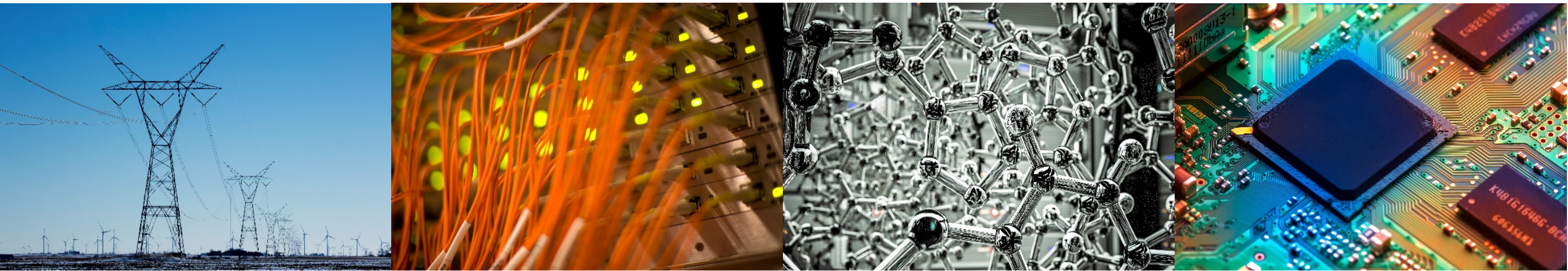


# TF Speech Recognition Challenge

Dongbo Wang, Kexin Hui, Zhaoheng Hu

ECE544 Final Presentation



# Introduction

- Tensorflow Speech Commands Dataset
- Preprocessing: Audio Feature Extraction
- LSTM
- Cross Entropy Loss

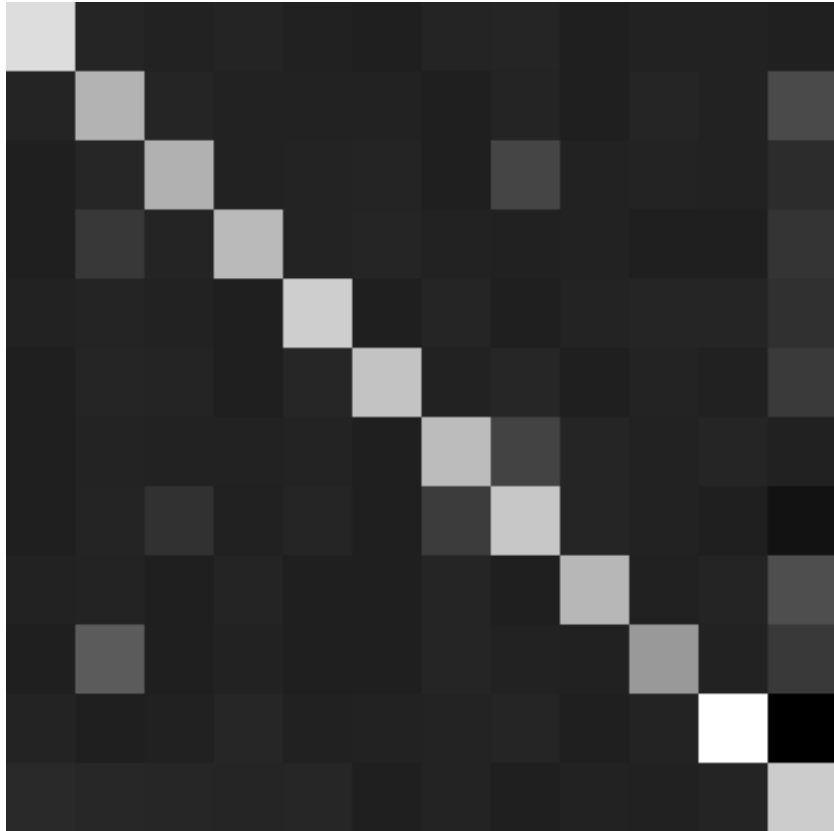
# Feature Extraction

- No Extraction
- Short Time Fourier Transform (STFT)
- Filter Banks (FBanks)
- Mel-frequency Cepstral Coefficients (MFCC)

# Results

- Unweighted vs. **Weighted** Loss Function
- None vs. STFT vs. **FBanks** vs. MFCC
- 64 vs. **128** vs. 256 Hidden Units

# Results



Overall accuracy: 73.101%

>>> class yes: 84.551%

>>> class no: 65.602%

>>> class up: 64.777%

>>> class down: 69.0%

>>> class left: 78.0%

>>> class right: 72.937%

>>> class on: 69.945%

>>> class off: 74.592%

>>> class stop: 67.539%

>>> class go: 54.0%

>>> class silence: 99.559%

>>> class unknown: 76.713%

# Thank you!