International Journal of Geo-Information

MDPI

Article

Identify Optimal Pedestrian Flow Forecasting Methods in Great Britain Retail Areas: A Comparative Study of Time Series Forecasting on a Footfall Dataset

Roberto Murcio [1,2,*,†] and Yujue Wang [1,†]

[1] School of Social Sciences, Birkbeck, London University, London WC1E 7HX, UK; y.wang@ucl.ac.ucl
[2] Centre for Advanced Spatial Analysis, University College London, London WC1E 6BT, UK
* Correspondence: r.murciovillanueva@bbk.ac.uk
† These authors contributed equally to this work.

Abstract: The UK retail landscape has undergone significant changes over the past decade, driven by factors such as the rise of online shopping, economic downturns, and, more recently, the COVID-19 pandemic. Accurately measuring pedestrian flows in retail areas with high spatial and temporal resolution is essential for selecting the most appropriate forecasting model for different retail locations. However, several studies have adopted a one-size-fits-all approach, overlooking important local characteristics that are only occasionally captured by the best global model. In this work, using data generated by the SmartStreetSensor project, a large network of sensors installed across UK cities that collect Wi-Fi probe requests generated by mobile devices, we examine the optimal forecasting method to predict pedestrian footfall in various retail areas across Great Britain. After assessing six representative time series forecasting models, our results show that the LSTM model outperforms traditional methods in most areas. However, pedestrian counts at certain locations with specific spatial characteristics are better forecasted by other algorithms.

Keywords: forecasting; human mobility; footfall

Academic Editors: Wolfgang Kainz and Godwin Yeboah

Received: 26 October 2024
Revised: 2 January 2025
Accepted: 22 January 2025
Published: 27 January 2025

Citation: Murcio, R.; Wang, Y. Identify Optimal Pedestrian Flow Forecasting Methods in Great Britain Retail Areas: A Comparative Study of Time Series Forecasting on a Footfall Dataset. *ISPRS Int. J. Geo-Inf.* **2025**, *14*, 50. https://doi.org/10.3390/ijgi14020050

1. Introduction

Functional characteristics of urban areas and local activity patterns are some of the main factors that dictate the suitability of a site for a store location on very granular scales [1,2]. The development of retailing also depends largely on changes in pedestrian flow, which cannot be understood simply as exploring the regional footfall pattern. Footfall (FF) plays an important role here, serving as a proxy for the activity and vitality of retail areas [3].

High street footfall can easily fall despite the population growth, and this was seen to be the case during the 2008 economic downturn, as Genecon reports a 10% footfall decline in the UK high streets, excluding Central London (2008–2011) [4]. In consequence, footfall, as such, can be used as a more accurate and up-to-date measure of the potential customer base around retail areas on a daily or weekly basis, helping retailers plan business activities, such as scheduling the opening/close time and personnel, production, and transportation, or formulate long-term business plans, such as risk management [5,6]. Footfall research is therefor a prerequisite for any business.

Footfall prediction can also help the government take emergency measures in advance to prevent overcrowding and reduce crowd risk. For example, overcrowding in Oxford Street, the world's biggest high street, led to temporary underground station closures

113 times in 2015 [7]. In [8,9], the authors explore different strategies for bottlenecks in underground stations. Especially in the context of the global COVID-19 pandemic, the risk of overcrowding has expanded to more than physical injuries.

However, the existing research on crowd flow prediction mainly focuses on short-term crowd behaviour modelling at the micro level. For example, Zhang et al. [10] proposed a cellular automata model for simulating pedestrian flow. Liu et al. [11] introduced a graph convolutional network (GCN) to predict the crowd flow in a walking street. Therefore, from a theoretical point of view, there is an important gap in the prediction of footfall in retail areas on a large scale due to the limitation of data acquisition in the past.

In this work, using data from a network of Wi-Fi sensors at retail locations across Great Britain, we applied different time series forecasting methods to analyse the pedestrian flow in hundreds of retail areas. We identified the behaviour of each model concerning specific spatiotemporal locations to provide a comprehensive guide for users to choose the optimal models for predicting footfall in retail areas.

The research question addressed here is as follows: What is the optimal pedestrian flow forecasting approach for retail areas in Great Britain? In an attempt to answer the research question, this paper also aims to address the following objectives:

1. What are the most commonly used and best-performing time series prediction models for pedestrian flow forecasting? How do we compare the performance of different forecasting models?
2. Does the best forecasting model differ for different retail areas in GB?
3. Is the difference among the optimal footfall prediction approaches related to some spatial characteristics of different locations?

Based on our objectives, the main contributions of this study are (1) to provide a framework for choosing an optimal model for forecasting footfall in Great Britain retail areas and (2) prove that the differences observed between optimal models can be attribute to the urban and social environment surrounding particular retail locations.

The rest of this paper is structured as follows: First, in the methods section, we discuss, analyse, and assess the footfall dataset compared with other human mobility datasets freely available. Then, we discuss the forecasting methods employed in this work, providing different metrics to evaluate their accuracy. Then, in the discussion section, we evaluate the statistical and spatial distribution of the footfall counts related to their spatial location and their influence on the accuracy of the different forecasting methods. Finally, we elaborate on conclusions, recommendations, and possible paths to extend this research.

This study can be used as a reference for future research on pedestrian flow prediction in GB retail areas.

2. Literature Review

In many fields of study, data are continuously collected over time. Technological advances in database systems, neural networks, and cloud computing have promoted the conversion of temporal data to useful information to support the decision-making process. Traditionally, forecasting models focused on AR and MA have been considered among the most accurate forecasting approaches in over a half-century [12]. More recently, and acknowledging the shortcomings of such parametric models, several studies have introduced the non-parametric approaches based on ML algorithms on the forecasting of temporal datasets [13,14] since they do not presuppose the nature of time series distribution. Empirical research has shown that ML models perform quite well on temporal data forecasting and usually outperform the statistical approaches [15–18]. Parmezan et al. applied various indicators, including MSE, Theil's U coefficient, POCID, and a multi-criteria performance measure, to evaluate the prediction performance of statistical and ML models. They demon-

strated that the SARIMA model is the only statistical time series prediction model that performs better than some ML models on several datasets [19]. Another comparative study by Gautam and Singh [20] indicates that parametric models are competitive with ML models and perform better for some datasets evaluated by symmetric mean absolute percentage error (sMAPE). Finally, Han et al. [21] conducted a comparative study on the prediction performance of various models on both generated and real-world data, adopting traditional ML models as benchmarks and MAE together with MAPE as accuracy indicators.

2.1. Classic Forecasting Methods

The approaches for time series forecasting [22,23] rely on the basic idea that historical data contains intrinsic patterns which convey the key information for the future description of the phenomenon investigated. Nevertheless, these patterns are usually non-trivial to identify, and their discovery becomes one of the primary goals of time series processing: the features the patterns found will repeat and which sort of changes they might suffer over time [24]. This idea led to the decomposition of the time series components, namely trend, seasonality, cyclic and residuals, which form the basis of many classic time series decomposition methods. Time series forecasting methods can be categorized by whether to use parameters as parametric or non-parametric methods [20], or by methodology as statistical and machine learning (ML) models [25].

The classic methods employed in this work are as follows:

- Seasonal Naïve [26];
- Moving Average [27];
- Holt–Winters [28,29];
- SARIMA Model [30].

It is worth mentioning that the literature about these methods is vast. In this work, we refer to [27] to develop these classical models.

2.2. Machine Learning Approaches

There is a close connection between the nature of time series forecasting and the regression analysis of the ML classification. The classic support vector machine [31], Bayesian network [32] and Gaussian process [33] have all achieved good results in time series forecasting. An early artificial neural network (ANN) is also used to obtain the long-term trend in time series. With the rise of neural networks, deep learning can also be regarded as an effective tool for forecasting time series.

Support Vector Machine (SVM), a machine learning method based on statistical learning theory, is mainly based on VC dimension theory and structural risk minimization principle [31]. It is also the first learning algorithm based on geometric distance.

A Bayesian network is essentially a directed acyclic graph, which uses probabilistic networks for uncertainty reasoning [32]. The use of Bayesian networks for time series forecasting is mainly to learn the Bayesian network structure for a given data set, and the directed arcs between nodes represent the direct dependencies between variables [34]. Using Bayesian networks for time series forecasting mainly aims to learn the Bayesian network structure for a given data set.

Traditional time series forecasting technology has limited ability to deal with high-dimensional big data and effectively express complex functions [35]. In addition, designing an effective machine learning system requires considerable data domain expertise. The Deep Learning models use multiple layers, learned from data rather than constructed by human engineers, to represent potential features at a higher and more abstract level. deep learning models have been successfully applied in many time series forecasting fields, including remote sensing [36] and multi-sensor analysis [37,38].

This work will use the Long-Short Term Memory (LSTM) recurrent network as our selected ML model. The LSTM overcomes the vanishing gradient of standard Recurrent Neural Networks (RNNs) and keeps the temporal information for a large number of timesteps, and has been successfully applied in time series analysis, classification and prediction [39–41]. We will discuss in detail the LSTM model in Section 3.6.

3. Materials and Methods

Figure 1 shows our workflow to analyse the FF data, define different FF forecasting model and relate them to the street type and output area classification around each location to understand why some methods performed better that others depending on the location. In the rest of this section, we elaborate on each one of these components.

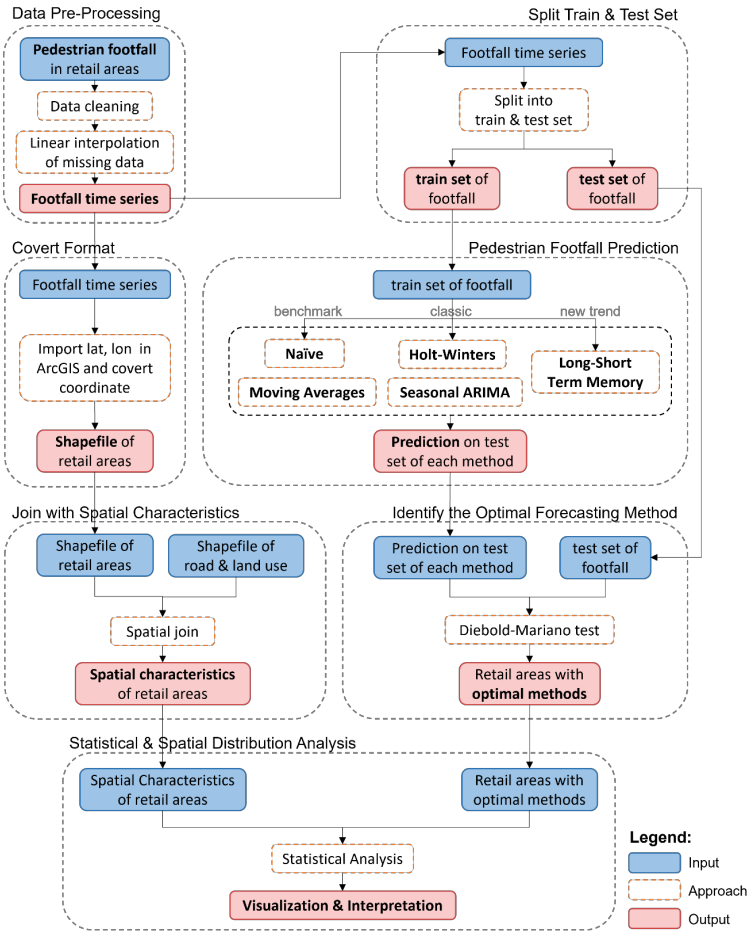

Figure 1. Proposed workflow to select the optimal footfall forecasting algorithm. Footfall derived from Wi-Fi signals.

3.1. Footfall Dataset

While retail geographical analysis has been covered in previous research [42,43], the analysis and prediction of pedestrian dynamics within large sample locations based on recorded activity patterns on a finer temporal scale is scarce. Research of micro-scale retail geography based on the demand side could only be conducted through manual surveying in the past, which required a costly and laborious process and without continuous data acquisition. However, these shortcomings have been addressed since the rapid development and wide-scale application of smartphones, Wi-Fi, GPS, and Bluetooth technologies. The Consumer Data Research Centre (cdrc.ac.uk), in collaboration with the Local Data Company (www.localdatacompany.com), set up the SmartStreetSensor project for collecting

and analysing Wi-Fi probe requests where a network of Wi-Fi sensors has been installed in retail establishments across the UK, including shopping centres, out-of-town retail parks and urban town centres [44].

Several studies have used these data for a wide range of applications [3,45–47]. Particularly, the research of Lugomer [3] demonstrated that the footfall patterns in retail areas exhibit different dynamic characteristics on different days of the week.

The footfall data comprise more than 1000 Wi-Fi sensors attached to retail areas across the UK, with the vast majority in London (Figure 2).

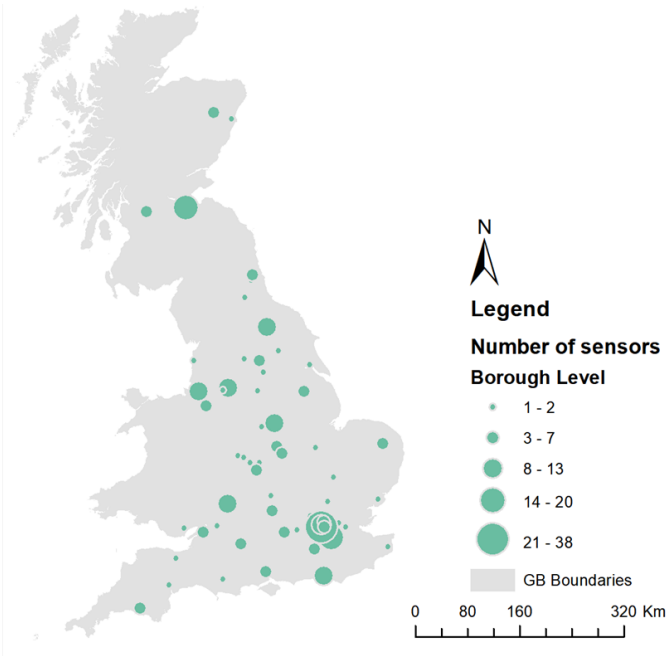

Figure 2. Spatial distribution of Wi-Fi sensors in Great Britain.

The project employs a proprietary sensor network that records mobile device signals to discover available Wi-Fi access points. Once captured, the data are processed into five-minute aggregated counts, and each record has the following structure: sensor, timestamp, footfall count, location, device, and address (latitude, longitude). We refer the reader to [48] for the full technical definition.

Figure 3 illustrates the time series of average footfall counts in Great Britain retail areas. The footfall shows a dramatic fluctuation since 2020 due to the COVID-19 epidemic. In addition, the data for some dates from September to October 2019 are missing due to technical issues in the data collection process.

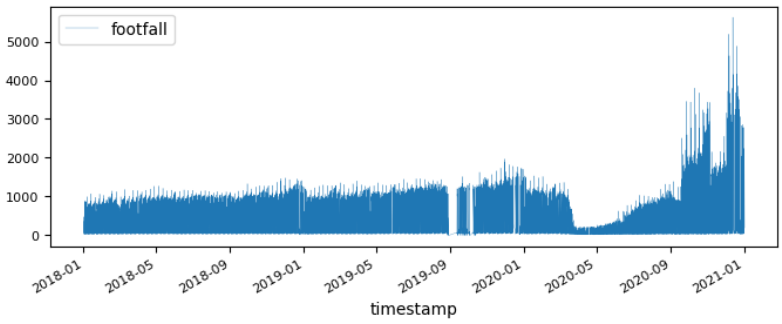

Figure 3. Average footfall counts every 5 min in GB retail areas.

3.2. Additional Datasets

In addition to the FF data, we use to other two datasets to associate the sensor's locations with their urban environment:

1. Open Road Data from Digimap, Ordnance Survey (digimap.edina.ac.uk). The data offer a high-level view of the road network with generalised geometry and network connectivity in Great Britain.
2. Output Areas 2011 Classification from the CDRC (data.cdrc.ac.uk/dataset/output-area-classification-2011). The 2011 Classification for Output Areas (2011 OAC) is a hierarchical geodemographic classification across the UK that identifies areas of the country with similar characteristics that contain the following breakdown of Output Areas into different groups and sub-groups.

After conducting the corresponding coordinate and projection conversion in the ArcGIS Desktop 10.8.2, we associated each sensor to its nearest road/Output Area centroid using a Euclidean distance.

3.3. Data Pre-Processing

Since it is hard to avoid the various technical problems of Wi-Fi sensors attached to each location during the data acquisition period, such as power outages or accidental damage, the footfall data are somehow inconsistent within most areas, which needs to be filtered and interpolated at the first stage. Moreover, the 5 min interval for the original aggregated footfall contains significant white noise interference. Therefore, the original footfall data were resampled to one-hour aggregated counts due to their observed stability (Figure 4) and because it has been suggested [49,50] that one-hour footfall predictions are more useful for retailers than any other short time period. With this aggregation, we avoid applying white-noise reduction techniques, such as a Kalman filter. After aggregating, the data still present null values for some hours, so we perform a simple linear interpolation procedure to have a complete dataset. We only interpolate values when no more than two hours were missed during extended business hours (6 a.m.–11 p.m.). Later in this section, we will discuss splitting the dataset into two groups: training and testing data. For now, it is worth mentioning that locations with more than 40% of missing data in the training set or more than 20% of missing data in the test set were removed. Finally, to avoid the influence of the COVID-19 pandemic and the missing data, the selected time frame for this study was from January 2018 to August 2019. After this data-cleaning procedure, 314 locations were used in this work.

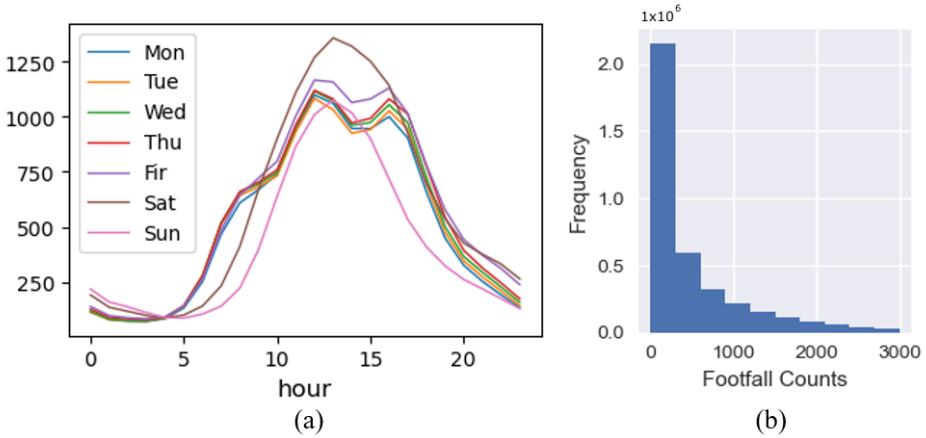

Figure 4. Hourly footfall distribution. (**a**) By day of the week. (**b**) Footfall counts—8×10^3 .

3.4. Footfall Data as Time Series

Figure 5 shows the FF data as a time series aggregated by different seasonalities. The sharp decay after 26 December is clear in all panels, while during Easter week, the fall in FF is only observable at precisely the weekly aggregation.

Figure 4 illustrates the average hourly distribution of footfall on a different day of the week, which shows a strong similarity in pattern between Monday to Thursday, compared with Friday, Saturday, and Sunday, which show a distinct distribution.

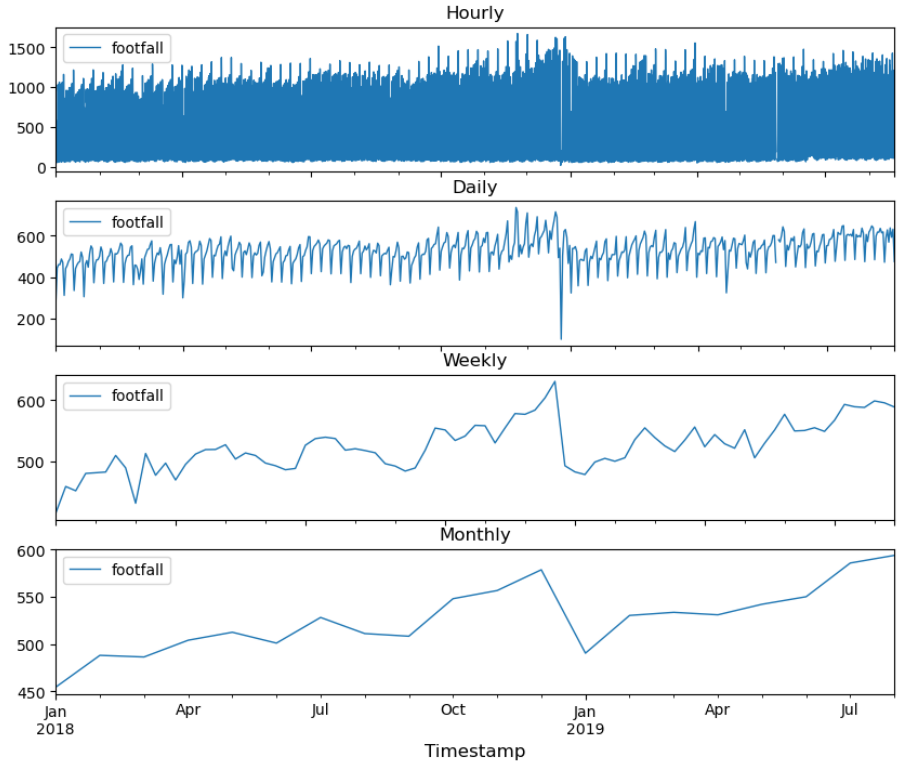

Figure 5. Footfall time series after cleansing.

3.4.1. Assessment

The FF signal's trend and seasonal components (Figure 5) depict the non-random nature of this data, and that is detecting the ups and downs in FF activity around different retail locations. To further assess these features, we compared our data with three external sources measuring human activity: (1) extraordinary events, (2) public transport figures, and (3) Google popular times.

Extraordinary Events

The Notting Hill Carnival is a massive yearly event in London that attracts thousands of visitors over three days. In 2018, the carnival occurred from Saturday, the 25th, to Monday, 27 August.

No sensor is installed around Notting Hill tube station. Still, there are two locations around High Street Kensington station, location 709 and location 957, which are popular locations for getting into the event. Figure 6 shows the change between an average Monday and Carnival Monday of location 709 in 2018. In the former, the typical morning/afternoon peaks at rush hour, plus a peek at lunchtime, are clear, while on Carnival day, there is only one peak at 2 p.m. and from 10 a.m. to 5 p.m., the volume of FF is in orders of more than 1500 people per hour. The data captured by these sensors successfully detect the change in FF around this area due to the unusual amount of activity and pedestrian flows.

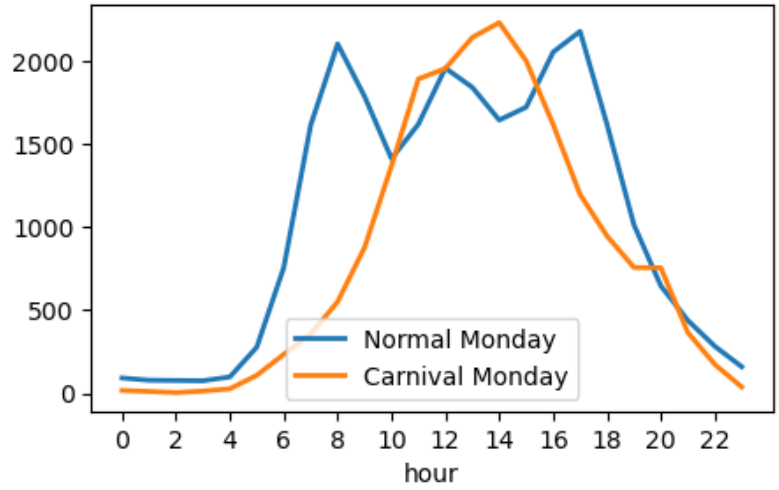

Figure 6. Hourly footfall distribution of location 709.

The second, easy-to-spot event is the decline in FF due to the COVID-19 pandemic. After the outbreak of the pandemic, public transport journeys were severely affected by the soaring number of cases; from February 2020 to May 2020 and from October 2020 to February 2021, there were two cliff-like drops in public traffic journeys, which coincided with two major lockdowns of London. By inspecting Figure 6, we can observe how it practically disappeared from the graph during the first major lockdown in the UK (March 2022).

Public Transport

We related the FF counts to the entry and exit figures for nearby underground stations to assess whether these sensors capture the activity trends around different areas. The data are open source, reported by the Transport for London (TfL) authority [51].

The annual counts for the London underground are collected and represent each day in autumn each year. Therefore, the footfall data for September, October and November 2018 are adopted for comparison. We conducted a simple Pearson correlation analysis of locations vs stations (Figure 7) over 88 retail areas where the distance location-station is less than one kilometre. The results indicate that the distribution of FF counts on different weekdays (Monday to Thursday, Friday, Saturday, and Sunday) is highly similar to the passenger counts of nearby tube stations, with a median correlation coefficient greater than 0.8.

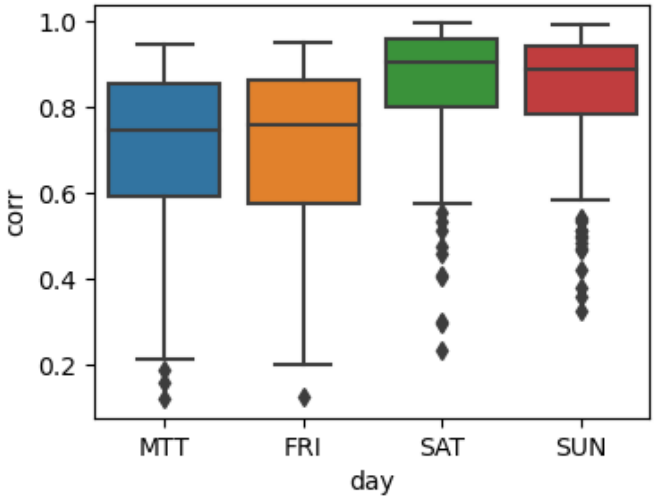

Figure 7. Pearson correlation between FF in retail areas and nearby underground flows on different days of the week. The label MTT represents Monday to Thursday measures.

In Appendix A, we showed an example of a particular popular station in London, Piccadilly Circus, to further show the value of the FF data. We also refer the reader to Appendix A for the Google Popular Times comparison.

3.5. Pedestrian Footfall Prediction

Before applying any forecasting methods, the standard approach is to split the dataset into training and test datasets. Our dataset spans over 84 weeks for the 314 studied locations. The training set was selected from 1 January 2018, to 28 July 2019 (81 weeks), and the validation set is defined from 29 July 2019 for three weeks.

3.6. Long Short-Term Memory Model (LSTM)

The LSTM recurrent network [39] is a special Recurrent Neural Network (RNN) designed to solve the traditional RNN problem when dealing with long-series data-designed gradient disappearing problems. By introducing a three-gating mechanism, a forget, input, and output gate, the flow of information is controlled through a single LSTM unit. This model has been successfully applied in time series analysis, classification and prediction [39–41] in both univariate and multivariate domains [52].

The standard LSTM, as described in [53], could be a state from the perspective of the Auto Regressive (AR) model:

$$x_{t+1} = d(x_t, x_{t-1}, x_{t-2}, ..., x_{t-N+1}), \tag{1}$$

As for the LSTM network a single timestep t, the i_t,f_t,u_t o_t and c_t, in this order, represent input gate, forget gate, update for cell state, output gate, and cell state. Two sets of weights are defined, namely $W = [W_i, W_f, W_u, W_o]$ and $U = [U_i, U_f, U_u, U_o]$. A random (bias) factor b is defined for each gate, which can be defined as follows:

1. Forget Gate f_t: The forget door step determines the time before memory in the current time step needs to be forgotten.

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f), \tag{2}$$

2. Input Gate: The input gate determines how much of the current input information will be kept in the memory cell.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i), \tag{3}$$

The updated formula of the memory content is as follows:

$$\tilde{C}_t = \tanh(W_C x_t + U_C h_{t-1} + b_C), \tag{4}$$

3. Memory cell state update: The state of the memory cell at the current time step is determined by both the results of the forget Gate and the input gate.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \cdot \tilde{C}_t, \tag{5}$$

4. Output (Output door Gate): The output door step determines the time to hide the status update, which is the final output of the LSTM hidden state.

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \tag{6}$$

The update formula for the hidden state is as follows:

$$h_t = o_t \cdot \tanh(C_t), \tag{7}$$

Together, these gates and the cell state form a complex memory system that allows LSTM to effectively process sequential data, making them well-suited for tasks like time series forecasting. In summary, the model determines which information from the previous time step is relevant and should be retained and which information can be discarded. Second, it learns new information from the current input and integrates it with the retained information. Finally, it outputs a portion of its updated state to the next time step. This entire process constitutes a single-time step in the LSTM.

3.7. Identifying the Optimal Forecasting Method

To statistically identify forecast accuracy equivalence for two sets of predictions, we applied the Diebold–Mariano (DM) Test [54]. With the assumption that the difference between the first list of predictions and the actual values is e_1 and the second list of forecasts and the actual value is e_2. The length of the time series is T. Then, d can be defined based on different criteria (crit), whereas in the measure of MSE,

$$d = {e_1}^2 - {e_2}^2 \tag{8}$$

The null hypothesis is $E[d] = 0$, and the test statistics follow the student-T distribution with degree of freedom $(T - 1)$.

The DM test is adopted to identify whether the forecasts of the two prediction methods have a significant distinction.

4. Results

4.1. Optimal Prediction Methods Statistics

The results shown in Table 1 confirm that the LSTM prediction model is the optimal method with the lowest RMSE for 83.76% of locations, where 76.81% of the predictions meet the DM test, which means the LSTM model outperforms other prediction methods on footfall forecasts within 64.33% of all locations with a statistical significance.

Table 1. Performance of different time series prediction models.

		Best Method		**Significance**	
Method	**RMSE**	**Counts**	**%**	**Counts**	**%**
Naive	254.24	5	1.59%	4	1.84%
MA	216.68	18	5.73%	3	1.38%
HW	254.24	15	4.78%	5	2.30%
SARIMA	221.61	13	4.14%	3	1.38%
LSTM	224.67	263	83.76%	202	93.10%
Total	–	314	100.00%	217	100.00%

The rest of the methods only account for a small proportion of the total, not exceeding 1.6%. In addition, Figure 8 indicates the MA approach does not perform well in predicting incomplete datasets.

To further study the prediction performance of other statistical time series prediction models, we counted the times each method performs best, excluding the LSTM model and those that satisfy the DM test.

Table 2 illustrates that MA, as the best prediction method with the lowest RMSE, has the most significant proportion among different locations, followed by SARIMA and HW, and the Naïve approach holds a minor proportion.

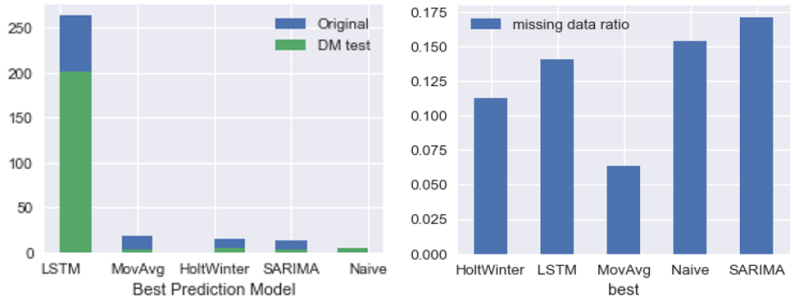

Figure 8. Footfall prediction results. The LSTM model emerge as the optimal for most of the locations with a high number scoring high in the DM test. In terms of missed data, we can observe how the MovAvg method performs poorly when many hourly FF counts are missed. (**a**) Optimal method counts. (**b**) Statistics of average missing data ratio.

Table 2. Performance of different time series prediction models without LSTM.

		Best Method		Significance	
Method	**RMSE**	**Counts**	**%**	**Counts**	**%**
Naive	254.24	26	8.28%	14	1.27%
MA	216.68	140	44.59%	73	0.96%
HW	221.61	74	23.57%	39	1.59%
SARIMA	224.67	74	23.57%	31	0.96%
Total	–	314	100.00%	157	100.00%

However, the number of the best prediction methods for different locations that meet the DM test only accounts for half of the total, meaning that half of the locations simultaneously have more than one optimal prediction method. As a result, different forecasting methods have no significant difference in the distribution of predicted pedestrian footfall in these retail areas. In addition, Figure 9 indicates these statistical approaches show little preference for datasets with different missing data ratios. Still, it should be noted that HW and MA locations show the smallest missing data ratio in both Figures 8 and 9.

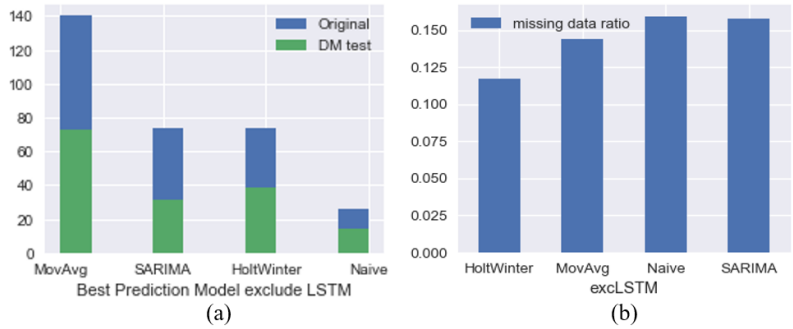

Figure 9. Footfall prediction results. (**a**) Optimal method counts, (**b**) statistics of average missing data ratio (without LSTM).

Pedestrian Footfall Distribution

Figure 10 shows the average footfall distribution of locations with different optimal methods by hourly, weekly, and monthly. Figure 10a indicates significant hourly distribution differences among these locations. The average hourly change is the most stable within locations where SARIMA is the optimal model. It also has the most negligible total flow, compared with locations with the best predicting method LSTM with the highest average

footfall, which can also be seen in the weekly and monthly figures. The peak of the average hourly flow of locations with the optimal method HW appears later in the day.

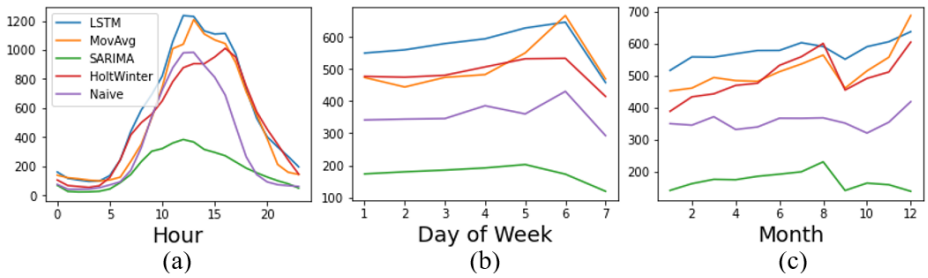

Figure 10. Average footfall distribution of locations with different optimal methods. We can observe how the different methods behave in the same way regardless of the time frame, confirming the robustness of each optimal method at their corresponding locations. (**a**) Hour. (**b**) Day of the week. (**c**) Month.

Figure 10b shows locations where MA is the optimal method have a more significant peak on Saturday in the average weekly footfall distribution. Also, the peak of locations with SARIMA appears on Friday, which is different from other locations. It can be seen from Figure 10c that locations with HW or MA as the best forecasting method show a more apparent seasonal trend in the average monthly footfall distribution, compared with Naïve, SARIMA and LSTM.

However, it should be noted that, except for locations where the optimal prediction method is LSTM, the number of other locations is relatively small, so the analysis of the results could be affected by some exceptional cases and might not be universal.

4.2. Spatial Characteristics Analysis

To further investigate the differences observed between optimal forecasting methods, we explored is such differences are related to the nearest road's type and aggregated social characteristics as defined in the Output Area Classification. After calculating the distance footfall sensor-road, we define a Random Forest Classifier (RFC) to assign the sensor into different classes. The RFC represents a family of models involved in building an ensemble of decision trees, where each tree is formed through a bootstrap sample of the data. Then, each tree node is split according to the best of a subset of randomly chosen predictors [55]:

RFC Algorithm

1. Select random data points from the training set;
2. Build decision trees for the selected data points;
3. Repeat steps 1 and 2;
4. For new data points, find the predictions of each decision tree;
5. DAssign the new data points to the most recurrent (popular) category.

The outputs from each tree are averaged to obtain an ensemble prediction of the target variable that can be used as classification tools [56].

The RFC model could help identify spatial factors highly correlated with the optimal prediction method of retail locations and illustrate the influence of different factors quantitatively.

4.2.1. Distance to Different Road Type

Figure 11 shows the average Euclidean distance from classified locations to different types of roads.

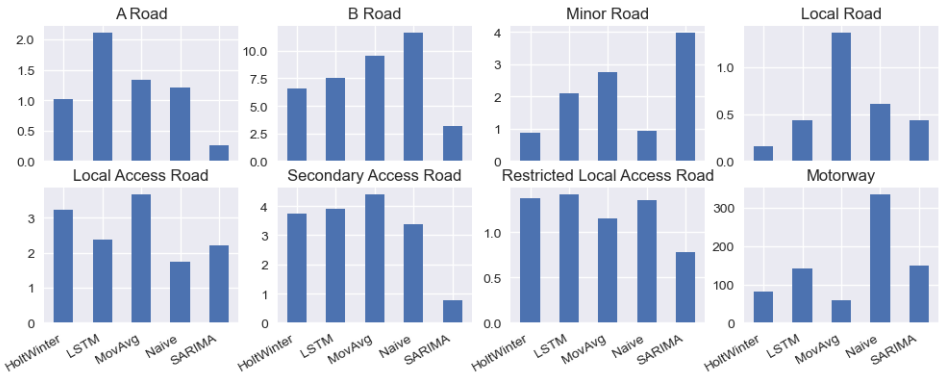

Figure 11. Average distance from different locations to roads. The top panel is A Road, B Road, and Minor Road; the below panel is Local Road, Secondary Road, and Restricted Access. The differences in optimal method vs. road type are evident and give some clues about the spatial influence that urban features have over forecasting.

Results indicate that locations with SARIMA, the optimal method, are closest to the main roads, including the A road and the B road, compared with the LSTM locations with the longest average Distance to the A road and the Naïve locations with the longest average Distance to the B road. In addition, The HW locations are nearest to the local roads. Also, the MA locations have the longest Distance to local roads and access roads, compared with the SARIMA locations close to the local roads and access roads. It can be seen from Figure 11 that Naïve locations have the longest average Distance to the motorway, compared with the MA locations that are closest to the motorway.

4.2.2. Output Area Classification (OAC)

Figure 12 illustrates the classification for different locations. Results show that the class Cosmopolitans (young adults, with a higher proportion of single adults and households without children with a comfortable style of life), accounts for the largest proportion in most locations except for Naïve locations. The LSTM locations cover a wide variety of classes.

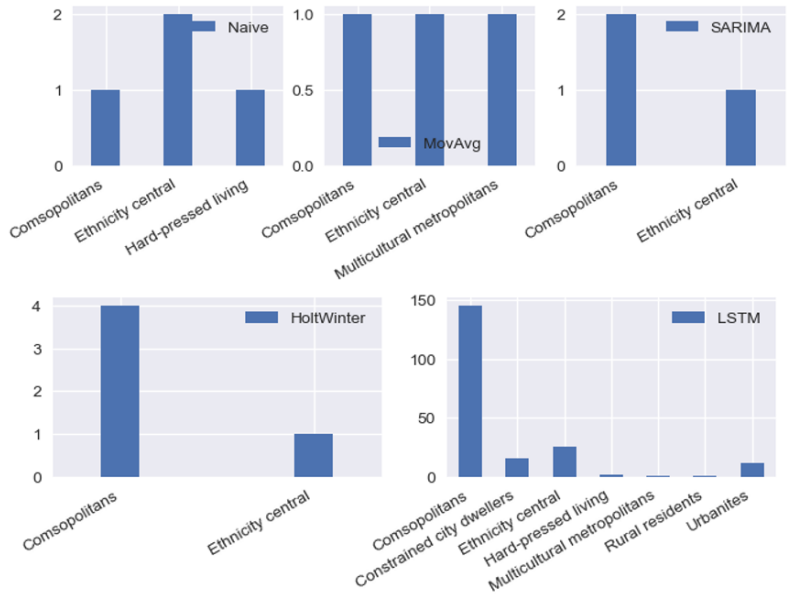

Figure 12. Land classification statistics on locations with different optimal prediction methods. Each panel shows a different forecasting methods and its relationship with the different classes defined in the OAC. The LSTM accounts for the most classes, only because this one is the optimal method at most locations.

These results could be driven by the fact that all locations are in retail areas, where Cosmopolitans are the predominant class. No particular marker indicates that the OAC class has some particular influence in the forecasting method, as the one observed for the road types.

4.2.3. Multi-Relationships with Spatial Features

Figure 13 shows the correlation matrix of continuous variables. No significant correlation among variables was observed.

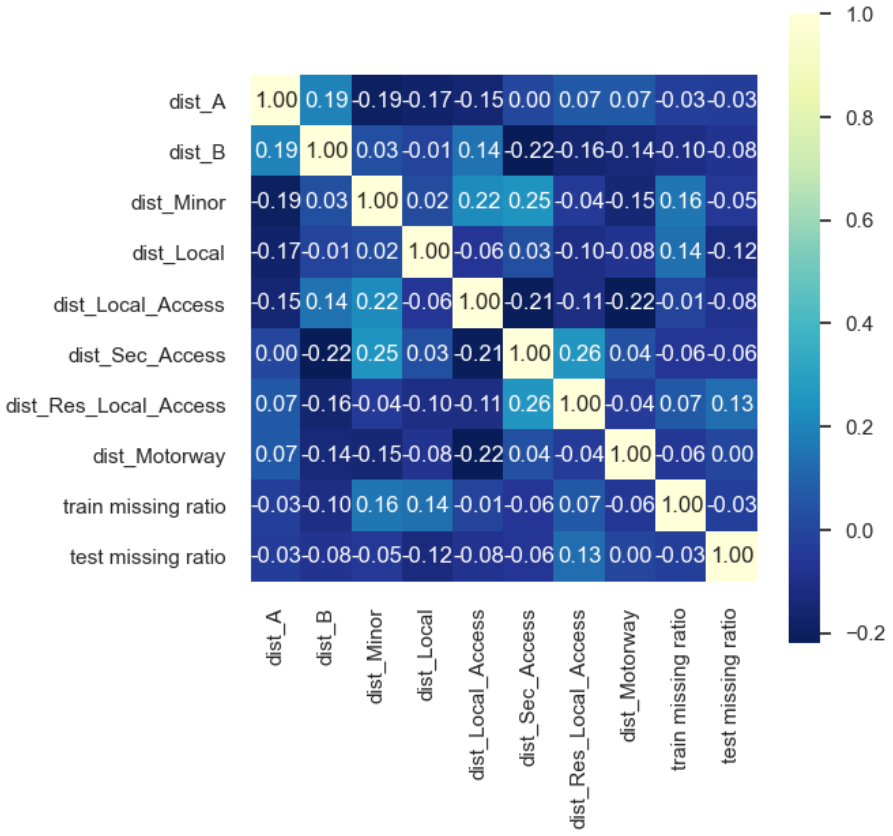

Figure 13. Correlation matrix of continuous variables.

The RFC model is not well-fitted for predicting the optimal method for each location through spatial characteristic variables. The training accuracy is 0.98, and the test accuracy is 0.61. Results shown in Figure 14 indicate that the optimal footfall prediction method of locations relates more to the Distance between locations on different roads.

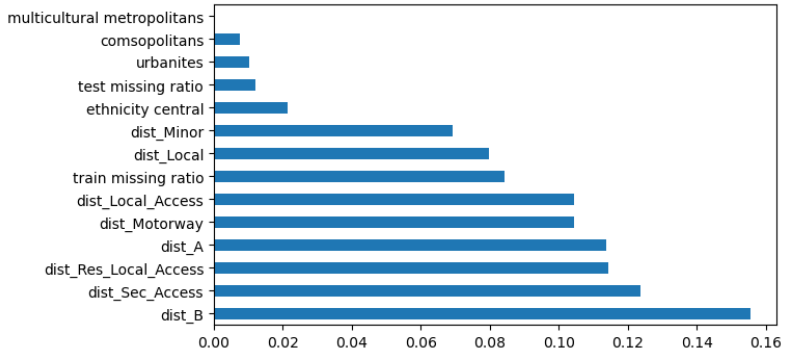

Figure 14. Importance of top 10 variables.

At Appendix B we showed a particular case of study for two locations on a famous high street in London, Oxford Circus.

5. Discussion

Results of the comparison among the performance of different prediction models show that the performance of the LSTM model is the most stable, even in retail areas where the LSTM is not the optimal footfall forecasting model, the prediction error is relatively small, which means that the LSTM model has a very good fit for the pedestrian flow prediction in all sorts of retail areas across the UK.

Other forecasting models apply to certain types of footfall data with different time series characteristics.

For example, in areas where the Naïve approach performs well, the flow distribution varies greatly in different periods within the research time range. This may be due to recent changes in the external environment that have affected the flow of people in a local area, resulting in a footfall distribution similar to a random walk time series. -Laura's paper- where can the random walk be trusted? The daily footfall distribution for MA locations has three distinctive peaks in the morning, noon, and evening. The distribution on weekdays is also different from other areas. For example, compared with other locations where Saturday is the highest pedestrian flow peak in a week, the peak in MA locations appears earlier on Friday. Moreover, the overall time series in the MA locations tends to be more stable, with no obvious trend at a large scale.

The footfall distribution in locations where the HW model performs well is similar to SARIMA locations. The difference is mainly in the distribution on working days. Within a week, the pedestrian flow changes in the SARIMA areas are more dramatic. Similarities in performance between HW and SARIMA models might be related to the common features shared among ES and ARIMA models. A statistical framework introduced by Geurts, Box and Jenkins indicates that the predictions produced by some linear ES models serve as a particular case of ARIMA [30].

The differences in footfall distribution in these retail areas might also be related to their spatial characteristics, indirectly leading to differences in optimal forecasting methods for locations. Therefore, this study mainly explores the differences in land and road types of locations with different optimal prediction methods.

To a certain extent, the statistical analysis explained the differences in the pedestrian flow distribution in these retail areas.

Locations with Naïve, the optimal method, are relatively farther from B roads and motorways and closer to minor roads and local access roads. The spatial characteristics of MA locations are just the opposite of the Naïve locations.

The spatial characteristics of HW and SARIMA locations are quite similar regarding their relations to traffic distribution. The only difference is that SARIMA locations are farther away from A and B roads. In contrast, HW locations are closer to them.

However, the RF model does not fit well, which means that the pedestrian flow characteristics/optimal prediction method of one certain retail location might be related to more sophisticated spatial characteristics.

Limitations

Reflecting on the optimal pedestrian flow forecasting approach for retail areas in Great Britain, our results have demonstrated that it differs in different locations despite the LSTM model being best for most areas. However, some limitations on the data and methodology have yet to be addressed.

One limitation is that the predictive performance might be related to the division of the test set for different locations. However, this shortcoming can be solved by introducing cross-validation, which requires a large amount of calculation and a sufficiently long time series period.

Another limitation is the need for more sample size. Although the LDC has installed over a thousand Wi-Fi sensors to capture pedestrian footfall in Great Britain, due to technical problems, only 311 locations can be used for prediction. At the same time, the more complex spatial features behind these locations need to be excavated.

In terms of the train/test split, future work could introduce cross-validation to identify the best prediction approach and make the conclusion more general. Different forecast periods could also be considered, corresponding to actual uses.

6. Conclusions

This research employs a comparative study approach to identify the optimal footfall prediction method in retail areas across Great Britain. This work responds to the need to further explore the footfall dataset in GB retail areas, especially in forecasting the temporal pedestrian dynamics, which has great practical importance for retailing in different formats in making business decisions as well as helping the government take emergency measures in advance to reduce the crowd risk. To conduct the empirical comparison, five time series prediction methods are selected, namely the Naïve, Moving Averages, Holt–Winters, SARIMA, and LSTM. In addition, MSE and Diebold–Mariano tests are used to test the forecasting performance. A significant statistical difference is shown in the footfall time series and spatial characteristics of locations classified through their optimal footfall forecasting methods.

As far as we know, this is the first study comparing forecasting methods for footfall counts, making the case that spatial characteristics impact the selected methods. Nevertheless, the literature for evaluating forecasting methods [57–59] is vast, where different parameters and criteria are applied. In recent years, the XGBoost model [60] has gained notoriety for its efficiency, flexibility and scalability capabilities, making it widely applicable to classification, regression, and ranking tasks. A key advantage of XGBoost is its proficiency in handling complex relationships. The input features in this study include only one variable, which is the main reason to exclude this popular model from this study.

In future work, multiple variables, such as precipitation and temperature, could be introduced further to reduce the prediction error in pedestrian flow modelling. Introducing a multivariable approach will allow us to compare our results against other ML models, such as XGBoost and other time series forecasting models; for example, fbProphet, the open-source algorithm of Facebook, performs well in dealing with outliers and missing values [61]. One limitation of this study is that for different locations, the predictive performance might be related to the division of the test set. This shortcoming can be solved by introducing cross-validation, which requires a large amount of calculation and a sufficiently long time series period.

Author Contributions: Conceptualization, Yujue Wang and Roberto Murcio; methodology, Roberto Murcio; software, Yujue Wang; validation, Roberto Murcio and Yujue Wang; resources, Yujue Wang and Roberto Murcio; data curation, Yujue Wang; writing—original draft preparation, Yujue Wang; writing—review and editing, Roberto Murcio; supervision, Roberto Murcio. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The footfall data presented in this article are not readily available because the CDRC labelled as Safeguarded data: data to which access is restricted due to license

conditions, but where data are not considered 'personally-identifiable' or otherwise sensitive (from: https://data.cdrc.ac.uk/protecting-data accessed on 21 January 2025). Access is available via a remote service with registration and project approval requirements (https://data.cdrc.ac.uk/using-our-data-services accessed on 21 January 2025). The road network presented in this article is not readily available because the Digimap Collections (https://digimap.edina.ac.uk/ accessed on 21 January 2025) provide maps and mapping data to UK colleges and universities. Institutions should subscribe to the service so their members can access the data (https://digimap.edina.ac.uk/help/about/eligibility/ accessed on 21 January 2025). The road network could be obtained from open sources, such as OpenStreetMap, and it would make no difference in the results obtained. The Output Area Classification (2011) is open data provided by the CDRC. The data are available at https://data.cdrc.ac.uk/dataset/output-area-classification-2011 accessed on 21 January 2025.

Acknowledgments: We acknowledge the support of the CDRC in approving our application for the footfall data.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Assessment

Appendix A.1. Picadilly Circus Underground Station

In Figure A1, the correlation between the two is more significant on weekends. However, for some well-known large retail areas, the distribution of footfall on weekends might be more related to the number of tube exits.

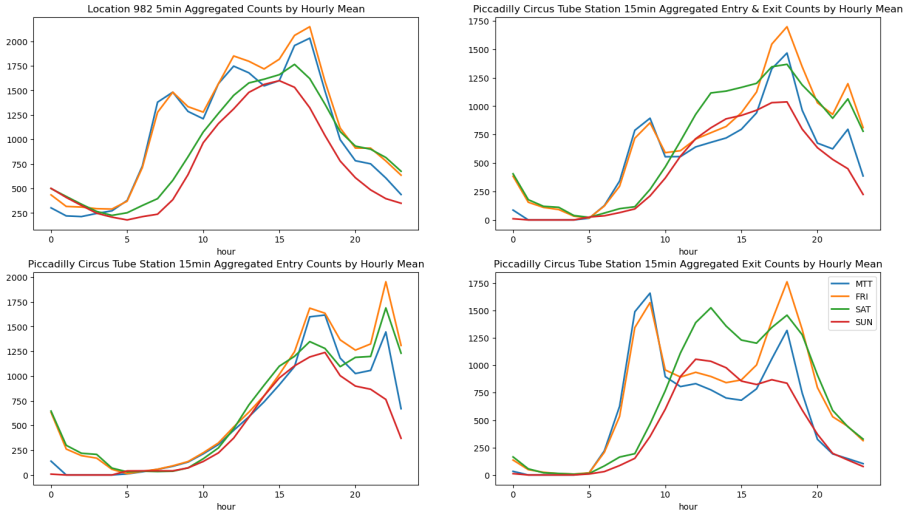

Figure A1. Hourly footfall distribution of location 982 and nearby tube station flows.

Appendix A.2. Google Popular Times

The distribution of the pedestrian flow of six selected retail areas in autumn 2018 is somewhat similar to the popular times of the Google map, with the overall median of Pearson correlation higher than 0.7. The differences in correlations across weekdays are also similar to the TfL data in that the correlation on weekends is higher than on workdays, which is reflected in Figure A2. Location 982 on Regent Street is about 100 m from the Piccadilly Circus station. As seen from Figure A3, the pedestrian flow data show similar troughs and peaks at around 5 a.m. and 5 p.m., respectively. However, due to the differences in the data acquisition time (Google Weekly popular Times uses an annual average and is updated to 2022), it can be seen that the pedestrian flow of some stations has undergone significant changes, which is mainly reflected in the decrease in the ratio of passengers on weekdays and the decrease in morning rush hour.

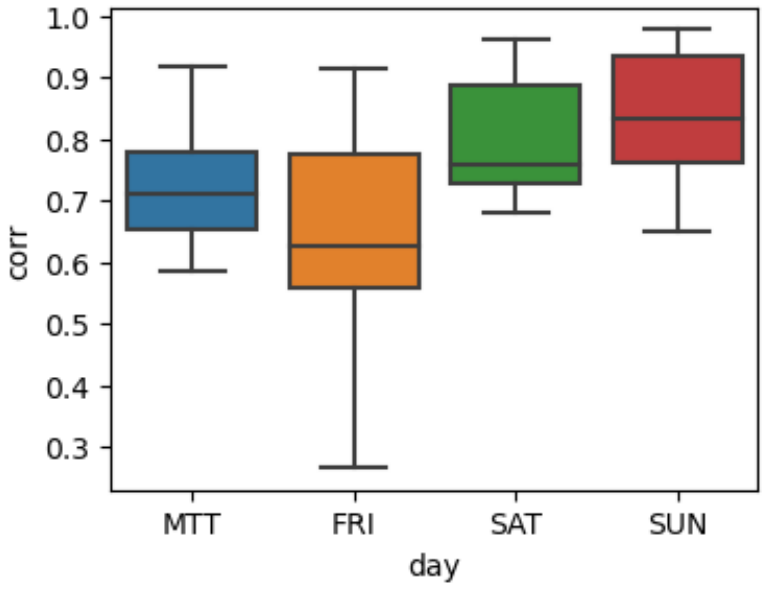

Figure A2. Box plot of Pearson correlation of FF in retail areas and Google popular times of nearby tube station on different days of the week. The MTT label refers to Monday-Thursday.

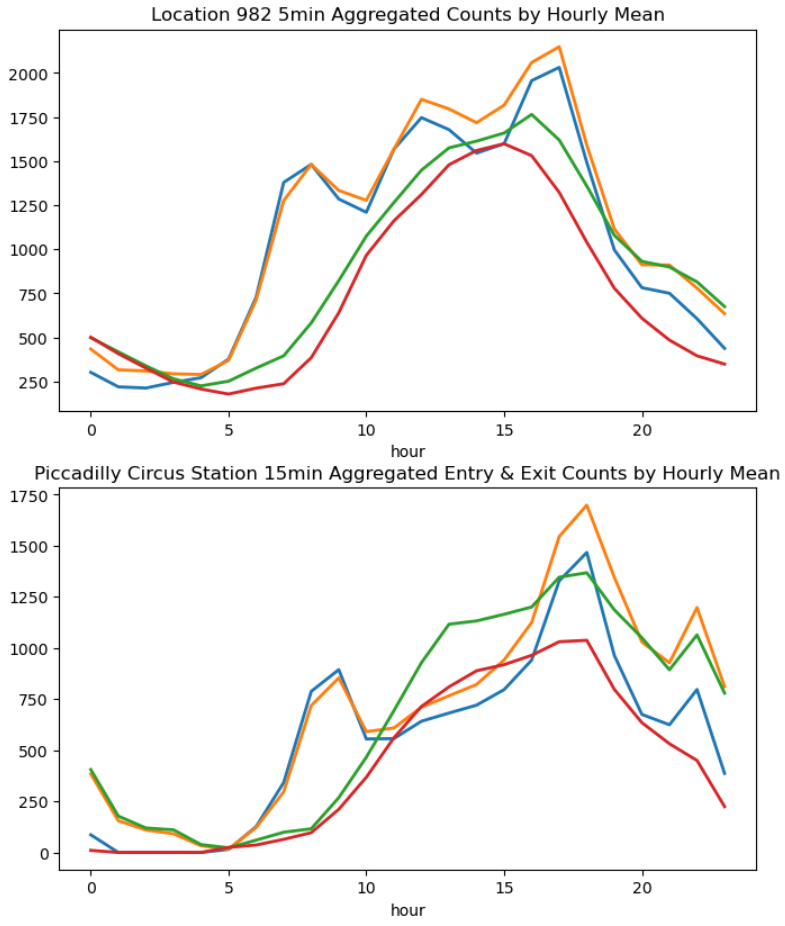

Figure A3. *Cont.*

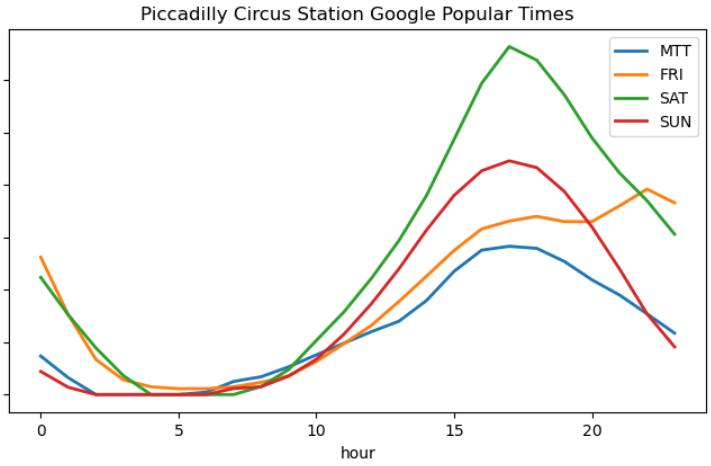

Figure A3. Hourly footfall distribution of location 982, passenger flow and Google popular times of nearby Piccadilly Circus Station.

Appendix B. Case of Study

Location 639, at 163 High Street, Greater London, is our case study to identify the best prediction method.

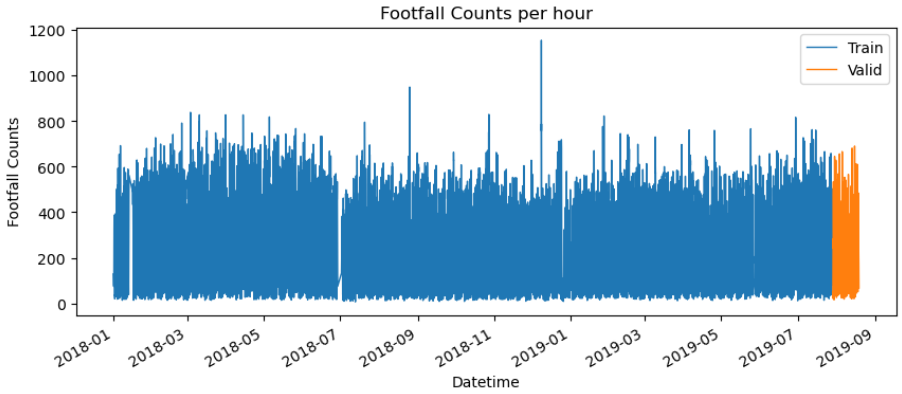

Figure A4. Footfall time series of location 639.

To estimate the parameters in HW and SARIMA models. For example, Figure A5 shows the components of footfall time series by daily mean. It can be seen from the seasonal component that in addition to week, there is also a complex periodicity for around half a year.

Figure A6 illustrates that the initial time series are non-stationary; the Dickey–Fuller test accepted the null hypothesis that a unit root is present. Figure 10 shows a visible fluctuation in the trend component. Therefore, the mean is not constant, and the variance is not stable throughout the series, indicating differentiation is needed in estimating the parameters of the SARIMA model.

The procedure of Auto-SARIMA indicates the SARIMA model with parameters (5, 1, 0) and (5, 1, 0) 7 has the lowest AIC, BIC and HQIC and shows a smaller RMSE than other parameter selections in empirical studies, indicating that these parameters are suitable for model fitting as well as prediction.

Figure A7 shows the train and test loss of the LSTM model. Results indicate that after 30 iterations, the loss curve drops to a stable stage and could be implemented to fit the time series.

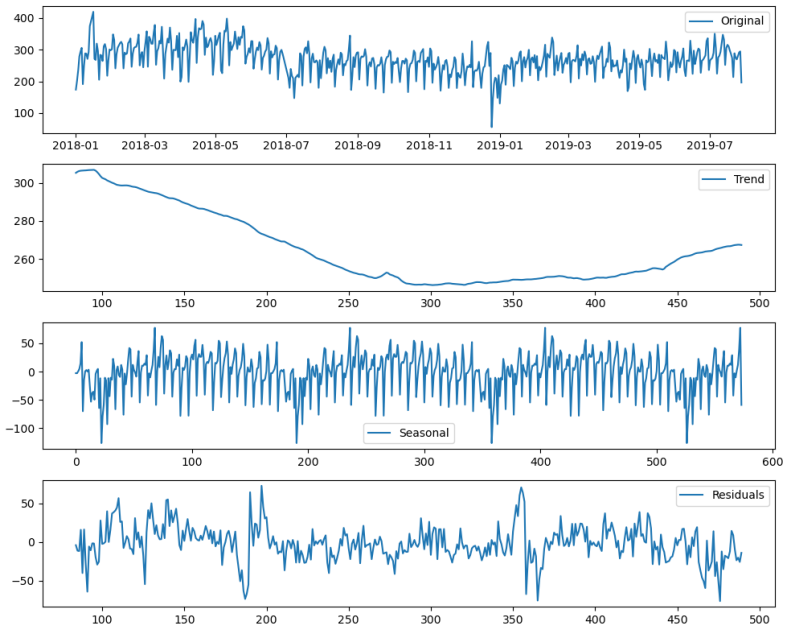

Figure A5. Decompose of footfall time series by daily mean of location 639.

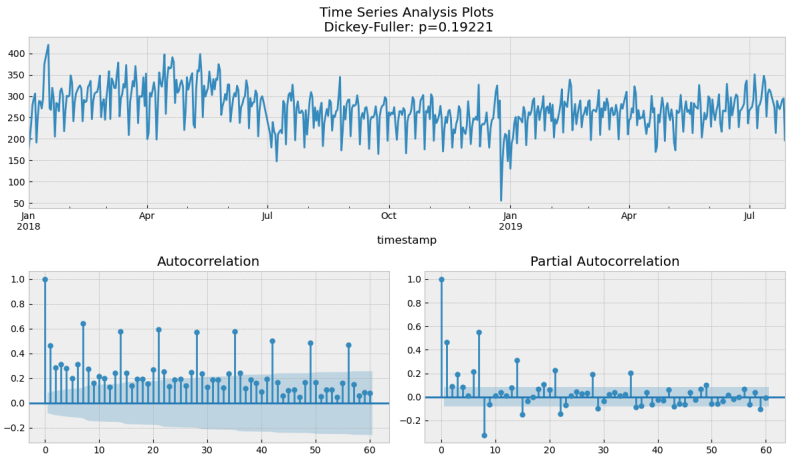

Figure A6. Stationary test on footfall time series by daily mean of location 639.

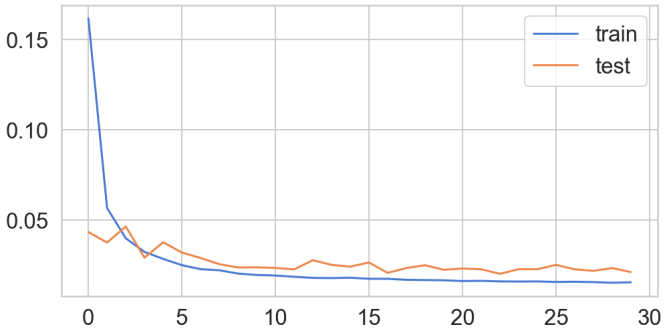

Figure A7. Train and test loss of LSTM of location 639.

Results demonstrate that LSTM outperforms other models and becomes the optimal footfall prediction method in location 639 with a statistical significance.

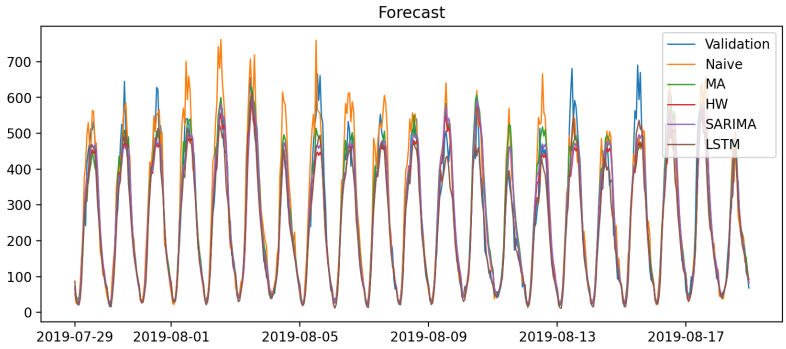

Figure A8. Forecasts on the test set—location 639.

References

1. Waddington, T.; Clarke, G.; Clarke, M.C.; Hood, N.; Newing, A. Accounting for temporal demand variations in retail location models. *Geogr. Anal.* **2019**, *51*, 426–447. [CrossRef]
2. Brown, S. Retail location theory: Evolution and evaluation. *Int. Rev. Retail. Distrib. Consum. Res.* **1993**, *3*, 185–229. [CrossRef]
3. Lugomer, K.; Longley, P. Towards a comprehensive temporal classification of footfall patterns in the cities of Great Britain. In *10th International Conference on Geographic Information Science (GIScience 2018)*; Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik: Dagstuhl, Germany, 2018; Volume 114, p. 43.
4. Ministry of Housing Communities & Local Government. 2019. Available online: https://assets.publishing.service.gov.uk/media/5fa4345de90e070420702a78/MHCLG_ARA_2019-20_Web_Accessible.pdf (accessed on 21 January 2025).
5. Parker, C.; Ntounis, N.; Millington, S.; Quin, S.; Castillo-Villar, F.R. Improving the vitality and viability of the UK High Street by 2020: Identifying priorities and a framework for action. *J. Place Manag. Dev.* **2017**, *10*, 310–348. [CrossRef]
6. Chapados, N.; Joliveau, M.; L'Ecuyer, P.; Rousseau, L.M. Retail store scheduling for profit. *Eur. J. Oper. Res.* **2014**, *239*, 609–624. [CrossRef]
7. BBC1. Georgi H, 2015. Available online: https://www.bbc.co.uk/news/uk-england-london-34997203 (accessed on 21 January 2025).
8. Luo, W.; Jiao, P.; Wang, Y. Pedestrian arching mechanism at bottleneck in subway transit hub. *Information* **2021**, *12*, 164. [CrossRef]
9. Luo, W.; Wang, Y.; Jiao, P.; Wang, Z. Improvement strategy at pedestrian bottleneck in subway stations. *Discret. Dyn. Nat. Soc.* **2022**, *2022*, 7258907. [CrossRef]
10. Zhang, P.; Li, X.Y.; Deng, H.Y.; Lin, Z.Y.; Zhang, X.N.; Wong, S.C. Potential field cellular automata model for overcrowded pedestrian flow. *Transp. A Transp. Sci.* **2020**, *16*, 749–775. [CrossRef]
11. Liu, M.; Li, L.; Li, Q.; Bai, Y.; Hu, C. Pedestrian flow prediction in open public places using graph convolutional network. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 455. [CrossRef]
12. De Gooijer, J.G.; Hyndman, R.J. 25 years of time series forecasting. *Int. J. Forecast.* **2006**, *22*, 443–473. [CrossRef]
13. Sapankevych, N.I.; Sankar, R. Time series prediction using support vector machines: A survey. *IEEE Comput. Intell. Mag.* **2009**, *4*, 24–38. [CrossRef]
14. Parmezan, A.R.S.; Batista, G.E. A study of the use of complexity measures in the similarity search process adopted by kNN algorithm for time series prediction. In Proceedings of the 2015 IEEE 14th International Conference on Machine Learning and Applications, ICMLA 2015, Miami, FL, USA, 9–11 December 2015 ; pp. 45–51. [CrossRef]
15. Cortez, P. Sensitivity analysis for time lag selection to forecast seasonal time series using neural networks and support vector machines. In Proceedings of the International Joint Conference on Neural Networks, Barcelona, Spain, 18–23 July 2010. [CrossRef]
16. Kandananond, K. A comparison of various forecasting methods for autocorrelated time series. *Int. J. Eng. Bus. Manag.* **2012**, *4*, 1–6. [CrossRef]
17. Ristanoski, G.; Liu, W.; Bailey, J. A time-dependent enhanced support vector machine for time series regression. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2013, Part F1288, Chicago, IL, USA, 11–14 August 2013; pp. 946–954. [CrossRef]
18. Zhang, X.; Zhang, T.; Young, A.A.; Li, X. Applications and comparisons of four time series models in epidemiological surveillance data. *PLoS ONE* **2014**, *9*, e88075. [CrossRef] [PubMed]
19. Parmezan, A.R.S.; Souza, V.M.; Batista, G.E. Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Inf. Sci.* **2019**, *484*, 302–337. [CrossRef]
20. Gautam, A.; Singh, V. Parametric versus non-parametric time series forecasting methods: A review. *J. Eng. Sci. Technol. Rev.* **2020**, *13*, 165–171. [CrossRef]

21. Weldegebriel, H.T.; Liu, H.; Haq, A.U.; Bugingo, E.; Zhang, D. A New Hybrid Convolutional Neural Network and eXtreme Gradient Boosting Classifier for Recognizing Handwritten Ethiopian Characters. *IEEE Access* **2020**, *8*, 17804–17818. [CrossRef]
22. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
23. Wei, W.W. *Multivariate Time Series Analysis and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2019.
24. Utlaut, T.L. *Introduction to Time Series Analysis and Forecasting*; Wiley-Interscience: Hoboken, NJ, USA, 2008; Volume 40, pp. 476–478. [CrossRef]
25. Sezer, O.B.; Gudelek, M.U.; Ozbayoglu, A.M. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Appl. Soft Comput. J.* **2020**, *90*, 106181. [CrossRef]
26. Ball, R.; Bartov, E. How naive is the stock market's use of earnings information? *J. Account. Econ.* **1996**, *21*, 319–337. [CrossRef]
27. Hyndman, R. *Forecasting: Principles and Practice*; OTexts: Melbourne, Australia, 2018.
28. Winters, P.R. Forecasting Sales by Exponentially Weighted Moving Averages. *Manag. Sci.* **1960**, *6*, 324–342. [CrossRef]
29. Holt, C.C. Forecasting seasonals and trends by exponentially weighted moving averages. *Int. J. Forecast.* **2004**, *20*, 5–10. [CrossRef]
30. Geurts, M.; Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*, rev. ed.; Holden-Day: San Francisco, CA, USA; London, UK, 1977; Volume 14, p. 269. [CrossRef]
31. Cristianini, N. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*; Cambridge University Press: Cambridge, UK, 2000.
32. Cooper, G.F.; Herskovits, E. A Bayesian Method for the Induction of Probabilistic Networks from Data. *Mach. Learn.* **1992**, *9*, 309–347. [CrossRef]
33. Rasmussen, C.E. Gaussian processes in machine learning. In *Summer School on Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 63–71.
34. Gámez, J.A.; Mateo, J.L.; Puerta, J.M. Learning Bayesian networks by hill climbing: Efficient methods based on progressive restriction of the neighborhood. *Data Min. Knowl. Discov.* **2011**, *22*, 106–148. [CrossRef]
35. Bengio, Y.; LeCun, Y. *Scaling Learning Algorithms Toward AI*; MIT Press: Cambridge, MA, USA, 2007.
36. Zhang, P.; Ma, X.; Zhang, W.; Lin, S.; Chen, H.; Yirun, A.L.; Xiao, G. Multimodal fusion for sensor data using stacked autoencoders. In Proceedings of the 2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Singapore, 7–9 April 2015; pp. 1–2.
37. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]
38. Gamboa, J.C.B. Deep learning for time-series analysis. *arXiv* **2017**, arXiv:1701.01887.
39. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
40. Graves, A.; Jaitly, N.; Mohamed, A.R. Hybrid speech recognition with Deep Bidirectional LSTM. In Proceedings of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2013, Olomouc, Czech Republic, 8–12 December 2013; pp. 273–278. [CrossRef]
41. Greff, K.; Srivastava, R.K.; Koutnik, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A Search Space Odyssey. *IEEE Trans. Neural Networks Learn. Syst.* **2017**, *28*, 2222–2232. [CrossRef]
42. Guy, C. Retail location analysis. In *Applied Geography*; Routledge: London, UK, 2002; pp. 450–462.
43. Wang, W.; Wang, L.; Wang, X.; Wang, Y. Geographical Determinants of Regional Retail Sales: Evidence from 12,500 Retail Shops in Qiannan County, China. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 302. [CrossRef]
44. cDRC. Consumer Data Research Centre. SmartStreetSensor Project. Available online: https://data.cdrc.ac.uk/dataset/local-data-company-ucl-smartstreetsensor-footfall-data-research-aggregated-data (accessed on 1 June 2024).
45. Soundararaj, B.; Cheshire, J.; Longley, P. Estimating real-time high-street footfall from Wi-Fi probe requests. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 325–343. [CrossRef]
46. Murcio, R.; Soundararaj, B.; Lugomer, K. Movements in Cities: Footfall and its spatio-temporal distribution. In *Consumer Data Research*; UCL Press: London, UK, 2018; pp. 84–95.
47. Trasberg, T.; Soundararaj, B.; Cheshire, J. Using Wi-Fi probe requests from mobile phones to quantify the impact of pedestrian flows on retail turnover. *Comput. Environ. Urban Syst.* **2021**, *87*, 101601. [CrossRef]
48. Soundararaj, B. Estimating Footfall From Passive Wi-Fi Signals: Case Study with Smart Street Sensor Project. Ph.D Thesis, UCL (University College London), London, UK, 2019.
49. Nunnari, G.; Nunnari, V. Forecasting monthly sales retail time series: A case study. In Proceedings of the 2017 IEEE 19th Conference on Business Informatics (CBI). Thessaloniki, Greece, 24–27 July 2017; Volume 1, pp. 1–6.
50. Fildes, R.; Ma, S.; Kolassa, S. Retail forecasting: Research and practice. *Int. J. Forecast.* **2022**, *38*, 1283–1318. [CrossRef]
51. TFL. TFL London Underground Passenger Counts Data. Available online: https://api-portal.tfl.gov.uk/ (accessed on 12 June 2019).

52. Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [CrossRef]
53. Fu, R.; Zhang, Z.; Li, L. Using LSTM and GRU neural network methods for traffic flow prediction. In Proceedings of the 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, China, 11–13 November 2016; pp. 324–328.
54. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **2002**, *20*, 134–144. [CrossRef]
55. Chen, W.; Zhang, S.; Li, R.; Shahabi, H. Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. *Sci. Total Environ.* **2018**, *644*, 1006–1018. [CrossRef]
56. Taalab, K.; Cheng, T.; Zhang, Y. Mapping landslide susceptibility and types using Random Forest. *Big Earth Data* **2018**, *2*, 159–178. [CrossRef]
57. Wang, S.; Chaovalitwongse, W.A. Evaluating and comparing forecasting models. In *Wiley Encyclopedia of Operations Research and Management Science*; Wiley: Hoboken, NJ, USA, 2011. [CrossRef]
58. Smith, B.L.; Demetsky, M.J. Traffic flow forecasting: Comparison of modeling approaches. *J. Transp. Eng.* **1997**, *123*, 261–266. [CrossRef]
59. Webby, R.; O'Connor, M. Judgemental and statistical time series forecasting: A review of the literature. *Int. J. Forecast.* **1996**, *12*, 91–118. [CrossRef]
60. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
61. Chikkakrishna, N.K.; Hardik, C.; Deepika, K.; Sparsha, N. Short-term traffic prediction using sarima and FbPROPHET. In Proceedings of the 2019 IEEE 16th India Council International Conference, INDICON 2019—Symposium Proceedings, Rajkot, India, 13–15 December 2019. [CrossRef]