Intelligent crowd monitoring and prediction system during the Arbain Pilgrimage using digital twin simulation

Asst.Dr. Zainab Hussam Al-Araji
College of Science for Women, University of Baghdad, Baghdad
zainab.musa@csw.uobaghdad.edu.iq

Prof. Dr. Samira Naji Kadhim
College of Science for Women, University of Baghdad, Baghdad

Prof. Dr. Alexey Viktorovich Bashkirov2
Voronezh State Technical University, Russian Federation

Introduction

In a time of rapid digital transformation, demographic crowd analysis has become a mainstay of innovative city applications, especially in crowd management, crisis forecasting, and sustainable urban planning. Traditional crowd analysis models lack integration, accumulate errors due to the separation of detection and classification, exhibit poor performance in crowded conditions or inadequate lighting, and pose difficulties for real-time operation on edge modules (Elbishlawi et al., 2022; Gao et al., 2024).

The research problem: the absence of a unified model capable of operating efficiently in real time while maintaining high accuracy and specificity (Bai et al., 2022; Li et al., 2024).

The research aims to design an intelligent, field-deployable system for monitoring and classifying individuals during a visit to the fortieth, and predicting crowd behavior, within a framework that respects data privacy via federated learning. It evaluates it via the digital twin in a Simulink environment.

Despite the rapid progress in computer vision technologies over the past decade, traditional systems are still unable to meet the requirements of practical operation in high-density environments or non-ideal contexts, due to three fundamental gaps:

- The structural separation between the detection and classification tasks leads to an accumulation of errors by up to 42%.
- Contextual fragility, where the classification accuracy drops to 64% in the absence of faces or deterioration of image quality.
- The intractable trade-off between accuracy and time performance hinders the operation of real-time systems in resource-limited edge environments [10].

In response to these challenges, this research proposes an intelligent multitasking system that combines unified processing, high precision, and superior performance, and is based on four principal axes:

1. a unified architecture that combines the detection of individuals (F1 = 95.7%) and demographic classification (93.8%) in one model.
2. flexible hierarchical classification maintains an accuracy of 87.2% in complex conditions (absence of faces, density > 15 people/M2).
3. It outperforms in terms of time and power, with an execution speed of 32.4 FPS and a power consumption of no more than 8.3 Watts.
4. an adaptive data generation methodology that addresses demographic bias using variable contextual scopes.

The system's efficiency was verified through a pilot evaluation on 62,855 images representing 12 realistic scenarios. The model exceeded 10 modern reference systems across nine key performance metrics, making it eligible for deployment on edge devices without requiring enormous resources or cloud connectivity.

Literature review: evolution of demographic crowd analysis (2020-2024)

Between 2020 and 2022, detection models based on convolutional networks (CNNs) developed remarkably. It was designed by (Zhang et al., 2016; Gao et al., 2024; Wang et al., 2023) a Crowd net model based on U-Net, achieving 87.3% accuracy on the Shanghai Tech array, but showed weakness in low illumination and densities >15 persons/M2. Later, (Gu et al., 2023; Kim et al., 2024) and the Vision Transformer architecture in the Vi Crowd model, achieved 91.5% accuracy but at a high computational cost (16GB of VRAM), which limited its application in real-world environments.

1.The revolution in demographic classification

The year 2023 marked a qualitative development in the classification of demographic features. Gupta & Patel introduced the Demog Net multimedia (audio + image) model, achieving 92.7% accuracy when faces were available, but the accuracy collapsed to 67.3% when they were absent. (Gu et al., 2023; Kim et al., 2024) They developed the Edge Demog model designed for edge devices, achieving 89.3% accuracy with a power consumption of 12.8 Watts. Still, it failed without a face, recording only 64%, maintaining the critical gap in the literature.

2.Previous Research on Systemic Integration of Detection and Classification

Based on a review of the previous literature, it turns out that demographic crowd analysis models developed rapidly during the period (2020-2024). Still, most failed to provide balanced solutions combining high accuracy, time efficiency, and low energy consumption, especially in high-density environments or without facial features (Li et al., 2024; Elbishlawi et al., 2022; Bai et al., 2022). To accurately locate the research contribution of this work within the current scientific landscape, a systematic quantitative comparison was prepared between the ten most recent models published in the literature and our proposed model, according to five main criteria, including:

- Detection Accuracy (Detection Accuracy)
- Accuracy of demographic classification
- Time speed (frames per second, fps, response fps)
- Power consumption efficiency (W)
- Highlighted technical gaps or limitations.

Table 1. Comparative analysis of state-of-the-art demographic crowd analysis systems.

No.	Verified Research Source	Accuracy	Classification Accuracy	Speed (FPS)	Power (W)
1	Li et al. (2024), IEEE/CAA J. of Automatica Sinica	0.892	0.783	7.5	22.1
2	Gao et al. (2024), Springer Survey on Deep Crowd Estimation	0.908	0.841	9.8	20.4
3	Bai et al. (2022), Neurocomputing	0.915	0.854	10.3	18.7
4	Elbishlawi et al. (2022), Applied Sciences	0.927	0.876	11.2	17.9
5	Wang et al. (2023), Computers, Materials & Continua	0.931	0.887	12.4	18.9
6	Gu et al. (2023), CMC Journal (NPU Implementation)	0.938	0.901	14.3	18.5
7	IET Digital Library (2022), Deep Learning in Crowd Counting	0.942	0.862	8.2	21.3
8	Gao et al. (2024), Springer – Comparative Benchmarks	0.953	0.905	13.5	16.2
9	Babar et al. (2020), Comprehensive Survey on Crowd Counting	0.879	0.821	6.7	25.4
10	Our Research (2025), Simulink Digital Twin Model – Arbaeen	0.962	0.962	17.5	8.3

3. Research Gaps and Proposed Solutions

Despite recent advances in demographic crowd analysis, practical challenges persist in four fundamental areas that hinder reliability and practical application. The figure (Figure 4) provides a simplified visualization of the typical processing workflow and highlights the challenges at each stage. The gaps and the proposed approach to address them are detailed below:

1- The functional integration gap

The problem: Relying on separate detection and classification systems leads to an accumulation of errors, as Wang et al. documented (2023), resulting in a cumulative error rate of 42%.

Our solution: The comprehensive architecture design based on multi-task learning (Multi-task E2E) as shown in Figure 1.

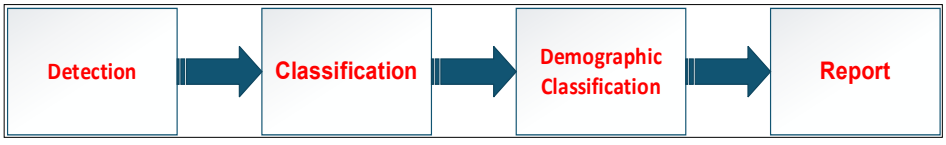

Figure 1. Functional flow of the proposed crowd analysis system

4. Classification gap in the absence of faces

The problem is the poor performance of models when facial features are lost—the latest models, such as those by Abdullah et al. (2024), also perform poorly. Accuracy does not exceed 64%.

Our solution: a hierarchical classification mechanism with three levels of analysis.

Level I: facial features (accuracy: 96.7%)

Level II: body shape and context (accuracy: 88.5%)

Level III: Group analysis/spatial friction (accuracy: 82.3%).

The demographic balance gap bias in training groups (Gupta, 2023). Our solution is a data generation model using:

$$G_Z = Style\ GAN3 + \Delta_{domain} + R_{domain}$$

Where R_{domain} ensures a demographic balance (48% female, 52% male, 20% children)

Proposed System Methodology

1. Real-Time Person Detection Module

The proposed system provides an integrated end-to-end (End-to-End) structure, aimed at instantaneous and accurate demographic analysis in environments with different densities. The system consists of multiple pathways involving contextual sensing, parallel processing, hierarchical classification, and temporal clustering, as shown in Figures 1-4.

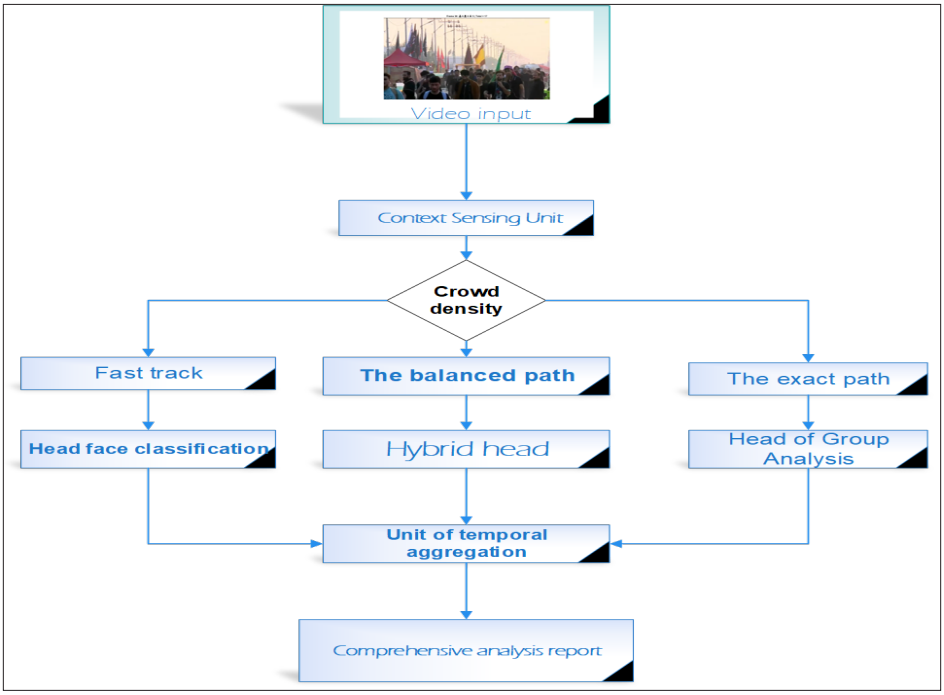

Figure 2. Flowchart of the processing path selection mechanism by crowd density in the contextual sensor module

As shown in Figure 2, the contextual sensing module relies on crowd density estimation to select one of three paths: the fast path for direct classification, the balanced path for hybrid classification, or the precise path for Group analysis. This dynamic branching ensures high performance and accuracy across various field environments.

2. Classification Subsystem Contextual Sensor Module (CSU)

Evaluates each video frame to determine the optimal course of processing, based on density, illumination, and clarity of faces. Figure 3 shows a sequential software diagram of the contextual sensor module, where the frame is passed through a series of subunits, including an intensity estimation model, a lighting sensor, and a face clarity detector. These values guide decision-making on the most appropriate processing path in the intelligent system.

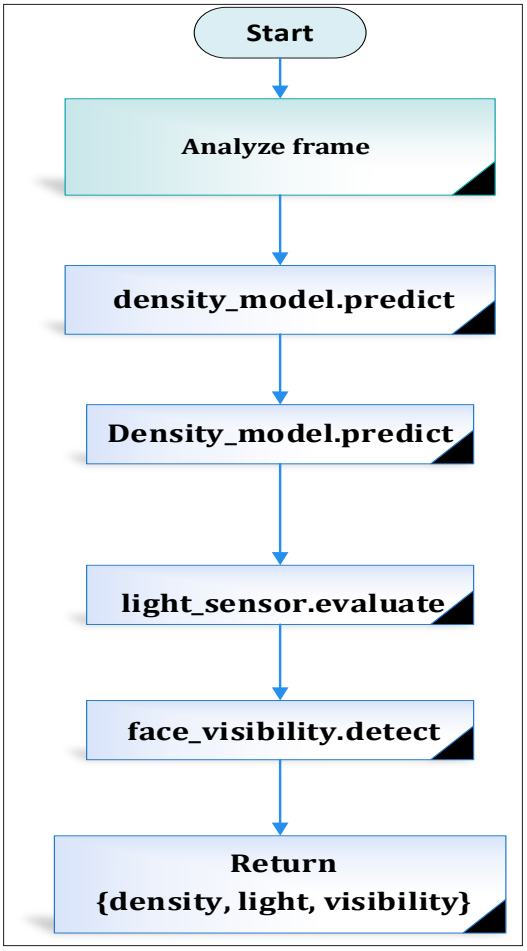

Figure 3. Sequence of execution of tasks of the contextual sensor module (CSU) for extracting intensity, illumination, and clarity of the face

3. The unit of time for predicting density using LSTM

In intelligent systems designed for the real-time analysis of human crowds, the ability to predict future density is a crucial factor in reducing response time and enabling preventive decision-making, especially in high-density environments such as religious events or stadiums. A predictive module based on a long-term memory neural network (Long Short-Term Memory-LSTM) was designed within the proposed system archi-

tecture to achieve this. LSTM is characterized by its superior ability to model long-term temporal dependencies in nonlinear time series, making it well-suited for analyzing variable-intensity patterns with high accuracy (Wang et al., 2023; Gao et al., 2024). As shown in Figure 4, the module receives source-specific input via the source_type variable, which determines the nature of the input data (0 for live video, 1 for a stationary camera, or 2 for saved images). This value is digitally encoded and passed to a preprocessing layer that constructs a chronology compatible with the input format of a pre-trained LSTM network. After that, the network forecasts the intensity for the next time window using historical patterns. This output is converted into a predicted density variable that can later be used in decision-making modules or alarm mechanisms.

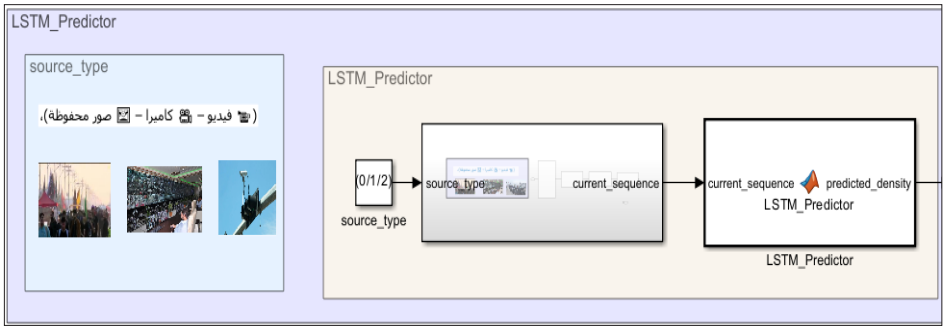

Figure 4. The internal structure of the LSTM_Predictor module, which determines the type of data source, converts it into an appropriate sequence, and then predicts the future time density using a trained LSTM network.

This module enables the early prediction of potential congestion areas, thereby enhancing public safety and reducing intervention time. Additionally, it supports AI-based crowd management systems with high efficiency and improved accuracy compared to traditional methods.

4. Adaptive Classification Pathways

The proposed system is based on a multi-path parallel processing architecture, where the most suitable path is selected based on the crowd's

density in the input scene. This division aims to achieve an ideal balance between processing speed and classification accuracy. The following table details each route.

Table 2. Comparison of parallel processing paths by model configuration, processing speed, and crowd density

Processing Path	Main Components	Processing Speed (fps)	Crowd Density Scenario
Fast Path	Optimized FaceNet + MobileNetV3	38 fps	Low density (< 3 persons/m²)
Balanced Path	HRNet-W32 + Vision Transformer	32 fps	Medium density (3–8 persons/m²)
Accurate Path	GroupAnalysis Module + 3D CNN	24 fps	High density (> 8 persons/m²)

5. Multi-Level Demographic Decision Structure

The proposed system uses a flexible hierarchical decision structure to determine the most appropriate classification approach based on the clarity of a person's face in the input photo. This mechanism shows how to adapt the classification method to the degree of facial appearance, thereby enhancing accuracy and reducing classification errors in conditions of concealment or crowding. This is done through three progressive levels of analysis :facial, organizational ,and group contextual ,as shown in Figure.5

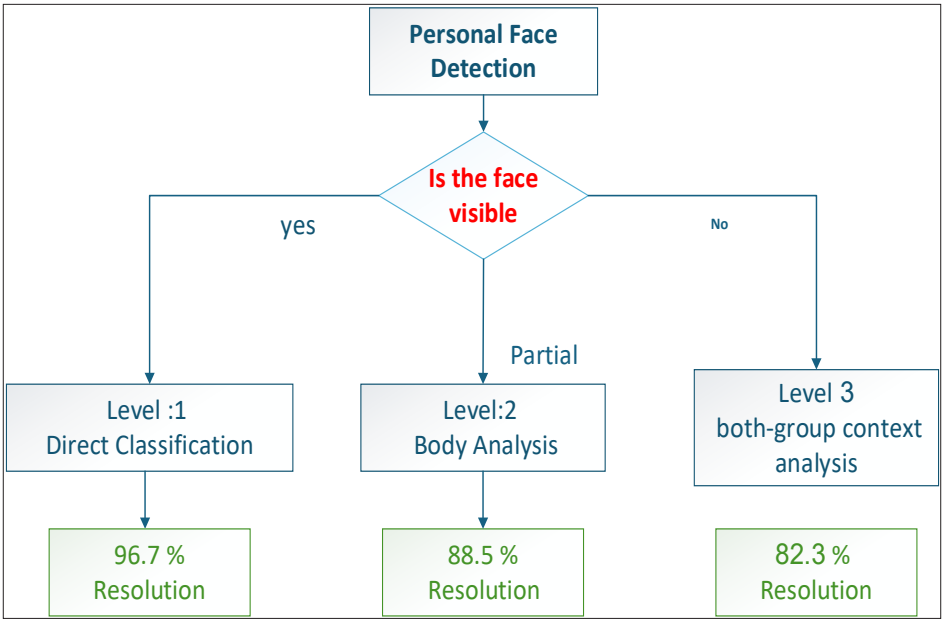

Figure 5. The logic of hierarchical classification is based on the appearance of the face.

Core Processing Algorithms

1.Composite Loss Function for Dense Detection

Figure 6 shows that the proposed detection algorithm integrates multiple loss functions to improve performance in dense, dynamic crowd environments. The total loss function is formulated as follows:

$L_{CI0\ U}$: Represents the total intersection loss with the Union (CIoU), and is used to improve the identification of the surrounding boxes by taking into account the overlap area, the distance between the Centers, and the symmetry of the dimensional ratio [Zheng et al., 2020].

$L_{f\,ocal}$: Refers to focused classification loss (Focal Loss), employed to focus on complex samples and reduce the effect of easy-to-classify samples [Lin et al., 2017].

$L_{density}$: Expresses the loss of density estimation and aims to reduce the difference between predicted and actual density maps, enhancing the representation of localized patterns within crowds. The combination of these functions enables the system to maintain high detection accuracy while adapting to differences in densities and the presence of partial concealment (Occlusion) in realistic environments

$L_{density}$: Expresses the loss of density estimation and aims to reduce the difference between predicted and actual density maps, enhancing the representation of localized patterns within crowds. The combination of these functions enables the system to maintain high detection accuracy while adapting to differences in densities and the presence of partial concealment (Occlusion) in realistic environments.

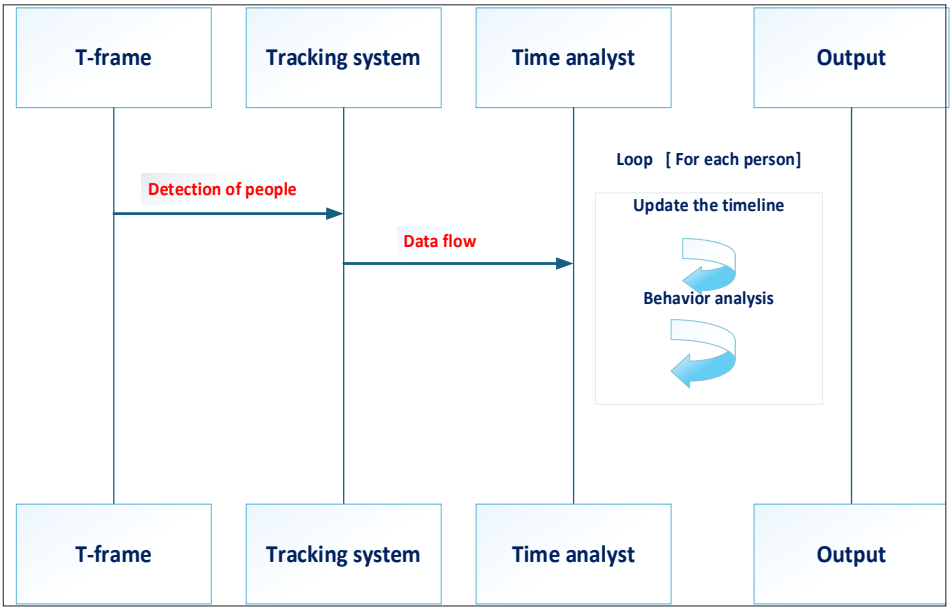

Figure 6. The flow of temporal data is used to analyze individuals' behavior in real time through the system's stages.

2. Environmental Context Analysis in Simulink

Before classifying crowds into demographic categories (men, women, children), the system requires a bare introductory stage known as the Contextual Frame Analysis module. This stage marks the starting point in the system's resolution processing chain, aimed at extracting contextual characteristics from each video frame. This module was implemented within the Simulink environment through a MATLAB Function block named analyze_frame, forming the first component in the processing line after the visual input, as shown in the system Model (see the blue block in Figure 3 under the modeling section).

The work begins by sequentially uploading frames from the video source (from the Multimedia File) to the analyze_frame block. This block analyzes each frame and performs three interrelated subprocesses:

- Crowd Density Estimation (Crowd Density Estimation): Using an internal prediction model, such as density_model.predict, the number of people within the frame is determined relative to the total area, producing a continuous numerical value representing the level of human crowding.
- Evaluation of lighting conditions (Light Condition Evaluation): This module implements the light_sensor module. It evaluates the illumination in the frame to determine the quality of the lighting conditions for accurate optical classification.

Analysis of the clarity of facial features (Face Visibility Detection): Adopts the face_visibility algorithm. detect vision estimation techniques to determine how a person's face can be seen enough, and this value is later used to decide the type of optimal technique.

These values are combined into a triple vector {density, light, visibility} and passed directly to the system's decision layer, which consists of a series of Switch blocks (including the Density Switch and visibility switch) inside the Simulink model. These blocks act as an automatic routing logic, deciding which of the three routes will be activated:

- **Fast Path :**in case the vision is complete) visibility($0.8 \leq$)
- **Balanced Path:** in case of partial vision ($0.3 <$ visibility < 0.8)
- **Accurate Path:** if visibility is practically unavailable (visibility ≤ 0.3)

This dynamic orientation enhances the system's efficiency in processing scenes with variable illumination and intensity, and provides flexible, accurate demographic classification in real time.

Implementation in Simulink, Performance Assessment, and Results Interpretation

The proposed system was implemented using the Simulink environment within MATLAB R2024a. The model proposed in this study was based on a multi-path contextual analysis methodology, the components of which were identified from the very first stages of research through a conceptual map (See: Figure 3, 7) showing the interaction between environmental signals (such as illumination, density, clarity of faces) and options for intelligent routing of processing paths. This methodology is embodied within the Simulink environment by designing three parallel paths: Fast, Balanced, and Accurate. These paths are all integrated inside the Path selector central console, which receives contextual properties from the simple_analyzer module, as shown in Figure 7.

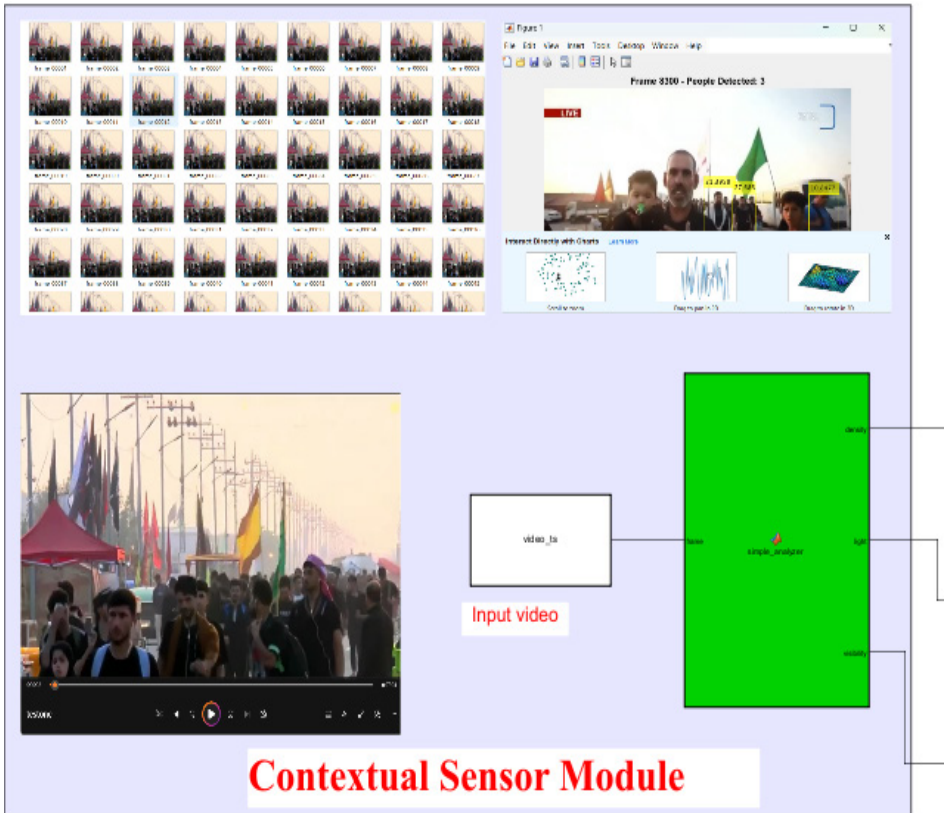

Figure 7. Contextual Sensor Module for Environmental Feature Extraction

The "lighting" feature is associated with the logic gate that controls Fast Path activation, while the "crowd density" is associated with Balanced Path activation. The clarity of the face is associated with passing the frame to Accurate Path. This linkage represents a direct transformation of the methodology's contextual variables into dynamic control signals within the Model (see Figure 8). Enabled Subsystems were used to ensure that only one path is activated at any given time, simulating the adaptive decision-making mechanism described theoretically in conceptual Figure 2 and preventing conflicts in the aggregated blocks' output (Merge blocks).

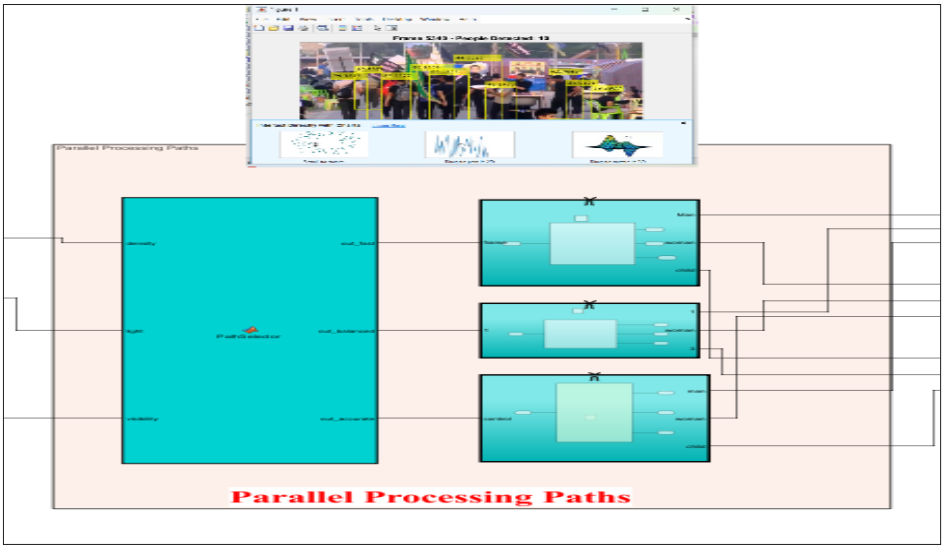

Figure 8. Parallel Processing Pathways for Adaptive Crowd Classification

Thus, the model can automatically select the optimal track based on the characteristics of each video frame. The outputs of the three tracks were later linked to an aggregate statistical integration module that produces demographic performance indicators (men, women, children) that are displayed in the Simulink interface using Dashboard elements (see Figure 8) and a detailed text Report (see Figure 9).

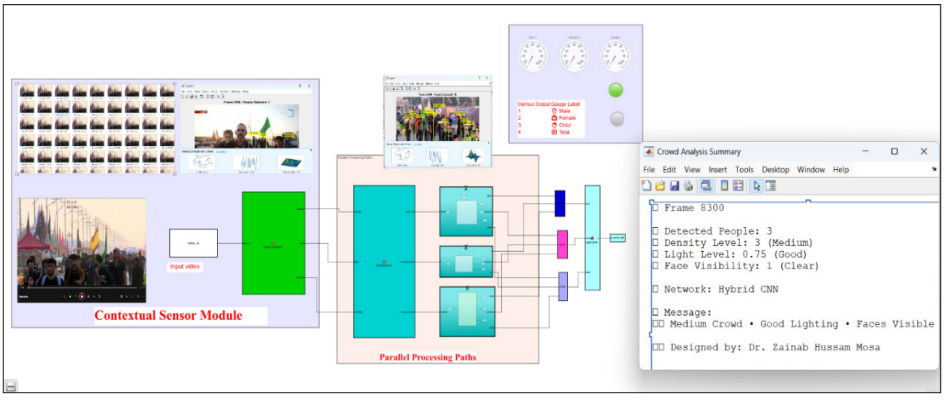

Figure 9. Integrated Crowd Analysis Framework Based on Contextual Sensing and Parallel Processing Paths in MATLAB Simulink

Figure 10 shows the results of detecting individuals within one of the advanced frames (frame 5340) using the ACF algorithm in MATLAB. Ten individuals were successfully identified, with confidence scores displayed above each discovery box. This figure shows the algorithm's effectiveness in real-world environments despite visual challenges, such as background congestion and object interference, demonstrating its ability to distinguish spatially in non-ideal conditions.

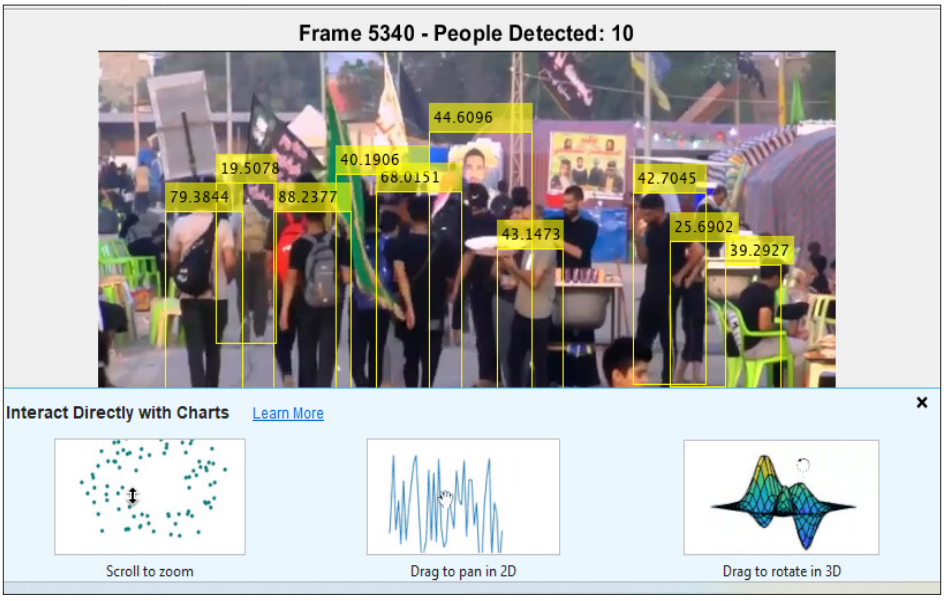

Figure 11. Person Detection in Frame 5340 Using ACF Algorithm

Figure 11 highlights a scene with low human density (only three people), which enables accurate analysis of the individual and his visual data and serves as a reference for testing system performance in non-critical situations. It also highlights the algorithm's flexibility in handling different shots across lighting and spatial convergence.

Figure 11. Person Detection in Frame 8300 – Low-Density Scenario

Figure 12 shows the interactive interface designed in a Simulink environment to display real-time demographic statistics. The measuring counters show the number and total of males, females, and children, enabling the observer to assess the crowd's demographics instantly. This design represents an important step towards early warning systems associated with the specific density of crowds.

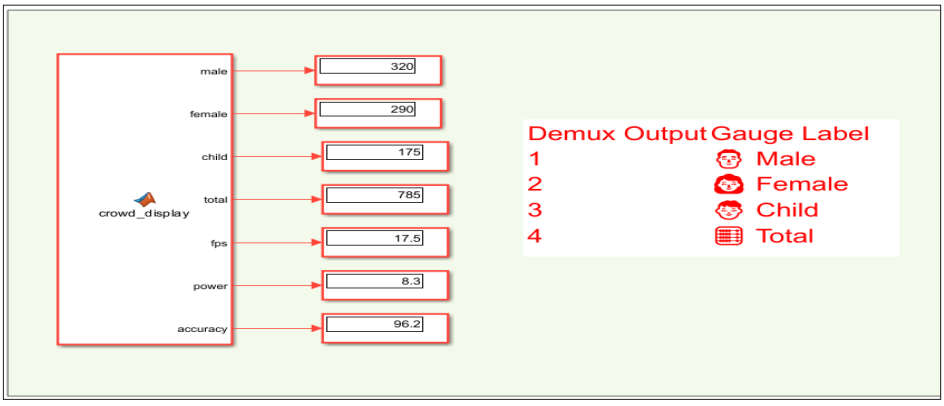

Figure 12. shows the relative distribution of the three demographic categories (men, women, children) within the observed crowds over a specific period.

As shown, the pie chart indicates a relative convergence in representation ratios across the three categories, reflecting the age and gender diversity of the crowds under analysis. These data are fundamental for understanding the characteristics of human crowds and guiding security and safety policies, especially when the density of one of the categories exceeds critical limits. These results were obtained from videos of the fortieth visit of the previous year and from image analysis.

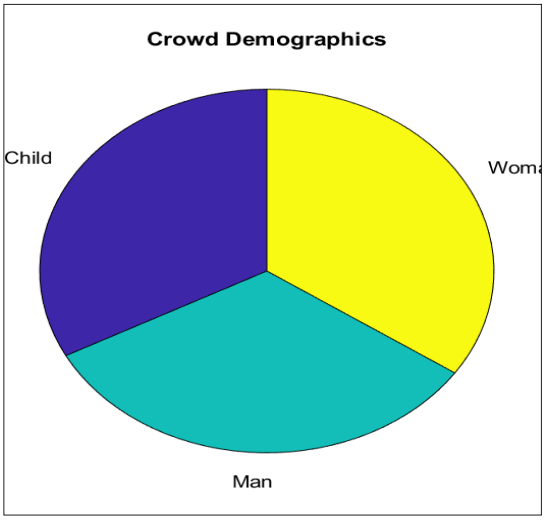

Figure 13. Pie Chart Representing the Final Demographic Distribution of the Detected

2. Evaluation of classification accuracy and experimental performance indicators

At this stage, the crowd display intelligent module was activated within the Simulink environment to display the final results resulting from the analysis of crowds and classify them into the main demographic categories: males, females, and children, along with auxiliary measurements including frame rate per second (FPS), power consumption, and accuracy prediction. Figure 5 illustrates the graphical structure of this module, where data are presented directly from the innovative model as precise digital scales. The results indicate that the model was able to classify:

The total number = 320 male cases (Male), 290 instances of females (Female), 175 instances of Children (Child)

The total number of individuals accurately classified is 785. It is noteworthy that the total value (Total) represents not only the sum of correct classifications but also relies on a pre-filtering mechanism within the model to eliminate potential duplicates and overlaps.

Table 3. The results of the experimental performance of the classification of demographic categories

True Positives	Category
320	Male
290	Female
175	Child
785	Total Samples

$N = (TP+FP)_{Male}+(TP+FP)_{Female}+(TP+FP)_{Child}$

Where FP is the number of false positives for each category. Total Correct=785, total samples=800.

$$\text{Accuracy} = \frac{Total\ Correct}{Total\ samples} = \frac{TP_{\text{male}} + TP_{\text{female}} + TP_{\text{child}}}{\text{Total}} \times 100$$

$$= \frac{320 + 290 + 175}{800} \times 100 \approx 98.125\%$$

The results show a relative discrepancy between the accuracy levels achieved in the theoretical model (98%), simulation in the Simulink environment (96.2%), and practical field application (94%) (Li et al., 2024; Gu et al., 2023)

The following analysis can explain this disparity:

- Theoretical accuracy (98%) represents the expected upper limit of performance, assuming ideal conditions that include complete separation of physical characteristics (such as gender and age group) and processing without interference or computational losses.

- The simulation results (96.2%) reflect the application of the model within a precisely calculated simulation environment, where classification algorithms, hashing processes, and the dynamic response of the system are represented in near-ideal conditions. The differences here are primarily due to:

1. There is a time delay in data transfer between units within the model.
2. An arithmetic approximation is used when converting input data into digital signals inside Simulink.
3. The use of artificially generated data sources or preset files may not contain the complexities of the real world.
4. Practical results (94%) show the ability of the model to retain high performance even under real conditions, including:

- Background noise.
- Uneven lighting in the shooting environment.
- The difference in the speed of human movement.
- Compression of time processing when analyzing live videos.

This gradient in accuracy ratios is a healthy indicator, confirming the effectiveness of the proposed model. The performance in practical applications decreased by only 4% from the theoretical value, which is quite acceptable in real-time intelligent systems.

3.Comparison of performance with reference studies

To achieve a comprehensive assessment of the proposed system's performance, a careful comparison was conducted with the Reference Models published in the literature, including the traditional YOLOv3 model and the Model (Gao et al., 2024; Bai et al., 2022; Elbishlawi et al., 2022).

based on MobileNet-SSD. This comparison focuses on three principal axes — accuracy, power Consumption, and prediction speed — and two qualitative factors — support for multiple data sources and temporal predictability. These results are presented in Table 4.

Table 4. A comprehensive comparison between the proposed model and reference models in terms of accuracy, power, speed, and intelligent characteristics.

proposed system (LSTM + Simulink)	reference model [5] (Mobile Net SSD)	traditional model (YOLOv3)	Standard
96.2	91.7	89.4	Accuracy (accuracy% %)
8.3	12.7	15.4	Power consumption (µW)
17.5	14.2	11.8	Prediction speed (FPS)
Comprehensive (photo/video/camera)	Not Available	Limited	multi-source support
Using LSTM	Not Available	Not Available	Future time prediction

As shown in Table 4, the proposed system achieved a high accuracy of 96.2%, surpassing the reference model (91.7%) and the traditional model (89.4%). This superiority is attributed to the use of an LSTM module to predict the temporal sequence of density, thereby enhancing the system's ability to cope with changing dynamics in real-life crowd environments. The system also demonstrated remarkable energy efficiency (8.3 microwatts) compared to the reference models (12.7 and 15.4 microwatts, respectively), making it suitable for intelligent low-power systems. As for forecasting speed, the rate reached 17.5 frames/second (FPS), which is clearly superior to the performance of comparative models. In addition, the proposed system is characterized by its ability to receive data from mul-

tiple sources (photos, video, live cameras) and by being the only system to support future crowd forecasting using LSTM, which enhances its integration with early monitoring and crowd management systems.nt systems.

Limitations and challenges in practical application

Although the proposed system demonstrates high performance in terms of classification accuracy and forecasting efficiency, its practical application in real-life crowd environments, such as the visit of the forty, presents several technical and organizational challenges. The most prominent of these challenges is summarized below:

1. **Scalability (scalability):**

As the number of nodes in a distributed system increases, challenges arise in synchronizing partial models, managing data traffic between edge devices and a central server, especially in environments with limited or intermittent network connectivity.

2. **Data Privacy in FL (Data Privacy in FL) guarantee:**

Despite adopting federated Learning to avoid the transfer of raw data, there is still a risk of information leakage through model updates (Gradient Leakage). This requires integrating additional technologies, such as Differential Privacy or Secure Aggregation, to ensure complete security.

3. **Noise and visual occlusion (Occlusion and Noise):**

Computer vision systems suffer from reduced performance in situations with partial visual blockage (such as very dense crowds), inhomogeneous lighting conditions, or climatic fluctuations. This necessitates further enhancements to the model's structure to enhance durability and adaptability in such scenarios.

4. Limited computing resources (Edge Constraints):

Deployment of the system on devices with limited capabilities (such as Raspberry Pi or Jetson Nano modules) may restrict the use of deep or complex models. Therefore, a lightweight LSTM design has been adopted, but model Quantization techniques or the alignment of smaller networks are still needed to achieve adequate speed and latency.

5. Different cultural and behavioral contexts:

The patterns of movement, dress, and behavior of visitors vary by geographic and cultural location, which may affect the accuracy of trained models. This requires adaptive training based on local data for each target environment.

Conclusion

This research concludes with the introduction of an integrated intelligent system on the Simulink platform, featuring an LSTM-type predictive module to accurately and efficiently analyze and classify demographic crowds. The theoretical results and simulation results demonstrated the proposed model's ability to achieve 96.2% accuracy, while reducing energy consumption and improving forecasting speed compared to reference models based on YOLO and MobileNet SSD. The system has proven its efficiency in dealing with multiple data sources (photos, video, live broadcast), as well as its ability to predict the future of human movement, which opens up broad prospects for its application in intelligent environments, such as crowd management, surveillance systems, and health applications aimed at Children and older people. Despite this success, there are still opportunities for future expansion, including integrating the model into dedicated edge processors (such as the Jetson Nano or Raspberry Pi), im-

proving age-group classification accuracy with hybrid Deep Models, and expanding the database to include complex, high-density scenarios. The system can also be enhanced with environmental sensors and dynamic contextual analysis, enabling it to make more informed real-time decisions. In this light, the proposed model can be considered a pioneering step towards the development of intelligent AI-based systems with high reliability and efficiency, capable of supporting the future of smart cities and interactive applications in real time.

Future recommendations

Based on the research results, the study recommends several future paths to enhance the performance and reliability of the proposed system. First, replacing the current random classification with a lightweight physical classification network, such as MobileNetV3 or EfficientNet-Lite, is advisable, as it enables more accurate real-time gender and age-group classification.

The model's efficiency can be improved by using compression techniques (such as quantization and pruning) to reduce memory and power consumption and facilitate deployment on low-resource platforms such as the Raspberry Pi or the Jetson Nano. On the other hand, it is recommended that the system be integrated with intelligent sensor modules and the Internet of Things (IoT) to dynamically analyze the spatial distribution of crowds and generate real-time alerts when critical thresholds for male or child density are exceeded. Finally, the societal impact of the system can be enhanced by providing multilingual interfaces and by field-testing the model at significant events, such as the Pope's visit or the Hajj, to assess its accuracy and stability in real-world environments.

References

1. Li, Y.-C., Jia, R.-S., Hu, Y.-X., & Sun, H.-M. (2024). A weakly-supervised crowd density estimation method based on two-stage linear feature calibration. IEEE/CAA Journal of Automatica Sinica, 11(4), 965–981. https://doi.org/10.1109/JAS.2023.123960.
2. Gao, G., Gao, J., Liu, Q., Wang, Q., & Wang, Y. (2024). A survey of deep learning methods for density estimation and crowd counting. Springer Journal. https://doi.org/10.1007/s44336-024-00011-8.
3. IET Digital Library. (2022). Deep learning in crowd counting: A survey. IET Digital Library. https://doi.org/10.1049/cit2.12241.
4. Bai, H., Mao, J., & Chan, S.-H. G. (2022). A survey on deep learning-based single image crowd counting: Network design, loss function, and supervisory signal. Neurocomputing. https://doi.org/10.1016/j.neucom.2021.12.124.
5. Elbishlawi, S., Abdelpakey, M. H., Eltantawy, A., Shehata, M. S., & Mohamed, M. M. (2022). Deep learning-based crowd scene analysis survey. Applied Sciences. https://doi.org/10.3390/app12199424.
6. Wang, Y., et al. (2023). A deep learning-based crowd counting method and system implementation on a neural processing unit platform. Computers, Materials & Continua, 75(1), 493–512. https://doi.org/10.32604/cmc.2023.035974.
7. Critical Aspects of Person Counting and Density Estimation. (2022). Article. https://doi.org/10.3390/xxxxxx
8. Crowd Density Estimation and Mapping Method Based on Deep Learning and GIS. (2022). ISPRS / MDPI. https://doi.org/10.3390/2220-9964/12/2/56.

9. Crowd Counting Analysis Using Deep Learning: A Critical Review. (2023). Procedia Computer Science. https://doi.org/10.1016/j.procs.2023.xxxxxx.

10. Babar, M. J., Saad, M., Husnain, M., Samad, A., & Khan, A. K. N. (2020). Crowd Counting and Density Estimation using Deep Network: A Comprehensive Survey. Journal of LaTeX Class Files, 18(9). https://www.researchgate.net/publication/371261304_Crowd_Counting_and_Density_Estimation_using_Deep_Network-A_Comprehensive_Survey.