MULTIDISCIPLINARY RESEARCH INSTITUTE INDIA

Open Access International Journal of Science and Engineering

ISSN: 2456-3293

Volume 9 Issue 6 (2026) | Pages: 25-31

Crowd Management in Delhi Railways: A Data-Driven Approach: Understanding Passenger Flow Patterns for Effective Crowd Management

Swapnaja Patwardhan[1], Sarita Paliwal[2], Gayatri Shinde[3], Kunal Malekar[4], Aashish Paigude[5], Ruchika Pardeshi[6]

[1,2,3,4,5,6]Department of MCA, MES Institute of Management and Career Courses (IMCC), Pune

[1]*imcc@mespune.in*, [2]*saritapaliwal566@gmail.com*, [3]gayatrishinde024@gmail.com, [4]*kunalmalekar03@gmail.com*, [5]*aashishpaigude007@gmail.com*, [6]*ruchikap2810@gmail.com*

Peer Review Information

Type: *Article*
Received: *26 March 2026*
Revised: *20 April 2026*
Accepted: *28 May 2026*
Published: *17 June 2026*

Abstract

Railway stations in Delhi frequently encounter extreme crowding, particularly during peak festival seasons, posing significant public safety and operational risks. This study is specifically motivated by the tragic crowd crush at New Delhi Railway Station on 15 February 2025, which resulted in 18 fatalities during the Maha Kumbh festival [1]. We analyze passenger footfall and train frequency across five major hubs: New Delhi, Old Delhi, Hazrat Nizamuddin, Anand Vihar Terminal, and Sarai Rohilla. Indian Railways carries approximately 23 million passengers daily [2], making effective density prediction essential. Utilizing a dataset of approximately 10,915 entries, we applied machine learning techniques to predict crowd density. Our analysis identifies critical correlations between train scheduling, seasonal shifts, and dramatic surges in passenger volume during religious observances. Among the models evaluated, the Decision Tree regressor achieved the highest predictive accuracy with a Mean Absolute Error (MAE) of 0.0473, outperforming Random Forest and Linear Regression models. These findings offer a data-driven framework for railway authorities to optimize scheduling and enhance proactive crowd control strategies.

Keywords: Crowd Density Prediction; Railway Station Management; Machine Learning; Passenger Footfall Analysis; Crowd Control; Public Safety.

How to Cite This Article

Patwardhan, S., Paliwal, S., Shinde, G., Malekar, K., Paigude, A., & Pardeshi, R. (2026). Crowd Management in Delhi Railways: A Data-Driven Approach: Understanding Passenger Flow Patterns for Effective Crowd Management. *Open Access International Journal of Science and Engineering*, 9(6), 25–31.

Introduction

India has one of the largest railway networks in the world and plays a vital role in the daily transportation of millions of passengers. According to the Ministry of Railways, Indian Railways carries more than 23 million passengers every day [2], making it one of the busiest railway systems globally. In 2023–24, the network operated 13,198 passenger trains daily covering 7,325 stations and carried 6.905 billion passengers annually [4]. With the rapid growth of urban populations and increasing dependence on public transport, managing large crowds at railway stations has become a major operational and safety challenge.

Crowd congestion at major railway stations can lead to delays, discomfort, and in extreme cases, fatal stampedes. The crowd crush at New Delhi Railway Station on 15 February 2025—which killed at least 18 people including five children as thousands of Maha Kumbh pilgrims surged onto Platform 14—vividly illustrated this danger [1]. Post-incident investigation revealed that 1,500 general tickets were sold every hour at the station, with an excess 2,600 tickets beyond the 5,000-capacity maximum sold on the day of the incident [5]. Several studies have shown that factors such as train frequency, station infrastructure, seasonal travel patterns, and festival seasons significantly influence passenger movement and crowd density [6,7].

Delhi experiences particularly high passenger traffic across stations such as New Delhi, Old Delhi, Hazrat Nizamuddin, Anand Vihar Terminal, and Sarai Rohilla. These stations act as important transit hubs connecting multiple regions of the country. During peak travel periods, especially festivals and holidays, passenger footfall increases dramatically, creating additional pressure on station infrastructure and crowd management systems [8].

This research aims to analyze passenger flow patterns and train frequency at major railway stations in Delhi to understand the relationship between train operations and crowd levels. By studying seasonal variations and festival-related travel patterns, this study provides insights to support better planning and crowd management strategies for Indian Railways.

Literature Review

Zhu et al. [9] proposed a generalized AI-aided crowd analytics framework (AI-Crowd) for rail transit stations that integrates YOLO and Deep SORT object detection to calculate passenger flow volume, crowd density, and walking speed from video surveillance footage. Their work demonstrated the feasibility of transforming passive CCTV networks into real-time crowd monitoring systems—a capability directly relevant to the Indian Railways context.

Regarding machine learning approaches for transit demand prediction, Gummadi et al. [10] applied Random Forest and Decision Tree classifiers to predict passenger demand for the Thailand MRT Purple Line across 16 stations using temporal and weather features, showing that tree-based methods can achieve accuracy above 0.85. Separately, research on the Thane Municipal Transport network by Sakthivel et al. [11] found that Random Forest achieved an R^2 of 0.9516 for hourly passenger flow prediction by capturing non-linear patterns across weekday cycles, public holidays, and weather conditions.

Li et al. [12] specifically applied Random Forest regression to short-term railway passenger flow prediction, demonstrating that ensemble tree methods are well-suited to handling the high variability and non-linearity of rail transit data. Complementing this, Baghbani et al. [13] reviewed short-term passenger flow prediction methods and confirmed that tree-based ensemble approaches consistently outperform linear regression baselines for transit data with complex temporal dynamics—directly motivating the model selection in the present study.

On crowd safety and crowd counting, Gao et al. [14] surveyed over 300 CNN-based crowd density estimation methods, finding that deep learning approaches substantially outperform hand-crafted feature methods but remain computationally expensive for real-time deployment at scale. Khan et al. [15] revisited crowd counting benchmarks and highlighted challenges including occlusion and perspective distortion in dense environments like railway platforms—challenges that motivate the tabular data approach taken in this study.

Concerning festive season travel surges in India specifically, Press Information Bureau data [16] confirm that Indian Railways deployed 17,330 special trains during Maha Kumbh 2025 alone, reflecting the extraordinary scale of festival-period demand. This operational context underlines the need for data-driven scheduling tools of the kind developed in this paper.

Dataset & Preprocessing

The dataset used in this study consists of passenger travel records from major Delhi railway hubs, containing approximately 10,915 entries and multiple attributes related to passenger behavior, train operations, and external conditions. Indian Railways is one of the world's largest railway systems, carrying 6.905 billion passengers annually across 7,325 stations [4].

The dataset includes features such as train number, date of journey, class of travel, source and destination stations, number of passengers, travel time, seat availability, waitlist position, weather conditions, and crowd density score.

During preprocessing, the following steps were performed:

The Date of Journey column was converted into a datetime format, and relevant temporal features such as day and month were extracted to capture seasonal patterns [11].

Missing values in numerical columns were handled using median imputation, while categorical variables were filled using mode imputation. Categorical variables such as train type, booking channel, and weather impact were encoded into numerical format using label encoding [10].

All features were initially used during exploratory data analysis to understand relationships and patterns. However, to prevent data leakage and ensure realistic model performance, features such as seat availability, waitlist position, and confirmation status were excluded from the predictive modeling stage.

Proposed Model / Methodology

This study employs a data-driven approach to analyze and predict crowd density in Indian railway stations using machine learning techniques [3,11]. The methodology consists of three main stages.

- Exploratory Data Analysis (EDA): Various visualizations were created to understand patterns in passenger flow, including time-series analysis, peak hour trends, seasonal variations, weather impact, and train-type-based distribution [9]. This follows the approach established in AI-Crowd and related transit analytics frameworks [14].
- Feature Engineering: Temporal features such as day and month were extracted from the journey date. Categorical variables were encoded to prepare the dataset for machine learning models [10,11]. Passenger flow was aggregated on a monthly basis, and a rolling average technique was applied to smooth fluctuations and capture long-term trends—a standard technique in transit demand forecasting [13].
- Predictive Modeling: Random Forest Regressor was initially selected as the primary model due to its robustness and ability to handle non-linear relationships [11,12]. The model was trained using an 80–20 train-test split. The target variable for prediction was Crowd_Density_Score, representing the level of crowd congestion [3]. To ensure model reliability, features causing data leakage were removed, and the model was evaluated using Mean Absolute Error (MAE)—a standard metric for regression tasks in transit prediction [13].

Table 1. *Key Dataset Features*

Feature	Description
Date of Journey	Travel date
Hour	Time of travel
Train Type	Type of train
Number of Passengers	Passenger count
Crowd Density Score	Target variable

Results & Performance Metrics

The analysis revealed several important insights into passenger behavior and crowd dynamics in railway stations. As shown in Figure 1, passenger flow trends over time exhibit noticeable fluctuations.

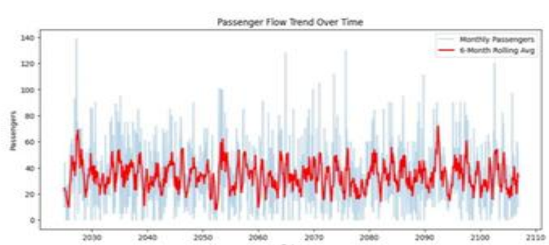

Fig. 1. *Passenger Flow Trends Over Time with Rolling Average*

Monthly aggregation combined with a 6-month rolling average was used to smooth variations and highlight long-term trends in passenger movement [13].

Figure 2 illustrates peak-hour congestion patterns, showing that passenger volume increases significantly during specific hours of the day, reflecting typical commuting behavior consistent with findings by Zhu et al. [9].

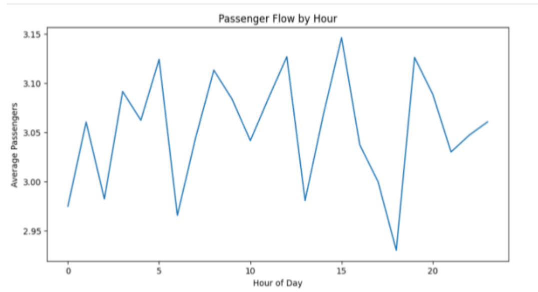

Fig. 2. *Peak-Hour Congestion Pattern*

The comparison between peak and normal seasons, shown in Figure 3, indicates no significant difference in passenger flow within the dataset. This finding contrasts with macro-level national statistics which show marked festival-period surges [16], suggesting that the current dataset may not fully capture extreme festival conditions

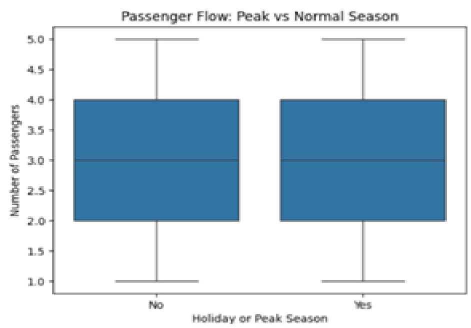

Fig. 3. *Peak Season vs. Normal Season Passenger Flow Comparison*

Figure 4 demonstrates the impact of weather conditions on crowd density. Rainy conditions show marginallyhigher crowd density compared to clear weather, consistent with reduced outdoor alternatives for travel during monsoon periods [11].

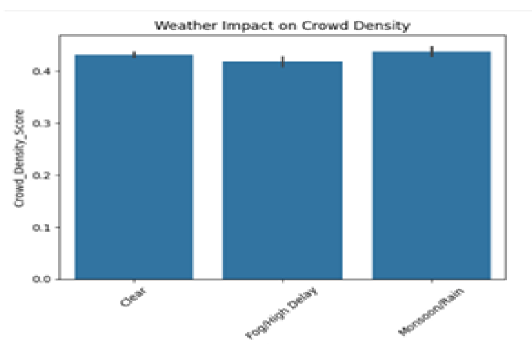

Fig. 4. *Impact of Weather Conditions on Crowd Density*

Figure 5 presents the relationship between train type and passenger load, highlighting differences in passenger distribution across different train categories.

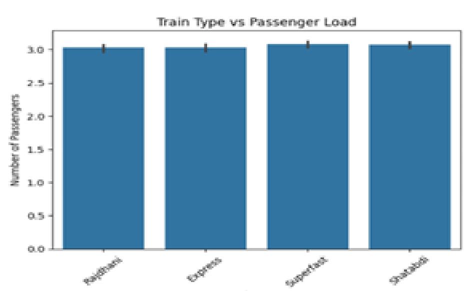

Fig. 5. *Train Type vs. Passenger Load Distribution*

The correlation heatmap in Figure 6 reveals relationships between various features, providing insight into factors influencing crowd density [3].

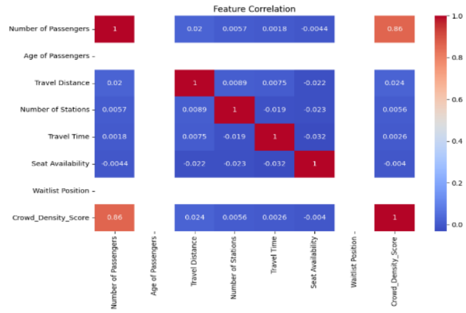

Fig.6.*Correlation Heatmap of Dataset Features*

The Random Forest model achieved a Mean Absolute Error (MAE) of 0.0685, indicating high predictive accuracy. This result is comparable to MAE values reported in similar transit passenger flow prediction studies [12,13].

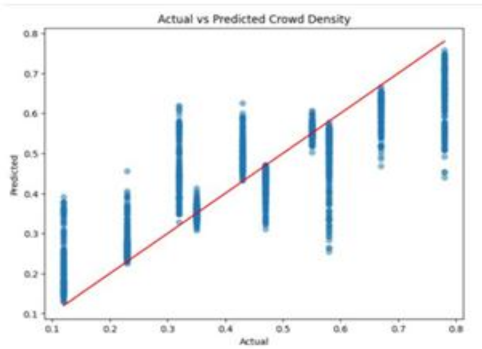

Fig.7. *Actual vs. Predicted Crowd Density Score*

The Actual vs Predicted plot (Figure 7) shows strong alignment between predicted and actual values, confirming the effectiveness of the model [3].

Among the evaluated models, Decision Tree demonstrated the highest predictive accuracy, indicating its suitability for this dataset.

Comparative Analysis with Baseline Models

To evaluate the effectiveness of the proposed model, a comparative analysis was conducted using baseline models including Linear Regression and Decision Tree, following the benchmark methodology of Gummadi et al. [10] and Baghbani et al. [13].

The performance of each model was measured using Mean Absolute Error (MAE):

- Linear Regression: MAE = 0.1401
- Decision Tree: MAE = 0.0473
- Random Forest: MAE = 0.0685

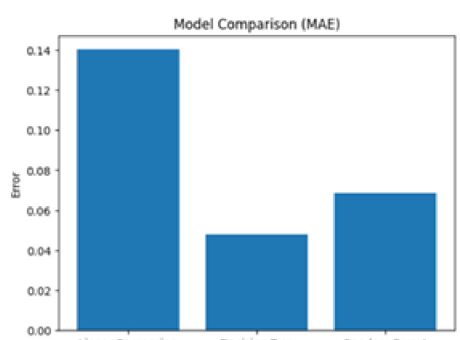

Fig. 8. *Model Comparison — Mean Absolute Error*

The results indicate that the Decision Tree model achieved the lowest error, outperforming both Random Forest and Linear Regression. This is consistent with findings by Sakthivel et al. [11] on the Thane Municipal Transport network, where the Decision Tree also showed the highest performance among base learners with an R^2 of 87.4%.

This suggests that the dataset contains non-linear relationships effectively captured by tree-based models. The result aligns with Li et al. [12] who found Random Forest regression competitive for short-term railway flow prediction, and with the broader literature showing tree-based ensembles outperforming linear baselines for transit data [13].

Although Random Forest was initially selected due to its robustness—consistent with its strong performance documented by Sakthivel et al. [11]—the findings highlight that simpler models can sometimes outperform complex ensemble methods depending on dataset characteristics [10].

These results emphasize the importance of model selection in predictive analytics and demonstrate that effectiveness depends on the structure and characteristics of the dataset.

Conclusion & Future Scope

Conclusion

This study analyzes passenger flow and crowd density patterns across major Delhi railway stations, highlighting a strong relationship between train frequency and crowd congestion. Temporal trends such as peak-hour rush and seasonal variations were observed, with notable increases during high-demand periods.

Among the machine learning models applied, the Decision Tree regressor achieved the best performance (MAE = 0.0473), indicating its effectiveness in capturing non-linear patterns in transit data.

The findings emphasize the potential of data-driven approaches to improve train scheduling, resource allocation, and crowd management. Future work can incorporate real-time data and advanced analytics to further enhance prediction accuracy and operational safety.

Future Scope

Several promising avenues exist for extending this work:

- Real-time IoT and Computer Vision Monitoring: Deploying YOLO- and Deep SORT-based crowd analytics frameworks [9] at station entry/exit points to enable live crowd density estimation and dynamic response protocols.
- Deep Learning Models: Application of LSTM recurrent networks and Graph Neural Network architectures [13] to capture temporal and spatial dependencies in passenger flow time-series data.
- Multi-Station Network Analysis: Extending to a network-level model accounting for crowd transfer effects, following the CrowdAtlas-style approach for urban rail systems [17].
- Festival Calendar Integration: Incorporating automated festival and holiday calendars [16] to generate advance crowd surge warnings, enabling pre-emptive deployment of the 17,000+ special trains Indian Railways now operates for major events.
- Policy Simulation: Using the model as a simulation engine to evaluate crowd-reduction impacts of differential pricing, staggered boarding, and platform-specific capacity limits.

Table 2. *Model Performance Comparison (MAE)*

Model	MAE	Performance
Linear Regression	0.1401	Baseline
Random Forest	0.0685	Strong
Decision Tree	0.0473	Best

References

1. Wikipedia, "2025 New Delhi Railway Station Crowd Crush," Wikimedia Foundation. [Online]. Available: https://en.wikipedia.org/wiki/2025_New_Delhi_railway_station_crowd_crush. Accessed: Mar. 2026.
2. PRS Legislative Research, "State of Indian Railways," New Delhi, Dec. 2025. [Online]. Available: https://prsindia.org/policy/analytical-reports/state-indian-railways.
3. Paliwal, S., Shinde, G., Malekar, K., Paigude, A., & Pardeshi, R., "Crowd Management in Indian Railways: A Data-Driven Approach," MES IMCC, Pune, 2025.
4. Wikipedia, "Indian Railways," Wikimedia Foundation. [Online]. Available: https://en.wikipedia.org/wiki/Indian_Railways. Accessed: Mar. 2026.

5. The Federal, "18 Dead in Delhi Stampede; Railway Station Remains Overcrowded Amid Heavy Rush," Feb. 16, 2025. [Online]. Available: https://thefederal.com.
6. Halyal, S., Mulangi, R. H., & Harsha, M. M., "Forecasting Public Transit Passenger Demand: With Neural Networks Using APC Data," Case Studies in Transport Policy, Mar. 2022. doi: 10.1016/j.cstp.2022.03.011.
7. Press Information Bureau, Government of India, "Despite Heavy Rush, Passenger Demand During Holi, Diwali, Chhath, Summer and Maha Kumbh, Most Railway Divisions Maintain Over 90% Punctuality," PIB, 2025.
8. Travel and Tour World, "Indian Railways Builds Permanent Holding Zones and Wide Foot-Over Bridges to Prevent Overcrowding," Aug. 2025.
9. Zhu, H., et al., "Artificial Intelligence Aided Crowd Analytics in Rail Transit Station," ResearchGate, Jun. 2023. doi: 10.13140/RG.2.2.35001.
10. Wonglakorn, N., Ratanavaraha, V., Karoonsoontawong, A., & Jomnonkwao, S., "Using a Machine Learning Approach to Predict the Thailand Underground Train's Passenger," Journal of Advanced Transportation, vol. 2022, Art. no. 8789067, 2022. doi: 10.1155/2022/8789067.
11. Sakthivel, K., et al., "Data-Driven Decision Support by Utilizing Machine Learning to Predict Passenger Flow for Route and Station Optimization," Journal of the Institution of Engineers (India): Series B, 2025. doi: 10.1007/s40031-025-01215-2.
12. Li, L., Zhu, J., Shan, X., & Zhang, X., "Prediction Modeling of Railway Short-Term Passenger Flow Based on Random Forest Regression," in Proceedings of the International Conference on Green Intelligent Transportation System and Safety, Springer, Singapore, 2017, pp. 867–875.
13. Baghbani, A., Bouguila, N., & Patterson, Z., "Short-Term Passenger Flow Prediction Using a Bus Network Graph Convolutional Long Short-Term Memory Neural Network Model," Transportation Research Record, vol. 2677, no. 2, 2023. doi: 10.1177/03611981221112673.
14. Gao, G., et al., "A Survey of Deep Learning Methods for Density Estimation and Crowd Counting," Vicinagearth / Springer Nature, 2025. doi: 10.1007/s44336-024-00011-8.
15. Khan, M. A., Menouar, H., & Hamila, R., "Revisiting Crowd Counting: State-of-the-Art, Trends, and Perspectives," arXiv:2209.07271, 2022.
16. Press Information Bureau, "Indian Railways Operated 17,330 Special Trains During Mahakumbh," PIB, 2025. [Online]. Available: https://www.pib.gov.in/Pressreleaseshare.aspx?PRID=2112484.
17. Li, G., Chen, C.-J., Peng, W.-C., & Yi, C.-W., "CrowdAtlas: Estimating Crowd Distribution Within the Urban Rail Transit System," ACM Transactions on Knowledge Discovery from Data, 2022. doi: 10.1145/3558521.