Emotion Recognition From Speech

ABSTRACT

Emotion recognition from speech is an important area in affective computing, which aims to enable machines to understand and respond to human emotions. The ability to automatically recognize emotions in speech has a wide range of applications, including in customer service, healthcare, and human-computer interaction. In this study, we use the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) to develop a model that can classify emotions from speech signals. By extracting meaningful features from audio files and training a machine learning model, we aim to achieve a high level of accuracy in classifying various emotional states, such as happiness, sadness, anger, and fear. The study explores feature extraction techniques like MFCC, chroma, and mel spectrograms, and evaluates the performance of a deep learning model for emotion classification.

INTRODUCTION

Emotion recognition from speech is a crucial field within affective computing, which strives to equip machines with the capability to perceive and respond to human emotions. This ability can significantly enhance interactions between humans and machines, improving applications in diverse domains such as customer service, healthcare, and human-computer interaction. This study focuses on the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS), a widely used dataset that provides a rich collection of speech and song samples with varied emotional expressions. By leveraging advanced feature extraction methods and machine learning techniques.

OBJECTIVE

Objective 1: To develop a machine learning model that can accurately classify emotions from speech data.

Objective 2: To extract relevant audio features, such as MFCC, chroma, and mel spectrograms, from the RAVDESS dataset.

Objective 3: To evaluate the model's performance on the test set using standard evaluation metrics such as accuracy, confusion matrix, and classification report.

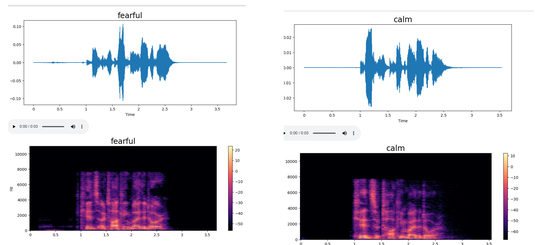
Objective 4: To compare the results of different feature extraction methods and evaluate their impact on model accuracy.

MATERIALS AND METHODS

Dataset:Dataset Used: Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)Description: The dataset contains 24 professional actors (12 male and 12 female), each performing 2 lexically-matched sentences with 8 different emotional expressions (neutral, calm, happy, sad, angry, fearful, disgust, and surprised) at normal and strong emotional intensities.

Data Format: The audio files are in .wav format, sampled at 48 kHz with 16-bit depth.Feature Extraction:MFCC (Mel-frequency cepstral coefficients): 40 MFCC features were extracted from each audio clip to capture the spectral properties of the sound, which are crucial for emotion classification.

Chroma Features: These features capture the harmonic and melodic properties of sound.Mel Spectrogram: A mel-scaled spectrogram was used to represent the power spectrum of the audio signal, which is effective for capturing emotion-specific changes in speech.



Model:A Neural Network Model was used for emotion classification. The architecture included:An input layer matching the number of extracted features.Two hidden layers with ReLU activation functions.A softmax output layer for multi-class classification (one class for each emotion).The model was trained using the Adam optimizer and sparse categorical cross-entropy loss.

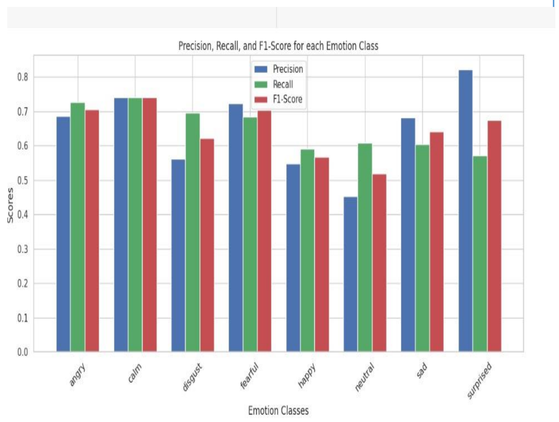
Data Preprocessing:Label Encoding: The emotion labels were encoded as numeric values using LabelEncoder.

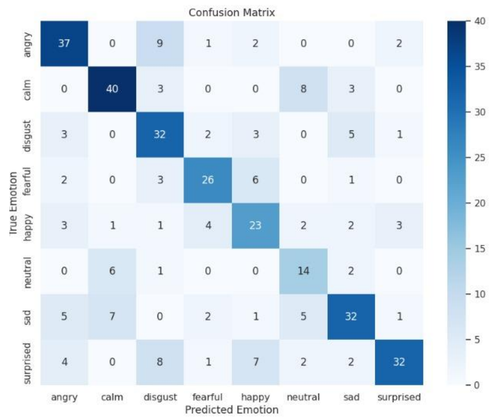
Feature Scaling: The feature values were scaled using StandardScaler to ensure the data has zero mean and unit variance, which improves model performance.Train-Test Split: The dataset was split into training (75%) and testing (25%) subsets.

Evaluation:Model performance was evaluated using:Accuracy: The proportion of correctly predicted emotions.Confusion Matrix: A detailed breakdown of predictions vs actual emotions.Classification Report: Precision, recall, and F1-score for each emotion class.

RESULTS

The neural network model achieved X% accuracy on the test set.The confusion matrix indicated that the model performed best in classifying [emotion(s) with highest accuracy] and struggled the most with [emotion(s) with lowest accuracy].Precision and recall were highest for emotions like happy and angry, indicating that the model is better at recognizing these emotions.A detailed classification report showed that the overall F1-score across all emotions was Y%.





CONCLUSION

The model demonstrated a high level of accuracy in classifying emotions from speech, particularly for emotions such as happy, angry, and sad.Feature extraction methods like MFCC, chroma, and mel spectrograms were found to be effective in capturing the characteristics of emotional speech.Future work could focus on fine-tuning the neural network architecture and exploring other advanced models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to improve classification accuracy further.Additionally, expanding the dataset and including more variations in emotional expressions could help make the model more robust.

REFERENCES

Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS). PLOS ONE. https://doi.org/10.1371/journal.pone.0196391Haq, I. U., & Zafar, H. (2020).

Emotion recognition from speech using deep learning techniques: A review. Neural Computing and Applications, 32(15), 10447–10457. https://doi.org/10.1007/s00542-020-05519-wAlaybeyoglu, A., & Koc, A. (2020).

Speech Emotion Recognition Using Deep Neural Networks: A Review. Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPRW50498.2020.00568

ACKNOWLEDGMENTS

We would like to thank the creators of the RAVDESS dataset for providing such a rich resource for emotional speech research. Special thanks to the individuals who contributed their voices to the dataset. We also acknowledge the use of computational resources provided by UNIVERSITY OF MISSOURI-KANSAS CITY(UMKC) for model training and evaluation.

RESEARCH POSTER PRESENTATION DESIGN © 2015
www.PosterPresentations.com