Paper Title:
A Review of multi-modal Speech Emotion Recognition and various techniques used to Solve Emotion Recognition on Speech Data
Paper Link:
https://ieeexplore.ieee.org/document/10220691

Summary:
1.1 Motivation:
The study focused on understanding why recognizing emotions from speech is important, especially in areas like healthcare and marketing. This study wants to improve how computers understand human emotions in speech, despite facing challenges like noise and limited data.

1.2 Contribution:
It contributes by exploring different ways to recognize emotions from speech using not just sound but also text and video. It talks about using fancy computer methods to make this recognition better. It also talks about using specific datasets to help train the computer to understand emotions better.

1.3 Methodology:
The methodology involves utilizing trimodal input (audio, video, text) for speech emotion recognition (SER), employing fusion techniques. Deep learning architectures like SVM, Bc-LSTM, RNN, TFN, RBF, and CNN, are utilized. Preprocessing enhancements include MFCCs for audio and EEG for video input. Models are designed to integrate multi-modal inputs effectively. Widely used datasets include CMU-MOSI, IEMOCAP, DEAP, SEED, MELD, SAVEE, RAVDESS, and TESS.

1.4 Conclusion:
In conclusion, the study provides a comprehensive overview of multimodal SER, highlighting the importance of merging various modalities for accurate emotion prediction. It discusses conventional and deep learning methods applied to SER, identifies challenges, and suggests areas for further research.

Limitations:
2.1 First Limitation:
One limitation of the study is that the computer methods it uses need a lot of data and can be hard to understand. This makes it hard to use them in real life where we need to know why the computer made a certain decision.

2.2 Second Limitation:
Another limitation is that the datasets the study uses might not be perfect. This means the computer might not learn emotions in the same way we do, which can make it less accurate in real life.

Synthesis:
On the contrary, The study underscores the importance of multi-modal SER for accurate emotion recognition, especially in complex real-world scenarios. It emphasizes the need for further research to address challenges such as dataset bias, model interpretability, and scalability to develop more robust and reliable SER systems.