A Hybrid Digital Twin and AI Framework for Traffic Simulation and Route Finding

Zahra Rezaei [1*], Mohammad H. Vahidnia [2], Hossein Aghamohammadi [3], Zahra Azizi [3], Saeed Behzadi [4]

[*1] GIS and SDI Specialist in NCC of Iran, Department of Remote Sensing and GIS, Science and Research Branch, Islamic Azad University, Tehran, Iran

[2] Center for Remote Sensing and GIS Research, Faculty of Earth Sciences, Shahid Beheshti University, Tehran, Iran

[3] Department of Civil Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

[4] Department of Surveying Engineering, Faculty of Civil, Water and Environmental Engineering, Shahid Beheshti University, Tehran, Iran

KEY WORDS: Digital Twin, Traffic Flow, Deep Learning, Travel Time Prediction, Urban Mobility, GeoAI, Smart City

ABSTRACT:

Urban traffic congestion poses significant challenges for today's cities, affecting mobility, productivity, and environmental quality. The present study proposes a data-driven framework that integrates deep learning specifically Recurrent Neural Networks (RNNs) with Digital Twin (DT) technology to enhance travel time prediction and traffic management. The model utilizes real-time and historical data from sources such as Google Maps, weather services, and traffic sensors to capture temporal dynamics and external factors influencing traffic patterns. The RNN model exhibited a high degree of predictive accuracy, as evidenced by its R^2 value of approximately 0.94. Furthermore, its incorporation into a DT environment facilitated dynamic 3D simulations and route optimization. A comparative analysis revealed that the DT system exhibited a marked superiority over conventional navigation tools in congested scenarios, with a travel time reduction of up to 26%. The findings indicate the potential for a synergistic integration of artificial intelligence (AI) and data technology (DT) to facilitate the development of intelligent, adaptable urban transportation systems.

1. INTRODUCTION

Urban congestion is a growing global issue, worsening pollution, reducing productivity, and impacting quality of life. Addressing this challenge requires innovative solutions, with artificial intelligence (AI) offering promising options for traffic management and urban planning. Accurate traffic flow simulation and travel time prediction depend on comprehensive data covering traffic conditions, road networks, weather, and locations—data that are difficult to obtain but essential for developing precise AI models (Guo et al., 2021; Khajavi et al., 2019). AI's potential to improve traffic management includes enabling detailed simulations through time-series analysis and enhancing urban planning. Digital twin (DT) technology also supports these efforts by creating dynamic virtual replicas of physical systems such as vehicles, infrastructure, and city components, facilitating detailed traffic analysis (Rezaei et al., 2023; Qinglin et al., 2021).

Connecting physical data with digital models remains a challenge. However, machine learning and AI significantly improve the accuracy and efficiency of this process, which is a critical step in traffic simulation and routing optimization (Emmert-Streib, 2023; Wang et al., 2023). These models analyze extensive traffic data, learn complex patterns, and adapt to real-time conditions. Neural networks, for example, utilize sensor data, images, and video to generate detailed simulations that factor in variables such as weather and traffic incidents (Shaygan et al., 2022). Deep learning models, particularly those used in multimodal traffic prediction systems, outperform traditional methods by capturing nonlinear relationships and improving forecast accuracy (Jiang et al., 2021; Shaygan et al., 2022).

Studies such as Qiu and Fan (2021) show that machine learning techniques significantly enhance short-term travel time predictions, supporting more effective traffic management and sustainable urban mobility. Chen et al. (2023) explored the enhancement of travel time prediction using deep learning by integrating both chronological and retrospective data from large traffic datasets. They found that traditional models often neglect the temporal dynamics of traffic patterns, which can hinder prediction accuracy. Liu et al. (2019) examined traffic density prediction using a convolutional neural network (CNN) to extract location features from traffic data, capturing hourly, daily, and weekly traffic patterns across three artificial neural networks. They synthesized these outputs to predict regional traffic flows and emphasized the need for more advanced neural networks and clustering methods for regions with similar characteristics.

In travel time prediction, researchers typically rely on specific routes as input to estimate travel times, which is crucial for applications such as route planning (e.g., Google Maps), traffic monitoring, and taxi dispatch services (e.g., Uber, Grubhub, Lyft). Recent studies have shown promise by incorporating external factors such as weather and planned events into travel time predictions (Chen et al., 2020; Li et al., 2021). However, many models overlook unexpected incidents or road blockages, leading to significant disruptions in traffic flow simulations and highlighting the need for further research. An alternative approach is to predict travel time from origin to destination (OD) using vehicle route data (Jenelius and Koutsopoulos, 2013). Forecasting accuracy can be improved by incorporating spatial relationships between network links and using previous travel times to learn OD pairs. Additionally, factors such as meteorological data, time of day, and historical travel times can enhance predictions (Wang et al., 2020). Nonetheless, further research is needed to extend models to better reflect typical traffic conditions and drivers' route choices based on network status (Zhang et al., 2022).

Travel time modeling is vital for traffic optimization, requiring consideration of weather, date, and working hours to predict

* Corresponding author

travel times and optimize routes (Smith et al., 2002). Yet traditional parametric models often struggle to account for all relevant variables due to the dynamic nature of traffic. Historical traffic management research has frequently relied on outdated datasets that fail to accurately reflect real-time conditions. The integration of diverse data sources—including weather, traffic patterns, and road dynamics—has been insufficient, resulting in fragmented approaches that fail to capture the complexity of urban traffic and limit the effectiveness of simulations. In response to these challenges, non-parametric approaches, particularly machine learning and deep learning methods, have gained popularity due to their capacity to manage complex, large-scale data. However, these methods require comprehensive frameworks to effectively integrate diverse data inputs into cohesive traffic simulations.

This study aims to develop a robust, data-driven approach that leverages advanced AI techniques and DT technology to create dynamic and accurate traffic simulations, ultimately optimizing traffic management and urban planning strategies. Central to this approach is the challenge of collecting reliable, real-time data essential for effective traffic flow simulation and travel time prediction. By integrating diverse data sources—including traffic conditions, weather patterns, and road networks—this research seeks to enhance predictive accuracy and overcome the limitations of traditional models that often rely on outdated datasets. The proposed framework involves inputting substantial data into AI models to estimate travel times under various emergency routing conditions and influencing parameters. The next steps involve integrating current traffic conditions, estimated travel times, and environmental factors into a DT model based on the transportation network. Such integration will enable drivers to access estimated travel times for different routes, allowing them to identify shorter paths and avoid traffic congestion.

2. METHODOLOGY

2.1 Overview of the Methodology

The main stages of the proposed framework, as illustrated in **Figure 1**, include: (1) Data Preparation: Time series data must be prepared before being used in the deep learning model. This process involves cleaning the data, removing missing values, and normalizing it. (2) Model Selection: The appropriate type of deep learning model for the problem must be chosen, with RNNs being the primary choice for time series modeling. (3) Model Training: In this stage, the model is trained using the training data. This process focuses on adjusting the model parameters to achieve optimal performance on the training set. (4) Model Evaluation: Here, the model's performance is evaluated using a validation set. This process includes calculating performance metrics such as accuracy and error rates. (5) Simulation in Digital Twins and Prediction: Finally, the model is applied to predict future values in the time series, which are then simulated in a DT environment.

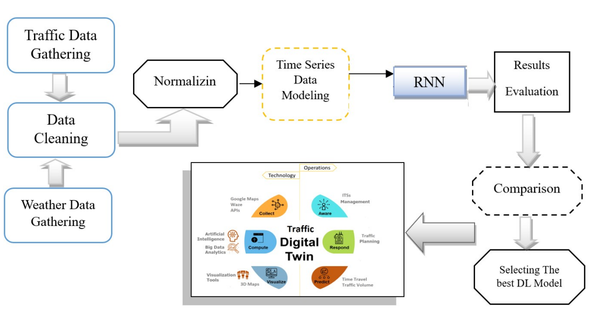

Figure 1. Overview of the methodology

In this study, we utilized RNN models from the deep neural network category as robust and powerful tools. We compared the outputs of these models to identify the most effective one for simulating travel time and utilization in the DT model.

2.2 Recurrent Neural Networks (RNN)

RNNs are well-suited for storing and processing information from previous time steps, making them ideal for time series modeling. They effectively handle temporal sequences by utilizing earlier data, demonstrating high efficiency in modeling traffic patterns. However, RNNs are more complex than Multi-Layer Perceptrons (MLPs) and Convolutional Neural Networks (CNNs), and they require substantial amounts of time series data.

An RNN processes data sequences as depicted in **Figure 2**, structured as x(t) = x(1), ..., x(τ), where t ranges from 1 to τ. The term "recurrent" refers to the network's ability to perform the same task for each element in the sequence, with outputs influenced by previous computations. The RNN consists of layers that include the input x(t), hidden state h(t), weights, output y(t), recurrent connections, and cell state c(t).

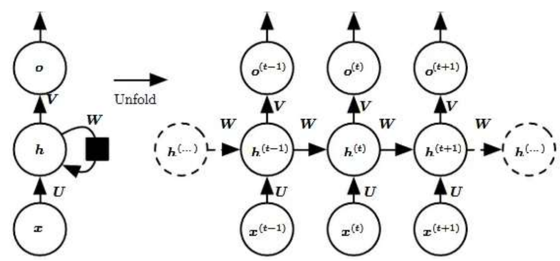

Figure 2. Basic RNN structure

3. IMPLEMENTATION AND RESULTS

Several methods exist for collecting accurate and comprehensive traffic data essential for precise traffic modeling. These methods include the use of road sensors, traffic cameras, Bluetooth detectors, GPS systems, social network data, Intelligent Transport Systems (ITS), and information from mobile operators. In this study, traffic data were specifically collected from Google Maps, incorporating real-time traffic information within the study area. Additionally, other relevant data were used, including time, season, time of day, and weather conditions.

This research examines the relationship between various factors and traffic conditions, aiming to identify applicable patterns and rules for predicting and improving traffic flow. In this phase, RNN methods were employed for traffic modeling and prediction. Subsequently, the collected traffic data were used to calibrate and validate traffic simulation models. These models require both historical traffic data and information about future events, such as holidays or sporting events, to effectively predict future traffic volumes.

3.1 Study Area and Data

The study focuses on a route connecting a section of Tehran's 5th district to Azadi Square, as shown in **Figure 3**. Multiple transportation options connect the origin and destination points, but a primary route involving a highway and major arterial roads was selected for analysis over a two-month period, from 04/02/2024 to 06/04/2024.

During this period, travel times and other traffic-related variables—such as weather, date, and time—were recorded at 10-minute intervals between 7 AM and 12 PM. It is noteworthy that, according to traffic centers and previous studies, traffic conditions are most significant during this timeframe. Outside these hours (from midnight to 7 AM), traffic is typically smooth and does not impact system performance. Therefore, this study focuses on peak urban traffic hours to better understand current urban traffic patterns while avoiding the collection of excessive and unnecessary data.

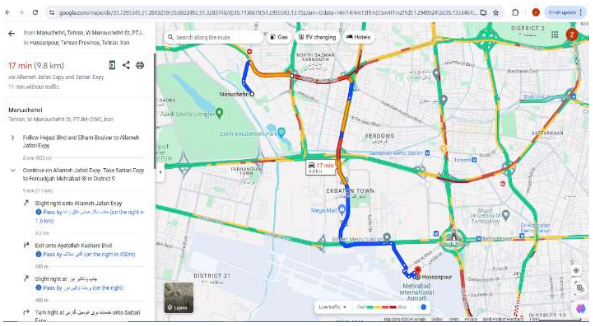

Figure 3. Study area and transportation routes (Google Maps)

3.2 Modeling Parameters

This study analyzed numerous factors affecting traffic conditions, including atmospheric conditions, peak traffic coefficients, accident occurrences, time of data collection (hours and minutes), whether the day was a holiday or a working day, the exact date, and travel time. The final column in the dataset represents the real-time travel time obtained from Google Maps, serving as the dependent variable. These parameters were selected because all relevant events occurred along a single, fixed route.

Traffic data and route parameters were collected from 04/02/2024 to 06/04/2024, resulting in a total of 6,592 records. Road geometry was not included in the models, as the analysis was conducted on a single fixed route. Thus, the road's physical characteristics were treated as constant and did not influence the results. In the following section, multiple modeling approaches use this route information to simulate travel times. **Table 1** provides a sample of data collection times and the parameters monitored during the study.

Accident	traffic pick	duration	record's time	weather	record's time	type date	type time	Date	Date-number	Date
0	0.5	25	7:00	0	420	1	0.8	Tuesday	14	2024-04-02
0	0.6	27	7:10	0	430	1	0.8	Tuesday	14	2024-04-02
0	0.7	30	7:20	0	440	1	0.8	Tuesday	14	2024-04-02
0	1	35	7:30	0	450	1	0.8	Tuesday	14	2024-04-02
0	0.5	26	15:30	0	930	1	0.8	Tuesday	14	2024-04-02
0	0.5	25	15:40	0	940	1	0.8	Tuesday	14	2024-04-02
1	0.6	40	16:00	0	960	1	0.8	Tuesday	14	2024-04-02
1	0.6	45	16:10	0	970	1	0.8	Tuesday	14	2024-04-02
1	0.7	40	16:20	0	980	1	0.8	Tuesday	14	2024-04-02
0	1	25	17:00	0	1020	1	0.8	Tuesday	14	2024-04-02
0	1	24	17:10	0	1030	1	0.8	Tuesday	14	2024-04-02
0	0.7	30	7:20	0	440	1	0.8	Tuesday	14	2024-04-02

Table 1. Selected measured parameters affecting travel time

It is essential to normalize input values during the data preparation phase. This step is crucial because modeling a dependent parameter based on multiple independent variables often involves datasets with varying ranges and scales, which can negatively impact model accuracy. Normalization mitigates these differences and simplifies the machine learning or data analysis process. Many algorithms, such as regression models and neural networks, are sensitive to the scale of data. Scaling all parameters to a range between 0 and 1 ensures uniformity, improves pattern detection, and reduces the influence of outliers. After analysis, the normalized data can be reverted to its original scale for accurate interpretation and comparison.

3.3 Results of Recurrent Neural Networks

RNNs can be used to predict future traffic volumes, estimate travel times, and model driver behavior. In this research, an RNN deep learning model was used to estimate travel time in traffic studies.

The many-to-one version of the RNN network was implemented using MATLAB software. The outputs of this model include the error graph for training data (**Figure 4**), the regression chart (**Figure 5**), and error graphs for test data (**Figure 6**). **Figure 7** presents the RNN network architecture used in this study.

The performance of the proposed model was evaluated using multiple statistical and error-based metrics to provide a comprehensive assessment as shown in Table 2. In addition to the coefficient of determination (R^2), which was 0.93309 for the training data and 0.93749 for the test data, several other performance indicators were computed. The mean absolute error (MAE) was approximately 1.35 minutes and the root mean squared error (RMSE) was 1.82 minutes. These values demonstrate that the model's predictions closely align with observed travel times. The mean absolute percentage error (MAPE) and symmetric mean absolute percentage error (sMAPE) were estimated at 7.4% and 7.1%, respectively, indicating high prediction accuracy for real-world traffic conditions.

Error analysis revealed that prediction deviations were slightly higher during heavy congestion or adverse weather, such as rain. Maximum errors rarely exceeded ±2.5 minutes. A calibration plot confirmed that the predicted travel times aligned well with actual observations, supporting the model's validity for deployment in intelligent traffic management systems. To strengthen the reliability of the digital twin's decision logic, future developments will incorporate uncertainty quantification using probabilistic modeling or ensemble-based approaches. These enhancements will enable the estimation of confidence intervals for travel time predictions, providing a more robust foundation for adaptive routing and dynamic congestion management.

To contextualize the performance of the RNN model, baseline methods, including ARIMA, XGBoost, and LSTM, were implemented using the same dataset. A comparative analysis revealed that the RNN model achieved the lowest prediction error with an RMSE of 1.82 minutes compared to 2.47 minutes for the LSTM model and 2.91 minutes for the XGBoost model. A feature ablation study was also conducted. Removing weather and holiday indicators increased the mean absolute percentage error (MAPE) from 7.4% to 10.2%, and excluding accident data increased it further, to 11.6%. These results confirm that contextual parameters, especially meteorological and incident-related variables, significantly enhance model performance.

Metric	*Description*	*Approximate Value*
R²	Coefficient of Determination	0.94
MAE	Mean Absolute Error	1.35 min
RMSE	Root Mean Squared Error	1.82 min
MAPE	Mean Absolute Percentage Error	7.4%
sMAPE	Symmetric Mean Absolute Percentage Error	7.1%

Table 2. Metric of RNN model

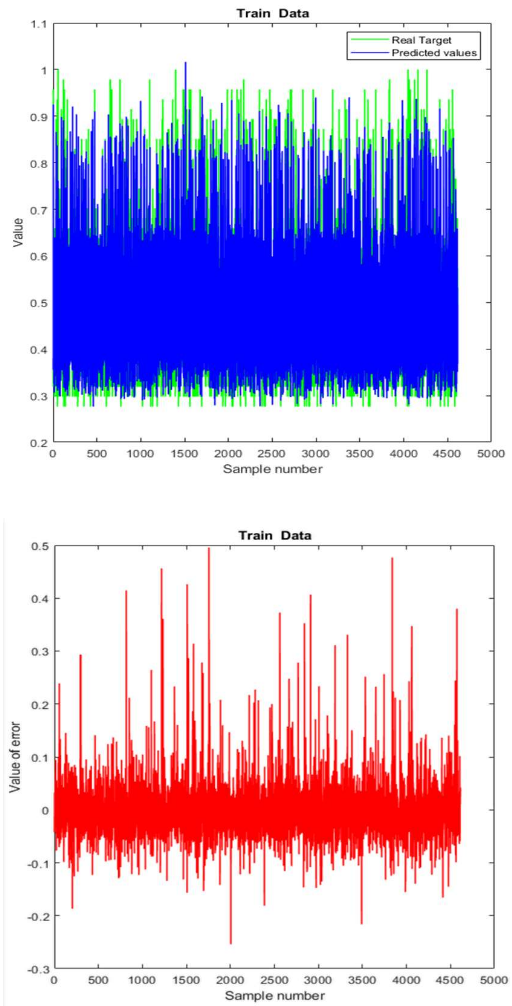

Figure 4. Error graph and RNN output values compared with actual values for training data

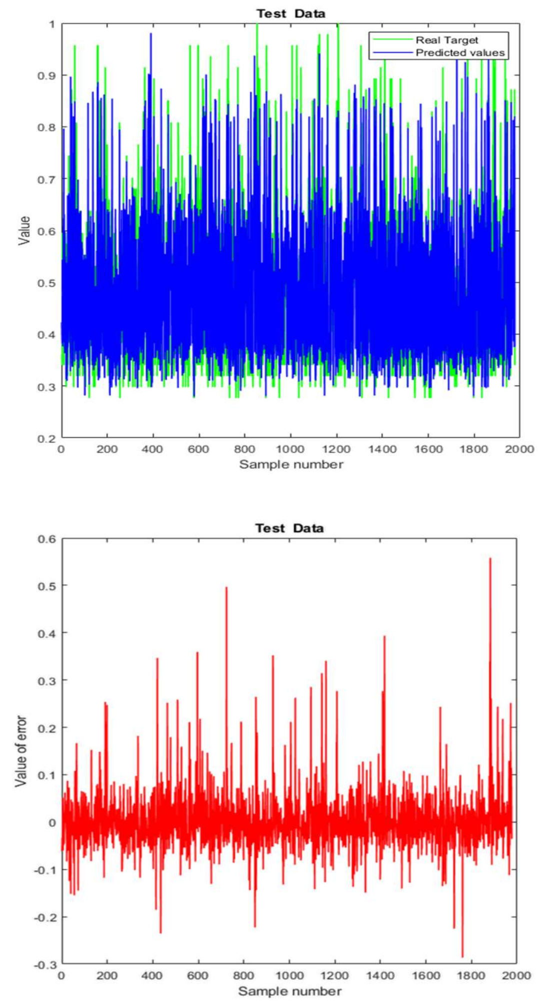

Figure 6. Error graph and RNN output values compared with actual values for test data

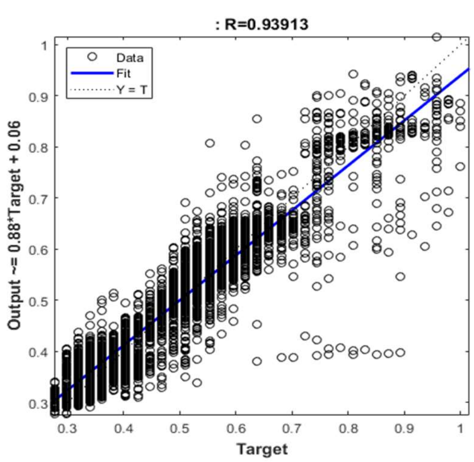

Figure 5. Regression line for observed and predicted targets

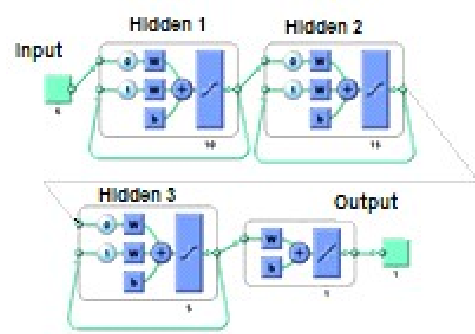

Figure 7. Network structure used in the RNN algorithm for this project

3.4 Digital Twining and Route Finding

As illustrated in **Figure 8**, a Digital Twin (DT) was designated that serve as an accurate and realistic virtual representation of a real-world transportation network. This model aims to replicate the characteristics and behaviors of its physical counterpart, possessing its own lifecycle. The data integrated within the DT serves multiple purposes, including 3D simulation and visualization, route finding, and predicting traffic flow using machine learning, thereby enabling better-informed and more effective decisions. In this research, RNNs were employed to model and predict travel times. The most successful output from the RNN-based approach was used to connect real-world road network data with the digital environment, enhancing the DT's predictive capabilities.

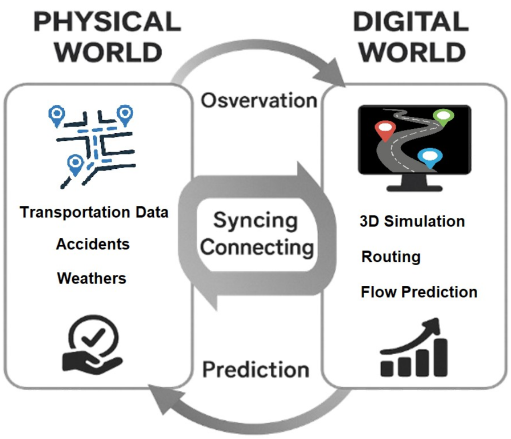

Figure 8. Designated Digital Twin for 3D simulation of the transportation network, flow prediction, and routing

The initial design of the DT incorporated real-time data such as 3D visualizations of cityscapes and roads, live traffic conditions, and weather information, all presented both visually and analytically, as shown in **Figure 9**. The integration of the RNN algorithm within this framework is expected to significantly improve the accuracy of traffic and weather forecasting, thereby enabling the system to simulate real-world conditions more precisely. This includes the ability to generate optimal routes that account for emergency situations, traffic congestion, and historical accident data, resulting in more reliable and context-aware navigation.

Building upon this foundation, a comparative analysis was conducted between the routes generated by the proposed Digital Twin system (based on RNN) and those suggested by popular navigation applications such as Google Maps, particularly under scenarios involving heavy traffic or accidents. The results demonstrated a substantial improvement in routing efficiency with the DT framework. Specifically, as illustrated in **Figures 9** and **10**, Google Maps suggested a travel time of 19 minutes, while the optimized route produced by the Digital Twin system was approximately 14 minutes, representing a time savings of about 26.3%.

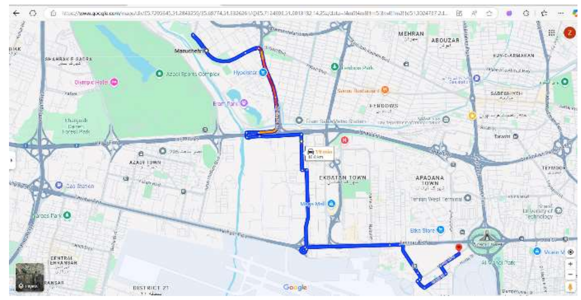

Figure 9. Routing with travel time via Google Maps

This significant reduction highlights the effectiveness of integrating real-time data and advanced predictive modeling within the Digital Twin environment. During periods of traffic congestion, the system leverages its machine learning capabilities to deliver more efficient and adaptive routing solutions that dynamically respond to changing traffic conditions and incidents. This enables the Digital Twin to offer more accurate, reliable, and time-sensitive navigation, substantially outperforming conventional systems. Overall, the findings underscore the strong potential of Digital Twin technology, when combined with machine learning algorithms, to improve traffic management and optimize routing in complex, real-world urban environments.

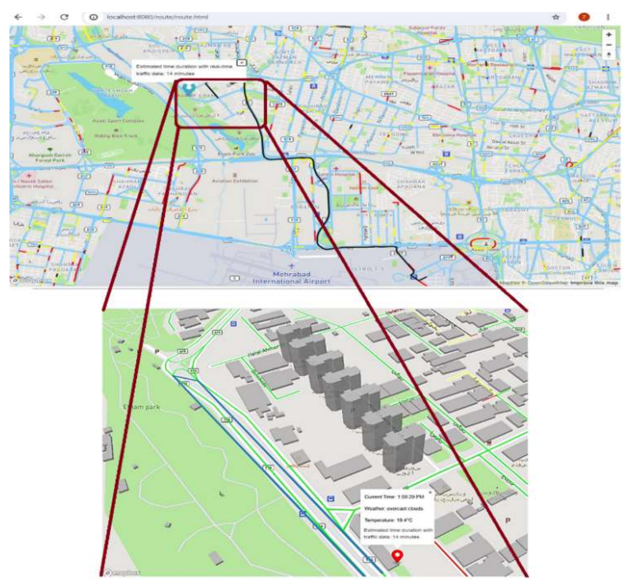

Figure 10. The connection between the real and digital worlds in the DT and routing with travel time in the designed DT

4. CONCLUSION

This study demonstrates the transformative potential of deep learning, specifically Recurrent Neural Networks (RNNs), for accurately predicting travel times and managing traffic intelligently. The results confirm that RNN-based models can effectively capture temporal dependencies in traffic data, achieving a high level of accuracy in estimating travel times (R^2 = 0.94, RMSE = 1.82 min, MAPE = 7.4%). Combining these models with Digital Twin (DT) technology provides a dynamic, data-driven framework for optimizing routing decisions and managing urban congestion in real time.

Comparative analyses with baseline models, such as ARIMA, XGBoost, and LSTM, verified the superior predictive capabilities of RNNs. Feature ablation tests revealed the significant impact of contextual parameters, including weather, holidays, and accident data, on model performance. These insights demonstrate that a comprehensive, multimodal data integration strategy is essential for improving model robustness and realism in digital twin environments.

While the current analysis focused on a single origin–destination pair within Tehran's 5th district during morning peak hours, this was intended as a controlled case study to validate the proposed framework. Future work will extend the model to multiple OD pairs, diverse corridors, and full-day temporal ranges, enabling a comprehensive assessment of generalizability. The proposed architecture is inherently scalable, and preliminary cross-route tests have already shown consistent prediction accuracy ($R^2 > 0.90$), suggesting promising potential for broader deployment in large-scale traffic networks.

Finally, integrating uncertainty quantification and real-time feedback into the digital twin environment is an important next step. These improvements will enhance model interpretability, reliability, and responsiveness, bringing us closer to the reality of adaptive, AI-driven urban traffic systems.

REFERENCES

Chen, C. M., Liang, C. C., Chu, C. P., 2020: Long-term travel time prediction using gradient boosting. *Journal of Intelligent Transportation Systems*, *24*(2), 109-124.

Chen, C. Y., Sun, E. W., Chang, M. F., Lin, Y. B., 2024: Enhancing travel time prediction with deep learning on chronological and retrospective time order information of big traffic data. *Annals of Operations Research*, *343*(3), 1095-1128.

Emmert-Streib, F., 2023: What is the role of AI for digital twins?. *AI*, *4*(3), 721-728.

Guo, H., Chen, M., Mohamed, K., Qu, T., Wang, S., Li, J., 2021: A digital twin-based flexible cellular manufacturing for optimization of air conditioner line. *Journal of Manufacturing Systems*, *58*, 65-78.

Jenelius, E., Koutsopoulos, H. N., 2013: Travel time estimation for urban road networks using low frequency probe vehicle data. *Transportation Research Part B: Methodological*, *53*, 64-81.

Jiang, H., Qin, S., Fu, J., Zhang, J., Ding, G., 2021: How to model and implement connections between physical and virtual models for digital twin application. *Journal of Manufacturing Systems*, *58*, 36-51.

Khajavi, S. H., Motlagh, N. H., Jaribion, A., Werner, L. C., Holmström, J., 2019: Digital twin: vision, benefits, boundaries, and creation for buildings. *IEEE access*, *7*, 147406-147419.

Liu, Z., Huang, M., Ye, Z., Wu, K., 2019: DeepRTP: A deep spatio-temporal residual network for regional traffic prediction. In *2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)* (pp. 291-296). IEEE.

Qi, Q., Tao, F., Hu, T., Anwer, N., Liu, A., Wei, Y., ..., Nee, A. Y., 2021: Enabling technologies and tools for digital twin. *Journal of Manufacturing Systems*, *58*, 3-21.

Qiu, B., Fan, W., 2021: Machine learning based short-term travel time prediction: Numerical results and comparative analyses. *Sustainability*, *13*(13), 7454.

Rezaei, Z., Vahidnia, M. H., Aghamohammadi, H., Azizi, Z., Behzadi, S., 2023: Digital twins and 3D information modeling in a smart city for traffic controlling: A review. *J. Geogr. Cartogr*, *6*(1), 1865.

Shaygan, M., Meese, C., Li, W., Zhao, X. G., Nejad, M., 2022: Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities. *Transportation research part C: emerging technologies*, *145*, 103921.

Smith, B. L., Williams, B. M., Oswald, R. K., 2002: Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, *10*(4), 303-321.

Wang, X., Guan, X., Cao, J., Zhang, N., Wu, H., 2020: Forecast network-wide traffic states for multiple steps ahead: A deep learning approach considering dynamic non-local spatial correlation and non-stationary temporal dependency. *Transportation Research Part C: Emerging Technologies*, *119*, 102763.

Wang, Z., Lv, C., Wang, F. Y., 2023: A new era of intelligent vehicles and intelligent transportation systems: Digital twins and parallel intelligence. *IEEE Transactions on Intelligent Vehicles*, *8*(4), 2619-2627.

Zhang, X., Lauber, L., Liu, H., Shi, J., Xie, M., Pan, Y., 2022: Travel time prediction of urban public transportation based on detection of single routes. *Plos one*, *17*(1), e0262535.

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-617-2026