# 1 Introduction

Given audio recordings of ten people reading the digits 0 to 9, we wish to train a computer to take the `.wav` files and accurately classify the digit being spoken. We follow [1] by converting the audio files into mel-frequency spectral coefficients, producing an image that can be inputted into a 2D convolutional neural network (CNN). We also compare this CNN with a recurrent neural network (RNN).

# References

[1] Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu. Convolutional neural networks for speech recognition. *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, 22(10):1533–1545, 2014.