Sign Language** hand poses.



MNIST image dataset

The dataset

28x28 pixels images of hand poses:

- 24 categories: full English alphabet excluding J and Z which require motion.
- 27455 training images:
  - 80% actual training.
  - 20% validation.
- 7172 test images.

Contents

1. Introduction
2. MNIST image dataset
3. Base architectures
   - LeNet5
   - Classifier 2
   - Classifier 3
4. Training procedure
   - Data augmentation
5. Results
   - LeNet5 results
   - Classifier 2 results
   - Classifier 3 results
6. References

Base architectures

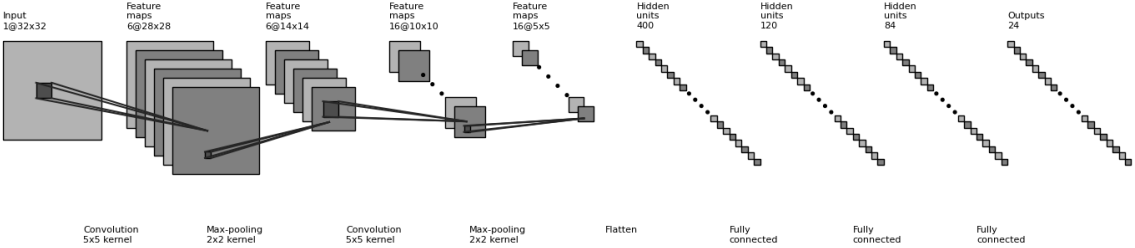
- "LeNet5" (1 architecture): LeNet5 architecture from original paper [1].
- "Classifier 2" (12 architectures): CNN with 2 convolutional layers.
- "Classifier 3" (24 architectures): CNN with 3 convolutional layers.

"Classifier 2" and "Classifier 3" architectures generated by varying:

- dropout layer positions;
- number of neurons in hidden layer.

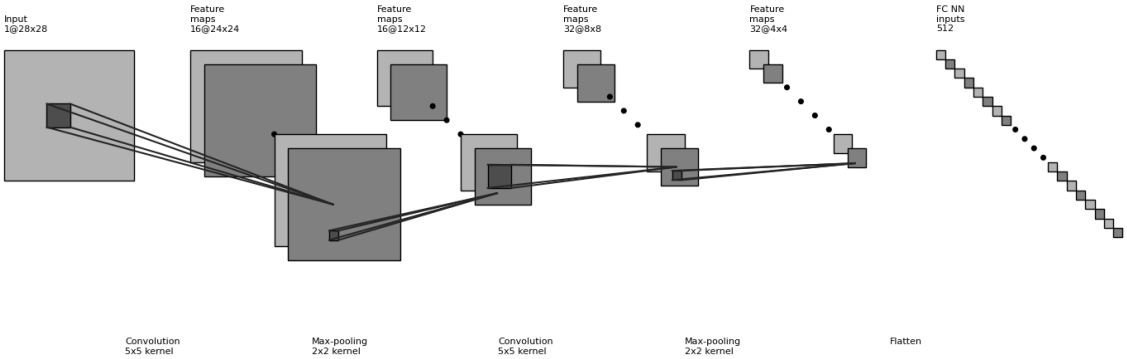
TOTAL: 37 architectures

LeNet5 architecture



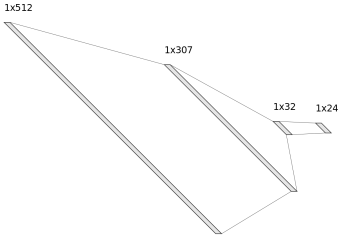
- 28x28 input image pre-transformed into 32x32 by zero padding.
- output layer modified from 10 units to 24 units.

Classifier 2 architecture

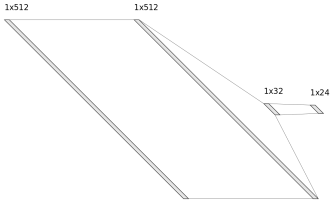


Dropout after second Max-pooling layer with probability 0.5 or 0.

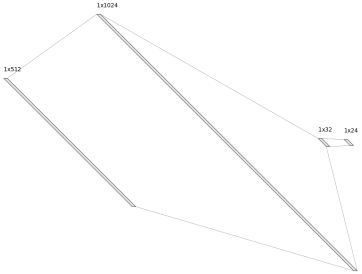
Classifier 2 Fully Connected layers



No. of Hidden Unit = 0.6 x inputs



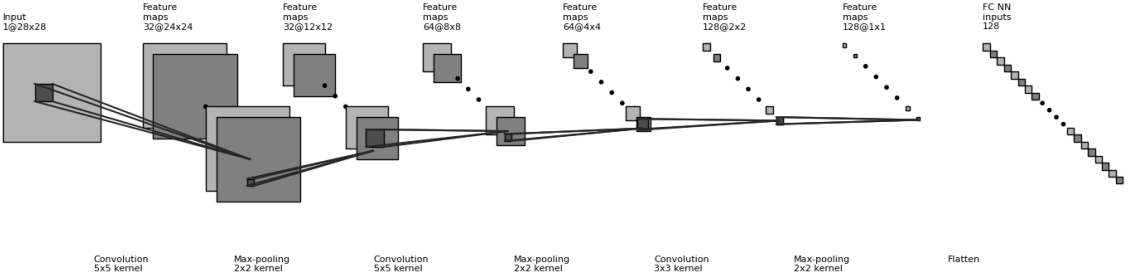
No. of Hidden Unit = inputs



No. of Hidden Unit = 2 x inputs

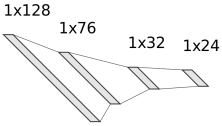
Dropout after first hidden layer with probability 0.5 or 0.

Classifier 3 architecture

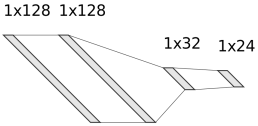


Dropout after second and third Max-pooling layers with probability 0.5 or 0.

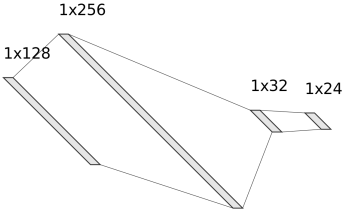
Classifier 3 Fully Connected layers



No. of Hidden Unit = 0.6 * inputs



No. of Hidden Unit = inputs



No. of Hidden Unit = 2 * inputs

Dropout after first hidden layer with probability 0.5 or 0.

Contents

1. Introduction
2. MNIST image dataset
3. Base architectures
   - LeNet5
   - Classifier 2
   - Classifier 3
4. Training procedure
   - Data augmentation
5. Results
   - LeNet5 results
   - Classifier 2 results
   - Classifier 3 results
6. References

Training Procedure and Hyperparameters tuning

All 37 architectures were trained using:

- 100 max epochs: all models converge to local optima in a few epochs;
- early stopping procedure (to prevent overfitting)
- two different optimizers: **ADAM** and **AMSGrad**.

Hyperparameters tuned for all 37 architectures:

- learning rate: [0.0005, 0.0001, 0.00001, 0.000001]
- batch size: [32, 64, 128, 256, 512]
- patience (early stopping): [5, 10, 15, 20]
- data augmentation percentage (on training data): [0, 0.25, 0.5, 0.75]

Training Procedure (details)

For each epoch:

1. train the model.
2. evaluate the model on validation data (calculate evaluation loss).
3. test the model on test data (only for recording reasons).
4. record:
   - the current epoch.
   - the current training and evaluation losses.
   - the training time in seconds.
   - the current test accuracy.
5. if the evaluation loss starts to increase for some epochs (patience), stop the training process.

Data augmentation

1. sample a certain percentage of images from training data;
2. apply random perspective with a distortion scale of 0.5 to half of these images;

3. apply random rotation with a maximum angle of 45 degrees to the other half;

4. add transformed images to training data.

Contents

1. Introduction
2. MNIST image dataset
3. Base architectures
   - LeNet5
   - Classifier 2
   - Classifier 3
4. Training procedure
   - Data augmentation
5. Results
   - LeNet5 results
   - Classifier 2 results
   - Classifier 3 results
6. References

Results structure

For each 'base architecture' the following points are covered:

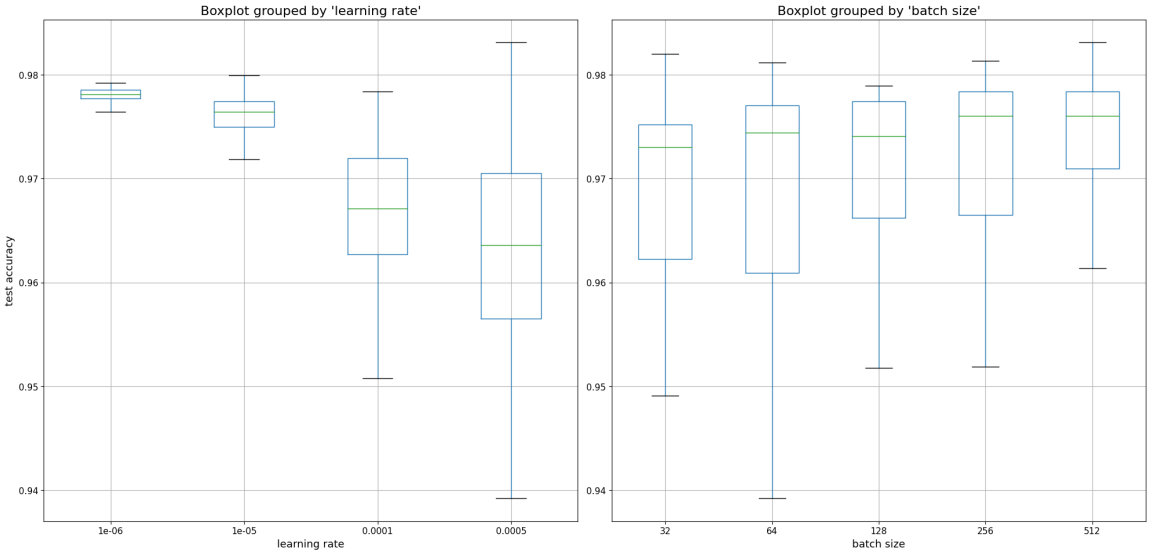
- architectures analysis
- hyperparameters analysis
- best model
- worst model

LeNet5 results: architectures analysis

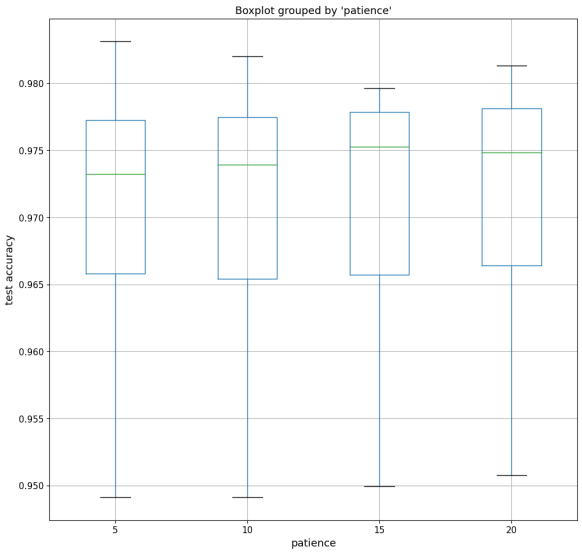
Single architecture tested corresponding to the original LeNet5 architecture. Follows an initial analysis on all the test accuracy results:

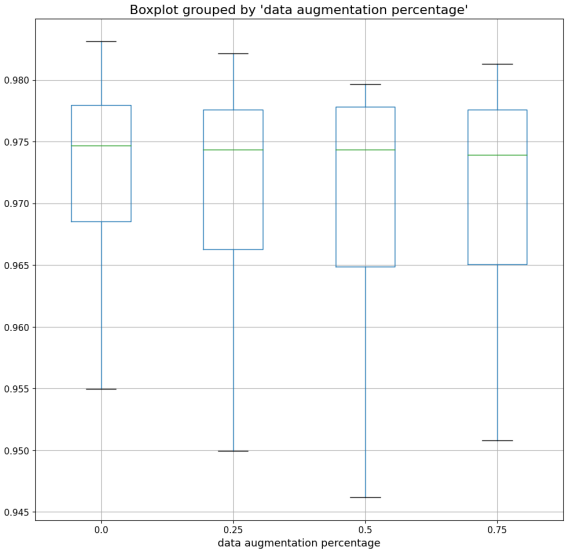
- Mean accuracy of 0.9706 with SD of 0.0096
- Best accuracy: 0.9831
- Worst accuracy: 0.9035

LeNet5 results: hyperparameters analysis I

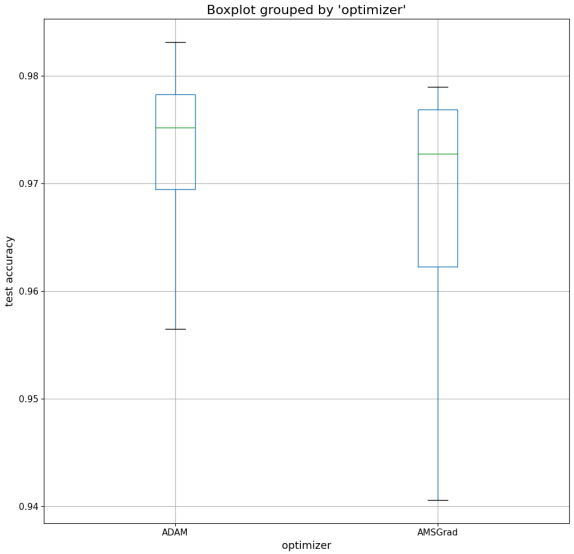


LeNet5 results: hyperparameters analysis II



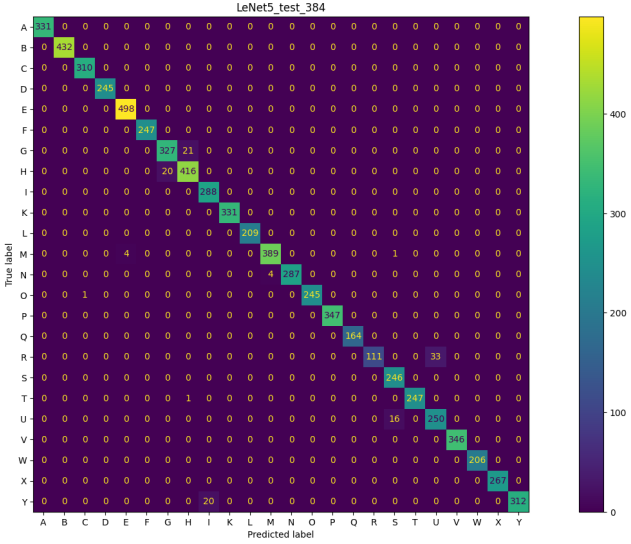


LeNet5 results: hyperparameters analysis III

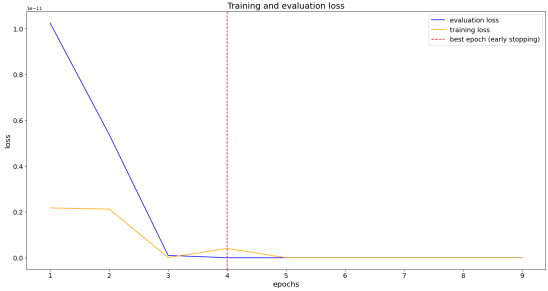


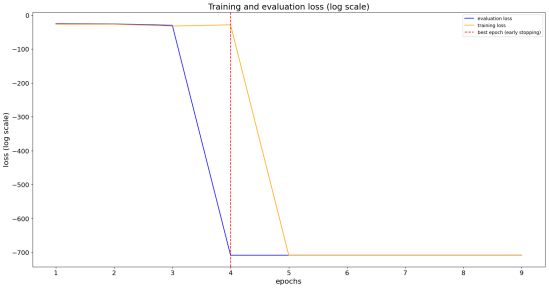
LeNet5 best model

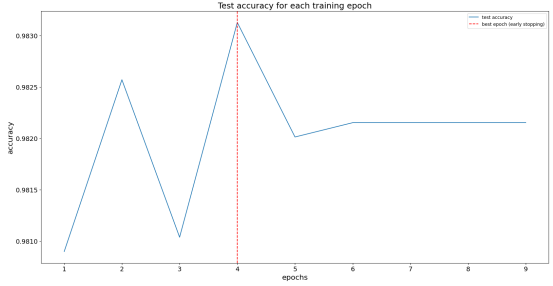
- Hyperparameters:
  - optimizer: ADAM
  - learning rate: 0.0005
  - batch size: 512
  - patience: 5
  - data augmentation percentage: 0
- Test accuracy: 0.9831
- Training time: 1.69s
- Test time: 0.03s



LeNet5 best model: training loss and accuracy

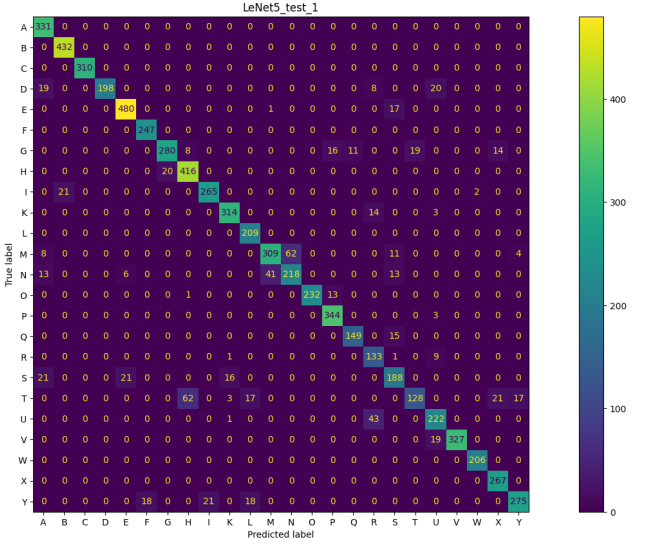




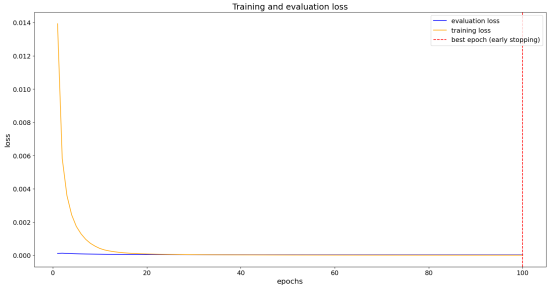


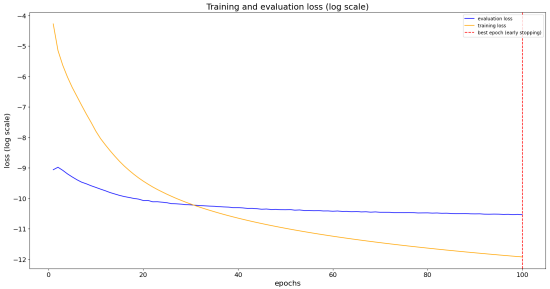
LeNet5 worst model

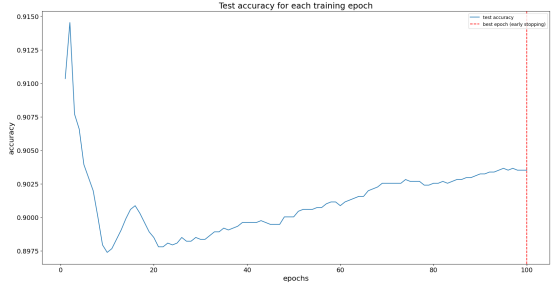
- Hyperparameters:
  - optimizer: AMSGrad
  - learning rate: 0.0005
  - batch size: 32
  - patience: 5
  - data augmentation percentage: 0.25
- Test accuracy: 0.9035
- Training time: 64.38s
- Test time: 0.06s



LeNet5 worst model: training loss and accuracy







Classifier 2 results: architectures analysis I

12 architectures tested. Follows an initial analysis:

- 'local minima' architecture (lowest test accuracy for all its models):
  Architectural features:
  - hidd. neurons molt. factor: 1.0,
  - dropout after: []

  Test accuracy results of all its models: 0.0201
- 'best' architectures (highest test accuracy on average):
  Architectural features:
  - hidd. neurons molt. factor: 1.0,
  - dropout after: ['Conv2', 'FC1']

  Test accuracy results:
  - Mean accuracy of 0.9947 with SD of 0.0045
  - Best accuracy: 0.9997
  - Worst accuracy: 0.9635

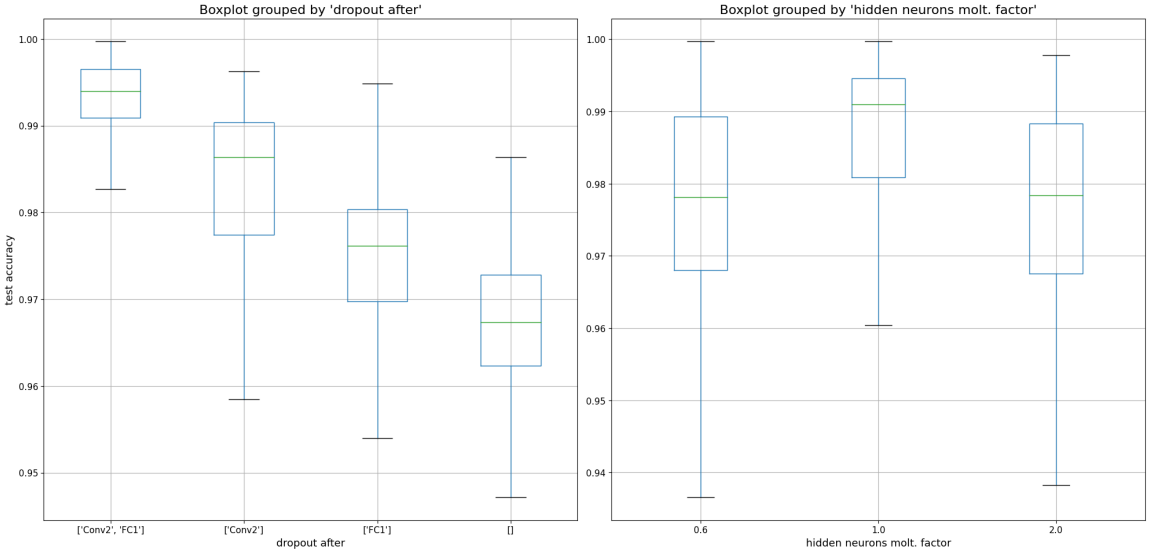
Classifier 2 results: architectures analysis II

- 'worst' architecture (lowest test accuracy on average):
  Architectural features:
  - hidd. neurons molt. factor: 0.6,
  - dropout after: ['Conv2']

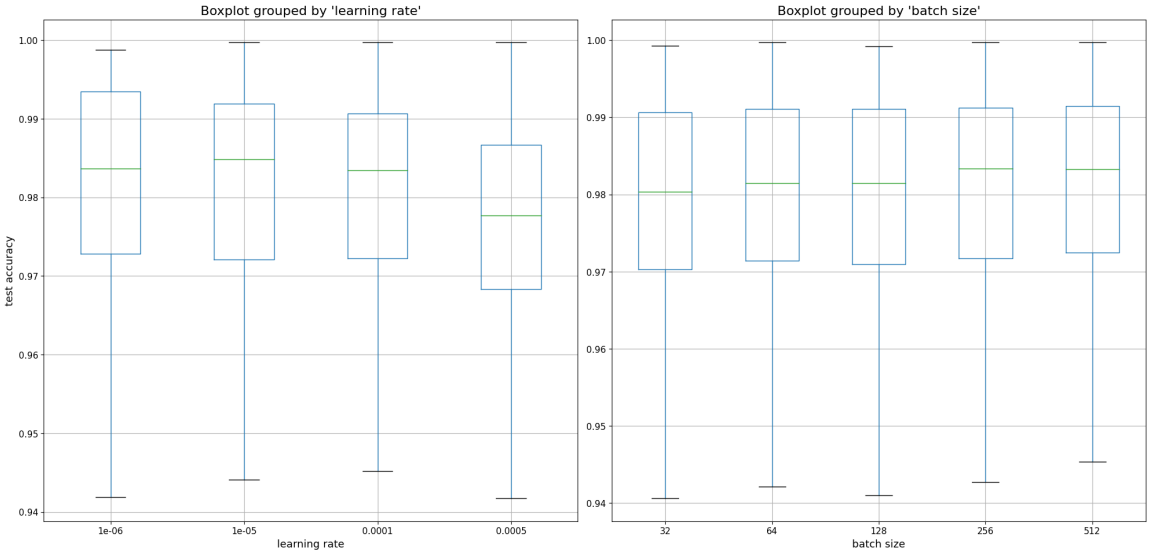
  Test accuracy results:
  - Mean accuracy of 0.9249 with SD of 0.0750
  - Best accuracy: 0.9927
  - Worst accuracy: 0.7460

Following analysis done by excluding the 'local minima' architecture.

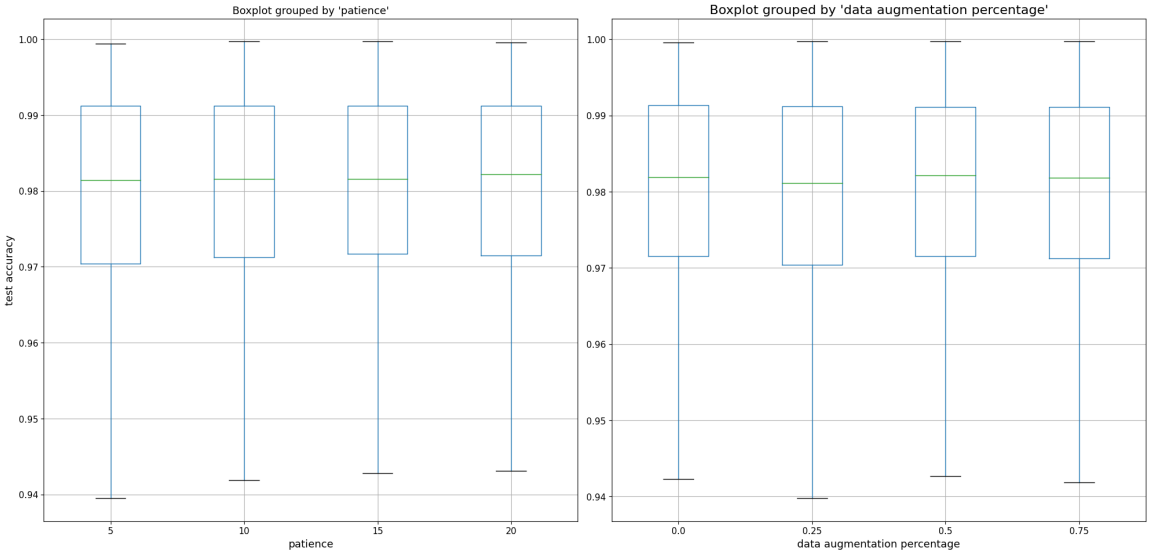
Classifier 2 results: architectures analysis III



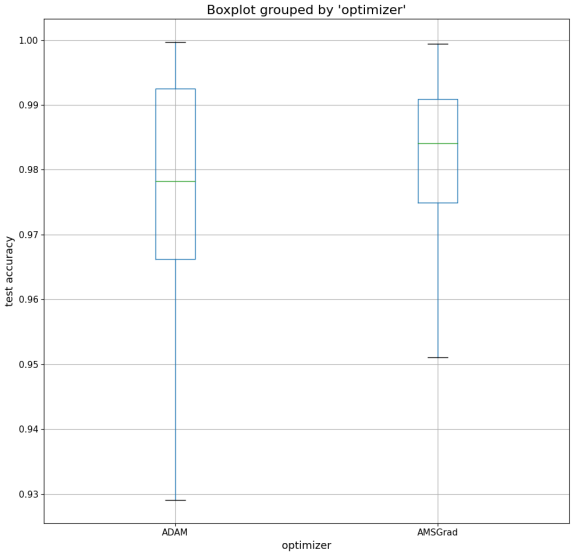
Classifier 2 results: hyperparameters analysis I



Classifier 2 results: hyperparameters analysis II

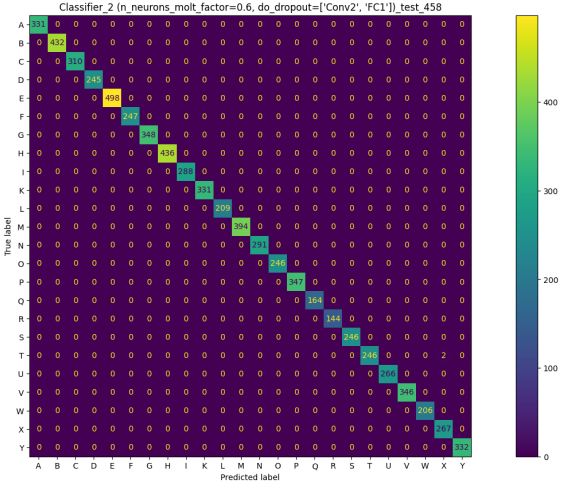


Classifier 2 results: hyperparameters analysis III

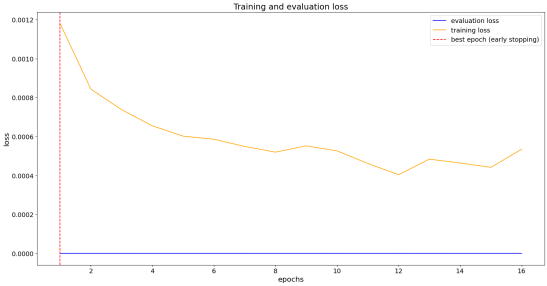


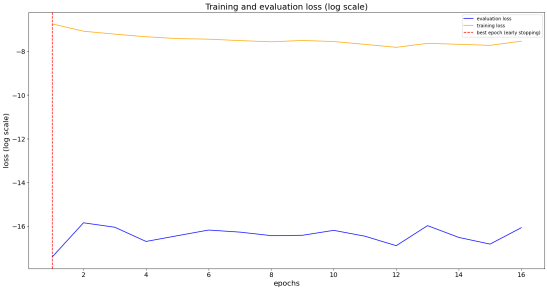
Classifier 2 best model

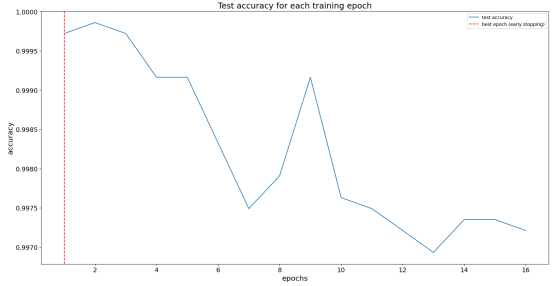
- Architectural features:
  - hidd. neurons molt. factor: 0.6
  - dropout after: ['Conv2', 'FC1']
- Hyperparameters:
  - optimizer: ADAM
  - learning rate: 0.0001
  - batch size: 256
  - patience: 15
  - data augmentation percentage: 0.5
- Test accuracy: 0.9997
- Training time: 5.64s
- Test time: 0.03s



Classifier 2 best model: training loss and accuracy

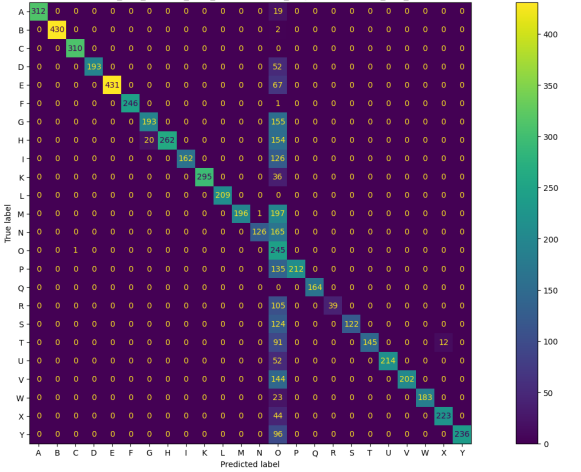




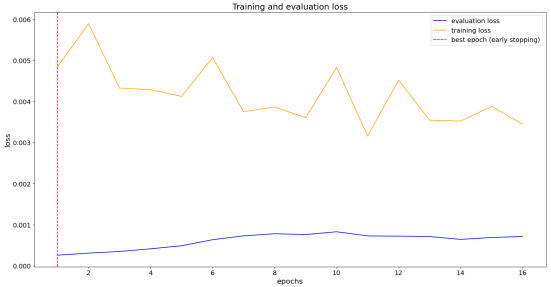


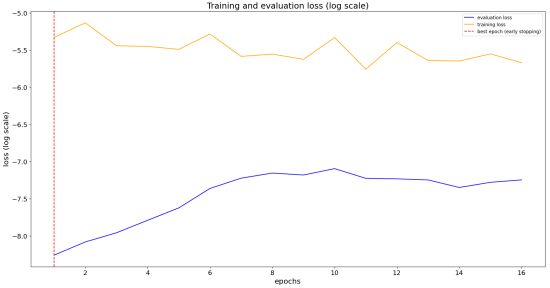
Classifier 2 worst model

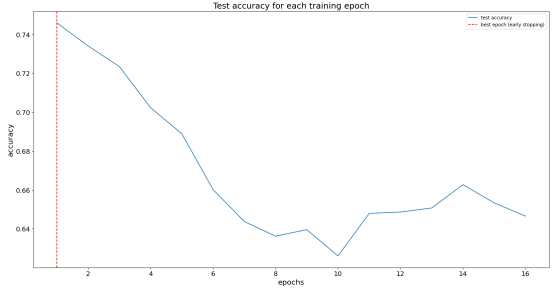
- Architectural features:
  - hidd. neurons molt. factor: 0.6
  - dropout after: ['Conv2']
- Hyperparameters:
  - optimizer: ADAM
  - learning rate: 0.00001
  - batch size: 512
  - patience: 15
  - data augmentation percentage: 0.75
- Test accuracy: 0.7459
- Training time: 6.00s
- Test time: 0.03s



Classifier 2 worst model: training loss and accuracy







Classifier 3 results: architectures analysis I

24 architectures tested. Follows an initial analysis:

- 2 'local minima' architecture (lowest test accuracy for all its models):
  1. Architectural features:
     - hidd. neurons molt. factor: 2.0,
     - dropout after: ['Conv2']

     Test accuracy results for all its models: 0.0201.
  2. Architectural features:
     - hidd. neurons molt. factor: 1.0,
     - dropout after: []

     Test accuracy results for all its models: 0.0201.
- 'best' architecture (highest test accuracy on average):
  Architectural features:
  - hidd. neurons molt. factor: 1.0,
  - dropout after: ['Conv2', 'Conv3']

Classifier 3 results: architectures analysis II

Test accuracy results:

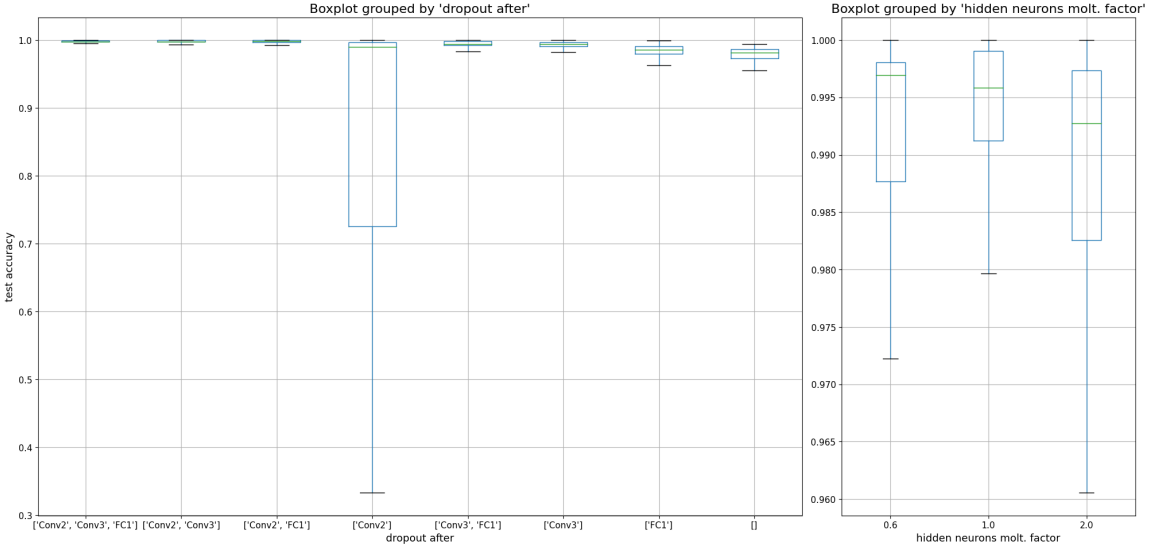
- Mean accuracy of 0.9996 with SD of 0.0011
- Best accuracy: 1.0000
- Worst accuracy: 0.9880

- 'worst' architecture (lowest test accuracy on average):
  Architectural features:
  - hidd. neurons molt. factor: 1.0,
  - dropout after: ['Conv2']

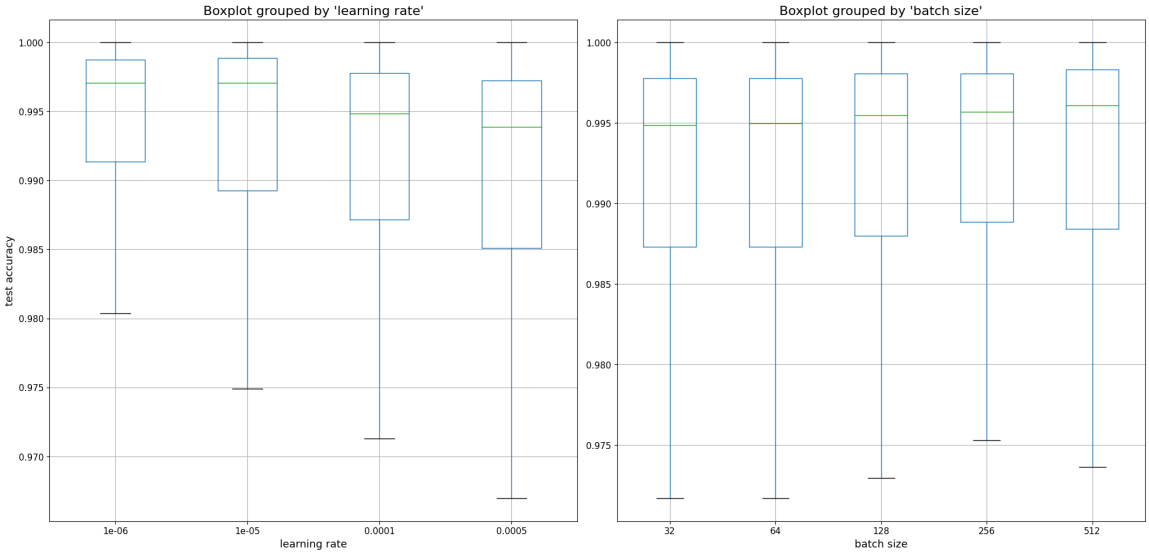
  Test accuracy results:
  - Mean accuracy of 0.6986 with SD of 0.3851
  - Best accuracy: 1.0000
  - Worst accuracy: 0.0809

Following analysis done by excluding the 'local minima' architectures.

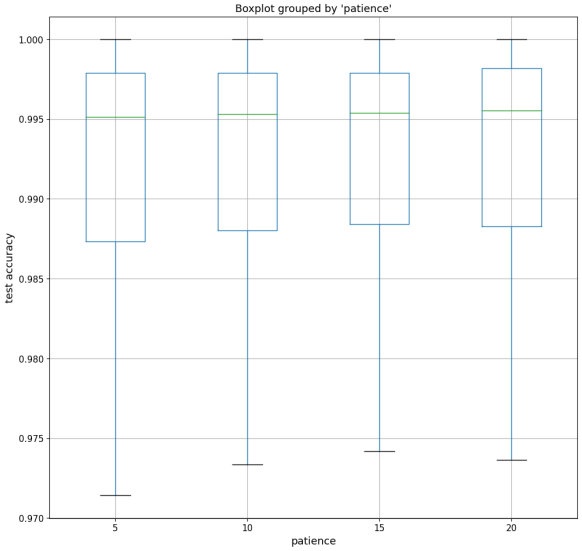
Classifier 3 results: architectures analysis III

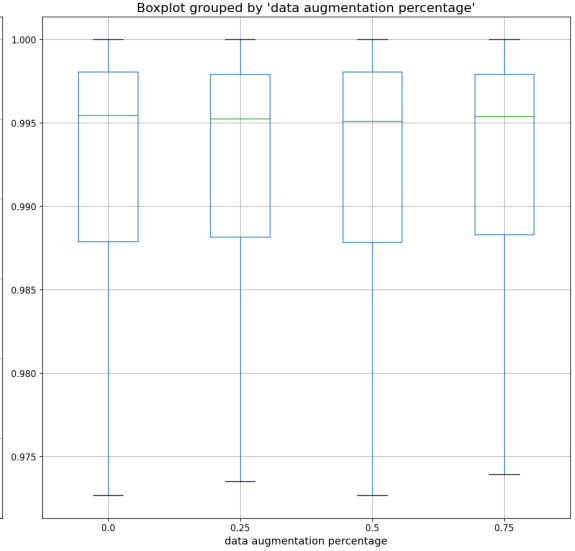


Classifier 3 results: hyperparameters analysis I

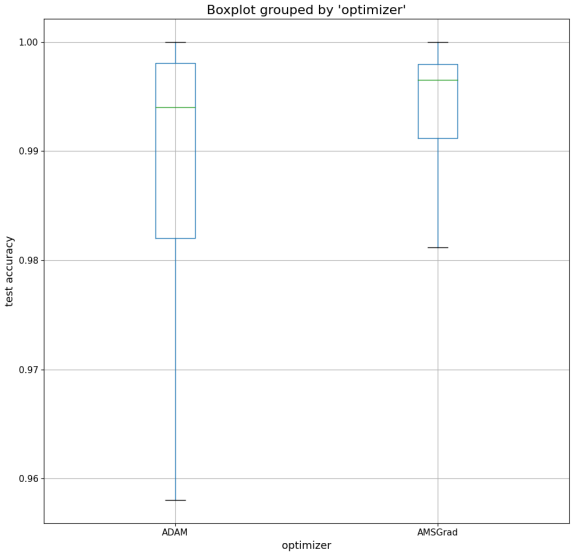


Classifier 3 results: hyperparameters analysis II



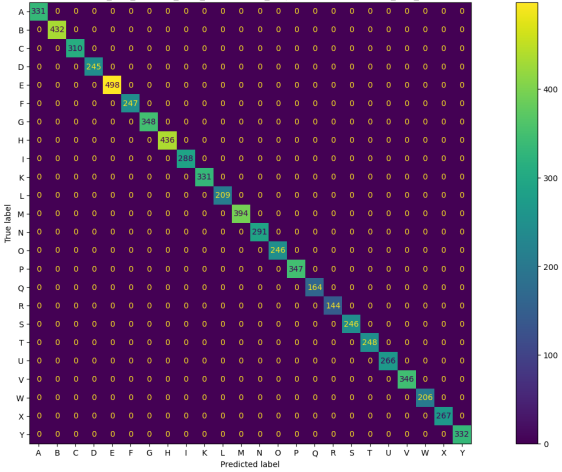


Classifier 3 results: hyperparameters analysis III

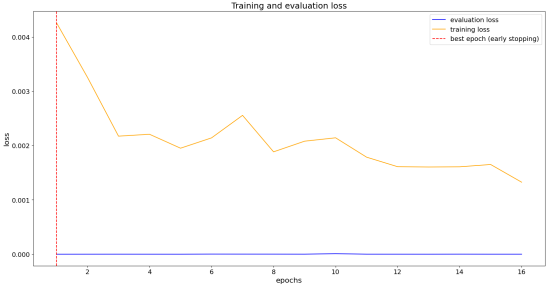


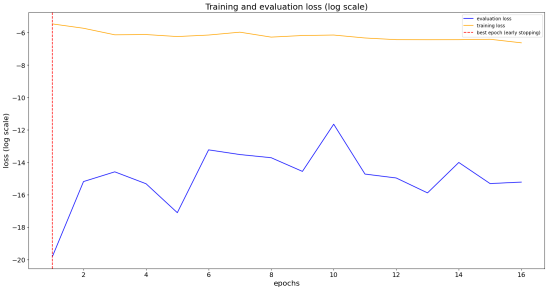
Classifier 3 best model

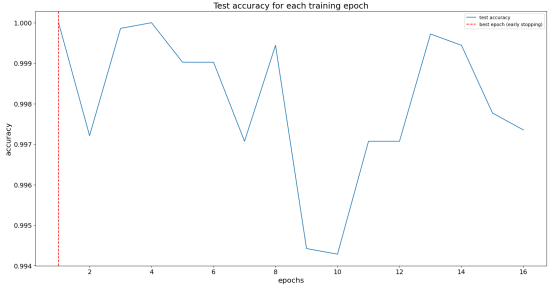
- Architectural features:
  - hidd. neurons molt. factor: 1,
  - dropout after: ['Conv2', 'FC1']
- Hyperparameters:
  - optimizer: ADAM
  - learning rate: 0.0005
  - batch size: 32
  - patience: 15
  - data augmentation percentage: 0.25
- Test accuracy: 1.0
- Training time: 14.43s
- Test time: 0.08s



Classifier 3 best model: training loss and accuracy

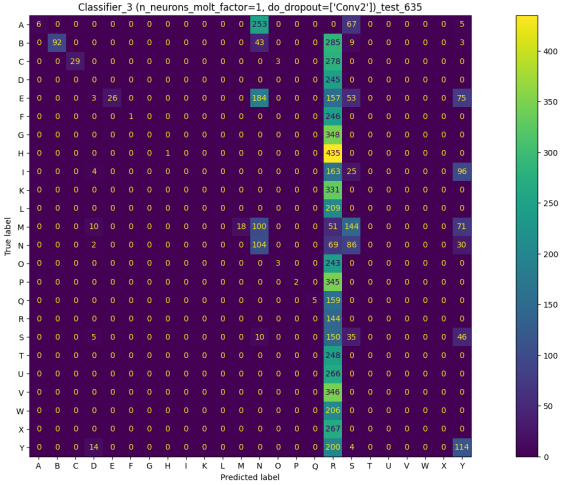




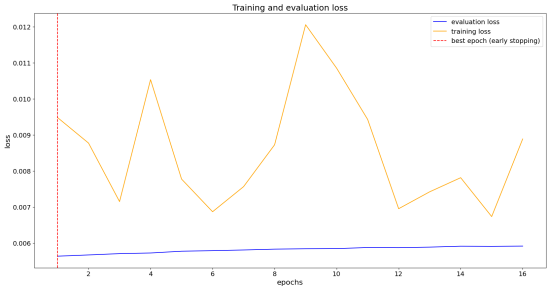


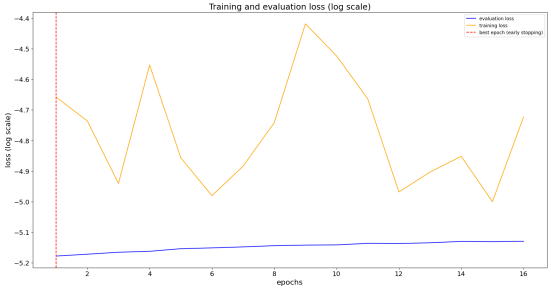
Classifier 3 worst model

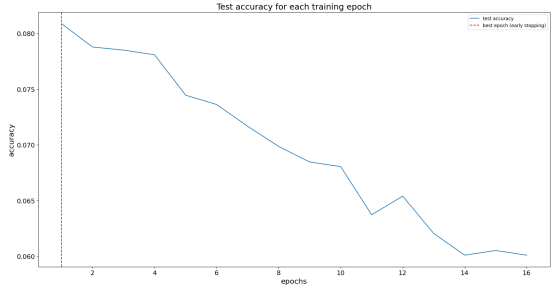
- Architectural features:
- hidd. neurons molt. factor: 1.0,
- dropout after: ['Conv2']
- Hyperparameters:
  - optimizer: ADAM
  - learning rate: 0.00001
  - batch size: 512
  - patience: 15
  - data augmentation percentage: 0.75
- Test accuracy: 0.0809
- Training time: 10.42s
- Test time: 0.04s



Classifier 3 worst model: training loss and accuracy







References

[1] Y. Lecun et al. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324. DOI: 10.1109/5.726791.