Demographics-Informed Neural Network for Multi-Modal Spatiotemporal Forecasting of Urban Growth and Travel Patterns Using Satellite Imagery

Eugene K. O. Denteh[1], **Andrews Danyo**[1], **Joshua Asamoah**[1], **Blessing Agyei Kyem**[1], **Armstrong Aboah**[1]

[1]Department of Civil, Construction and Environmental Engineering,
North Dakota State University, Fargo, ND 58102
eugene.denteh@ndsu.edu, andrews.danyo@ndsu.edu, joshua.asamoah@ndsu.edu,
blessing.agyeikyem@ndsu.edu, armstrong.aboah@ndsu.edu

Abstract

Spatiotemporal forecasting of urban growth requires models that explain not only where change occurs but why, linking built form to demographic dynamics and mobility outcomes. However, most models treat these signals in isolation, relying on static projections. To address this problem, we present a Demographics-Informed Neural Network (DINN) that integrates multiyear satellite imagery with demographic data for spatiotemporal prediction of urban growth. The study also leverages these learned demographic-spatial representations for travel behavior forecasting. DINN couples a DenseNet-style image predictor with gated residual connections and a demographic encoder fused at the bottleneck; a separately pre-trained demographic predictor serves as a frozen consistency regularizer during training, while its encoder transfers to a travel-behavior head predicting 16 mobility features. A multi-objective loss balances image fidelity, demographic consistency, and semantic consistency. Using satellite images from 2012-2023 paired with county-level American Community Survey data, DINN improves image quality (SSIM ≈ 0.83) and demographic coherence (Demo-loss ≈ 0.14), achieves strong demographic prediction (overall $R^2 \approx 0.80$, $>$0.93 for core population metrics), and delivers accurate travel behavior forecasts (overall $R^2 \approx 0.91$). To validate the relevance of each architectural component in DINN, we conduct comprehensive ablation studies which effectively highlights the relevance of each model component. This study shows that the framework accurately forecasts spatiotemporal urban change and its associated demographics, linking where change occurs to its drivers and to resulting travel behavior.

1 Introduction

Spatiotemporal forecasting of urban growth requires an understanding of the fundamental relationships that exist between built environments, demographic characteristics, and travel behavior patterns [1–3]. These three components are fundamentally interconnected, as population dynamics drive land-use change, infrastructure development shapes residential patterns, and transportation networks structure mobility behaviors; yet conventional urban spatiotemporal analysis methods often treat them as isolated variables[4–6]. This fragmented approach has contributed to costly planning failures, exemplified by projects such as the construction of Interstate 95 through Miami's Overtown neighborhood, which displaced 10,000-12,000 residents and destroyed 40 blocks of established community infrastructure, resulting in substantial social disruption and economic losses[7]. Cur-

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

rent urban growth models predominantly rely on static projections that fail to capture the dynamic temporal relationships between demographic shifts, spatial transformations, and transportation patterns, limiting their effectiveness for sustainable development planning [8–10]. To address these limitations, this study introduces the Demographics-Informed Neural Network (DINN), a novel deep learning framework that integrates satellite imagery sequences with demographic data through an encoder-decoder architecture forecast future urban settings. DINN employs temporal gated residual connections to capture spatiotemporal dynamics, incorporates a frozen demographic predictor to enforce spatial-demographic consistency, and demonstrates transferability of learned representations through a travel behavior prediction network that utilizes frozen encoder weights.

2 Literature Review

Spatiotemporal urban growth studies range from cellular-automata models that simulate diffusion of land use [11–13] to deep spatiotemporal predictors that learn from image sequences using CNN–LSTM/ConvLSTM variants atop encoder–decoder backbones such as U-Net [14–16], yet most approaches append demographics rather than embed them in the representation, weakening cross-modal coherence. Building on evidence that built form strongly conditions travel demand and mode choice [17], recent mode-choice studies show that statistical and neural models with post-hoc explanation such as SHAP [18] outperform classical logit baselines [19, 20], yet they typically rely on tabular demographics and network summaries rather than spatial data, leaving the connection to imagery-derived representations underused. Moreover, recent multimodal fusion studies in remote sensing identify attention-based architectures as effective for combining heterogeneous inputs, while also noting persistent challenges in aligning modalities and enforcing cross-modal consistency [21]. Complementing this, studies that fuse sociodemographic attributes with satellite imagery via deep hybrid models show consistent gains when numeric and image features are embedded in a shared latent space, reinforcing the case for representation-level integration [22]. Therefore these studies highlight the need for a framework that (i) fuses imagery and demographics within the encoder–decoder, (ii) regularizes forecasts with demographic consistency, and (iii) transfers the learned demographic representation to travel-behavior prediction.

3 Problem Formulation

Despite the significant advancements in data-driven transportation planning methodologies, effectively modeling the dynamic and temporal relationship between geographic changes in satellite imagery and socio-demographic patterns remains a persistent challenge. This paper proposes the utilization of satellite image sequences and corresponding demographics to predict future spatial representations and also predict the future demographics and travel behavior. To effectively model the relationship between geographic changes in satellite images and demographic factors, we formalize the problem as follows. Specifically, let $\{x_{t-n}, \ldots, x_t\}$ denote a temporal sequence of $(n+1)$ historical satellite images, where each image $x_i \in \mathbb{R}^{H \times W \times C}$ comprises RGB channels, C with spatial dimensions $H \times W$ (height and width, respectively). Correspondingly, let $\{d_{t-n}, \ldots, d_t\}$ represent the associated demographic feature vectors, each $d_i \in \mathbb{R}^f$ containing f socio-demographic variables. The objective is to predict the future satellite image $\hat{x}_{t+1} \in \mathbb{R}^{H \times W \times C}$ and the future demographic vector $\hat{d}_{t+1} \in \mathbb{R}^f$ by using the temporal sequence of past satellite images, $\{x_{t-n}, \ldots, x_t\}$ together with their corresponding demographics, $\{d_{t-n}, \ldots, d_t\}$ as input into a network that predicts the future year. The general methodology adopted to achieve the objective is described in section 4. This study also leverages the strong relationship that exists between demographics and travel behavior [22, 23] to develop a travel behavior prediction network that leverages transfer learning by using a pretrained demographic feature encoder along with a specialized decoder to predict travel behavior.

4 Method

Demographics-Informed Neural Network (DINN): As displayed in Figure 1, DINN integrates satellite image sequences $\{x_{t-n}, \ldots, x_t\}$ (each image $x_i \in \mathbb{R}^{H \times W \times C}$) and the corresponding demographic feature vectors $\{d_{t-n}, \ldots, d_t\}$ (each vector $d_i \in \mathbb{R}^f$) through a DenseNet-based encoder-decoder with gated residual connections. The encoder extracts hierarchical spatial features $E = \{E_1, \ldots, E_7\}$, while the demographic sequence is embedded as $e_d \in \mathbb{R}^{512}$ and fused at

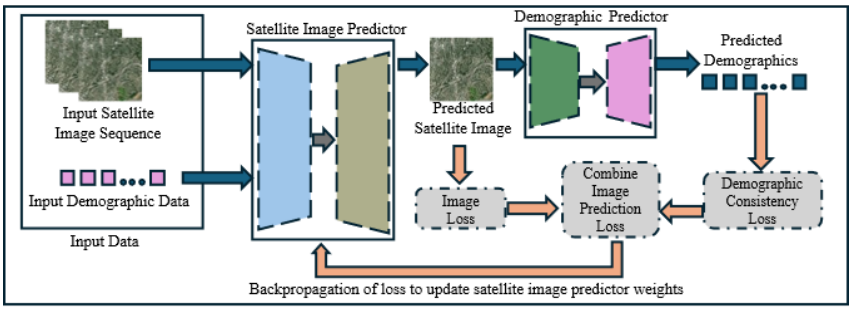

Figure 1: Demographics-Informed Neural Network (DINN)

the bottleneck as $E_{\text{fused}} = \text{Conv}_{1\times1}\big([E_{\text{bottleneck}}, E_d]\big)$, where $E_{\text{bottleneck}} \in \mathbb{R}^{H'\times W'\times C_b}$ and $E_d = \text{tile}(e_d; H', W') \in \mathbb{R}^{H'\times W'\times 512}$ is the spatial broadcast of e_d across the bottleneck height and width (H', W'). During decoding, gated residual connections $G_i = \sigma(f_\theta([E_i, \text{Up}(D_{i+1})]))$ regulate feature flow between encoder and decoder, where $\sigma(\cdot)$ denotes the sigmoid function, f_θ is a 1×1 convolution with batch normalization, and $\text{Up}(\cdot)$ represents bilinear upsampling. The predictor outputs the future satellite image $\hat{x}_{t+1}$, which is passed into a frozen demographic predictor pre-trained to map imagery into demographics, producing $\hat{d}_{t+1}$. Training minimizes the combined loss $\mathcal{L} = \alpha\mathcal{L}_{\text{image}}(\hat{x}_{t+1}, x_{t+1}) + \beta\mathcal{L}_{\text{demo}}(\hat{d}_{t+1}, d_{t+1})$, where $\mathcal{L}_{\text{image}}$ blends perceptual and SSIM terms for visual fidelity, $\mathcal{L}_{\text{demo}}$ is a scale-normalized mean squared error enforcing demographic alignment, and α, β weight their contributions. This design ensures that spatial predictions are both visually realistic and demographically consistent, with the frozen demographic predictor acting as the key innovation that constrains the model to cross-modal coherence.

Travel Behavior Prediction Network: The travel behavior prediction network builds on the demographic predictor by adopting its pre-trained encoder as a frozen feature extractor, thereby preserving robust spatial–demographic mappings while avoiding overfitting. Specifically, the frozen encoder $f_{\text{frozen}} : \mathbb{R}^{H\times W\times C} \to \mathbb{R}^{h'\times w'\times c'}$ provides fixed bottleneck features that encode both spatial context and demographic structure, ensuring that the downstream model inherits representations already aligned with population characteristics. A trainable decoder reconstructs transportation-specific patterns through transposed convolutions, while a global context pathway aggregates region-wide signals using fully connected layers to capture broader mobility trends. The outputs are concatenated and passed through a predictor head to produce the travel behavior vector $\hat{t}_{t+1} \in \mathbb{R}^m$, which includes attributes such as mode choice, vehicle availability, and travel times. Training minimizes the objective $\mathcal{L} = \mathcal{L}_{\text{travel}}(\hat{t}_{t+1}, t_{t+1}) + \gamma\mathcal{L}_{\text{semantic}}$, where $\mathcal{L}_{\text{travel}}$ is a scale-normalized mean squared error across all m travel variables and $\mathcal{L}_{\text{semantic}}$ preserves alignment between the decoder's embedding and the frozen encoder bottleneck. The key innovation lies in freezing the demographic encoder which constrains the model to reuse learned demographic representations, the network ensures that travel behavior forecasts remain demographically grounded while leveraging transfer learning to improve generalization.

5 Experiments and Evaluation Metrics

We evaluate DINN on a multimodal dataset (2012–2023) comprising satellite imagery, demographics, and travel behavior data for U.S. counties. Training used Adam (lr $= 3 \times 10^{-4}$, batch size 8) for 200 epochs on NVIDIA A100 GPU. Performance metrics include MSE for demographic/travel predictions, SSIM/PSNR for image fidelity, and R^2 for predictive accuracy. We also conduct ablation studies and change heatmap evaluation to assess temporal robustness and spatial consistency.

6 Results and Discussion

Demographic and Travel Behavior Predictor Performance: As shown in Figure 2, the demographic predictor demonstrates strong performance across most population categories, achieving R^2 values exceeding 0.90 for total and gender-specific populations, though comparatively lower accuracy

is observed for income- and inequality-related variables. This robust performance enables the demographic predictor to successfully enforce spatial-demographic consistency during training. Similarly, the travel behavior predictor achieves high predictive accuracy across most mobility attributes, with mode choice and vehicle availability well captured ($R^2 > 0.95$) but walking remaining the most challenging attribute due to higher variance in spatial distribution. These results demonstrate successful representation transfer from the frozen demographic encoder to behavioral forecasting, validating the framework's ability to leverage learned spatial-demographic relationships for cross-modal prediction tasks.

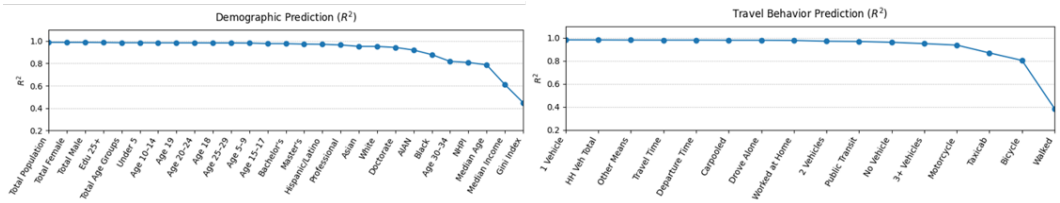

Figure 2: Line plots summarizing demographic predictor (Left) and Travel behavior Predictor (Right) test performance by feature.

Ablation Study Results: As illustrated in Figure 3, removing DenseNet blocks or gated skip connections increases missed-change regions and spurious artifacts in the predictions (revealed through the change heatmaps), while dropping the demographic predictor further disrupts demographic coherence. The quantitative trends in Table 1 mirror these effects with lower SSIM/PSNR and higher Demo-Loss for ablated variants, whereas the full DINN attains the best scores across all metrics, confirming each component's contribution. Beyond architectural validation, the framework successfully predicts subtle urban spatiotemporal changes that are typically difficult to identify through visual inspection alone, leveraging change heatmaps that compute pixel-level differences between temporal image sequences to quantify transformation patterns.

Table 1: Ablation study Results

Model Variant	Encoder Type	Gated Skip	Demographic Predictor	SSIM	PSNR (dB)	Demo-Loss
Baseline	2DConv	✗	✗	0.73	25.40	0.95
DINN-V1	2DConv	✓	✓	0.75	25.60	0.62
DINN-V2	Dense	✗	✓	0.78	25.77	0.33
DINN-V3	Dense	✓	✗	0.81	25.63	0.58
DINN	Dense	✓	✓	**0.83**	25.17	**0.14**

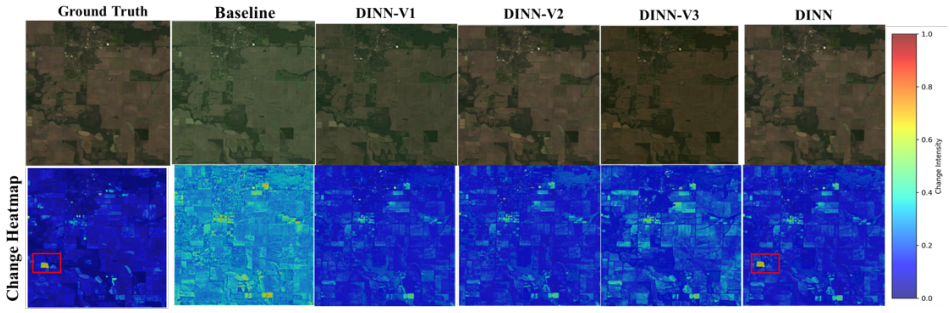

Figure 3: Qualitative ablation: top:inputs/predictions; bottom:change heatmaps. Red boxes highlight regions where ablations miss subtle changes; DINN preserves them.

7 Conclusion

This study demonstrates that integrating demographic context into spatiotemporal urban prediction significantly enhances prediction accuracy, with the proposed DINN framework achieving remark-

able performance and validating co-evolutionary theories linking built environments and population characteristics. The strong predictive accuracy across demographic and travel behavior patterns confirms that spatial configurations encode sufficient information to forecast urban dynamics, providing planners with a practical tool that explicitly models demographic-spatial relationships for more equitable development strategies. The framework's ability to predict both where urban change occurs and it's corresponding demographics (why it occurs) represents a significant advancement toward evidence-based planning that can mitigate costly retrofitting and infrastructure obsolescence. Future research should extend validation beyond US metropolitan areas to diverse international contexts where demographic-spatial correlations may differ substantially.

References

[1] Reid Ewing and Robert Cervero. Travel and the built environment. *Journal of the American Planning Association*, 76(3):265–294, 2010. doi: 10.1080/01944361003766766. URL https://doi.org/10.1080/01944361003766766.

[2] Michael Wegener. Overview of land use transport models. In *Handbook of transport geography and spatial systems*, pages 127–146. Emerald Group Publishing Limited, 2004.

[3] Michael J Clay. Developing an integrated land-use/transportation model for small to medium-sized cities: case study of montgomery, alabama. *Transportation planning and technology*, 33 (8):679–693, 2010.

[4] Robert Cervero and Kara Kockelman. Travel demand and the 3ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, 2(3):199–219, 1997. ISSN 1361-9209. doi: https://doi.org/10.1016/S1361-9209(97)00009-6. URL https://www.sciencedirect.com/science/article/pii/S1361920997000096.

[5] John Bates and Jan Oosterhaven. Review of land-use/transport interaction models. *Department of the Environment, transport and the Regions*, 1999.

[6] Ransford A Acheampong and Elisabete A Silva. Land use–transport interaction modeling: A review of the literature and future research directions. *Journal of Transport and Land use*, 8(3): 11–38, 2015.

[7] Raymond A Mohl. Interstating miami: Urban expressways and the changing american city.". *Tequesta*, 68:193–226, 2008.

[8] Maher Milad Aburas, Yuek Ming Ho, Mohammad Firuz Ramli, and Zulfa Hanan Ash'aari. The simulation and prediction of spatio-temporal urban growth trends using cellular automata models: A review. *International Journal of Applied Earth Observation and Geoinformation*, 52:380–389, 2016. ISSN 1569-8432. doi: https://doi.org/10.1016/j.jag.2016.07.007. URL https://www.sciencedirect.com/science/article/pii/S0303243416301143.

[9] Muhammad Asif, Jamil Hasan Kazmi, Aqil Tariq, Na Zhao, Rufat Guluzade, Walid Soufan, Khalid F. Almutairi, Ayman El Sabagh, and Muhammad Aslam. Modelling of land use and land cover changes and prediction using ca-markov and random forest. *Geocarto International*, 38(1):2210532, 2023. doi: 10.1080/10106049.2023.2210532. URL https://doi.org/10.1080/10106049.2023.2210532.

[10] J. Ronald Eastman and Jiena He. A regression-based procedure for markov transition probability estimation in land change modeling. *Land*, 9(11), 2020. ISSN 2073-445X. doi: 10.3390/land9110407. URL https://www.mdpi.com/2073-445X/9/11/407.

[11] Michael G McNally. The four-step model. In *Handbook of transport modelling*, volume 1, pages 35–53. Emerald Group Publishing Limited, 2007.

[12] Agung Wahyudi and Yan Liu. Cellular automata for urban growth modelling: A review on factors defining transition rules. *International Review for Spatial Planning and Sustainable Development*, 4(2):60–75, 2016.

[13] Anthony GO Yeh, Xia Li, and Chang Xia. Cellular automata modeling for urban and regional planning. In *Urban informatics*, pages 865–883. Springer, 2021.

[14] Wadii Boulila, Hamza Ghandorh, Mehshan Ahmed Khan, Fawad Ahmed, and Jawad Ahmad. A novel cnn-lstm-based approach to predict urban expansion. *Ecological Informatics*, 64: 101325, 2021. ISSN 1574-9541. doi: https://doi.org/10.1016/j.ecoinf.2021.101325. URL https://www.sciencedirect.com/science/article/pii/S1574954121001163.

[15] Jeong-Min Kim, JS Park, CY Lee, and SG Lee. Predicting of urban expansion using convolutional lstm network model: the case of seoul metropolitan area, korea. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10:113–118, 2022.

[16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[17] Robert Cervero and Kara Kockelman. Travel demand and the 3ds: Density, diversity, and design. *Transportation research part D: Transport and environment*, 2(3):199–219, 1997.

[18] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

[19] Li Tang, Chuanli Tang, Qi Fu, and Changxi Ma. Predicting travel mode choice with a robust neural network and shapley additive explanations analysis. *IET Intelligent Transport Systems*, 18(7):1339–1354, 2024.

[20] Victoria Dahmen, Simone Weikl, and Klaus Bogenberger. Interpretable machine learning for mode choice modeling on tracking-based revealed preference data. *Transportation Research Record*, 2678(11):2075–2091, 2024.

[21] Jiaxin Li, Danfeng Hong, Lianru Gao, Jing Yao, Ke Zheng, Bing Zhang, and Jocelyn Chanussot. Deep learning in multimodal remote sensing data fusion: A comprehensive review. *International Journal of Applied Earth Observation and Geoinformation*, 112:102926, 2022. ISSN 1569-8432. doi: https://doi.org/10.1016/j.jag.2022.102926. URL `https://www.sciencedirect.com/science/article/pii/S1569843222001248`.

[22] Qingyi Wang, Shenhao Wang, Yunhan Zheng, Hongzhou Lin, Xiaohu Zhang, Jinhua Zhao, and Joan Walker. Deep hybrid model with satellite imagery: How to combine demand modeling and computer vision for travel behavior analysis? *Transportation Research Part B: Methodological*, 179:102869, 2024.

[23] Xuedong Lu and Eric I. Pas. Socio-demographics, activity participation and travel behavior. *Transportation Research Part A: Policy and Practice*, 33(1):1–18, 1999. ISSN 0965-8564. doi: https://doi.org/10.1016/S0965-8564(98)00020-2. URL `https://www.sciencedirect.com/science/article/pii/S0965856498000202`.