

Detecting stress anxiety depression from voice tone and text responses

Summary on research papers

Exploring the Effectiveness of Advanced Machine Learning Models in Speech Emotion Recognition

The paper presents a comparative study of traditional and deep learning algorithms in identifying emotions from speech using the RAVDESS dataset. Recognizing emotions from voice is crucial in fields like healthcare, security, and human-machine interaction. The study emphasizes the importance of feature extraction using Mel-Frequency Cepstral Coefficients (MFCCs) and compares models such as Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting Machine (GBM), Convolutional Neural Networks (CNN), and Long Short-Term Memory networks (LSTM).

Through detailed preprocessing, including data augmentation techniques like noise injection, voice shifting, and pitch modification, the models were trained and tested. Evaluation metrics such as accuracy, precision, recall, F1-score, ROC curves, and confusion matrices were used to assess performance. Among the models, **LSTM achieved the highest accuracy (92.3%)** and F1-score (0.91), closely followed by **CNN (91% accuracy)**, showcasing their strength in capturing temporal and spatial features respectively.

The ROC curves for CNN, LSTM, and SVM showed AUC values nearing 1.0, indicating excellent classification capabilities. The study concludes that deep learning models, particularly LSTM, outperform traditional machine learning models in speech emotion recognition due to their ability to learn from sequential audio data.

The authors recommend further enhancement of the models using larger and more diverse datasets to improve generalizability and robustness. The research reinforces the growing potential of AI in emotional understanding and its real-world application in emotion-aware systems.