Postgraduate Seminar Series

Venue: Graduate Seminar Room
Date & Time: September 19, 2026 at 2:30 PM

Speaker Information

Mahuwa Paul (Std No. 1018052097) is a part-time M.Engg. student in the Department of CSE, BUET. She completed her undergraduate studies from Rajshahi University of Engineering & Technology (RUET) in 2017. She is currently working as an Assistant Manager, IT and Communication, at Nuclear Power Plant Company Bangladesh Limited (NPCBL). Her research interests lie in the field(s) of Machine Learning, Deep Learning, and Speech Emotion Recognition. She is currently conducting her postgraduate thesis under the supervision of Dr. Mahmuda Naznin. She will be speaking about her ongoing research in this talk.

GEN-SER: Leveraging Transfer Learning for Gender Dependent Speech Emotion Recognition Model

Speech Emotion Recognition (SER) aims to automatically identify a speaker's emotional state from speech and has important applications in human–computer interaction, healthcare, education, customer service, and affect-aware intelligent systems. Despite significant progress, SER models often experience performance degradation when training and evaluation conditions differ in language, corpus, speaker characteristics, and recording environment. This research presents GEN-SER, a gender-dependent Speech Emotion Recognition framework based on transfer learning for single-domain, multi-domain, and cross-lingual evaluation. Six widely used emotional speech corpora—CREMA-D, Emo-DB, RAVDESS, SAVEE, TESS, and BASER—are considered, covering English, German, and Bangla speech. MFCC, Chroma, and Spectral Contrast features are extracted and fused to construct the acoustic representation, while time stretching and pitch shifting are employed for data augmentation. The fused features are normalized and transformed into a 128×128×3 VGG16-compatible representation. A pretrained VGG16 network with a frozen convolutional backbone is used for transfer-based feature extraction, followed by task-specific dense layers and gender-dependent emotion classification using available speaker-gender information. Experimental results demonstrate that gender-dependent modeling improves Unweighted Average Recall (UAR) across the directly comparable mixed-gender corpora. In the multi-domain experiment, UAR increases from 93.92% to 96.59%, corresponding to an absolute improvement of 2.67 percentage points. For CREMA-D, Emo-DB, RAVDESS, and BASER, gender-dependent modeling provides absolute UAR improvements of 3.45, 2.11, 2.98, and 1.11 percentage points, respectively. The VGG16-based approach also outperforms the conventional CNN baseline across all evaluated single-domain corpora, with an average absolute UAR improvement of approximately 9.55 percentage points. Cross-lingual experiments involving English-to-Bangla, English-to-German, and Bangla-to-German conditions further demonstrate that emotion recognition performance varies substantially across source-target combinations and target-domain data conditions. Overall, the findings demonstrate the effectiveness of combining fused acoustic features, VGG16-based transfer learning, multi-domain data, and gender-dependent modeling for SER under the evaluated experimental conditions.