

Peer Review

Review of: "Integrating Travel Behavior Forecasting and Generative Modeling for Predicting Future Urban Mobility and Spatial Transformations"

Yuyan (Annie) Pan¹

1. Pennsylvania State University, United States

Integrating Travel Behavior Forecasting and Generative Modeling for Predicting Future Urban Mobility and Spatial Transformations

The manuscript proposes a two-stage framework combining a Temporal Fusion Transformer (TFT) for travel behavior forecasting with a Generative Adversarial Network (GAN) to generate future satellite images of urban areas. The integration of behavioral and spatial modeling presents a novel and practical approach to sustainable infrastructure planning. While the methodology is promising, the paper would benefit from a clearer articulation of its contributions, enhanced validation, and a more in-depth discussion of its practical applicability.

1. The authors should more clearly differentiate their proposed method from existing GAN-based urban modeling and travel behavior forecasting work. What specific innovations are introduced by integrating TFT and GAN in this context? The contribution of this manuscript should be stated more specifically.
2. Have you considered incorporating exogenous variables to enhance the adaptability of your model?
3. The related work section could be better organized by themes, such as “urban spatial forecasting,” “behavioral modeling,” and “GAN applications in planning,” to improve clarity. Some references could be expanded, especially regarding recent advancements in transportation prediction for complex traffic environments. For example:

Li, A., Xu, Z., Li, W., Chen, Y., & Pan, Y. (2025). *Urban signalized intersection traffic state prediction: A*

spatial–temporal graph model integrating the cell transmission model and transformer. Applied Sciences, 15(5), 2377.

4. It is recommended to use consistent terminology. For instance, terms like “mobility trends” and “travel behavior” should be clearly defined and used consistently throughout the paper.
5. The resolution of Fig. 6 is low; it could be improved, and annotations could be added to highlight critical spatial or temporal insights.
6. The limitations of the model and the directions for future research should be elaborated more clearly in the article.
7. What unique temporal or feature-level attention patterns are captured by TFT that LSTM+Attention failed to learn? Were any interpretability modules of TFT (e.g., variable importance, attention maps) used to analyze key demographic drivers of the prediction? If so, visualization would be helpful.
8. The model is trained on California tracts. Could it generalize to other urban regions with different spatial characteristics or travel cultures (e.g., New York City, Dallas)? Was transfer learning or regional stratification considered?
9. The manuscript contains numerous instances of non-standard language usage, grammatical errors, redundant sentence structures, and awkward phrasing. A professional language editing service is strongly recommended to improve the academic clarity and rigor of the paper.

Declarations

Potential competing interests: No potential competing interests to declare.