Journal of Mathematics and Data Science

[ISSN 3005-544X]
Volume: 04 Issue: 01 | February-2026

JMDS

Predicting Future Pedestrian Movement Patterns: An LSTM Approach Using Historical Footfall Data from Dublin

Authors

Mercy Jemursoi Koech [(1)]; Nasim Sobhani [(2)]

Main author's email: mercyjemursoi@gmail.com

(1.2) Griffith College Dublin, Ireland.

Cite this article in APA

Koech, M. J., & Sobhani, N. (2026). Predicting future pedestrian movement patterns: An LSTM approach using historical footfall data from Dublin. *Journal of mathematics and data science, 4*(1), 15-22. https://doi.org/10.51317/jmds.v4i1.924

OPEN ACCESS A publication of Editon Consortium Publishing (online)

Article history

Received: 2025-12-15

Accepted: 2026-01-20

Published: 2026-02-27

Scan this QR to read the paper online

Copyright: ©2026 by the author(s). This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0).

Abstract

Accurate forecasting of pedestrian footfall is essential for understanding temporal patterns in urban mobility. Existing statistical and machine learning approaches, like regression-based models, struggle to capture the nonlinear temporal dependencies and complex seasonal patterns common in urban pedestrian datasets, hence creating a need for more robust sequence learning methods. To address this limitation, this study implements and evaluates a Long Short-Term Memory (LSTM) neural network to predict daily pedestrian counts across Dublin City Centre using three years of sensor data (2022–2024) collected via PYRO-BOX counters. Hourly data were aggregated to daily totals, pre-processed to handle missing values, and assessed for stationarity. A 30-day sliding window approach was applied to construct sequential input-output pairs, and the dataset was partitioned into 80 per cent training and 20 per cent testing subsets. The LSTM model was trained with dropout regularisation and early stopping to prevent overfitting. Predictive performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2). The model demonstrated a strong predictive performance, with an MAE of 26,069.72, an MAPE of 8.82 per cent, and an R^2 of 0.7778, capturing both seasonal trends and complex temporal dependencies. A 30-day forecast starting January 1, 2025, closely aligned with historical footfall data, demonstrating the model's ability to project future values from historical patterns. These findings contribute both theoretically and practically by demonstrating the effectiveness of deep learning based temporal modelling for urban pedestrian dynamics and by providing an evidence-based forecasting framework that can support urban planning, transport policy decisions, such as crowd management and event planning.

Keywords: LSTM, pedestrian footfall forecasting, time series prediction, urban mobility.

1.0 INTRODUCTION

Urban pedestrian mobility provides essential insight into the patterns and dynamics of human movement within city environments. Reliable forecasting of pedestrian footfall depends on accurate historical data and advanced modelling techniques. The increasing availability of high-frequency sensor data enables predictive modelling of these movement patterns, offering a pathway beyond descriptive analysis toward forward-looking forecasts.

Dublin City Centre is a dynamic urban area with high variability in pedestrian activity across commercial, transport, and recreational streets. Footfall data were captured using PYRO-BOX passive infrared counters, which record hourly pedestrian counts and transmit validated data to centralised systems (Traffic Technology International, n.d.). Infrared pedestrian counting technologies have been empirically validated in academic research, demonstrating strong agreement with manual counts in urban field settings (Ryan & Benton, 2023). Aggregating these hourly counts into daily totals allows for the modelling of broader temporal trends while maintaining the resolution necessary to detect seasonal and weekly patterns.

Pedestrian movement exhibits temporal dependencies influenced by seasonality, holidays, infrastructure disruptions, and other temporal factors. Traditional time series approaches, including ARIMA and Holt–Winters methods, have been widely applied to model such patterns (Box & Jenkins, 1976; Bermudez et al., 2007). However, these approaches may be limited in handling irregular, non-linear variations and structural breaks common in real-world urban footfall data. Recent research highlights the increasing adoption of deep learning architectures, including hybrid LSTM-based frameworks for modelling complex urban mobility patterns and pedestrian dynamics (Fu et al., 2023; Zhang et al., 2025), reinforcing the relevance of advanced neural approaches for footfall forecasting. Contemporary methods, such as Prophet, improve flexibility by incorporating seasonal and holiday effects, yet deep learning models have increasingly demonstrated superior capacity to capture complex temporal dependencies.

Long Short-Term Memory (LSTM) networks, designed to address long-term dependency issues in sequential data, are particularly suited to forecasting pedestrian footfall (Hochreiter & Schmidhuber, 1997; Lindemann et al., 2021). By learning temporal representations directly from historical sequences, LSTMs can adapt to irregular and evolving patterns without relying on predefined functional forms. Despite growing applications of deep learning in mobility forecasting, several gaps remain in the literature. Limited evidence of LSTM-based pedestrian forecasting applied to Irish urban environments where mobility patterns differ due to local socio-economic and spatial characteristics. Prior studies often focus on hourly or single-location predictions with less attention given to modelling daily aggregated pedestrian counts that support policy and planning decisions. Few studies explore a multi-street modelling framework that produces both street-level forecasts and aggregate city-scale projections from independently trained models (Liu et al., 2021; Quan et al., 2021).

This study addresses these gaps by implementing an LSTM-based framework to forecast daily pedestrian footfall across multiple streets in Dublin City Centre. Each street is modelled independently, with results aggregated to produce city-level projections. The approach demonstrates the application of deep learning for generating accurate short- and medium-term forecasts based solely on historical footfall data, directly addressing the study's objective of predicting future pedestrian movement.

2.0 LITERATURE REVIEW

Pedestrian forecasting is an important aspect of understanding urban mobility patterns and is increasingly supported by high-frequency sensor networks. In modern cities, such sensor systems, including infrared counters and mobile-based data collection, provide large-scale datasets that enable predictive modelling of human movement (Cecaj et al., 2021; Komar & James, 2024). Beyond purely technical applications, pedestrian forecasting also contributes to broader smart city planning and urban analytics by informing infrastructure design, crowd management and mobility governance decisions. Insights derived from pedestrian data can support policy interventions related to public space utilisation and transport integration, highlighting the interdisciplinary relevance of accurate footfall prediction. Despite these technological advances, pedestrian flows remain influenced by nonlinear and context-specific factors such as weather, public events, and local disruptions. Accurate prediction of pedestrian volumes is therefore essential for analysing and anticipating short-term movement patterns.

Early research in time series forecasting relied on traditional statistical approaches, such as ARIMA and Holt–Winters methods, due to their interpretability and theoretical grounding (Box & Jenkins, 1976; Bermudez et al., 2007). These models provided a foundation for forecasting structured temporal data, but they typically assume stable seasonality and linear relationships, limiting their ability to handle irregular patterns and structural changes common in urban pedestrian data. Methodological developments introduced seasonal extensions such as SARIMA to better model periodic behaviour (Milenkovic et al., 2018). Similarly, tools like Prophet incorporate seasonal and holiday effects but rely on predefined functional forms, which may reduce flexibility in complex urban settings (Rafferty, 2021). These limitations have motivated a shift toward more flexible modelling approaches capable of learning nonlinear dependencies directly from data.

Deep learning approaches, particularly Long Short-Term Memory (LSTM) networks, offer a powerful alternative for capturing complex temporal dependencies in sequential data (Hochreiter & Schmidhuber, 1997). LSTM's gated architecture allows the network to selectively retain information from past observations, making it well-suited to learn patterns from historical footfall data. Studies applying LSTM to pedestrian and mobility forecasting have reported improved predictive performance over classical methods, particularly in contexts involving high-frequency, nonlinear data (Cecaj et al., 2021; Murcio & Wang, 2025). Research in transportation and other dynamic systems further demonstrates LSTM's ability to model non-stationary sequences (Xu et al., 2024).

Reliable forecasting requires robust validation, often using train-test splits and evaluation metrics such as RMSE, MAE, and MAPE to assess model accuracy (Kreinovich & Kosheleva, 2020; Hyndman & Koehler, 2006). Reviews of neural network applications also note that mainstream LSTM-based approaches achieve strong accuracy for short-term prediction but are rarely tested in longer-horizon urban mobility forecasting frameworks, highlighting a gap in practical deployment for planning purposes (Huang et al., 2024). This indicates a gap between methodological performance evaluation and practical forward forecasting applications for urban mobility planning. This study addresses that gap by applying an LSTM-based framework to historical daily footfall data in Dublin City Centre. By modelling each street individually and using past sequences to predict future counts, the research demonstrates the practical application of LSTM networks for accurate short-term pedestrian forecasting.

3.0 METHODOLOGY

This study uses a quantitative research design using a deep learning approach to forecast pedestrian footfall patterns. Secondary data were sourced from Dublin City Council and the National Transport Authority, consisting of hourly pedestrian counts recorded by PYRO-Box infrared sensors across 15 streets in Dublin City Centre from January 2022 to December 2024. All streets with consistent data records after preprocessing were included in the analysis

Data preparation involved cleaning, removal of redundant variables, forward-fill imputation of missing values, and aggregation of hourly counts into daily totals. The final dataset comprised daily footfall observations across the study period. Prior to modelling, data were normalised using Min-Max scaling to improve computational stability. A sliding window technique with a 30-day look-back period was used, based on both empirical findings in urban mobility forecasting and the typical duration over which pedestrian patterns exhibit weekly and monthly cycles. This window size allows the model to capture short-term fluctuations as well as broader monthly patterns, providing sufficient context for accurate day-ahead predictions.

The dataset was partitioned chronologically into training (80%) and testing (20%) subsets. A Long Short-Term Memory (LSTM) neural network was implemented using a multi-layer architecture with dropout regularisation to mitigate overfitting. The model was developed in Python 3.11 using Tensor Flow 2.14 and Keras 2.14 on a workstation equipped with an NVIDIA RTX A2000 with integrated RAMDAC. The Adam optimiser was used to train the model, and the Mean Squared Error loss function, with early stopping applied to prevent performance degradation. Forecast accuracy was assessed using MAE, RMSE, and MAPE on the test set.

Quantitative Performance Evaluation

The model was evaluated using a combination of absolute and relative error metrics, as well as a variance explaining score. Results are summarised in table 1.

Table 1: Evaluation Metrics for LSTM Model (Combined Train and Test Sets)

Metric	Value
Mean Absolute Error (MAE)	26,069.72
Root Mean Squared Error (RMSE)	35,285.11
Mean Absolute Percentage Error (MAPE)	8.82%
R-squared (R^2)	0.7778

DISCUSSION

Mean Absolute Error (MAE): 26,069.72

This shows that, on average, the model's predictions deviate from exact footfall counts by around 26,070. Considering the footfall range (approximately 56,000 to 520,000), this level of error is moderate and acceptable for prediction using time series.

Root Mean Squared Error (RMSE): 35,285.11

The RMSE penalises larger deviations more heavily than MAE. A value of roughly 35,285 implies the model occasionally experiences higher errors during volatile or outlier periods, yet remains effective for general trend tracking.

Mean Absolute Percentage Error (MAPE): (8.82%)

A MAPE under 10 per cent is considered strong in forecasting contexts as reported in prior studies on time series prediction (Hyndman & Athanasopoulos, 2018). This value reflects that predictions of the model deviate from actual values by less than 9 per cent on average, demonstrating high relative accuracy across varying scales.

R-Squared (R2): 0.7778

The coefficient of determination reveals that the model explains around 77.8 per cent of the variation in foot traffic. While this suggests substantial explanatory power, it does not definitively confirm a strong fit. Compared with similar pedestrian forecasting studies, this level of R-squared is within the range typically observed, though improvements may be possible through the inclusion of exogenous variables or more advanced model architectures.

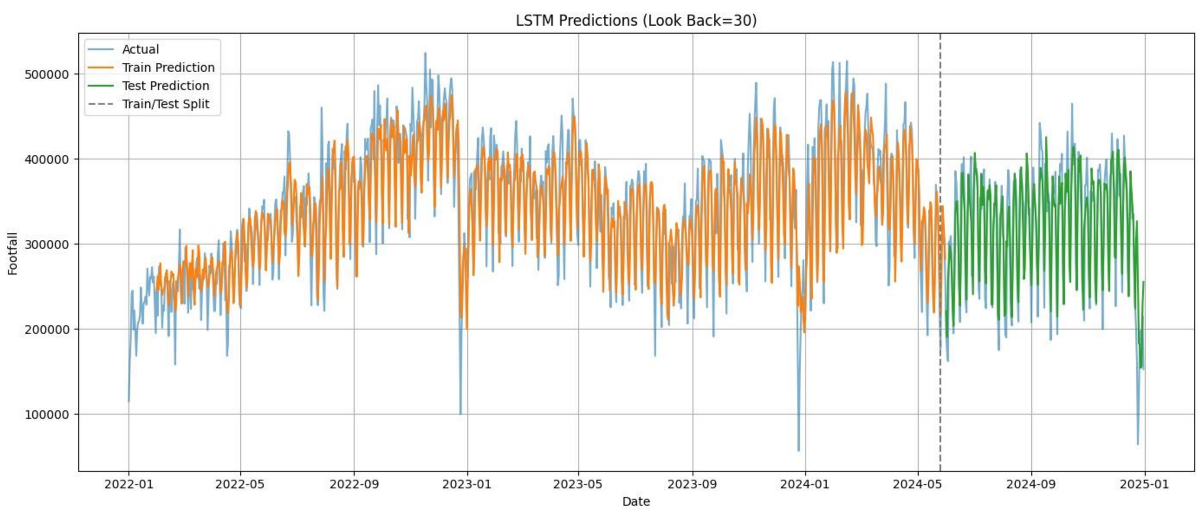

Figure 1: LSTM Predictions vs Actual Daily Footfall (Look Back = 30)

As illustrated in Figure 1, the LSTM model effectively tracks both seasonal trends and the overall trend present in footfall data. The blue line shows the actual footfall across the entire timeline. The orange line corresponds to predictions during the training period, closely following actual values, especially in capturing regular fluctuations. The green line represents predictions on the testing set, where the model continues to demonstrate good alignment with true values, despite minor deviations during high-variance periods. A vertical dashed line marks the train/test split, highlighting that the model's performance remains stable even on unobserved data. Temporal alignment and preservation of seasonal cycles—particularly weekly patterns—are evident throughout the forecast, further supporting the model's robustness and effectiveness.

Forecasting Future Footfall Using the LSTM Model

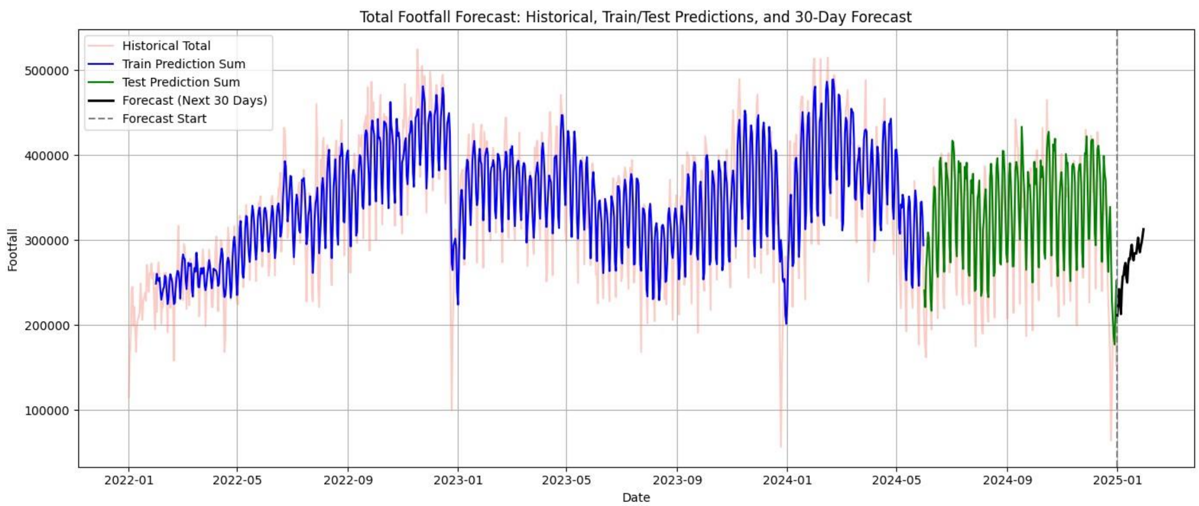

Figure 2: Footfall Forecast with Train/Test Predictions and 30-Day Future Outlook Using LSTM

In fulfilment of the study's key objectives—to forecast footfall for the subsequent 30 days—the LSTM model was employed due to its outstanding results on every evaluation metric. Figure 2 displays the historical footfall data, the model's predictions on the training and test sets, and a 30-day forecast starting from January 1, 2025.

The forecast section, highlighted in black, depicts the model's projected footfall values over the next 30 days. Prior sections of the graph incorporate the performance of the model on test and training data, which closely match real observations, confirming the model's correctness and ability for generalisation. To forecast the next 30 days, a recursive approach was used. The final observed sequence for each location was used iteratively to generate future values one step at a time. Individual forecasts are then aggregated to obtain an estimate of total footfall across all monitored streets. The forecasted period continues to reflect the established seasonal and periodic patterns observed in the historical data, indicating that the LSTM model effectively captured and extrapolated these underlying trends.

Notably, the transition into the future forecast period is clearly marked by a vertical dashed line, separating known values from model predictions. The predicted values do not exhibit abrupt shifts, suggesting the robustness and dependability of the model. This prediction is a useful instrument for anticipating short-term footfall, with potential applications in operational planning, resource allocation, and strategic decision-making for businesses or public venues that rely on accurate foot traffic predictions. All results were saved in a structured CSV file format to facilitate future usage in analytical exploration. The coherence and stability of the forecast reinforce the LSTM model's suitability for real-world time series forecasting tasks and demonstrate its potential for deployment in environments where accurate prediction is essential.

5.0 CONCLUSION AND RECOMMENDATIONS

Conclusion: This study demonstrated the effectiveness of a Long Short-Term Memory (LSTM) neural network in forecasting pedestrian footfall in Dublin City Centre. The LSTM model achieved strong predictive performance, with an R^2 value of 0.7778, indicating its ability to capture long-term dependencies and nonlinear temporal patterns in pedestrian movement. The results affirm the suitability of deep learning architectures for complex urban forecasting tasks where traditional assumptions of linearity and fixed seasonality may not hold.

Recommendations: Further research may explore the use of advanced neural architectures that have demonstrated potential in time series forecasting by effectively modelling temporal relationships and irregularities. Real-time forecasting systems supported by continuous data ingestion and adaptive model updating could improve the responsiveness of urban infrastructure and commercial operations to changing pedestrian dynamics.

6.0 REFERENCES

1. Bermudez, J. D., Segura, J. V., & Vercher, E. (2007). Holt–Winters forecasting: An alternative formulation applied to UK air passenger data. *Journal of Applied Statistics, 34*(9), 1075–1090.
2. Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (Revised ed.). Holden-Day.
3. Cecaj, A., Lippi, M., Mamei, M., & Zambonelli, F. (2021). Sensing and forecasting crowd distribution in smart cities: Potentials and approaches. *IoT, 2*(1), 33–49.
4. Fu, X., Wu, M., Ponnarasu, S., & Zhang, L. (2023). A hybrid deep learning approach for real-time estimation of passenger traffic flow in urban railway systems. *Buildings, 13*(6), 1514. https://doi.org/10.3390/buildings13061514
5. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.
6. Huang, S., Wei, R., Lian, L., Lo, S., & Lu, S. (2024). Review of the application of neural network approaches in pedestrian dynamics studies. *Heliyon, 10*(10), e30659. https://doi.org/10.1016/j.heliyon.2024.e30659
7. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts. https://otexts.com/fpp2/
8. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688.
9. Komar, T., & James, P. (2024). Orchestrating urban footfall prediction: Leveraging AI and batch-oriented workflow for smart city application. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48*, 119–124.
10. Kreinovich, V., & Kosheleva, O. (2020). Why 70/30 or 80/20 relation between training and testing sets (Technical Report UTEP-CS-20-20). University of Texas at El Paso, UTEP ScholarWorks.
11. Lindemann, B., Müller, T., Vietz, H., Jazdi, N., & Weyrich, M. (2021). A survey on long short-term memory networks for time series prediction. *Procedia CIRP, 99*, 650–655.
12. Liu, M., Li, L., Li, Q., Bai, Y., & Hu, C. (2021). Pedestrian flow prediction in open public places using a graph convolutional network. *ISPRS International Journal of Geo-Information, 10*(7), 455. https://doi.org/10.3390/ijgi10070455

13. Murcio, R., & Wang, Y. (2025). Identify optimal pedestrian flow forecasting methods in Great Britain retail areas: A comparative study of time series forecasting on a footfall dataset. *ISPRS International Journal of Geo-Information, 14*(2), 50.

14. Rafferty, G. (2021). *Forecasting time series data with Facebook Prophet: Build, improve, and optimise time series forecasting models using the advanced forecasting tool*. Packt Publishing.

15. Ryan, D. J., & Benton, J. S. (2023). Using automated active infrared counters to estimate footfall on urban park footpaths: Behavioural stability and validity testing. *International Journal of Behavioural Nutrition and Physical Activity, 20*(1), 49. https://doi.org/10.1186/s12966-023-01438-w

16. Traffic Technology International. (n.d.). Pyro-box pedestrian traffic counter. Retrieved May 14, 2024, from https://staging.traffictechnology.co.uk/blog/portfolio/pyro-box-pedestrian-traffic-counter/

17. Xu, J., Wang, Z., Li, X., Li, Z., & Li, Z. (2024). Prediction of daily climate using a long short-term memory (LSTM) model. *International Journal of Innovative Science and Research Technology*, 83–90.

18. Zhang, H., Deng, J., Xu, Y., Deng, Y., & Lin, J.-R. (2025). An adaptive pedestrian flow prediction model based on first-order differential error adjustment and a hidden Markov model. *Buildings, 15*(6), 902. https://doi.org/10.3390/buildings15060902