Team:

Project title: Speech Emotion Recognition (SER)

Project summary: The capability to discern emotions from spoken words is crucial for improving the interaction between humans and computers. The speech signal offers a wealth of information about the speaker, including their age, gender, ethnicity, health status, emotions, and thoughts. Speech processing techniques are vital in various human-computer interaction domains, such as safety systems, computer education tools, in-car systems, automated translation services, call center technologies, monitoring and diagnosis of psychological conditions, voice mail systems, telecommunications, assistive devices, and audio analysis.

Research has explored a monitoring technology aimed at addressing crimes like robbery, harassment, domestic and gender violence, and kidnapping, all while preserving individual privacy. Some real-time models utilize speech recognition, natural language understanding, and pattern identification to autonomously identify situations of insecurity.

Traditionally, the most common classification methods include techniques like Hidden Markov Models (HMM), Gaussian Mixture Models (GMM), the Naive Bayes classifier, and Support Vector Machines (SVM). However, deep learning has recently emerged as a powerful branch of machine learning, showing remarkable achievements across various sectors, particularly in computer vision, speech recognition, and natural language understanding. Systems for recognizing emotions in speech that leverage these advanced deep learning frameworks have demonstrated substantial enhancements compared to traditional techniques.

Approach:

Our approach to enhancing speech emotion recognition includes exploring and implementing state-of-the-art deep learning architectures. We plan to investigate several promising models and techniques, including:

Convolutional Neural Networks (CNNs) for extracting robust, discriminative features from raw audio signals or spectrograms. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, for capturing the temporal dynamics of speech. Attention Mechanisms to focus on emotion-relevant parts of the speech signal, potentially improving the model's sensitivity to emotional nuances.

Initially, we plan to benchmark these methods using standard datasets like the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) and the Emo-DB database. We aim to reproduce existing baselines and then incrementally introduce our innovations to measure performance improvements.

Resources/Related Work:

Kim, T. W., & Kwak, K. C. (2024). Speech Emotion Recognition Using Deep Learning Transfer Models and Explainable Techniques. *Applied Sciences, 14*(4), 1553.

Issa, D., Demirci, M. F., & Yazici, A. (2020). Speech emotion recognition with deep convolutional neural networks. *Biomedical Signal Processing and Control, 59*, 101894.

Aouani, H., & Ayed, Y. B. (2020). Speech emotion recognition with deep learning. *Procedia Computer Science*, *176*, 251-260.

Siam, A. I., Soliman, N. F., Algarni, A. D., Abd El-Samie, F. E., & Sedik, A. (2022). Deploying machine learning techniques for human emotion detection. *Computational intelligence and neuroscience*, *2022*.

Data sets: **https://zenodo.org/records/1188976#.XrC7a5NKjOR**
https://github.com/CheyneyComputerScience/CREMA-D

Team members:

Yicun Deng
Thanushan Kirupairaja
Dean Gladish
Subanky Suvendran