Real-Time Crowd Anomaly Detection with Predictive Alerts.

Dhanush TC[1], Nandish Sajjan[2], Nikhil Gowda CR[3], Thilak Gowda YS[4], Dr. Shankara Gowda SR[5],R Sivakumar[6]
[1,2,3,4]Undergraduate Student, Department of Information Science and Engineering
[5 ,6]Associate Professor, Department of Information Science and Engineering

ABSTRACT Monitoring crowded places like railway stations, airports, shopping malls, stadiums, places of worship, public transport, and large gatherings can be a difficult task. Dangerous situations may evolve quickly, such as crowding, stampedes, fights, unusual running or walking, overcrowding, people walking or running in opposite directions, and unusual noises. It may be hard to detect these situations with only traditional CCTV surveillance since humans can become tired or take a long time to react properly, potentially leading to dangerous situations.. This paper proposes real-time crowd anomaly detection with predictive alerts. This paper presents a technique for detecting abnormal crowd behavior using computer vision and deep learning models, audio processing, modal fusion, and an edge server for on-device processing. This approach monitors live CCTV videos and microphone audio feeds over a local area network. The proposal detects objects in the video using YOLO and estimates crowd density, then extracts features from the video using Vision Transformer or Swin Transformer. The audio feed is converted to a spectrogram and processed using an appropriate audio classification model for detecting irregular activities. Both processed feeds are fused in a multimodal fusion module to make decisions on the status of the crowd.

KEYWORDS Crowd Behavior Analysis, Anomaly Detection, Computer Vision, Deep Learning, Real-Time Surveillance, Crowd Density, Alert Sound System, Public Safety.

I. INTRODUCTION

Crowd management and security have become an important issue in a variety of modern public places involving a huge number of people. Such places as railway stations, airports, stadiums, shopping malls, religious places, public functions, and large-scale events can experience a rapid convergence of people. Herein, an unforeseen problem like stampede, panic, violence, fall, surging, counterflow, or other anomalous events can take place. Such situations require immediate attention to avoid hazardous outcomes; thus, it is necessary to continuously monitor such public places and recognize such anomalous events at an early stage. Most of the current surveillance systems utilize Closed-Circuit Television (CCTV) cameras for monitoring such places in order to detect any irregularity. However, although continuous surveillance is possible using CCTV cameras, the analysis of recorded footage is usually operator-dependent. Moreover, when multiple CCTV cameras are used, it can be impossible for a human operator to continuously watch over multiple screens. As a result, important footage can be ignored. Furthermore, anomaly recognition from the recorded video data needs a lot of time and effort. This limitation reduces the effectiveness of conventional surveillance systems, especially in crowded environments where events can occur within a few seconds. These challenges motivate researchers to design Artificial Intelligence (AI)-based surveillance systems for automated analysis of a large amount of visual information. Such systems allow automated anomaly recognition and can continuously monitor surveillance footage without depending completely on human observation. In particular, recent advances in deep learning have offered several approaches for understanding complex crowd behavior. Convolutional Neural Networks (CNNs) can be employed for feature extraction from a given video frame, while Long Short-Term Memory (LSTM) networks are suitable for learning temporal patterns from a set of video frames. You Only Look Once (YOLO) object detection algorithms can be used for detecting people and other objects in a given video frame, while optical flow can offer useful information about the motion and direction of people.

Modern approaches also utilize Transformer-based encoder-decoder networks for learning relationships between video frames. Such networks are appropriate for learning long-range dependencies between objects and activities occurring at different locations and time-intervals. One of the important factors to be considered in crowd monitoring is the density of a given crowd, which can offer useful information for predicting the likelihood of an anomalous event. Under normal circumstances, most people present in a given crowd generally move in a similar direction and maintain relatively comparable movement patterns. However, when an anomalous event takes place, these patterns can be substantially changed. By analyzing the apparent motion of objects between consecutive video-frames, optical-flow methods can estimate the direction and velocity of movement within a crowd. There are also several other challenges associated with building an efficient surveillance system based on computer vision. First, such a system should be able to work in different environments. Due to variations in background, camera angles, illumination, weather conditions, image quality, and perspectives, AI-based models can produce different results when processing different videos. A model trained on one

environment may not always offer the same level of performance when processing videos from another environment. Thus, the proposed system needs to consider environmental variations and should be designed in a way to offer reliable results under changing surveillance conditions. Another important limitation of conventional video-only surveillance is that visual information may not always be sufficient to understand the complete situation. Certain events may produce strong audio-signals even when the corresponding visual information is unclear. For example, screams, alarms, unusual shouting, breaking sounds, emergency announcements, or sudden loud noises may offer additional evidence of a dangerous situation. Thus, integrating audio information with video analysis can improve the capability of the system to recognize abnormal events. Audio analysis can be conducted by converting sound signals into suitable representations and applying machine-learning or deep-learning models to classify abnormal acoustic patterns. The integration of multiple data-sources is commonly referred to as multimodal analysis. In the proposed system, video and audio information are analyzed separately and then are combined using a multimodal fusion mechanism. The video branch can offer information related to crowd density, people detection, movement direction, speed, and abnormal behavior, whereas the audio branch can offer information about unusual environmental sounds. A crowd anomaly detection system is useful only when it can process incoming information quickly enough to support timely decision-making.

Sending every video-frame and audio-signal to a centralized cloud server can introduce communication delays and increase network requirements. To overcome this problem, Edge AI can be incorporated into the proposed architecture. Edge computing allows part of the AI processing to be performed close to the location where the data is generated. Thus, this can reduce latency, decrease network traffic, and improve the response-time of the surveillance system. The proposed system further considers the importance of explainability in AI-based security applications. When an automated system generates a high-risk alert, security personnel need to know the reason behind that decision. Thus, an explainability module can identify the important visual or audio factors that led to an alert. For instance, the system may indicate that an alert was generated because of increasing crowd density, abnormal movement direction, high motion variation, or detection of an unusual sound. Such explanations can assist operators in making faster and more informed decisions.

Another important component is automated alert generation and event logging. When the calculated risk-exceeds a predefined threshold, the decision engine can classify the situation into different risk-levels, such as normal, low risk, medium risk, high risk, or critical. Based on the risk level, the system can generate appropriate notifications through a monitoring dashboard, mobile-application, SMS, alarm, or other communication mechanisms. At the same time, important information such as the detected event, timestamp, camera location, risk-score, and relevant evidence can be stored in an event log. The use of AI in crowd surveillance can reduce the dependence on continuous manual observation. Thus, this combination of automated analysis and human decision-making can improve the overall efficiency of security operations.

Despite the advantages of AI-based crowd monitoring, the development of a reliable system is challenging. A surveillance model must deal with large variations in crowd size, camera perspectives, illumination, occlusion, background conditions, movement patterns, and environmental noise. Thus, accuracy, speed, robustness, scalability, and explainability must be considered together when designing the system. The proposed work is focused on addressing these challenges through a Real-Time Multimodal Crowd Anomaly Detection and Predictive Alert System. The system combines CCTV-based video analysis with audio analysis to obtain complementary information about crowd situations. The video-processing component utilizes person detection and crowd-density estimation to understand the number of people and their spatial distribution. Behavior analysis techniques, including Transformer-based approaches and optical-flow information, are used to identify irregular movement patterns. Meanwhile, the audio-processing component analyzes environmental sounds to identify unusual acoustic events.

The outputs of the video and audio analysis modules are combined through multimodal fusion to obtain a more reliable representation of the current situation. The fused information is then passed to a predictive risk-engine that analyzes both current and changing situations. Based on the estimated risk, a decision-engine determines whether an alert should be generated. When a potentially dangerous condition is recognized, the system can provide an automated alert along with relevant information and supporting evidence. Thus, the proposed work is focused on addressing several important aspects of intelligent crowd surveillance, including crowd monitoring using CCTV cameras, utilizing audio data for anomaly detection, estimating crowd density, detecting anomalous crowd activities, multimodal fusion for obtaining more complete situational information, predictive risk assessment, Edge AI for efficient real-time processing, explainable AI for understanding system decisions, and automated alerts and event logging.

Overall, the proposed system is aimed at providing a proactive and intelligent approach to crowd safety. Rather than relying completely on human operators or detecting an incident only after it has occurred, the system attempts to continuously understand crowd conditions, identify deviations from normal behavior, estimate the level of risk, and provide early warnings. Such a system can be particularly useful in railway stations, airports, stadiums, shopping malls, religious gatherings, public events, and smart-city surveillance environments. By combining computer vision, audio analysis, deep learning, multimodal fusion, predictive analytics, edge computing, and explainable AI, the proposed approach has the potential to improve the speed, reliability, and effectiveness of real-time crowd safety and security management

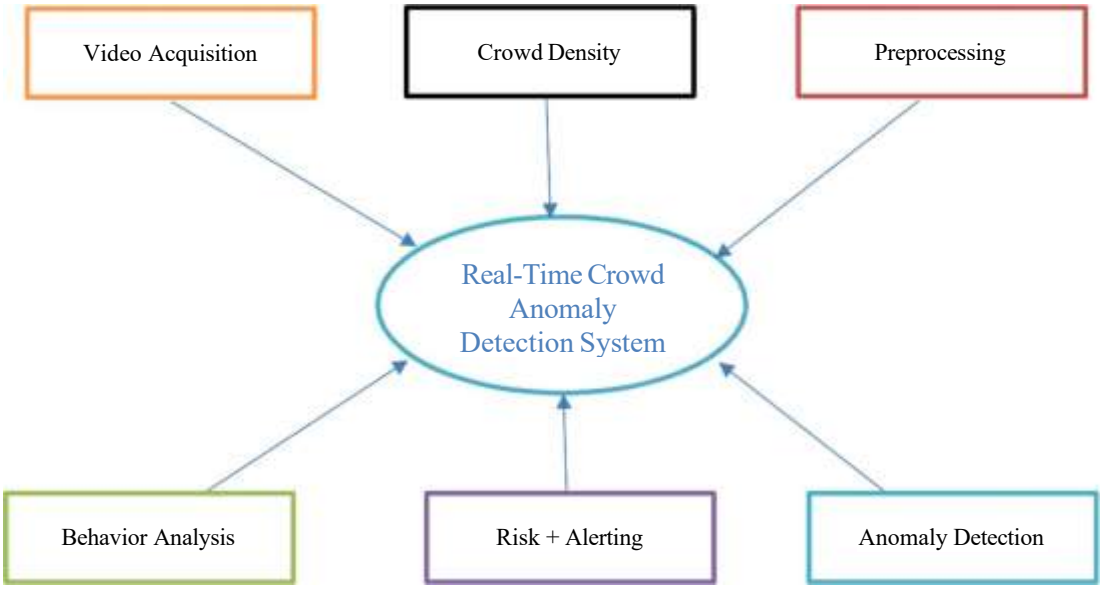


FIGURE 1. Taxonomy with the main themes and sub-themes.

II. LITERATURE REVIEW METHODOLOGY

A. *LITERATURE SEARCH AND SELECTION*

The articles were collected mainly by exploring research works related to crowd anomaly detection and intelligent surveillance. The research papers were reviewed to retrieve information on the existing techniques for crowd abnormality detection, crowd density estimation, violence detection, people's flow detection or monitoring with safety alerts. Of primary importance were the recent research works involving AI and Deep Learning methods since they are used for automatic real-time surveillance.

The collected literature was reviewed, and those research papers were included in the final list that addressed various issues related to intelligent video surveillance of crowds. All the included works either addressed the problems of person detection and counting using YOLO-based approaches or applied optical-flow, 3D convolution, temporal, or transformer-based networks for analyzing abnormal behaviors in crowds. Crowd-density estimation was also of particular interest due to the hazards of overcrowding and the importance of the density as an indicator of potential crises. After removing the articles that did not address the issues related to the goal of the research, all the remaining works were analyzed in-depth following the below-described structure. Every paper was evaluated according to the objectives, methodology, dataset, algorithms, real-time processing ability, and results with conclusions. Finally, each article was grouped based on the main issues addressed and the used methodologies.

B. *INCLUSION AND EXCLUSION CRITERIA*

The literature was selected according to its relevance to one or more of the major research areas considered in this study. Papers were included when they contributed substantially to at least one of the following:

- Research papers related to crowd anomaly detection, crowd monitoring, abnormal behavior detection or similar fields.
- Studies that make use of Artificial Intelligence (AI), Machine Learning (ML), Deep Learning, or Computer Vision techniques.
- Research involving crowd counting, person detection, density estimation, or movement analysis.
- Studies that aim at detecting abnormal activities like violence, crowd turbulence, counter-flow, sudden movements, or overcrowding.
- That involve real-time or near-real time surveillance and public safety applications.

The Research papers which did not contribute to any module of this research themes were removed and are not included in this final analysis of our literature survey paper.

C. *THEMATIC CLASSIFICATION OF THE INCLUDED RESEARCH PAPEPS*

To make comparative analysis, the selected studies were classified according to their primary technological contribution. The literature broadly falls into six interconnected themes:

- crowd anomaly and abnormal behavior detection
- deep learning-based video analysis.

- YOLO-based person detection.
- optical flow and motion analysis.
- transformer-based behavior recognition.
- multimodal and predictive safety systems.

This classification enables the review to move beyond individual paper summaries and examine how existing approaches address different stages of the proposed forest-monitoring and traceability framework. The reviewed studies are subsequently compared in terms of their data sources, methodologies, detection capabilities, limitations, and research gaps.

TABLE 1. A summary of the identified studies

OBJECTIVES	STUDIES
Develop real-time systems for detecting abnormal crowd behavior from live surveillance video.	[1], [7], [8], [9], [10]
Detect and analyze unusual crowd movements such as panic, turbulence, counter-flow, and rapid gathering.	[1], [5], [8], [9]
Improve crowd detection and density estimation for enhanced public safety.	[3], [4], [6], [12]
Use deep-learning models to capture spatial and temporal features for accurate anomaly detection.	[5], [7], [8], [11]
Develop efficient person detection and tracking methods for crowded environments.	[3], [4], [6], [13]
Reduce surveillance response time through automated real-time alert generation.	[1], [4], [9], [10]
Improve anomaly detection under challenging conditions such as occlusion, frame discontinuity, and dense crowds.	[6], [7], [11], [12]
Integrate multiple surveillance sources and contextual information for improved threat assessment.	[2], [10]
Enhance the reliability and interpretability of AI-based crowd anomaly detection systems.	[6], [10], [13]
Support proactive public-safety monitoring by identifying potential threats before they develop into serious incidents.	[1], [2], [4], [5], [9]

The above table summarizes the objectives that are gathered from this study and they are mentioned with the related paper for that particular objective or focus of that paper.

TABLE 2. Themes and their relation to the relevant papers.

THEME	STUDIES
Real-Time Crowd Monitoring and Anomaly Detection	[1], [5], [7], [8], [9], [10], [11]
Crowd Density Estimation and Person Detection	[3], [4], [6], [12]
Crowd Movement and Abnormal Behavior Analysis	[1], [5], [8], [9], [11]
Deep Learning and Intelligent Surveillance	[5], [7], [8], [11], [13]
Predictive Alerts and Public Safety	[1], [2], [4], [9], [10]

III. CRITICAL ANALYSIS OF THE PAPERS

THEME I – REAL TIME CROWD MONITORING AND ANOMALY DETECTION

V. Devi et al. [1], Linbo Luo et al. [5], Maira Khalid et al. [7], AngelJulie J. et al. [8], Manu Y. M. et al. [10]. The first theme deals with developing surveillance systems for real-time crowd monitoring and building automatic anomaly detection systems. Typically, CCTV surveillance involves security staff constantly watching over multiple camera feeds, which can be challenging, especially with a high number of concurrent cameras, and some abnormal activity may not be noticed in time. The papers reviewed under this theme attempt to tackle automatic anomaly detection using Artificial Intelligence and Deep Learning algorithms. V. Devi et al. [1] proposed an AI-based system for real-time crowd surveillance, capable of identifying abnormal crowd activities such as running, panic movement, abnormal gathering, crowd density, and aggressive behavior. Moreover, the system allows for incorporating automated audio announcements to notify security staff of abnormal activities. Luo et al. [5] focused on crowd-level abnormality detecting, namely, counter-flow and crowd turbulence. The paper describes an approach that takes into account motion consistency at different scales and tries to estimate the degree of abnormality [5]. Crowd-level abnormality is relevant when motions of several people deviate from the expected motion pattern. Khalid et al. [7] applied 3D ResNet with optical flow to encode both spatial and time-wise information of the video. AngelJulie J. et al. [8] utilized a Convolutional Neural Network with Long Short-Term Memory (CNN-LSTM) architecture to recognize abnormal activities by extracting features from multiple video frames, highlighting the importance of considering time-wise information in video content analysis. Manu Y.
M. et al. [10] also worked on real-time crowd and video surveillance with automatic anomaly detection for improved safety. The paper confirms that video content analysis is valuable for automating safety inspection in public places.

Research Gap as mentioned above, the authors of the reviewed papers primarily focus on developing automatic anomaly detection systems for crowd surveillance. In contrast, less attention is paid to predicting imminent dangerous situations such as stampede. The project can address this gap by introducing a risk prediction module that analyzes deviations in crowd behavior over time.

THEME II – CROWD DENSITY ESTIMATION AND PERSON DETECTION

Dr. G. Adiline Macriga et al. [3], Aswathy Leelamony et al. [4], Deevesh Chaudhary et al. [6], Md Roman Bhuiyan et al. [12]. The second theme deals with person detection, crowd counting, crowd density estimation, and crowd dynamics. Estimating the number of people in a specific area is critical for the purpose of safety control since a high concentration of people brings the risks of congestion, impediment, and crowd disaster. Dr. G. Adiline Macriga et al. [3] explored the application of artificial intelligence and machine learning to crowd management and public safety. The article discusses the potential of automatic monitoring and predictive analysis in improving the control of large crowds. Aswathy Leelamony et al. [4] proposed Deep Track Net, an approach that leverages on Deep SORT and YOLO-based object detection for crowd counting and crowd dynamic analysis. According to the author, detecting and tracking persons' movements enable a better understanding of crowd behavior compared to just counting them. Deevesh Chaudhary et al.
[6] addressed the concept of intelligent anomaly detection with crowd-density estimation. The article demonstrated the possibility of partitioning crowd density into categories as a critical parameter for detecting anomalous events. Md Roman Bhuiyan et al. [12] presented a vision-based database for crowd monitoring during the hajj. The authors argued that specific databases are essential for analyzing crowd behavior, as their characteristics may differ from one region to another. This idea is critical in addressing the theme since it acknowledges the variations in crowd behavior and the need for differentiating them. Critical Analysis person detection is a key element in the proposed project because the count can be used to estimate the crowd density. In this regard, using an algorithm such as YOLO is appropriate because it allows for fast detection and can be used in real-time surveillance systems. However, the detection process faces several challenges, including occlusion, low-resolution videos, low-quality cameras, poor lighting, and crowd size. These factors make it hard to detect persons' movements and behaviors accurately. While density estimation is a critical aspect of the project, it cannot be used as the sole parameter for determining the level of crowd danger. For instance, a large number of persons in a football stadium may not be dangerous if they are seated far from each other. On the other hand, a moderate number of people in a small area can lead to a disaster if they start walking unsystematically.

THEME III – CROWD MOVEMENT AND ABNORMAL BEHAVIOR ANALYSIS

V. Devi et al. [1], Linbo Luo et al. [5], AngelJulie J. et al. [8], Hina Hashmi et al. [9], and Karthicka K. et al. [11]. The third theme focuses on analyzing crowd motion patterns and abnormal behavior. Crowd emergencies are usually related to crowd motion pattern changes. Panic running, counter-flow, sudden gathering, crowd dispersion, abnormal movement direction, and violent activity are examples of such changes. V. Devi et al. [1] considered panic movement pattern, sudden running, aggressive gestures, and unexpected gathering as potential candidates for abnormal behavior. On the other hand, counter-flow and crowd turbulence were considered abnormal collective behavior by Luo et al. [5]. AngelJulie J. et al. [8] used convolutional neural networks with long-short term memory (CNN-LSTM) to analyze spatial and temporal patterns. It should be noted that crowd motion pattern analysis was not sufficient when single-frame processing was used. Multiple consecutive frame observations were necessary to identify whether the crowd was behaving abnormally or its motion pattern changed. Hina Hashmi et al. [9] focused on abnormal movement patterns, sudden gathering, and intrusion detection using light-weight artificial intelligence suitable for secure regions. Karthicka K. et al. [11] solved the video frame disconnection problem and developed a temporal-aware approach for violence recognition. Critical Analysis one of the key techniques for analyzing crowd motion patterns is optical flow, which allows obtaining information about the direction and magnitude of motion. The combination of optical flow and deep-learning algorithms can effectively detect sudden changes in crowd motion pattern. It should be noted that temporal analysis plays a special role in predictive systems, as answering the question "is the current frame abnormal?" is not enough. Instead, the following questions should be asked. The proposed project can address this gap by tracking changes in movement patterns over time. These changes, along with density and audio information, can be used to calculate the predictive risk score.

THEME IV –DEEP LEARNING, EXPLAINIABLE AND ROBUST SURVEILLANCE

Linbo Luo et al. [5], Maira Khalid et al. [7], AngelJulie J. et al. [8], Karthicka K. et al. [11], and Melisa Gozet et al. [13]. The fourth theme deals with the use of deep learning and intelligent computer-vision methods for crowd anomaly detection. Traditional image-processing techniques often require manual design of features, while deep-learning models can learn important features directly from training data. Luo et al. [5] have applied multi-scale motion information and attention-based feature fusion for crowd-level abnormal behavior detection. This is useful since anomalies can appear at varying spatial scales and involve either small subsets or the entire crowd. Khalid et al. [7] have used 3D ResNet to learn spatial and temporal characteristics of a video. Compared to conventional image-based architectures, 3D convolution can process information at different frames, suitable for activities involving motion. AngelJulie J. et al. [8] have applied CNN-LSTM, where CNN extracts visual features and LSTM processes temporal information. This is useful for understanding the time evolution of crowd behavior. Karthicka
K. et al. [11] have introduced a temporal-aware framework

considering video-frame discontinuities. This is relevant to real-world surveillance since network failures, camera problems, occlusion, and corrupted frames can interrupt video streams. Melisa Gozet et al. [13] have focused on explainable anomaly detection using YOLOv11 and attention-based visualization techniques. Explainability is valuable since security personnel needs to understand why the system has produced an alarm. Critical Analysis deep-learning models can offer high detection performance, but they bring in several practical challenges. Large models demand powerful GPUs and can increase system latency. For a real-time surveillance system, the model needs to process incoming videos fast enough to provide timely alerts. Another concern is dataset dependency since anomalies can appear differently under varying camera resolutions, weather, crowd behaviors, lighting, and background. Explainability is another concern. A security operator would be hesitant to act on an anomaly if the system simply states "Anomaly Detected." Perhaps an improved surveillance system can state "Increased Density + Unusual Movement + Unusual Audio." Research Gap there is a need for a real-time, computationally efficient, explainable, and robust system that can operate under various surveillance conditions. The proposed project can address this by using multiple detection modules instead of a single deep-learning model.

THEME V– RPREDICTIVE ALERTS AND PUBLIC SAFETY

V. Devi et al. [1], Rohith Amalan J. et al. [2], Aswathy Leelamony et al. [4], Hina Hashmi et al. [9], Manu Y. M. et al. [10]. The fifth theme is related to the main objective of the proposed research work, which is to provide automated alerts for increased public safety. Providing automated alerts is only one step of an intelligent video surveillance system. The information about the detected anomalous events needs to be translated into an appropriate alert to inform the security personnel. V. Devi et al. [1] added automated sound alert for detecting abnormal crowd behavior. Rohith Amalan J. et al. [2] utilized the information from multiple cameras, presence of weapons, size of the crowd, and prioritization to develop meaningful security alerts. Aswathy Leelamony et al. [4] focused on crowd alert and analysis whereas Hina Hashmi et al. [9] proposed a concept for lightweight AI monitoring for secured zones. Manu Y. M. et al. [10] explored the concept of real-time anomaly detection for improved public safety. Critical Analysis. The main benefit of such automated alert systems is that they can reduce the time to reaction, which means the time between the event identification and human initiation of an appropriate response. Depending on the concept, the alert can be visualized in a monitoring dashboard, sent as push notification to a mobile application, as an SMS, or by triggering a siren or other visual and audio alarms.

IV. CRITICAL ANALYSIS OF RESEARCH PAPERS

A. *COMPARATIVE ANALYSIS*

The reviewed research papers present the development of AI and deep-learning algorithms for intelligent crowd monitoring and abnormality recognition. The papers [1-3] have explored various aspects of crowd-level abnormal activity recognition, such as counter-flow and crowd turbulence detection [1], crowd management, density estimation, and public-safety-related tasks [2], and abnormal event detection by applying 3D ResNet along with the information on optical flow [3]. The papers [4-7] have proposed different solutions, including using YOLO and Deep-SORT for crowd-dynamics estimation [4], introduction of the crowd density estimation framework applicable for very large crowds, such as at the Hajj [5], simultaneous consideration of crowd-density estimation and abnormality detection and location [6], and developing lightweight algorithms for smart-city crowd surveillance [7].
The papers [8-11] have presented different architectures and approaches, including an explainable anomaly detection framework based on YOLOv11 and attention mechanism [8], application of CNN-LSTM for simultaneous consideration of spatial and temporal data [9], and motion-based local and global abnormal crowd event detection [10]. The paper [11] suggested the real-time deep-learning approach to anomaly detection, while the computational complexity of the chosen model might become an issue for some applications. The papers [12-14] have introduced various crowd abnormality detection solutions. In particular, the authors [12] suggested analyzing unusual crowd behavior using crowd-count-based information and image occupancy without focusing on individual persons' recognition. The authors [13] explored motion-based approaches for local and global crowd abnormality detection, while the authors [14] have emphasized the importance of the explainable AI for crowded scene analysis. The papers [15-20] have greatly contributed to the development of video processing and crowd analysis algorithms as the basis for the creation of an intelligent surveillance system. In particular, Ito et al. [15], Ye et al. [16], and Chrisandy et al. [17] investigated various video-based approaches for crowd surveillance, while Alsabei et al. [18] and Chaudhary and Kumar [19] emphasized the importance of applying AI-based solutions. Finally, the paper by Hu et al.
[20] introduced visual analysis algorithms for intelligent video surveillance. All the authors mentioned above have made significant contributions to the development of automated video processing systems.Overall Comparison based on the review, it is possible to conclude that most authors tend to focus on either crowd-density estimation, person detection and tracking, motion analysis, abnormality classification, temporal analysis, or applying explainable artificial intelligence or approaches applicable for real-time processing. Applying YOLO-based architectures has been proved to be efficient for person detection. Using CNN and LSTM-based approaches allows for efficient consideration of spatial- and temporal-related information contributing to crowd abnormality detection. The application of motion-based algorithms, including optical flow estimation, helps recognize unusual patterns of movement. Crowd-density estimation approaches help determine the density distribution and better understand how crowds behave. Moreover, the approaches based on explainable AI are efficient to make sure that the crowd surveillance system works correctly. Algorithms based on a lightweight architecture allow for efficient processing of information at the edge level. However, most authors address only one problem or explore a small part of the issue. Most solutions based on video information only react to an already occurred event. Several issues, such as occlusion, crowd density, a lack of appropriate annotated data,computational complexity of the architectures, false alarms, changes in

illumination or weather conditions, the lack of generalizability between different data sets, and a lack of consideration of audio information, are still critical points that need to be addressed. As a result, it is important to make sure that an intelligent system is capable of efficiently processing different types of information, such as video and audio, to ensure the detection of various types of abnormal activity and predict potential adverse outcomes. Thus, the thesis statement of this paper considers the necessity of creating a unified system that takes advantage of the latest developments in the field of automated video processing and applies them for real-time abnormality detection and risk prediction. In contrast to solutions that only address already occurred events, the proposed solution would be able to learn the current state of the crowd and predict potential risks.

B. IDENTIFIED RESEARCH GAPS

However, several limitations are still present. Most existing systems concentrate on detecting abnormal activities from video, while other useful information such as sound, crowd density, and changes in movement is not always considered together. The following research gaps were identified:

1. **Limited Use of Multiple Data Sources:**
Most existing methods mainly depend on surveillance video. Sounds such as screams, alarms, or other unusual noises can provide important clues about dangerous situations, but they are not commonly combined with video analysis.

2. **Lack of Early Prediction:**
Many systems detect an abnormal event only after it has already started. There is a need for systems that can study the gradual changes in crowd density and movement and predict a possible threat before it becomes serious.

3. **Lack of an Integrated System:**
Existing studies often focus on a particular task, such as person detection, crowd counting, movement analysis, or anomaly classification. A complete system that combines all these functions in one framework is still limited.

4. **Difficulty in Handling Highly Crowded Areas:**
In dense crowds, people frequently overlap or become partially hidden from the camera. This makes person detection and tracking difficult and can reduce the accuracy of anomaly detection.

5. **Limited Generalization Across Environments:** Many studies are tested using specific datasets or particular locations. A model that performs well in one environment may not provide the same performance in railway stations, stadiums, airports, shopping malls, or religious gatherings.

6. **Computational Challenges:**
Advanced deep-learning models can provide good detection performance but may require high processing power. This can make real-time implementation difficult, especially on edge devices with limited computational resources.

7. **Possibility of False Alerts:**
Normal crowd activities can sometimes look similar to abnormal behavior. For example, people running during an ordinary event may be incorrectly identified as a panic situation. Therefore, more reliable risk assessment and adaptive thresholds are required.

8. **Limited Explanation of Alerts:**
Some systems can identify an anomaly but do not clearly explain why the alert was generated. Providing information about the location, type, severity, and possible cause of an event would help security personnel respond more effectively.

9. **Limited Real-World Testing:**
A large number of existing approaches are evaluated using benchmark datasets rather than continuous real-world surveillance. More practical testing is required under different lighting conditions, camera angles, crowd sizes, and environmental situations.

10. **Need for Proactive Safety Monitoring:**
The major gap is the lack of a unified system that can **detect, understand, and predict crowd anomalies in real time**. Integrating video analysis, audio detection, crowd-density estimation, behavior analysis, multimodal fusion, and predictive risk assessment can help provide early and more reliable safety alerts.

Overall Research Gap

In summary, the existing studies provide strong foundations for crowd anomaly detection, but most of them address individual aspects of the problem. There is still a need for a **real-time, multimodal, predictive, and practical system** that can combine visual and audio information to understand the current crowd situation and estimate the possibility of an upcoming threat.

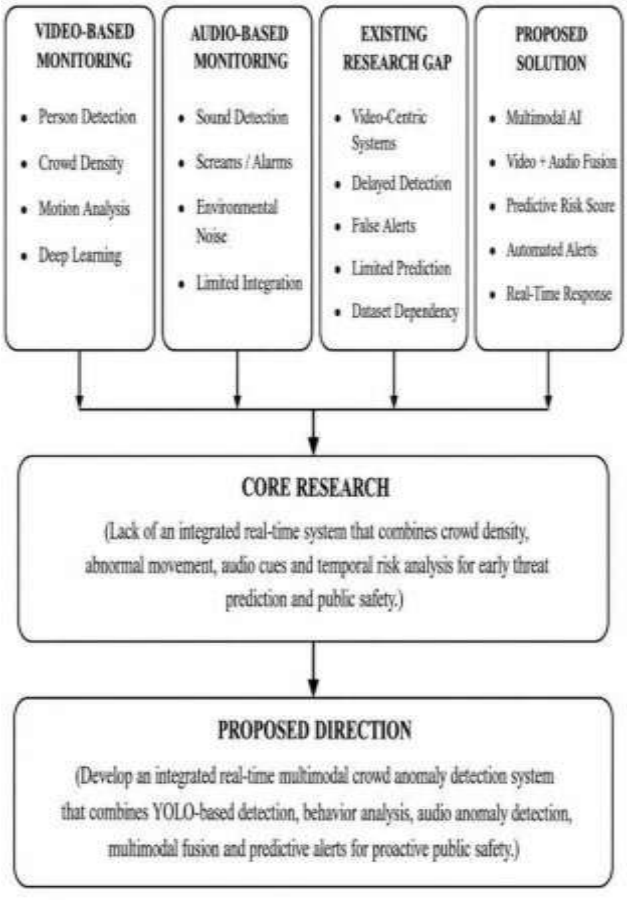


FIGURE 2. Research gap framework

V. FUTURE RESEARCH AND ADVANCEMENT

The suggested Real-Time Crowd Anomaly Detection With Predictive Alerts system could be improved upon by implementing advanced deep-learning and transformer-based models for the detection of crowds and their behaviors. Future research could focus on incorporating video and audio data to identify panic, irregular sounds, alarms, and other anomalies. The predictive risk engine could be improved to monitor the changes in density, movement, and anomaly score to issue alerts before the occurrence of a dangerous crowd event. The system could also utilize multi-camera fusion and edge computing to increase the speed and accuracy of the detection while minimizing latency. Explainable AI could be implemented to make detection and issuing of alerts interpretable to guarantee trustworthiness. These ideas could improve the system's accuracy, reliability, speed, and applicability in the real world.

VI .CONCLUSION

The Real-Time Crowd Anomaly Detection With Predictive Alerts project demonstrates an AI-based concept for improving safety and security in crowds in a public environment. The designed system constantly analyzes surveillance videos and evaluates the crowd's status, detecting abnormal activities, such as sudden movements, overcrowding, counterflow, panic, and other irregular behaviors. The YOLO-based person detector and crowd-density estimator provide the algorithm with insights on the number of people and their distribution, while analyzing movements and behaviors detect deviations from the norm. The proposed approach utilizes audio anomaly detection, allowing the system to receive additional information from unusual sounds, such as screams or alarms. By processing both video and audio, the crowd-monitoring system, using multimodal fusion, improves reliability and decreases fallibility in favor of a system that utilizes visual data only. Lastly, through a predictive risk engine, the algorithm assesses the rate of change of density, movement, and anomaly score, predicting the likelihood of an incident. The system is capable of classifying the context of crowds and issuing early warning alerts via a security dashboard, mobile application, SMS, alarm, or other available means. This reduces the response time of security personnel and eliminates the need for constant monitoring of CCTV feeds. Overall, this project aims for the paradigm shift of traditional surveillance systems from reactive anomaly detection to proactive crowd-safety management. By providing real-time insights, utilizing deep learning and multimodal processing, implementing an automated risk assessment engine, and issuing predictive alerts, the proposed system will offer a more reliable and faster alternative to the current safety solutions. The concept can be expanded upon with the addition of multiple camera feeds, edge computing, explainable AI, extensive data sets, and advanced predictive algorithms. The application of the designed system will be useful in railway stations, airports, stadiums, shopping malls, places of worship, smart cities, and other large public gatherings.

REFERENCES

[1] L. Luo, S. Xie, H. Yin, C. Peng, and Y.-S. Ong, "Detecting and Quantifying Crowd-Level Abnormal Behaviors in Crowd Events," IEEE Transactions on Information Forensics and Security , vol. 19, pp. 6810-6823, 2024.

[2] G. A. Macriga, S. Bavyatha, S. H. Abhinidhi, and N. Jagadeesh, "Crowd Management Using AI & ML," in 2024 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)2024,pp.1-5.

[3] M. Khalid, A. R. Mohsin, J. Chandroth, J. Ali, and B.-H. Roh, "Real-Time Anomaly Detection in Crowd Surveillance Using 3D ResNet and Optical Flow," in 2025 International Conference on Information Networking (ICOIN) , 2025, pp. 513-51.

[4] A. Leelamony, A. Y. Ju Yang, S. Fallon, E. Fallon, R. Young, and M. E. Curran, "DeepTrackNet: Advanced Crowd Dynamics Estimation with YOLO and Deep Sort Integration," in 2025 International Conference on Computer Science and Software Engineering (CSASE) , 2025, pp. 80-86.

[5] M. R. Bhuiyan, J. Abdullah, N. Hashim, F. Badie, F. A. Farid, B. Duraisamy, and J. Uddin, "Towards intelligent crowd monitoring during Hajj: a novel dataset for density estimation and anomaly detection in the Tawaf area (2015-2019)," in 2025 Multimedia University Engineering Conference (MECON) , 2025.

[6] D. Chaudhary, S. Kumar, and V. S. Dhaka, "Intelligent Anomaly Detection and Localization With Crowd Density Estimation for Improved Public Safety," IEEE Access , vol. 13, pp. 169912-169928, 2025.

[7] H. Hashmi, A. Kumar, and P. Singh, "Real-Time Crowd Monitoring Using Lightweight AI for Secure Zones," 2025.

[8] M. Gozet, M. Karakose, and A. E. Yilmaz, "YOLOv11-Based Explainable Framework For Anomaly Detection In Crowded Scenes Using Attention Fusion," in 2025 15th International Conference on Advanced Computer Information Technologies (ACIT) , 2025, pp. 859-862.

[9] J. Angeljulie, M. Vishal, J. Sathish, and V. Vignesh, "Real-Time Crowd Anomaly Detection in Surveillance Utilizing ML," in Proceedings of the 10th International Conference on Communication and Electronics Systems (ICCES) , 2025, pp. 1601-1605.

[10] M. Akpulat and M. Ekinci, "Anomaly Detection in Crowd Scenes Via Cross Trajectories," Applied Intelligence , vol. 55, article 525, 2025.

[11] K. Chidananda and A. P. Siva Kumar, "An Efficient Real Time Anomaly Detection in Surveillance Videos Using PRU-DPCN Classifier," SN Computer Science , vol. 6, no. 1, article 31, 2025.

[12] A. Dionis-Ros, J. Vila-Francés, R. Magdalena-Benedito, F. Mateo, and A. J. Serrano-López, "Multimodal Video Analysis for Crowd Anomaly Detection Using Open Access Tourism Cameras," Applied Sciences , vol. 14, no. 23, article 11075, 2024.

[13] M. Akpulat and M. Ekinci, "Anomaly Detection in Crowd Scenes Via Cross Trajectories," Applied Intelligence , vol. 55, article 525, 2025.

[14] M. Gozet, M. Karakose, and A. E. Yilmaz, "YOLOv11-Based Explainable Framework For Anomaly Detection In Crowded Scenes Using Attention Fusion," in 2025 15th International Conference on Advanced Computer Information Technologies (ACIT) , 2025, pp. 859-862. .

[15] Y. Ito, T. Sasaki, and S. Kondo, "Crowd Anomaly Detection from Drone and Ground," IEEE Access, vol. 13, pp. 1–15, 2025.

[16] H. Ye, J. Wang, and F. Ye, "K-fold Matching Model for Crowd Behavioral Anomaly Detection from Discontinuous Inputs," Expert Systems with Applications, vol. 261, article 125504, 2025.

[17] N. Chrisandy, K. R. R. Wardani, I. Martina, and H. Heryanto, "You Only Look Once v5 and Long Short-Term Memory Implementation for Crowd Anomaly Detection," Jurnal Teknik Informatika (JUTIF), vol. 6, no. 2, pp. 951–964, 2025.

[18] A. A. Alsabei, T. M. Alsubait, and H. H. Alhakami, "Enhancing Crowd Safety at Hajj: Real-Time Detection of Abnormal Behavior Using YOLOv9," IEEE Access, vol. 13, pp. 37748– 37761, 2025.

[19] R. Chaudhary and M. Kumar, "Optimized Deep Maxout for Crowd Anomaly Detection: A Hybrid Optimization-Based Model," Network: Computation in Neural Systems, vol. 36, no.1, pp. 148–173, 2025.

[20] Y. Hu, Y. Lin, H. Yang, B. Liu, G. Wan, J. Hong, C. Xie, W. Wang, and X. Lu, "CLDE-Net: Crowd Localization and Density Estimation Based on CNN and Transformer Network," Multimedia Systems, vol. 30, no. 3, article 120, 2024.9132.