**MODELS**

What to try:

* ATTNN paper
* ATT NN paper but with transformers
* ATT NN paper but with dsconv e gdsconv
* Other papers????

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **#** | **Dataset** | **Acc** | **Bias** | **Var** | **Params** | **Network** | **LR** | **OPT** | **Reg** | **batchS** | **Comments** | **File** |
| 1 | 10 cmd  10k-1k-1k  40mfcc +delta | 0.771 | 0 | 0.227 | 1,465,226 | LeNet5 con elu | / | Nadam | / | 32 | 0 Bias  High Variance  Overfitting |  |
| 2 | 10 cmd  10k-1k-1k  40mfcc +delta | 0.687 | 0.013 | 0.333 | 123,542 | LeNet5 con elu  Senza i 2 FC layers | / | Nadam | / | 32 | 0 Bias  High Variance  Overfitting |  |
| 3 | 10 cmd  10k-1k-1k  40mfcc +delta | 0.656 | 0 | 0.358 | 123,630 | LeNet5 con elu  Senza i 2 FC layers  Con Batch Norm | / | Nadam | / | 32 | Peggio del 2 | 2020-07-19\_15-13\_LeNet5-elu-1FC-BN |
| 4 | 10 cmd  10k-1k-1k  40mfcc +delta | 0.691 | 0.002 | 0.312 | 123,542 | LeNet5 con elu  Senza i 2 FC layers  Regolariz | / | Nadam | L2  1e-4 | 32 | Leggermente meglio del 2  Overfitting | 2020-07-19\_15-36\_LeNet5-elu-1FC-Reg |
| 5 | 10 cmd  10k-1k-1k  40mfcc +delta | 0.707 | 0.081 | 0.235 | 123,542 | LeNet5 con elu  Senza i 2 FC layers  Regolariz | / | Nadam | L2  1e-3 | 32 | Some Bias  High Variance  Leggermente meglio del 2 e 3  Overfitting | 2020-07-19\_15-44\_LeNet5-elu-1FC-Reg |
| 6 | 10 cmd  10k-1k-1k  40mfcc +delta | 0.796 | 0 | 0.223 | 1,455,422 | LeNet5 con elu  Con FC 120  Regolariz | / | Nadam | L2  1e-3 | 32 | 0 Bias  High Variance  Meglio dei precedenti  Overfitting | 2020-07-19\_16-02\_LeNet5-elu-2FC-Reg |
| 7 | 10 cmd  10k-1k-1k  40mfcc +delta | 0.796 | 0 | 0.223 | 1,455,422 | LeNet5 con elu  Con FC 120  Regolariz | / | Nadam | L2  1e-2 | 32 | Cambiato nulla dal 6 | 2020-07-19\_16-44\_LeNet5-elu-2FC-Reg |
| 8 | 10 cmd  10k-1k-1k  12mfcc +delta | 0.757 | 0 | 0.254 | 224,341 | ATTNETWORK paper | / | Nadam | / | 32 | 0 Bias  High Variance  Overfitting | 2020-07-19\_18-07\_AttRNNSpeechModel |
| 9 | 10 cmd  10k-1k-1k  80 mel | 0.755 | 0 | 0.251 | 224,341 | ATTNETWORK paper | decay | Nadam | / | 32 | 0 Bias  High Variance  Overfitting  Confusion Matrix seems slightly better Than 8 | 2020-07-20\_14-58\_AttRNNSpeechModel |
| 10 | 10 cmd  10k-1k-1k  80 mel Normalized |  |  |  | 224,341 | ATTNETWORK paper | decay | Nadam | / | 32 | Vary Bad |  |
| 11 | 10 cmd  10k-1k-1k  80 mel | 0.802 | 0 | 0.223 | 224,341 | ATTNETWORK paper | decay | adam | / | 32 | 0 Bias  High Variance  Overfitting  Better than 6 | 2020-07-20\_15-34\_AttRNNSpeechModel |
| 12 | 10 cmd  10k-1k-1k  40mfcc | 0.773 | 0 | 0.232 | 224,341 | ATTNETWORK paper | decay | adam | / | 32 | 0 Bias  High Variance  Overfitting | 2020-07-20\_16-40\_AttRNNSpeechModel |
| 13 | 10 cmd  20k-1k-1k  40mfcc +delta | 0.773 | 0.029 | 0.205 | 224,341 | ATTNETWORK paper | / | adam | / | 32 | Little Bias  High Variance  Overfitting  Little variance improvement | CLUSTER DEI  2020-07-20\_22-28\_AttRNNSpeechModel |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

**Datasets**:

* 10 cmd
* 10 cmd + silence +unknown
* 20 cmd + unknown

**Preprocessing**:

* No. Directly train on the 16000-element vector
* Mel spectrogram with 80mels
* 40 MFCC
* 40MFCC +delta +deltadelta
* Solo 13 MFCC?????

**Architectures:**

* ????

**Metrics**:

* Accuracy (in a problem where there is a large class imbalance, a model can predict the value of the majority class for all predictions and achieve a high classification accuracy. So, further performance measures are needed such as F1 score and Brier score, but since in our dataset the classes are balanced we can still use Accuracy)
* Prediction speed (ms)

Project (60 points):

* originality (10)
* data preprocessing techniques (10)
* learning architectures (20)
* comparison against other/existing approaches (10)
* live demo of the code (10)

Written report (40 points):

* clarity of exposition (10)
* completeness (10)
* analysis of results (number and type of metrics used) (20)

Oral exposition (20 points):

* duration (your talk must be shorter than 25 minutes, using slides) (10)
* clarity of exposition (10)

The final grade will be computed as grade = (points\*30)/100