All the results using Mel-spectrogram as the input feature on the Librispeech dev-clean data:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Diff. test ratio | Precision (M/F) | Recall (M/F) | F-Score (M/F) | Overall Accuracy |
| 0.9:0.1 | 0.95/0.85 | 0.84/0.95 | 0.89/0.90 | 0.89 |
| 0.8:0.2 | 0.94/0.83 | 0.81/0.95 | 0.87/.88 | 0.87 |
| 0.75:0.25 | 0.96/0.86 | 0.85/0.96 | 0.90/0.91 | 0.90 |
| **0.7:0.3** | **0.96/0.86** | **0.86/0.96** | **0.91/0.90** | **0.91** |
| 0.65:0.35 | 0.95/0.86 | 0.86/0.95 | 0.90/0.90 | 0.90 |
| 0.6:0.4 | 0.95/0.82 | 0.80/0.90 | 0.87/0.88 | 0.88 |

 

Fig. 1. T-SNE plots of the (a) data distribution and (b) test result of the model