Real-Time Rumor Verification System using AI-Powered Video Surveillance Based Threat Detection

Aurchi Chowdhury[1], Raiyan Siddiqui[2], Sadia Binte Sayed[3], Sakif Naieb Raiyan[4]
Department of CSE, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh
Email: [1]2105083@ugrad.cse.buet.ac.bd, [2]atgraiyan02@gmail.com, [3]sadiasayed2020@gmail.com and [4]sakif.raiyan128@gmail.com

Background

One of the problems plaguing our country in recent times is **rumors**. The spread of false information has been on a **scary rise** in recent times, and we have next to **no reliable rumor detection systems** to counter it; certainly none which can do it swiftly and in real time. Concerned by the **mass panic and incited violence** rumors leave in its wake, we developed a solution that **combines social media text analysis with real-time surveillance footage** from government CCTV systems. This integrated approach allows us to **detect, categorize, and validate potential threats** swiftly. Even in today's turbulent political landscape, rumors continue to fuel chaos, making our system crucial for mitigating the spread of misinformation.

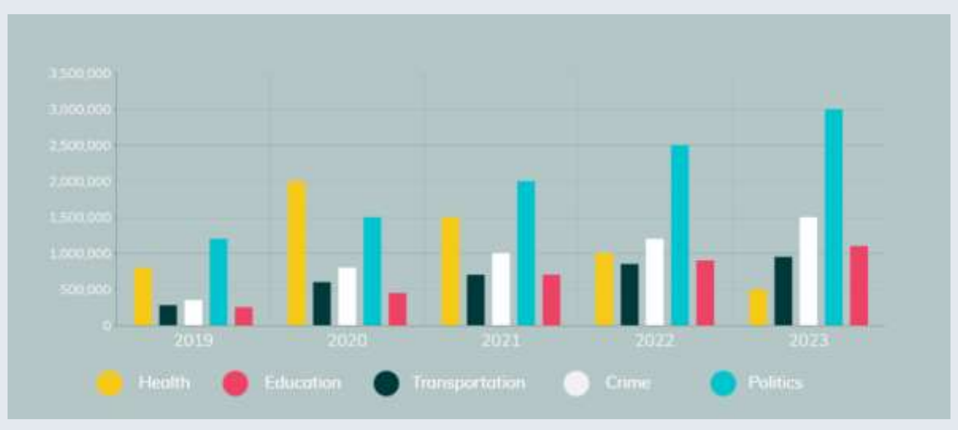

Figure 1: Prevalence of Rumor Types in Bangladesh Over the Last Five Years (2019-2023)

Motivation and Problem Formulation

- Our endeavor is fueled by the **tragic passing of Abrar Fahad**. The effectiveness of CCTV footage to dispel any and all efforts to save the perpetrators by generalizing false explanation of events caught our attention. Yet, we felt a **more proactive** approach was required in this regard.
- Most rumor detection systems nowadays are **uni-modal**, which **rely on text patterns** solely to detect fake news. That's why our approach is **multi-modal; cross referencing social media posts with the camera feed of the location and time mentioned to accurately identify any event** occurring there.
- Our **goal** is **to implement this system nationwide**. For now, we will be **using BUET's surveillance system** to refine our system and make it ready for greater scale deployment.

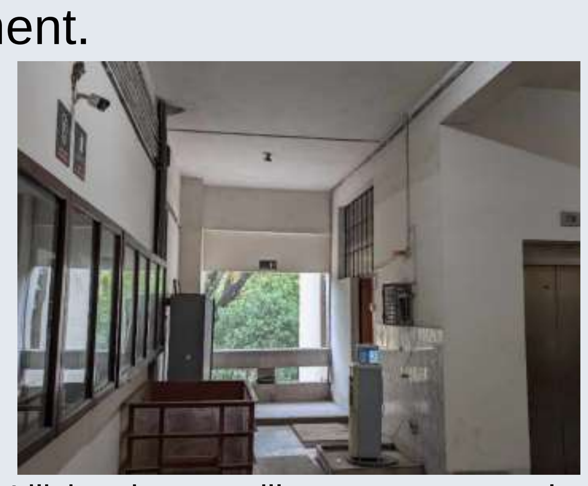

Figure 2: Utilizing the surveillance system employed at BUET

Proposed Methodology

- Our system **continuously monitors and processes** a fixed number of social media posts per hour, prioritizing them based on **two main factors**: (i) the **engagement rat**e (likes, comments, shares) and (ii) the **confidence scores** of the accounts posting.
- We classify events into **16 predefined categories**: Abuse, Burglary, Explosion, Shooting, Fighting, Shoplifting, Road Accidents, Arson, Robbery, Stealing, Assault, Vandalism**[3]**, added with **Police Brutality, Riot/Protests, Road Blockade, and Weapon Display**, plus one "safe" category indicating none of these events.
- If a post identifies a threat (e.g., "Robbery at X"), we verify it against CCTV footage to check for **matching anomalies**.
- Our **fine-tuned LLaVa NeXT Video** model converts the surveillance footage into text description.
- The **text description** data from the LLaVa model as well as the cleaned text from **social media posts** are used separately to get **classifications** from our **fine-tuned LLaMa 3.2** model.
- Accounts start with a **100% confidence score**, which decreases for sharing rumors.

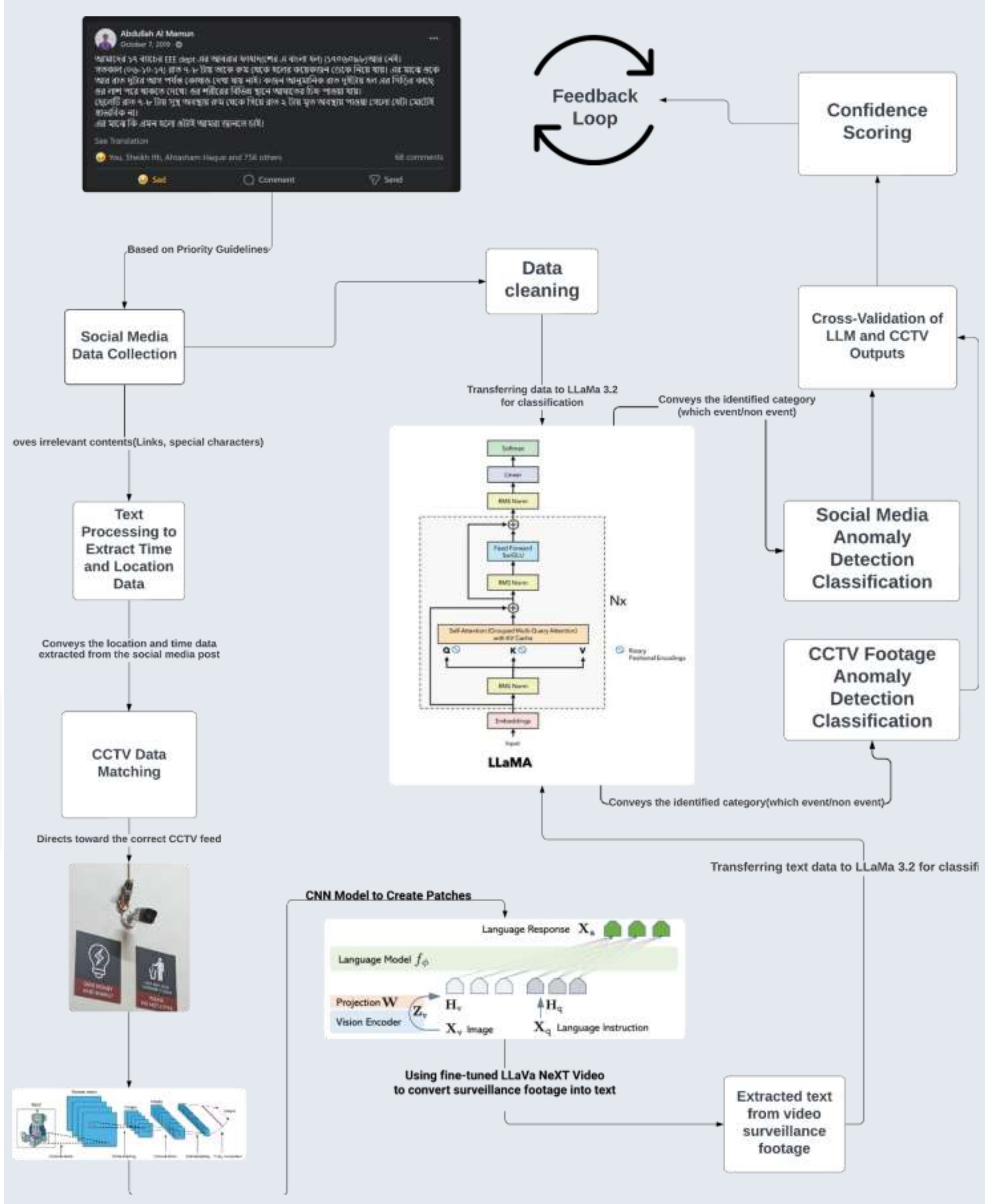

Figure 3: Integrated Framework for Social Media and CCTV-Based Anomaly Detection

Experimentation

- Our system integrates **two advanced modules**: the **multi-patched LaVA-NeXT-video** model, which **generates text descriptions from video data**, and the **LLaMA large language model (LLM)** for natural language processing.

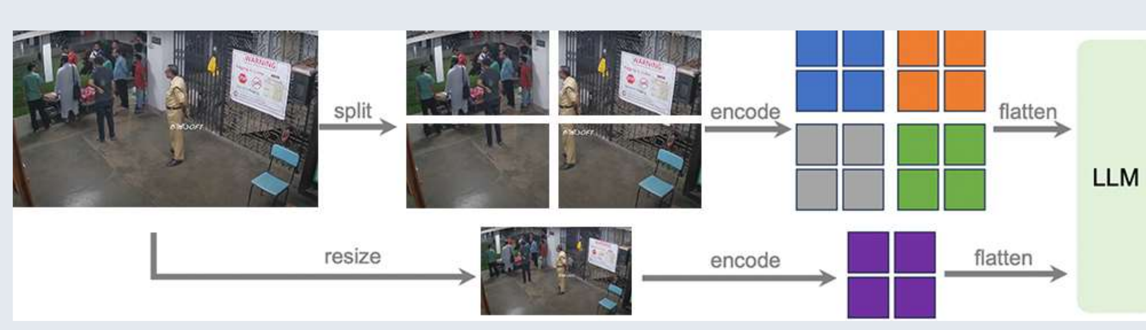

Figure 4: Multi-patched LaVa-NeXT-video model

- The LaVA-NeXT model processes CCTV footage and provides descriptive outputs, enabling real-time, automated monitoring of potential security threats.
- The **LLaMA model** further **classifies these descriptions into relevant categories**, streamlining threat identification and data organization.
- We **fine-tune our integrated system** using datasets like **the UCF Crime Dataset**, a benchmark dataset containing videos **categorized into 12 unsafe and 1 safe category**, to enhance the model's ability to detect and classify potential security risks effectively.
- We also plan to use BUET's CCTV footage to create a dataset to fine-tune our model even further

Expected Findings

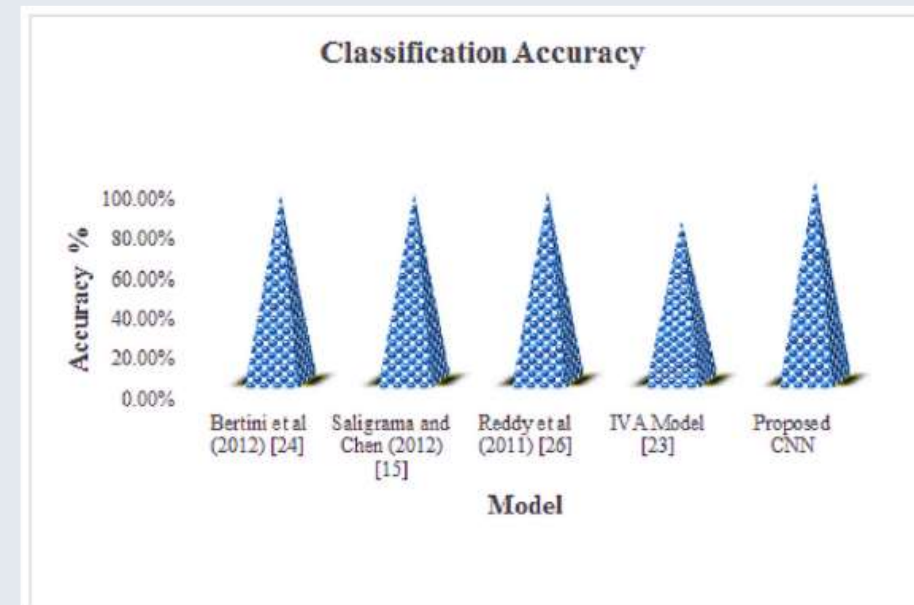

Figure 5: Comparison of Classification Accuracy Across Different Models [2]

- In existing works**[2]**, models like **VGG-16** were used for **feature extraction** from video data, **achieving the accuracy** shown in **Figure 5**.
- However, our system leverages the **more advanced multi-patched LaVA-NeXT-video model**, which **processes video frames** by dividing them into **smaller patches**, enabling **enhanced feature extraction** and generating **more accurate** and **detailed text descriptions** of visual content.
- This updated architecture is expected to yield **improved accuracy** compared to VGG-16.
- By integrating **the LLaMA model's robust NLP** capabilities and **fine-tuning** it on datasets like the UCF Crime Dataset, our system ensures **precise, context-aware threat classification, surpassing prior approaches**.

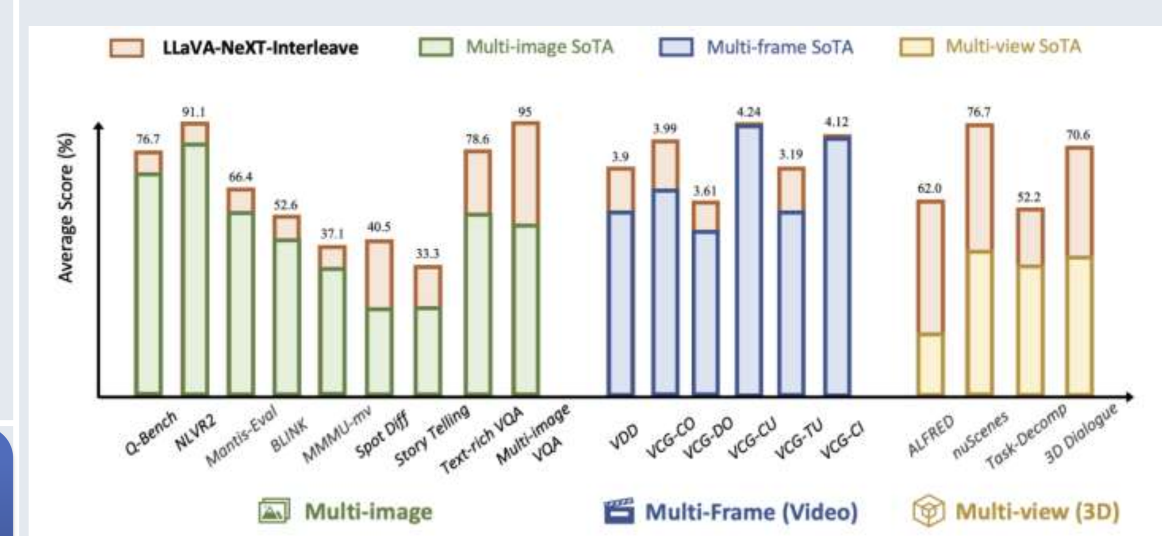

Figure 6: Performance comparison in three interleaved scenarios, including multi-image, multi-frame (video), and multi-view (3D). **[1]**

Conclusion and Future Work

Our rumor verification system, utilizing the **LLaMA model for NLP** and the **LaVA-NeXT-video model for video anomaly detection**, offers a proactive solution to **enhance campus safety at BUET**. By combining social media analysis with CCTV footage, it swiftly identifies potential threats and reduces misinformation. Future work will focus on **optimizing** these models for **BUET's multilingual data**, **ensuring ethical deployment** with strong attention to privacy and security, and **expanding implementation beyond BUET**, and eventually, **outside Bangladesh.**

References

[1] *Haotian Liu, Chunyuan Li, Linjie Zeng, Axiong Zeng, Ronghang Hu, Luowei Zhou, Chaowei Tang, Shizhe Chen, Jianfeng Wang, Zhe Chen, Pengchuan Zhang, Kai-Wei Chang, and Lu Wang.* ***Llava-next-interleave: Tackling multi-image, video, and 3d in large multimodal models****. arXiv preprint arXiv:2410.12450, 2024.*
[2] *Edeh Michael Onyema, Sundaravadivazhagn Balasubaramanian, Kanimozhi Suguna S, Celestine Iwendi, B.V.V. Siva Prasad, and Chinecherem Deborah Edeh.* ***Remote monitoring system using slow-fast deep convolution neural network model for identifying anti-social activities in surveillance applications.*** *Measurement: Sensors, 27:100718, 2023.*
[3] *Virender Singh, Swati Singh, and Pooja Gupta.* ***Real-time anomaly recognition through cctv using neural networks****. Procedia Computer Science, 173:254–263, 2020. International Conference on Smart Sustainable Intelligent Computing and Applications under ICITETM2020.*

11th NSysS 2024, December 19-21, 2024, Khulna, Bangladesh