

**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**

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

<b>Aspects</b>	<p><b>Paper # 24 HMRM: A Hybrid Motion and Region-Fused Mamba Network for Micro-Expression Recognition)</b> <b>[Published Year: 2025]</b> <b>Publisher: Elsevier (Digital Signal Processing)  </b></p>
<b>Problem Statement</b>	<p>Micro-expression recognition (MER) is extremely difficult because facial micro-expressions are low-intensity, short-duration, and easily affected by noise and low motion amplitude. Existing MER models struggle to capture subtle motion, handle limited datasets, and model long-range temporal dependencies efficiently. The authors aim to build a lightweight but highly accurate MER framework that can capture fine motion, enhance data diversity, and fuse region-specific features effectively..</p>
<b>Key Contributions</b>	<p>The core contributions of this paper are:</p> <p>(i) a Hybrid Motion Feature Augmentation (HMFA) module combining a GRU-Attention Optical Flow Estimator and MotionMix augmentation,</p>

	<p>(ii) a Grained Mamba Encoder for efficient sequence modeling with low computational cost,</p> <p>(iii) a Region Feature Fusion Strategy (RFFS) that fuses multi-scale regions for better discrimination, and</p> <p>(iv) achieving state-of-the-art performance on multiple MER benchmark datasets.</p>
<b>Methodology/Theory/Framework</b>	The authors proposed the HMRM framework, consisting of a GRU-Attention Optical Flow Estimator to generate clean and high-quality motion features, followed by MotionMix augmentation for region-level motion mixing. The processed flow maps are then encoded using a Grained Mamba Encoder, which models long-range dependencies efficiently. Finally, a Region Feature Fusion Strategy integrates coarse and fine region-level features using multi-head attention, and classification is performed with a Multi-Scale Weighted Cross-Entropy loss.
<b>Software Tools/Setup Details</b>	Ubuntu 20.04.1 operating system, Intel Xeon Gold 6271C CPU, NVIDIA RTX 4090 GPU, Python, PyTorch framework, Dlib for landmark extraction, and MTCNN for face alignment. GRU-AOFE was pretrained on FlyingChairs and FlyingThings datasets
<b>Test/Experiment Analysis</b>	<p>Experiments were conducted using the Leave-One-Subject-Out (LOSO) protocol on CASME II, SMIC-HS, SAMM, and the Composite Dataset. Evaluation metrics included UF1 and UAR.</p> <p>Training used AdamW optimizer, learning rate 0.0005 with dynamic scheduling, weight decay 0.01, 1000 epochs, and Mamba settings: embedding dimension 192, depth 4, state dimension 16. Region-based analysis was performed using coarse (4-region) and fine (16-region) partitions. GRU-AOFE and MotionMix showed strong improvements in ablation experiments.</p> <p>Accuracy:</p> <ul style="list-style-type: none"> <li>• CASME II: 0.9561 UF1, 0.9588 UAR</li> <li>• SAMM: 0.8909 UF1, 0.9017 UAR</li> <li>• Composite: 0.8788 UF1, 0.8906 UAR</li> <li>• SMIC-HS: 0.7491 UF1, 0.7759 UAR</li> </ul>

<b>Test Data/Dataset Source</b>	CASME II: 247 samples (200 fps, 280×340) SMIC-HS: 164 samples (100 fps, 150×190) SAMM: 159 samples (high-resolution 2040×1088) Composite Dataset: MEGC2019-based merged dataset with 442 samples (Negative, Positive, Surprise classes)
<b>Final Result (Assessment Criteria Wise)</b>	<ul style="list-style-type: none"> <li>The model achieved state-of-the-art performance in MER using motion enhancement (GRU-AOFE, MotionMix), efficient sequence modeling (Mamba), region-aware fusion (RFFS), and multi-scale loss design. The framework demonstrated strong accuracy, robust region-based discrimination, improved motion quality, and computational efficiency across all datasets.</li> </ul>
<b>Limitations</b> (List the limitations the authors mentioned in the article)	<ul style="list-style-type: none"> <li>The model performs poorly on low-resolution datasets, relies on accurate face alignment and apex frame detection, and uses fixed region partitions. It has no explainability module, limited real-world evaluation, and relies on landmark accuracy.</li> </ul>
<b>Final Summary</b>	S. Guo et al. proposed HMRM, a motion-enhanced and region-fused micro-expression recognition model combining GRU-Attention optical flow, MotionMix augmentation, and a Grained Mamba Encoder. The framework efficiently captures subtle facial motions while maintaining low computational cost. Using LOSO evaluation, the model achieved state-of-the-art accuracy on CASME II, SAMM, and the Composite dataset. Region Feature Fusion further improved fine-grained discrimination, and ablation studies verified the importance of each module. However, performance decreases on low-resolution data, and the model relies heavily on accurate apex frame detection and landmarks. The system lacks explainability and real-world testing. Overall, HMRM is a highly effective and efficient MER framework offering strong motion modeling and region-based learning.