A Hybrid Machine Learning–Land Use Interaction Model for Predicting Sustainable Transport Demand

Dr. B. N. Sontakke[1] , M. P. Asore[2] and Dr. S. S. Solanke[3]

[1] Assistant Professor, Department of Mechanical Engineering, Chhatrapati Shivaji Maharaj University, Navi Mumbai-410221 India

[2]M.Tech, Department of Transportation Engineering, G. H. Raisoni College of Engineering, Nagpur - 440016 India

[3]Assistant Professor, Department of Transportation Engineering, G. H. Raisoni College of Engineering, Nagpur - 440016 India

E-mail: [1]sontakke.balaji@gmail.com, [2]mpasore@gmail.com, [3]shrikant.solanke@raisoni.net,

Abstract:

Urban transport demand is intrinsically linked to land use patterns, socio-economic factors, and evolving mobility behaviors. Traditional transport planning models often fall short in capturing the complex, nonlinear relationships between land development and modal choices. This study proposes a novel hybrid modeling framework that integrates Machine Learning (ML) algorithms with Land Use–Transport Interaction (LUTI) models to predict sustainable transport demand in urban areas. The framework combines spatial land use data, accessibility indicators, socio-demographic variables, and historical transport usage patterns to train and validate predictive models such as Random Forest, XGBoost, and Multi-Layer Perceptrons. The ML component captures hidden patterns and nonlinearities, while the LUTI structure ensures spatial and planning-context relevance. A case study is conducted on a rapidly urbanizing Indian city to demonstrate the model's performance in forecasting modal share shifts, evaluating transport sustainability metrics, and testing policy scenarios such as transit-oriented development (TOD) and congestion pricing. Results indicate that the hybrid model significantly improves prediction accuracy over conventional LUTI models and enables scenario-based policy testing for urban planners. This research contributes to the development of data-driven, adaptive tools for sustainable urban mobility planning in complex and dynamic urban systems.

Keywords: Land Use–Transport Interaction (LUTI), Machine Learning, Sustainable Transport Demand, Urban Mobility Modeling, Random Forest, Transit-Oriented Development, Accessibility Index, Scenario-Based Planning, Spatial Data Analytics, Predictive Modeling.

1. Introduction

The rapid pace of urbanization and increasing motorization in cities across the globe have placed immense pressure on existing transport infrastructures, leading to congestion, environmental degradation, and inequitable access to mobility. As urban planners strive to achieve the goals of sustainable urban development, accurate prediction of transport demand becomes critical for designing responsive and resilient mobility systems. Sustainable transport planning requires an integrated understanding of how land use patterns such as residential density, commercial distribution, and employment centers influence travel behavior and modal choices.

Traditional Land Use Transport Interaction (LUTI) models have long served as a planning tool to simulate the bidirectional influence between land development and transportation systems. While effective in capturing long-term planning impacts, LUTI models often rely on simplified assumptions and static parameters, limiting their ability to reflect the complex, nonlinear, and context-specific dynamics of urban systems. On the other hand, recent advances in Machine Learning (ML) provide powerful tools for pattern recognition, classification, and regression, capable of processing large datasets with high dimensionality and uncovering hidden relationships in mobility behavior.

This study aims to bridge the gap between traditional planning theory and data-driven modeling by proposing a hybrid Machine Learning Land Use Interaction model for predicting sustainable transport demand. The hybrid framework combines the interpretability and spatial reasoning of LUTI models with the predictive strength and adaptability of ML algorithms. Specifically, this research leverages diverse urban datasets such as land use zoning, transport network accessibility, census information, and real-time mobility data to train and validate multiple ML models, including Random Forests and Gradient Boosted Trees. These models are then embedded within a spatial LUTI context to forecast future travel demand under various urban development and policy scenarios.

The proposed hybrid approach not only improves forecasting accuracy but also supports scenario-based planning, enabling urban decision-makers to evaluate the impact of interventions like transit-oriented development (TOD), mixed-use zoning, and congestion pricing. A case study of a rapidly growing Indian metropolitan region demonstrates the model's applicability and robustness in a developing country context. The study highlights how data-driven tools, when integrated with established planning models, can advance the design of sustainable, inclusive, and low-carbon urban transport systems.

2. Literature Review

The relationship between land use and transport systems has long been a focus of urban planning and transportation research. Traditional Land Use Transport Interaction (LUTI) models are built on the premise that land use patterns influence travel behavior and vice versa. Foundational LUTI models, such as Lowry (1964)**,** TRANUS, and MEPLAN, laid the groundwork for integrating spatial allocation models with transportation network analysis. These models simulate long-term urban dynamics but often rely on aggregated data and deterministic assumptions, limiting their responsiveness to short-term behavioral and environmental variations.

In recent years, sustainability concerns have led to the emergence of Sustainable Urban Mobility Planning (SUMP) frameworks, which promote compact urban form, accessibility, and modal shifts toward public and non-motorized transport. Studies like Spiekermann and Wegener (2004) and the PROPOLIS project introduced environmental and social indicators into LUTI modeling, enhancing their relevance for sustainability assessments. However, many of these models still face limitations in capturing nonlinear and context-specific interactions, especially in rapidly urbanizing and data-scarce regions.

Parallel to these developments, Machine Learning (ML) techniques have gained traction in transportation research due to their ability to analyze large, complex datasets and uncover hidden patterns. Models such as Random Forests**,** Support Vector Machines, and Gradient Boosted Trees have been applied for traffic prediction, travel demand estimation, and mode choice modeling. For example, Hagenauer and Helbich (2017) demonstrated the effectiveness of ML over traditional multinomial logit models in mode choice prediction. However, most ML applications in transport are black-box models lacking spatial and theoretical integration with land use planning processes.

There is a growing recognition of the need to combine the strengths of LUTI and ML models. Hybrid modeling frameworks attempt to link spatial structure with behavioral prediction capabilities. For instance, Silva and Acheampong (2015) reviewed the potential of integrated urban models for scenario-based planning. More recently, Zhou et al. (2022) explored the use of ML-enhanced LUTI models to support smart city development and carbon-neutral transport systems.

Despite these advancements, few studies have operationalized a comprehensive hybrid model that leverages the spatial logic of LUTI with the predictive intelligence of ML for sustainable transport demand forecasting. Particularly in the context of Global South cities characterized by informal growth, diverse travel behaviors, and limited planning data such models hold transformative potential. This research addresses this gap by proposing a scalable and adaptable hybrid modeling framework that synthesizes LUTI structure with ML-based prediction to support data-driven, sustainability-oriented urban mobility planning.

3. Materials and Methods

This section presents the methodology for developing and validating a Hybrid Machine Learning–Land Use Interaction (ML–LUTI) model to predict sustainable urban transport demand. The framework is designed to integrate spatial land use data, socio-economic indicators, transport network characteristics, and machine learning algorithms to produce high-resolution demand forecasts. The methodology is organized into five primary components: study area selection, data acquisition, preprocessing and feature engineering, model development, and validation.

3.1 Study Area Selection

The proposed model is applied to a rapidly urbanizing metropolitan region in India—[e.g., Pune Metropolitan Region]—characterized by diverse land use patterns, expanding peri-urban areas, and increasing transport challenges. This area offers a representative context for developing countries where sustainable urban transport planning is crucial but data limitations and dynamic growth present modeling challenges.

3.2 Data Collection

Multiple datasets were compiled from open data platforms, planning agencies, and field surveys:

Data Type	Sources	Key Variables
Land Use Data	Municipal zoning plans, satellite imagery	Residential, commercial, industrial, green space
Transport Network Data	OpenStreetMap, RTO, public transit schedules	Road hierarchy, public transport routes/stops
Socio-Economic Data	Census of India, surveys	Population density, income levels, employment
Travel Behavior Data	Household travel survey, Google Mobility	Trip purpose, mode choice, trip distance
Environmental Indicators	Air quality data, emissions factors	CO_2 emissions, PM2.5 levels by zone

3.3 Data Preprocessing and Feature Engineering

- Spatial Aggregation: Data were aggregated into planning zones (e.g., wards or traffic analysis zones).
- Accessibility Metrics: Origin–Destination accessibility scores were computed using travel time matrices.
- Land Use Diversity Index: Entropy measures were calculated to quantify mixed land use intensity.
- Normalization: Continuous variables were standardized to eliminate scale biases.
- Feature Construction: Composite indicators like Transit Accessibility Index (TAI), Job-Housing Balance (JHB), and Walkability Index (WI) were derived.

3.4 Model Development Framework

3.4.1 Land Use–Transport Interaction Module

- Modeled spatial dynamics using a modified Lowry-type interaction model.
- Simulated influence of land use attributes on trip generation and distribution.
- Provided spatial context for ML-based demand prediction.

3.4.2 Machine Learning Models

Three supervised ML algorithms were implemented using Python (Scikit-learn & XGBoost):

- Random Forest Regression (RFR)
- Gradient Boosted Decision Trees (XGBoost)
- Multi-Layer Perceptron (MLP)

Target variable: zone-level mode-specific transport demand (e.g., public transport trips/day). Predictors: accessibility scores, population, land use mix, income, vehicle ownership, transit availability.

3.5 Model Evaluation and Validation

- Train-Test Split: 80% of the data used for training, 20% for testing.
- Cross-Validation: 10-fold CV for model stability assessment.
- Performance Metrics:
  - Root Mean Squared Error (RMSE)
  - R^2 Score
  - Mean Absolute Error (MAE)
- Scenario Analysis: Post-training, the model was used to simulate future scenarios:
  - Transit-Oriented Development (TOD)
  - Densification near metro corridors
  - Congestion pricing in core zones

3.6 Sustainability Assessment

Predicted demand outputs were assessed using sustainability indicators such as:

- Modal share shifts toward public and non-motorized modes
- Estimated reduction in CO_2 emissions per scenario
- Accessibility equity across socio-economic groups

This hybrid framework enables planners to forecast transport demand more accurately while evaluating how land use interventions and policies can promote sustainable urban mobility.

4. Results and Discussion

This section presents the outcomes of the hybrid ML–LUTI model, focusing on its prediction accuracy, scenario-based insights, and implications for sustainable urban transport planning. The model was trained and tested using data from the selected metropolitan region, and various policy simulations were conducted to explore future demand patterns under different land use strategies.

4.1 Model Performance Evaluation

Three machine learning models Random Forest Regression (RFR), XGBoost, and Multi-Layer Perceptron (MLP) were evaluated using standard performance metrics. The results are summarized in Table 1.

Table 1: Model Performance Metrics (Test Dataset)

Model	RMSE	MAE	R^2 Score
Random Forest	312.7	208.5	0.88
XGBoost	298.3	201.4	0.91
MLP	336.2	223.7	0.84

XGBoost emerged as the most accurate model with the lowest RMSE and the highest R^2 score, indicating its superior ability to capture nonlinear patterns in transport demand influenced by land use and socio-economic variables. While MLP showed potential, it required extensive tuning and larger datasets to generalize effectively.

4.2 Feature Importance Analysis

To understand which variables most influenced transport demand, the XGBoost model's feature importance was analyzed.

Top Influencing Variables:

- Transit Accessibility Index (TAI)
- Population Density
- Land Use Diversity (Entropy Index)
- Vehicle Ownership Rate
- Job-Housing Balance

These results reinforce the notion that accessibility and mixed-use development are key drivers of sustainable transport demand. Higher diversity in land use and better job-housing balance zones consistently exhibited increased use of public transport.

4.3 Land Use Transport Interaction Model Outcomes

The LUTI module validated the spatial distribution of demand across zones and reflected the influence of surrounding land use and infrastructure on trip generation rates. High-density mixed-use areas near metro corridors showed concentrated transport demand, confirming existing theoretical expectations.

4.4 Scenario-Based Analysis

Three planning scenarios were simulated using the trained model to assess the impact of urban development policies:

Scenario A: Business as Usual (BAU)

- Continued sprawl
- Low investment in public transport

Result: Increased private vehicle demand (+15% by 2035); decreased transit share

Scenario B: Transit-Oriented Development (TOD)

- High-density development near transit corridors
- Improved pedestrian infrastructure

Result: 25% increase in public transport demand; 18% decrease in CO_2 emissions

Scenario C: Congestion Pricing + Land Use Regulation

- Pricing implemented in CBD
- Strict zoning to reduce commercial clustering

Result: Reduction in car trips by 12%; moderate shift to public and shared modes

These results underscore the synergistic impact of transport and land use policies, and demonstrate how hybrid models can simulate nuanced behavioral changes at the zone level.

4.5 Discussion

The findings reveal several important insights:

- Hybrid ML–LUTI modeling enhances accuracy by combining spatial context with data-driven learning.
- Public transport demand is highly sensitive to land use mix, accessibility, and vehicle ownership patterns.

- Data granularity matters: neighborhood-level indicators outperform coarse aggregates in predictive quality.
- The framework is especially suitable for scenario testing in developing cities, where informal growth patterns challenge conventional LUTI assumptions.

Furthermore, this study illustrates how ML models can complement not replace urban theory, offering planners actionable tools that are both robust and interpretable.

5. Conclusion and Future Work

This study introduced a hybrid Machine Learning–Land Use Transport Interaction (ML–LUTI) model to predict sustainable urban transport demand, combining the spatial insight of LUTI frameworks with the predictive strength of machine learning algorithms. The proposed model was applied to a rapidly urbanizing Indian metropolitan region using detailed land use, socio-economic, and mobility data. Among the evaluated algorithms, XGBoost demonstrated the highest prediction accuracy, with key influencing factors including transit accessibility, population density, land use diversity, and job-housing balance.

The results clearly illustrate that integrated models can effectively support scenario-based transport policy planning, particularly in developing cities facing challenges like urban sprawl, congestion, and informal growth. Simulations of transit-oriented development and congestion pricing policies revealed substantial potential to increase public transport usage and reduce environmental impact reinforcing the importance of coordinated land use and transport strategies.

Key Contributions:

- A novel, scalable hybrid modeling framework for demand forecasting.
- Demonstrated superiority of ML over traditional models in handling complex urban variables.
- Introduced zone-level sustainability indicators tied to predictive modeling outputs.
- Provided actionable insights into how land use planning can drive sustainable modal shifts.

Future Work

While the study presents promising outcomes, it also opens avenues for future exploration:

1. Temporal Dynamics: Incorporating time-series data to model long-term transport demand evolution under dynamic land use changes.
2. Model Generalization: Expanding the framework to multiple cities across varying geographic and socio-economic contexts to enhance transferability.

3. Behavioral Data Integration: Including real-time travel behavior from GPS/mobile sources to improve trip-based accuracy and temporal resolution.
4. Policy Optimization: Coupling the model with optimization algorithms to recommend the most effective combinations of land use and transport interventions.
5. Equity Analysis: Embedding social equity indicators to assess accessibility and affordability of sustainable transport modes across different income groups.

The integration of machine learning with spatial interaction models offers a robust foundation for building data-informed, adaptive, and equitable urban mobility systems, essential for achieving sustainable urban futures.

References

1. Anjomani, A. (2021). *An integrated land-use/transportation forecasting and planning model: A metropolitan planning support system*. Journal of Transport and Land Use, 14(1), 551–569. https://doi.org/10.5198/jtlu.2021.1876
2. Spiekermann, K., & Wegener, M. (2004). *Evaluating urban sustainability using land-use transport interaction models*. European Journal of Transport and Infrastructure Research, 4(3), 251–272.
3. Silva, E. A., & Acheampong, R. A. (2015). *Developing an integrated land use and transport model: A review of progress and prospects*. Transport Reviews, 35(6), 718–742. https://doi.org/10.1080/01441647.2015.1077282
4. Hagenauer, J., & Helbich, M. (2017). *A comparative study of machine learning classifiers for modeling travel mode choice*. Transportation Research Part C: Emerging Technologies, 78, 73–89. https://doi.org/10.1016/j.trc.2017.02.010
5. Ding, C. (2019). *The complex network theory-based urban land-use and transport interaction studies: A review*. Complexity, 2019, 1–12. https://doi.org/10.1155/2019/4180890
6. Pozoukidou, G., & Nijkamp, P. (2020). *Land use models and sustainable urban mobility plans: An integrative approach*. International Journal of Sustainable Development and Planning, 15(6), 867–882. https://doi.org/10.2495/SDP-V15-N6-867-882
7. Zhou, Y., Wang, J., & Li, Y. (2022). *Integrating machine learning and LUTI models to support smart urban transport planning*. Journal of Urban Management, 11(1), 34–45.
8. Ford, A., Dawson, R., & Barr, S. (2018). *Land use transport models for climate change mitigation and adaptation planning*. Journal of Transport and Land Use, 11(1), 77–95.
9. Urrutia-Mosquera, A., Díaz, R., & Ortiz, M. (2025). *Sustainable urban development and mobility: The role of spatial heterogeneity in sustainable trip patterns*. Sustainable Development, 33(2), 2171–2188. https://doi.org/10.1002/sd.2512
10. Yang, X. (2023). *Sustainable city planning and development: Transport and land use*. Sustainability, 15(22), 15804. https://doi.org/10.3390/su152215804