Title: Enhancing Multimodal Reasoning for Crisis Event Analysis with Vision Language Models and Attention Modules

PRESENTER:
Nusrat Munia

BACKGROUND:

- Social media provides real-time crisis information, but posts vary in informativeness and clarity.
- Timely identification of relevant crisis posts can support humanitarian response, resource allocation, and situational awareness.

METHODS

- **Caption Augmentation:** Use LLaVA to generate detailed, image-grounded captions for tweets.
- **Cross-Feature Module (CFM):** Fuse original tweet text with generated captions via cross-attention.
- **Guided Cross Attention:** Fine-grained alignment between visual and textual features.
- **Decision Module:** Apply Differential Attention over the fused features, then a classification head to produce the final prediction.

RESULTS:

- CapFuse-Net achieved SOTA performance across all splits on CrisisMMD dataset.
- Original split: Predefined train/val/test partition.
- Stratified split: Balances class distributions, reducing bias.
- Event-wise split: No overlap of disaster events, testing generalization.

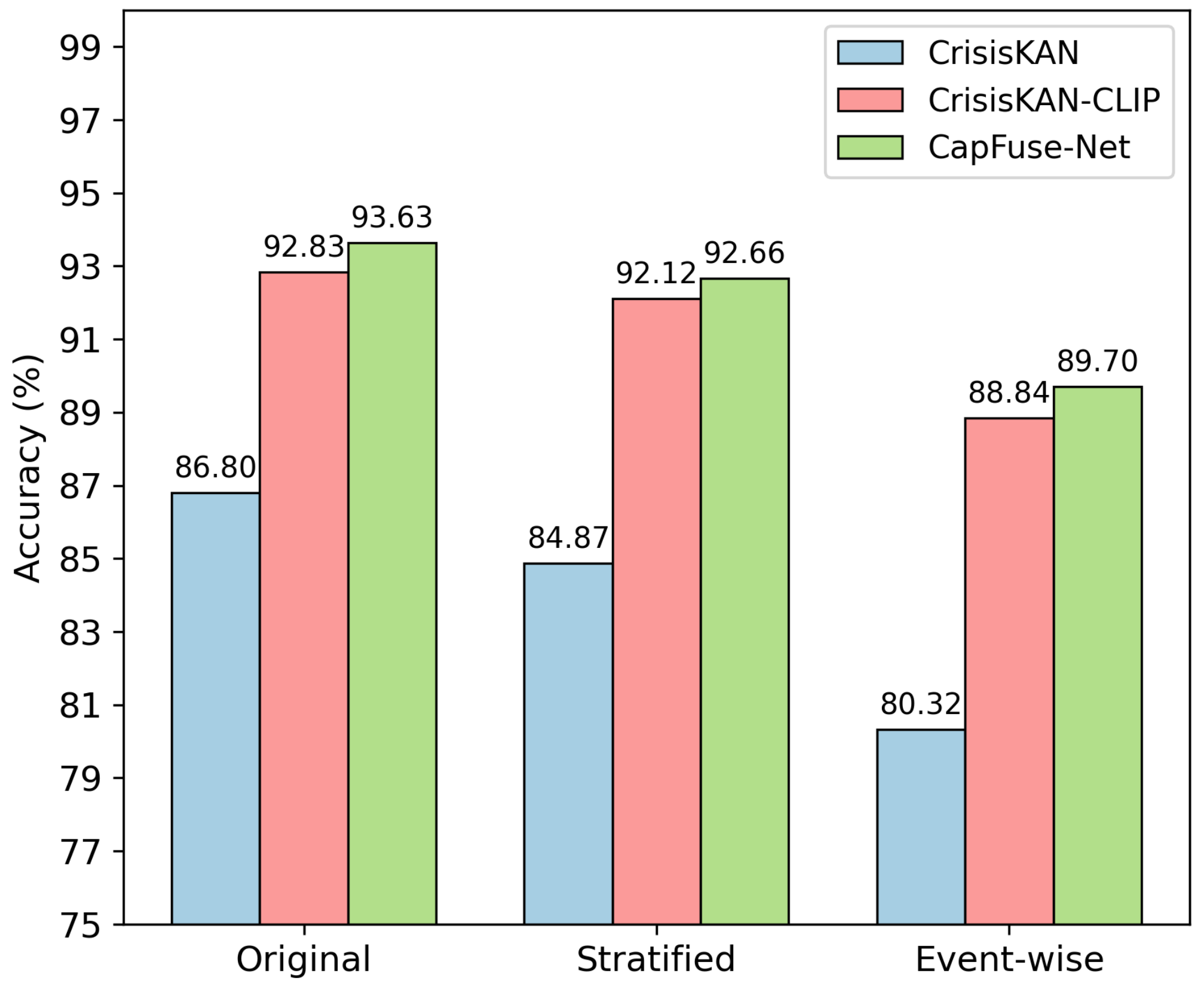

Fig. : Task 1 — Informativeness classification accuracy.

This material is based upon work supported by the National Science Foundation under Cooperative Agreement No. 2344533.

Harnessing Social Media with Vision Language Models for Disaster Response

AI-powered multimodal analysis to identify and classify critical disaster information

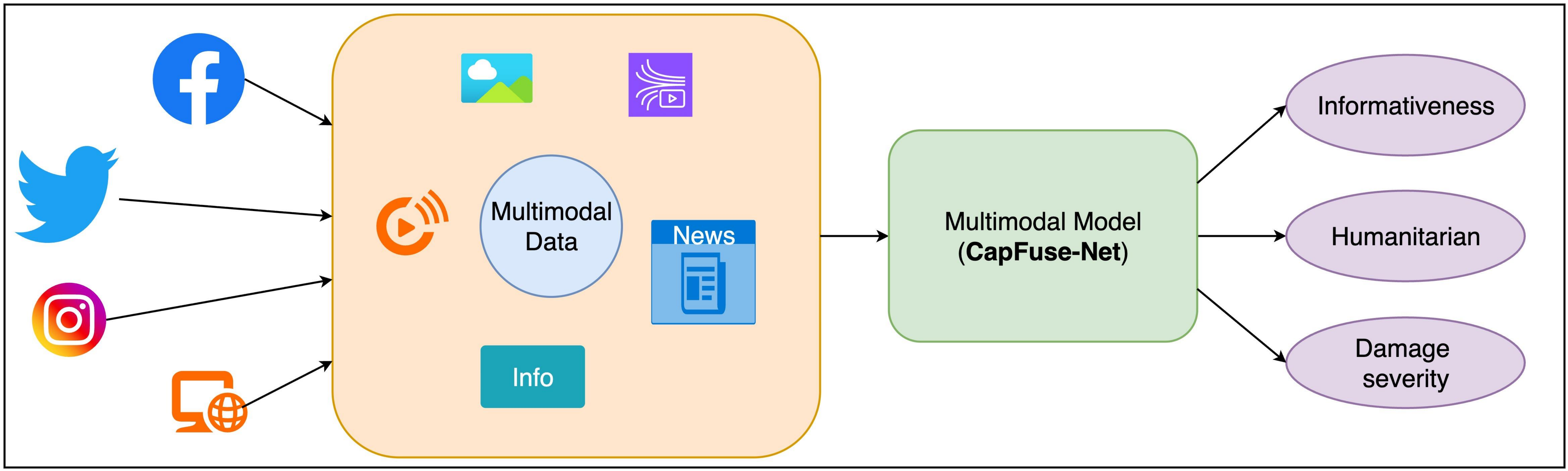

Fig.: Overview of the disaster response classification pipeline. Multimodal data from social media and news sources is processed by Multimodal Model (CapFuse-Net) to classify posts based on informativeness, humanitarian category, and damage severity.

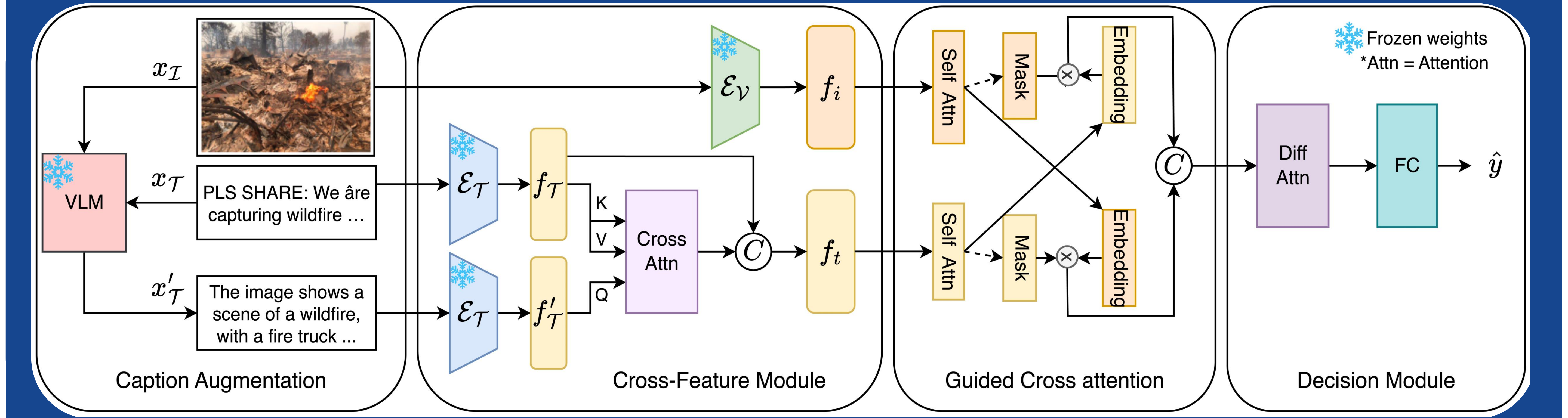

Fig.: CapFuse-Net (Caption-Augmented Multimodal Feature Fusion Network), a multimodal architecture integrating Caption Augmentation, Cross-Feature Module, Guided Cross Attention, and a Decision Module for improved vision-language understanding in disaster response tasks.

Download the short paper here

References:

- Alam, F. et al. CrisisMMD: Multimodal Twitter Datasets from Natural Disasters. ICWSM, 2018.
- Liu, H. et. al. Visual instruction tuning. Advances in Neural Information Processing Systems, 2023.
- Gupta, S. et. al. Crisiskan: Knowledge-infused and explainable multimodal attention network for crisis event classification. ECIR, 2024.

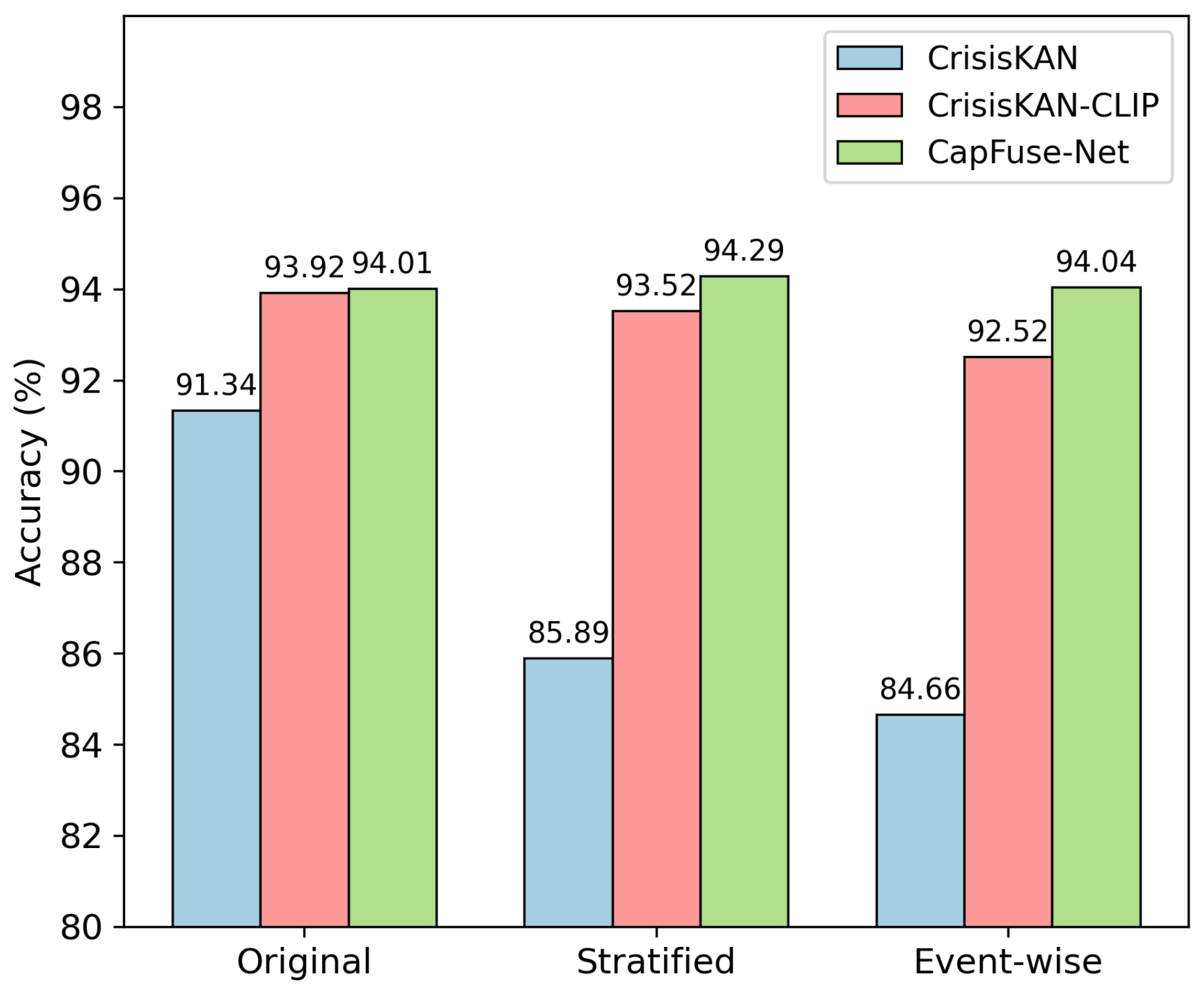

Fig. : Task 2, Humanitarian classification accuracy on the CrisisMMD dataset using original, stratified, and event-wise splits.

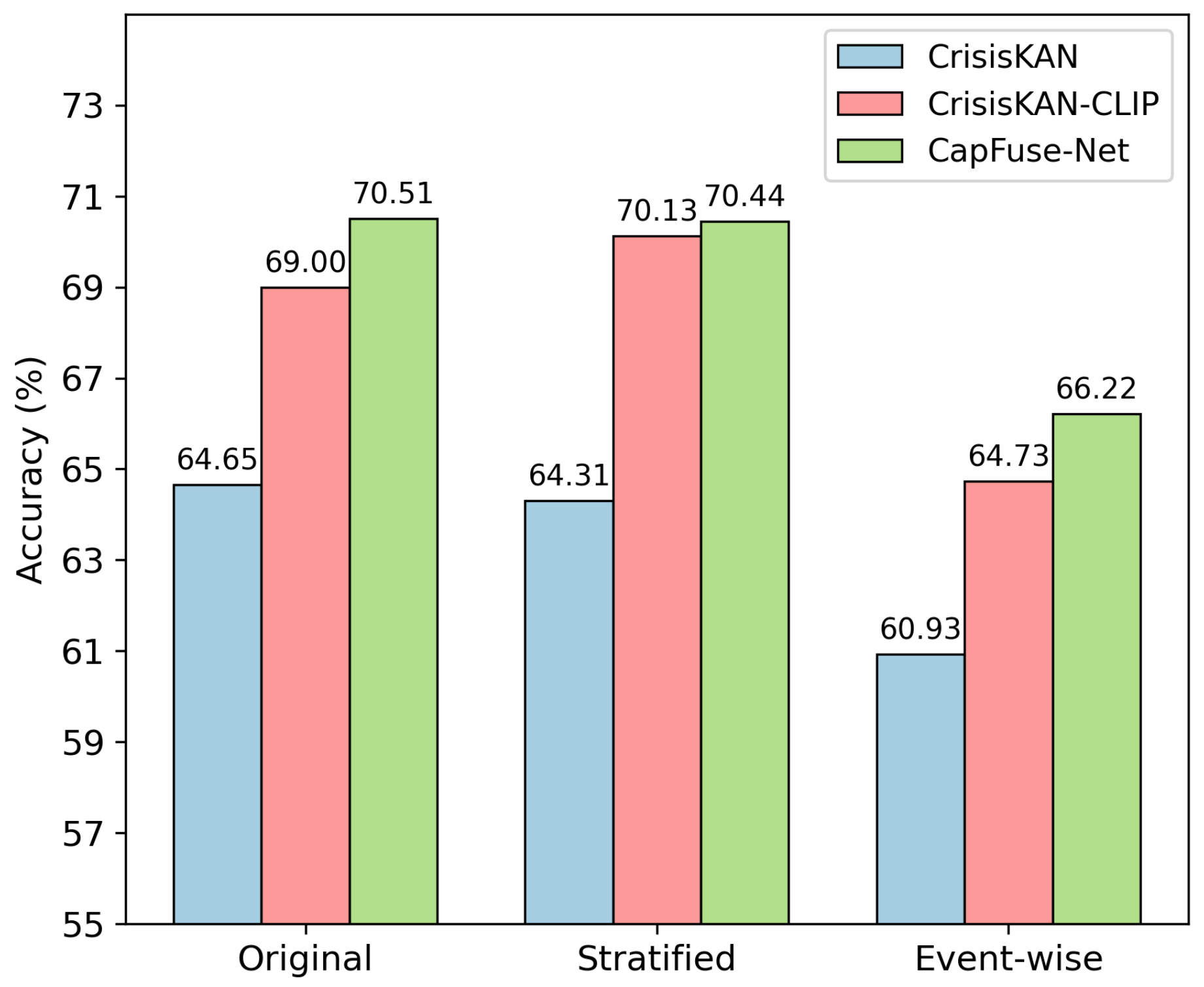

Fig. : Task 3, Damage severity classification accuracy on the CrisisMMD dataset using original, stratified, and event-wise splits.

PROJECT NUMBER:

- Project 7: Integrate Artificial Intelligence

Nusrat Munia,
Junfeng Zhu,
Olfa Nasraoui,
Abdullah-Al-Zubaer Imran

CLIMBS KENTUCKY CLIMATE RESILIENCE
KENTUCKY NSF EPSCoR ADVANCING GEOGRAPHIC DIVERSITY IN STEM
University of Kentucky
UofL