Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

Deep Learning-Based Crowd Safety and Risk Prediction: A Review of Density Estimation, Behavioral Anomaly Detection, and Early Warning Approaches

Jansi S
Department of CSE,
Achariya College of Engineering,
Technology Puducherry, India

Dharanya D
Department of CSE,
Achariya College of Engineering
Technology Puducherry, India

Mahalakshmi A
Department of CSE,
Achariya College of Engineering
Technology Puducherry, India

Mohanasri S
Department of CSE,
Achariya College of Engineering
Technology Puducherry, India

Sivabarathi S
Department of CSE,
Achariya College of Engineering
Technology Puducherry, India

***Abstract* - Ensuring public safety in crowded environments such as railway stations and transportation hubs, religious services, shopping centers and public events is an important topic. Higher density, less-predictable crowd movement, occlusion and changing environmental conditions complicate manual observation and can delay the recognition of dangerous conditions. Deep learning approaches have enabled progress in methods for counting and estimating density in crowds, detecting and tracking human individuals and detecting abnormal behavior but many methods focus on their specific tasks and are not linked with an integrated mechanism for risk anticipation and timely intervention. This article reviews recent work on artificial intelligence-based crowd monitoring with focus on methods for estimating crowd density, lightweight models, detecting behavioral anomaly, predictive crowd analysis and privacy-aware edge-based deployment. The collected studies are reviewed according to major technical categories, and are compared in terms of their goals, deployment issues and reported limitations. The review reveals recurring areas of difficulty concerning extreme crowd density and robustness to environmental changes, generalization across scenes, separate pipelines for different tasks, real-time operation and privacy. With regard to these issues, conceptual Intelli-Watch architecture is sketched to link crowd-state estimation, movement analysis, and risk forecasting with human-oversight and alerting. The architecture is presented as a research direction and has not yet been implemented or validated.**

Index Terms* - *Crowd safety, deep learning, crowd density estimation, anomaly detection, risk prediction, video analytics, edge artificial intelligence, early warning.

I. INTRODUCTION

Ensuring public safety in crowded environments such as railway stations and transportation hubs, religious services, shopping centres and public events involves continuous monitoring for signs of high congestion, abnormal movement and conditions that may escalate into incidents. The density and movement patterns of a crowd in a transportation hub, a religious space, a commercial complex or a large public gathering can change rapidly over time. If this is detected at a late stage of development, there may be a very limited window of time for security personnel to act. Manual visual observation is important, but it is difficult to keep a consistent view of multiple camera feeds and large areas.

Computer vision and deep learning have enabled the surveillance video to be analysed automatically. There are several related tasks to which existing methods have been applied: detecting individual people in images, estimating the crowd count in an image, generating a density map of people, tracking their movement and detecting people who are engaged in an unusual or dangerous activity. The ability to estimate density is important because the number of people can overlap heavily and cannot always be reliably detected separately. People tracking and movement analysis give further information on their movement trajectories and the flow of the crowd in particular directions, while behavioural anomaly recognition seeks to identify patterns which deviate from what is expected.

More recent approaches consider lightweight architectures and edge-based processing to facilitate practical deployment [1]–[4].

Despite this progress, the task of crowd monitoring is still challenging when the crowd has high density or is relatively uncontrolled. Occlusion, changes in the illumination conditions, camera perspective or background appearance, and the visually-similar appearance of multiple people can all decrease the reliability of a model. Further, detecting an abnormality which has already been developing in an image does not exactly equate to the ability to predict a potentially dangerous condition before it becomes visible. A density-estimation model which does not analyse crowd movement might miss evidence of directional crowd conflict, while an individual behavioural detector without a density-aware component might not understand an emerging high-density situation.

The literature indicates three closely related shortfalls: Limited ability to be robust in extremely high-density crowd scenes, insufficient efforts to bring reactive detection and predictive risk analysis together, and an inability to combine

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

density estimation, tracking and behavioural anomaly detection into a single pipelined solution [1], [2], [5]. Deploying a model for crowd monitoring raises challenges related to processing delay, available hardware and computer networks, and privacy [3], [6].

This article reviews some recent work on AI-based crowd monitoring, and examines what limitations of the reviewed studies could guide and inform work in the Intelli-Watch project, which is focused on intelligent crowd safety, crowd-risk prediction and emergency response. The key contributions are: (i) a thematic organization of recent work relating to density-estimation models, lightweight deployment, behavioural anomaly detection, and integrated crowd monitoring; (ii) a discussion of the comparative limitations which recur through the reviewed material; (iii) a table summarizing research gaps which links each limitation to the reviewed studies; (iv) a conceptual Intelli-Watch architecture, which attempts to link crowd-state estimation and movement analysis, and risk forecasting, to human-oversight and alerting; and (iv) a set of research directions for future investigation.

II. REVIEW SCOPE AND METHODOLOGY

This review attempts to synthesize a set of 14 research papers which are collected in the Intelli-Watch literature survey project. These are survey articles and individual system proposals from 2026. They are concerned with crowd counting and density estimation, detection and tracking, behavioural anomaly recognition, and edge-based systems; and they discuss crowd-safety applications. The authors of this article did not perform any additional database search.

The review organizes the literature into categories according to the primary technical contribution of each piece of work. The survey articles are used to summarize the broader trends and recurring limitations of the collected work. The architecture-oriented studies are analysed for how they address changes in density, the way in which feature extraction works and the associated efficiency. The system-oriented papers are discussed in terms of the extent to which they are combined into an integrated architecture, the alerts, privacy issues and deployment concerns they address. Descriptions of individual studies are limited to what is indicated by the title and other abstract-level descriptions; data-sets, metrics and numerical results are thus reported as "Not reported" rather than being inferred.

The present review is a structured narrative review, not a formalized systematic review. A systematic review would additionally require a description of the databases and search-strings used to locate studies, date limitations for searching, inclusion and exclusion criteria, a formalized screening process and other considerations which should be documented and made reproducible. The observations should thus be viewed as interpretation from the literature the authors have collected, rather than as an exhaustive evaluation of the entire body of crowd-monitoring research. Where they go beyond the studies cited, their statements are the authors' interpretation, and are identified as such.

There are three limitations to this review, which it is necessary to state. First, the evidence we refer to is small and confined to a single publication year; trends that seem obvious such as the emphasis on lightweight and edge-oriented models could reflect more of the composition of the collection than the progress of the field. Second, because the comparisons rely on the title- and abstract-level information, the strengths and weaknesses of the work which are attributed to individual studies are cautious and may be adjusted by reading the full text of a study. Third, the research gaps described in Section VI are a synthesis of the observed recurring themes, and have not been validated by experimentation.

III. DEEP LEARNING APPROACHES FOR CROWD MONITORING

A. Crowd Density Estimation and Counting

Crowd density estimation offers information on the number and spatial arrangement of people in a scene. Detection-oriented methods identify individuals or people's heads and aggregate the number of detected people or heads to estimate the crowd count. These methods may perform well in relatively clear crowd scenes, but it is more difficult to localize individual people with heavy overlapping and occlusion, as with extremely dense crowds.

An alternative to this is provided by density-map regression, in which a spatial representation of the crowd density is predicted instead of requiring each person to be separated and detected. Multiscale feature extraction is important, as the apparent size of people in an image varies with the distance and perspective of the camera. A variety of architectures have been tried to represent these variations, including multicolumn networks and later density-regression models with feature extraction.

New studies on multiscale gated feature aggregation and progressive refinement examine how counting networks can be combined for feature extraction while addressing the potential problems of parameter redundancy and background robustness [7]. CPMS-Net, a recent study, explores an efficient cooperative-perception architecture with multiscale feature extraction for crowd-counting [4]. A recent survey discusses the challenges of performing crowd counting for complex environments, and the emergent trends [2]. These studies illustrate the ongoing effort to find a suitable design for balancing spatial detail, model complexity and counting accuracy.

Nonetheless, the problem of density estimation remains vulnerable to a distribution shift between different scenes. A model which was trained with a particular camera perspective or crowd environment may not achieve the same performance in another scene, and the problem is more difficult if the image quality is lower or there are extremely dense crowd areas in the scene. The ability to estimate density should not only be assessed by an average counting error but should also be considered across levels of crowd and environmental conditions.

B. Human Detection, Tracking, and Movement Analysis

Human detection and tracking can augment the information provided by density estimation. Detection offers information on people or parts of their body which may be relevant to the situation, while tracking methods associate the observations from one video frame to the next. The resulting trajectories may be used to analyse the direction and speed of movement, consistency of direction, and how neighbouring groups interact with each other.

An integrated system for human detection and tracking, in combination with crowd analysis, trajectory and heat-map estimation for a commercial complex or public transport hub takes people localization into account alongside crowd

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

analysis, trajectories and a crowd heat-map [8]. Such capabilities could help a monitoring system represent how a crowd is moving instead of relying simply on the static number-of-people estimate.

However, tracking can become difficult if people frequently occlude each other in the image or if they move in and out of the camera's field of view. There can be an accumulation of identity association errors over time, and a camera cannot always provide a good overview of the movement across the whole venue. The information provided by tracking needs to be interpreted with knowledge about the camera geometry, missing observations, and the associated uncertainty.

The directional flow is of special importance to crowd safety. The estimated density of two scenes may be comparable, but in one scene people may be moving in an orderly fashion whereas in the other they may be moving in opposing or converging directions, increasing the risk of an incident. This motivates a combination of the density-map with additional features of the movement over a sequence of images, instead of treating the crowd-count estimate as the sole indicator of the situation.

C. *Behavioral Anomaly Detection*

The purpose of behavioural anomaly detection is to identify people who are engaged in a movement pattern or activity that deviates from a pattern which has been learned or defined, such as sudden group movement, unusual directional changes, or localized movement irregularity. The method may be based on frame-level appearance features, motion features, tracked trajectories, or temporal representations which are extracted from the video.

Recent work has examined a real-time crowd behavioural anomaly detection and an alert sound system [9]. Other work combines crowd-surveillance and anomaly analysis into a single system [5]. These studies demonstrate the importance of linking an understanding of video to an operational alerting system.

The detection of an abnormality does not automatically equate to a prediction of a dangerous situation. It may be possible that an unusual pattern of movement has not caused a dangerous situation, and conversely that a dangerous situation develops as a slowly increasing crowd density without any clearly abnormal action. A practical crowd-safety system should be able to distinguish between unusual observations and a sign that a dangerous situation is emerging. It should also consider the number of false alarms and missed events, the confidence in each predicted event, and contextual information to help understand the circumstances of the event.

D. *Crowd Risk Prediction and Early Warning*

The goal of predictive crowd analysis is to forecast how the current conditions might develop over a period of time. Relevant data could include the recent density estimate, the change in occupancy over time, movement direction, irregularity in flow and the persistence of an anomaly that has been detected.

A study on directional anomaly detection in the dense crowd during the Hajj as an aid to proactive safety management discusses the need to move beyond reactive anomaly detection, and to investigate directional flow as an indicator of the development of high-density areas [10]. This direction of study is important because safety personnel benefit from some warning before a location passes into a critical state.

Nonetheless prediction requires temporally representative data and should be carefully validated. Models which have been trained on a limited set of scenarios may overfit to particular locations or types of movement. The prediction horizons should also be carefully defined: if a model classifies the current frame as normal, this should not be described as a prediction of a future event. Future systems should provide information on the warning lead time, prediction error, false-alarm rate and performance in different conditions.

IV. LIGHTWEIGHT MODELS, EDGE AI, AND PRIVACY

A. *Efficiency and Real-Time Operation*

A crowd-monitoring model should be able to process a stream of video fast enough for its intended purpose. The computational demands depend on the input resolution, network architecture, the number of cameras, available hardware and frequency of prediction. A high-capacity model can offer a rich representation but it may be difficult to deploy it in a low-power, or computing-resource-restricted environment.

Studies on adaptive crowd-density estimation for real-time video analytics emphasize the deployment-oriented concerns of the system in terms of latency and throughput [3]. Lightweight multiscale architectures also try to decrease the computational cost while representing varying crowd sizes [4]. These efforts show the trade-off between architectural design features. The most accurate model on a benchmark is not necessarily the most suitable for continuous operational monitoring.

A fair comparison should describe the counting quality or detection ability, alongside the latency, throughput, memory utilization and hardware configuration. A statement like "real-time" operation should be supported by a measured processing speed under stated conditions rather than inferred from the architecture.

B. *Edge-Based and Privacy-Aware Processing*

A crowd-monitoring model should be able to process a stream of video fast enough for its intended purpose. The computational demands depend on the input resolution, network architecture, the number of cameras, available hardware and frequency of prediction. A high-capacity model can offer a rich representation but it may be difficult to deploy it in a low-power, or computing-resource-restricted environment.

Studies on adaptive crowd-density estimation for real-time video analytics emphasize the deployment-oriented concerns of the system in terms of latency and throughput [3]. Lightweight multiscale architectures also try to decrease the computational cost while representing varying crowd sizes [4]. These efforts show the trade-off between architectural design features. The most accurate model on a benchmark is not necessarily the most suitable for continuous operational monitoring.

A fair comparison should describe the counting quality or detection ability, alongside the latency, throughput, memory utilization and hardware configuration. A statement like "real-time" operation should be supported by a measured processing speed under stated conditions rather than inferred from the architecture.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

V. COMPARATIVE ANALYSIS OF THE LITERATURE

The reviewed methods differ in their goals and operating principles. Density-estimation systems emphasize the crowd-count or crowd density; tracking systems place more emphasis on identity association and trajectory; anomaly detectors concentrate on detecting unusual activity; predictive systems aim to provide an estimate of future conditions; and edge-oriented studies emphasize deployment efficiency and privacy concerns. There are application-oriented studies for railway stations [12], people localization in high-resolution drone imagery [13], and temple environments with multi-channel alerts [14]. The main categories and their recurring limitations are summarized in Table I, and the individual studies are listed in Table II.

TABLE I. THEMATIC COMPARISON OF CROWD-MONITORING APPROACHES

Approach	Main objective	Recurring limitation
Detection-based counting	Localize people or heads and estimate counts	Occlusion and overlap in dense scenes
Density-map regression	Estimate spatial crowd concentration	Domain shift and environmental variation
Multi-scale lightweight models	Balance density estimation and efficiency	Accuracy loss in complex scenes
Detection and tracking	Associate people over time; analyze trajectories	Occlusion, identity switches, incomplete coverage
Behavioral anomaly detection	Identify unusual activity or motion	False alarms; limited predictive context
Temporal risk prediction	Forecast future crowd conditions	Data requirements; scenario-dependent generalization
Edge video analytics	Reduce central processing and transmission	Hardware limits, cost, privacy governance

The comparison reveals that it is not possible to judge a model based on a single objective function. A model can make an accurate count but fail to offer any useful directional information. An anomaly detector can flag the presence of an unusual movement but not give any clue as to whether a dangerous level of density is in danger of being passed. A lightweight edge model can be practical to deploy but can still require testing under extremely high-density conditions.

The comparison should therefore consider several aspects when evaluating a system: density-estimation error, detection performance, trajectory consistency, anomaly precision and recall, warning lead-time, processing latency, and processing and memory usage. It is also necessary to consider the differences between the studies, the data-set and camera perspectives, annotations and hardware. The numerical results from two different types of experimental set-ups should not be ranked directly, unless the evaluation condition is known to be similar.

TABLE II. SUMMARY OF THE REVIEWED STUDIES

Ref.	Research focus (from title/abstract)	Data, metrics, results; relevance to Intelli-Watch
[1]	Review of AI-based crowd density prediction and control	Not reported; broad survey of density prediction and control
[2]	Survey of deep crowd counting in complex environments	Not reported; challenges and trends for dense-scene counting
[3]	Adaptive density estimation for real-time video analytics	Not reported; latency and throughput considerations
[4]	Lightweight cooperative-perception multi-scale counting network	Not reported; efficient multi-scale density estimation
[5]	Crowd surveillance with anomaly detection	Not reported; linking surveillance and anomaly analysis
[6]	Serverless privacy-preserving edge AI (traffic monitoring)	Not reported; edge/privacy design, different domain
[7]	Counting with multi-scale gated aggregation, progressive refinement	Not reported; feature fusion and background robustness
[8]	Human detection and tracking in commercial/transport hubs	Not reported; trajectories and heatmaps for movement analysis
[9]	Real-time behavioral anomaly detection with alert sound	Not reported; alert generation after detection
[10]	Directional anomaly detection in dense Hajj crowds	Not reported; proactive safety through directional flow
[11]	Privacy-aware edge video analytics for crowd management	Not reported; scalable, privacy-aware architecture
[12]	Crowd density analysis for railway stations	Not reported; station-specific density monitoring
[13]	People localization in high-resolution drone imagery	Not reported; localization for crowd management
[14]	Density monitoring with multi-channel alerts, temples	Not reported; alert delivery in a venue setting

VI. CROSS-CUTTING RESEARCH GAPS

These gaps are interpretations based on the thematic review; the table links them to the underlying studies (Table III). Descriptors such as “limited” mean that the collected material offers insufficient evidence, not that a subject is unexplored.

TABLE III. RESEARCH GAPS AND RELEVANCE TO INTELLI-WATCH

Research area	Remaining gap	Relevance to Intelli-Watch
Dense-scene robustness [2], [4], [7]	Limited evaluation across density levels, occlusion, and unseen scenes	Density module needs multi-level, cross-scene testing
Predictive safety [9], [10]	Reactive detection dominates; warning lead time rarely reported	Risk forecasting with defined horizon and lead time
Pipeline integration [5], [8]	Insufficient integration of density, tracking, and anomaly outputs	Common workflow combining all signals
Generalization and explainability [3], [5]	Transfer across venues and cameras requires further investigation	Alerts with trend, region, and confidence
Deployment and privacy [3], [6], [11]	Accuracy–speed trade-offs; privacy governance beyond edge processing	Lightweight, edge-oriented, data-minimizing design
Alert delivery [9], [14]	Link between risk detection and useful emergency alerts is limited	Human-supervised alerting

A. Robustness Under Extreme Density

Occlusion and overlapping people remain an issue for dense crowds. Variation in the illumination condition, camera perspective, image quality and background appearance can also decrease robustness. A model which performs well for an ordinary crowd scene may not maintain performance during an extremely high-density event. Future evaluation should

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

include multiple levels of density, difficult environmental conditions, and different scenes that were not used during training.

B. From Reactive Detection to Predictive Safety

Many systems seek to identify the crowd-state which is currently in effect, or detect an anomaly which can be seen in a scene, and a dangerous situation may have little time for people to act after it is detected. Directional flow, changes over time in the density estimate and the persistence of irregularity in the movement could offer signals to estimate the development of a high-density area. The research question would be whether signals could offer sufficiently early and reliable warnings without over-producing false alarms.

C. Fragmented Processing Pipelines

Crowd-density estimation, tracking systems, behavioural anomaly detection and forecasting are usually designed to operate as independent tasks. Different pipelines can use different data sets, representations and evaluation metrics, and it can be difficult to interleave the result of these pipelines into a coherent risk-assessment system. An integrated system could investigate ways of sharing knowledge between the parts which each relies on their own reliability while also providing a manageable computational effort.

D. Generalization and Explainability

The model performance can decrease when it is transferred into a new location, camera configuration or environmental condition, which is critical when applied to safety-critical applications. An alert should also provide human operators with some useful information to understand the situation, such as the affected region, the density trend, the movement pattern, and a confidence value in the model prediction, so that the human operator can make their own assessment of the situation rather than taking the model output as a given.

E. Deployment, Privacy, and Operational Cost

Big models, multiple cameras and a need for continuous video transmission to a remote location increase the hardware and infrastructure requirements. An edge-based system with inference at the local level reduces the network requirements, but low-power hardware will have its own limitations in terms of a model's accuracy. The requirements which can be practical to deploy should also consider privacy and access controls, data retention, and the implications of missed or mislabelled alerts. These requirements must be considered at the design stage rather than being added later.

F. Emergency Alerting and Operational Integration

Identifying a problematic situation is of no benefit if that signal cannot be conveyed to individuals who have the means to address it. Of the studies reviewed, alerting mechanisms are treated in a sparse and domain-dependent manner: One publication combines anomaly-detection with an audible warning [9], while a second outlines multi-channel alerting for temple environments [14].

At the level of detail available in abstracts, these publications demonstrate that alerting may be linked to a vision pipeline but cannot provide answers to how alerting thresholds are determined, the rate of false alarms, or how alerts are addressed by operators.

Another consideration in emergency response is the existence of parameters that are not described by the vision model, such as staffing levels, exit capacities, and event schedules. To be a practical solution, a system should treat its own output as a single component of the decision-making process, while also prioritizing alerts to avert excessive re-prompting by the model of low confidence alerts. Whether an alerting system that uses grades or alert escalations or operator feedback leads to improved response times in operational venues is an avenue for investigation.

VII. PROPOSED INTELLI-WATCH CONCEPTUAL FRAMEWORK

From these limitations, the Intelli-Watch conceptual framework is proposed to provide an idea of how crowd-state estimation could be integrated with movement and risk-driven alerting (Fig. 1). The purpose of this design would be to support human operators by structuring evidence available to them in a way that common practice in monitoring. The framework is prospective and is not intended to provide a history of implemented systems and experimental validation of the framework's design; distinctions are made between published studies' capabilities, proposed components here, and future work throughout.

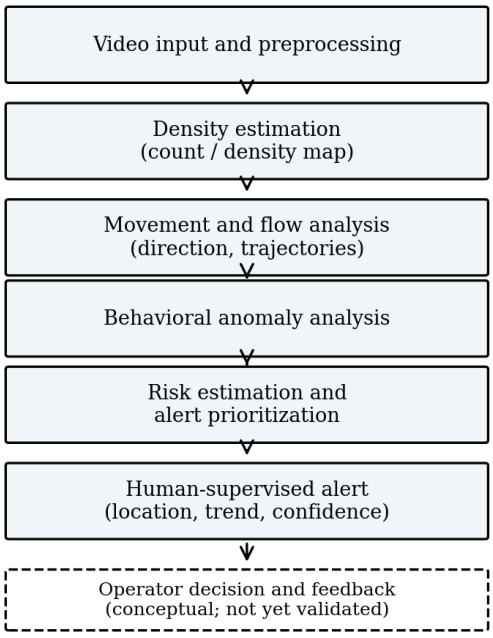


Fig. 1. Conceptual Intelli-Watch workflow. The pipeline is a proposed design and has not been implemented or validated.

A. Framework Components

1) Video input and preprocessing: Authorized cameras are drawn from frames in input, and pre-processing may consist of resizing, frame sampling, image quality, and other tests. Camera placement should be considered when interpreting density and movement; viewing geometry will affect interpretation and should also be considered.

2) Crowd density estimation: A Crowd concentration is estimated using a vision model that provides a crowd count, density map, or region-level estimates, with confidence information where available.

3) Movement and flow analysis: Movement direction and temporal variance is analyzed on consecutive frames. Where trackable, trajectories may provide evidence of converging traffic patterns, opposing traffic, or persistent flow irregularities.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

4) Behavioral anomaly analysis: A temporal module that detects patterns that deviate from expected movements. Output from this component would be interpreted in conjunction with density and flow information, not as an exclusive marker of potential danger.

5) Risk forecasting and alert prioritization: A future implementation may combine density trends, flow information features, anomaly scores, and persistence to provide an estimate of risk-level. Horizons, thresholds, and model calibration would need to be determined and evaluated empirically.

6) Human-supervised alerting: Based on prespecified criteria, the framework may issue an alert containing the location and recent trends contributing signals and confidence in the detection. Relevant and authorized people would be awareness of and would analyze the evidence, determining whether any action is appropriate. The outputs of the model would assist, not replace, the operator's judgment.

B. Expected Research Value

The research value of the design of the Intelli-Watch system is that it attempts to explore the potential for disparate capabilities in monitoring to be chained into an effective process that aids decisions about safety. Its potential benefits are hypotheses to be tested, not outcomes achieved. Experiments would be intended to measure the reliability of warnings based on combinations of density information and movement information versus baselines of density-only and anomaly-only warning. Lightweight inference, minimal transmission of unprocessed footage, and clear and operator-focused explanations may improve the viability of the system for a real-world deployment, but they will depend on how the system is implemented, the hardware used, the datasets available for evaluation, and operational testing.

C. Evaluation Plan

A future implementation should benefit from evaluation of component and of the integrated system as a whole. Density estimation could be evaluated using mean absolute error and squared error measures where ground truth is available. Detection and tracking could be evaluated using task-specific measures of precision and recall, though metrics appropriate to tracking would be needed. Anomaly detection would be task for which precision and recall, along with false-alarm rate and missed-event rate would be reported. Risk forecasting might be evaluated in terms of prediction error and warning lead time. Evaluation of a deployment would benefit from measuring latency and throughput, as well as memory consumed, hardware used, and communications involved.

Ablation studies varying between density-only and anomaly-only and movement-aware variants, as well as between scenes to explore the degree to which the framework generalizes to other venues, and to other camera conditions should be carried out. Until such experiments are carried out, no improvement can be quantitatively reported, nor real-world benefit to safety claimed.

D. Feasibility, Privacy, and Governance

The feasibility of the framework will depend on the camera coverage of a venue, the compute power available, and network circumstances. A staged deployment may be warranted, with density estimates and simple analysis of movement being implemented first, then tracking, anomaly scores and forecasting as each is proven valid. Carrying out inference at or close to the camera, discarding raw frames once they have been analyzed by inference models, and storing only aggregated indicators to record and disseminate, may minimize the exposure to potential privacy issues, as recommended in [6] and [11]. These do not substitute for the need of legal and organizational protections for privacy, including a defined purpose, control of access, limits on retention, and transparency to the public about the nature of the monitoring.

Finally, the framework bears the weaknesses of its components; errors in density estimation propagate to estimates of risk, and failures in tracking in a dense scene could skew features used for the flow analysis. Uncertainty should be propagated forward through the pipeline, and failure of any component of the system should degrade it gracefully rather than suppressing or silently fabricating an alert.

VIII. FUTURE RESEARCH DIRECTIONS

Several lines of research arise from the limitations identified. First, the design of models could incorporate multi-task learning to allow density, movement, and anomaly detection tasks to share representations without negative interference. Second, domain adaptation and evaluation on new scenes can determine whether the models generalize to environments for which they were not trained.

Third, adaptive inference and model compression might improve performance in a scene by devoting more compute resources to uncertainty or rapidly changing scenes, and more for normal operation. Fourth, temporal risk models should be tried for calibration, lead time of warnings, and ability to be robust against rare events. Fifth, explanability and uncertainty estimation should be built into the alert interface so that operators understand the basis for the warning.

Lastly, deployment in a privacy-aware fashion will require a systematic analysis of how to minimize the data held, how to ensure security, how long to retain the data, and how much processing should occur at the edge. Common evaluation protocols and a broad range of datasets will help to compare systems under a common set of conditions, and to determine whether improvements generalize to the environments of public venues in the real world.

IX. CONCLUSION

This review analyzed deep learning approaches for crowd density estimates, detection and tracking, behavioral anomaly detection, predictive safety analysis, and edge-oriented video monitoring. The collected literature suggests that a number of challenges have not been eliminated through progress on individual components but are instead intertwined. Extreme densety and environmental variation may limit the reliability of models; anomaly detection cannot provide advance warning by itself; and independent pipelines cannot support fusion of crowd-state and movement information. The ability to perform in real time, generalize across scenes, provide explainability, and deploy with respect to privacy remain deployment considerations.

The Intelli-Watch conceptual framework addresses these challenges by connecting density estimation, movement analysis, anomaly identification, risk forecasting, and human-supervised alerts in a workflow. Its value must be demonstrated through implementation, baseline comparison, cross-scene evaluation, and analysis of false alarms and warning lead time. The overall trend is to shift from individual

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 10 , October - 2026

crowd-analysis functions to integrated, efficient, privacy-conscious, and explicable decision support for crowd safety.

REFERENCES

[1] G. J. Chhajed, V. N. Madhure, P. S. Bansode, P. D. Wagh, and D. R. Dange, "AI-based crowd density prediction and control system: A comprehensive review," in *Proc. 2026 Int. Conf. Emerging Trends and Innovations in ICT (ICEI)*, 2026, pp. 1–6, doi: 10.1109/ICEI65890.2026.11447564.

[2] H. F. Elsepae, H. M. El-Hoseny, E. K. I. Hamad, and E.-S. M. El-Rabaie, "Deep learning for crowd counting in complex environments: Challenges and novel trends," *Discover Comput.*, vol. 29, no. 1, Art. no. 101, 2026, doi: 10.1007/s10791-026-09928-8.

[3] T. P. Patel, "Adaptive AI-based crowd density estimation for real-time video analytics in smart public spaces," in *Proc. 2026 IEEE Int. Conf. Advanced Systems and Emergent Technologies (IC_ASET)*, 2026, pp. 1–6, doi: 10.1109/IC_ASET69920.2026.11502455.

[4] G. Shang, H. Weng, and X. Lu, "CPMS-Net: A lightweight cooperative perception multi-scale network for crowd counting," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 38, 2026, doi: 10.1007/s44443-026-01052-3.

[5] S. K. Singh, N. Maurya, R. Kaushik, S. Sisodia, D. Sharma, and R. Sandhu, "AI based crowd surveillance and anomaly detection system," in *Proc. 2026 Int. Conf. NextGen Data Science and Analytics (ICNDSA)*, 2026, pp. 1–6, doi: 10.1109/ICNDSA68777.2026.11652341.

[6] S. Bhunia, A. Pakhira, A. Chanda, D. Das, and B. Saha, "CrowdLense: A serverless, privacy-preserving edge AI framework for sustainable crowd-sourced real-time traffic monitoring," in *Proc. 18th Int. Conf. Communication Systems and Networks (COMSNETS)*, 2026, pp. 1261–1263, doi: 10.1109/COMSNETS67989.2026.11418200.

[7] S. Xu, W. Xu, and S. Zhang, "Crowd counting network based on multi-scale gated aggregation and progressive refinement," *IEEE Access*, vol. 14, pp. 82315–82325, 2026, doi: 10.1109/ACCESS.2026.3698077.

[8] N. S., T. K. Kumar, I. Mallikarjun, M. Sri, and K. Jayaprakash, "AI-driven human detection and tracking system for commercial complexes and public transport hubs," in *Proc. 2026 Int. Conf. Machine Learning and Autonomous Systems (ICMLAS)*, 2026, pp. 951–957, doi: 10.1109/ICMLAS67792.2026.11483876.

[9] V. Devi, D. Satheesh Kumar, K. S., R. M., S. A., and V. S., "AI-powered real-time crowd behavioral anomaly detection with alert sound system," in *Proc. 6th Int. Conf. Pervasive Computing and Social Networking (ICPCSN)*, 2026, doi: 10.1109/ICPCSN68523.2026.11543710.

[10] R. Alturki, M. Wedyan, and A. Nasayreh, "AI-powered directional anomaly detection in dense Hajj crowds for proactive safety management," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 38, no. 6, 2026, doi: 10.1007/s44443-026-00793-5.

[11] D. Dawadi, U. S. Yandamuri, S. Kumar V, J. L. Aldo Stalin, and R. Naveenkumar, "Privacy-aware edge-based intelligent video analytics for scalable crowd management in smart cities," in *Proc. 3rd Int. Conf. Integrated Intelligence and Communication Systems (ICIICS)*, 2026, pp. 1–7, doi: 10.1109/ICIICS67880.2026.11483444.

[12] S. Alex David, K. Yashwanthi, K. Harshitha, and G. Shanmukh Pavan Kumar, "AI-based crowd density analysis for railway stations," in *Proc. IEEE Int. Students' Conf. Electrical, Electronics and Computer Science (SCEECS)*, 2026, pp. 1–6, doi: 10.1109/SCEECS68810.2026.11430145.

[13] B. Ptak and M. Kraft, "Enhancing people localization in drone imagery for better crowd management by utilizing every pixel in high-resolution images," *IEEE Access*, vol. 14, pp. 86534–86550, 2026, doi: 10.1109/ACCESS.2026.3700840.

[14] S. Singh, L. S. Maurya, A. Maurya, P. K. Srivastava, and A. Prasad, "Real-time AI-based crowd density monitoring and multi-channel alert system for temple environments," in *Proc. 2026 Int. Conf. Green Technologies and Sustainable Solutions (GTSS)*, 2026, pp. 501–507, doi: 10.1109/GTSS68743.2026.11679619.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)