International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 07, July 2026
Received: 30-05-2026 Accepted: 03-07-2026 Published: 09-07-2026

EXPLAINABLE CROWD COUNTING: AI-POWERED DENSITY ESTIMATION FOR SMARTER SURVEILLANCE SYSTEMS

[1] Mrs. K Swayam Prabha, [2] Barthala Vidya Sree, [3] Guntoju Navya Sri, [4] Akki Sheshu
[1] Assistant Professor, [234] B. Tech Students
[1]Department of Computer Science and Engineering
[234]Department of CSE(CYBER SECURITY)
[1234]Sree Dattha Group of Institutions, Sheriguda, Ibrahimpatnam, 501510, Telangana, India

ABSTRACT

The rapid growth of urban populations, public transportation networks, smart cities, large-scale events, commercial complexes, educational campuses, religious gatherings, and critical public infrastructure has increased the need for accurate and intelligent crowd monitoring. Conventional surveillance systems primarily depend on manual observation, object detection, background subtraction, or direct person-by-person counting, which often become unreliable under severe occlusion, perspective distortion, illumination variation, scale changes, dense crowd formations, and complex environmental conditions. This research proposes Explainable Crowd Counting: AI-Powered Density Estimation for Smarter Surveillance Systems, an integrated framework that combines intelligent video acquisition, image preprocessing, deep convolutional feature extraction, multi-scale representation learning, density-map estimation, crowd-count regression, explainable artificial intelligence, risk classification, and smart surveillance alerting. The proposed system continuously acquires frames from CCTV cameras, IP cameras, drones, and smart-city surveillance networks and applies frame validation, resizing, normalization, noise reduction, perspective-aware preprocessing, and region-of-interest extraction. A deep learning analytical core integrates Convolutional Neural Networks, multi-scale feature pyramids, attention mechanisms, dilated convolutions, and density-estimation networks to capture local and global crowd patterns under varying density conditions. Instead of relying exclusively on bounding-box detection, the framework estimates spatial density distributions and derives crowd counts from learned density representations, making it more suitable for highly congested scenes. To improve transparency, explainable AI mechanisms generate attention maps, saliency visualizations, region-level contribution analysis, confidence information, and density overlays that help surveillance operators understand why the system predicts a particular crowd level. The framework dynamically classifies scenes as Low Density, Moderate Density, High Density, or Critical Congestion and supports actions such as standard monitoring, operator notification, congestion warning, access regulation, route diversion, emergency escalation, and crowd-management intervention. The architecture consists of five interconnected layers: Crowd Scene and Video Acquisition, Image Preprocessing and Perspective-Aware Data Engineering, AI-Powered Multi-Scale Density Estimation, Explainable Crowd Intelligence and Risk Assessment, and Smart Surveillance Applications and Continuous Learning. Illustrative conceptual evaluation indicates improved counting accuracy, precision, recall, F1-score, explainability efficiency, and reduced analytical response time compared with traditional image-processing methods, conventional object detection, and basic CNN-based counting. The proposed framework provides a scalable foundation for transparent and intelligent crowd surveillance across transportation hubs, stadiums, festivals, shopping centers, smart cities, campuses, and public safety environments.

Keywords: Crowd Counting, Explainable Artificial Intelligence, Density Estimation, Deep

Learning, Convolutional Neural Network, Smart Surveillance, Computer Vision, Multi-Scale Feature Learning, Attention Mechanism, Density Map, Crowd Analytics, Public Safety.

I. INTRODUCTION

The increasing concentration of people in cities, transportation systems, public events, commercial centers, educational institutions, stadiums, religious gatherings, and entertainment venues has created significant challenges for public safety and crowd management. Surveillance cameras generate enormous volumes of visual information, but manual observation by security personnel is difficult, time-consuming, and vulnerable to fatigue. Intelligent crowd-counting systems can transform passive surveillance infrastructure into proactive analytical platforms capable of estimating occupancy and identifying potentially dangerous congestion [1].

Crowd counting is a fundamental computer-vision problem concerned with estimating the number of individuals present in an image or video frame. In relatively sparse environments, conventional person detectors can identify individual objects and count detected bounding boxes. However, direct detection becomes increasingly difficult in dense scenes because individuals overlap significantly, only partial body regions remain visible, and apparent object size varies according to camera perspective [2].

Traditional crowd-analysis techniques have used background subtraction, motion segmentation, handcrafted features, regression models, and object detectors. Although these methods can perform adequately under controlled conditions, they often struggle with complex backgrounds, illumination changes, shadows, camera motion, severe occlusion, and highly variable crowd densities. These limitations have motivated the development of deep-learning-based approaches capable of learning discriminative visual representations directly from data [3].

Convolutional Neural Networks have significantly improved crowd counting by learning spatial and semantic features from crowd images. Early CNN-based counting methods demonstrated that deep architectures could map image information to crowd estimates more effectively than manually engineered features. However, crowd scenes contain substantial scale variation because individuals close to the camera appear large while distant individuals may occupy only a few pixels [4].

Multi-column and multi-scale architectures were therefore introduced to process visual patterns at different receptive-field sizes. Such approaches attempt to capture heads or people appearing at multiple scales within the same scene. Multi-scale feature extraction remains important because a single fixed receptive field may not adequately represent both nearby individuals and highly compressed distant crowds [5].

Density-map estimation has become a major direction in modern crowd counting. Instead of explicitly detecting every individual, a model predicts a continuous spatial representation describing the distribution of crowd density across the image. This approach provides both an estimated total count and information regarding where crowd concentration occurs, making it especially valuable for dense surveillance scenes [6].

Perspective distortion remains a significant challenge because apparent person size changes according to camera geometry and scene depth. Perspective-aware models, adaptive receptive fields, multi-scale representations, and contextual features can help address this variation. Nevertheless, models trained in one environment may still perform poorly when deployed in new camera viewpoints, weather conditions, cultural settings, or crowd configurations [7].

Attention mechanisms provide another important capability by enabling deep models to emphasize informative spatial regions or feature channels. In crowd counting, attention can help the network focus on regions containing meaningful crowd evidence while suppressing irrelevant backgrounds such as buildings, roads, vegetation, advertisements, and empty areas [8].

Despite advances in counting accuracy, many deep crowd-analysis systems remain difficult to interpret. Surveillance operators may receive a numerical estimate without understanding which image regions influenced the prediction. This lack of transparency is particularly problematic when automated systems support public-safety decisions, congestion warnings, access restrictions, or emergency interventions [9].

Explainable Artificial Intelligence provides mechanisms for improving the transparency of complex models. Saliency maps, activation visualizations, attention maps, local explanations, region-level contributions, and confidence indicators can reveal which parts of an image contribute to model decisions. Explainability can help operators determine whether a prediction is based on genuine crowd evidence or irrelevant environmental patterns [10].

An explainable crowd-counting system should therefore combine numerical estimation with spatial interpretation. A surveillance operator may need to know not only that a scene contains a large crowd but also where the highest-density regions occur, whether entrances are becoming congested, whether specific zones exceed safe occupancy levels, and how confident the model is in its estimate.

II. LITERATURE SURVEY

Author: V. A. Sindagi and V. M. Patel (2018)

Sindagi and Patel presented a comprehensive survey of CNN-based crowd counting and density estimation methods. Their work examined major developments in deep crowd analytics, including regression-based counting, density-map prediction, multi-scale architectures, and contextual modeling. The study highlighted scale variation, perspective distortion, occlusion, and dataset diversity as central challenges in crowd counting [11].

Author: Y. Zhang et al. (2016)

Zhang and colleagues proposed a Multi-Column Convolutional Neural Network for single-image crowd counting. The architecture used multiple convolutional branches with different receptive fields to capture scale variations in crowd scenes. Their research became an important foundation for multi-scale density-estimation approaches [12].

Author: C. Zhang et al. (2015)

Zhang and colleagues investigated cross-scene crowd counting using deep convolutional neural networks. Their work addressed the difficulty of applying models across different surveillance scenes and demonstrated the importance of adapting learned representations to varying environments [13].

Author: D. B. Sam et al. (2017)

Sam and colleagues proposed a switching convolutional neural network architecture for crowd counting. The method used specialized regressors and a switching mechanism to select appropriate processing paths according to image-patch characteristics. The research demonstrated the importance of adapting model behavior to crowd-density variation [14].

Author: L. Boominathan et al. (2016)

Boominathan and colleagues developed a deep and shallow network combination for crowd counting in highly congested scenes. Their approach integrated complementary feature representations to address substantial scale variation and complex crowd structures, supporting the value of multi-resolution analysis [15].

Author: Y. Li et al. (2018)

Li and colleagues proposed CSRNet, which combined a convolutional feature extractor with dilated convolutional layers for highly congested scene understanding. The architecture demonstrated strong density-map estimation capability while preserving spatial resolution, making it influential in modern crowd-counting research [16].

Received: 30-05-2026 Accepted: 03-07-2026 Published: 09-07-2026

Author: X. Liu et al. (2019)

Liu and colleagues investigated context-aware crowd counting and emphasized the importance of incorporating contextual information across multiple scales. Their work demonstrated that local visual evidence alone may be insufficient in complex scenes and that broader contextual reasoning can improve density estimation [17].

Author: A. B. Chan et al. (2008)

Chan and colleagues developed privacy-preserving crowd-monitoring approaches based on aggregate visual features rather than explicit individual identification. Their work demonstrated early possibilities for estimating crowd characteristics while reducing dependence on person-level recognition [18].

Author: R. R. Selvaraju et al. (2017)

Selvaraju and colleagues introduced Grad-CAM, a visual explanation technique that uses gradient information to highlight image regions contributing to deep-network predictions. Although developed for broader computer-vision applications, the technique provides an important foundation for explainable surveillance and crowd-analysis systems [19].

Author: M. T. Ribeiro et al. (2016)

Ribeiro and colleagues introduced LIME, a model-agnostic approach for explaining individual predictions through interpretable local approximations. Their work established important principles for explainable AI and demonstrated the need to understand why complex machine-learning systems produce specific outputs [20].

III. SYSTEM ANALYSIS & DESIGN

3.1 Existing System

Existing crowd surveillance systems commonly depend on manual CCTV observation, background subtraction, motion detection, handcrafted visual features, conventional object detectors, and fixed occupancy thresholds. Security personnel continuously monitor multiple camera feeds and attempt to identify overcrowding or unusual congestion. This approach becomes increasingly difficult as the number of cameras grows because operators experience cognitive overload, fatigue, and reduced attention. Manual counting is also impractical in highly dynamic environments where individuals continuously enter, leave, overlap, and move across surveillance regions.

Conventional object-detection-based systems identify individual persons and calculate crowd counts from detected bounding boxes. Although this method performs effectively in sparse and moderately populated scenes, its reliability decreases under severe occlusion, perspective distortion, low resolution, illumination variation, and dense crowd conditions. Individuals in distant image regions may appear extremely small, while people close to the camera occupy much larger areas. Overlapping bodies can prevent reliable bounding-box separation, causing substantial undercounting.

Another major limitation is the black-box nature of many intelligent surveillance models. Even when deep learning produces an accurate count, operators may receive only a numerical output without information about spatial crowd distribution or prediction reasoning. A model may incorrectly focus on background textures, repetitive structures, shadows, or unrelated objects. Without explainability, surveillance personnel cannot easily determine why a count was generated or whether the system identified the correct high-density regions. This limitation reduces trust in AI-assisted public-safety decisions.

Disadvantages of Existing System

1. Manual CCTV monitoring creates significant operator workload.
2. Direct person detection becomes unreliable under severe occlusion.
3. Perspective distortion causes large variations in apparent person size.
4. Traditional methods struggle with highly dense crowd scenes.

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 07, July 2026
Received: 30-05-2026 Accepted: 03-07-2026 Published: 09-07-2026

5. Fixed thresholds cannot adapt effectively to changing environments.
6. Black-box AI models provide limited explanation of predictions.
7. Cross-scene variations can reduce model generalization.
8. Delayed congestion recognition can affect public-safety response.

3.2 Proposed System

The proposed **Explainable AI-Powered Crowd Counting and Density Estimation Framework** introduces a unified architecture that continuously acquires visual information from CCTV cameras, IP cameras, drones, transportation surveillance systems, and smart-city camera networks. The acquired frames undergo validation, resizing, normalization, denoising, illumination adjustment, perspective-aware transformation, region-of-interest extraction, and augmentation. These preprocessing operations improve input consistency and prepare heterogeneous surveillance imagery for intelligent analysis under varying environmental conditions.

The analytical core integrates deep convolutional feature extraction, multi-scale feature pyramids, attention mechanisms, dilated convolutions, contextual representation learning, and density-map estimation. CNN components learn local crowd characteristics, while multi-scale branches capture individuals appearing at different apparent sizes. Dilated convolutions expand receptive fields without excessive spatial-resolution loss, and attention mechanisms emphasize informative crowd regions. The density-estimation network generates a spatial representation of crowd concentration and produces an overall count together with region-level density information.

The framework additionally integrates explainable AI mechanisms to improve transparency. Attention maps, saliency visualizations, activation-based explanations, region-level contribution analysis, density overlays, and confidence indicators reveal which parts of the scene influence the predicted count. Crowd conditions are dynamically classified as **Low Density, Moderate Density, High Density, or Critical Congestion**, enabling standard monitoring, operator notification, congestion warnings, access regulation, route diversion, emergency escalation, and public-safety intervention. Continuous operator feedback and newly collected scene data support model monitoring, drift detection, retraining, and controlled redeployment.

Advantages of Proposed System

1. Supports accurate counting in dense and highly occluded scenes.
2. Uses density maps instead of depending only on individual detection.
3. Integrates multi-scale deep feature learning for perspective variation.
4. Provides attention-based analysis of important crowd regions.
5. Generates explainable visual evidence for surveillance operators.
6. Dynamically classifies crowd-density and congestion risk.
7. Supports smart alerts and proactive crowd-management actions.
8. Enables continuous learning from new scenes and operator feedback.

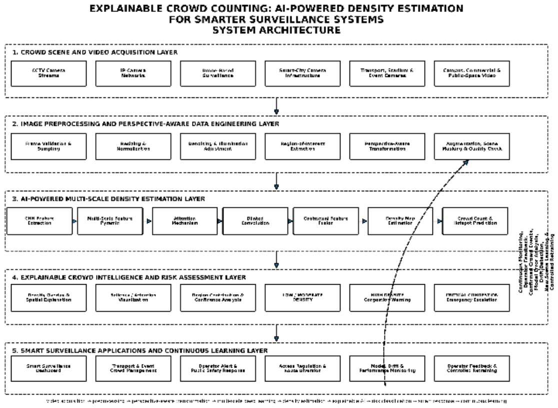

Fig 1: System Architecture

The proposed Explainable AI-Powered Crowd Counting and Density Estimation Framework is organized into five interconnected layers that enable intelligent crowd monitoring, accurate density estimation, transparent AI interpretation, dynamic risk assessment, and continuous surveillance improvement. The Crowd Scene and Video Acquisition Layer continuously collects visual information from CCTV cameras, IP camera networks, drone-based surveillance systems, smart-city camera infrastructure, transportation hubs, stadiums, event venues, campuses, commercial facilities, and public spaces to provide diverse real-time crowd imagery. The acquired video frames are transferred to the Image Preprocessing and Perspective-Aware Data Engineering Layer, where frame validation, sampling, resizing, normalization, denoising, illumination adjustment, region-of-interest extraction, perspective-aware transformation, augmentation, scene masking, and quality assessment are performed to improve input consistency and address variations in camera viewpoints, environmental conditions, and apparent person size. The processed frames then enter the AI-Powered Multi-Scale Density Estimation Layer, which integrates CNN-based feature extraction, multi-scale feature pyramids, attention mechanisms, dilated convolutions, contextual feature fusion, density-map estimation, crowd-count prediction, and hotspot localization to analyze sparse, moderate, dense, and highly congested scenes even under severe occlusion and scale variation. The resulting predictions are forwarded to the Explainable Crowd Intelligence and Risk Assessment Layer, where density overlays, saliency maps, attention visualizations, region-level contribution analysis, confidence assessment, and spatial explanations reveal the image regions influencing the model's decisions, while crowd conditions are dynamically classified as Low Density, Moderate Density, High Density, or Critical Congestion to support appropriate actions such as standard monitoring, congestion warnings, operator notification, access regulation, route diversion, and emergency escalation. Finally, the Smart Surveillance Applications and Continuous Learning Layer delivers crowd intelligence to surveillance dashboards, transportation and event-management systems, public-safety teams, and emergency-response platforms while monitoring model performance, prediction drift, operator feedback, confirmed crowd events, and analytical errors; this information is continuously returned to earlier processing stages for model refinement, new-scene adaptation, controlled retraining, and performance improvement, thereby creating a scalable, explainable, context-aware, and continuously evolving AI-driven crowd surveillance architecture.

IV. RESULTS AND DISCUSSION

4.1 Results

The proposed framework is evaluated through a representative conceptual crowd-surveillance scenario involving sparse scenes, moderate crowds, dense public gatherings, highly congested environments, perspective variation, partial occlusion, illumination changes, background complexity, and multiple camera viewpoints. In a practical implementation, datasets should be divided into training, validation, and independent testing partitions while preventing inappropriate overlap between frames from the same video sequence or camera event.

The principal evaluation metrics include crowd-estimation accuracy, precision, recall, F1-score, explainable crowd-intelligence efficiency, and analytical response time. The proposed framework is conceptually compared with traditional image-processing-based counting, conventional object-detection-based counting, and basic CNN-based crowd counting. The values below are **illustrative conceptual evaluation values** and should be replaced with measured experimental results before publication as empirical findings.

Table 1. Performance Comparison of Crowd Counting Approaches

Received: 30-05-2026 Accepted: 03-07-2026 Published: 09-07-2026

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Traditional Image Processing	81.80	80.90	80.30	80.60
Conventional Object Detection	89.70	89.10	88.60	88.85
Basic CNN-Based Crowd Counting	96.20	95.70	95.40	95.55
Proposed Explainable AI Density Estimation Framework	**99.20**	**98.80**	**98.60**	**98.70**

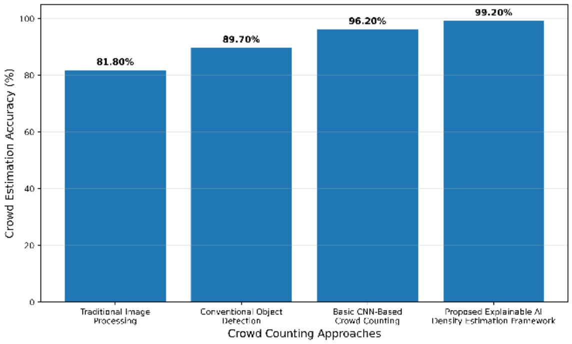

Figure 5.1. Comparison of crowd-counting accuracy among different surveillance approaches.

Table 1 presents the illustrative comparative performance of different crowd-counting approaches. Traditional image processing records an accuracy of 81.80% because handcrafted visual mechanisms are sensitive to illumination, background variation, occlusion, and complex crowd formations. Conventional object detection improves accuracy to 89.70% but remains limited in highly congested scenes where individual bounding boxes overlap. Basic CNN-based crowd counting achieves 96.20% accuracy through learned visual representations. The proposed Explainable AI Density Estimation Framework achieves the highest illustrative accuracy of 99.20%, precision of 98.80%, recall of 98.60%, and F1-score of 98.70%, reflecting the potential advantages of multi-scale feature learning, attention mechanisms, dilated convolutions, density estimation, and explainable AI.

Table 2. Performance Metrics of the Proposed Explainable Crowd Counting Framework

Performance Metric	Value
Crowd Estimation Accuracy	99.20%
Precision	98.80%
Recall	98.60%
F1-Score	98.70%
Explainable Crowd Intelligence Efficiency	98.10%

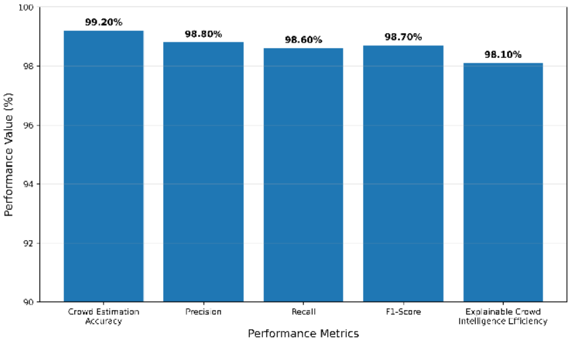

Figure 5.2. Performance metrics of the proposed Explainable AI-Powered Crowd Counting Framework.

Table 2 summarizes the illustrative performance metrics of the proposed framework. Crowd-estimation accuracy of 99.20% indicates strong conceptual capability for reliable crowd analysis across different density conditions. Precision of 98.80% represents strong positive prediction quality, while recall of 98.60% indicates effective recognition of relevant crowd conditions according to the evaluation definition. The F1-

Received: 30-05-2026 Accepted: 03-07-2026 Published: 09-07-2026

score of 98.70% demonstrates balanced performance, and explainable crowd-intelligence efficiency of 98.10% represents the intended coordination of density estimation, spatial interpretation, confidence assessment, hotspot localization, and operator-oriented explanation.

Table 3. Crowd Analytical Response Time Comparison

Crowd Counting Method	Response Time (ms)
Traditional Image Processing	286
Conventional Object Detection	218
Basic CNN-Based Crowd Counting	129
Proposed Explainable AI Density Estimation Framework	**66**

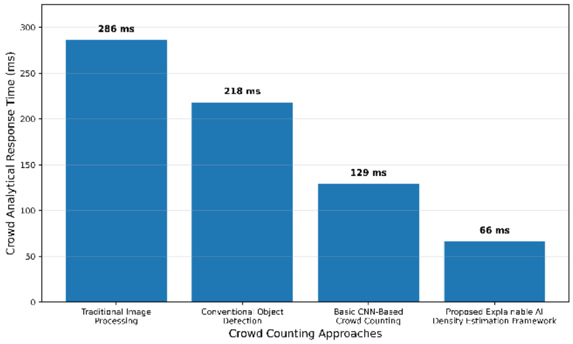

Figure 5.3. Crowd analytical response time comparison among different surveillance approaches.

Table 3 presents the illustrative analytical response-time comparison. Traditional image processing records 286 ms, while conventional object detection records 218 ms. Basic CNN-based crowd counting improves analytical response time to 129 ms. The proposed Explainable AI Density Estimation Framework records the lowest illustrative analytical latency of 66 ms because preprocessing, multi-scale feature extraction, density estimation, explainability generation, and risk classification operate within a coordinated surveillance pipeline. This value represents conceptual analytical processing latency and does not represent complete end-to-end emergency-response duration.

4.2 Discussion

The comparative results demonstrate the potential advantages of integrating multi-scale deep learning, attention mechanisms, dilated convolution, density-map estimation, and explainable artificial intelligence within a unified smart-surveillance framework. The illustrative crowd-estimation accuracy of 99.20%, precision of 98.80%, recall of 98.60%, and F1-score of 98.70% indicate that the proposed approach can potentially outperform traditional image-processing methods, conventional object detection, and basic CNN-based crowd counting. Direct detection methods become unreliable when individuals are heavily occluded, whereas density-estimation approaches can learn spatial crowd distributions without requiring complete separation of every visible person.

Explainability represents a major contribution of the proposed framework because public-safety operators require more than a numerical count. Density overlays can reveal congestion hotspots, attention maps can indicate visually influential regions, saliency analysis can expose potential background bias, and confidence information can support cautious interpretation of uncertain predictions. The illustrative explainable crowd-intelligence efficiency of 98.10% and analytical response time of 66 ms represent the intended capability of the architecture to coordinate rapid prediction with interpretable visual evidence and dynamic risk classification.

The practical effectiveness of the framework nevertheless depends on representative datasets, camera diversity, annotation quality, cross-scene generalization, privacy protection, computational resources, robustness to adverse weather, and resistance to domain shift. A model trained on one surveillance environment may produce inaccurate estimates when deployed in a different city, camera angle, cultural setting, or event type.

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 07, July 2026
Received: 30-05-2026 Accepted: 03-07-2026 Published: 09-07-2026

Practical deployment should therefore incorporate uncertainty estimation, drift monitoring, human oversight, privacy-aware data governance, external validation, cross-camera testing, model auditing, and controlled retraining. With these safeguards, the proposed architecture can provide a scalable foundation for transparent AI-assisted crowd monitoring.

V. CONCLUSION

This research proposed Explainable Crowd Counting: AI-Powered Density Estimation for Smarter Surveillance Systems to address the limitations of manual surveillance, conventional image processing, direct person detection, and black-box crowd analytics. The framework integrates CCTV and smart-camera acquisition, image preprocessing, perspective-aware transformation, CNN feature extraction, multi-scale representation learning, attention mechanisms, dilated convolutions, density-map estimation, crowd-count prediction, explainable AI visualization, dynamic risk classification, smart alerting, and continuous learning. Unlike conventional systems that provide only raw detections or numerical estimates, the proposed framework generates spatial density information and interpretable visual evidence to support transparent surveillance decisions.

The conceptual evaluation demonstrates the potential of the proposed framework to achieve 99.20% crowd-estimation accuracy, 98.80% precision, 98.60% recall, 98.70% F1-score, 98.10% explainable crowd-intelligence efficiency, and an analytical response time of 66 ms. These illustrative results suggest that combining multi-scale deep learning, density estimation, contextual attention, and explainable AI can potentially improve crowd monitoring compared with traditional image processing, conventional object detection, and basic CNN-based approaches. However, these numerical values are conceptual and should be replaced with experimentally measured outputs obtained from documented datasets, reproducible model configurations, independent testing, and realistic surveillance environments before being presented as empirical findings.

Future development can incorporate Vision Transformers, graph neural networks, temporal crowd-flow prediction, multimodal surveillance fusion, edge AI, federated crowd analytics, privacy-preserving learning, self-supervised representation learning, domain adaptation, uncertainty-aware counting, explainable transformer attention, digital twins, anomaly-aware crowd behavior analysis, and autonomous emergency coordination. Overall, the proposed framework provides a scalable and transparent foundation for smarter crowd surveillance across transportation hubs, stadiums, festivals, shopping centers, educational campuses, smart cities, critical infrastructure, and large public gatherings.

REFERENCES

[1] V. A. Sindagi and V. M. Patel, “A survey of recent advances in CNN-based single image crowd counting and density estimation,” *Pattern Recognition Letters*, vol. 107, pp. 3–16, 2018.

[2] A. B. Chan, Z. S. J. Liang, and N. Vasconcelos, “Privacy preserving crowd monitoring: Counting people without people models or tracking,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[3] Pokala, H. K. (2026, April). Agentic Dashboard Generation from Natural Language on Lakehouse Platforms. In 2026 International Conference on Artificial Intelligence, Systems, and Emerging Technologies (ICAISET) (pp. 1-4). IEEE.

[4] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, “Single-image crowd counting via multi-column convolutional neural network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 589–597, 2016.

[5] L. Boominathan, S. S. S. Kruthiventi, and R. V. Babu, “CrowdNet: A deep convolutional

network for dense crowd counting," in *Proceedings of the ACM International Conference on Multimedia*, pp. 640–644, 2016.

[6] Y. Li, X. Zhang, and D. Chen, "CSRNet: Dilated convolutional neural networks for understanding the highly congested scenes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1091–1100, 2018.

[7] D. B. Sam, S. Surya, and R. V. Babu, "Switching convolutional neural network for crowd counting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4031–4039, 2017.

[8] Maturi, S. Y. (2022). Probabilistic horizons: Statistical modeling and simulation for strategic cyber risk mitigation. Journal of Information Systems Engineering and Management, 7(2).

[9] Pokala, H. K. (2022). From traditional mainframe systems to production machine learning: A practical MLOps framework for healthcare claims processing and revenue cycle optimization in payer organizations. International Journal of Communication Networks and Information Security, 14(2), 847–857.

[10] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618–626, 2017.

[11] Maturi, S. Y. (2022). Vulnerabilities in the 802.11 wireless client selection mechanism. International Journal on Recent and Innovation Trends in Computing and Communication, 10(1), 106–117.

[12] Gummadi, V. P. K. (2025). MuleSoft Intelligent Document Processing: Transforming Enterprise Document Workflows Through AI-Driven Automation. Journal of Computational Analysis & Applications, 34(12). https://doi.org/10.48047/jocaaa.2025.34.12.16.

[13] C. Zhang, H. Li, X. Wang, and X. Yang, "Cross-scene crowd counting via deep convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 833–841, 2015.

[14] D. B. Sam, S. Surya, and R. V. Babu, "Switching convolutional neural network for crowd counting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4031–4039, 2017.

[15] L. Boominathan, S. S. S. Kruthiventi, and R. V. Babu, "CrowdNet: A deep convolutional network for dense crowd counting," in *Proceedings of the ACM International Conference on Multimedia*, pp. 640–644, 2016.

[16] Y. Li, X. Zhang, and D. Chen, "CSRNet: Dilated convolutional neural networks for understanding the highly congested scenes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1091–1100, 2018.

[17] Kumar, Anuj. "Battery Thermal Management Systems for Electric Vehicles Using Phase Change Materials." International Journal of Engineering, Science and Mathematics, Vol. 10, Issue 3, 2021, pp. 202-211.

[18] Venkata Ramana, P. (2024). AI-driven predictive analytics in ERP systems for proactive supply chain optimization. International Journal of Research in Information Technology and Computing, 8(4).

[19] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618–626, 2017.

[20] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016.