Morphpublishing
J. Comput. Intell. Mach. Reason. Decis.-Mak.
open access academic journal

Peer reviewed

Big Data Analytics for Predictive Modeling of Indian Public Transit Passenger Demand Patterns

Rupam Choudhury[a]

Abstract: Big Data Analytics for Predictive Modeling of Indian Public Transit Passenger Demand Patterns synthesizes a multifaceted framework that amalgamates advanced data processing techniques, statistical learning theory, and linear algebra–based modeling to forecast transit usage dynamics. This study employs an integrative approach whereby heterogeneous datasets, acquired from automated fare collection systems, sensor networks, and mobile applications, are preprocessed, harmonized, and subsequently analyzed via both classical econometric models and modern machine learning algorithms. Emphasis is placed on the development and refinement of complex linear algebra models that leverage matrix decompositions and eigenvalue analyses to capture the intrinsic structure of high-dimensional data. A novel linear lagbera modeling approach is introduced to incorporate temporally lagged variables, thereby encapsulating delayed effects and intertemporal dependencies that are critical for accurately forecasting passenger demand. The methodology is underpinned by rigorous mathematical formulations—including singular value decomposition (SVD), principal component analysis (PCA), and regularized regression techniques—to mitigate noise and enhance model interpretability. Simulation experiments and theoretical analyses substantiate the performance improvements attributable to the integration of advanced linear algebraic constructs. The resultant models demonstrate substantial predictive accuracy, computational efficiency, and resilience to data heterogeneity. By bridging the methodological gap between traditional time-series forecasting and state-of-the-art linear algebraic frameworks, the research offers substantive insights for optimizing transit planning, resource allocation, and strategic decision-making in rapidly evolving urban transit networks.

1. Introduction

Urbanization in India has accelerated at an unprecedented rate, imposing significant challenges on the management and operation of public transit systems. The rapid growth of urban centers, driven by economic opportunities and demographic shifts, has led to an increase in population density, vehicular congestion, and rising demands for efficient and reliable public transportation infrastructure. As India transitions towards a more urbanized society, its cities struggle to maintain sustainable mobility solutions in the face of expanding metropolitan boundaries and diverse commuter needs.

1.1. Demographic and Spatial Growth of Indian Cities

The urban population of India has surged dramatically in recent decades. According to census data, the proportion of urban dwellers increased from 17.3% in 1951 to approximately 34.9% in 2021, with projections suggesting further expansion in the coming decades. This growth has been accompanied by spatial expansion, wherein cities have extended beyond their traditional cores, leading to complex and often unplanned suburbanization. The rise of satellite towns and peri-urban areas has necessitated extensive transit connectivity, yet many regions remain underserved due to inadequate public transit infrastructure.

Furthermore, India's top-tier metropolitan areas, such as Delhi, Mumbai, Bengaluru, and Kolkata, have experienced particularly high urban agglomeration rates, often surpassing infrastructural capacity. Mid-sized cities, including Pune, Lucknow, and Coimbatore, are also undergoing rapid urbanization, necessitating scalable transportation models. These shifting demographics have introduced new challenges in planning, executing, and maintaining transit services that are accessible, efficient, and resilient.

1.2. Increasing Demand and Strain on Public Transit Infrastructure

The intensifying demand for public transit systems in India has placed immense pressure on existing infrastructure. The increase in urban populations has led to overcrowded buses, metro systems, and suburban rail networks, often operating beyond their intended capacities. The Mumbai suburban railway, for example, carries over 7.5 million passengers daily, leading to extreme congestion, delays, and safety concerns. Similarly, metro systems in Delhi, Chennai, and Bengaluru have seen exponential growth in ridership but struggle with peak-hour saturation.

Table 1 provides a comparative overview of public transit ridership statistics in major Indian cities, highlighting the rising demand and stress on current systems.

Table 1. Urban Public Transit Ridership in Major Indian Cities (2023)

City	Daily Bus Ridership (millions)	Daily Metro Ridership (millions)	Suburban Rail Ridership (millions)
Delhi	4.5	6.2	–
Mumbai	3.2	4.5	7.5
Bengaluru	2.8	2.5	–
Chennai	1.9	2.1	1.5
Kolkata	1.5	1.8	2.2

Beyond overutilization, the lack of infrastructure modernization exacerbates these issues. Many bus fleets in Tier-II and Tier-III cities remain outdated, resulting in frequent breakdowns, lower fuel efficiency, and increased environmental pollution. The absence of sufficient multimodal connectivity further hinders seamless commuter experiences, compelling people to rely on inefficient and unregulated transit options.

1.3. Traffic Congestion and Environmental Implications

The exponential rise in private vehicle ownership in Indian cities has further strained public transit systems. With increasing urbanization, the number of registered motor vehicles has surged, leading to severe traffic congestion and delays in public transport operations. In metropolitan areas like Bengaluru, peak-hour traffic speeds have dropped to as low as 10 km/h, significantly hampering the efficiency of public bus services. This congestion not only increases travel times but also contributes to rising fuel consumption and operational costs.

Additionally, air pollution levels in major Indian cities have reached critical thresholds due to vehicular emissions. Public transportation, if adequately managed, serves as a viable alternative to reduce dependency on private vehicles; however, the limitations of existing systems have deterred a modal shift. Table 2 outlines the air quality index (AQI) levels in select urban centers, emphasizing the environmental ramifications of inadequate transit infrastructure.

Table 2. Air Quality Index (AQI) in Major Indian Cities (2023)

City	AQI (Annual Average)	Primary Pollutant	Impact on Health
Delhi	211	PM2.5	Severe
Mumbai	168	NOx	Moderate
Bengaluru	140	PM10	Moderate
Chennai	135	SO2	Moderate
Kolkata	190	PM2.5	Unhealthy

The direct correlation between traffic congestion, emissions, and declining air quality presents a significant public health crisis. Prolonged exposure to high pollution levels has been linked to respiratory ailments, cardiovascular diseases, and reduced life expectancy. Given that public transit remains a crucial mechanism for reducing urban emissions, its inefficiencies contribute to escalating environmental hazards.

1.4. Institutional and Governance Challenges

The governance and management of public transportation in India remain highly fragmented, involving multiple agencies with overlapping responsibilities. In many cases, urban transport falls under the jurisdiction of state transport authorities, municipal corporations, and independent transit operators, leading to inefficiencies in planning and coordination. The lack of a unified transit policy further exacerbates operational inefficiencies, resulting in inconsistent fare structures, non-integrated ticketing systems, and uncoordinated service schedules.

Financial constraints further impede the expansion and modernization of transit networks. Many state-run bus corporations operate at a deficit due to fare subsidies, high maintenance costs, and low revenue generation. Additionally, public-private partnerships (PPPs) in urban transit have faced hurdles due to regulatory uncertainties and investment risks. The absence of strategic long-term policies limits the ability of cities to adopt sustainable and scalable transit solutions [1, 2].

1.5. Equity and Accessibility Concerns

Public transportation in India also faces critical issues related to accessibility and equity. Many marginalized populations, including low-income groups, the elderly, and persons with disabilities, encounter barriers in using public transit due to inadequate infrastructure, poor last-mile connectivity, and safety concerns. Women, in particular, report significant challenges in accessing secure and reliable transit options, with instances of harassment and lack of gender-sensitive transport policies deterring their participation in the workforce.

Additionally, informal settlements and peripheral urban areas often remain underserved, forcing residents to rely on inefficient or unsafe transportation modes. The disproportionate distribution of transit infrastructure exacerbates socio-economic disparities, reinforcing spatial segregation within cities.

The unprecedented pace of urbanization in India has placed extraordinary pressure on public transit systems, revealing critical gaps in infrastructure, governance, and equity. While the demand for efficient urban mobility continues to rise, the existing public transit framework remains insufficient to address the challenges posed by congestion, environmental degradation, and institutional inefficiencies. A comprehensive understanding of these issues is essential for formulating sustainable transit strategies that ensure accessibility, efficiency, and long-term viability. Rapid population growth, spatial expansion of metropolitan areas, and escalating commuter demand necessitate innovative approaches to forecast passenger flow and optimize service delivery [3–5]. Advanced predictive modeling techniques, supported by the voluminous data streams generated by transit networks [6], have emerged as critical tools for addressing these challenges. The integration of big data analytics with rigorous mathematical and econometric models offers a pathway to unravel the complex interplay between temporal, spatial, and socio-economic factors influencing transit demand.

The present study examines the confluence of classical statistical methods and modern machine learning algorithms to construct robust forecasting models[7]. Central to this research is the incorporation of advanced linear algebraic methods that facilitate dimensionality reduction, noise suppression, and the extraction of latent structures from high-dimensional datasets [8]. In addition, the introduction of a linear lagbera modeling approach allows for the systematic integration of lagged variables, capturing delayed effects inherent in transit data. Such methodological innovations not only enhance predictive performance but also provide interpretable insights into the dynamics of passenger demand. The ensuing sections delineate the theoretical underpinnings, advanced mathematical formulations, and empirical strategies that collectively define the proposed analytical framework.

2. Methodology and Theoretical Framework

The methodological framework underlying this research is characterized by the systematic integration of data acquisition, preprocessing, feature engineering, model specification, and rigorous validation protocols. Data sources encompass a diverse array of digital records, including automated fare collection logs, sensor-generated time series, and mobile device trajectories [9, 10]. Initial preprocessing endeavors focus on addressing data irregularities—such as missing values, noise, and temporal misalignments—through normalization, imputation, and robust filtering techniques. Subsequent feature engineering transforms raw inputs into structured predictors, including temporal indicators, spatial coordinates, and exogenous variables (e.g., weather and public events) [11].

Model specification is informed by both classical econometric theory and modern machine learning paradigms. Standard time-series models such as autoregressive integrated moving average (ARIMA) and vector autoregression (VAR) are deployed as benchmarks, while machine learning approaches—including decision trees, random forests, and support vector machines—are evaluated for their capacity to capture nonlinear dependencies. The theoretical framework is rooted in statistical learning theory, which governs the bias–variance tradeoff and underpins regularization techniques that prevent overfitting. Regularized regression methodologies, including ridge regression and LASSO, are implemented to constrain model complexity and enhance generalizability [10, 12].

A distinguishing facet of the framework is the incorporation of temporally lagged variables, operationalized through a novel linear lagbera modeling approach. Lag operators are formally defined, and their inclusion in regression models allows for the quantification of delayed responses within the transit system. Diagnostic tools—such as autocorrelation and partial autocorrelation functions—are systematically employed to determine appropriate lag orders. The resulting framework is further refined by cross-validation procedures, sensitivity analyses, and Monte Carlo simulations that collectively ensure robustness across heterogeneous data segments [13]. Overall, the methodology represents a synthesis of rigorous statistical methods and contemporary computational techniques tailored to the forecasting of complex transit demand dynamics.

3. Advanced Mathematical Modeling and Linear Algebraic Analysis

Central to the analytical framework is an advanced mathematical modeling component that employs linear algebra techniques to elucidate the structural properties of high-dimensional transit data. Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ denote the data matrix, where n is the number of observations and p the number of predictor variables. Singular Value Decomposition (SVD) is applied to $\mathbf{X}$ to decompose it into orthogonal matrices:

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T,$$

where $\mathbf{U} \in \mathbb{R}^{n \times n}$ and $\mathbf{V} \in \mathbb{R}^{p \times p}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{n \times p}$ is a diagonal matrix of singular values. This decomposition facilitates dimensionality reduction by allowing the retention of the top k singular values and corresponding singular vectors, thus projecting the data onto a lower-dimensional subspace that preserves maximal variance.

Eigenvalue decomposition is employed to examine the covariance matrix $\mathbf{C} = \frac{1}{n-1}\mathbf{X}^T\mathbf{X}$, where the eigenvalues λ_i and eigenvectors $\mathbf{v}_i$ satisfy

$$\mathbf{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i.$$

Such analyses provide insight into the intrinsic dimensionality of the dataset and identify directions of greatest variability, which are crucial for the subsequent construction of predictive models. Moreover, the application of Principal Component Analysis (PCA) leverages these eigen-decompositions to extract principal components, effectively reducing multicollinearity among predictors.

The forecasting model is formalized in a matrix regression framework given by

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{Y} \in \mathbb{R}^n$ is the response vector representing passenger demand, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the coefficient vector, and $\boldsymbol{\varepsilon}$ is the error term. The estimation of $\boldsymbol{\beta}$ is obtained by minimizing the least squares loss:

$$\min_{\boldsymbol{\beta}} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2^2.$$

In scenarios of high multicollinearity, regularized regression techniques are invoked. For instance, ridge regression introduces an ℓ_2-penalty:

$$\min_{\boldsymbol{\beta}} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_2^2,$$

while LASSO employs an ℓ_1-penalty to encourage sparsity:

$$\min_{\boldsymbol{\beta}} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1.$$

Further complexity is introduced through the formulation of the data fusion model. Given multiple data sources, we define $\mathbf{X}^{(i)}$ for $i = 1, 2, \ldots, m$ representing distinct feature matrices. These are concatenated to form an augmented matrix:

$$\mathbf{Z} = \left[\mathbf{X}^{(1)} \,|\, \mathbf{X}^{(2)} \,|\, \cdots \,|\, \mathbf{X}^{(m)}\right].$$

The resulting model can then be expressed as

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\theta}$ is the coefficient vector associated with the augmented feature space. The interplay between these matrices is further examined via Kronecker products and tensor decompositions when addressing multi-modal data [14].

A comprehensive analysis also entails the computation of condition numbers for the design matrices to assess numerical stability. The condition number $\kappa(\mathbf{X})$ is defined as

$$\kappa(\mathbf{X}) = \frac{\sigma_{\text{max}}(\mathbf{X})}{\sigma_{\text{min}}(\mathbf{X})},$$

where σ_{max} and σ_{min} are the largest and smallest singular values of $\mathbf{X}$, respectively. High condition numbers necessitate the adoption of regularization to ensure stable coefficient estimates. Collectively, these advanced linear algebra techniques underpin the extraction of latent features and the stabilization of model estimations in the context of high-dimensional transit datasets.

4. Predictive Modeling Techniques and Data Analysis

Robust predictive modeling in the domain of public transit demand necessitates the deployment of both conventional time-series techniques and contemporary machine learning algorithms. Data analysis is conducted in a multistage process, commencing with exploratory data analysis (EDA) to characterize statistical properties such as central tendency, dispersion, and distributional asymmetries across the dataset [15]. Visualization tools—including histograms, scatter plots, and heat maps—are employed to detect anomalies, seasonality, and correlation structures. These initial insights inform the selection and transformation of variables prior to model fitting.

Feature extraction involves both manual engineering and automated selection methods. Temporal features, such as hour-of-day and day-of-week indicators, are complemented by spatial features derived from geospatial clustering algorithms. External exogenous variables, including meteorological conditions and special event markers, are incorporated to capture their influence on passenger behavior. Recursive feature elimination (RFE) and mutual information criteria are applied to distill the most salient predictors, thereby reducing dimensionality and mitigating multicollinearity.

Predictive models are constructed within a dual framework. First, classical econometric models such as ARIMA and VAR are deployed to capture linear dependencies and seasonal patterns inherent in time-series data. These models are refined through parameter tuning using information criteria (e.g., AIC, BIC) and diagnostic tests for residual autocorrelation and heteroskedasticity. Second, machine learning methods—including decision trees, random forests, support vector regression (SVR), and neural networks—are utilized to capture nonlinear interactions among predictors. Ensemble methods, such as boosting and bagging, are also investigated to enhance robustness by aggregating diverse model predictions [16].

The calibration of these models is achieved through rigorous cross-validation procedures, typically via k-fold cross-validation, to assess out-of-sample performance. Error metrics such as mean absolute error (MAE), root mean squared error (RMSE), and the coefficient of determination (R^2) are computed to quantify predictive accuracy. In addition, sensitivity analyses and bootstrapping techniques are employed to ascertain the stability of parameter estimates under varying sampling conditions. The integration of these modeling strategies yields a comprehensive framework that not only forecasts passenger demand with high fidelity but also provides interpretable insights into the causal relationships underpinning transit system dynamics.

5. Linear Lagbera Modeling Approach

In addressing the temporal dependencies inherent in transit demand data, the linear lagbera modeling approach is introduced to systematically integrate lagged variables into the predictive framework. The model posits that current passenger demand y_t is influenced not only by contemporaneous predictors $\mathbf{x}_t$ but also by past values of the dependent variable. Formally, the model is expressed as

$$y_t = \beta_0 + \sum_{i=1}^{p} \beta_i x_{i,t} + \sum_{j=1}^{q} \gamma_j y_{t-j} + \varepsilon_t,$$

where β_0 represents the intercept, β_i are coefficients associated with the contemporaneous predictors $x_{i,t}$, γ_j denote the coefficients of the lagged dependent variable at lag j, and ε_t is the error term assumed to be normally distributed with zero mean and constant variance.

The incorporation of the lag operator L simplifies the notation, where $L^j y_t = y_{t-j}$. Consequently, the model may be succinctly rewritten as

$$y_t = \beta_0 + \sum_{i=1}^{p} \beta_i x_{i,t} + \sum_{j=1}^{q} \gamma_j L^j y_t + \varepsilon_t.$$

Parameter estimation is performed via ordinary least squares (OLS) under the assumption that the predictors, including the lagged terms, are exogenous. However, when multicollinearity is detected among the lagged variables or between contemporaneous and lagged predictors, regularized estimation methods such as ridge regression or LASSO are applied. The optimization problem for ridge regression is formulated as

$$\min_{\beta,\gamma} \sum_{t=1}^{T} \left(y_t - \beta_0 - \sum_{i=1}^{p} \beta_i x_{i,t} - \sum_{j=1}^{q} \gamma_j y_{t-j} \right)^2 + \lambda \left(\sum_{i=1}^{p} \beta_i^2 + \sum_{j=1}^{q} \gamma_j^2 \right),$$

where λ is the regularization parameter that penalizes large coefficients, thereby enhancing model stability.

The selection of the optimal lag order q is informed by statistical diagnostics, including the examination of autocorrelation functions (ACF) and partial autocorrelation functions (PACF), which reveal significant lag structures in the residuals of preliminary models. Furthermore, Granger causality tests are employed to statistically assess whether lagged variables contribute significantly to the predictive power of the model.

Advanced formulations extend the linear lagbera model to incorporate interaction terms and polynomial functions of lagged variables, thereby capturing potential nonlinearities in the delayed effects. In matrix notation, consider a vector $\mathbf{y}$ containing T observations and construct the lagged matrix $\mathbf{Y}_L$ as

$$\mathbf{Y}_L = \begin{bmatrix} y_0 & 0 & \cdots & 0 \\ y_1 & y_0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ y_{T-1} & y_{T-2} & \cdots & y_{T-q} \end{bmatrix}.$$

The extended model then takes the form

$$\mathbf{y} = \beta_0 \mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \mathbf{Y}_L \boldsymbol{\gamma} + \boldsymbol{\varepsilon},$$

where $\mathbf{1}$ is an T-dimensional vector of ones, $\mathbf{X}$ is the matrix of contemporaneous predictors, and $\boldsymbol{\gamma}$ is the vector of lag coefficients. The estimation of this model leverages linear algebra techniques and may employ generalized inverse methods if the design matrix exhibits singularity issues.

The linear lagbera modeling approach thus offers a coherent framework for embedding temporal dynamics within predictive models. Its explicit integration of lagged variables not only improves forecast accuracy by accounting for historical dependencies but also facilitates a nuanced understanding of how past system states influence current transit demand. This model, augmented by regularization and interaction extensions, represents a significant advancement in the analytical toolkit available for forecasting in complex, data-rich urban environments [13].

6. Conclusion

The research presented herein delineates an integrative framework that combines big data analytics with advanced predictive modeling techniques to forecast passenger demand in Indian public transit systems. Through the deployment of sophisticated linear algebraic methods—encompassing singular value decomposition, eigenvalue analysis, and matrix regression—the study has established a robust methodology for processing high-dimensional data and extracting latent structures that underpin transit dynamics. The incorporation of a novel linear lagbera modeling approach has proven effective in capturing the temporal dependencies inherent in transit data, thereby enhancing both forecast precision and interpretability.

The multi-stage analytical framework, which encompasses rigorous data preprocessing, feature engineering, and model validation, demonstrates that the fusion of traditional econometric models with state-of-the-art machine learning techniques can yield significant improvements in predictive performance. The advanced mathematical formulations presented in this study not only address issues of multicollinearity and dimensionality but also provide a foundation for the systematic integration of lagged variables into the forecasting model.

Implications for urban transit planning are profound, as the enhanced predictive capabilities facilitate optimized scheduling, resource allocation, and strategic decision-making in rapidly evolving metropolitan environments. Future research may extend these methodologies by incorporating additional nonlinear dynamics and leveraging emerging data sources. In conclusion, the integration of big data analytics with advanced linear algebraic and lag-based modeling techniques offers a promising avenue for advancing the precision and applicability of predictive models in the domain of public transit management, thereby contributing substantially to the literature on transportation analytics and data-driven urban mobility solutions.

References

[1] P. P. F. Balbin, J. C. Barker, C. K. Leung, M. Tran, R. P. Wall, and A. Cuzzocrea, "Predictive analytics on open big data for supporting smart transportation services," *Procedia computer science*, vol. 176, pp. 3009–3018, 2020.

[2] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big data analytics in intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 383–398, 2018.

[3] S. Bhat, "Optimizing network costs for nfv solutions in urban and rural indian cellular networks," *European Journal of Electrical Engineering and Computer Science*, vol. 8, no. 4, pp. 32–37, 2024.

[4] A. I. Torre-Bastida, J. Del Ser, I. Laña, M. Ilardia, M. N. Bilbao, and S. Campos-Cordobés, "Big data for transportation and mobility: recent advances, trends and challenges," *IET Intelligent Transport Systems*, vol. 12, no. 8, pp. 742–755, 2018.

[5] A. S. Alic, J. Almeida, G. Aloisio, N. Andrade, N. Antunes, D. Ardagna, R. M. Badia, T. Basso, I. Blanquer, T. Braz *et al.*, "Bigsea: A big data analytics platform for public transportation information," *Future generation computer systems*, vol. 96, pp. 243–269, 2019.

[6] S. Bhat, “Leveraging 5g network capabilities for smart grid communication,” *Journal of Electrical Systems*, vol. 20, no. 2, pp. 2272–2283, 2024.

[7] K. Lu, J. Liu, X. Zhou, and B. Han, “A review of big data applications in urban transit systems,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 5, pp. 2535–2552, 2020.

[8] S. V. Bhaskaran, “Tracing coarse-grained and fine-grained data lineage in data lakes: Automated capture, modeling, storage, and visualization,” *International Journal of Applied Machine Learning and Computational Intelligence*, vol. 11, no. 12, pp. 56–77, 2021.

[9] C. K. Leung, J. D. Elias, S. M. Minuk, A. R. R. de Jesus, and A. Cuzzocrea, “An innovative fuzzy logic-based machine learning algorithm for supporting predictive analytics on big transportation data,” in *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2020, pp. 1–8.

[10] M. D. Jackson, C. K. Leung, M. D. B. Mbacke, and A. Cuzzocrea, “A bayesian framework for supporting predictive analytics over big transportation data,” in *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 2021, pp. 332–337.

[11] S. Bhat and A. Kavasseri, “Multi-source data integration for navigation in gps-denied autonomous driving environments,” *International Journal of Electrical and Electronics Research*, vol. 12, no. 3, pp. 863–869, 2024.

[12] A. Fonzone, J.-D. Schmöcker, and F. Viti, “New services, new travelers, old models? directions to pioneer public transport models in the era of big data,” pp. 311–315, 2016.

[13] S. M. Bhat and A. Venkitaraman, “Hybrid v2x and drone-based system for road condition monitoring,” in *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*. IEEE, 2024, pp. 1047–1052.

[14] T. F. Welch and A. Widita, “Big data in public transportation: a review of sources and methods,” *Transport reviews*, vol. 39, no. 6, pp. 795–818, 2019.

[15] S. V. Bhaskaran, “Automating and optimizing sarbanes-oxley (sox) compliance in modern financial systems for efficiency, security, and regulatory adherence,” *International Journal of Social Analytics*, vol. 7, no. 12, pp. 78–91, 2022.

[16] ——, “Enterprise data ecosystem modernization and governance for strategic decision-making and operational efficiency,” *Quarterly Journal of Emerging Technologies and Innovations*, vol. 8, no. 2, pp. 158–172, 2023.