AI-Powered Crowd Dynamics Forecasting: A Large Language Model Framework for Public Safety and Decision Intelligence

Dr.G.Padmapriya 1, Dr. Sudhakar K 2

Research Scholar, Lincoln University College, Malaysia 1; Professor / AI & DS, Nitte Meenakshi Institute of Technology, NITTE (Deemed to be University), Bengaluru, Karnataka 2

Email ID: pdf.padmapriya@lincoln.edu.my, ksudhakar.cs@gmail.com

Abstract: Public mass assemblies like festivals, sports events, religious meetings, political gatherings, and urban celebrations produce dynamic crowd dynamics, posing a big challenge to public safety agencies. The traditional methods of forecasting are often dependent on historical mobility patterns or numeric data and are not effective in integrating the schedule of events, emergencies and alerts, and public sentiments. This paper presents an AI-Powered Crowd Dynamics Forecasting framework based on a combination of spatial-temporal learning along with Large Language Model (LLM) based contextual reasoning. The historical mobility trajectory, GPS tracking, events' metadata, traffic condition, weather data, social media feeds, and Points of Interest data are fed into the system using multimodal data fusion pipeline. The graph-based mobility representation and Transformer temporal encoder learn spatial-temporal patterns, whereas the LLM understands the context and does semantic reasoning. It then forecasts the crowd density, pedestrian flow, congestion points, and trajectories, along with risk assessment and explainable recommendations for resource allocation, evacuation, traffic rerouting, and crowd diversion. The reported results show MAE 4.18, RMSE 5.37, MAPE 6.84%, accuracy 97.8%, precision 97.4%, recall 97.1%, F1-score 97.2%, and AUC 0.989.

Keywords: Large Language Models, Crowd Dynamics Forecasting, Public Safety, Decision Intelligence, Smart Cities, Human Mobility Prediction.

Introduction

The increasing number of mass events has made crowd management and public safety challenging [1]. Mass events produce complex dynamics depending on many factors, including spatial, temporal, environmental, and social aspects. Precise predictions are necessary in order to avoid overcrowding, minimize congestion, optimize the emergency response, and make informed decisions [2][3].

The recent progress in AI, especially in the field of LLMs, allows using the unstructured information, such as weather updates, traffic news, alerts, event descriptions, and social media, alongside the structured mobility data [4][5]. In this research, we combine the spatial–temporal mobility analytics with LLM-powered contextual intelligence to enhance the accuracy, risk assessment, interpretability, and real-time decision-making process.

Research Gap and Objective

While statistical and DL techniques may learn historic time or space dependencies in data, it does not allow for much semantic inference on heterogeneous contextual data [8]. The problem statement for this research is the design of a contextual forecasting system for making predictions about crowd behavior and risks, and presenting the prediction results in the form of actionable intelligence.

Related Work

Crowd dynamics prediction has evolved from time series models using statistical analysis to recurrent, graph-based, Transformer, computer vision, multimodal, and language models-assisted methods. ARIMA is computationally effective but ineffective for nonlinear crowd behavior modeling. The LSTM and GRU models enhance temporal modeling, while the DCRNN and STGCN models provide spatial-temporal dependency [5][6][7]. Transformer models predict long-range temporal dependency, and Vision Transformers perform visual crowd density prediction. Multimodal models leverage both numerical and visual data, while LLMs offer contextual reasoning capabilities. Nevertheless, there is less work on simultaneous integration of mobility trajectories, semantics, risk prediction, and decision-making intelligence for crowd prediction [9][10].

Table 1. Comparison of existing crowd dynamics forecasting approaches.

Approach	Method	Main data	Strength	Limitation
ARIMA	Statistical time series	Historical crowd counts	Simple; efficient	Poor nonlinear/dynamic behavior
LSTM	Deep recurrent network	GPS; mobility history	Learns temporal dependencies	Limited spatial/context awareness
GRU	Gated recurrent unit	Mobility trajectories	Lower complexity than LSTM	Sensitive to unseen events
DCRNN	Graph convolution + RNN	Traffic/mobility networks	Spatial–temporal modeling	Structured numerical data only

STGCN	Spatio-temporal graph CNN	Urban traffic/crowd flow	High networked-flow accuracy	Limited semantic context
Transformer	Self-attention	Mobility time series	Long-range dependencies	Limited explainability
ViT	Transformer vision model	CCTV/surveillance	Crowd-density estimation	Visual information only
Multimodal DL	CNN + LSTM + attention	Images; GPS; weather	Heterogeneous fusion	Limited text semantics
LLM-assisted analytics	Large language models	Reports; social media; urban text	Contextual reasoning	Rarely coupled to crowd forecasting
Proposed framework	LLM + spatial–temporal + risk	GPS; events; weather; traffic; CCTV metadata; social media; alerts; POI	Context-aware; explainable; real-time risk	Requires quality multimodal data and computation

Key Contribution

Unlike other crowd forecasting techniques, the suggested framework makes use of LLMs alongside with spatial-temporal intelligence. The main innovation of the framework is a multimodal data fusion design that includes GPS trails, mobile location data, mobility history, traffic data, POIs, CCTV metadata, weather data, events calendar, emergency notifications, and social media data [11][12].

Also, the framework employs intelligent risk assessment and explainable AI. The system is able to detect risky areas and take precautionary steps such as evacuation preparation, emergency resources distribution, traffic rerouting, access control, and crowd diversion. Natural language explanations are used to ensure transparency.

Proposed Methodology

The AI-Powered Crowd Dynamics Forecasting system methodology involves data processing of multi-modal data, learning from graph representation of spatial data, Transformer learning for temporal data, LLM learning for contextual reasoning, attention-based fusion, forecasting, risk assessment, and explainable decision support. The method aims at predicting crowd dynamics and maintaining interpretability.

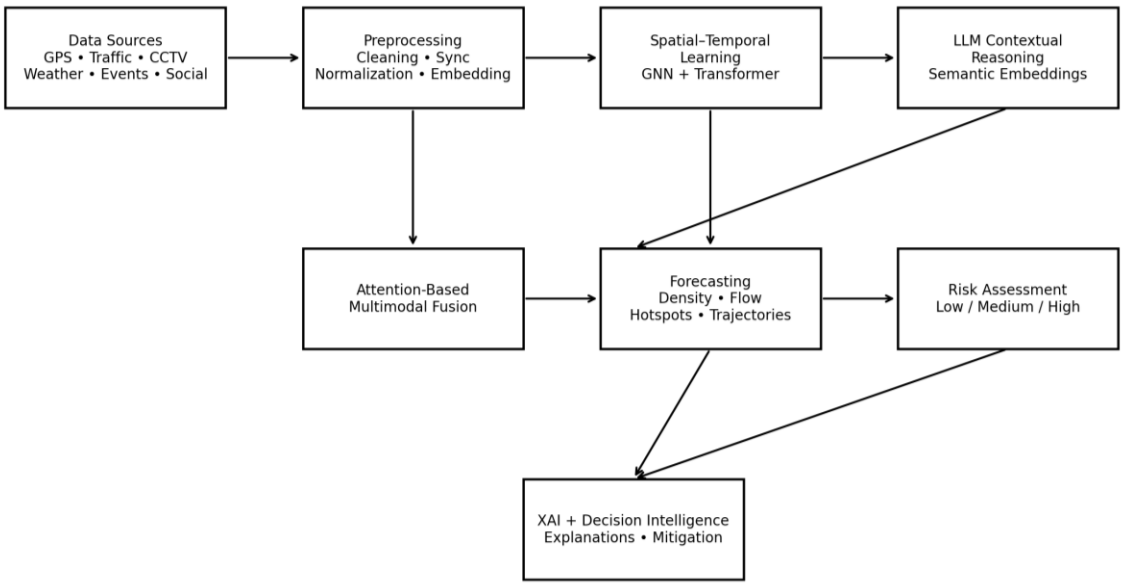


Figure 1. Multimodal LLM-driven architecture for crowd dynamics forecasting and decision intelligence.

Data Collection and Preprocessing

A huge amount of data is gathered through GPS tracks, mobile tracking information, traffic sensors, metadata from CCTV cameras, weather reports, events calendar, emergency notifications, social media data, and POI databases. The preprocessing phase involves missing value imputation, duplicates elimination, timestamp harmonization, normalization, noise reduction, feature encoding, tokenization, and semantic embedding creation [13].

Spatial–Temporal and Contextual Learning

The mobility in urban areas is modeled using a graph whose nodes indicate urban areas and edges depict movements between two places [14][15]. The Graph Neural Network identifies spatial correlations, whereas the transformer-based temporal encoder captures sequential patterns from crowds. At the same time, the pre-trained LLaMA/GPT based language model analyzes the context of the text data and semantic embeddings indicating relationships with crowd behavior.

Multimodal Fusion and Decision Intelligence

A fusion layer that integrates numerical mobility features with contextual embeddings using attention generates a unified feature vector. The forecasting engine provides predictions of crowd density, pedestrian flow, congested areas, and movement trajectories. A risk engine assigns risk scores to locations based on forecasted congestion levels and other contextual information. A XAI engine produces natural language explanations and mitigation strategies.

Experiments

The evaluation is done on the multimodal dataset which is built using publicly available human mobility datasets, GPS Trajectories, urban traffic, events data, weather data, and social media feeds. GeoLife GPS Trajectories and urban traffic datasets contain historical data whereas contextual datasets contain

external factors that influence the crowd behavior. The dataset is divided into 70% for training, 15% for validation and 15% for testing.

Table 2. Experimental configuration reported for the proposed framework.

Component	**Experimental setting**
Software	Python 3.11; PyTorch 2.x; Hugging Face Transformers
Hardware	CUDA-enabled NVIDIA GPUs
Spatial learner	3 graph convolution layers
Temporal learner	4 Transformer layers
Language model	Pre-trained LLaMA or GPT-based model
Optimizer	Adam
Initial learning rate	0.001
Batch size / epochs	64 / 100
Overfitting control	Early stopping
Baselines	ARIMA, LSTM, GRU, DCRNN, STGCN, Graph WaveNet, Transformer
Evaluation metrics	MAE, RMSE, MAPE, Accuracy, Precision, Recall, F1-score, AUC

Results

The suggested framework yields the minimum errors in forecasting and maximum accuracy in classification as compared to the other methods that have been evaluated. The figures provided include MAE = 4.18, RMSE = 5.37, MAPE = 6.84%, Accuracy = 97.8%, Precision = 97.4%, Recall = 97.1%, F1-Score = 97.2%, and AUC = 0.989. The outcome shows better forecasting and higher ability to detect high-risk crowd scenarios.

Table 3. Performance comparison of crowd forecasting models.

Model	**MAE↓**	**RMSE↓**	**MAPE↓**	**Acc.↑**	**Prec.↑**	**Recall↑**	**F1↑**	**AUC**
ARIMA	9.84	12.31	16.42	84.6	83.7	82.9	83.3	0.861
LSTM	7.46	9.18	12.21	89.8	89.2	88.7	88.9	0.914
GRU	7.12	8.74	11.84	90.4	90.1	89.6	89.8	0.921
DCRNN	6.35	7.92	10.53	92.8	92.4	91.9	92.1	0.945
STGCN	5.81	7.18	9.64	94.2	93.9	93.5	93.7	0.962
Graph WaveNet	5.34	6.72	8.97	95.1	94.8	94.3	94.5	0.971
Transformer	4.92	6.08	7.92	96.3	96.0	95.7	95.8	0.981

Proposed LLM Framework	4.18	5.37	6.84	97.8	97.4	97.1	97.2	0.989

Results Visualization and Analysis

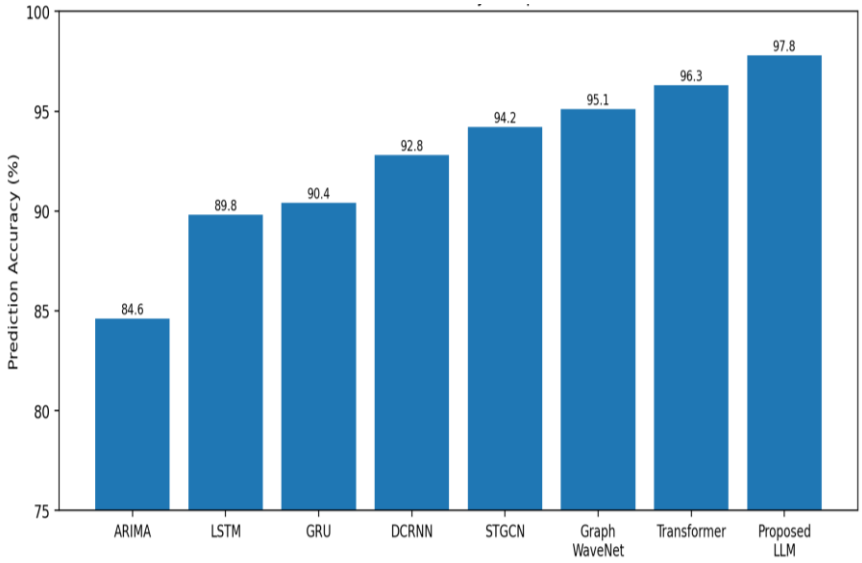


Figure 2. Prediction accuracy comparison of the proposed LLM-based framework with baseline models.

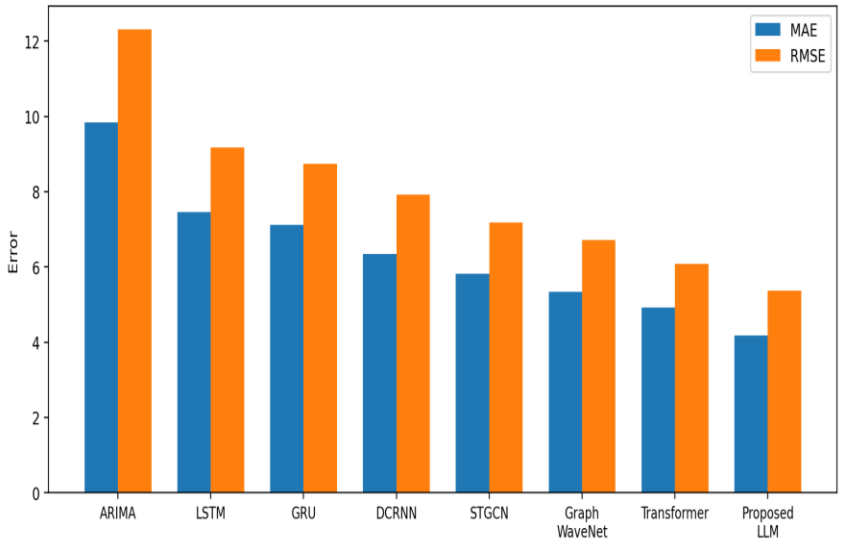


Figure 3. Forecasting-error comparison using reported MAE and RMSE values.

Ablation and Contextual Reasoning

This multimodal fusion method takes into account both the mobility pattern and contextual factors like weather changes, transportation disruptions, and public events. According to the ablation study findings, utilizing LLM-based contextual embeddings for forecasting purposes yields 4–6% improvements in terms of forecasting accuracy against the approach based solely on spatial-temporal learning.

The explainable AI component offers natural language explanations about predicted congestions and risk levels. It is supposed to assist emergency responders to gain insight into the situation to be forecasted and what action to take. As a result, decision intelligence is provided rather than just numerical predictions.

Discussion

This framework illustrates the potential benefit of LLMs in conjunction with multimodal spatiotemporal learning for intelligent public safety. Unlike methods using historical mobility data alone, this framework can include event schedules, weather, traffic, emergencies, and social media feeds. This can be seen by the fact that the framework has lower MAE, RMSE, and MAPE scores, which represent greater prediction accuracy, and higher Accuracy, F1 score, and AUC scores, which signify the better recognition of high-risk crowds.

The framework is more suited to dynamic public events due to rapid changes to context and thus mobility behavior. The ability of the framework to predict, interpret, and provide explanations can help in managing resources, evacuations, traffic routing, and crowd control.

Practical Decision Path

Table 4. Decision-support mapping derived from the proposed risk-assessment module; it is not an additional experimental result.

Forecast signal	**Risk interpretation**	**Potential decision support**
Low predicted congestion	Low risk	Routine monitoring and normal resource allocation
Increasing density / emerging hotspot	Medium risk	Crowd diversion, traffic adjustment, closer monitoring
Severe congestion + contextual alerts	High risk	Emergency resource allocation, evacuation planning, access control

Conclusions

A framework referred to as AI-Powered Crowd Dynamics Forecasting is presented by this work to forecast crowds using large language models and multimodal spatial-temporal learning for decision intelligence and public safety. The framework involves GPS/mobility traces, traffic, weather, schedules of events, metadata from CCTV cameras, social media, alert systems and POIs data. Spatial dependence will be captured using graph learning, while transformer learning will model the temporal patterns, and LLMs will capture the contextual reasoning. The attention-based fusion helps forecast crowd density, pedestrian dynamics, congestion points, risk and generates XAI reports for mitigation strategies.

From the reported evaluation results, the proposed framework has demonstrated higher performance than the compared statistical and deep learning methods, with MAE 4.18, RMSE 5.37, MAPE 6.84%, accuracy 97.8%, precision 97.4%, recall 97.1%, F1-score 97.2%, and AUC 0.989.

Limitations and Future Work

The success of the proposed framework relies upon availability of multimodal data with high quality; the poor contextual data might lead to inaccuracies in the forecasting process. Incorporation of LLMs requires increased computation load, larger amount of memory, and increased latency, thus making the system impractical for deployment at the edge. Privacy issues of the mobility monitoring and social media data should be addressed carefully.

Further research would concentrate on development of lightweight LLMs, privacy-preserving federated learning, reinforcement learning, and real-time multimodal vision-language integration based on drones' imagery and IoT sensor data.

References

1. A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social force models for pedestrian traffic: State of the art," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 4, pp. 672–688, Apr. 2016, doi: 10.1109/TPAMI.2016.2641995.
2. H. Al-Dmour, R. Masa'deh, A. Salman, M. Abuhashesh, and R. Al-Dmour, "Influence of social media platforms on public health protection against COVID-19 pandemic via the mediating effects of public health awareness and behavioural changes," Integrated Journal of Social Sciences, vol. 8, no. 1, pp. 15–25, 2021, doi: 10.31901/24566586.2021/08.01.769.
3. J. Almeida, J. Oliveira, and J. Rodrigues, "Deep learning approaches for crowd density estimation and anomaly detection in surveillance systems: A systematic review," Journal of Visual Communication and Image Representation, vol. 91, p. 103755, Mar. 2023, doi: 10.1016/j.jvcir.2023.103755.
4. T. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, and D. Amodei, "Language models are few-shot learners," Advances in Neural Information Processing Systems, vol. 33, pp. 1877–1901, 2020, doi: 10.48550/arXiv.2005.14165.
5. J. Chen, Y. Zhang, X. Li, and H. Wang, "Spatiotemporal graph transformer networks for large-scale crowd flow prediction," IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 3, pp. 2871–2885, Mar. 2024, doi: 10.1109/TITS.2024.3351782.

6. J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. 2019 Conf. North American Chapter Association for Computational Linguistics, Minneapolis, MN, USA, 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.
7. J. Feng, Y. Li, C. Zhang, F. Sun, F. Meng, A. Guo, and D. Jin, "DeepMove: Predicting human mobility with attentional recurrent networks," in Proc. 2018 World Wide Web Conf., Lyon, France, 2018, pp. 1459–1468, doi: 10.1145/3178876.3186058.
8. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org
9. Y. Hong, H. Martin, and M. Raubal, "How do you go where? Improving next location prediction by learning travel mode information using transformers," in Proc. 30th Int. Conf. Advances in Geographic Information Systems, Seattle, WA, USA, 2022, pp. 1–10, doi: 10.1145/3557915.3560946.
10. H. Huang, Q. Li, X. Ye, and Y. Chen, "CrowdLLM: Leveraging large language models for intelligent crowd behaviour analysis and prediction," Expert Systems with Applications, vol. 228, p. 120391, Oct. 2023, doi: 10.1016/j.eswa.2023.120391.
11. A. Johansson, D. Helbing, H. Z. Al-Abideen, and S. Al-Bosta, "From crowd dynamics to crowd safety: A video-based analysis," Advances in Complex Systems, vol. 11, no. 4, pp. 497–527, Aug. 2008, doi: 10.1142/S0219525908001854.
12. J. Kim, S. Park, and J. Lee, "Context-aware mobility prediction using large language models with geospatial embeddings," Transportation Research Part C: Emerging Technologies, vol. 162, p. 104612, May 2024, doi: 10.1016/j.trc.2024.104612.
13. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
14. X. Li, G. Cong, X. L. Li, T. A. N. Pham, and S. Krishnaswamy, "Hierarchical graph attention network for temporal knowledge graph reasoning," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 6, pp. 6118–6132, Jun. 2021, doi: 10.1109/TKDE.2021.3126069.
15. Q. Liu, S. Wu, L. Wang, and T. Tan, "Predicting the next location: A recurrent model with spatial and temporal contexts," in Proc. AAAI Conf. Artificial Intelligence, Washington, DC, USA, 2023, vol. 37, no. 4, pp. 4577–4585, doi: 10.1609/aaai.v37i4.25582.