Format Prepared by: Sudipto Chaki, Assistant Professor, CSE, BUBT

Department of Computer Science and Engineering

Bangladesh University of Business and Technology (BUBT)

CSE 498A: Literature Review Records

Student's Id and Name	22234103382, Umma Sumaiya Riya
Capstone Project Title	High-Efficiency Micro-Expression Recognition for Automated Human Behavior Analysis Systems.
Supervisor Name & Designation	Md. Mijanur Rahman, Assistant Professor
Course Teacher's Name & Designation	Sudipto Chaki, Assistant Professor

Aspects	**Paper # 21 HMRM: A Hybrid Motion and Region-Fused Mamba Network for Micro-Expression Recognition [Published Year: 2025 Publisher: Digital Signal Processing (Elsevier)]**
Problem Statement	The paper addresses the challenge of accurately recognizing facial micro-expressions, which are extremely subtle, low-intensity, and short-duration facial muscle movements. Existing MER methods struggle with (i) capturing fine-grained motion details, (ii) modeling long-range spatiotemporal dependencies efficiently, (iii) limited and imbalanced datasets, and (iv) high computational cost of optical flow and Transformer-based models. The authors aim to design a lightweight but highly accurate MER framework that enhances motion representation and effectively models region-specific facial dynamics.
Key Contributions	The core contributions of this paper are:

	1)Proposes a **Hybrid Motion Feature Augmentation (HMFA)** module combining GRU-attention-based optical flow estimation and MotionMix regional motion augmentation. 2)Introduces a **Grained Mamba Encoder** for efficient long-range spatiotemporal modeling using selective state space models. 3)Designs a **Regions Feature Fusion Strategy (RFFS)** that integrates multi-scale regional features for improved discrimination. 4)Achieves **state-of-the-art MER performance** on CASME II, SAMM, and composite datasets while maintaining low computational cost.
Methodology/Theory/Framework	The authors propose the **HMRM framework**, which consists of: ● **GRU-Attention Optical Flow Estimation (GRU-AOFE):** Generates high-quality, noise-suppressed optical flow using correlation volumes and attention-guided GRU updates. ● **MotionMix Enhancement:** Performs landmark-guided patch mixing (eyes & mouth) between two same-class flow maps to increase motion diversity. ● **Grained Mamba Encoder:** Splits flow maps into patches, encodes them using a bidirectional Mamba module with convolutional gating, and extracts multi-scale temporal dependencies. ● **RFFS:** Combines coarse and fine region features using multi-head self-attention. The final fused features are classified using a fully connected layer with a multi-scale weighted cross-entropy loss.
SoftwareTools/Setup Details	**OS:** Ubuntu 20.04.1 **GPU:** NVIDIA RTX 4090 **CPU:** Intel Xeon Gold 6271C **Frameworks:** Python, PyTorch **Optical Flow Pretraining:** FlyingChairs & FlyingThings datasets **Landmark detection:** MTCNN **Face alignment:** Dlib 68-point landmarks

Test/Experiment Analysis	Experiments used LOSO protocol with UF1 and UAR metrics on CASME II, SMIC-HS, SAMM, and Composite datasets. Training used AdamW, LR 0.0005, 1000 epochs, and a Mamba Encoder (192 dim, depth 4). Optical flow was refined using GRU-attention, and model efficiency was measured by parameters/FLOPs**.** **Accuracy:** ● CASME II: 0.9561 / 0.9588 ● SAMM: 0.8909 / 0.9017 ● Composite: 0.8788 / 0.8906 ● SMIC-HS: 0.7491 / 0.7759
Test Data/Dataset Source	CASME II Dataset — 247 spontaneous micro-expression video samples; 200 fps; resolution 280×340. SMIC-HS Dataset — 164 high-speed micro-expression samples; 100 fps; resolution 150×190. SAMM Dataset — 159 high-resolution micro-expression samples; resolution 2040×1088.
Final Result (Assessment Criteria Wise)	● CASME II: UF1 = 0.9561, UAR = 0.9588 ● SAMM: UF1 = 0.8909, UAR = 0.9017 ● Composite Dataset: UF1 = 0.8788, UAR = 0.890 ● SMIC-HS: UF1 = 0.7491, UAR = 0.7759 (2nd best) HMRM achieves the best overall performance across most datasets, surpassing Transformer, CNN, and GNN-based methods. It also shows strong efficiency with significantly fewer parameters compared to other high-performing models.
Limitations (List the limitations the authors mentioned in the article)	The model performs poorly on low-resolution data and is limited by small, imbalanced datasets. It cannot handle global head movement, and apex frame approximation may add noise. It also lacks explainability and uncertainty estimation
Final Summary	S. Guo et al. introduced HMRM, a lightweight micro-expression recognition model combining enhanced motion features, region-based modeling, and efficient Mamba-based sequence encoding. It uses GRU-attention optical flow and MotionMix to improve motion quality, while the fusion strategy strengthens regional dynamics. The model achieves state-of-the-art results

	on CASME II, SAMM, and composite datasets using LOSO evaluation. Although highly effective on high-resolution data, performance drops on low-resolution datasets and the system lacks explainability. Overall, HMRM is a strong and efficient MER framework.