Postgraduate Seminar Series

Venue: Graduate Seminar Room
Date & Time: November 21, 2026 at 2:30 PM

Speaker Information



Shahrin Khan (Student ID: 1024054005) is a part-time PhD student in the Department of Computer Science and Engineering (CSE), BUET. She is also a Lecturer in CSE, Daffodil International University. She completed her BSc and MSc in CSE at Jahangirnagar University. Her research interests include Artificial Intelligence (AI), Machine Learning (ML), Computer Vision (CV), and Image Processing. She is currently pursuing her Doctor of Philosophy under the supervision of Dr. Sadia Sharmin. She will be speaking about her ongoing research in this talk.

FlashViT: A Memory-Efficient and Explainable Vision Transformer for Multi-Class Sweet Orange Disease Classification

Accurate and computationally efficient plant disease detection is essential for scalable precision agriculture systems. While convolutional neural networks (CNNs) have achieved strong performance in crop disease classification, their localized receptive fields may limit modeling of long-range spatial dependencies in complex field environments. Vision Transformers (ViTs) address this limitation through global self-attention; however, the quadratic memory complexity of standard attention mechanisms can restrict practical deployment, particularly in resource-constrained agricultural settings. This study presents FlashViT, an efficiency-aware Vision Transformer framework for multi-class sweet orange disease classification applied to a dataset of 5,675 images spanning 11 disease categories. FlashViT integrates FlashAttention-2 (via PyTorch's scaled_dot_product_attention) and a mid-network Token Pruning module within a ViT-Small/8 backbone, reducing the active token sequence from 785 to 589 tokens in the deeper encoder blocks while preserving exact attention computation in the full sequence. The dataset was stratified into training (70%), validation (15%), and testing (15%) partitions prior to model development. Class imbalance was addressed exclusively within the training partition using weighted sampling and class-weighted loss, ensuring strict prevention of data leakage. Comprehensive evaluation includes held-out testing, five-fold cross-validation, Bonferroni-corrected paired statistical testing, Matthews Correlation Coefficient (MCC) analysis, and computational efficiency benchmarking. Experimental results demonstrate that FlashViT achieves high multi-class discrimination performance (99.88% on the original test set and 99.92% under balanced training, using a keep ratio selected via a held-out validation split and evaluated once on the untouched test set) and low variance across folds, indicating stable generalization. An external validation study on an independent citrus dataset with a different class taxonomy shows a notable generalization gap (87.13% accuracy), highlighting the value of dataset-specific adaptation before deployment. Statistical analysis confirms significant improvements over GlobalViT, LocalViT, and ResNet101 after Bonferroni correction ($p < 0.0125$), while the improvement over EfficientNet-B5 does not reach significance at the corrected threshold; given the limited number of cross-validation folds ($n = 5$), these statistical comparisons should be regarded as preliminary rather than conclusive evidence of general superiority. Efficiency analysis shows that FlashViT achieves approximately 33% higher inference throughput compared to standard GlobalViT and substantially lower memory consumption compared to window-based LocalViT, while maintaining near-equivalent accuracy across all transformer variants. Grad-CAM heatmaps are spatially consistent with expected disease regions, supporting model interpretability. These findings suggest that IO-aware attention with token reduction can preserve predictive robustness while offering practical efficiency advantages, supporting deployment-aware transformer-based solutions for agricultural monitoring.