A black and purple background

Description automatically generated

A purple and blue waves on a black background

Description automatically generated

Word Count: XXXX

Date: DD/MM/YYYY

KV6003 – Individual Computing Project

Panagiotis Theodorou – W20020633

Supervisor: Dr Hua Mao

2nd Marker: Dr Nic Whitton

A Comparative Analysis of Machine Learning and Deep Learning Models

Speech Emotion Recognition (S.E.R)

**2023 - 2024**

Pages

**Title: Speech Emotion Recognition using Machine Learning and Deep Learning Models**

**Abstract:**

Provide a concise summary of your report, including the purpose, methodology, key findings, and implications of your research.

# Introduction

## Background

## Objectives

## Scope and Significance

# Literature Review

## Overview of Speech Emotion Recognition

The Speech Emotion Recognition (SER) domain is a multidisciplinary field encompassing various key terms, concepts, and techniques to comprehend emotions conveyed through human speech (Wani et al., 2021). Paralinguistic information, such as gender, personality, emotion, aim, and state of mind, is integral in decoding the hidden significance of utterances within SER (Wani et al., 2021). Initial emphasis in classical automatic speech recognition systems, as highlighted in Wani et al.'s review, was on linguistic content, with less attention to paralinguistic information (Wani et al., 2021). However, the evolution of SER acknowledges the profound impact of meagreness in paralinguistic awareness on communication, especially in children, leading to substandard social skills and potential psychopathological manifestations (Wani et al., 2021). Modern SER systems aim to create coherent and human-like communication machines by comprehending paralinguistic data, especially emotional nuances, to enhance communication effectiveness.

Methodologies employed in SER encompass diverse approaches, including traditional classifiers like Gaussian Mixture Models (GMM), Hidden Markov Models (HMM), Support Vector Machines (SVM), k-Nearest Neighbour (kNN), and contemporary techniques like Deep Neural Networks (DNN) (Wani et al., 2021). These illustrate the evolution of classification methods for emotion recognition. Feature extraction techniques in SER, as discussed in both articles, range from prosodic and spectral features to voice quality features and Teager Energy Operator (TEO)-based features, showcasing the comprehensive nature of information considered (Wani et al., 2021).

Jaiswal et al. (2020) outlined the processing steps involved in speech signal analysis, including framing, windowing, voice activity detection, normalization, and noise reduction. The significance of feature selection and dimension reduction techniques is underscored to mitigate challenges associated with high-dimensional data (Jaiswal et al., 2020). Global features, representing supra-segmental or long-term characteristics, and local features, capturing segmental or short-term dynamics, are critical components in feature extraction (Jaiswal et al., 2020). Various spectral features, such as Mel Frequency Cepstral Coefficients (MFCC), Linear Prediction Cepstral Coefficients (LPCC), and Gammatone Frequency Cepstral Coefficients (GFCC), offer insights into the frequency domain characteristics of speech signals (Jaiswal et al., 2020). The importance of classifiers, both traditional and deep learning-based, is evident in the choice of models like Gaussian Mixture Model (GMM), Hidden Markov Model (HMM), Support Vector Machine (SVM), k-Nearest Neighbour (kNN), and Deep Neural Networks (DNN) for accurate emotion classification (Jaiswal et al., 2020). These fundamental elements collectively contribute to the rich landscape of Speech Emotion Recognition, reflecting the interdisciplinary nature of the field, from linguistics and psychology to signal processing and machine learning.

## Studies and Models

### Deep Learning Models

#### ANN

#### CNN

#### MLP

#### ResNet

#### RNN+LSTM

#### Transformer

### Machine Learning Models

#### GMM

#### HMM

#### KNN

#### RF

#### SVM

### Ensemble Models

#### Lorem Ipsum 1

#### Lorem Ipsum 2

#### Lorem Ipsum 3

#### Lorem Ipsum 4

#### Lorem Ipsum 5

## Challenges and Limitations

### Datasets

#### CREMA-D

#### RAVDESS

#### SAVEE

#### EMO-DB

### Emotions

### Authenticity

### Cultural and Gender biases

### Subjectivity and Variability

# Methodology

## Feature Extraction

## Augmentation

## Performance

## Model Selection

# Own Approach

## Lorem Ipsum 1

## Lorem Ipsum 2

## Lorem Ipsum 3

## Lorem Ipsum 4

## Lorem Ipsum 5

# Implementation

## Lorem Ipsum 1

## Lorem Ipsum 2

## Lorem Ipsum 3

## Lorem Ipsum 4

## Lorem Ipsum 5

# Results

## Lorem Ipsum 1

## Lorem Ipsum 2

## Lorem Ipsum 3

## Lorem Ipsum 4

## Lorem Ipsum 5

# Discussion

## Lorem Ipsum 1

## Lorem Ipsum 2

## Lorem Ipsum 3

## Lorem Ipsum 4

## Lorem Ipsum 5

# Conclusion

## Lorem Ipsum 1

## Lorem Ipsum 2

## Lorem Ipsum 3

## Lorem Ipsum 4

## Lorem Ipsum 5

**References:**

* + Wani, T. M., Gunawan, T. S., Qadri, S. A. A., Kartiwi, M., & Ambikairajah, E. (2021). A Comprehensive Review of Speech Emotion Recognition Systems. IEEE Access, 9, 47795-47814. doi: 10.1109/ACCESS.2021.3068045.
  + Jaiswal, A., Tiwari, P., Kumar, A., & Tiwari, S. (2020). Speech Emotion Recognition: A Review of Contemporary Techniques. Journal of King Saud University-Computer and Information Sciences. doi: 10.1016/j.jksuci.2020.06.035.
  + P. Tzirakis, J. Zhang and B. W. Schuller, "End-to-End Speech Emotion Recognition Using Deep Neural Networks," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018, pp. 5089-5093, doi: 10.1109/ICASSP.2018.8462677.
  + W. Lim, D. Jang and T. Lee, "Speech emotion recognition using convolutional and Recurrent Neural Networks," 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Jeju, Korea (South), 2016, pp. 1-4, doi: 10.1109/APSIPA.2016.7820699.
  + S. Mirsamadi, E. Barsoum and C. Zhang, "Automatic speech emotion recognition using recurrent neural networks with local attention," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 2017, pp. 2227-2231, doi: 10.1109/ICASSP.2017.7952552.
  + C. Zhang and L. Xue, "Autoencoder With Emotion Embedding for Speech Emotion Recognition," in IEEE Access, vol. 9, pp. 51231-51241, 2021, doi: 10.1109/ACCESS.2021.3069818.
  + Lijiang Chen, Xia Mao, Yuli Xue, Lee Lung Cheng, Speech emotion recognition: Features and classification models, Digital Signal Processing, Volume 22, Issue 6, 2012, Pages 1154-1160, ISSN 1051-2004, <https://doi.org/10.1016/j.dsp.2012.05.007>.
  + Tin Lay Nwe, Say Wei Foo, Liyanage C De Silva, Speech emotion recognition using hidden Markov models, Speech Communication, Volume 41, Issue 4, 2003, Pages 603-623, ISSN 0167-6393, https://doi.org/10.1016/S0167-6393(03)00099-2.

**Acknowledgments:**

* Acknowledge any individuals, organizations, or resources that contributed to your research.

**Appendices:**

* Include any additional material such as code snippets, detailed model architectures, and supplementary information.