**First Model**

**2D Parallel CNN**

**Original model Using RAVDESS Dataset**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Augmentation | Learning rate | Extracted features | Dataset | Accuracy | Number of emotions |
| 2D parallel CNN with four transformer layer (self attention layer) | White gaussian noise  on train  &  vaild and test | 0.01 | MFCC | Speech | 64.87% | 8 |
| 2D parallel CNN with four transformer layer  (self attention layer) | White gaussian noise  on train  &  vaild and test | 0.01 | MFCC | Speech &  song | 73% | 8 |

**Our Models**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Augmentation | Learning rate | Extracted features | Dataset | Accuracy | Number of emotions |
| Same original | White gaussian noise  on train  &  vaild and test | 0.01 | MFCC | Speech  &  Song | 76% | 6 |
| change in dropout ratio | white gaussian noise  on train only | 0.01 | MFCC | Speech  &  Song | 79.9% | 6 |
| Same original | white gaussian noise  on train only | 0.01 | MFCC | Speech  &  Song | 79.4% | 6 |
| Same original | white gaussian noise  on train only | 0.01 | MFCC | Speech  &  Song | 84.96% | 4 |
| Same original | white gaussian noise  on train only | 0.01 | MFCC | Speech  &  Song | 80.12% | 5 |
| Same original | White gaussian noise  on train  &  vaild and test | 0.01 | MFCC | Speech  &  Song | 72% | 6 |
| 3 parallel CNN Blocks | white gaussian noise  on train only | 0.01 | MFCC | Speech  &  Song | 77.51% | 6 |
| 2 transformer layers instead of 4 | white gaussian noise  on train only | 0.01 | MFCC | Speech  &  Song | 77.99% | 6 |
| 1 Block CNN Instead of 2D Parallel CNN | white gaussian noise  on train only | 0.01 | MFCC | Speech  &  Song | 77.03% | 6 |
| Change only in Drop ratio | white gaussian noise  on train only | 0.01 | MFCC | Speech  &  Song | 77% | 6 |
| Adding layer to original model  (4 layers in each Block) | white gaussian noise  on train only | 0.01 | MFCC | Speech  &  Song | 75.12% | 6 |
| Same original | white gaussian noise  on train only | 0.001  &  weight decay  =  1e-6 | MFCC | Speech  &  Song | 71.62% | 7 |
| Same original | white gaussian noise  on train only | 0.01 | MFCC | Speech  &  Song | 73.36% | 7 |
| Same original | Without augmentation | 0.01 | MFCC | Speech  &  Song | 72.05% | 7 |
| Same original | White gaussian noise  on train  &  vaild and test | 0.001  &  weight decay  =  1e-6 | MFCC | Speech  &  Song  &  Dataset  60,20,20 | 72.57% | 7 |
| Same original | white gaussian noise  on train only | 0.01 | MFCC | Speech  &  Song | 75% | 6 |
| Same original | White gaussian noise  on train  &  vaild and test | 0.01 | MFCC | Speech | 68% | 7 |
| Same original | white gaussian noise  on train only | 0.01 | MFCC | Speech  &  Song | 78.95% | 5 |
| Drop second CNN layer by 0.8 ratio | white gaussian noise  on train only | 0.01 | MFCC | Speech  &  Song | 69.86% | 6 |
| Same original | white gaussian noise  on train only | Lr  Decrease with time | MFCC | Speech  &  Song | 80.86% | 6 |

**Second Model**

**CNN**

****

**Third Model**

**CNN**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Accuracy** | **Emotions** | **Batch size** | **Epoch** | **Feature extraction** | **Dataset** | **Model** |
| 61 | 4 | 20 | 500 | MFCC | RAVDESS Speech | **CNN** |
| 60 |
| 65 |
| 75 | 4 | 20 | 500 | RAVDESS Speech  &  Song |
| 77 | 4 | 10 | 500 |
| 73 | 4 | 5 | 500 |
| 75 | 6 | 4 | 500 |
| 74 | 6 | 24 | 200 |
| 75 | 6 | 26 | 200 |
| 81 | 6 | 10 | 500 | MFCC & Mel &chroma |

**Conclusion**

The highest accuracy we achieved is on the second model using Batch normalization and also using augmentation.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | MODEL | AVG Accuracy | Best Accuracy | Worst Accuracy |
| RAVDESS Speech  &  song | CNN | 83.13% | 86.47% | 81% |
| SAVEE | CNN |  | 72.22% | 69% |