



GUJARAT TECHNOLOGICAL UNIVERSITY

(Established by Government of Gujarat under Gujarat Act No. : 20 of 2007)

ગુજરાત ટેકનોલોજીકલ યુનિવર્સિટી

(ગુજરાત સરકારના ગુજરાત અધિનિયમ ક્રમાંક : ૨૦/૨૦૦૭ દ્વારા સ્થાપિત)

Accredited with A++ Grade by NAAC

Abstract of the Thesis



Name: Prajapati Yogeshkumar Jethabhai

Enrollment No.: 219999913019

Faculty: Engineering and Technology

Discipline/Branch: COMPUTER/IT Engineering

Title of the Thesis: Design and Implementation of Specific Novel
Architecture for Bilingual Speech Emotion Recognition

Abstract

The research presents an alternative bilingual speech emotion recognition system that combines patch-encoded VGG-16 spectral cues and a complete Vision Transformer pipeline to train the affective cues on English and Gujarati speech. Mel-spectrograms are first fed into a frozen VGG-16 backbone to extract high-level spatial spectral features, and the features are further divided into regular patches and mapped to an embedding space to be employed to represent it in a transformer-based global attention model. It is tested on four reference English emotional speech datasets, namely, RAVDESS, CREMA-D, SAVEE, and TESS, and all have accuracy of 99%. To test the strength on data on non-controlled, a hand-collected bilingual corpus of student speech was developed, with the model achieving the accuracy of English and Gujarati speech on 90 percent and 88 percent respectively. These results indicate that the contextual learning with convolutional spectral extraction by transformers is an effective method of learning cross-lingual variation in emotions and performing better than the conventional convolution-based or transformer-based models. The bilingual findings also indicate that it is possible to stabilize the model whereby languages with differing phonetics and prosody are used and so is applicable in scalable and inclusive emotion sensitive speech technology practicing through interactive assistants, Call-Centre analytics and affect sensitive human-machine interfaces.



GUJARAT TECHNOLOGICAL UNIVERSITY

(Established by Government of Gujarat under Gujarat Act No. : 20 of 2007)

ગુજરાત ટેકનોલોજીકલ યુનિવર્સિટી

(ગુજરાત સરકારના ગુજરાત અધિનિયમ ક્રમાંક : ૨૦/૨૦૦૭ દ્વારા સ્થાપિત)

Accredited with A++ Grade by NAAC

This PhD Thesis will help in

This PhD thesis would be useful for advancing bilingual speech emotion recognition systems for English and Gujarati languages using deep learning. It can support the development of intelligent human computer interaction, virtual assistants, call center analytics, mental health monitoring, e-learning platforms, and emotion-aware communication systems by providing accurate and scalable cross-lingual emotion recognition.

List of Publication(s):

1. Yogeshkumar J. Prajapati, Dr. Priyesh P. Gandhi and Dr. Shesang Degadwala, "A Review - ML and DL Classifiers for Emotion Detection in Audio and Speech Data," 2022 International Conference on Inventive Computation Technologies (ICICT), Nepal, 2022, pp. 63-69, doi: <https://doi.org/10.1109/ICICT54344.2022.9850614>.
2. Y. J. Prajapati, Dr. P. Ghandhi, and Dr. S. D. Degadwala, Trans., "Systematic Evaluation of Deep Learning Paradigms for Speech Emotion Recognition Using Diverse Audio Sources", Int J Sci Res Sci & Technol, vol. 12, no. 3, pp. 1318–1330, Jun. 2025, doi: <https://doi.org/10.32628/IJSRST25123148>.
3. Yogeshkumar Prajapati, Priyesh Gandhi, & Sheshang Degadwala. (2025). Bilingual Spectral Emotion Learning Through Patch-Encoded VGG-16 Features and a Full Vision Transformer Pipeline. Journal of Computing & Biomedical Informatics, 10(01). <https://doi.org/10.56979/1001/2025/1146>.

Patents

1. Yogeshkumar Jethabhai Prajapati, Priyesh Gandhi, and Sheshang Degadwala, "Voice emotion interpretation device," Indian Design No. 422532-001, Jul. 07, 2024.
2. Yogeshkumar Jethabhai Prajapati, Priyesh Gandhi, and Sheshang Degadwala, "Emotion detecting smart microphone device," Indian Design No. 447721-001, Feb. 11, 2025.