Pattern Recognition Assignment 3

Names: ID:

Manar Amr Mohammed   18011842
Nour Waled Ali   18012006
Youssef Mohamed Hussein   18015031

Data Preprocessing

Feature space

Each sample has 88 features, 44 for zero-crossing-rate and 44 for Energy (rms).
audio sampling rate is adapted to generate a fixed number of features for all samples.
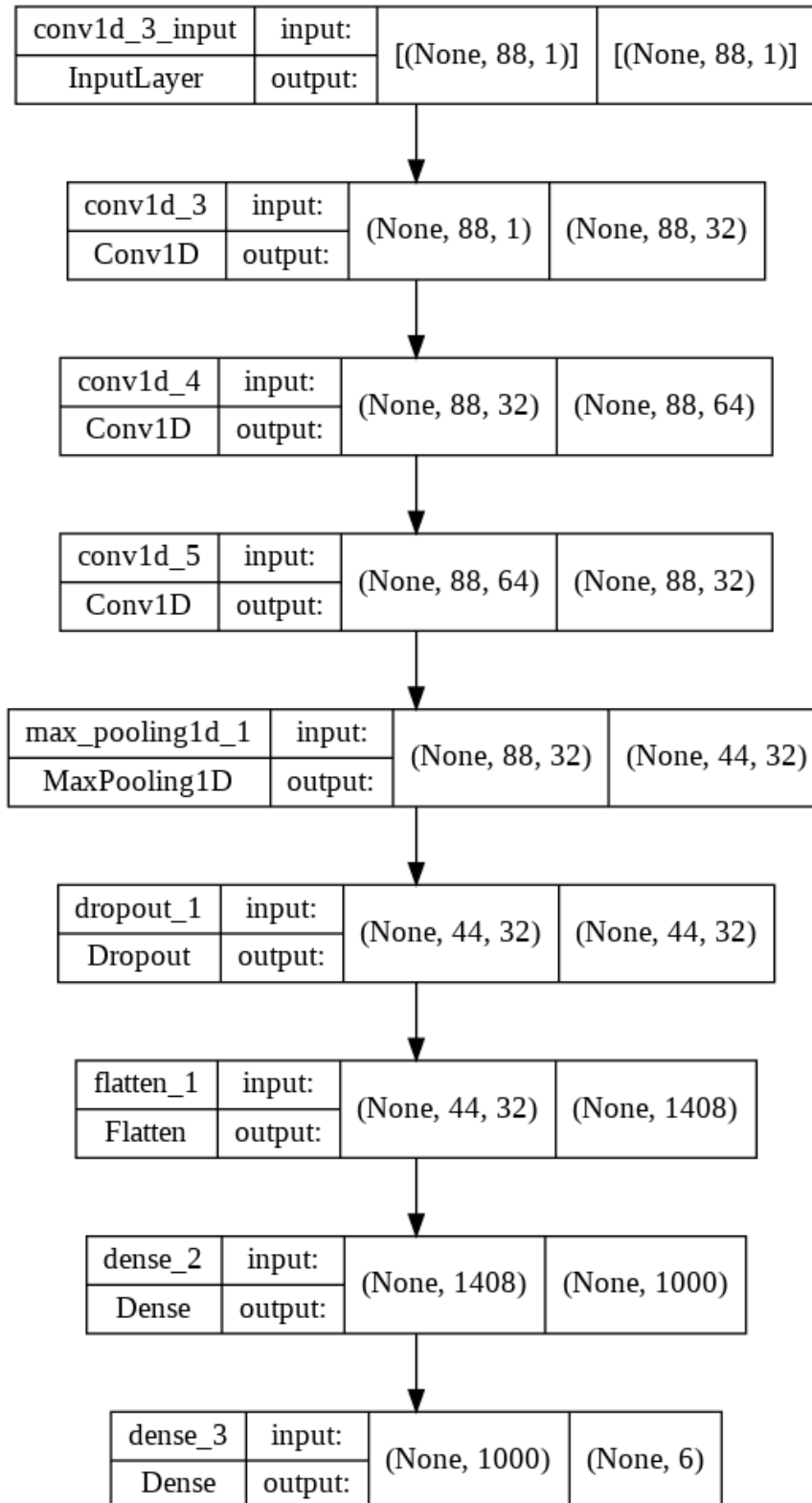
Data Augmentation

3 additional datasets (audio with noise, audio shifted to left, and audio shifted to right) are generated to increase data samples and reduce overfitting.

Original data	7442 samples
Original data + noise	7442 samples
Original data + random shift left	7442 samples
Original data + random shift right	7442 samples

Tuning and Validation

- Model 1

conv1d_3_input	input:	[(None, 88, 1)]	[(None, 88, 1)]
InputLayer	output:		

conv1d_3	input:	(None, 88, 1)	(None, 88, 32)
Conv1D	output:		

conv1d_4	input:	(None, 88, 32)	(None, 88, 64)
Conv1D	output:		

conv1d_5	input:	(None, 88, 64)	(None, 88, 32)
Conv1D	output:		

max_pooling1d_1	input:	(None, 88, 32)	(None, 44, 32)
MaxPooling1D	output:		

dropout_1	input:	(None, 44, 32)	(None, 44, 32)
Dropout	output:		

flatten_1	input:	(None, 44, 32)	(None, 1408)
Flatten	output:		

dense_2	input:	(None, 1408)	(None, 1000)
Dense	output:		

dense_3	input:	(None, 1000)	(None, 6)
Dense	output:		

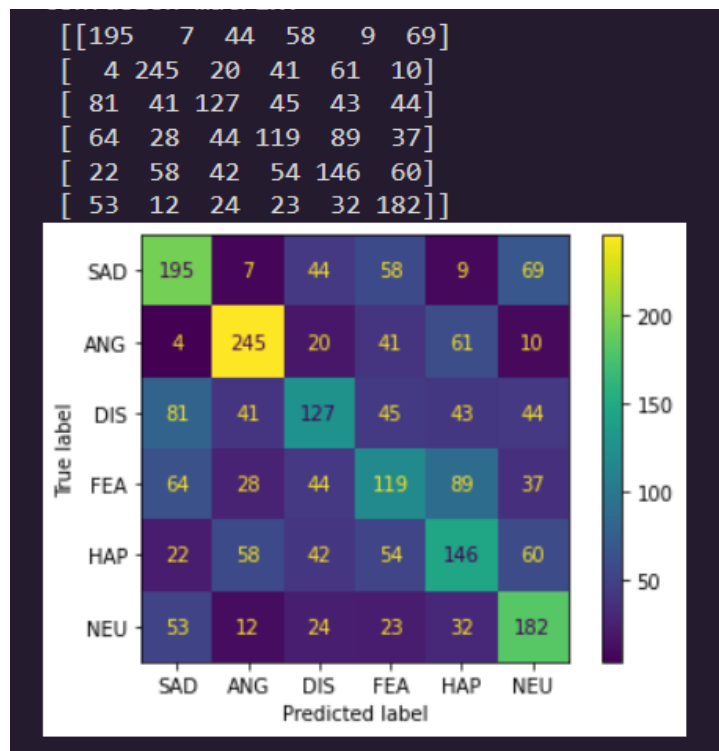| Dataset	Dropout	training accuracy	validation accuracy
original	0.1	0.99015	0.42599
original + noise	0.1	0.988	0.465
original + shift right	0.1	0.991	0.499
original + shift_left	0.1	0.99215	0.514688
original + noise + shift_right	0.5	0.992	0.514
original + noise + shift_left	0.5	0.9744554	0.5247016
original + noise + shift left + shift right	0.5	0.965	0.533

Results of the best parameters:
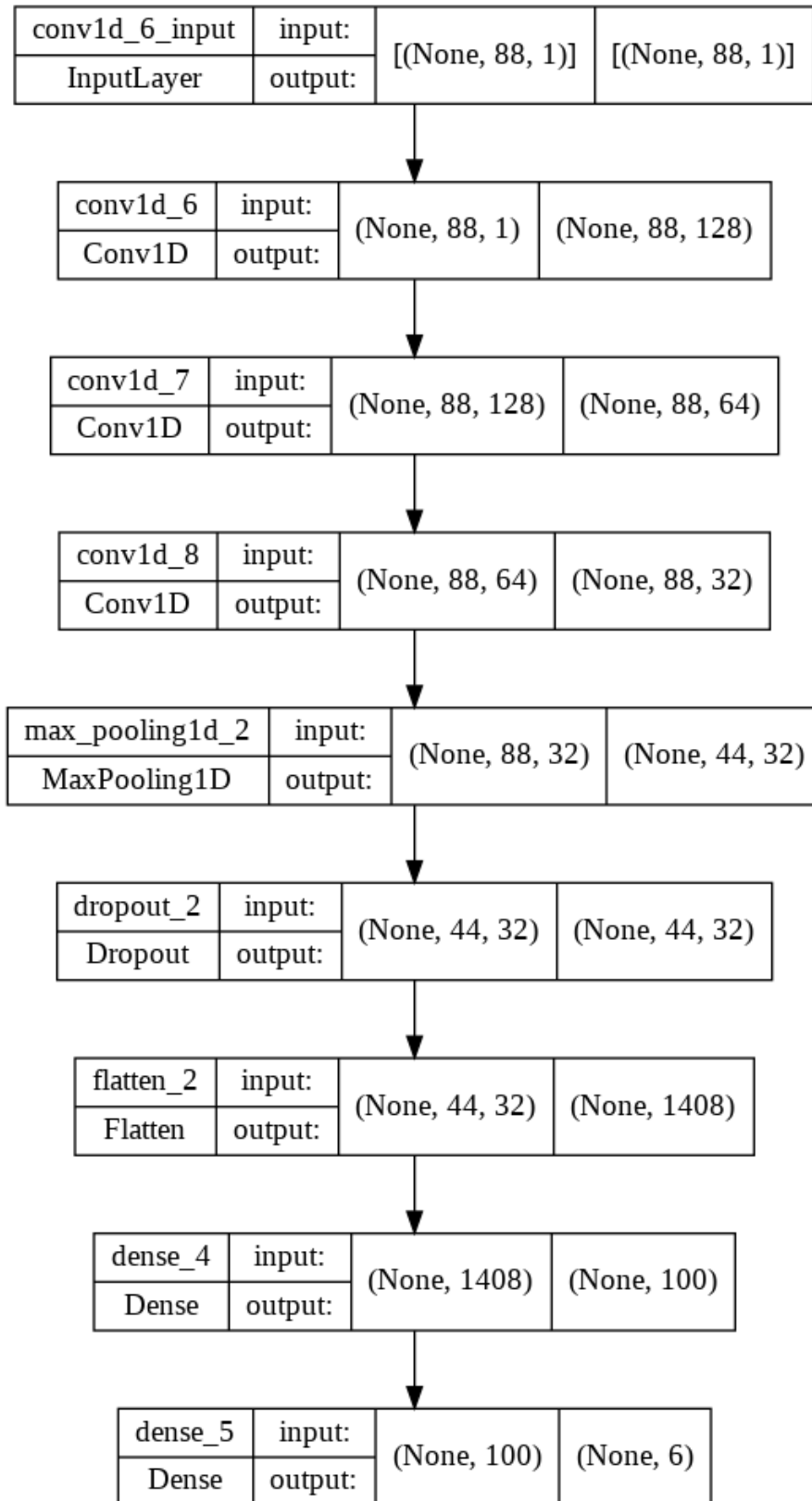   train accuracy = 88%
    testing accuracy = 45.4%

f1-scores

SAD	ANG	DIS	FEA	HAP	NEU
0.487	0.635	0.372	0.33	0.383	0.5

confusion matrix

```
[[195    7   44   58    9   69]
 [   4  245   20   41   61   10]
 [  81   41  127   45   43   44]
 [  64   28   44  119   89   37]
 [  22   58   42   54  146   60]
 [  53   12   24   23   32  182]]
```

- Model 2

conv1d_6_input	input:	[(None, 88, 1)]	[(None, 88, 1)]
InputLayer	output:		

conv1d_6	input:	(None, 88, 1)	(None, 88, 128)
Conv1D	output:		

conv1d_7	input:	(None, 88, 128)	(None, 88, 64)
Conv1D	output:		

conv1d_8	input:	(None, 88, 64)	(None, 88, 32)
Conv1D	output:		

max_pooling1d_2	input:	(None, 88, 32)	(None, 44, 32)
MaxPooling1D	output:		

dropout_2	input:	(None, 44, 32)	(None, 44, 32)
Dropout	output:		

flatten_2	input:	(None, 44, 32)	(None, 1408)
Flatten	output:		

dense_4	input:	(None, 1408)	(None, 100)
Dense	output:		

dense_5	input:	(None, 100)	(None, 6)
Dense	output:		

- The number of filters is increased to extract more low level features compared with model one.

Tuning

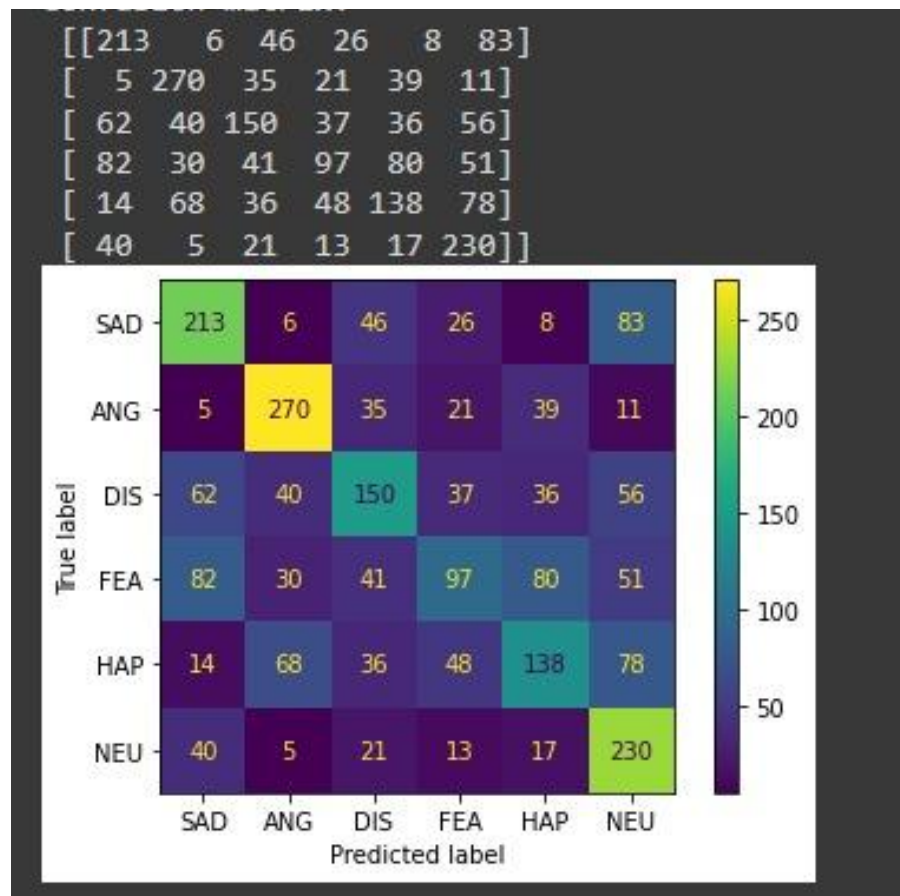| Dataset	Dropout	training accuracy	validation accuracy
original	0.1	0.94378	0.420975
original + noise	0.1	0.90858	0.452006
original + shift right	0.1	0.934768	0.469954
original + shift_left	0.1	0.945302	0.474755
original + noise + shift_right + shift_left	0.1	0.743265	0.515237
original + noise	0.5	0.75394	0.46718
original + shift_right	0.5	0.79298	0.5068
original + noise + shift left + shift right	0.5	0.738	0.517

Results of best parameters
train accuracy = 61.6%
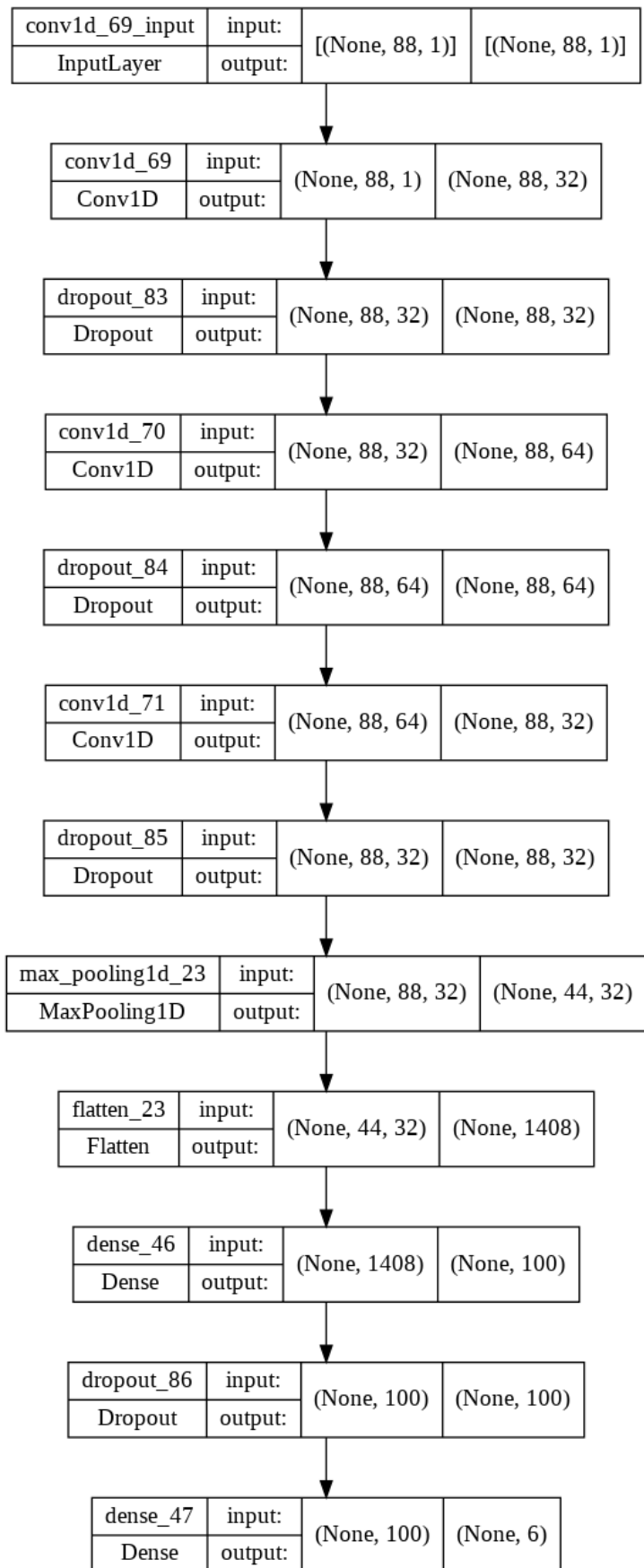testing accuracy =49%

f1-scores

SAD	ANG	DIS	FEA	HAP	NEU
0.533	0.675	0.422	0.311	0.394	0.55

confusion matrix

- model 3

conv1d_69_input	input:	[(None, 88, 1)]	[(None, 88, 1)]
InputLayer	output:		

conv1d_69	input:	(None, 88, 1)	(None, 88, 32)
Conv1D	output:		

dropout_83	input:	(None, 88, 32)	(None, 88, 32)
Dropout	output:		

conv1d_70	input:	(None, 88, 32)	(None, 88, 64)
Conv1D	output:		

dropout_84	input:	(None, 88, 64)	(None, 88, 64)
Dropout	output:		

conv1d_71	input:	(None, 88, 64)	(None, 88, 32)
Conv1D	output:		

dropout_85	input:	(None, 88, 32)	(None, 88, 32)
Dropout	output:		

max_pooling1d_23	input:	(None, 88, 32)	(None, 44, 32)
MaxPooling1D	output:		

flatten_23	input:	(None, 44, 32)	(None, 1408)
Flatten	output:		

dense_46	input:	(None, 1408)	(None, 100)
Dense	output:		

dropout_86	input:	(None, 100)	(None, 100)
Dropout	output:		

dense_47	input:	(None, 100)	(None, 6)
Dense	output:		

To reduce overfitting:
- kernel weights initializer is set to random normal.
- L1 regularization is added.
- More dropout layers are added with dropout ratio = 0.1.

Tuning

Dataset	Dense	regularization factor	training accuracy	validation accuracy
original	100	1e-5	0.761	0.46
original + noise + shift_left	100	1e-5	0.585	0.485
original + shift_left	100	1e-7	0.71717	0.49625
original + noise + shift left + shift right	100	1e-7	0.646	0.488
original + noise	500	1e-5	0.533895	0.3971

Results of best parameters
train accuracy = 56.17%
testing accuracy = 47.5%

f1-scores

SAD	ANG	DIS	FEA	HAP	NEU
0.51	0.643	0.415	0.342	0.408	0.523

confusion matrix

```
[[199    2   23   67   12   79]
 [   5  244   43   29   52    8]
 [  70   38  142   50   33   48]
 [  66   37   31  128   83   36]
 [  17   52   41   60  151   61]
 [  41    5   23   33   27  197]]
```
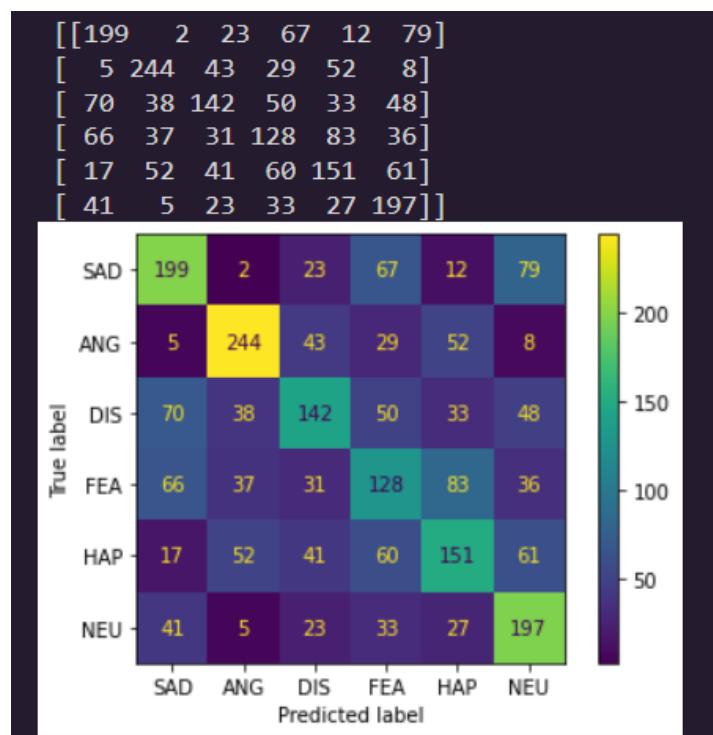

Best Model

Best Model : model 2
model 2 and 3 have close testing accuracy.

Big Picture

Most confusing classes:
- Happy with Fear
- Sad with Neutral.

In general, Fear has high values and gets misclassified as other classes
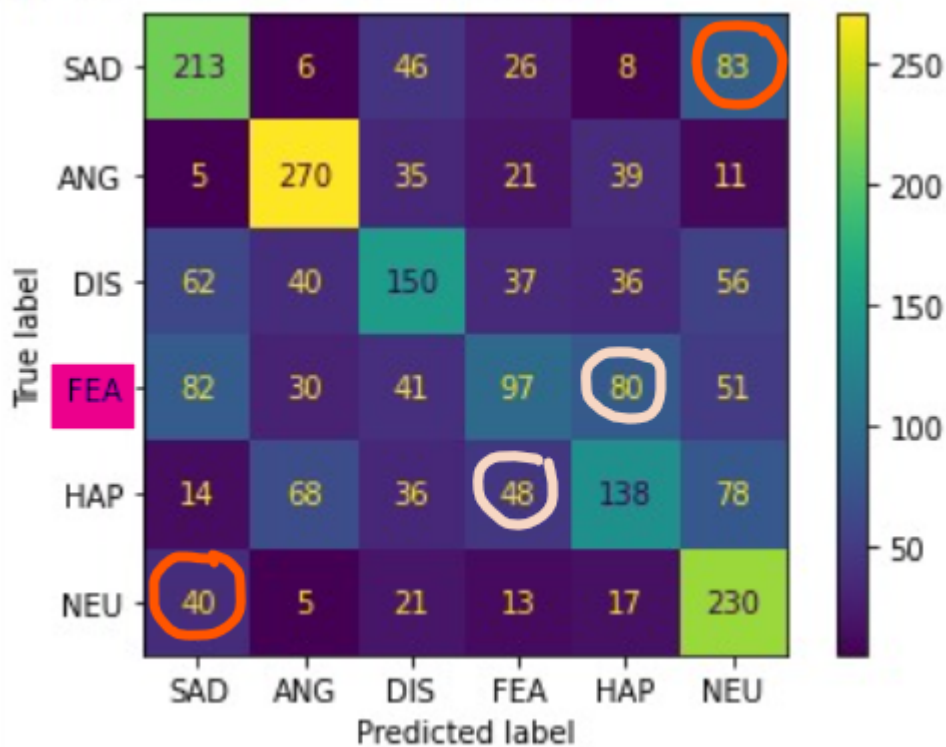
Data Preprocessing

Feature space

Each sample is represented by (128x44) array representing its mel spectrogram (an offset 0.5 second is used)
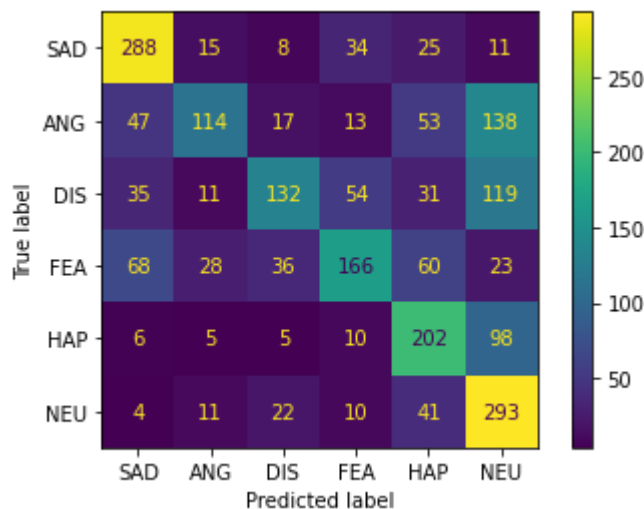Audio sampling rate is adapted to generate a fixed array size for all samples.

Best Model

[PNG of model](#)

```
loss: 2.2840 - accuracy: 0.5352
```

Big Picture

f1-scores

SAD	ANG	DIS	FEA	HAP	NEU
0.69	0.40	0.44	0.50	0.55	0.55

Confusing cases:
- Misidentifying angry,disappointed and happy as neutral.