Predictive Risk Analysis for the Arbaeen Pilgrimage Crowds

Asst.Lect. Saif Saad Shihab
Saif.Saad@sc.uobaghdad.edu.iq

Asst. Prof. Suhad Faisal Shehan Maan
suhad.f@sc.uobaghdad.edu.iq

Asst.Prof. Sarmad Makki Mohammed Al-Qaisi
sarmad.garib@sc.uobaghdad.edu.iq

Asst.Dr.Ahmad Hashim Hussein
ahmedhashem@pgiafs.uobaghdad.edu.iq

College of Science – Department of Computer Science – University of Baghdad

Abstract

Effective crowd management is essential for maintaining public safety during large-scale events. Traditional approaches often prove inadequate in addressing real-time changes and the complex dynamics of high-density crowds. This challenge is particularly evident during the Arbaeen pilgrimage in Karbala, Iraq—one of the largest annual human gatherings—where millions of pilgrims converge within confined areas. The absence of predictive, data-driven systems to monitor and guide crowd movement increases the risk of congestion, delays, and potential casualties. This study introduces an intelligent framework that leverages machine learning and real-time analytics to enhance crowd management efficiency. An artificially generated dataset was constructed to simulate real-world conditions, including features such as visitor numbers, crowd pressure, average speed, environmental conditions, and emergency events. The system calculates a dynamic «risk degree» index, classifying areas into three behavior categories: Normal (73.7%), Suspicious (5.0%), and High Risk (21.2%)—enabling early detection and intervention. To support strategic planning, a linear regression model is employed to forecast trends in visitor numbers, pressure, and risk levels. While the model effectively captures overall patterns, it is limited in predicting abrupt fluctuations, underscoring the importance of real-time data monitoring. For optimal path planning, Dijkstra's algorithm is applied with risk-weighted edge costs. The resulting smart route significantly outperforms the traditional path in both safety and efficiency, with a total cost of 11.18 compared to 85.70, and an average risk of 4.92 versus 8.26. An interactive web-based dashboard supports visualization and decision-making through real-time alerts, heatmaps, and exportable analytical tables.

Keywords: Crowd Management, Predictive Risk Analysis, Route Optimization, Crowd Behavior Modeling, Arbaeen Pilgrimage

Introduction

The Arbaeen pilgrimage represents an exceptional humanitarian and spiritual event, and is considered one of the most prominent religious events and the largest human gatherings in the world. Every year, Iraqis and millions of Pilgrims from countries around the world commemorate this occasion by participating in huge processions on foot, covering vast distances towards the holy city of Karbala, where the shrine of Imam Hussein (peace be upon him) is located. This walk is considered a true embodiment of the deep values of loyalty and love associated with this occasion (Bhardwaj,S,Dwivedi, A., Pandey,A,Perwej, D.Y, &Khan,2023)(Hashjin, Z.G,& Khanghahi, 2019). Naturally, organizing this global event requires logistical and organizational tasks at the best levels of quality, as it presents huge collective efforts weeks in advance to ensure the success of this major humanitarian event and its safe and smooth flow.

However, this widespread form of participation poses very complex logistical and security challenges in terms of crowd management, as unprecedented numbers of Pilgrimsinherently complicate the process of ensuring smooth movement and achieving effective and continuous crowd control (Choi, Y. W., & Eltahir, 2022). In addition, extreme heat or heavy rainfall (unstable environmental factors) may affect the safety of Pilgrims and significantly disrupt their movement dynamics. Also, to ensure optimal safety and prevent potential threats in crowded environments, various security measures should be applied. Crowd management and orientation pose a complex challenge, especially given sudden changes in individuals behavior and surrounding conditions. Indeed, traditional methods are often unable to respond quickly or predict potential risks, which limits their effectiveness in sensitive situations. To overcome these limitations, it is important to use smart systems based on artificial intelligence and

data analytics, as they provide predictive and analytical capabilities that contribute to improving decision-making and directing crowds more efficiently towards shelter or exit areas, whether through traditional or alternative methods (Soltani, A., Aram, M.,Alaeddini, F,& Marzaleh, 2021) (Karampourian, A.,Ghomian,Z,& Khorasani-Zavareh,2018). Subsequently, this study is based on an artificial dataset that simulates reality for assessing crowd dynamics, predicting future density, and measuring the pressure level and risk at Arbaeen Pilgrims gathering areas, which may be implemented using machine learning (ML) techniques. It also takes into account the influence of environmental factors such as weather or physical obstacles. A particularly valuable aspect of the system is its ability to classify crowd behavior into three categories: «normal,» «suspicious,» and «high-risk,» based on flexible thresholds. This classification provides real-time alerts supporting concerned parties to intervene early, and address congestion or security problems quickly and effectively. As well as, it suggests the optimal safety path for the Pilgrimage crowds.

RELATED WORK

To contextualize this study, several researches on crowd analysis have been published over the past decade, reflecting an increasing interest in handling large gatherings effectively. While some focus narrowly on specific subtopics such as crowd counting or behavior recognition, others provide a broader examination of the field as a whole. This study synthesizes the most significant contributions relevant to the Arbaeen Pilgrimage, one of the world largest annual religious gatherings, and explores the potential of ML techniques for crowd movement analysis and dynamic path suggestion in this context. Although direct research on Arbaeen remains limited, insights from analogous large-scale events, such as the Hajj pilgrimage, offer a foundation for addressing its unique challenges.

A notable example of recent advancements in intelligent routing is presented by (Geng, Y., Liu, E., Wang, R., Liu, Y., Rao, W., Feng, S., ... & Chen, n.d.,2021), who propose a Dynamic Adjustable Route Planning (DARP) algorithm based on Deep Reinforcement Learning (DRL) for pedestrians, which addresses the limitations of current methods that require prior knowledge of the road network and face dynamic traffic congestion. Their approach relies on a competitive deep Q network to enable the agent to learn optimal paths by interacting with a dynamic virtual environment, using an automatic time series model to predict pedestrian flow. So, it allows the agent to adapt to real-time traffic, demonstrating a significant travel time savings of 52% under congested conditions compared to traditional shortest path planning. While (Shah, 2024) proposes an ML model to classify crowd density in real-time in Hajj video footage, addressing the urgent need to identify moderate, crowded, and very dense crowd conditions in a timely manner, rather than just detecting abnormal behavior. The approach relies on an Integrated Gradient Boosting Classifier (GBC) with structural features such as edge density, local binary pattern texture, and crowd area coverage, in addition to data augmentation and class balancing to achieve robustness. Thus, it achieves an accuracy rate of 87% with a low error rate of 2.14%, enabling efficient crowd management and increased safety during large events. In response to growing attention for dynamic crowd analysis, (Moulaei, K., Bastaminejad, S., & Haghdoost, 2024) conducted a comprehensive review to identify health issues, and facilitating factors during the Arbaeen pilgrimage, to address a gap in the general literature on this important event. Its methodology included a comprehensive search in databases and a thematic classification of the extracted data, revealing 61 distinct challenges and 40 facilitative measures. Hence, its main contribution is a comprehensive understanding of health risks such

as the outbreak of infectious diseases and inadequate facilities, along with solutions like tailored training for pilgrims and coordinated efforts among stakeholders, ultimately guiding strategies to improve pilgrims› welfare and enhance event management. Another relevant contribution is presented by (Hu, Y., Fang, Z., Zou, X., Zhong, H., & Wang, 2023) that proposed a new two-phase tourist trip design problem with a crowd dynamics framework (TTDP-CD) for dynamic and personalized route recommendations in major cities. Its main innovation lies in integrating crowd dynamics indicators—derived from mobile phone tracking data—into a multi-objective optimization model to reduce perceived congestion while increasing attraction value and reducing travel distance. The improved non-dominated sorting genetic algorithm II (INSGA-II) was used to solve this complex problem, demonstrating superior performance and solution quality by reducing real-time congestion by 7% compared to previous methods. To push the boundaries of realism and interpretability in crowd modeling, (Chen,B,Guo,R,Zhang,Q,Zhao,Y,Wang,X,&Zhu,2025) introduces an innovative crowd simulation framework driven by data that combines Physics-informed Machine Learning (PIML) with navigation potential fields. Its approach addresses the limitations of traditional rule-based models in dealing with complex human behaviors and data-based approaches in generalization and interpretability, by designing the PI-STGCN to predict pedestrian movement trends under physical constraints. The performed hybrid framework greatly enhances the accuracy and fidelity of simulated crowd movements, outperforms baselines and improves interpretability by leveraging the strengths of both deep learning and physical models. In the pursuit of more accurate crowd behavior forecasting, (Bhardwaj, S., Dwivedi, A., Pandey, A., Perwej, D. Y., & Khan, 2023) proposes a novel Multi Column Convolutional Neural Network (MCNN) based technique

for predicting crowd behavior, aiming to mitigate overcrowding disasters by forecasting secret organized groups of criminals (Mob) activity. The method preprocesses input images to extract low-level features and generate density maps, which are then linearly mapped and quantified using the MCNN algorithm. Evaluated on the ShanghaiTech dataset, this approach demonstrates higher accuracy in crowd counting and density analysis compared to existing methods, showing improved prediction capabilities for abnormal crowd events. Eventually, table 1 summarizes the presented related works that are mentioned in this section.

Table (١): Summary Comparative The Existing Crowd Management and Analysis Techniques

Author(s) & Year	Research Objective	Methodology	Key Findings
Geng et al. (2020)	Minimize travel time for pedestrians; dynamic route planning; avoid congestion.	Deep Reinforcement Learning (DRL), Dueling Deep Q Network (DQN), ARIMA for pedestrian flow, agent-based simulation.	DARP algorithm saves 52% time under congestion compared to the shortest path and learns dynamic traffic.
Shah (2024)	Classify crowd density (moderate, overcrowded, very dense) in Hajj video frames; provide real-time alerts.	Machine Learning, Gradient Boosting Classifier (GBC), structured feature extraction (Edge Density, LBP texture, crowd area coverage), data augmentation, class balancing.	Achieved 87% accuracy, 93% precision, 87% recall, 84% F1-score; 2.14% error rate. Effective for real-time monitoring.

Moulaei et al. (2024)	Examine health challenges and identify facilitators for Arba'in Pilgrimage.	Scoping review of databases (Web of Science, PubMed, Scopus, Google Scholar), data extraction, thematic categorization.	Identified 61 health challenges (e.g., infectious disease, poor waste management) and 40 facilitators (e.g., customized training, stakeholder coordination).
Hu et al. (2023)	Develop dynamic, personalized tour routes for tourists by integrating crowd dynamics.	Two-stage route strategy ("global optimization first, local update later"), INSGA-II (improved non-dominated sorting genetic algorithm II), mobile tracking data.	Suggested method outperforms NSGA-II, MOPSO, MOACO, WSM in solution quality and decreases real-time crowding by 7%.
Chen et al. (2025)	Propose data-driven crowd simulation framework integrating physics-informed machine learning (PIML) and navigation potential fields.	PI-STGCN (Physics-informed Spatio-temporal Graph Convolutional Network) for movement prediction; physical model of navigation potential fields.	Framework outperforms rule-based methods in accuracy and fidelity; 10.8% improvement in trajectory similarity.
Bhardwaj et al. (2023)	Predict mob behavior for crowd management using machine learning.	Multicolumn Convolutional Neural Network (MCNN); image preprocessing; ShanghaiTech dataset.	MCNN accurately predicts crowd count/density on ShanghaiTech dataset; effective for forecasting potential incidents.

Methodology

In complex event environments data-driven approaches are increasingly vital, a major challenge requires advanced analytical tools to ensure Pilgrims safety, and improve emergency responses. This section outlines the employed methodology in the design and development of the proposed intelligent system. It presents basic theoretical concepts, mathematical models, and system architecture in detail. The discussion also addresses the specific methods applied in data handling, continuous risk analysis, classification of evolving crowd behaviors, predictive modelling and adaptive path planning. This comprehensive account forms the scientific and technical basis of the system and supports its role in improving public safety and improving crowd management.

1. Observed Dataset

The limitations about acquiring real dataset for this study compelled us to generate a predetermined Artificial Dataset. The observed dataset is composed of artificially generated .csv file of 18 columns and 5000 records (Shihab et al., 2025) as in Figure 1. A custom script, utilizing the Pandas, NumPy, and random Python libraries, had been developed to generate this dataset. The data generation process for record building was guided by a set of procedural rules and heuristic formulas, designed to simulate realistic scenarios.

m_Barrier	from_Event	from_Latitude	from_Longitude	from_Pressure	from_Speed	from_Visitors	from_Weather	from_area	to_Barrier	to_Event	to_Latitude	to_Longitude	to_Pressure	to_Speed	to_Visitors	to_Weather	to_area
FALSE	-	31.84357376	43.75962913	1.68	4.76	1669	Dusty	Street_106	TRUE	Food Distributic	30.39497571	46.31033283	3.64	0.92	7642	Dusty	Street_148
TRUE	-	32.07833235	45.35566592	1.74	2	5144	Dusty	Crossroad_30	FALSE	-	32.69474921	44.89675941	0.98	1.24	3590	Dusty	Village_56
TRUE	Food Distributic	30.91461434	44.35106743	3.13	0.72	2270	Rainy	Crossroad_52	TRUE	Major Processic	30.71404054	44.05732844	1.9	1.61	4926	Sunny	Crossroad_55
FALSE	-	32.32519727	47.7054399	2.43	4.32	6747	Sunny	Village_48	FALSE	Food Distributic	31.35361147	44.02641631	1.29	3.47	4385	Sunny	Square_121
FALSE	-	33.29076164	44.79213194	1.78	3.51	2695	Dusty	Crossroad_109	FALSE	-	33.33320798	45.94119492	1.65	3.96	3290	Dusty	Crossroad_60
TRUE	-	32.55499444	44.98766912	2.92	0.92	7096	Rainy	Street_16	TRUE	Gathering	32.20551586	47.56618967	1.93	0.57	4118	Rainy	Village_34
TRUE	-	32.20063213	43.94659961	3.29	1.47	3352	Sunny	Crossroad_101	FALSE	-	31.84357376	43.75962913	1.68	4.76	1669	Dusty	Street_106
FALSE	Food Distributic	30.18743965	45.80665048	2.46	2.47	6310	Rainy	Crossroad_88	TRUE	-	31.71321771	43.62267136	2.62	0.87	4560	Dusty	Street_14
FALSE	-	32.30102482	44.8528791	1.85	3.2	3718	Rainy	Crossroad_9	TRUE	-	30.87229655	46.21879032	2.4	3.53	4542	Dusty	Street_96
FALSE	-	33.09065786	47.67883619	1.96	0.56	3245	Sunny	Village_46	TRUE	Food Distributic	31.59147551	46.23218971	2.09	4.37	2770	Sunny	Street_116
FALSE	Food Distributic	32.03965107	43.55561457	1.99	1.16	7001	Dusty	Street_58	TRUE	-	32.55409444	44.98766912	2.92	0.92	7096	Rainy	Street_16
TRUE	Major Processic	32.65078763	45.81303641	3.02	1.19	3300	Dusty	Square_118	FALSE	Gathering	30.09788096	47.87576663	2.94	3.82	11676	Dusty	Square_72
FALSE	Food Distributic	30.18743965	45.80665048	2.46	2.47	6310	Rainy	Crossroad_88	TRUE	Gathering	30.61466065	45.84719528	4.02	0.99	9818	Dusty	Square_76
TRUE	-	30.505374	47.818228	4.77	1.49	18198	Dusty	Basra	FALSE	-	32.32519727	47.7054399	2.43	4.32	6747	Sunny	Village_48
FALSE	-	33.21563719	44.4217009	2.45	4.34	5033	Rainy	Village_103	FALSE	Major Processic	30.24912243	47.47864523	1.49	0.74	2944	Sunny	Village_2
TRUE	Food Distributic	32.3294293	46.73460906	1.96	2.03	3012	Rainy	Village_143	FALSE	Food Distributic	31.80850499	45.39100534	1.55	3.6	7428	Rainy	Village_82
FALSE	-	32.32519727	47.7054399	2.43	4.32	6747	Sunny	Village_48	FALSE	Food Distributic	31.996162	44.341144	4.58	1.87	39940	Rainy	Najaf
FALSE	Major Processic	31.27057421	46.80300622	1.39	3.19	1334	Dusty	Street_142	TRUE	-	31.86550603	47.09627985	3.15	5.21	5767	Sunny	Street_140
FALSE	Gathering	30.22596438	46.18423639	1.73	3	9177	Sunny	Square_67	FALSE	Food Distributic	30.79896309	46.36475013	2.14	0.59	9493	Sunny	Square_131
TRUE	-	33.28531274	45.31175765	3.57	3.41	6627	Rainy	Village_15	TRUE	Food Distributic	31.19054936	44.38924655	2.23	4.42	4061	Rainy	Village_12
FALSE	Food Distributic	31.72777954	44.24985134	1.94	1.08	4842	Sunny	Square_108	TRUE	Food Distributic	32.62187027	45.07113362	2.09	5.35	3616	Sunny	Crossroad_127
FALSE	Food Distributic	33.24669681	43.85118901	0.86	4	2944	Rainy	Crossroad_3	FALSE	Major Processic	33.19861652	47.26368715	2.42	4.69	6627	Dusty	Crossroad_89
TRUE	-	33.067215	44.39604898	3.21	5.15	6718	Rainy	Street_137	FALSE	Major Processic	30.24912243	47.47864523	1.49	0.74	2944	Sunny	Village_2
TRUE	Food Distributic	33.12606475	44.48115703	1.49	3.47	580	Sunny	Crossroad_91	FALSE	Gathering	30.60134247	45.50084967	1.44	1.83	5194	Dusty	Village_134
TRUE	-	30.48451873	45.93156248	3.4	5.36	3850	Rainy	Street_130	FALSE	-	31.08785473	43.70502251	1.84	3.09	7258	Rainy	Square_126
FALSE	Major Processic	33.03193718	44.81629877	2	2.02	3279	Rainy	Crossroad_38	TRUE	-	32.88286724	45.6641359	3.1	2.77	6649	Rainy	Crossroad_144
TRUE	-	32	45	1.78	1.81	5308	Rainy	Diwaniyah	FALSE	-	30.43837343	46.43288674	2.15	1.32	3278	Sunny	Crossroad_73
TRUE	Major Processic	33.17574893	44.29144563	2.51	1.7	8949	Sunny	Mahmudiyah	FALSE	Food Distributic	30.5397103	45.75797387	1.73	2.3	5379	Rainy	Square_6
TRUE	Gathering	31.6717293	44.66118003	3.31	2.61	7322	Rainy	Square_65	TRUE	Food Distributic	31.93294593	45.14260658	2.74	4.28	7833	Dusty	Village_114
FALSE	-	32.14636712	45.90546753	2.73	4.91	8311	Sunny	Square_78	TRUE	Major Processic	33.17574893	44.29144563	2.51	1.7	8949	Sunny	Mahmudiyah
FALSE	-	31.04560998	45.77143153	2.11	3.06	3020	Dusty	Crossroad_135	TRUE	-	31.86550603	47.09627985	3.15	5.21	5767	Sunny	Street_140
TRUE	-	32.78849611	44.87533247	3.19	0.97	5231	Rainy	Square_41	TRUE	Gathering	30.66815025	44.15203942	1.36	2.38	2475	Dusty	Crossroad_93
FALSE	Food Distributic	33.24669681	43.85118901	0.86	4	2944	Rainy	Crossroad_3	TRUE	-	33.28531274	45.31176765	3.57	3.41	6627	Rainy	Village_15

Figure 1: Observed artificial generated dataset features

2. THE PROPOSED MODEL

This subsection provides precise definitions for the mathematical models and computational logic. These frameworks are fundamental for converting observed data into actionable insights, thereby supporting effective crowd mobility management. For each presented formula, a clear description is provided, alongside exact definitions for its variables and relevant contextual details concerning its implementation within the system, as in the following and shown in Figure 2 that presents a general perspective of the proposed model for this study:

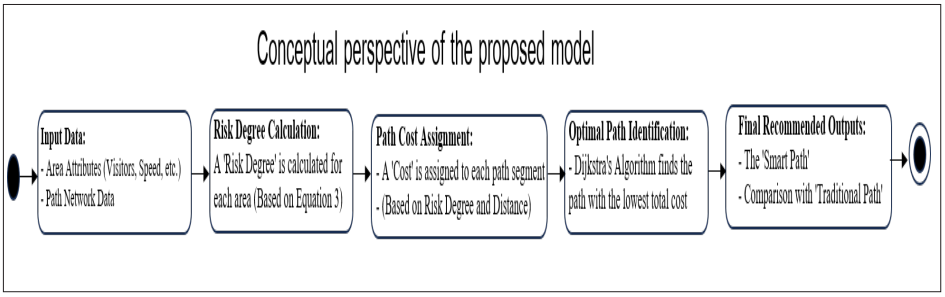


Figure :2 Conceptual perspective of the proposed model

A. Calculating crowd status measures and behavioral analysis

1. Crowd Pressure and Density Metrics

In the study of crowd dynamics, Crowd Density (D) (Kui, Q.,Li,J.,& Zhang, n.d,2009) is a foundational metric, defined as the number of individuals (N) per unit area (A), as shown in Equ.(1):

$$D = \frac{N}{A}(individuals/m^2) \dots\dots\dots (1)$$

To assess crowd concentration within the constraints of the proposed simulation data, an operational index was developed. This Estimated Pressure Index ($P_{index,i}$), which uses the Number of Visitors (Vi) as an alternative to the theoretical crowd density measure (D). Based on (Liang, H,Yang,L,

Du,J,Shu,C.W, & Wong, 2024) in Crowd Modeling, the basic concept has been extended to meet the specific requirements by incorporating additional critical factors, namely the effect of events (Fevent), and the effect of physical barriers (Fbarrier). The estimated pressure index ($P_{index,i}$), which measures the effective impact of individuals on the local environment for each area (i), is thus calculated according to Equ. (2):

$$P_{index,i} = \left(\frac{V_i}{C_1} + R_i^{rand}\right) \cdot F_{E,i}^{gen} \cdot F_{B,i}^{gen} \ \ldots\ldots\ldots (2)$$

Where, ($P_{index,i}$) represents the Estimated Pressure Index, which corresponds to the value stored in the 'Pressure' column within the data, (Vi) represents the visitor numbers, C1 is an empirical scaling constant (10,000) for normalization, and Rirand is a random component introduced to simulate real world variability. The final value is modulated by a generation-specific Event Factor($F_{E.i}^{gen}$)as 1.3 for a major procession, and a Barrier Factor as 1.5 for a barrier. Hence, (serves as a robust proxy for crowd pressure, which is acknowledged that for a full-scale deployment, integrating directly measured Crowd Density (D) would be a critical enhancement for achieving higher precision in risk assessment.

2. Risk Level (R_1)

To provide a comprehensive indicator of potential risk, a composite Risk Level (R_i) is defined for each area i. Inspired by foundational works in multiscale crowd modeling (Bellomo, N, Bellouquid, A., & Knopoff, 2013), and multi-agent system control (Gong, X., Herty, M., Piccoli, B., & Visconti, 2023), a tailored multiplicative risk model has been developed. As formulated in Equ.(3), this model integrates crowd pressure ($P_{index,i}$), the inverse of crowd speed (1/Si), and a set of environmental factors as Fweather, Fevent, Fbarrier to quantify the situational risk. Therefore, a higher value of R_i cor-

responds to an increased probability of adverse events occurring

$$R_i = P_{idx,i} \times \left(\frac{1}{S_i+\epsilon}\right) \times F_{weather,i} \times F_{event,i} \times F_{barrier,i} \text{ ------>} (3)$$

Where, (Pindex,i) is the Pressure Index for area I as defined in earlier, Si represents the Average Crowd Speed (2.5 units/second), since a lower speed signifies higher risk. As well as very small constant, ϵ (0.001) is added to S_i to ensure numerical stability by preventing division by zero. Also, the final score is modulated by three dimensionless factors: The Weather Factor ($F_{weather,i}$) as 1.2 if adverse conditions like rain else 1.0 otherwise. The Event Factor ($F_{event,i}$) as 1.3 for a major procession; and the Barrier Factor ($F_{barrier,i}$) as 1.5 if a barrier is present.Actual Behavior Classification

3. The system classifies the observed crowd behavior within each area (i) into one of three distinct simplified categories: ″Normal″,″Suspicious″ or «High Risk.» This rule-based classification facilitates rapid situational comprehension and targeted alerting for operators. The classification is determined by evaluating the Risk Degree (R_i), the presence of a Barrier ($Barrier_i$), the Crowd Speed (S_i), and the Type of Event ($Event_i$). Actually, the hierarchical logic for this classification is formally described in Algorithm 1.

Algorithm 1: Crowd Behavior Classification.

Input:Risk_Score(Ri),Barrier_Status(Barrier_i),Crowd_Speed(Si), Event_Type (Event_i) Output: Behavior_Classification

1. IF (Ri > T_HighRisk) AND (Barrier_i is True) AND (Si < S_Stagnation) THEN
2. RETURN "High Risk"
3. ELSE IF (Ri > T_Suspicious) OR (Event_i = "موكب كبير") THEN
4. RETURN "Suspicious"
5. ELSE
6. RETURN "Normal"
7. END IF

However, $T_{HighRisk}$ represents the High-Risk Classification Threshold as 5.0, However, $T_{HighRisk}$ represents the High-Risk Classification Threshold as 5.0, $T_{Suspicious}$ is the Suspicious Behavior Classification Threshold as 3.0, where $T_{Suspicious} < T_{HighRisk}$ always. Also, $S_{Stagnation}$ denotes a speed threshold as 2 units/second when crowd movement is considered stagnant. As well as, the logical operators $\wedge$ and $\vee$ represent "AND" and "OR," respectively. Thus, an area is classified into "High Risk" only if its risk score is critical, a barrier is present, and movement is severely impeded. Otherwise, it is classified as "Suspicious" if its risk score is moderately high or if a large procession event is active. Finally, if neither of these conditions is met, then the behavior is deemed "Normal."

B. Predictive Modeling for Crowd Dynamics

A key capability of the proposed system is its predictive analytics, which is designed to provide proactive insight into the emerging conditions of crowds. And based on trend analysis methodologies in pedestrian

flow modelling as in (Das, P., Parida, M., & Katiyar, 2015), the forecasting module utilizes simple linear regression principles to estimate future trends for vital features (visitors number, pressure, and risk degree). So, the computational efficiency and high interpretability in extrapolating linear patterns from serial data, providing a clear framework for crowd dynamics prediction as defined in Equ.(4)

$$\hat{Y}_t = \beta_0 + \beta_1 t \quad \text{..... (4)}$$

Where, $\hat{Y}_t$ Represents the Predicted Value of a given features (visitor numbers, pressure, or risk degree) at a future time or sequential step t. The Intercept Coefficient β_0 is the baseline value of the metric when the independent variable t=0. While, the Slope Coefficient $\beta_1$is the average rate of change in the metric of each unit increasing in temporal or ordinal index (t). For model training, t corresponds to the ordered sequence of currently available data points N= {0,1,…,N−1}. For forecasting, t is extended to future steps, {N,N+1,…,N+M−1}, Where, M is the number of steps to be predicted. The model is trained using the current metric values as the target output and their corresponding indices as input, then used to extrapolate future values.

C. Risk Alerting and High-Risk Zone

In order to identify immediate and actionable intelligence, the framework incorporates a dual-layered mechanism for risk notification. This mechanism is designed to clearly identify and visually highlight areas where risk levels exceed critical thresholds. The first layer involves a direct alert trigger based on the calculated risk degree (). An alert status for each area (i) is determined by comparing its against a predefined operational threshold (). So, ″High Alert″ status is assigned to an area if its surpasses this threshold, as in the formulated rule in Equ.(5):

$$Alert_i = High\ Alert \qquad if\ R_i > T_{Alert} \ \ldots\ldots\ldots (5)$$

Where, R_1 is the Risk Degreefor area (i), and T_{Alert} is the configurable-High-Risk Alert Threshold(5.0), which establishes the critical point for triggering an explicit alert.

By the other hand, the second layer provides an intuitive visual mapping on the system interactive display, where the marker color for each area is directly determined by its Actual Behavior classification: a red marker is assigned if the behavior is "High Risk", an orange marker if it is "Suspicious Behavior", and a green marker if it is "Normal". This dual approach ensures that operators can rapidly identify and locate high-risk zones through both the explicit "High Alert" designation and the unambiguous color-coded visual cues on the geographical display, thereby facilitating swift and informed decision-making.

D. Smart Route Optimization

This phase provides advanced pathfinding capabilities designed to guide crowds along safer and more efficient paths. It achieves this by performing a comparative analysis and optimization of crowd movement, dynamically balancing traversal risk with geographical distance.

1. Edge Weight Function

The core of the smart routing phase is based on the sophisticated Edge Weight Function (W(u,v)), which defines the composite «cost,» W(u,v), of crossing a path segment. This cost is not merely based on geographical distance. The formulation of a cost functions as a linear weighted sum of multiple objectives, integrating both traversal risk and distance, which is a standard approach in multi-objective pathfinding as in (Alwan, N. A., 2014). Thus, a lower weight would be more desirable path segment as determined in Equ. (6)

$$W(u,v) = (\alpha \cdot R_v) + \big(\beta \cdot Dist(u,v)\big) \ldots\ldots\ldots 6$$

Where, W(u,v) is the Composite Weight (Cost) of the edge. And, α is a configurable Risk Weight Factor ranging from 0.0 to 10.0, which controls the relative importance of the destination Risk Score (R_v), a higher α value prioritizes paths through lower-risk areas. Similarly, β is a configurable Distance Weight Factor ranging from 0.0 to 1.0 that controls the importance of the Geographical Distance between nodes, Dist(u,v), while a higher β prioritizes shorter geographical paths.

2. Geographical Distance Calculation

Where, (Dist(u,v)) is the edge weight function representing the geographic distance, calculated using the standard Euclidean distance formula (Tatit, P., Adhinugraha, K., & Taniar, 2024). This method is used as a computationally efficient alternative to actual road network distances. To ensure a more accurate representation, the coordinates are converted to radians before applying Equ.(7).

$$Dist(u,v) = \sqrt{(lat_v^{rad} - lat_u^{rad})^2 + (lon_v^{rad} - lon_u^{rad})^2} \dots\dots\dots 7$$

Where, (lat_u^{red},lon_u^{re}) and ((lat_v^{red},lon_v^{re}) are the respective latitude and longitude coordinates of nodes u and v after their conversion to radians. The data for these coordinates is sourced from the df_coords DataFrame within the model.

3. Dijkstra›s Algorithm for Smart Path Determination

To determine the optimal "smart path," the system employs Dijkstra's algorithm, an efficient greedy algorithm to find the least-cost path from a single source node to a destination node in a weighted graph with non-neg-

ative edge weights. Actually, the "least cost" corresponds to the path with the minimum cumulative composite weight, as calculated by the (W(u,v)). Then, based on dynamically constructed graphs (G), the observed areas are the nodes and their interconnections are the weighted edges. Regarding source(S) and a destination node (D) computing the optimal sequence of nodes that form the smart path and its total cumulative cost.

4. Traditional Path Definition and Cost

For comparative analysis, the "Traditional Path" is defined as a predetermined sequence of nodes representing a conventional route, with its relevant segments dynamically extracted between the user-selected source and destination. The total cost of this path ($Cost_{Traditional}$) in Equ.8 (Alwan, N. A., 2014).

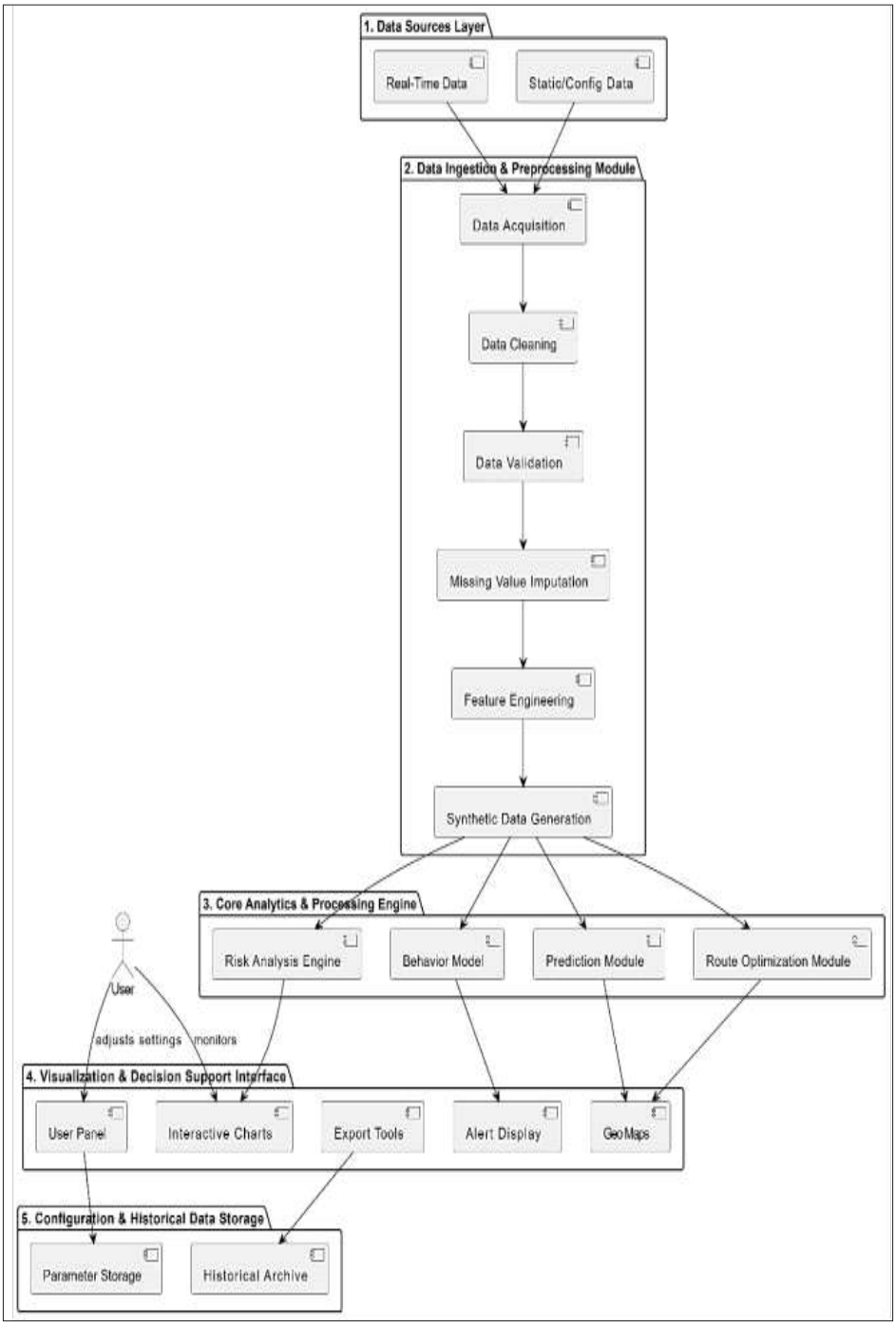
$$Cost_{Traditional} = \sum_{(u,v)\in Traditional\ Path\ Segments} W(u,v) \quad \ldots\ldots\ldots 8$$

5. Smart and Traditional Path Priority

A major aim of the study is to compare and analyze the performance of the «SmartPath» against the predefined «TraditionalPath» that enables informed decision-making. This comparative analysis is achieved by a visualization strategy that allows to assess immediately the benefits of the smart routing algorithm.

In general, the proposed intelligent crowd analysis and prediction system is designed in a modular and multi-layered architecture with a focus on flexibility, scalability, and real-time data processing power. This architecture facilitates the ingestion of diverse crowd data, comprehensive analytical processing of crowd behavior, predictive modeling of future conditions, and the generation of recommendations for optimal routes, so the general structure flow is illustrated in Figure 3.

Figure 3: The Conceptual Diagram of the proposed model



3. Results and Discussion

Data Initialization and Processing Module, acts as the main data entry gateway to the system, responsible for taking the input data and converting them into a structured, clean and consistent format, making them ready for advanced analysis. In addition, the module extracts higher-level features from the raw data architecture. Figure 4 shows the side control interface that allows the user to load data and configure system parameters.

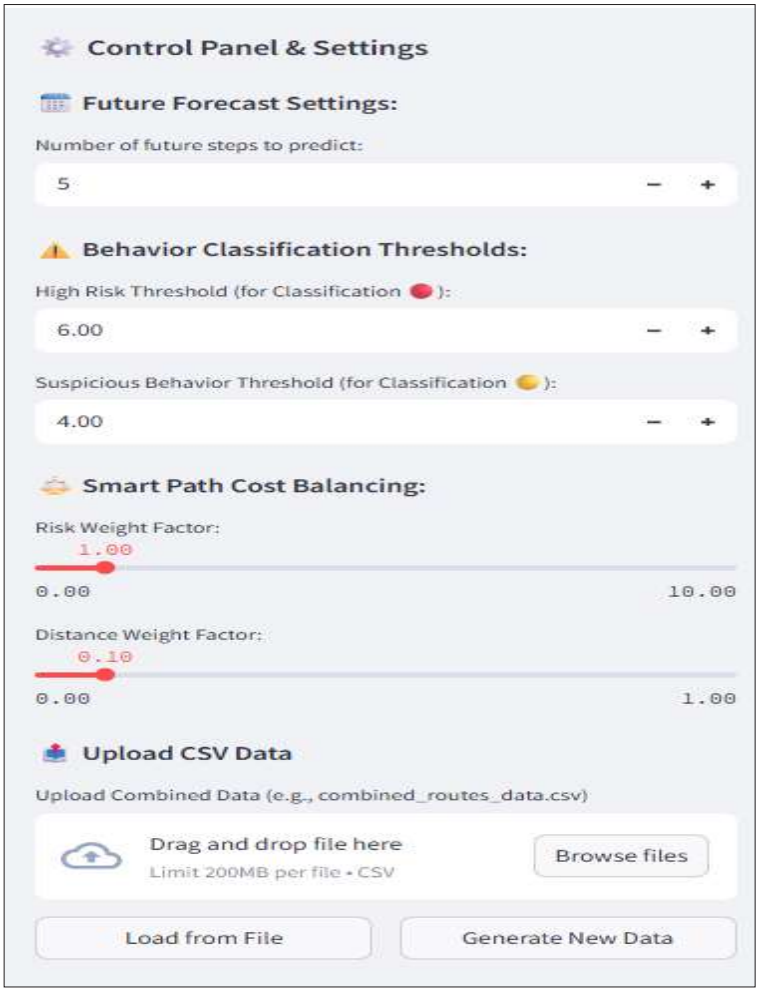


Figure 4: Control interface for load data and configure system parameters

1.Centralized Analysis and Processing Engine

This engine represents the operational brain of the system and is designed to work through several parallel analytical streams to provide a comprehensive, multi-dimensional view of the current state of the crowd by processing different aspects of the data simultaneously.

Regarding the risk analysis and behavior classification, it has been focused on calculating the pressure index (and the advanced risk level () for each observed area. The dataset features are presented ina comprehensive analytical table for direct review, as shown in table 2. It shows the main analytical table, where risks are classified based on dynamic thresholds. Also, it can be seen in figure 5, Row 34 (Sequare_97) shows a very high-risk level that has reached risk degree 13.76, which exceeds the threshold of 6.0, due to barriers existing, and the speed of crowds is very low of 1.21 Km/h. In turn, Row 38 (Village_110) has a risk level of 4.09 that could be considered as high value (suspicious) of threshold falls between the (4.0-6.0), thus classified as “suspicious behavior”. Therefore, the model distinguishes between critical and normal threats.

Figure 5: Area-Level Analysis Table of Visitor Count, Pressure, and Risk

Visitor, Pressure, and Risk Data by Area

	Barrier	Event	Pressure	Speed	Visitors	Weather	Area	Weather_Factor	Event_Factor	Barrier_Factor	Risk_Degree	Actual_Behavior	Alert	Overall Trend
32	✓	Gathering	4.01	2.04	3889	Rainy	Karmat Bani	1.2	1	1.5	3.54	Normal	Normal	10523
33		Major Procession	1.48	4.28	1617	Dusty	Crossroad_1(	1.2	1.3	1	0.54	Normal	Normal	10464
34	✓	Major Procession	7.12	1.21	7338	Rainy	Square_87	1.2	1.3	1.5	13.76	High Risk	High Risk	10404
35	✓	.	4.49	2.86	4180	Sunny	Village_135	1	1	1.5	2.35	Normal	Normal	10345
36	✓	.	3.06	2.35	4748	Rainy	Crossroad_1(	1.2	1	1.5	2.34	Normal	Normal	10285
37	✓	Major Procession	4.53	3	2510	Rainy	Crossroad_2(	1.2	1.3	1.5	3.53	Normal	Normal	10225
38	✓	.	4.98	2.19	5657	Rainy	Village_110	1.2	1	1.5	4.09	Suspicious Behavior	Suspicious Behavior	10166
39		Food Distribution	2.56	1.73	11298	Sunny	Square_71	1	1	1	1.48	Normal	Normal	10106
40	✓	Food Distribution	3.45	3.27	7188	Rainy	Crossroad_3(	1.2	1	1.5	1.9	Normal	Normal	10047
41	✓	Major Procession	14.68	1.6	23376	Sunny	Samawah	1	1.3	1.5	17.88	High Risk	High Risk	9987

As a continuation of the analysis, the model classifies the actual crowd behavior according to predefined rules in Algorithm 1, along with the corresponding alert status, which is presented in a risk degree analysis table, as shown in Figure 6. It demonstrates an intelligent alert mechanism, and automatically identifies any zone with a Risk_Degree exceeding the initial warning threshold of 5.0. It is of great importance to note the distinction between this alert threshold and the more stringent behavior classification threshold of 6.0. An area of Diwaniyah with a Risk_Degree of 5.71 is immediately flagged in the high-risk watchlist, and it is still classified as "Suspicious Behavior". In contrast, an area of Hillah with a Risk_Degree of 31.89 is not only on the watchlist, but also classified as "High Risk" since it has surpassed the higher threshold and met other critical conditions.

	Area	Pressure	Speed	Weather	Event	Barrier	Risk_Degree	Alert
0	Hatwah		4.52	2.44 Sunny	Gathering	☐		1.85 Normal
1	Hillah		24.41	1.79 Dusty	Major Procession	☑		31.89 High Risk
2	Diwaniyah		6.27	1.14 Sunny	-	☐		5.5 Suspicious Behavior
3	Latifiyah		2.83	2.91 Rainy	-	☐		1.17 Normal
4	Al Khayrat		2.85	2.17 Dusty	Gathering	☐		1.58 Normal
5	Kerbala		9.31	0.87 Dusty	-	☐		12.83 High Risk
6	Iskandariya		11.28	1.07 Sunny	Major Procession	☑		20.54 High Risk
7	Basra		3.3	3.46 Dusty	Major Procession	☐		1.49 Normal
8	Musayib		15.92	1.71 Rainy	Gathering	☑		16.75 High Risk
9	Mahmudiyah		7.11	1 Rainy	Gathering	☑		12.79 High Risk

High-Risk Areas (Risk Degree > 5.0):

	Area	Risk_Degree	Alert
1	Hillah	31.89	High Risk
2	Diwaniyah	5.5	Suspicious Behavior
5	Kerbala	12.83	High Risk
6	Iskandariya	20.54	High Risk
8	Musayib	16.75	High Risk
9	Mahmudiyah	12.79	High Risk
11	Hashimiyah	11.34	High Risk
12	Baghdad	11.57	High Risk
13	Square_142	11.09	High Risk
25	Kut	19.93	High Risk

Figure 6: Risk level and alert status analysis table by region

Figure 7 illustrates the distribution of visitor counts across the different areas. The red line, representing the actual observed visitor numbers, shows significant fluctuations with sharp peaks in certain key locations. These peaks usually appear in major urban centers or at important crossroads used for major religious events, which attract much larger crowds than secondary streets or smaller villages. In contrast, the blue line shows the overall linear trend, which acts as a simple average for all locations. This trend line helps to better understand the sharp differences in the red line, highlighting areas with unusually high or low visitor traffic compared to the average.

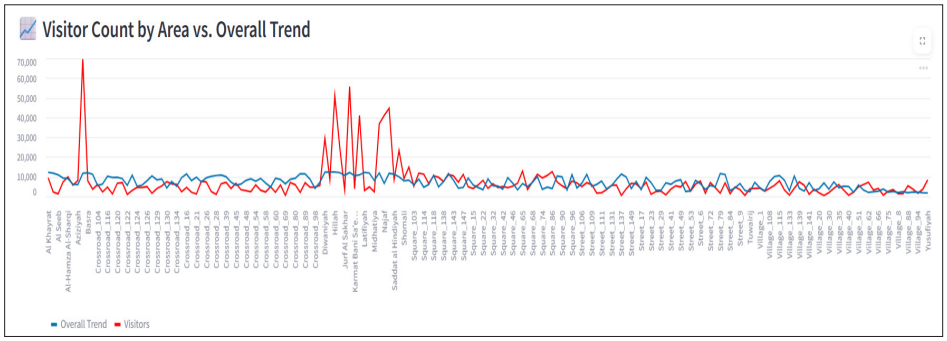


Figure 7: Visitor Distribution by Area vs. Overall Trend Line

To summarize the overall risk distribution across all areas, the system classifies all monitored areas into three distinct behavioral categories. The distribution of these categories is visualized in the pie chart in Figure 8.

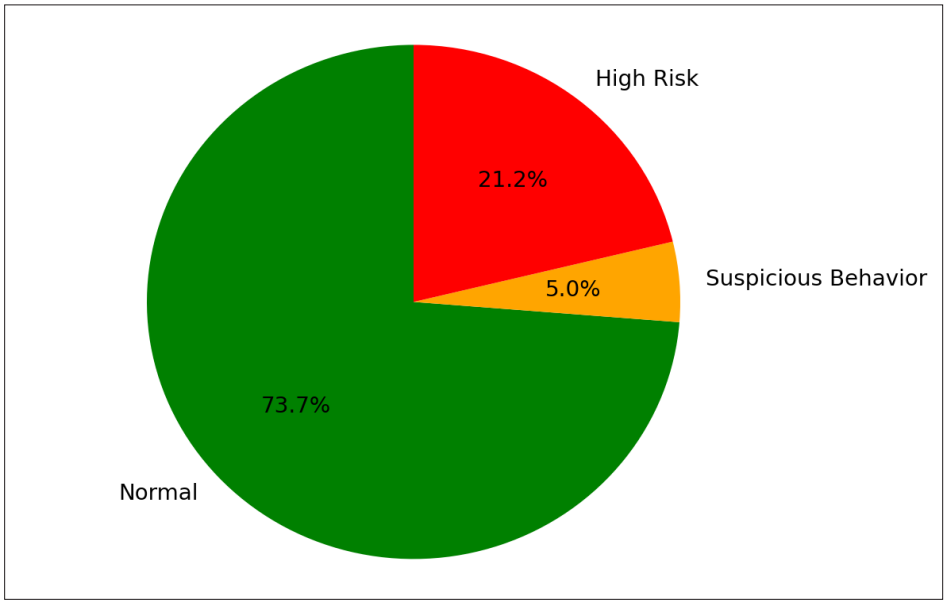


Figure 8: Classified behavior distribution (Normal, Suspicious, High Risk)

The results indicate that while a large majority of areas (73.7%) operate under "Normal" conditions, the system successfully flags two levels of potential threats. A small but significant 5.0% of areas are categorized as "Suspicious Behavior" warranting closer observation. Critically, the model isolates the 21.2% of areas designated as "High-Risk". This crucial segment represents locations where multiple risk factors have converged, allowing authorities to prioritize resources and deploy immediate interventions effectively. Furthermore, an interactive point map that displays all observed areas with color-coded markers that reflect the alert status in real time, and features interactive pop-ups that provide detailed information, as in Figure 9. It displays the interactive geographical map, which serves as the primary visual interface for real-time situational awareness.

Each observed area is represented by a colored marker, based on the configured thresholds like green for "Normal" if risk < 4.0, orange for "Suspicious" if risk 4.0-6.0, and red for "High Risk" if risk > 6.0 and other conditions met.

The Heat Map is used to provide a clear visual representation of the spatial distribution for the crowd pressure along the route, helping to identify congestion hotspots, as shown in Figure 10. It presents a continuous and aggregate view of crowd pressure across the observed region. So, the color intensity, ranging from blue (low pressure) to bright red (high pressure), reflects the calculated Pressure Index (P_(index,i)). Indeed, it identifies broad congestion patterns and "hotspots" where crowd density is highest. The bright red "hotspot" of Figure 10 corresponds to an orange marker on the risk map of Figure 9 indicates that despite high density, other factors (good crowd speed) are currently preventing the situation from becoming critically dangerous.

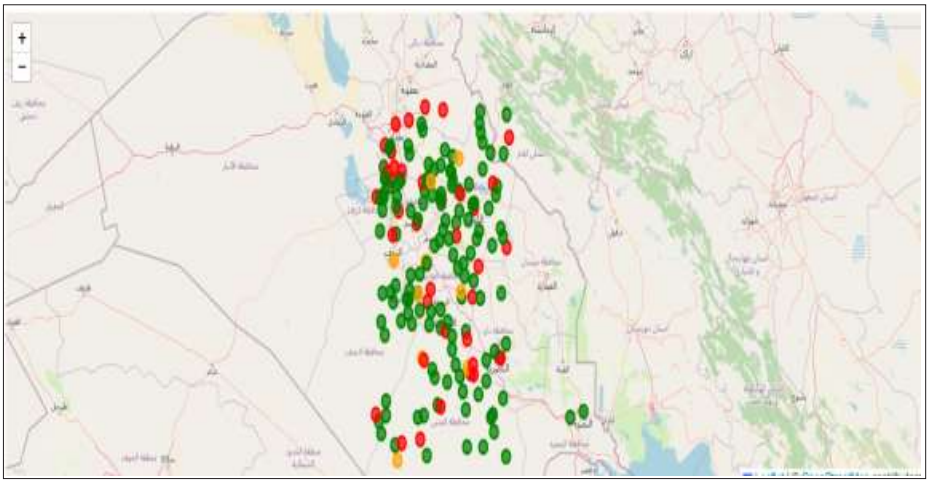


Figure 9: Interactive map of areas with coded risk level markers

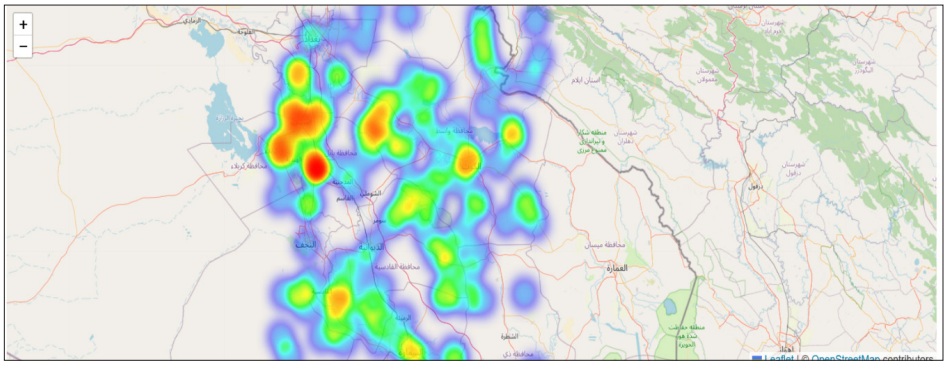


Figure 10: Heatmap of pressure distribution along the route

Figure 11 presents a path comparison interface, designed to analyze and compare different routes on a geographic map, supported by a detailed summary of the analysis. While custom paths can be defined by selecting the start and destination points, the Baghdad to Karbala route is presented as the default case for this demonstration. The main feature is the map itself, which visually plots both the predefined "Traditional Route" (in gray) and the generated "Smart Route" (in blue). This overlay provides an immediate visual comparison of the two paths. To support the analysis, the interface also includes a "Path Comparison Details" panel that shows the calculated costs and key metrics for each route, along with expandable sections that list the full sequence of areas for both the traditional and smart paths.

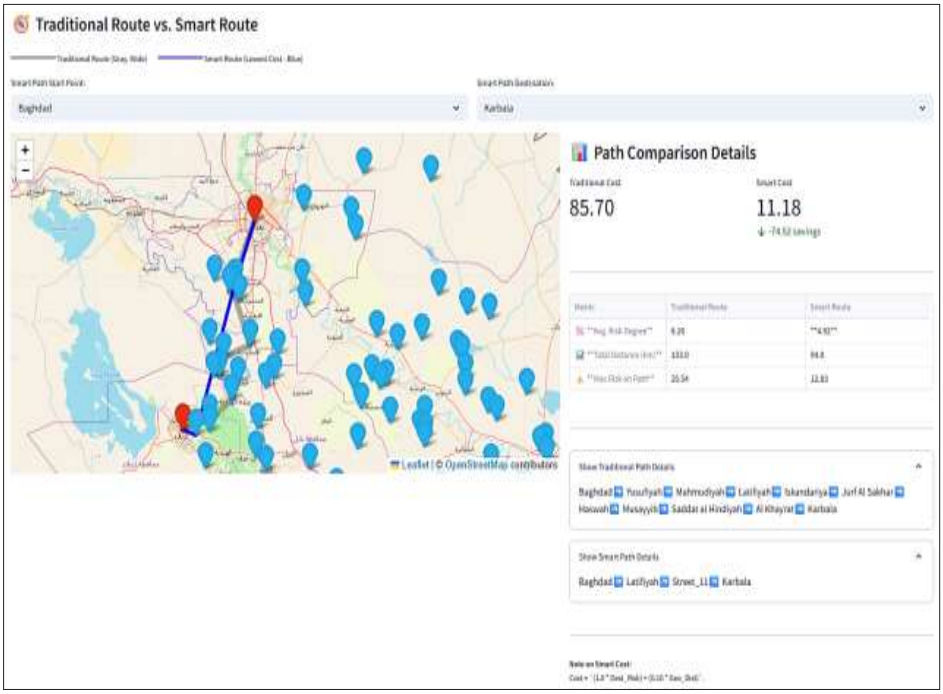


Figure 11: Visual comparison of the proposed Smart Path (blue) and the Traditional Path (grey)

These results clearly show how the model improves route selection. The most important finding is the large reduction in total cost, from 85.70 for the Traditional Route down to just 11.18 for the Smart Route. This significant saving is achieved by finding a safer path. The data shown in the "Path Comparison Details" panel confirms this: the Smart Route has an average risk of only 4.92, much lower than the Traditional Route's risk of 8.26. The map illustrates this strategy visually: the Smart Route (blue line) deviates from the main traditional path to navigate through a sequence of alternative, lower-risk locations. By combining this visual evidence with the numerical data, it is clear that the model successfully identifies intelligent routes that improve crowd safety. Furthermore, intuitive visualizations of future predictions through graphs that display projected trends for key indicators such as the visitors numbers, pressure levels, and risk degree, as illustrated in Figures 12, 13, and 14.

These results clearly show how the model improves route selection. The most important finding is the large reduction in total cost, from 85.70 for the Traditional Route down to just 11.18 for the Smart Route. This significant saving is achieved by finding a safer path. The data shown in the "Path Comparison Details" panel confirms this: the Smart Route has an average risk of only 4.92, much lower than the Traditional Route's risk of 8.26. The map illustrates this strategy visually: the Smart Route (blue line) deviates from the main traditional path to navigate through a sequence of alternative, lower-risk locations. By combining this visual evidence with the numerical data, it is clear that the model successfully identifies intelligent routes that improve crowd safety. Furthermore, intuitive visualizations of future predictions through graphs that display projected trends for key indicators such as the visitors numbers, pressure levels, and risk degree, as illustrated in Figures 12, 13, and 14.

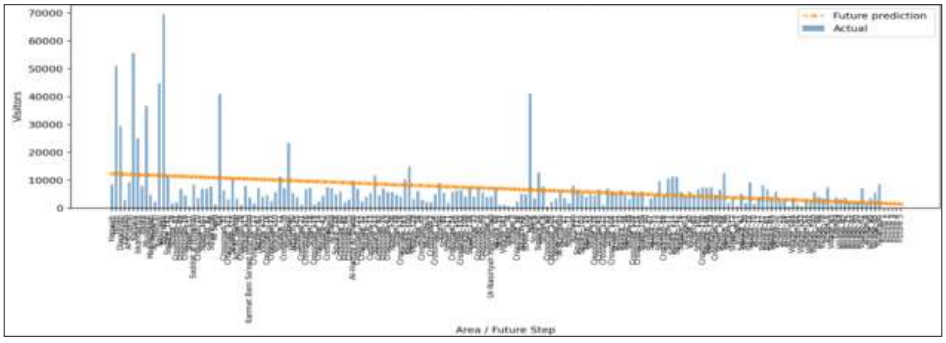


Figure :12 Future prediction for Visitor numbers

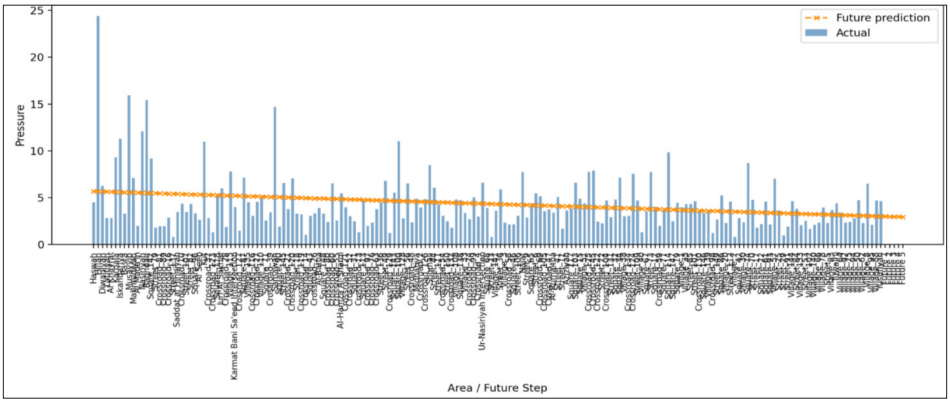


Figure 13: Future prediction for Pressure Level

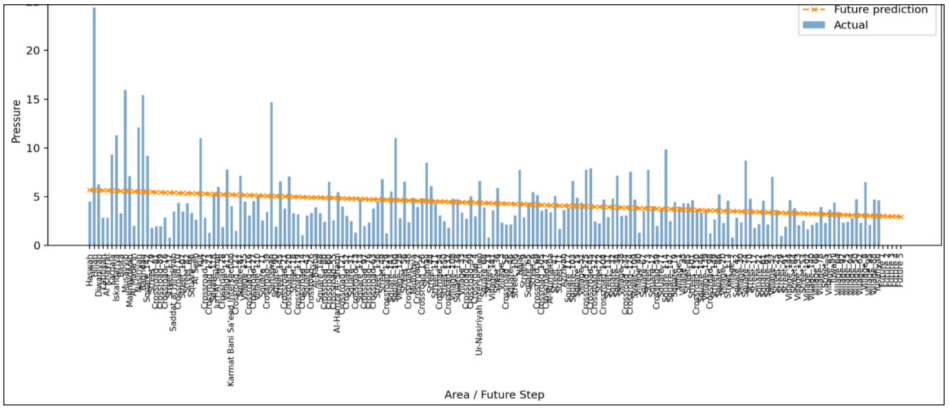


Figure 14: Future prediction for Risk Score

They collectively demonstrate the predictive model utility and limitations .The linear prediction) orange line (effectively captures the average baseline for visitor numbers ,pressure ,and risk degree ,making it suitable for high-level strategic planning .However ,its inability to predict the sharp ,localized spikes seen in the actual data) blue bars (underscores the necessity of complementing these long-term predictions with real-time monitoring to manage immediate ,high-variance incidents .Consequently, Data Export Capabilities has been added to download processed data and analytical results in CSV format ,as demonstrated in Figure.15

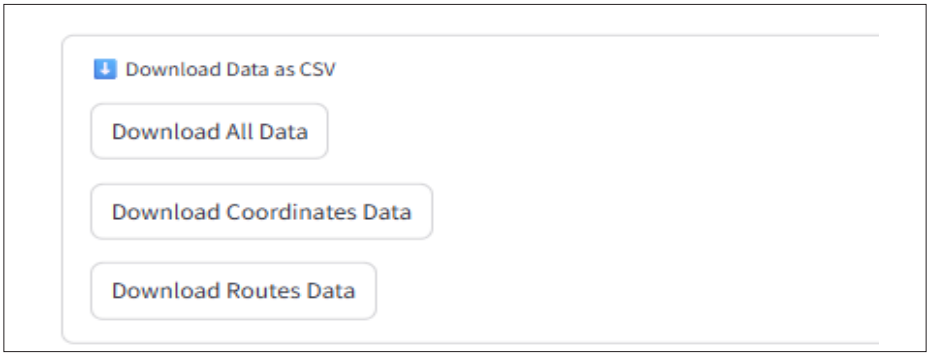


Figure 15: Interface for exporting data in CSV format

Conclusions and Future Works

This study demonstrates an intelligent, modular framework for the real-time analysis and optimization of crowd movement. It establishes a unified system that integrates risk assessment, behavioral classification, predictive analytics, and adaptive pathfinding. So, the key contribution lies in a composite risk model and an efficient graph-based Dijkstra's algorithm, which collectively identify safer and more efficient paths than conventional alternatives. The practical significance of this work is its capacity to shift crowd management paradigms from reactive to proactive, data-driven strategies, thereby enhancing public safety.

Also, the current model, validated with synthetic data and a linear predictive function, provides a strong foundation. In the future, it will focus on advancing its real-world applicability and sophistication. The immediate priority is live validation through integration with real-world data streams from Closed-circuit television (CCTV) and IoT sensors. Furthermore, to enhance predictive accuracy, it could implement more advanced non-linear models, such as Long Short-Term Memory (LSTM) networks. And, could use Reinforcement Learning (RL) for the dynamic, adaptive calibration

of routing parameters. Finally, exploring multi-objective optimization will be pursued to generate a set of Pareto-optimal routes, offering a nuanced range of choices that balance safety, travel time, and distance. These future enhancements are poised to significantly increase the system robustness and practical utility in smart city public safety initiatives.

References

1. Bellomo, N., Bellouquid, A., & Knopoff, D. (2013). From the microscale to collective crowd dynamics. Multiscale Modeling & Simulation, 11(3), 943–963. https://doi.org/https://doi.org/10.1137/130904569
2. Bhardwaj, S., Dwivedi, A., Pandey, A., Perwej, D. Y., & Khan, P. R. (2023). Machine Learning-Based Crowd behavior Analysis and Forecasting. International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN, 2456-3307.
3. Chen, B., Guo, R., Zhang, Q., Zhao, Y., Wang, X., & Zhu, Z. (2025). A data-driven crowd simulation framework integrating physics-informed machine learning with navigation potential fields. IEEE Transactions on Computational Social Systems.
4. Choi, Y. W., & Eltahir, E. A. (2022). Heat stress during Arba'een foot-pilgrimage (World's largest gathering) projected to reach "dangerous" levels due to climate change. Geophysical Research Letters, 49(19).
5. Das, P., Parida, M., & Katiyar, V. K. (2014). Performance Analysis of Dijkstra-Based Weighted Sum Minimization Routing Algorithm for Wireless Mesh Networks. Modelling and Simulation in Engineering, 1, 658408.
6. Das, P., Parida, M., & Katiyar, V. K. (2015). Analysis of interrelationship between pedestrian flow parameters using artificial neural network. Journal of Modern Transportation, 23, 298–309.
7. Geng, Y., Liu, E., Wang, R., Liu, Y., Rao, W., Feng, S., ... & Chen, Y. (n.d.). Deep reinforcement learning based dynamic route planning for minimizing travel time. In 2021 IEEE International Conference on Communications Workshops (ICC Workshops), IEEE, 1–6.

8. Gong, X., Herty, M., Piccoli, B., & Visconti, G. (2023). Crowd dynamics: Modeling and control of multiagent systems. Annual Review of Control, Robotics, and Autonomous Systems, 6(1), 261–282.

9. Hashjin, Z. G., & Khanghahi, M. M. (2019). Messages of Arbaeen Walking as Media. International Journal of Multicultural and Multireligious Understanding, 6(6).

10. Hu, Y., Fang, Z., Zou, X., Zhong, H., & Wang, L. (2023). Two-stage tour route recommendation approach by integrating crowd dynamics derived from mobile tracking data. Applied Sciences, 13(1), 596.

11. Karampourian, A., Ghomian, Z., & Khorasani-Zavareh, D. (2018). Exploring challenges of health system preparedness for communicable diseases in Arbaeen mass gathering: a qualitative study. F1000Research, 7, 1448.

12. Kui, Q., Li, J., & Zhang, B. (n.d.). A Quantitative Model to Describe the Crowd Massing Risk Based on FIST Model. In 2009 International Conference on Information Engineering and Computer Science, IEEE, 1–4.

13. Liang, H., Yang, L., Du, J., Shu, C. W., & Wong, S. C. (2024). Modelling crowd pressure and turbulence through a mixed-type continuum approach. Transportmetrica B: Transport Dynamics, 12(1), 2328774.

14. Moulaei, K., Bastaminejad, S., & Haghdoost, A. (2024). Health challenges and facilitators of arbaeen pilgrimage: a scoping review. BMC Public Health, 24(1), 132.

15. Shah, A. A. (2024). A machine learning model for crowd density classification in Hajj video frames. International Journal of Advanced Computer Science and Applications (IJACSA), 15(12). https://doi.org/https://doi.org/10.14569/IJACSA.2024.0151231

16. Soltani, A., Aram, M., Alaeddini, F., & Marzaleh, M. A. (2021). Challenges of health services during Arbaeen Pilgrimage in 2019. Diabetes, 3(19), 0067.

17. Tatit, P., Adhinugraha, K., & Taniar, D. (2024). Navigating the maps: Euclidean vs. road network distances in spatial queries. Algorithms, 17(1), 29.

18. Shihab, S. S., Behadili, S. F., Makki, S., & Hussein, A. H. (2025). A Synthetic Dataset for Predictive Risk Analysis and Path Optimization of Arbaeen Pilgrimage Crowds (Version 1.0.0) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.15781743