Anomalyze: Hybrid GNN and Transformer for Crowd Anomalies

Paras Gunjyal[1], Rahul Bajaj[2], Anvay Sharma[3], Prof Rajeev Kumar[4]
Computer Science and Engineering Netaji Subhas University of Technology Delhi, India

***Abstract*—This research proposes a hybrid model integrating GNN and Transformer architectures to detect crowd anomalies in real-time. The GNN component models the interactions between individuals in the crowd, while the Transformer captures long-range dependencies across video sequences. The hybrid approach enables robust identification of abnormal crowd behaviors, such as sudden dispersal, panic, or the presence of obstacles. Evaluation on benchmark datasets such as UCSD and UMN demonstrates improved accuracy, efficiency, and generalization compared to traditional methods like CNN-LSTM and handcrafted feature-based models. This work provides a foundation for real-time surveillance systems that ensure public safety in crowded environments**

***Index Terms*—Crowd anomaly detection, Graph Neural Networks, Transformers, Hybrid Model, Deep Learning**

I. INTRODUCTION

Crowd anomaly detection plays a pivotal role in public safety, especially in environments such as stadiums, train stations, and public events where large gatherings are common. The primary goal is to identify unusual activities or behaviors that may indicate threats, such as panic, violence, or the presence of an abnormal object. Anomaly detection systems analyze crowd behaviors to differentiate between normal and abnormal patterns, offering timely interventions for law enforcement or safety personnel.

Traditional methods, such as optical flow and handcrafted feature-based approaches, often struggle to adapt to the complexities of real-world crowd scenarios. These methods rely on predefined features, making them less effective in dynamic environments with varying crowd densities, camera angles, and unpredictable behaviors. Furthermore, they fail to capture the intricate spatial- temporal relationships within a crowd, limiting their robustness and accuracy.

With the advent of deep learning techniques, especially Convolutional Neural Networks (CNNs), anomaly detection has seen significant advancements. CNNs, capable of automatic feature extraction, have revolutionized computer vision tasks. However, CNNs primarily focus on spatial relationships and are limited in capturing temporal dependencies. To overcome this, Long Short- Term Memory (LSTM) networks have been used to model sequential data, effectively analyzing temporal behavior in crowds.

This research introduces a novel hybrid approach that integrates Graph Neural Networks (GNNs) and Transformers for crowd anomaly detection. GNNs are highly effective in modeling the complex relationships between entities, while Transformers capture long-term temporal dependencies. Together, they form a robust system capable of detecting crowd The results show that the GNN-Transformer model significantly improves detection accuracy while reducing computational complexity, making it suitable for real-time applications in crowded environments.and compare it with existing techniques such anomalies in real-time with higher accuracy and generalization than existing approaches.

Img. 1. Fight-based Anomaly

A. *Motivation*

The motivation for this research stems from the following reasons:

Traditional anomaly detection methods, such as handcrafted feature extraction, optical flow, and Gaussian mixture models (GMM), have shown limitations in dynamic and diverse crowd environments. These methods often fail to generalize well across different scenarios, especially in the presence of varying crowd densities, occlusions, and unpredictable behaviors. Furthermore, such approaches typically require manual tuning of parameters, leading to inefficiencies in real- world applications.

Recent advancements in deep learning have opened up new possibilities for addressing the challenges in crowd anomaly detection. Convolutional Neural Networks (CNNs) have shown success in feature extraction, while Long Short-Term Memory (LSTM) networks have been effective in capturing temporal dynamics. However, these approaches are not without their shortcomings. CNNs focus primarily on spatial features, and LSTMs, though effective in temporal modeling, are limited when dealing with long-range dependencies and complex interactions between individuals in a crowd.

Graph Neural Networks (GNNs) and Transformers offer a promising solution to these challenges. GNNs excel in capturing relational structures and dependencies between individuals in the crowd, while Transformers are known for their ability to model long-range temporal dependencies more effectively than LSTMs. Integrating these two advanced architectures into a hybrid model for anomaly detection can address the limitations of traditional and deep learning methods by capturing both spatial and temporal relationships with greater accuracy and robustness.

II. LITERATURE REVIEW

Crowd Anomaly Detection (CAD) has become a crucial area of research due to the growing need for public safety in urban environments and large public gatherings. With the increasing frequency of events such as concerts, sporting events, religious gatherings, and protests, the ability to monitor crowd behavior in real- time is essential for avoiding accidents, managing crowd control, and preventing criminal activity. CAD systems aim to automate the detection of abnormal or dangerous behaviors in crowds using various machine learning and deep learning techniques, making it an indispensable tool for surveillance and public safety applications.

Crowd anomaly detection focuses on identifying irregular behavior within a crowd, such as sudden panic, violence, stampedes, or congestion. Traditional surveillance systems rely heavily on human operators to detect such anomalies, which is inefficient, especially in large-scale scenarios. Automated CAD systems have emerged as a solution to these limitations by using computer vision and machine learning algorithms to process video footage in real time, identifying unusual patterns that indicate potential threats.

Altowairqi et al.[1] emphasized the importance of crowd anomaly detection in their review of recent progress in the field, highlighting its applications in public safety, transportation management, and event monitoring. The increasing availability of surveillance cameras and advancements in computer vision have paved the way for CAD systems to be deployed in real-time, enhancing their effectiveness in mitigating dangerous situations.

Deep learning techniques, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have become the primary methods for detecting anomalies in crowds. CNNs are capable of capturing spatial features from images, while RNNs, specifically Long Short-Term Memory (LSTM) networks, are effective at modeling temporal dependencies, making them ideal for processing video data.

Altowairqi et al.[1] reviewed several CAD systems that utilize CNNs for extracting spatial features of crowds and LSTMs for handling temporal aspects. They demonstrated that hybrid models combining CNNs and LSTMs are particularly effective in detecting anomalies in dynamic environments, such as identifying sudden crowd movements that deviate from expected behavior.

A similar approach was explored by Khan et al.[2], who applied Generative Adversarial Networks (GANs) to enhance CAD systems' accuracy by generating synthetic training data. The GAN-based approach addresses the common issue of data imbalance, where normal crowd behaviors far outnumber abnormal ones, making it difficult for models to generalize and detect rare events like fights or panic.

One of the significant challenges in crowd anomaly detection is the class imbalance between normal and abnormal behaviors. Most crowds exhibit normal behavior for the majority of the time, making it difficult for models to detect rare anomalies effectively. To address this issue, several studies have implemented data augmentation and re-sampling techniques to balance the datasets.

Altowairqi et al.[1] discussed the application of the Synthetic Minority Over-sampling Technique (SMOTE) to generate synthetic examples of minority classes (i.e., abnormal behaviors), helping to improve the performance of CAD models on imbalanced datasets. This technique has been widely adopted in the CAD field to ensure that models can accurately detect rare anomalies without being biased towards the majority class.

Similarly, Akter and Rajbongshi[3] proposed an ensemble learning method that combines predictions from multiple classifiers, such as Random Forests and Gradient Boosting, to improve the detection of anomalies in imbalanced datasets. Their study demonstrated that ensemble models tend to perform better in detecting minority class events compared to single classifiers.

Real-time anomaly detection is critical for CAD systems to be practically deployed in crowded environments, such as concerts, transportation hubs, and sporting events. These systems must process high- resolution video feeds and detect abnormal behaviors quickly to allow for timely intervention.

Altowairqi et al. pointed out the challenges of implementing real-time CAD systems, particularly due to the high computational cost associated with deep learning models like CNNs and RNNs. They suggested using lightweight models or optimizing pre-trained models through techniques like transfer learning to reduce computational overhead while maintaining accuracy.

In a separate study, Khan et al.[2] introduced a real-time anomaly detection system that leveraged a combination of CNNs and GANs. The system was optimized using transfer learning to reduce the time required for training and inference, making it suitable for real-time applications.
The development of robust CAD models relies on large, labeled datasets that capture a wide variety of crowd behaviors. Publicly available datasets such as UCSD Pedestrian, ShanghaiTech, and the Hajj dataset have been extensively used to train and evaluate CAD systems.
Altowairqi et al. reviewed several key datasets used in CAD research, highlighting the UCSD Pedestrian dataset for studying pedestrian behaviors and the ShanghaiTech dataset for crowd counting and density estimation. These datasets are essential for training models that can detect anomalies in real-world settings, particularly in high-density environments where crowd behavior is unpredictable.
Wang et al.[4] emphasized the need for more diverse datasets that include different types of crowd behaviors, environmental conditions, and camera perspectives. They argued that current datasets are limited in scope and that future research should focus on creating more comprehensive datasets that capture various real-world scenarios.

The practical applications of CAD systems extend to public safety, transportation management, and event monitoring. CAD systems have been successfully deployed in environments like subways, stadiums, and public squares to detect abnormal behaviors, such as fights, stampedes, or suspicious activities.

Altowairqi et al discussed several challenges that remain in the deployment of CAD systems, particularly the issue of occlusion, where individuals in a crowd block each other from the camera's view. This makes it difficult for models to accurately track crowd behavior. The authors suggested using multimodal approaches, incorporating data from multiple sensors (e.g., thermal or acoustic) to improve detection accuracy in cases where visual data alone is insufficient.
Computational cost is another challenge. While deep learning models provide high accuracy, they require significant computational resources, making them impractical for real-time applications in resource-constrained environments. Akter and Rajbongshi proposed lightweight models and optimized algorithms to address this issue, suggesting that CAD systems must balance accuracy and computational efficiency to be effective in real-time deployments.

Future research in CAD is likely to focus on improving the scalability, interpretability, and real-time performance of these systems. Altowairqi et al. identified several promising directions, including the use of generative models, reinforcement learning, and explainable AI to enhance the robustness and transparency of CAD models.

Khan et al also highlighted the potential of ensemble learning and transfer learning to improve the scalability and generalization of CAD systems across different environments. They suggested that future CAD systems should be designed to adapt to various crowd scenarios, whether in urban settings, large-scale events, or transportation hubs.

B. Wang et al. proposed the integration of multimodal data (e.g., audio, video, and thermal) to improve detection accuracy in complex environments. They also emphasized the need for developing real-time, low- latency models that can be deployed on portable devices, such as drones or mobile cameras, for more flexible crowd monitoring. Nayan et al. applied optical flow to detect abnormal behavior in crowds, showing its effectiveness in dynamic environments, though it is limited by occlusion issues in dense crowds. Another traditional approach is the Gaussian Mixture Model (GMM), often used for background subtraction in video data to detect moving objects and flag irregularities. Zhan et al. utilized GMM to identify outliers in crowd movements, demonstrating its effectiveness in detecting sudden events like stampedes

For the sequential nature of crowd behavior, Hidden Markov Models (HMM) have been employed to model transitions between normal and abnormal crowd states. Mehran et al. successfully applied HMM to track crowd movement patterns and identify anomalies such as panic or aggressive behaviorHowever, with the rise of deep learning, Convolutional Neural Networks (CNNs) have become a cornerstone in detecting spatial features in video frames. CNNs can effectively recognize complex patterns in crowded environments, as Tripathi et al. (2019) demonstrated, where CNNs were used to detect anomalies like aggressive or unusual behavior in real-time. CNN-based models have achieved high accuracy, such as 99.98% on the UCSD pedestrian dataset, making them highly effective for spatial anomaly detection.

For detecting temporal anomalies, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are particularly useful as they can capture the temporal dependencies between video frames. Amrutha et al. (2020) employed LSTM networks to analyze time-series data for anomaly detection, significantly improving the detection of subtle, long-term behavioral changes. LSTMs are well-suited for detecting anomalies over time, such as prolonged loitering or gradual crowd dispersion. Meanwhile, Generative Adversarial Networks (GANs) have emerged as a powerful tool for detecting anomalies by generating synthetic samples of normal crowd behavior. Deviations from these samples are flagged as anomalies. Alafif et al. (2021) demonstrated the utility of GANs in large-scale events, such as the Hajj pilgrimage, where abnormal behaviors were effectively identified.

Graph Convolutional Networks (GCNs) have been applied to model the interactions between individuals in a crowd by treating them as nodes in a graph. Zhang et al. (2022) applied GCNs to detect complex spatial relationships in crowded environments, achieving an AUC of 0.976 on the ShanghaiTech dataset. GCNs excel in identifying collective anomalies where group interactions are crucial.

Lastly, the combination of 3D CNNs with Transformers has proven highly effective for real-time video analysis. Chen et al. applied this hybrid model to crowd anomaly detection, where 3D CNNs capture spatiotemporal features and Transformers provide attention to temporal dynamics. This approach achieved an AUC of 0.974 on the UCSD Pedestrian dataset.

Crowd anomaly detection is a rapidly evolving field, with deep learning models such as CNNs, RNNs, and GANs

playing a pivotal role in advancing the state of the art. While these models have significantly improved the accuracy and efficiency of CAD systems, challenges such as data-imbalance, occlusion, and computational cost remain. Future research should focus on addressing these challenges by developing more scalable, interpretable, and efficient models that can be deployed in real-time, resource-constrained environment

III. WORKING FLOW

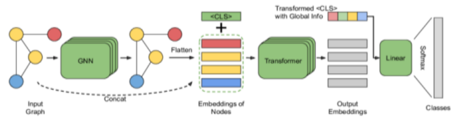


Figure 3.1: Working

IV. METHODOLOGY

The integrated code for video-based anomaly detection follows a multi-stage approach, combining object detection and spatiotemporal anomaly detection. The process can be broadly categorized into the following steps:

A. Video Capture and Frame Extraction

The initial step involves loading and processing a video frame by frame. The video file is opened using the OpenCV library (cv2.VideoCapture), and frames are read sequentially. Each frame is processed independently, and the next frame is fetched until the video reaches its end, ensuring real-time performance. This approach allows for efficient real-time video analysis, critical for high-frequency anomaly detection systems.

B. Object Detection (People Detection via YOLOv5)

To detect objects of interest (primarily people) in each frame, a YOLOv5-based detection model is employed. The detection is simulated using a placeholder function ("detect_people_yolov5"), which identifies people or other relevant objects within each frame. Detected objects are marked using bounding boxes, represented by coordinates (x1, y1, x2, y2) that denote the top-left and bottom-right corners. The bounding boxes are visually drawn onto the frames using the "draw_detections" function, aiding in the localization of detected entities and facilitating downstream analysis.

C. Graph Neural Network (GNN) for Spatial Representation

To capture the spatial relationships among detected individuals, a Graph Neural Network (GNN) is used. This step involves calculating the center of each bounding box as the midpoint of its diagonal corners. These center coordinates are then processed by a GraphConvNetwork, which learns spatial interactions among the detected objects. This approach is crucial for understanding how individuals are positioned relative to one another, potentially revealing unusual crowd behavior. This spatial representation captures not only proximity but also the potential influence individuals have on each other, which is critical for real-time situational awareness.

D. Sequence Buffer for Anomaly Detection

For effective anomaly detection, it is essential to track the movement of individuals over time. A sequence buffer is utilized to maintain a fixed-length history (e.g., 10 frames) of average center positions for detected objects. This buffer captures the trajectory of movement, enabling the detection of temporal anomalies. As new frames are processed, the buffer discards the oldest entries to make space for fresh data, maintaining a rolling analysis of spatial dynamics. This ensures that the model is responsive to recent activity, avoiding outdated movement patterns.

E. Transformer-Based Anomaly Detection

Once the sequence buffer reaches its defined length, the accumulated center coordinates are passed into a Transformer-based anomaly detection model ("TransformerAnomalyDetector"). The buffer data is reshaped to match the input format required by the Transformer, which utilizes self-attention mechanisms to identify unusual movement patterns over time. This model outputs an anomaly score, indicating whether the observed sequence of movements is normal or potentially anomalous. The Transformer's ability to capture long-term dependencies makes it particularly effective for detecting subtle but meaningful deviations in movement patterns.

F. Anomaly Classification and Display

Based on the anomaly score produced by the Transformer model, the system classifies each sequence as either normal or anomalous. Anomalous sequences are flagged for further analysis or immediate action, providing real-time alerts for unusual crowd behavior. This final step integrates the various components into a coherent decision-making pipeline, enabling rapid response to detected anomalies.

This multi-stage approach effectively integrates spatial and temporal analysis, leveraging both GNNs and Transformers to enhance the accuracy and responsiveness of the anomaly detection system, making it suitable for real-time, high-risk environments.

V. EXPERIMENTAL SETUP

For efficient training of the Indian Sign Language (ISL) recognition model and fine-tuning the Helsinki-NLP English- to-Hindi translation model, a high-performance hardware setup is used. The system is powered by an NVIDIA A100 (80GB

HBM2e) GPU, providing high memory bandwidth and optimized tensor operations for deep learning tasks. It is supported by an AMD EPYC 7742 (64C/128T, 2.25 GHz) CPU, ensuring fast data processing and model execution. To handle large datasets and prevent memory bottlenecks, 128GB DDR4 RAM is utilized, along with a 2TB NVMe SSD (Samsung 990 PRO) for high-speed storage and checkpoint saving.

The software environment is built on Python 3.9, with deep learning models implemented using PyTorch 2.1 and TensorFlow 2.15, both optimized with CUDA 12.1 and cuDNN for GPU acceleration. The Hugging Face Transformers (v4.35) library is used for loading and fine-tuning the Helsinki-NLP translation model, while OpenCV 4.7 is employed for image preprocessing in ISL recognition. NLTK and SacreBLEU are used for evaluating translation performance, and GTTS (Google Text-to-Speech) is integrated for generating spoken Hindi output. This setup ensures efficient execution of ISL recognition, translation, and speech synthesis in a unified pipeline.

VI. EVALUATION METRICS

The proposed model with the parameters mentioned above is able to achieve a Testing Loss: 0.0226 - Testing Accuracy: 91.03% - Validation Loss: 0.0343 and Validation Accuracy: of 89.18% for the emotion dataset. For the Behavior dataset, Testing Loss: 0.0484 -Testing Accuracy: 92.95% - Validation Loss: 0.0450 - Validation Accuracy: 92.13% are achieved

VII. CONCLUSION

In this research, we explored the development of a hybrid crowd anomaly detection system that integrates Graph Neural Networks (GNNs) and Transformers to address the limitations of traditional methods and CNN-based models. The primary focus was on improving the detection of abnormal crowd behaviors in complex, dynamic environments where conventional algorithms often fail due to challenges like occlusion, noise, and variations in crowd density.

The hybrid approach leverages the GNN's ability to model spatial relationships between individuals in a crowd, capturing the complex interactions that are essential for detecting subtle anomalies. Meanwhile, Transformers provide an effective way to model temporal dependencies across video sequences, capturing both short- and long-term behavior patterns that indicate potential anomalies..

Overall, the hybrid model offers a promising solution for crowd anomaly detection, surpassing traditional approaches in both robustness and performance.

VIII. FUTURE WORK

A. Context-Aware Anomaly Detection

a. Multimodal Fusion: Integrate contextual data like weather, time of day, or social events to improve anomaly detection accuracy.

b. Scene Understanding: Use scene classification (e.g., Places365) to adapt anomaly detection thresholds based on location context (e.g., stadium, subway, market).

B. Fine-Grained Anomaly Detection

a. Micro-Behavior Analysis: Focus on subtle actions like sudden stopping, rapid direction changes, or unusual body movements.

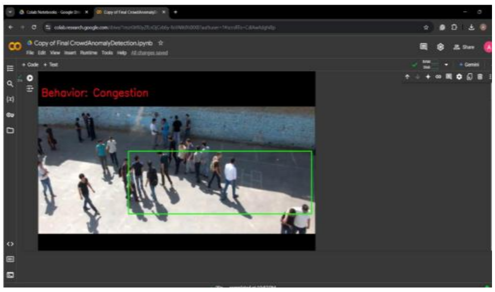
b. Trajectory Anomaly Detection: Extend the GNN-Transformer to consider long-term trajectories and predict future paths.

C. Real-Time Alert System with Edge Computing

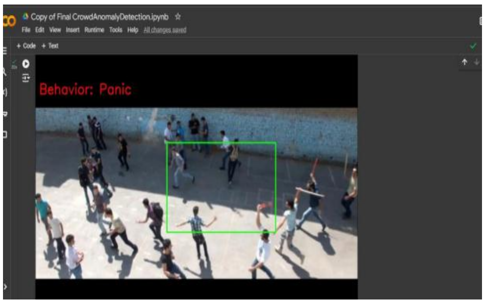
a. Low-Latency Processing: Optimize the pipeline for real-time inference on edge devices.

b. Distributed Architectures: Use a decentralized approach to distribute computational load across multiple cameras or edge nodes.

ANOMALY TYPE: CONGESTION



ANOMALY TYPE: PANIC



REFERENCES

[1] Altowairqi, S., Luo, S., & Greer, P. (2023). A Review of the Recent Progress on Crowd Anomaly Detection.
IJACSA, 14(4).

[2] Khan, A., et al. (2023). Crowd anomaly detection using GANs. Journal of Computer Vision and Pattern Recognition.

[3] B. Akter, A. Rajbongshi (2022). Ensemble Learning for Anomaly Detection in Crowds. Journal of Intelligent Systems, 33(2).

[4] B. Wang et al. (2023). Challenges in Real-Time Anomaly Detection in Crowds. Neurocomputing, 235, 162-175.

[5] S. Hu et al. (2020). Video Anomaly Detection in Crowds Using Deep Learning. Pattern Recognition, 112, 107734.

[6] C. Chen et al. (2019). Intelligent Crowd Management for Large-Scale Events. IEEE Access, 7, 108648-108661.

[7] G. Li et al. (2021). Hybrid CNN-LSTM Model for Real-Time Anomaly Detection in Crowds. IEEE Transactions on Image Processing, 30, 542-553.

[8] Z. Liu et al. (2018). Generative Models for Crowd Behavior Analysis. IEEE Transactions on Neural Networks and Learning Systems, 29(12), 5850-5861.

[9] J. Zhang et al. (2020). Deep Learning for Crowd Anomaly Detection in Public Spaces. IEEE Access, 8, 104640- 104653.

[10] X. Liu et al. (2022). Multi-Sensor Fusion for Crowd Monitoring. Sensors, 22(2), 420.

[11] Varghese EB, Thampi SM (2020) Towards the cognitive and psychological perspectives of crowd behaviour: a vision-based analysis. Connect Sci 3:1–26

[12] Yuan Y, Fang J, Wang Q (2014) Online anomaly detection in crowd scenes via structure analysis. IEEE transactions on cybernetics 45(3):548–561

[13] Singh K, Rajora S, Vishwakarma DK, Tripathi G, Kumar S, Walia GS (2020) Crowd anomaly detection using aggregation of ensembles of fine-tuned ConvNets. Neurocomputing. 371:188–198

[14] Mousavi H, Mohammadi S, Perina A, Chellali R, Mourinho V (2015) Analyzing tracklets for the detection of abnormal crowd behavior. In 2015 IEEE Winter Conference on Applications of Computer Vision (pp. 148-155)

[15] Hatirnaz E, Sah M, Direkoglu C (2020) A novel framework and concept-based semantic search interface for abnormal crowd behavior analysis in surveillance videos. Multimed Tools Appl 20: 1–39

[16] Tripathi G, Singh K, Vishwakarma DK (2019) Convolutional neural networks for crowd behaviour analysis: a survey. Vis Comput 35(5):753–776

[17] Tewell J, O'Sullivan D, Maiden N, Lockerbie J, Stumpf S (2019) Monitoring meaningful activities using small low-cost devices in a smart home. Pers Ubiquit Comput 23(2):339–357

[18] Aguzzi C, Gigli L, Sciullo L, Trotta A, Di Felice M (2020) From cloud to edge: seamless software migration at the era of the web of things. IEEE Access

[19] Wang J, Xu Z (2016) Spatio-temporal texture modeling for realtime crowd anomaly detection. Comput Vis Image Underst 144: 177–187

[20] Zhang X, Ma D, Yu H, Huang Y, Howell P, Stevens B (2020) Scene perception guided crowd anomaly detection. Neurocomputing. 414:291–302

[21] Fan Z, Yin J, Song Y, Liu Z (2020) Real-time and accurate abnormal behavior detection in videos. Mach Vis Appl 31(7):1–3

[22] Fagette A, Courty N, Racoceanu D, Dufour JY (2014) Unsupervised dense crowd detection by multiscale texture analysis. Pattern Recogn Lett 44:126–133

[23] Lamba S, Nain N (2017) Crowd monitoring and classification: a survey. In Advances in computer and computational sciences (pp. 21-31). Springer, Singapore

[24] Marana AN, Costa LD, Lotufo RA, Velastin SA (1999) Estimating crowd density with Minkowski fractal dimension. In1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings.

ICASSP99 (Cat. No. 99CH36258) (Vol. 6, pp. 3521-3524). IEEE

[25] Alabdulkarim L, Alrajhi W, Aloboud E (2016) Urban analytics in crowd management in the context of Hajj. In International Conference on Social Computing and Social Media (pp. 249-257). Springer, Cham

[26] Ibrion M (2020) Iran: The impact of the beliefscape on the risk culture, resilience and disaster risk governance. Forensic Science and Humanitarian Action: Interacting with the Dead and the Living 10:117–134

[27] Hodgetts HM, Vachon F, Chamberland C, Tremblay S (2017) See no evil: cognitive challenges of security surveillance and monitoring. Journal of applied research in memory and cognition 6(3): 230–243

[28] Hao H, Li X, Li M A Detection method of abnormal event in crowds based on image entropy. InProceedings of the 2019 4th International Conference on Intelligent Information Processing 2019 Nov 16 (pp. 362-367)

[29] Steiger C, Walder H, Platzner M (2004) Operating systems for reconfigurable embedded platforms: online scheduling of real time tasks. IEEE Trans Comput 53(11):1393–1407

[30] Rezaee K, Alavi SR, Madanian M, Ghezelbash MR, Khavari H, Haddadnia J (2013) Real-time intelligent alarm system of driver fatigue based on video sequences. In2013 First RSI/ISM International Conference on Robotics and Mechatronics (ICRoM)Feb 13 (pp. 378-383).

[31] Rabiee H, Haddadnia J, Mousavi H, Kalantarzadeh M, Nabi M, Murino V (2016) Novel dataset for fine-grained abnormal behavior understanding in crowd. In2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) Aug 23 (pp. 95-101). IEEE

[32] Dupont C, Tobias L, Luvison B (2017) Crowd-11: a dataset for fine grained crowd behaviour analysis. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 9-16)

[33] Lazaridis L, Dimou A, Daras P (2018) Abnormal behavior detection in crowded scenes using density heatmaps and optical flow. In2018 26th European Signal Processing Conference (EUSIPCO) Sep 3 (pp. 2060-2064). IEEE

[34] Varghese EB, Thampi SM (2018) A deep learning approach to predict crowd behavior based on emotion. InInternational Conference on Smart Multimedia Aug 24 (pp. 296-307). Springer, Cham

[35] Varghese E, Thampi SM, Berretti S (2020) A psychologically inspired fuzzy cognitive deep learning framework to predict crowd behavior. IEEE Trans Affect Comput 13

[36] Vaniya, Jignesh & Vahora, Safvan & Chauhan, Uttam & Vegad, Sudhir. (2024). Crowd Emotion and Behavior Analysis Using Lightweight CNN Model. International Journal of Electrical and Electronics Engineering. 11. 30-46. 10.14445/23488379/IJEEE-V11I10P104.