IOCBE 2024 Conference
The 1st International Online Conference on Bioengineering
16–18 October 2024 | Online
MDPI

A Robust Approach for Emotional Assessment: The Employment of Power-Normalized Cepstral Coefficient and Stacked Classifiers

Michele Giuseppe Di Cesare, David Perpetuini, Daniela Cardone and Arcangelo Merla
Department of Engineering and Geology, University G. D'Annunzio of Chieti-Pescara, 65127 Pescara, Italy

INTRODUCTION & AIM

Emotional assessment is a pivotal area of research across various fields, driven by its ability to reflect individuals' real-time emotional states. Among the key indicators of emotional status, voice stands out as a rich source of information. Affective computing, which leverages voice recordings, plays a crucial role in domains such as healthcare and human-computer interaction. By analyzing vocal cues, affective computing systems can effectively identify emotional states, offering valuable insights into an individual's mental health and well-being. This study presents a novel machine learning approach for recognizing emotions from vocal recordings, utilizing Power Normalized Cepstral Coefficients (PNCC)[1] to capture essential vocal features, thereby advancing the field of emotion analysis in audio data.

METHOD

Emotion classification was performed using audio recordings from the EMOVO dataset, which includes fourteen semantically neutral sentences in Italian spoken by six actors (three male and three female) across seven emotions: neutral, disgust, anger, surprise, fear, sadness, and joy. The recordings were made with professional tools, and from these, 20 PNCC were extracted to represent vocal features. The dataset was randomly divided into a training set comprising 80% of the data (67 samples per emotion) and a testing set containing 20% (17 samples per emotion), ensuring balanced representation of each emotion. A data-driven approach led to the construction of a stacked classifier utilizing SVM and kNN as base models, with Extreme Gradient Boosting as the final evaluation model, chosen for its effectiveness in managing the high dimensionality of the feature set.

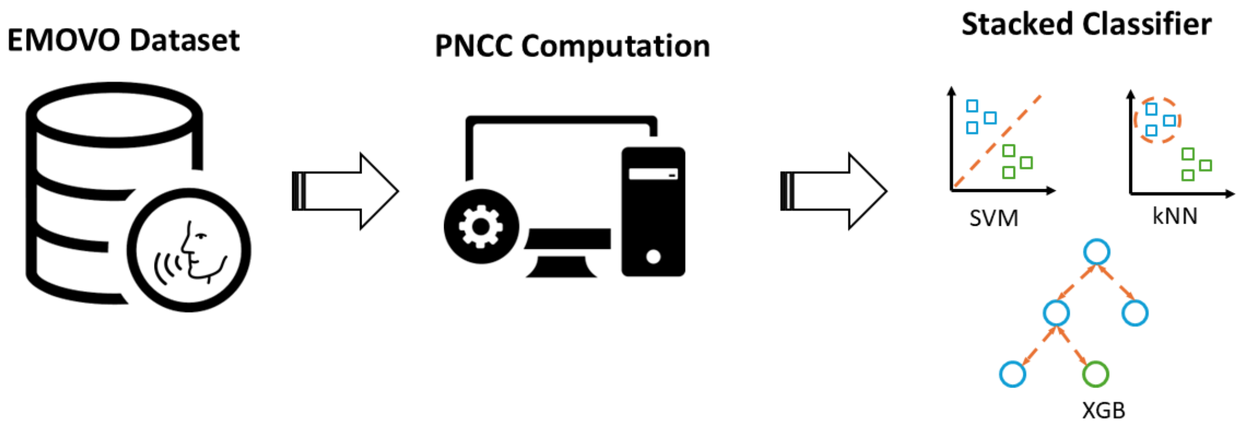

Figure 1. Processing pipeline of the study.

RESULTS & DISCUSSION

The proposed stacked classifier demonstrated strong performance, achieving 90% accuracy on the training set and 88% accuracy on the test set. The effectiveness of PNCC is highlighted by the high accuracy, even with training and testing sets split solely by emotion labels, without considering the intrinsic characteristics of the speaker or the sentence content.
This study confirms the remarkable utility of cepstral coefficients in emotion recognition, contributing to a growing body of research where these features have been used for vocal health analysis, neurodegenerative disease detection, and emotional state recognition[2].
While the dataset's use of semantically neutral phrases, varied in length and form, enhances classification realism, it also presents a limitation. Future work should focus on extending this approach to datasets with semantically neutral phrases spoken in different languages by non-professional speakers.

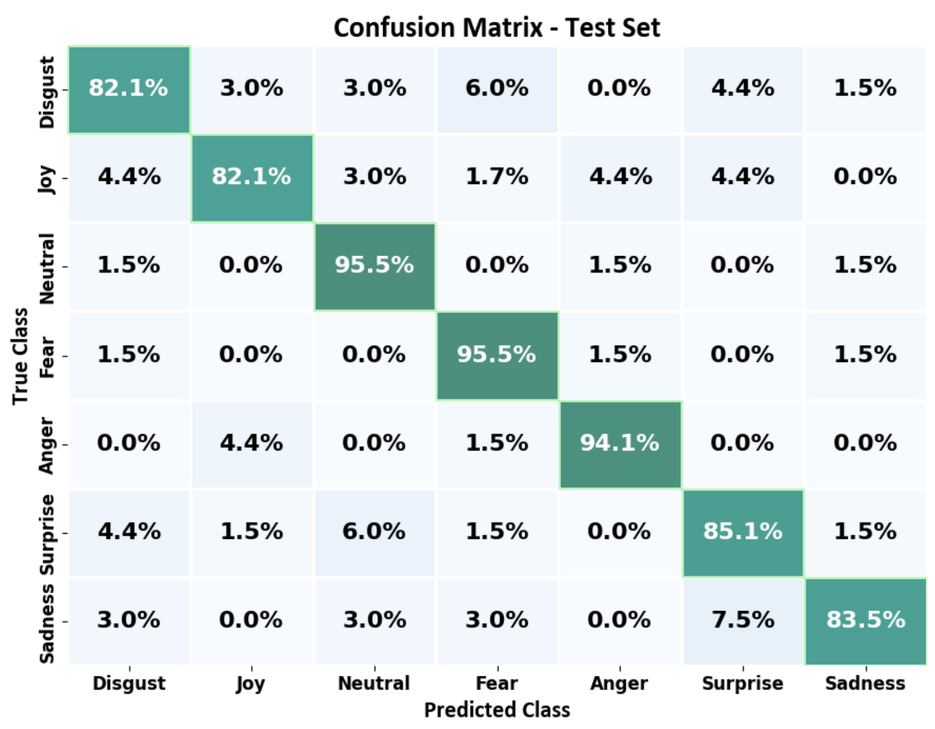

Figure 2. Confusion Matrix for the testing set.

CONCLUSION

This study demonstrates the effectiveness of PNCC and stacked classifiers in emotion recognition from semantically neutral vocal recordings, achieving an 88% accuracy on the test set and exploring new avenues of research in the field of vocal analysis.

REFERENCES

[1]. C. Kim and R. M. Stern, «Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition», doi: 10.1109/TASLP.2016.2545928
[2]. Di Cesare, M.G., Perpetuini, D., Cardone, D., Merla, A. «Unveiling Age-Related Patterns in Vocal Expression of Emotions: A Machine Learning Approach with Mel and Gammatone Frequency Cepstral Coefficients», doi: 10.1007/978-3-031-61628-0_40