

Speech Emotion Recognition using Machine Learning With Real-time Audio Analysis

Prof. Sharda Dabhekar¹, Shivam K. Yadao², Tanay R. Tiwari³, Pranjal Zode⁴, Shubham B. Vaidya⁵

Guide, Department of Computer Science Engineering¹

Students, Department of Computer Science Engineering²

Rajiv Gandhi College of Engineering Research and Technology, Chandrapur, Maharashtra, India

shivamyadao284@gmail.com, tanaytiwari21@gmail.com,

pranjalzode007@gmail.com, shubhamvaidya599@gmail.com

Abstract: *This research paper presents a Speech Emotion Recognition (SER) system utilizing a Multilayer Perceptron (MLP) classifier and real-time audio analysis. The system records audio samples, extracts relevant features, and employs machine learning techniques to predict emotions in spoken language. The study focuses on the development of an intuitive Graphical User Interface (GUI) using the Kivy framework, providing a user-friendly platform for real-time emotion analysis.*

In shortly In this project, we attempt to detect underlying emotions such as (sad, happy, neutral, angry, disgust, surprised, fearful and calm) in recorded speech by analysing the acoustic features of the audio data of recordings and Created an application to implement the same on user input.

Keywords: python, python libraries, emotion recognition, Ravdess dataset, kivy framework for python

REFERENCES

- [1]. Han, K., et al. "Speech Emotion Recognition Using Deep Neural Network and Extreme Learning Machine." Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2014.
- [2]. Badshah, A. M., et al. "Emotion Speech Recognition System for Smart Affective Services Based on Deep Functionality." IEEE Access, 2019.
- [3]. Saito, Y., et al. "Statistical Parametric Speech Synthesis Incorporating Generative Adversarial Networks." IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2018.
- [4]. Zhou, X., et al. "Deep Learning Based Affective Model for Speech Emotion Recognition." IEEE Transactions on Affective Computing, 2017.
- [5]. Zhang, T., et al. "Speech Emotion Recognition with i-vector Feature and MN Model." IEEE Transactions on Multimedia, 2016.
- [6]. RAVDESS Dataset: Livingstone, S. R., Russo, F. A. "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English." PLoS ONE, 2018.

Future Work in SER:

- [7]. Schuller, B., et al. "The INTERSPEECH 2019 Computational Paralinguistics Challenge: Styrian Dialects, Continuous Sleepiness, Baby Sounds & Orca Activity." Proceedings of Interspeech, 2019.
- [8]. Eyben, F., et al. "The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing." IEEE Transactions on Affective Computing, 2016.
- [9]. Real-Time Audio Analysis: Marchi, E., Schuller, B. "Real-time Speech and Emotion Interaction: State of the Art and a Prototype." Proceedings of the IEEE, 2020.
- [10]. Multimodal Integration: Dhall, A., et al. "From Individual to Group Emotion Recognition: Emotiw 5.0." Proceedings of the 23rd ACM International Conference on Multimedia, 2015.

- [11]. Transfer Learning and Generalization: Pan, S. J., Yang, Q. "A Survey on Transfer Learning." IEEE Transactions on Knowledge and Data Engineering, 2010.
- [12]. User-Centric Design: Desmet, P. M. A. "Measuring Emotions: Development and Application of an Instrument to Measure Emotional Responses to Products." Human-Computer Interaction, 2003.
- [13]. Benchmarking and Comparative Studies: Schuller, B., et al. "The INTERSPEECH 2018 Computational Paralinguistics Challenge: Atypical & Assisted Communication, Crying & Galician." Proceedings of Interspeech, 2018.
- [14]. <https://doi.org/10.22214/ijraset.2022.41112>
- [15]. <https://www.doi.org/10.56726/IRJMETS43659>