NATIONAL INSTITUTE OF TECHNOLOGY
TIRUCHIRAPPALLI – 620 015

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CS677: M.TECH. PROJECT WORK

ZEROTH REVIEW - REPORT

1. Name of the student: Shivendu Mishra

2. Name of the Guide: Dr. R. Bala Krishnan

3. Broad area of the Project: Deep Learning, Affective Computing

4. Sub-domain of the Project: Emotion Recognition, Multimodal Data Fusion

5. Tentative Title of the Project: Emotion-Aware Intelligent System Using Transformer-Based Multimodal Data Fusion

6. Proposed Abstract: Unlike traditional unimodal approaches, the system integrates complementary modalities from datasets such as WESAD, DREAMER, etc., to capture diverse physiological and contextual cues for improved accuracy. The proposed methodology involves preprocessing and synchronizing multiple modalities (e.g., ECG, EDA, respiration, EEG, facial features) and extracting modality-specific representations. These are fused using a Transformer framework with multi-head self-attention to learn inter-modality dependencies and long-range temporal relationships.

- Problem Statement and Objectives: To create a robust, context-aware emotion recognition system. The objective of this project is to develop an advanced emotion recognition system using multimodal physiological and behavioural signals, with a strong focus on Transformer-based architectures for feature fusion and temporal modelling.

- **Justification for the problem:** Accurate emotion recognition is essential for applications in healthcare, adaptive learning, and human–computer interaction. Emotions are dynamic, context-dependent, and expressed through multiple physiological and behavioural cues. Leveraging advanced architectures like Transformers can enable more precise modelling of these complex patterns.

- **Existing Solution:** Existing systems range from unimodal models to basic multimodal frameworks. While unimodal models capture limited cues, current multimodal approaches often depend on simple feature fusion or sequential processing, which struggle with complex inter-modality interactions and adapting to diverse, real-world environments.

- **Gaps in the existing solution:** There is a lack of frameworks that can simultaneously integrate diverse data sources, model intricate relationships between them, and capture long-term temporal patterns. This gap limits the robustness, adaptability, and accuracy of emotion recognition systems in practical, real-time scenarios.

- **Gaps to be addressed in this project:** This project addresses the absence of a robust, unified framework that can seamlessly integrate multiple physiological and behavioural signals, capture complex inter-modality relationships, and model long-range temporal dependencies

- **How to fill up the gap?:** This project addresses the limitations of unimodal and simple-fusion multimodal systems by developing a Transformer-based framework capable of capturing complex inter-modality relationships and long-range temporal dependencies for robust, real-world emotion recognition.

- **Implementation environment:** The project will be implemented in Python 3.8 using PyTorch 2.0 as the primary deep learning framework for model development and training. Supporting libraries will include NumPy for numerical computations, Scikit-learn for evaluation metrics and utilities, and Matplotlib and Seaborn for visualization. The models will be trained on an NVIDIA GPU with CUDA support to accelerate computations. Development and experimentation will be conducted in Jupyter Notebook to enable interactive coding, analysis, and result documentation.

- **Performance Metrics:** The performance of the proposed UAED model will be evaluated using four key metrics: Accuracy (ACC), which measures the proportion of correctly classified samples; Precision (PRE), which indicates the proportion of correctly predicted positive cases among all predicted positives; Recall (REC), which reflects the proportion of correctly predicted positive cases among all actual positives; and the F1-score (F1), which is the harmonic mean of precision and recall, providing a balanced measure of both. These metrics collectively offer a comprehensive assessment of the model's classification effectiveness and enable fair comparison with baseline approaches.

7. Learning Outcomes of the project: Upon completion, this project will yield proficiency in:

• Advanced concepts of affective computing and its role in emotion-aware intelligent systems.

• Synchronization of multimodal physiological and behavioural signals such as ECG, EDA, respiration, EEG, and facial features.

• Design and implementation of Transformer-based architectures for multimodal feature fusion and temporal sequence modelling.

• Integration and utilization of benchmark affective computing datasets as WESAD, DREAMER, etc., for system development and validation.

• Translation of research outcomes into a real-time deployable emotion recognition system.

8. References:

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "*Bidirectional Encoder Representations from Transformers (BERT): Pre-training of deep bidirectional transformers for language understanding*," arXiv preprint arXiv:1810.04805, 2018.

[2] M.-G. Kim and S. B. Pan, "*Deep learning based on 1-D ensemble networks using ECG for real-time user recognition*," **IEEE Transactions on Industrial Informatics**, vol. 15, no. 10, pp. 5656–5663, Oct. 2019.

[3] P. Sarkar and A. Etemad, "Self-supervised ECG representation learning for emotion recognition," **IEEE Transactions on Affective Computing**, vol. 13, no. 3, pp. 1541-1554, Jul.–Sep. 2022.

[4] A. Gu, K. Goel, and C. Ré, "*Efficiently modelling long sequences with structured state spaces*," in Proc. Int. Conf. Learn. Representations, 2022.

[5] J. Zhu, F. Deng, J. Zhao, D. Liu, and J. Chen, "*Unsupervised Abnormal Emotion Detection (UAED): Unsupervised abnormal emotion detection network based on wearable mobile device*," **IEEE Transactions on Network Science and Engineering**, vol. 10, no. 6, pp. 3682–3696, Nov./Dec. 2023.

Signature of student **Signature of Guide with date**