modelling

MDPI

Article

Predicting Cyclist Speed in Urban Contexts: A Neural Network Approach

Ricardo Montoya-Zamora [1], Luisa Ramírez-Granados [1], Teresa López-Lara [1], Juan Bosco Hernández-Zaragoza [1] and Rosario Guzmán-Cruz [2,*]

[1] Facultad de Ingeniería, Universidad Autónoma de Querétaro, Cerro de las Campanas s/n, Col. Las Campanas, Querétaro 76010, Mexico; ricardo.montoya@uaq.mx (R.M.-Z.); luisa.ramirez@uaq.mx (L.R.-G.)
[2] Facultad de Ingeniería, Campus Amazcala, Universidad Autónoma de Querétaro, Carr. Chichimequillas-Amazcala Km 1 s/n, Amazcala, Querétaro 76265, Mexico
* Correspondence: rosario.guzman@uaq.mx

Abstract: Bicycle use has become more important today, but more information and planning models are needed to implement bike lanes that encourage cycling. This study aimed to develop a methodology to predict the speed a cyclist can reach in an urban environment and to provide information for planning cycling infrastructure. The methodology consisted of obtaining GPS data on longitude, latitude, elevation, and time from a smartphone of two groups of cyclists to calculate the speeds and slopes through a model based on a recurrent short-term memory (LSTM) type neural network. The model was trained on 70% of the dataset, with the remaining 30% used for validation and varying training epochs (100, 200, 300, and 600). The effectiveness of recurrent neural networks in predicting the speed of a cyclist in an urban environment is shown with determination coefficients from 0.77 to 0.96. Average cyclist speeds ranged from 6.1 to 20.62 km/h. This provides a new methodology that offers valuable information for various applications in urban transportation and bicycle line planning. A limitation can be the variability in GPS device accuracy, which could affect speed measurements and the generalizability of the findings.

Keywords: prediction; recurrent neural network; cyclist speed; urban area

Citation: Montoya-Zamora, R.; Ramírez-Granados, L.; López-Lara, T.; Hernández-Zaragoza, J.B.; Guzmán-Cruz, R. Predicting Cyclist Speed in Urban Contexts: A Neural Network Approach. *Modelling* **2024**, *5*, 1601–1617. https://doi.org/10.3390/modelling5040084

Academic Editor: Ivan Dimov

Received: 9 October 2024
Revised: 28 October 2024
Accepted: 31 October 2024
Published: 5 November 2024

1. Introduction

Currently, bicycle use has increased because of its health benefits and because it has become the most accessible, sustainable, and environmentally friendly means of transport. Therefore, it is essential to have adequate and safe cycling infrastructure that encourages the use of more sustainable means of transport, like bicycles, and reduces the use of motorized transport, especially in large cities with environmental pollution, traffic congestion, and road accident issues.

To achieve this, studies have been conducted using smartphones, cameras, and web platforms for data collection, transmission, processing, and analysis. These tools enable improved cycling mobility in smart cities, optimizing cycling infrastructure and traffic management, supporting the choice of more suitable routes, and promoting sustainable transport [1–4]. However, when collecting information from users, it is crucial to prioritize computer security and data privacy to build trust and ensure the successful implementation of these technologies. Research on the vulnerability of biometric recognition to identity theft underlines this need [5–7]. Thus, improving urban mobility aligns with the principles of smart city development, promoting a safer and more efficient urban environment.

Therefore, constructing cycling infrastructure requires proper planning of the cycling network in any city. Planning a cycling network is not a simple issue since factors such as the following must be considered [8]:

- Conducting preliminary studies (surveys, traffic counts, identification of priority routes, and accident analysis).

- Identification of suitable roads (traffic volume, speed schemes, road functionality, geometric surveying, opportunities for new links, and analysis of conflict points).
- Design and implementation (choice of type of cycling infrastructure, design of plans, implementation, and supervision of the work).
- Monitoring (maintenance works and continuous improvement).

A valuable tool in planning cycling infrastructure is modeling to identify optimal routes through speed analysis. This allows for analysis of the advantages and disadvantages of implementation before economic resources are allocated to construction. In this sense, it is essential to study the impact of bicycle speed on urban infrastructure and vice versa, as it influences both the safety of cyclists and the efficiency of traffic in general, especially in cities that seek to use bicycles as a sustainable transportation option.

Since cyclist movement influences vehicular traffic, studying the factors affecting cyclists is also important. These factors are broad and range from infrastructure conditions to weather conditions. For example, stress, slope, congestion, connectivity, or interaction with motor vehicles affect movement [9,10].

In this regard, studies show that external conditions, cultural context, safety, sociodemographic characteristics [11], social environment, residential location, socioeconomic level [12], the width of the bike lane, or the use of exclusive lanes [13] also influence speed.

In recent years, various studies have been conducted to determine how the use of bicycles affects vehicular traffic and the factors that influence cyclists when choosing a route when moving from one place to another. Also, the influence of infrastructure on cyclist movement is how the width of the lane, the slope, or the curvature determines the cyclist's speed.

Ref. [14] investigated the influence of bicycle use on vehicular traffic on urban streets in Nanjing, China. They used the cumulative curve method with video data to estimate and analyze data periods; they then used a multiple linear regression model for vehicular movement delays based on vehicular flow. This allowed them to find that increased bicycle density can significantly reduce vehicle speed.

Ref. [15] compared speeds between cyclists and vehicles in Montreal, Canada, in traffic assignment zones (TAZ), finding that bicycle use can compete with cars in dense areas, and the speed difference between both modes decreases during peak hours. Ref. [16] applied first- and last-mile studies and found that cyclists prefer more extended travel to a train station if this means not transferring; that is, they prefer comfort over speed.

In environments with little interaction with vehicles, it has been studied that gender, age, type of bicycle, lane width, and lateral position of the bicycle influence free-flow speed, suggesting design speeds of 25 km/h for lanes less than 3.5 m wide and 30 km/h for wider lanes [17]. On the other hand, [18] found that painted bike lanes increase cyclist speed since they allow cyclists to move an average of 31 cm away from the sidewalk compared to streets without painted lanes.

Studies like [19] have found that speed is influenced by reflective vests or the cyclists' experience on e-bikes. Ref. [20] used variables such as the cyclist's weight, wind speed, and slope to determine speed on electric bikes. Other studies consider safety, slope, and surface conditions as variables affecting cyclists' movement performance [21] or using segregated bike lanes and bike paths [22]. It should not be overlooked that route choice also influences cyclists' behavior, their interactions with drivers [23], and mental and physical comfort [24].

Regarding infrastructure, various authors have found that cyclist speeds are influenced by the width of the path, traffic conditions, type of bicycle, and cyclist characteristics [25]; the level of traffic, speed limits, the presence of heavy vehicles, and the number of intersections also influence [26].

Some studies have focused on measuring speed variations of regular and non-regular cyclists using the STRAVA app [27]. Studies related to GPS data collection focus on obtaining the quality of service experienced by cyclists when using the infrastructure [28] or using georeferenced data to represent the level of comfort and safety in cyclist environments [29], estimate the annual average daily volume on rural roads [30], or study accidents on urban roads [31].

Among the studies with GPS data, [32] is the only one focused on predicting the speed of bicycles based on slope and horizontal curvature using a Markov model. This suggests that the design characteristics of the bike paths influence cyclist speed. To date, no studies have been found that involve other techniques, such as neural networks, for estimating a cyclist's speed in an urban environment.

Ref. [33] conducted a study in Davis, California, and estimated that the average speeds of cyclists in bicycle lanes range from 17.7 km/h to 20.1 km/h. Ref. [34] found average speeds between 20.3 and 24.9 km/h with students at the University of Michigan. In Sweden, [35] studied the 85th percentile free flow speed of bicycles, which is between 16 and 28 km/h. The free flow speed in Canada was 25 km/h [36]. In China, the average speeds varied from 10 to 16 km/h [37]. Another study in China estimated average speeds between 12 and 16.3 km/h [38].

Since cycling infrastructure is a crucial factor in the choice of cycling route, it determines the speed of movement of an average cyclist. This study aims to develop a methodology to estimate cyclist speed based on slope, longitude, latitude, elevation, and GPS time, measured from the smartphones of two groups of cyclists in the urban area of Querétaro, using recurrent neural networks. Finally, it provides a new methodology that offers valuable information for various transportation and urban cycling planning applications.

2. Materials and Methods

The method utilized in this study involved the implementation of a long short-term memory (LSTM) recurrent neural network (RNN) in forecasting cyclist speed for two reasons: firstly, due to the scarcity of studies related to cyclist speed forecasting, with only Markov models and linear regression being employed [39]; secondly, because in recent years, neural networks have demonstrated high performance in forecasting complex systems.

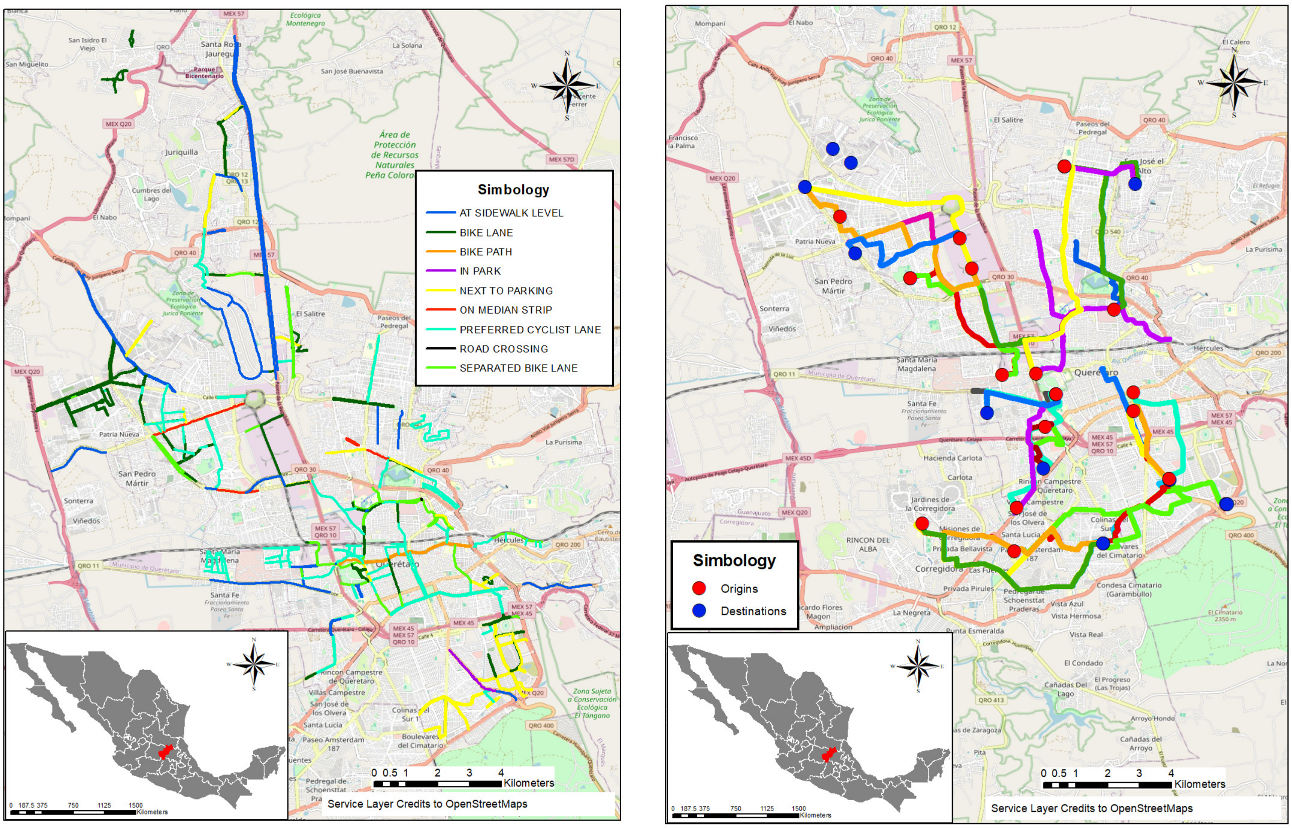
2.1. Experiment Design

This study was conducted in Queretaro, Mexico, through the publication of a call for various cycling groups. They were invited to participate in pre-established routes to identify their selected routes and their achieved speeds. They were informed that the gathered information would be used for this study. Cyclists interested in participating register through a form designed in Google Forms to gather information from the participants, with prior authorization, such as a cycling group, age, gender, height, weight, and frequency of bicycle use. Each cyclist installed an app on their smartphone to collect GPS information.

Sixteen origin points and sixteen destination points known to the cyclists were defined, and the shortest routes between these points were found. From the total number of routes found, journeys in the ranges of 2–3, 4–5, 6–7, and 8–9 km were selected, and 40 origin–destination pairs were generated (10 per range), which were published for the participants' knowledge.

Each route was sought to start and end in parking lots, parks, and centers for reference and accessible locations, as shown in Figure 1b. Each participant was asked to choose four routes (one route in each range) to be completed before going to work or heading home and without a companion from their cycling group to avoid any influence on the measured travel speeds.

Participants were asked to record their routes in the STRAVA app or any other app that would record the path of their choice. Subsequently, each participant shared their route information on the Strava website for later download. As shown in Figure 1a, some of the routes chosen by the participants do not align with the cycling infrastructure in Querétaro, which highlights the importance of having models that support the planning of cycling infrastructure.



(**a**) (**b**)

Figure 1. Comparison between the current cycling infrastructure and the travel preferences of cyclists in the urban area of Querétaro. (**a**) shows the current classification and connectivity of the infrastructure. (**b**) represents the GPS routes the cyclists choose for trips between the different origins and proposed destinations.

2.2. Data Normalization

To input the data into the artificial neural network (ANN), it was essential to ensure a uniform representation of slope and speed, as these variables were not directly obtained from the GPS (Global Positioning System). Data normalization is essential preprocessing when using artificial neural networks because it reduces variability and improves model efficiency during training [40].

The slope was calculated by dividing the elevation difference by the horizontal distance between two consecutive GPS points, using the formula slope = Δ elevation/Δ distance. Similarly, speed was calculated as the distance traveled between two points divided by the time elapsed: speed = Δ distance/Δ time. This normalization process allows the ANN to interpret the inputs consistently, regardless of scale variations among different measurements.

For this study, a Min–Max normalization technique was applied, scaling the values between 0 and 1 to ensure compatibility across various GPS sources in different smartphone types and minimizing issues with outliers. The formula for scaling is as follows:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

where:

X_{scaled} is the speed value scaled.
X is the GPS speed value.
X_{min} is the minimum value of speed.

X_{max} is the maximum value of the speed in the dataset.

Each cyclist's data were obtained in gpx format directly from their smartphone, and routes with missing time data or incorrect georeferenced values were discarded to maintain data quality. Additionally, the autocorrelation of the time series was calculated to understand the data distribution over time and to identify potential seasonal or trend patterns in the data.

2.3. Design of the Recurrent Neural Network

Recurrent neural networks (RNN) are a type of artificial neural network specialized for analyzing data sequence patterns and can also be applied to time series data. These networks have connections that form cycles, allowing them to retain information from previous states. This ability makes them ideal for tasks where the context or the order of data is crucial, such as in natural language processing or in the prediction of time series.

Figure 2 shows a typical LSTM network with four layers of interacting neural networks and details the structures of the hidden layer, forget gate, input gate, output gate, and memory block. Recurrent neural networks (RNNs) require historical information (previous states) to predict future states and can be estimated with one or more input variables. In LSTM networks, information flow passes from the cell state through the entire chain (with some minor linear interactions), and the information flows unchanged (Figure 2a).

The LSTM network, with its unique architecture, can dynamically add or remove information in the cell state through specialized components known as gates. These gates, each consisting of an artificial neural network layer with its respective activation function, play a pivotal role in controlling the neuron's state. The first of these gates, the forget gate, selectively removes information from the cell state, providing a clear understanding of the network's architecture (Figure 2b) [41].

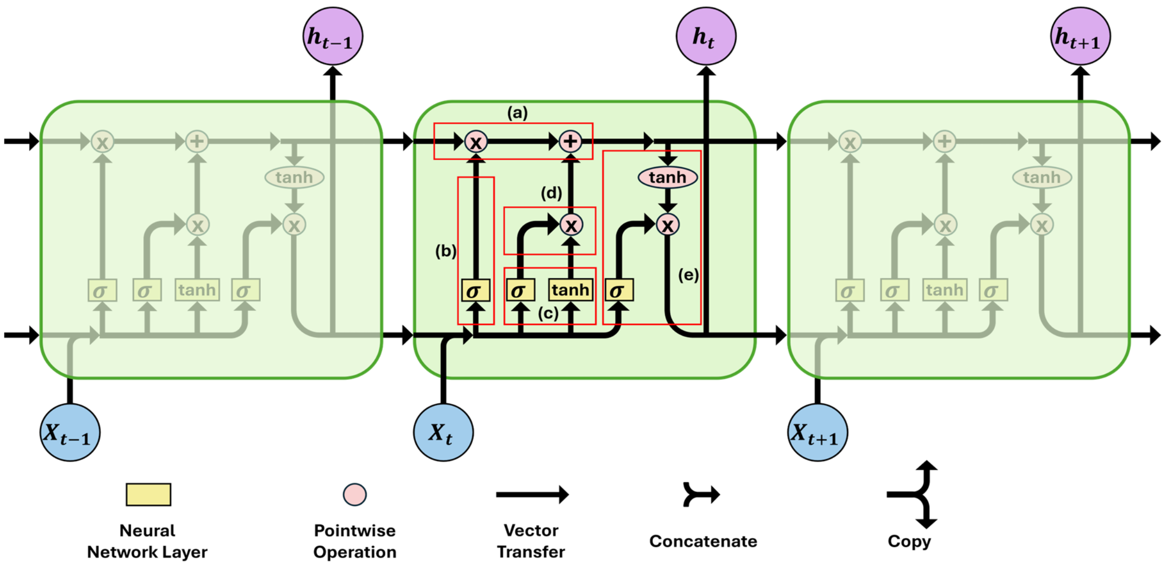


Figure 2. Repetitive module of the LSTM-type neural network with four layers. (**a**) Information flow, (**b**) forget gate, (**c**) information added, (**d**) update information, and (**e**) output data [42].

The forget gate decides which information will be retained from the cell state. The next gate, called the input gate, decides which values should be updated. It does this by considering the current input and the previous output. Together with a hidden layer (with a hyperbolic tangent function as the activation function), it creates new values that are candidates to be added to the current state. Updating the cell state is a key operation in the LSTM network (Figure 2c).

Once the new values are created (and the decision to add them or not has been made), the operations (Figure 2d) are performed to update the previous cell state as the new one.

Finally, the output data, culminating in the LSTM network's intricate operations, are not just based on the updated cell state. It undergoes a crucial filtering process through a sigmoid layer, which intelligently decides what part of the information passes to the next layer, illuminating the network's decision-making process (Figure 2e) [41].

The neural network was developed using the Python programming language version 3.9, one of the most widely used softwares for training artificial neural networks. The Scikit-learn and Keras libraries were imported to facilitate the construction of the LSTM.

2.4. Prediction Based on a Single Variable

The previous speed was input for the single variable case to estimate the current speed. Since RNNs require historical data, scenarios were designed considering intervals of 5, 10, 15, and 20 previous records (in seconds), allowing the models to learn from different time horizons. These intervals were considered to evaluate the forecast's confidence level in the long term. Finally, it was decided to design a neural network with three hidden layers called the "base network", in which the number of neurons in the hidden layers was varied to measure the forecast's performance. The hyperparameters used in this study for the neural network models include:

1. Batch Size. This parameter determines the number of training samples to be used in one iteration of the model training. Variations in batch size were tested to evaluate their impact on model performance (16 or 32).
2. Twenty neurons in each hidden layer. The neural network architecture includes hidden layers, and the number of neurons in these layers can significantly affect the model's ability to learn complex patterns. Different configurations were explored to find the optimal setup.
3. Historical data. The model utilized 20 s of historical data for training, which were critical in predicting cyclist speeds based on previous records of slope and speed.
4. Dropout. A dropout rate of 0.2 was used to prevent overfitting by randomly setting a fraction of the input units to 0 during training.
5. 'Adam' optimizer. It was employed to train the model; it is known for its efficiency and effectiveness in handling sparse gradients.
6. Mean squared error loss. The mean squared error (MSE) was used as the loss function to measure the average of the squares of the errors between predicted and actual values.

Finally, as part of the scenarios, evaluating the data behavior as a time series to determine trend and seasonality patterns through autocorrelation diagrams was considered. Since predicting a cyclist's speed is desired, this article focuses on generating a recurrent neural network model that allows predicting the speed of an average cyclist with data from other cyclists.

2.5. Prediction Based on Two Variables

In the case of the two-variable model, slope and speed were used as input data, and speed was used as the output or prediction data. For the evaluation of the RNN with two variables, it was decided to use 20 previous records of slope and speed to evaluate the forecast performance compared to the evaluation with one variable. The structure of this network was the same as that of a single variable.

2.6. Model Validation

The model validation process consists of evaluating the estimated results from data other than those used in the methodology to prove that it leads to the expected result. As mentioned above, 30% of the data were used for validation. Some statistical parameters were used to analyze the results obtained, such as the coefficient of determination (R^2), whose value must be near one. Additionally, the mean absolute error (MAE), the root mean square error (RMSE), the mean absolute percentage error (MAPE), and the mean squared error (MSE) were calculated; these values must be near zero [43].

3. Results

3.1. Cyclist Response

Out of all the cyclist groups in Querétaro, two cycling groups called "Saca la Bici" and "Libre a Bordo" responded, from which 135 cyclists were registered. Finally, 35 cyclists completed the four full routes with GPS information.

Table 1 shows the statistical data collected from the cyclists from two sources: a Google form and data collected from the Strava app. It can be seen that the average follows a trend to move on asphalt, on flat terrain with a slope close to 1% at 14.53 km/h in the urban area of Querétaro. Despite encountering obstacles along the way, on average, 88% of the trip remains in motion.

Table 1. Cyclist variables during study rides.

Variable	Max	Min	Mean
Age (y)	65	18	34.3
Height (m)	1.8	1.53	1.7
Weight (kg)	90	55	72.04
Frequency (days/week)	7	1	3.92
Speed (km/h)	20.7	6.8	14.53
Slope (%)	5.1	−1.2	1.15
Percentage of the route on asphalt (%)	100	5	78.29
Percentage of climb in the route (%)	71.9	0	31.2
Percentage of downhill in the route (%)	28.9	0	6.95
Percentage of flat terrain in the route (%)	98.3	15.7	58.52
Motion relationship	1	0.59	0.88

3.2. Normalized Data

The point records were downloaded every second from the GPS devices and the Strava app in gpx format, and a database was finally generated with 35,194 records containing data on longitude, latitude, altitude, and time within the urban area of Querétaro during June, July, and August 2021. Once the data were collected, the Geographic Information System (GIS), Qgis, was used to calculate the distances between consecutive GPS records (Δd), the time difference (Δt), and the altitude difference (ΔA) to calculate the speed and slope between two continuous records.

Figure 3 partially shows the cyclists' behaviors. As can be seen, there is no regular pattern of behavior in the cyclist's speed; this makes it more challenging to estimate the speed forecast since, during the rides, cyclists brake for pedestrian crossings, traffic lights, or directional turns.

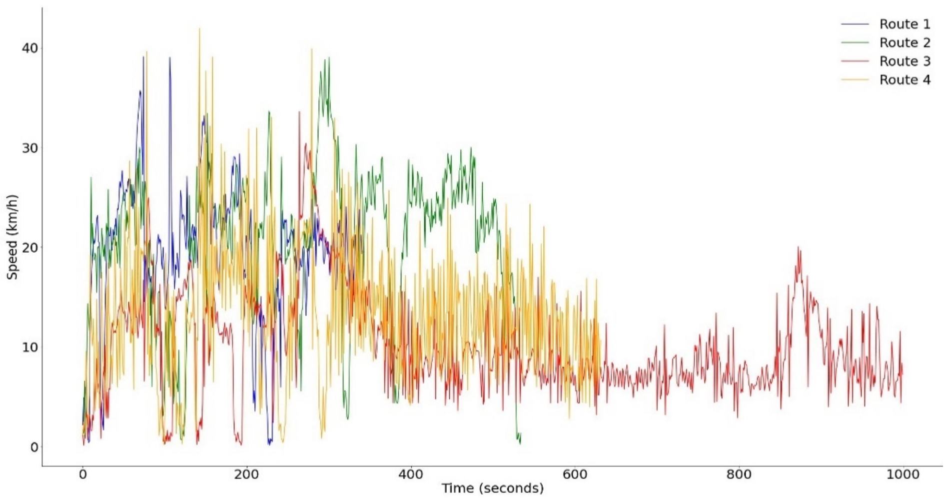


Figure 3. Variation of speed along the route for four selected cyclists on four routes.

Figure 4 shows the autocorrelation calculation of cyclist speeds, aiming to find any seasonal or trend pattern. However, the correlation coefficients could be higher, which makes it very difficult to forecast speed with time series techniques such as autoregressive models, ARIMA models, or SARIMA.

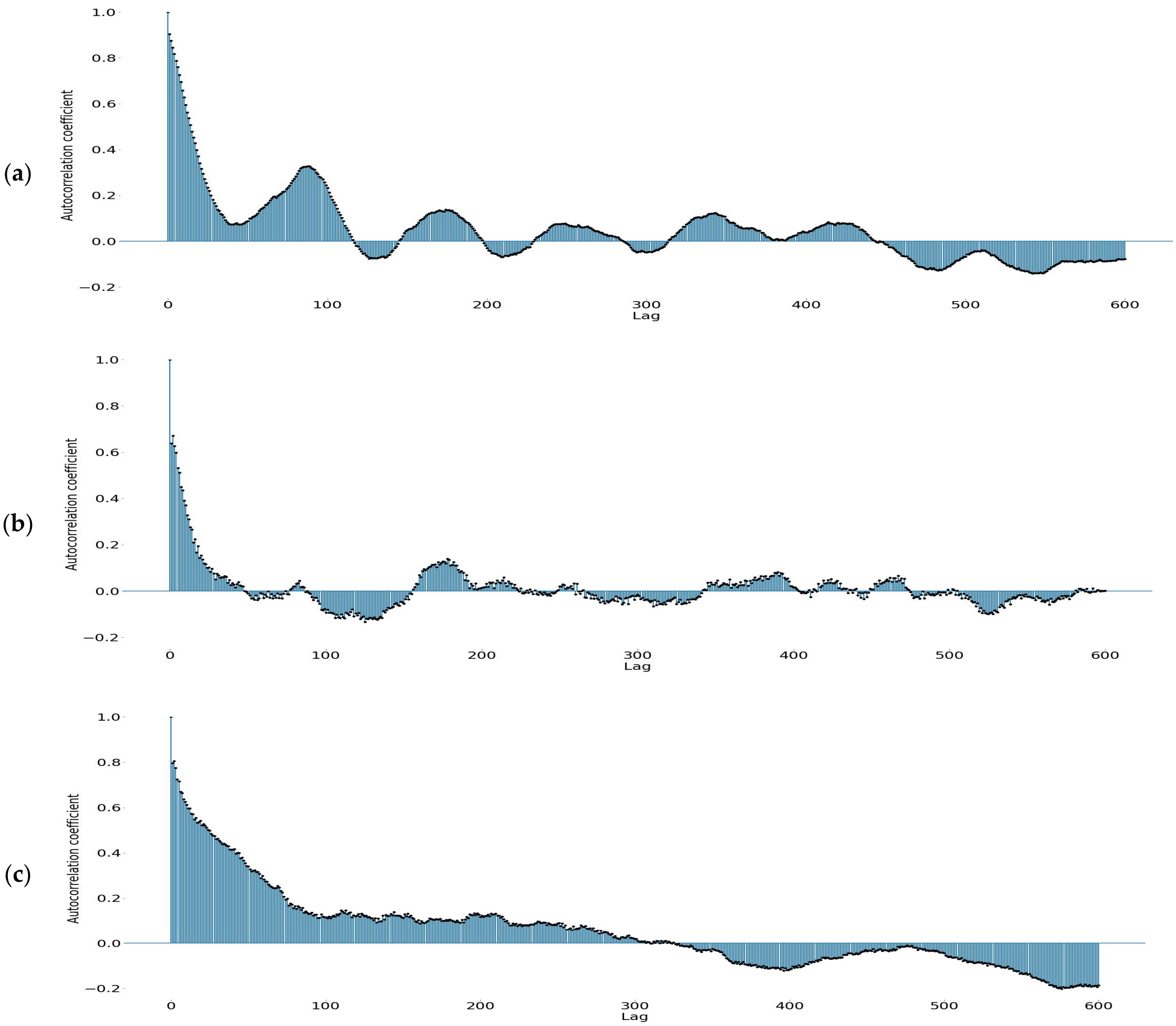


Figure 4. Autocorrelation graph of three selected cyclists on three different routes. (**a**) Cyclist 1, (**b**) Cyclist 2, and (**c**) Cyclist 3.

3.3. Forecast with One Variable

The LSTM-type RNN was designed to train with 70% of the data and the remaining 30% for validation. The initial calculations of the cyclist speed, where the network learns the behavior of the same cyclist, showed correlation coefficients ranging from 0.55 to 0.74 (Figure 5).

As mentioned, the research objective was to generate a methodology for estimating a cyclist's speed in an urban area; therefore, this methodology was validated with data from other cyclists to analyze their behavior. Due to the uncertainty about the network performance, random tests were performed on the cyclist data, varying the number of training epochs. These variations were 100, 200, 300, and 600 epochs. The coefficient of determination improved (from 0.43 to 0.67) in some cases but decreased in others, so it was decided to perform the first approaches with 100 epochs.

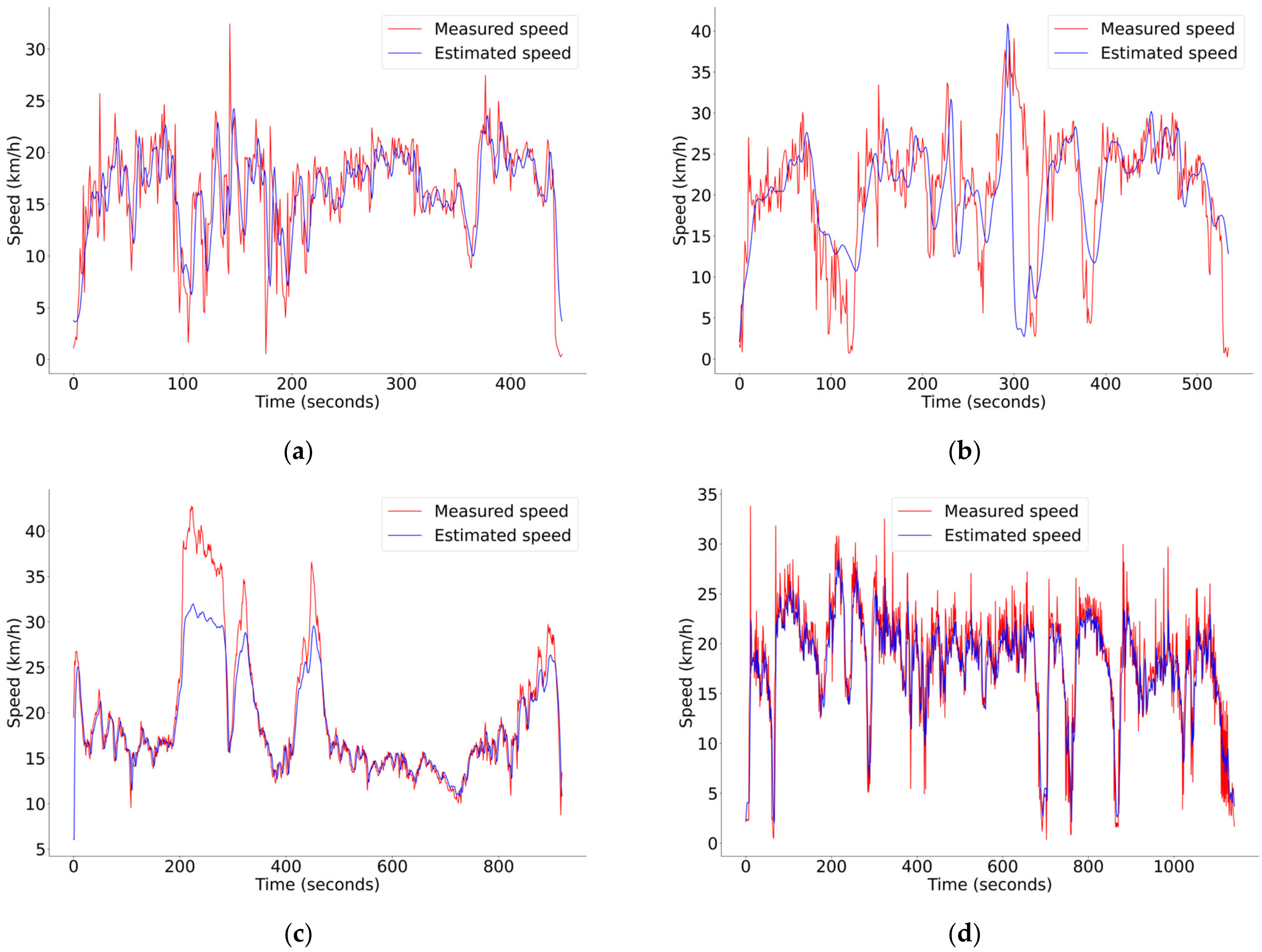


Figure 5. Recurrent neural network that learns the cyclist's behavior and predicts their speed with a coefficient of determination of 0.7. (**a**) 2–3 km, (**b**) 4–5 km, (**c**) 6–7 km, and (**d**) 8–9 km.

3.4. Validation

Table 2 shows that the LSTM-type base network can reliably predict with little historical data ranging from 5 to 20 s. The results showed that the historical time data did not significantly influence the model's prediction. The same happened with the batch size and the number of hidden neurons; the variation did not significantly affect the model estimation.

Table 2. Comparison of the performance of the network to predict the cyclist's speed, varying the batch size and the number of neurons in the hidden layers.

Previous Data	Neurons in the Hidden Layer	Batch Size	MAE	MSE	RMSE	MAPE	R^2
5	10	16	2.35	10.25	3.20	0.21	0.69
10	10	16	2.41	10.66	3.26	0.20	0.68
15	10	16	2.25	9.71	3.12	0.21	0.70
20	10	16	2.32	10.00	3.16	0.22	0.70
5	10	32	2.28	9.89	3.14	0.21	0.70
10	10	32	2.44	10.80	3.29	0.22	0.67
15	10	32	2.43	10.68	3.27	0.22	0.68
20	10	32	2.51	11.34	3.37	0.21	0.66
5	20	16	2.22	9.46	3.08	0.21	0.71

Table 2. *Cont.*

Previous Data	Neurons in the Hidden Layer	Batch Size	MAE	MSE	RMSE	MAPE	R^2
10	20	16	2.60	12.06	3.47	0.21	0.63
15	20	16	2.45	10.81	3.29	0.23	0.67
20	20	16	2.56	11.54	3.40	0.23	0.65
5	20	32	2.31	9.93	3.15	0.21	0.70
10	20	32	2.38	10.60	3.26	0.20	0.68
15	20	32	2.30	9.93	3.15	0.21	0.70
20	20	32	2.29	9.96	3.16	0.20	0.70
60	20	32	2.57	11.44	3.38	0.22	0.65
60	50	32	2.62	12.33	3.51	0.22	0.62

The results showed acceptable long-term predictions when the models were applied to data from other cyclists, as can be seen in the determination coefficient matrix in Figure 6, which shows ten models trained and applied to ten different routes. The matrix shows an average determination coefficient of 0.82, suggesting that the models estimated with one variable can represent the behavior of various cyclists, provided that there are 20 historical speed data points on the roads.

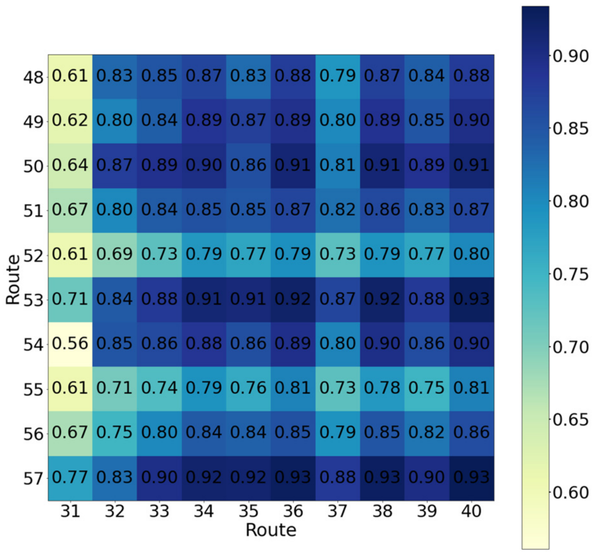


Figure 6. The matrix of determination coefficients between ten models was estimated using data from 10 cycling routes and applied to data from other cyclist routes using speed as previous data.

3.5. Forecast with Two Variables

The same scheme as the one-variable model was followed to obtain results with the two-variable model (70% of the data for training and 30% for validation). The results showed that the determination coefficients ranged from 0.77 to 0.96. Figure 7 shows the prediction result.

Moreover, taking the experience analyzed in the one-variable model, it was decided to use the two-variable model with 20 s of historical data for training. As in the one-variable model, the batch size, the number of hidden neurons, and historical data are insignificant. In that case, they also should not significantly influence the two-variable model, and this can be seen in Figure 8, where it can be observed that for the two-variable model, the determination coefficients improve in all cases. Finally, Figure 9 compares the forecast of the model with one variable and the model with two variables, showing the improvement in the prediction fit when using two input variables.

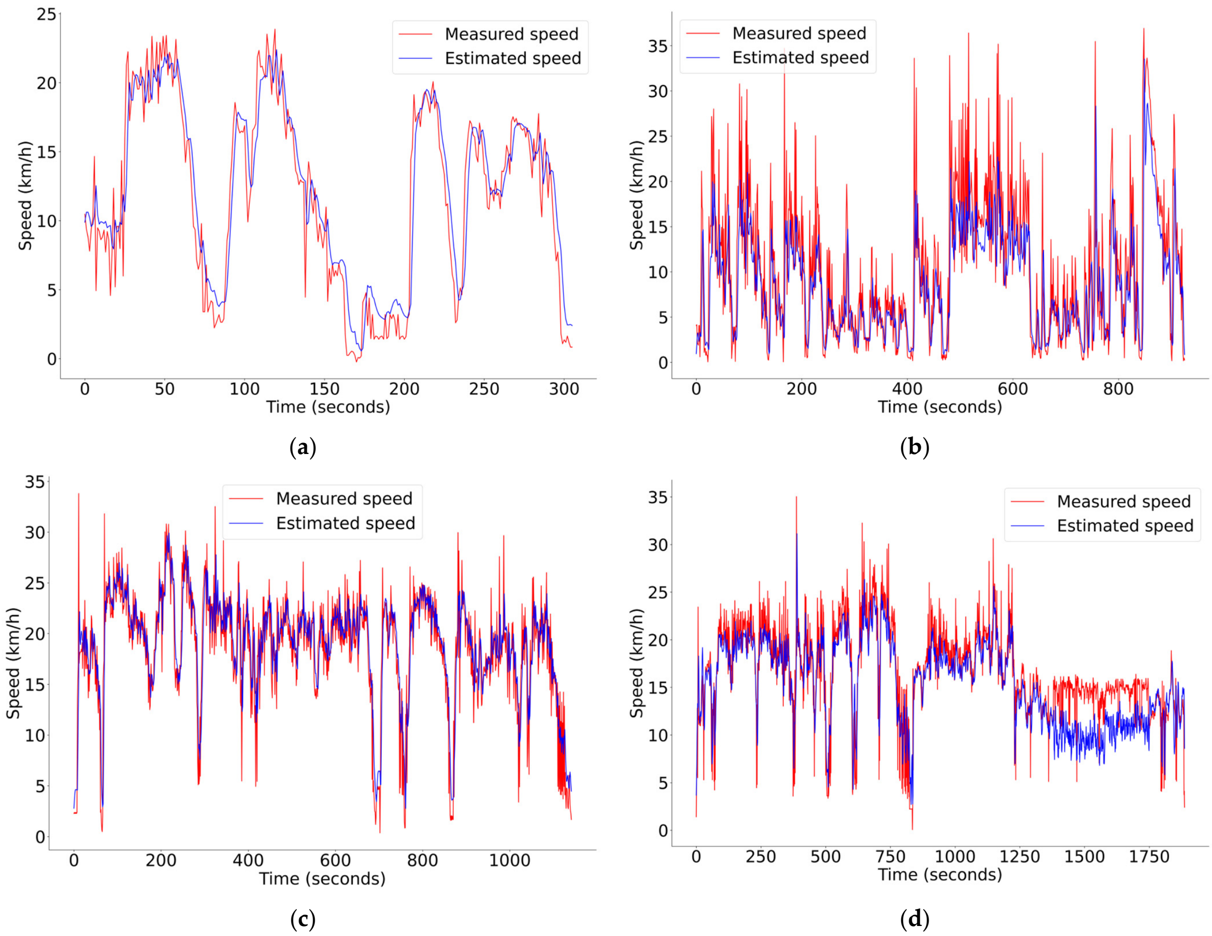


Figure 7. Prediction of cyclist's speed with a coefficient of determination of 0.96. (**a**) 2–3 km, (**b**) 4–5 km, (**c**) 6–7 km, and (**d**) 8–9 km.

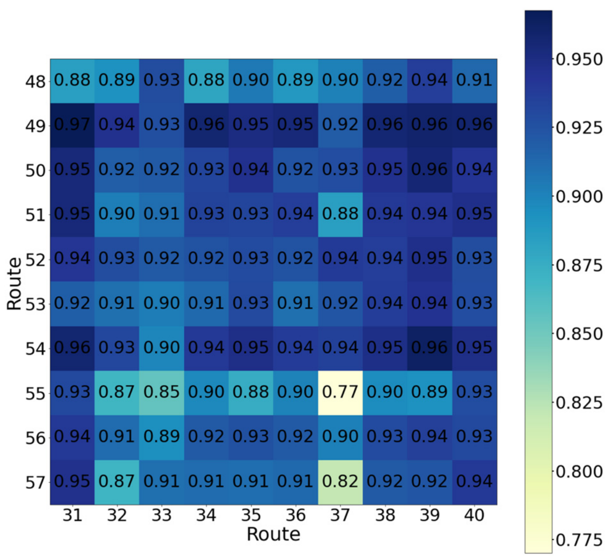


Figure 8. The matrix of determination coefficients between ten models was estimated using data from 10 cycling routes and applied to data from other cyclist routes using speed and slope as previous data.

Variations were also made in the historical data, batch size, and the number of hidden neurons to validate the model with two variables, finding no significant variation in the forecast results, as shown in Table 3.

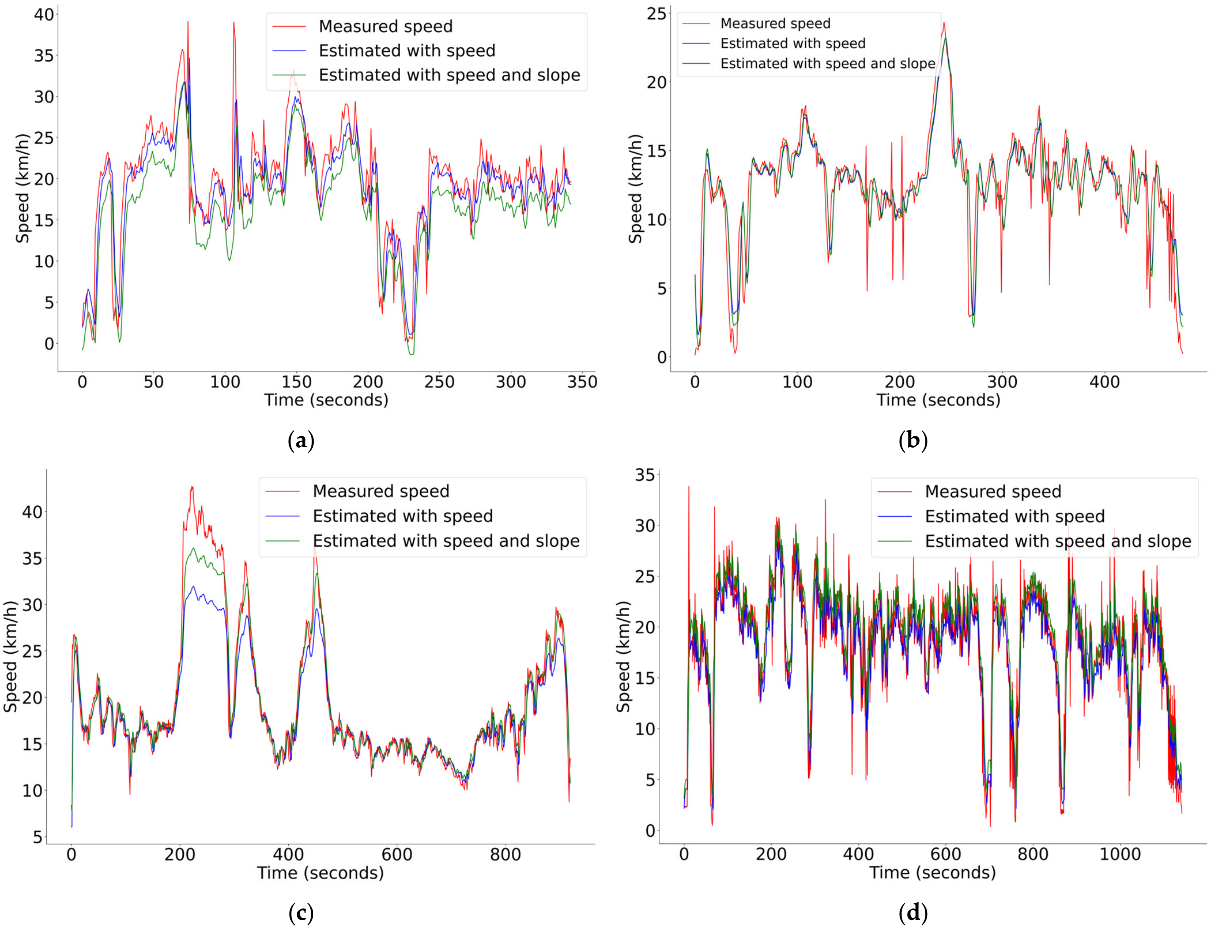


Figure 9. Cyclist speed prediction with one variable versus two variables. (**a**) 2–3 km, (**b**) 4–5 km, (**c**) 6–7 km, and (**d**) 8–9 km.

Table 3. Comparison of the performance of the network to predict the cyclist's speed, with two variables varying the batch size and the number of neurons in the hidden layers.

Previous Data	Neurons in the Hidden Layer	Batch Size	MAE	MSE	RMSE	MAPE	R^2
5	10	16	2.56	10.75	3.28	0.76	0.69
10	10	16	2.46	10.26	3.20	0.64	0.70
15	10	16	2.45	10.15	3.19	0.71	0.70
20	10	16	2.57	10.60	3.26	0.84	0.69
5	10	32	2.50	10.52	3.24	0.69	0.69
10	10	32	2.52	10.64	3.26	0.67	0.69
15	10	32	2.55	10.90	3.30	0.90	0.68
20	10	32	2.51	10.54	3.25	0.76	0.69
5	20	16	2.38	10.12	3.18	0.68	0.70
10	20	16	2.37	9.77	3.13	0.62	0.71
15	20	16	2.45	10.37	3.22	0.76	0.70
20	20	16	2.48	10.36	3.22	0.73	0.70
5	20	32	2.37	10.32	3.21	0.59	0.70
10	20	32	2.42	10.08	3.17	0.70	0.70
15	20	32	2.41	10.13	3.18	0.74	0.70
20	20	32	2.44	10.41	3.23	0.80	0.70
60	20	32	2.59	10.98	3.31	0.89	0.69
60	50	32	2.42	10.47	3.24	0.70	0.70

The results show that the estimated models can predict the speeds of several cyclists on different routes. These findings are consistent with the research of [32]. However, our study differs because it involves learning a neural network, which allows us to forecast more accurately and with only 20 s of historical data, as shown in Figure 10.

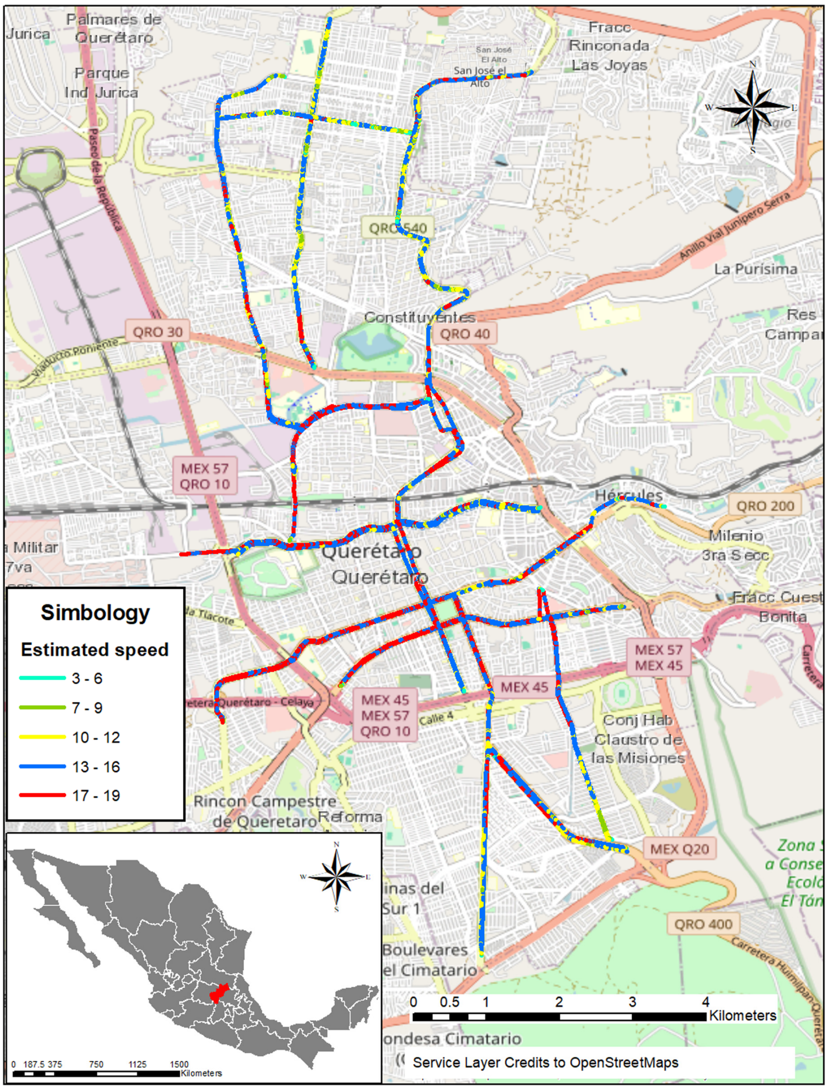


Figure 10. Using quantile classification, estimated speeds on different avenues in Querétaro are based on slope and the trained recurrent neural network.

The average speed cyclists observe on selected roads varies from 6.1 to 20.62 km/h, which is similar to the results presented by [27,33,35–38].

On average, women appear to choose routes with higher flat slopes (69.99% compared to 54.02% for men), which could explain why women have higher average speeds than men. The routes women select have smaller upward slopes than those chosen by men (22.59% vs. 40.14%), which could also explain why women achieve higher average speeds.

Despite these significant findings, it is essential to recognize some limitations of this study, such as the accuracy of the GPS devices, which could influence the generalizability of the results presented here. As well as, external conditions, such as weather and traffic, were not accounted for, which can significantly influence cycling speed.

Given the increasing emphasis on cycling as a sustainable mode of transportation, understanding and predicting cyclist behavior can help urban planning and infrastructure development. Hence, the rationale behind this research was to develop a methodology for accurately estimating cyclist speeds using minimal historical data in urban settings.

Using LSTM-type RNNs offers a sophisticated approach to modeling temporal patterns in speed data, enabling important traffic management and safety predictions. The findings contribute to the existing literature and present practical applications for improving urban mobility strategies.

4. Discussion

As can be seen, the trends in the predictions are the same when one or two variables are used with this methodology, so considering a single variable would be more convenient if it wants to have a lower computational cost.

Applying an LSTM-type RNN leverages deep learning capabilities, allowing for effective modeling of temporal patterns in cyclist speed data and a robust model design. Using different training epochs (100, 200, 300, and 600) shows a thorough approach to optimizing model performance, allowing to balance training time and accuracy. Hence, the results indicated that variations in batch size and the number of hidden neurons did not significantly impact predictions, suggesting the model's robustness to changes in these hyperparameters.

Furthermore, the models showed predictive solid performance, with determination coefficients ranging from 0.55 to 0.74 for one-variable forecasts and 0.77 to 0.96 for two-variable forecasts, indicating effective speed estimation due to the slope is a variable that directly affects speed, which coincides with the study by [39] for real-world studies. On the other hand, the ability to predict cyclist speeds using only 5 to 20 s of historical data suggests the model's efficiency and practical applicability in real-time scenarios. Previous speed states have been shown to influence their future state, as shown by [32], although they incorporate an advanced Markov model.

This model estimates speed based on the slope and enables the continuity of cyclists' routes through infrastructure connectivity, as suggested by [9,10]. It also anticipates the impact on vehicle speeds when cyclists' speed is shared with that of cars.

Although several studies have been conducted to determine the factors that affect or influence cycling speed, these have mainly focused on sociodemographic or socioeconomic characteristics [11,12]. Others have focused on determining how bicycle use affects, interacts, or influences vehicular traffic [14]. However, they sought to determine how the slope influences cycling speed to know the route choice a cyclist makes when traveling from one place to another. This will allow for better planning of the cycling infrastructure since, with this knowledge, they can select paths or roads with characteristics similar to those used by cyclists to build cycle paths, particularly in the city of Queretaro.

In this study, data were taken from cyclists' smartphone GPS, processed, and neural networks were applied to estimate cyclists' speeds on any route. This processing could be performed in real time using edge computing in navigation aids or safety alerts, as well as cloud computing that can handle large volumes of data from multiple cyclists, making it suitable for training neural network-based models, taking advantage of its computational power.

Unfortunately, implementing edge computing may require an initial investment in local hardware and software capable of running neural network models, which can be a barrier for some organizations. On the other hand, cloud services with a pay-per-use system can be cheaper to process large datasets without a significant initial investment.

Several factors must be considered to evaluate the cost–benefit of edge computing versus cloud computing when estimating cyclist speed using neural networks, such as the need for real-time processing, the volume and complexity of data, the training of complex models, infrastructure and resources, and long-term goals.

5. Conclusions

The study aimed to design a model to predict a cyclist's speed in an urban environment using recurrent neural networks. Historical data on slope and speed significantly influenced the prediction, more so than the one-variable model, since the speed forecasts from the one-variable model are different from the speeds of the model with two variables. Therefore, the slope has a significant influence on speed prediction.

The average coefficient of determination for one variable was 0.825. Using the same data and network configuration, the slope variable (combined with speed) increased to 0.921 in the forecast.

The mean absolute error decreased from 3.11 to 2.27 km/h from the one-variable to the two-variable model, which was considered acceptable for the forecast.

The number of hidden layers and neurons did not significantly influence the speed model's prediction. Nor did the optimizer or dropout.

Using smartphones for GPS data collection was sufficient for estimating speeds. In 96% of the cases, a model estimated for one cyclist could be applied to predict the speed of another cyclist even if the route was different.

The findings of this study contribute to understanding how RNNs can be used to predict a cyclist's speed and their potential application in urban bicycle lane planning. Further, the methodology was validated using data from various cyclists, enhancing the reliability and generalizability of the findings beyond individual behavior.

This study proposes a new way of predicting a cyclist's speed in an urban environment that would be difficult to model with traditional time series models such as autoregressive (AR), ARIMA, and SARIMA models.

This research can help urban planning and infrastructure development, improving cycling environments, and promoting cycling as a sustainable transportation option.

The low correlation coefficients in the results suggest that other variables, such as age or fatigue from traveled distances, should be considered in future work to estimate new models.

Future work should focus on applying the model to an urban network to broaden the effect of estimating the speed of average cyclists. This would consider the slope on the roads and generate maps of cyclist travel time based on the current slope and the speed and slope in previous states.

Author Contributions: Conceptualization, R.M.-Z.; methodology, R.M.-Z.; software, R.M.-Z.; validation, L.R.-G. and R.G.-C.; formal analysis, R.M.-Z., R.G.-C. and L.R.-G.; investigation, T.L.-L., J.B.H.-Z., R.M.-Z. and R.G.-C.; writing—original draft preparation, R.M.-Z. and R.G.-C.; writing—review and editing, T.L.-L., J.B.H.-Z., R.M.-Z. and R.G.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data used in this investigation can be found online: https://github.com/rmontyz/Cyclist-data.git (accessed on 1 October 2024).

Acknowledgments: The authors acknowledge the support of the two cycling groups, "Saca la Bici" and "Libre a Bordo," for their availability and participation in this study.

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

References

1. Stamatiadis, N.; Pappalardo, G.; Cafiso, S. Use of technology to improve bicycle mobility in smart cities. In Proceedings of the 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Naples, Italy, 26–28 June 2017. [CrossRef]
2. Tekouabou, S.C.K. Intelligent management of bike sharing in smart cities using machine learning and Internet of Things. *Sustain. Cities Soc.* **2021**, *67*, 102702. [CrossRef]
3. Behrendt, F. Cycling the smart and sustainable city: Analyzing EC policy documents on internet of things, mobility and transport, and smart cities. *Sustainability* **2019**, *11*, 763. [CrossRef]
4. Namiot, D.; Sneps-Sneppe, M. On bikes in smart cities. *Autom. Control Comput. Sci.* **2019**, *53*, 63–71. [CrossRef]
5. Li, L.; Correia, P.L.; Hadid, A. Face recognition under spoofing attacks: Countermeasures and research directions. *IET Biom.* **2018**, *7*, 3–14. [CrossRef]
6. Mizinov, P.V.; Konnova, N.S.; Basarab, M.A.; Pleshakova, E.S. Parametric study of hand dorsal vein biometric recognition vulnerability to spoofing attacks. *J. Comput. Virol. Hacking Tech.* **2023**, *20*, 383–396. [CrossRef]

7. Tome, P.; Marcel, S. On the vulnerability of palm vein recognition to spoofing attacks. In Proceedings of the International Conference on Biometrics (ICB), Phuket, Thailand, 19–22 May 2015; pp. 319–325. [CrossRef]
8. Guía Práctica Infraestructura Ciclista. Available online: https://changing-transport.org/wp-content/uploads/Guia-practica-infraestructura-ciclista.pdf (accessed on 22 May 2024).
9. McNally, D.; Tillinghast, R.; Iseki, H. Bicycle accessibility GIS analysis for bike master planning with a consideration of level of traffic stress (LTS) and energy consumption. *Sustainability* **2022**, *15*, 42. [CrossRef]
10. Parkin, J.; Rotheram, J. Design speeds and acceleration characteristics of bicycle traffic for use in planning, design and appraisal. *Transp. Policy* **2010**, *17*, 335–341. [CrossRef]
11. Jahanshahi, D.; Costello, S.B.; Dirks, K.N.; Chowdhury, S.; Wee, B.V. Understanding perceptions of cycling infrastructure provision and its role in cycling equity. *Transp. Res. Rec. J. Transp. Res. Board* **2022**, *2677*, 036119812211178. [CrossRef]
12. Barajas, J.M. Supplemental infrastructure: How community networks and immigrant identity influence cycling. *Transportation* **2018**, *47*, 1251–1274. [CrossRef]
13. Gössling, S.; McRae, S. Subjectively safe cycling infrastructure: New insights for urban designs. *J. Transp. Geogr.* **2022**, *101*, 103340. [CrossRef]
14. Pu, Z.; Li, Z.; Wang, Y.; Ye, M.; Fan, W. Evaluating the interference of bicycle traffic on vehicle operation on urban streets with bike lanes. *J. Adv. Transp.* **2017**, *2017*, 6973089. [CrossRef]
15. Saunier, N.; Chabin, V. Should I bike or should I drive? Comparative analysis of travel speeds in Montreal. *Transp. Find.* **2020**, 11900. [CrossRef]
16. Van Mil, J.F.P.; Leferink, T.S.; Annema, J.A.; Van Oort, N. Insights into factors affecting the combined bicycle-transit mode. *Public Transp.* **2020**, *13*, 649–673. [CrossRef]
17. Yan, X.; Chen, J.; Bai, H.; Wang, T.; Yang, Z. Influence factor analysis of bicycle free-flow speed for determining the design speeds of separated bicycle lanes. *Information* **2020**, *11*, 459. [CrossRef]
18. Louro, T.V.; Ubirajara, J.; Torresin, G.; Alberto, C. Factors influencing lateral distance and speed of motorized vehicles overtaking bicycles. *Transp. Res. Rec.* **2023**, *2677*, 51–61. [CrossRef]
19. Simović, S.; Ivanišević, T.; Trifunović, A.; Čičević, S.; Taranović, D. What affects the e-bicycle speed perception in the era of eeco-sustainable sobility: A driving simulator study. *Sustainability* **2021**, *13*, 5252. [CrossRef]
20. Sha, A.M.; Thankachan, A.; Hari, P.S.; Poudel, M.; George, G.A.; Thomas, T. Design of electrical bicycle using Matlab-Simulink. In Proceedings of the International Conference on Circuit Power and Computing Technologies (ICCPCT), Kollam, India, 10–11 August 2023; pp. 1038–1043. [CrossRef]
21. Gavalas, D.; Gerodimos, T.; Zaroliagis, C. Context-Aware bicycle route planning. In *Smart Energy for Smart Transport*; Springer: Cham, Switzerland; Volos, Greece, 2023; pp. 765–776.
22. De Jong, T.; Böcker, L.; Weber, C. Road infrastructures, spatial surroundings, and the demand and route choices for cycling: Evidence from a GPS-based mode detection study from Oslo, Norway. *Environ. Plan. B Urban Anal. City Sci.* **2022**, *50*, 239980832211414. [CrossRef]
23. Iuliano, J.E.; Keith, L. Near misses and split routes: Comparing rider behavior, driver interaction, and route choice for cyclists. *J. Transp. Saf. Secur.* **2022**, *15*, 1148–1171. [CrossRef]
24. Berghoefer, F.L.; Vollrath, M. Cyclists' perception of cycling infrastructure—A Repertory Grid approach. *Transp. Res. Part F Traffic Psychol. Behav.* **2022**, *87*, 249–263. [CrossRef]
25. Xu, C.; Li, Q.; Qu, Z.; Jin, S. Predicting free flow speed and crash risk of bicycle traffic flow using artificial neural network models. *Math. Probl. Eng.* **2015**, *2015*, 212050. [CrossRef]
26. Chatziioannou, I.; Bakogiannis, E.; Karolemeas, C.; Kourmpa, E.; Papadaki, K.; Vlastos, T. Urban environment's contributory factors for the adoption of cargo bike usage: A systematic literature review. *Future Transp.* **2024**, *4*, 92–106. [CrossRef]
27. Pazdan, S.; Kiec, M. Bicycle free-flow speed estimation based on GPS data–comparison of bikesharing system and Strava data. *Arch. Transp.* **2023**, *68*, 77–90. [CrossRef]
28. Barrero, G.A.; Rodriguez-Valencia, A. Asking the user: A perceptional approach for bicycle infrastructure design. *Int. J. Sustain. Transp.* **2021**, *16*, 246–257. [CrossRef]
29. Joo, S.; Oh, C.; Jeong, E.; Lee, G. Categorizing bicycling environments using GPS-based public bicycle speed data. *Transp. Res. Part C Emerg. Technol.* **2015**, *56*, 239–250. [CrossRef]
30. Camacho-Torregrosa, F.J.; Llopis-Castelló, D.; López-Maldonado, G.; García, A. An examination of the Strava usage rate—A parameter to estimate average annual daily bicycle volumes on rural roadways. *Safety* **2021**, *7*, 8. [CrossRef]
31. Ugan, J.; Abdel-Aty, M.; Cai, Q.; Mahmoud, N.; Al-Omari, M. Effect of various speed management strategies on bicycle crashes for urban roads in Central Florida. *Transp. Res. Rec. J. Transp. Res. Board* **2021**, *2676*, 544–555. [CrossRef]
32. Manum, B.; Arnesen, P.; Nordstrom, T.; Gil, J. Improving GIS-based models for bicycling speed estimations. *Transp. Res. Procedia* **2019**, *42*, 85–99. [CrossRef]
33. Allen, D.P.; Rouphail, N.; Hummer, J.E.; Milazzo, J.S. Operational analysis of uninterrupted bicycle facilities. *Transp. Res. Rec. J. Transp. Res. Board* **1998**, *1636*, 29–36. [CrossRef]
34. Opiela, K.S.; Khasnabis, S.; Datta, T.K. Determination of the characteristics of bicycle traffic at urban intersections. *Transp. Res. Rec. J. Transp. Res. Board* **1980**, *743*, 30–38.

35. Vagverk, S. Bicycle traffic facilities. In *Swedish Capacity Manual*; National Swedish Road Administration: Borlänge, Sweden, 1977; Volume 19.
36. Navin, F. Bicycle Traffic Flow Characteristics: Experimental Results and Comparisons. *ITE J. Inst. Transp. Eng.* **2014**, *64*, 31–37.
37. Liu, Y. The capacity of highway with mixture of bicycle traffic. In *Highway Capacity and Level of Service*, 1st ed.; Routledge EBooks: London, UK, 1991; pp. 253–257. [CrossRef]
38. Liu, X.; Shen, L.D.; Ren, F. Operational Analysis of Bicycle Interchanges in Beijing, China. *Transp. Res. Rec. 1396* **1993**, 18–21. Available online: https://onlinepubs.trb.org/Onlinepubs/trr/1993/1396/1396-004.pdf (accessed on 1 October 2024).
39. Arnesen, P.; Malmin, O.K.; Dahl, E. A forward Markov model for predicting bicycle speed. *Transportation* **2020**, *47*, 2415–2437. [CrossRef]
40. September, M.A.K.; Passino, F.S.; Goldmann, L.; Hinel, A. Extended Deep Adaptive Input Normalization for Preprocessing Time Series Data for Neural Networks. *arXiv* **2023**. [CrossRef]
41. Montenegro-Díaz, Á.M.; Montenegro-Reyes, D.M. Fundamentos de IA y AP. 2022. Available online: https://aprendizajeprofundo.github.io/Libro-Fundamentos/tutorial.html#span-style-color-4361ee-desarrollo-de-tecnologia-span (accessed on 13 May 2024).
42. Olah, C.; Understanding LSTM Networks. Github.io. 2015. Available online: https://colah.github.io/posts/2015-08-Understanding-LSTMs/ (accessed on 1 October 2024).
43. Chicco, D.; Warrens, M.J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [CrossRef] [PubMed]