A GIS and Machine Learning based Predictive Framework for Optimizing Highway Bus Scheduling in Sri Lanka: A Systematic Literature Review

G. A. Pramodaya, and R.M.K.T. Rathnayaka

Abstract Highway bus services are a central part of Sri Lanka's public transport network, yet face operational challenges such as fixed route planning, peak-hour service limitations, and the lack of passenger demand forecasting. This systematic literature review (SLR) examines how Geographic Information Systems (GIS) and Machine Learning (ML) have been applied for transport planning and scheduling, with the aim of identifying their suitability for Sri Lankan highway bus operations. Following PRISMA 2020 guidelines, studies published between 2015 and 2025 were reviewed through an analysis of 40 papers. The literature was categorized into six themes, including GIS applications, ML-based demand forecasting, integrated GIS-ML systems, transport modelling in developing countries, highway bus management, and data-related challenges. The analysis identified three research gaps: (1) limited attention to highway-specific integrated predictive frameworks, (2) dependence on data-rich environments with high-resolution datasets, and (3) dominance of reactive rather than proactive scheduling. To address limitations, this study proposes a hybrid GIS-ML predictive framework that combines spatial analysis with demand forecasting to support proactive bus scheduling under sparse data conditions. The findings highlight the potential of GIS and ML integration for capturing travel patterns while emphasizing the need for low-cost solutions suitable for resource-constrained environments such as Sri Lanka.

***Index Terms* - Geographic Information Systems, Highway Buses, Machine Learning, Predictive Scheduling, Public Transport**

I. Introduction

PUBLIC transport supports economic activity, social inclusion, and sustainable mobility, and investment in it generates wider economic and social benefits [1], [2]. In Sri Lanka, road congestion already imposes substantial economic losses, particularly around Colombo [3], and buses remain the main mode of motorized travel for a large share of the population [4]. Highway bus services play an important role in connecting major regions and supporting long-distance passenger mobility. However, these services still face operational challenges related to fixed scheduling, changing passenger demand, and limited forecasting capabilities.

Unlike urban bus services, highway operations work under different conditions, including longer travel distances, stronger day-of-week and holiday demand variations, and greater sensitivity to weather disruptions [5], [6], [7]. Current scheduling approaches mainly depend on fixed timetables, which may not effectively respond to dynamic changes in passenger demand and traffic conditions. Emerging technologies, specifically Geographic Information Systems (GIS) and Machine Learning (ML), offer transformative potential for transport planning [8], [9]. GIS provides spatial understanding of population distribution, service coverage, route accessibility, and demand patterns, while ML models such as Random Forest (RF), XGBoost, and Long Short-Term Memory (LSTM) networks have demonstrated effectiveness in forecasting passenger demand and congestion patterns [10], [11], [7].

However, most reported studies have been developed in data-rich environments using smart-card systems, automatic vehicle location systems, and other high-resolution datasets. Their cost and data requirements create challenges for direct application in developing countries such as Sri Lanka [12], [13]. Therefore, there is a need to investigate approaches that can integrate GIS and ML techniques while operating effectively under limited data availability.

G. A. Pramodaya is an undergraduate from the Department of Computing and Information Systems, Sabaragamuwa University of Sri Lanka. (Email: gapramodaya@std.appsc.sab.ac.lk)

R.M.K.T. Rathnayaka is a Professor attached to the Department of Physical Sciences and Technology, Sabaragamuwa University of Sri Lanka. (Email: kapilar@appsc.sab.ac.lk)

ISSN: 2961-5992

This paper presents a systematic literature review (SLR) of studies published between 2015 and 2025 on GIS and ML approaches for bus planning and scheduling. The review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines and the SLR procedures described by Kitchenham and Charters. The purpose of this review is to identify methodological trends, research gaps, and possible approaches for developing a predictive scheduling framework suitable for Sri Lankan highway bus operations.

This study has the following research objectives:

- RO1: To examine the spatial distribution of bus routes, stops, and passenger demand using GIS in order to identify underserved and high-demand areas.
- RO2: To investigate machine learning techniques for predicting short-term passenger demand and congestion trends on major highway routes.
- RO3: To explore approaches for integrating GIS spatial analysis with ML predictions to develop a hybrid framework for dynamic and proactive bus scheduling in resource-constrained settings such as Sri Lanka.

The rest of the paper is organized as follows. Section II provides the background, describes the Sri Lankan highway bus context, and reviews related work. Section III explains the review methodology. Section IV presents the thematic analysis results and discussion. Section V discusses the identified research gaps, Section VI presents the proposed framework and future implementation, and Section VII concludes the paper.

II. Background and Related Works

A. Background

In early days bus schedules were set in a static manner. These approaches often led to inefficiencies such as bus bunching and peak-hour delays. In the late 1990s, threshold-based control strategies emerged as an initial step toward dynamism [14]. Intelligent transport systems (ITS) later added automatic data collection, real-time passenger information, and computer-aided dispatching [15], [16]. In Sri Lanka, traditional road networks often led to significant inefficiencies, including time wastage due to congestion and unreliable service delivery. The integration of GIS and ML represents the current boundary in this evolution. While GIS provides the necessary spatial grounding for route mapping and accessibility analysis, ML offers the predictive intelligence required to forecast complex passenger demand patterns.

B. The Sri Lankan Highway Bus Context

The Southern Expressway (E01) is the first expressway in Sri Lanka which opened in 2011 between Kottawa and Galle and extended to Matara in 2014. Dedicated highway bus services now run along this corridor, mainly from the Makumbura Multimodal Centre in Kottawa. Services are operated by the Sri Lanka Transport Board (SLTB) and private companies under route permits issued by the National Transport Commission (NTC).

Operating conditions on this corridor differ sharply from the data-rich cities studied in most of the literature. Ticketing is still largely manual, so demand records exist as paper counts and simple point-of-sale logs rather than smart-card transactions. Automatic vehicle location coverage is partial, and automatic passenger counting is absent.

Useful data nevertheless exist. Manual ticket summaries and operator logbooks record boarding volumes, part of the fleet carries GPS devices, and open sources supply census population layers, OpenStreetMap (OSM) road data, weather records, and the national holiday calendar. Local studies already show what such sources can support: night-time light (NTL) imagery combined with open datasets tracks urbanization in Sri Lanka [17], GIS network analysis has organized school bus routing in Kandy [18], and collaborative, low-cost information infrastructure has improved traffic awareness in Metro Manila under similar constraints [19]. At the same time, limited budgets, procurement rules, institutional data sharing, and connectivity gaps restrict what can realistically be deployed [2]. Any framework proposed for this corridor has to respect these conditions.

C. Related Works

The integration of GIS and ML has emerged as a transformative approach for modern transport planning. GIS methodologies provide essential spatial grounding for network analysis and accessibility mapping. Network analysis has been used to compute travel times and service areas for school bus routing [18], spatial clustering has identified travel demand corridors [20], and buffer and catchment analysis has supported station configuration [21]. Accessibility modeling has measured the equity impacts of bus rapid transit (BRT) expansion [22], GIS-based surveys have informed urban transport strategy [9], and NTL imagery has tracked urban growth in Sri Lanka [17]. Recent work also shows that open-source GIS combined with aggregated cell-phone data can design feeder routes without costly sensing infrastructure [23], and that GIS-based analytics can strengthen public-sector decision-making more broadly [9].

ML research has moved from historical averages to models that learn demand patterns directly. RF and LSTM models forecast multimodal passenger flows from smart-

ISSN: 2961-5992

card records [10], multilayer perceptron (MLP) networks predict weather-dependent ridership [7], and direct demand models capture non-linear travel behavior that the traditional four-step process misses [11]. Spatiotemporal forecasting has been combined with vehicle dispatching for on-demand mobility platforms [24], blended statistical and ML approaches have improved ridership forecasts [25], and ML regression has predicted the operating costs of local bus services [26].

Current state-of-the-art "location-aware" systems integrate GIS data fusion with ML optimization. Notable innovations include "floating bus" dynamic routing [27], AFC-based flow estimation, and adaptive clustering for disruptions [28]. However, as many frameworks remain reactive, there is a critical need for low-cost, proactive predictive models tailored to the unique operational constraints of Sri Lanka's highway corridors.

III. METHODOLOGY

This research adopts a quantitative, analytical, and data-driven research design to build a hybrid framework for optimizing Sri Lanka's highway bus transport sector by combining spatial analysis (GIS) and predictive modeling (ML). This approach aims to improve departure schedules, commuter satisfaction, and overall efficiency. The GIS component analyzes geographical accessibility, route coverage, and demand patterns, while the ML component forecasts passenger flows, travel times, and peak congestion trends. The combination of these two components allows for the development of a predictive, adaptive scheduling model that meets the needs of passengers in real time.

A. Research Questions Specification

- RQ1: Which spatial factors (e.g. land use, population distribution, or employment centers) have the strongest impact on the demand of the public bus services in Sri Lanka?
- RQ2: How can open-source GIS software be used to overlay bus route and bus stop data with the population density layers to visually identify and quantify underserved suburban regions towards better journey planning?
- RQ3: What can transfer learning methods do to reconsider the pre-trained ML models of data abundant global datasets using sparse, local data (e.g., manual ticketing and telecommunication mobility traces) to forecast passenger demand on the highway buses in short-term scenarios of the Colombo-Galle-Matara southern expressway?
- RQ4: How effectively can ensemble ML techniques including monsoon-induced flooding and festival crowdsourcing data increase predictive stability in bus congestion patterns in the heterogeneous corridors of Colombo as quantified by measures of uncertainty?
- RQ5: How can the GIS spatial layers and ML demand forecast be combined into one structure to be used in dynamic bus allocation?
- RQ6: To what extent can the efficiency of the service delivery and mitigation of crowds be improved with the proposed GIS-ML integrated system in comparison with the current, fixed, method?

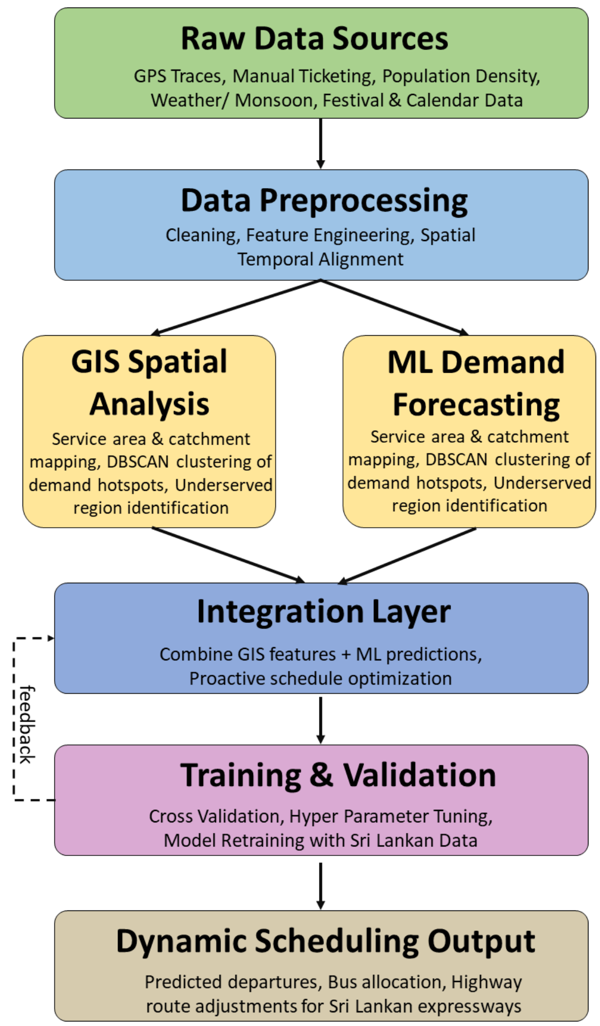

Fig. 1: Proposed Methodology Flowchart for the GIS and Machine Learning Based Predictive

The next figure shows the general approach to the proposed hybrid framework. There is pre-processing of raw multi-source data (GPS traces, manual ticketing records, population density, weather/monsoon data, and festival information). The architecture then splits into the parallel GIS spatial and ML demand forecasting modules. These are combined in an integration layer that is fed on a training and validation loop. The end product produces proactive

ISSN: 2961-5992

optimized dynamic bus schedules specific to the Sri Lankan highway expressways like Colombo-Galle-Matara Road.

B. Search Strategy

There are a number of research repositories and digital libraries that have been used to ensure the comprehensive coverage of multidisciplinary research. Some of them are, Elsevier / ScienceDirect, IEEE Xplore, SpringerLink, MDPI Journals, Taylor & Francis Online, SAGE Journals, ACM Digital Library, and etc. In the below table it has been included the search area and search terms.

TABLE I
SEARCH AREAS AND SEARCH TERMS

Search area	Search terms
Public transport and bus operations	"public transport", "highway bus", "long-distance bus", "bus demand", "bus route planning", "passenger forecasting", "transport optimization", "dynamic bus allocation"
GIS and spatial analysis	"GIS-based analysis", "geospatial transport planning", "route accessibility", "spatial modeling", "catchment area analysis", "transport coverage mapping", "population density GIS", "urban mobility spatial analysis"
Machine learning and prediction	"machine learning transport", "bus demand forecasting", "time-series prediction public transport", "Random Forest demand prediction", "XGBoost transport forecasting", "LSTM traffic prediction", "predictive analytics public transport"
General	"developing countries transport", "Sri Lanka transport", "data-driven mobility planning", "integrated GIS and ML framework"

It has used a series of Boolean combinations to narrow the search results depending on relevance. They have gone through the below sectors. "GIS with ML with Bus Transport", "Highway Bus with Prediction", "Transport Planning with ML", "GIS with Transport Optimization", "Developing Countries Focus"

C. Select and Evaluate

To select the articles which only related to the particular topic it has been applied set of inclusion and exclusion criteria for filter only relevant and quality studies. The articles were selected by following 5 inclusion and 5 exclusion criteria.

Exclusion Criteria

- EC1: Publications which not related to public bus transportation sector.
- EC2: Not related to GIS or ML
- EC3: Simulated only studies (Studies not used real or semi-real transport data)
- EC4: Low research transparency.
- EC5: Poor quality or non-academic sources

Inclusion Criteria

- IC1: Relevance to Highway/Bus Transport
- IC2: Studies focus on GIS techniques or ML forecasting, as well as integrated GIS and ML frameworks.
- IC3: According to publication type which in Peer-reviewed journal articles, conference papers, published theses, and relevant government reports.
- IC4: Publications between 2015 and 2025.
- IC5: Studies which used real-world or open datasets, validated case studies, and provide clear method descriptions with measurable outcomes.

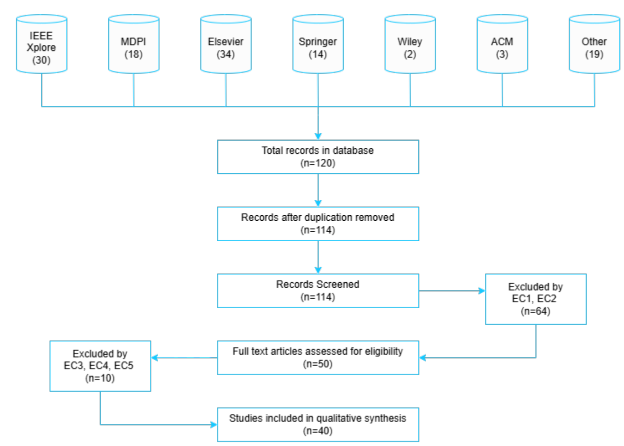

Fig. 2: PRISMA flow of study identification, screening, eligibility assessment, and inclusion.

The Selection procedure is conducted in a sequential step, followed by inclusion and exclusion criteria. In the first step, duplicate publications were removed. The procedure followed PRISMA guidelines. From 120 retrieved articles, 6 duplicates were removed. After applying exclusion criteria 74 papers were excluded. Figure 2 shows the PRISMA flow of section.

D. Eligibility Assessment

50 studies were fully reviewed to verify methodological relevance and alignment with the scope of this SLR. When reviewing the papers it has been considered about following areas. During this phase, 10 studies were excluded. Reasons

ISSN: 2961-5992

included: No usable data, Focus only on theory without analytical contribution, Method not applicable to highway bus systems, Missing details in methodology or results.

IV. Results and Discussion

The 40 included studies correspond to references [1]-[40]. Each study was coded with one primary theme and, where relevant, one or more secondary themes. Six themes emerged from this coding: GIS applications in transport planning, ML for passenger demand forecasting, integrated GIS–ML systems, transport modeling in developing countries, highway and intercity bus management, and data limitations and practical challenges. Fig. 3 shows how often each theme appears; because a study can carry several themes, the counts sum to more than 40.

The below table shows Comparison of Candidate ML Models for Highway Bus Demand Forecasting.

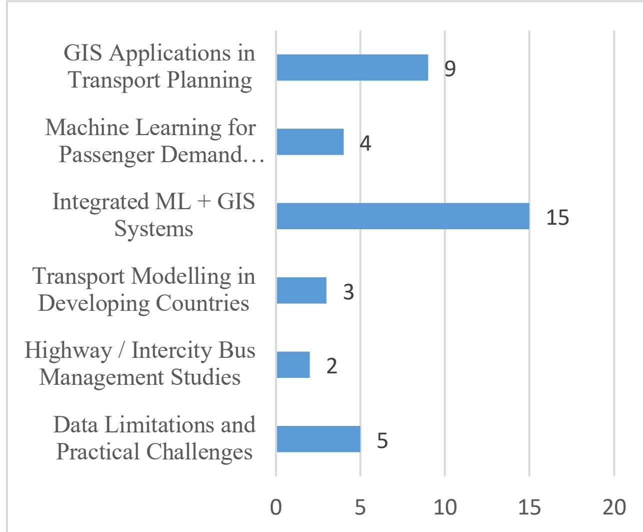

Fig. 3: Distribution of the six research themes across the 40 included studies (primary and secondary themes combined).

TABLE II

COMPARISON OF CANDIDATE ML MODELS FOR HIGHWAY BUS DEMAND FORECASTING

Model	**Strengths**	**Limitations**	**Fit for sparse highway-bus data**
RF [10]	Robust to noise and outliers; needs little tuning; handles mixed feature types; provides feature importance	Cannot extrapolate beyond the training range; large forests are slow to score	Strong baseline; works with the modest hourly datasets that manual ticketing can supply
XGBoost [15]	High accuracy on tabular data; built-in regularization; handles missing values; fast training	Sensitive to hyperparameter choices and to data leakage; less interpretable without added tools	Primary candidate for hourly stop-level demand forecasting on the corridor
LSTM [10]	Learns long temporal dependencies directly from sequences	Needs large training sets and more computation; harder to interpret; overfits short, sparse histories	Secondary option; becomes attractive once longer digital records accumulate
MLP [7]	Simple non-linear mapper; effective with well-engineered inputs such as weather variables	No built-in temporal memory; performance depends heavily on feature quality	Useful baseline and component in hybrid or ensemble designs

A. *Research Interest*

The core interest of this research is to address the systemic limit in Sri Lanka's public transport, specifically the critical issue of overcrowding and supply-demand imbalance. Current transport management system in Sri Lanka lacks predictive capabilities and based on that it makes inadequate service during peak hours of the day. Throughout this SLR it will highlight the requirement of a hybrid framework that integrates ML with GIS to create a proactive, data-driven scheduling model.

The primary goal is to develop a decision-support system capable of recommending dynamic bus counts and optimized schedules based on predicted surges. This research aims to bridge the significant gap between traditional spatial planning and modern Intelligent Transport Systems (ITS) within the Sri Lankan context.

B. *GIS Applications in Transport Planning*

When referring to all 40 papers, this theme was the core theme of most of them. In [8], [17], [18], [20], [22], and [29] papers, this theme was the primary theme. The rest of the papers mostly treated this as a secondary theme. Under this theme, location data can be turned into practical planning tools, usually by using open data sources.

It can be seen that the most important GIS techniques are discussed in several papers, and researchers have been using open data sources there. Paper [18] has discussed network analysis to calculate travel times and service areas, and paper [20] discussed spatial clustering methods like

ISSN: 2961-5992

DBSCAN to group demand points into demand corridors. In paper [13], buffer density analysis was used to define catchment zones. In paper [17], night time light (NTL) satellite imagery is combined with other open datasets such as Road Network Coverage (RNC), Mobile Network Coverage (MNC), and vegetation index layers to track urbanization in Sri Lanka. These studies show how combining different spatial layers can provide a better understanding of urban growth patterns and transport accessibility.

Paper [18] discusses how to apply GIS techniques focusing on narrow roads and morning congestion at school buses in Kandy, Sri Lanka. Paper [22] has applied GIS in transport planning by modeling accessibility impacts of Bus Rapid Transit expansion in Rio de Janeiro. Paper [18], which is about School Bus Routing in Kandy, and Paper [22], which is about Future Accessibility Impacts of BRT in Rio, focus on optimizing transport using fixed-location GIS and policy scenarios, while Paper [21], about Hong Kong station configuration, applies predictive machine learning with Generative Adversarial Networks (GANs) to forecast how new stations could reshape population density.

Paper [21] shows some detailed maps of Road Network Coverage (RNC), Mobile Network Coverage (MNC), Night-Time Light (NTL) imagery, and Normalized Difference Vegetation Index (NDVI). It proves that putting all these map layers together gives us a much better idea of how towns are actually growing and where transport demand is increasing. Other studies also extend GIS applications beyond traditional planning. Paper [23] shows that open-source GIS combined with aggregated cell-phone data can be used to design feeder routes and select stops without costly sensing infrastructure, while paper [9] highlights how GIS-based analytics can support broader public-sector decision-making.

Here are some common limitations that were apparent after reading the papers under this theme. Many models are static and do not handle dynamic (real-time) changes in traffic or demand (Paper [18]). Manual digitizing and coarse data reduce accuracy when scaling to bigger networks. In developing countries like Sri Lanka, reliance on estimated or secondary sources is common, and some advanced techniques from data-rich cities need careful adaptation before applying them to local transport corridors.

C. Machine Learning for Passenger Demand Forecasting

Theme 'Machine Learning for Passenger Demand Forecasting' is a common theme in most of the 40 papers. This theme was the primary theme in papers [10], [11], [24], and [7]. It can be found as a secondary theme in several papers. These studies move forward from simple historical averages to build models that can learn demand patterns and predict future passenger demand. Demand can be changed daily or weekly based on several factors such as time, weather, and other external conditions.

When building the methods, researchers have applied ensemble techniques such as Random Forests (RF) and deep learning models such as Long Short-Term Memory (LSTM) for forecasting passenger flows using smart-card records [10]. These models have been used for both short-term and long-term demand forecasting. In paper [7], Multilayer Perceptron (MLP) networks have been applied for demand prediction considering weather conditions. There is a model called the Traditional Four-Step Model (Steps: Trip generation → Trip distribution → Mode choice → Traffic assignment). To overcome limitations such as difficulties in handling nonlinear relationships and modern mobility data, direct demand models have been applied as an alternative to the traditional four-step travel demand modeling process in paper [11].

To improve accuracy, most studies have included temporal features and ensemble methods. Different machine learning models show different advantages depending on the available data. Tree-based models such as RF and XGBoost are suitable for structured transport datasets because they can handle multiple features and require comparatively less training data. LSTM models are effective for learning long-term temporal patterns but require larger datasets and higher computational resources. MLP models can capture nonlinear relationships but their performance mainly depends on effective feature selection. These differences are important when selecting suitable models for transport systems with limited local data availability. In papers [11] and [7], it can be seen that machine learning models demonstrate practical value by using openly available data and supporting better matching of supply to demand.

Several limitations are found when referring to these papers, and some of them are given here. In paper [11], daily time bins have been considered, and this often misses critical hourly demand peaks, which are essential for operational planning and resource allocation. The forecasting models used in paper [24] heavily depend on predefined variables such as Points of Interest (POI) distributions, weather, and time-of-day. This can limit the ability to automatically uncover hidden correlations in the data. Some studies have also extended machine learning applications beyond passenger demand forecasting, such as predicting bus operating costs [26], combining statistical and machine learning approaches for ridership forecasting [25], and developing passenger flow models [30]. Weather-related prediction has also been considered, where anomaly-based methods have been used to identify low-visibility conditions [5], showing the importance of treating weather as an important factor in transport forecasting.

In summary, this theme represents a move from traditional statistical methods to advanced data-driven models that can effectively capture complex travel patterns. The use of ensemble methods and deep learning techniques, alongside enhanced temporal and external features, has led to improved predictive accuracy for transportation planning. However, challenges such as coarse temporal granularity, dependence on predefined variables, and limited availability of local datasets highlight the need for adaptive models that can improve the robustness and real-time effectiveness of highway bus demand forecasting.

ISSN: 2961-5992

D. Integrated ML + GIS Systems

It can be seen that the theme, Integrated ML + GIS Systems, is the primary theme in most of the research papers out of the considered 40 papers, including [11], [12], [21], [24], [27], [28], [31], [32], [33], [34], [35], and many others. They propose location-aware solutions for problems in the transport sector by combining spatial analysis with machine learning.

In methodologies, ML has been applied for prediction and optimization, while GIS has been used for data fusion, network analysis, and mapping. GIS provides geographical grounding through networks, stations, routes, and catchment areas, while machine learning provides predictive intelligence through demand prediction, density estimation, and clustering. It can be seen Passenger Flow Clustering with GIS Mapping in [33]. ML and GIS dynamically adjust bus line lengths and routes in real time, ensuring demand-responsive operations and spatially efficient service coverage [31].

Studies share methods for fusing spatial and demand data but differ in their goals. Some studies focus on immediate dispatching decisions [27], while others focus on proactive forecasting and repositioning based on future demand [24]. AFC (Automatic Fare Collection) and GPS provide a solution for real-time passenger flow estimation by combining boarding information with vehicle location data [28]. However, practical issues such as hardware synchronization mismatches between boarding events and bus location data can reduce accuracy [28]. Smartphone sensing combined with transport data provides commuter-specific insights [32], and adaptive clustering with Modified Adaptive Large Neighborhood Search with Node-based clustering (MALNSN) ensures efficient evacuation routing under disruptions [33]. Other studies extend this combined approach through prediction-based dispatching [35], ant colony optimization for identifying weak-demand areas [34], and agent-based methods for matching transport supply and demand [36].

In paper [32], we can see a limitation as its performance heavily depends on the accuracy of demand clustering and predefined parameters, which may reduce robustness when applied to highly dynamic or uncertain real-world transport scenarios. Similarly, some advanced smart-city applications require large amounts of sensing data and supporting infrastructure, which may not be directly available in developing countries. This theme acts as a bridge between GIS and ML, and it shows that combined spatio-temporal approaches can be applied for Sri Lankan expressway buses to create dynamic schedules that adjust to real-time conditions. However, the main challenge is developing such systems with limited local data availability and operational constraints.

E. Transport Modeling in Developing Countries

The theme Transport Modelling in Developing Countries is the primary and secondary theme in [18], [30], [8], [22], and [6]. These studies were conducted in Sri Lanka, the Netherlands, Romania, Brazil, and India respectively. These studies demonstrate how standard transport modelling techniques perform in resource-limited environments and how they can be adapted to different operational conditions.

It has been applied GIS network analysis to local topography [18]. Paper [22] shows that accessibility measurement combined with spatial regression can effectively evaluate the equity impacts of Bus Rapid Transit expansion in Rio de Janeiro. Paper [6] focuses on estimating commuter travel costs under varying levels of crowding in the Mumbai suburban rail system, highlighting the economic burden of overcrowding and its implications for transport policy. Similarly, congestion cost analysis has demonstrated the significant economic impact of traffic congestion in Colombo [3]. Paper [19] also shows that collaborative information infrastructure can improve transport management in developing cities without requiring heavy investment, while paper [1] highlights the wider economic and social benefits of investing in public transport systems.

In paper [18], the model does not fully account for dynamic traffic conditions and real-time variability in road networks, which can affect the efficiency of bus scheduling. GIS-based surveys for public transport strategy rely heavily on static spatial data and predefined planning assumptions. These may not adequately reflect dynamic travel behavior or rapidly changing urban traffic conditions [8].

In summary, these studies prove that standard transport models are very useful for addressing major transport problems in developing countries. However, most of these approaches are based on static data and limited time frames. Since traffic conditions and travel behavior continuously change, future transport modelling approaches should incorporate dynamic and real-time data to improve planning accuracy and operational effectiveness.

F. Highway / Intercity Bus Management Studies

The theme of highway and intercity bus management can be seen in both [27] and [4] as the primary theme, while a few other papers treat this theme as a secondary topic. These two papers are the only studies that directly focus on highway or intercity bus management among the reviewed literature. The particular studies have addressed challenges in changing passenger demand, long travel times, and the need for reliable scheduling on corridors such as Sri Lanka's highways.

Paper [27] shows the concept of floating buses through dynamic route planning and passenger allocation based on real-time demand. According to [27], CO_2 emissions can be reduced through dynamic scheduling of floating buses by minimizing unnecessary travel and optimizing passenger allocation. Paper [4] studies the stated and revealed preferences of passengers towards bus transport in Sri Lanka. It highlights punctuality, vehicle condition, and

ISSN: 2961-5992

driver attitude as key service quality attributes influencing user satisfaction.

The key difference is that paper [27] focuses on designing a new, flexible bus system that changes its routes and stops based on real-time passenger demand, while paper [4] focuses on surveying passenger preferences to understand how to improve the quality of existing bus services. Therefore, the two studies complement each other, where one demonstrates how demand-responsive operations can be implemented, while the other identifies what passengers value most in the Sri Lankan context. These findings give clear direction for designing a dynamic scheduling approach for Sri Lankan highway buses.

Limitations can be seen in the current studies under this theme. In [4], the focus on stated and revealed passenger preferences does not fully incorporate dynamic travel behavior or real-time operational factors, which may limit the applicability of the findings for improving bus services. In paper [27], although the floating bus approach relies on dynamic scheduling and real-time demand data, it does not fully address challenges such as unpredictable traffic conditions and infrastructure constraints, which may affect the consistency of service delivery.

This theme fits well with integrated ML and GIS systems and shows their suitability for developing countries. However, the limited number of studies in this area highlights that highway and intercity bus operations remain comparatively underexplored, especially when compared with urban transport systems. The main takeaway is that to improve the bus system, both ideas discussed in these two papers should be combined by building a smart, real-time scheduling system while ensuring that buses remain reliable and comfortable for everyday passengers.

G. Data Limitations & Practical Challenges

The theme Data Limitations and Practical Challenges appears throughout all 40 papers, while this theme is the primary or secondary theme in [2], [5], [12], [13], [16], [25], [26], [28], [29], [32], [37], [38], and [39]. Every study mentioned its limitations within the paper. The key limitations were data quality, availability, data fusion, and real-world implementation.

The study in paper [37] highlights that weather extremes, economic conditions, and real-time information influence passenger demand, while also showing the difficulty of separating the impact of real-time information from external factors. Paper [28] discusses the recurring issue of modelling cyclical passenger flow patterns and emphasizes the difficulty of balancing routine travel behavior with irregular disruptions in real-time forecasting. The study in [26] notes the difficulty of implementing complex ML models as a key limitation in predicting bus transport service costs. A systematic review in [13] also highlights that large smart-card, AFC, and GPS datasets are mainly available in developed countries, making it difficult to directly apply these methods in developing countries.

Paper [28] proposes a real-time framework to address data limitations and practical challenges by effectively managing incomplete passenger flow data in urban bus systems. Other studies suggest practical approaches to overcome these limitations. Paper [38] combines expert knowledge with machine learning to improve anomaly detection using noisy sensor data, while paper [16] identifies the type of real-time information that passengers actually require. Paper [2] further highlights that successful intelligent transport systems depend not only on technology but also on collaboration among different transport agencies.

Current studies suggest the need for creative, low-cost solutions that use open data sources, participatory smartphone sensing, and careful fusion of limited ticketing and GPS data for developing countries like Sri Lanka [32], [29], [19]. Researchers have carried out several important studies under this theme by combining multiple data sources to overcome data limitations. These findings provide a realistic understanding of what data-driven transport systems can achieve in resource-limited environments and highlight the importance of designing practical solutions that can operate with limited and incomplete data.

V. IDENTIFIED RESEARCH GAPS

The reviewed literature provides a strong and comprehensive foundation for GIS and ML based dynamic approaches in the public transport sector. However, several critical research gaps can be identified through the thematic analysis, particularly concerning the unique operational landscape of Sri Lanka's highway bus systems. The identified research gaps are discussed under the following subtopics.

- Absence of Highway-Specific Integrated Predictive Frameworks.
- Overreliance on Data-Rich Environments.
- Reactive Rather Than Proactive Scheduling Approaches.

Absence of Highway-Specific Integrated Predictive Frameworks: Several studies have used a combined approach of GIS and ML [27], [31], [28], [32], [33], [15], [21], [34], [35], [12], [24]. However, these studies were mainly designed for metropolitan bus networks, metro systems, school transport, or on-demand mobility platforms. Only a very limited number of studies focus directly on highway or intercity bus operations. Long-distance highway services have different operational characteristics, such as longer travel distances, stronger day-of-week and holiday demand variations, corridor-based travel patterns, and different infrastructure constraints. Current studies give limited attention to these characteristics. Therefore, there is still a lack of an integrated GIS and ML framework specifically designed for highway bus scheduling in a developing country such as Sri Lanka.

Overreliance on Data-Rich Environments: Most current studies depend on high-resolution datasets such as smart-card or AFC systems, GPS traces, automated

ISSN: 2961-5992

monitoring infrastructure [27], [37], [28] detailed spatial datasets [21], and rich trip or POI data [24], [40]. However, such datasets are not easily available in developing countries like Sri Lanka, where ticketing systems are still largely manual and vehicle tracking is limited. Although a few studies have demonstrated the use of alternative data sources such as smartphone sensing [32] and aggregated cell-phone data with open-source GIS [23], only a limited number of studies have investigated predictive transport systems under genuinely low-data conditions.

Reactive Rather Than Proactive Scheduling Approaches: Most existing studies mainly follow a reactive approach rather than a proactive approach. That means they respond to operational issues after they occur instead of preventing problems before they happen. For example, dynamic routing reacts to real-time passenger requests [27], flexible route adjustments respond to observed passenger demand [31], and disruption management reacts to unexpected operational events [33]. Although these approaches improve operational efficiency, they provide limited support for forecasting future demand and optimizing schedules in advance. Only a few studies consider prediction-based repositioning [24], while proactive scheduling based on demand forecasting remains largely unexplored for highway bus operations.

VI. Proposed Framework and Future Implementation

The figure below presents the proposed conceptual framework for GIS and Machine Learning Based Bus Demand Prediction and Scheduling Optimization. The proposed framework is designed to address the identified research gaps by supporting proactive rather than reactive scheduling, utilizing sparse local data commonly available in Sri Lanka, and providing a framework specifically tailored for long-distance highway bus operations.

TABLE II

COMPARATIVE ANALYSIS OF REPRESENTATIVE APPROACHES AND THEIR TRANSFERABILITY TO SRI LANKAN HIGHWAY BUS OPERATIONS

Ref.	Approach	Data requirements	Key limitation	Transferability to the target corridor
[28]	Real-time flow estimation from AFC and GPS	AFC transactions plus GPS streams	Requires fare-collection hardware; synchronization mismatches between boarding and location data	Low – the required hardware is absent
[27]	“Floating bus” dynamic routing and allocation	Live passenger demand feeds	Reactive; assumes real-time request data; traffic and infrastructure constraints not fully treated	Low to moderate – concept useful, data feeds missing
[31]	Real-time optimization of flexible line lengths	Real-time operations data	Reactive; designed for dense urban geometry	Low to moderate
[10]	RF and LSTM forecasting of multimodal flows	Large smart-card record sets	Data volumes unavailable outside AFC systems	Moderate – methods transfer, data must be replaced
[21]	GAN-based station planning from geodata	NTL, road and mobile coverage, NDVI layers	Heavy multi-layer data fusion; metro-oriented	Moderate – open layers exist for Sri Lanka [17]
[24]	Spatiotemporal forecasting with dispatching	High-frequency trips, POI, weather	Car-hailing context; limited to predefined variables	Moderate – feature design transfers
[7]	MLP prediction of weather-dependent ridership	Ridership plus weather records	City-bus focus; no spatial module	High – weather data are openly available
[23]	Feeder route design from cell-phone data with open-source GIS	Aggregated cell-phone data	Needs operator cooperation for data access	High – closest low-cost analog to the target setting
[18]	GIS network analysis for school bus routing	Road network and facility locations	Static; no demand forecasting	High – GIS component proven locally

ISSN: 2961-5992

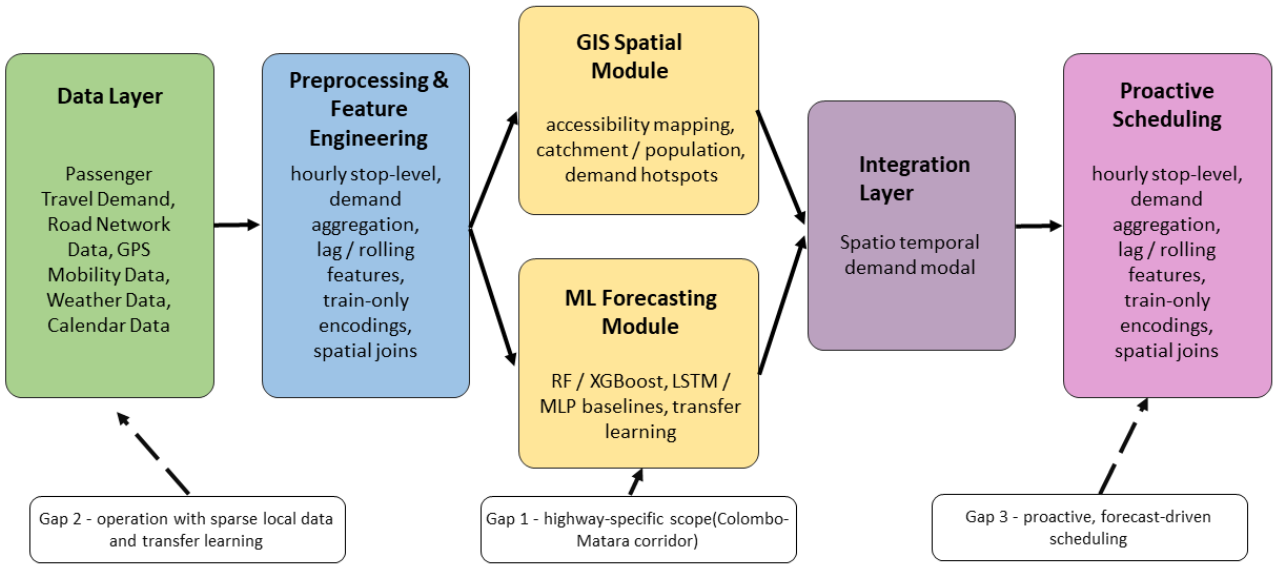

Fig. 4: Conceptual framework for GIS and ML based demand prediction and proactive highway bus scheduling.

VII. Conclusion

This SLR evaluated 40 studies published between 2015 and 2025 to explore how GIS and ML approaches can optimize public transport, with a specific focus on Sri Lankan highway bus operations. Highway bus systems face several challenges, including static route planning, demand uncertainty during peak periods, and the absence of reliable demand forecasting mechanisms.

According to this review, several key limitations can be identified within existing studies. Most current approaches depend on high-resolution and data-rich environments, including smart-card systems, automated monitoring infrastructure, and dense GPS datasets, which are not easily available in developing countries such as Sri Lanka. Furthermore, existing studies have given limited attention to long-distance highway bus operations and mainly focus on urban transport systems. Many current approaches also follow reactive strategies that respond to operational issues after they occur rather than proactively predicting future demand.

GIS provides essential geographical grounding by supporting route analysis, accessibility measurement, spatial mapping, and demand hotspot identification, while ML techniques, including ensemble methods and deep learning models, provide the ability to capture complex passenger demand patterns. The integration of these two approaches can create a powerful location-aware system capable of supporting dynamic and predictive scheduling decisions.

Standard transport models applied in developing regions often depend on static spatial data and fixed planning assumptions. Their limitations highlight the importance of innovative, low-cost data solutions that can combine available ticketing records, open spatial datasets, and alternative data sources instead of depending only on expensive automated systems.

This review identified three major research gaps affecting Sri Lanka's highway bus network. First, existing GIS and ML integrated frameworks have rarely focused on long-distance highway bus operations in developing countries. Second, most current models are designed for data-rich environments and cannot be directly transferred to low-data conditions. Third, existing scheduling approaches are mainly reactive and lack forecasting mechanisms for proactive demand-based planning.

To address these limitations, this study proposes a GIS-ML based hybrid framework specifically designed for Sri Lankan highway bus operations. The proposed framework aims to operate effectively under limited data availability and support proactive scheduling by integrating spatial analysis with predictive demand modelling.

This systematic literature review confirms the need for a proactive and predictive scheduling model tailored to Sri Lankan highway buses. Future research should focus on implementing and validating the proposed framework using real operational data from Sri Lankan highway corridors. A dynamic, predictive, and low-cost scheduling solution can support passengers through reduced waiting times, assist operators with improved resource utilization, and provide policymakers with a practical approach for improving public transport systems in developing countries.

VIII. Declaration of AI Usage

During the preparation of this manuscript, the authors utilized ChatGPT, Claude, Perplexity AI, and Grammarly for various tasks including language editing and summarization. However, the authors independently

ISSN: 2961-5992

conducted all study selection and analyses, taking full responsibility for the manuscript's accuracy and integrity. AI tools are not acknowledged as authors or co-authors.

References

[1] J. A. K. Alhassan and P. Anciaes, "Public transport investments as generators of economic and social activity," J. Transp. Heal., vol. 41, no. January, p. 101989, 2025, doi: 10.1016/j.jth.2025.101989.

[2] M. Visan, S. L. Negrea, and F. Mone, "Towards intelligent public transport systems in Smart Cities; Collaborative decisions to be made," Procedia Comput. Sci., vol. 199, pp. 1221–1228, 2021, doi: 10.1016/j.procs.2022.01.155.

[3] S. A. C. S. Jayasooriya and Y. M. M. S. Bandara, "Measuring the Economic costs of traffic congestion," in 2017 Moratuwa Engineering Research Conference (MERCon), IEEE, May 2017, pp. 141–146. doi: 10.1109/MERCon.2017.7980471.

[4] T. M. Rengarasu and J. A. S. Dulanga, "Stated and Revealed Preferences Towards Bus Transport in Sri Lanka," J. South Asian Logist. Transp., vol. 4, no. 2, pp. 85–107, 2024, doi: 10.4038/jsalt.v4i2.94.

[5] W. Qian, J. C.-H. Leung, Y. Chen, and S. Huang, "Applying Anomaly-based Weather Analysis to the Prediction of Low Visibility Associated with the Coastal Fog at Ningbo-Zhoushan Port in East China," Adv. Atmos. Sci., vol. 36, no. 10, pp. 1060–1077, Oct. 2019, doi: 10.1007/s00376-019-8252-5.

[6] P. K. Sahu, G. Sharma, and A. Guharoy, "Commuter travel cost estimation at different levels of crowding in a suburban rail system: a case study of Mumbai," Public Transp., vol. 10, no. 3, pp. 379–398, Dec. 2018, doi: 10.1007/s12469-018-0190-6.

[7] Z. H. Farahmand, K. Gkiotsalitis, and K. T. Geurs, "Predicting bus ridership based on the weather conditions using deep learning algorithms," Transp. Res. Interdiscip. Perspect., vol. 19, p. 100833, May 2023, doi: 10.1016/j.trip.2023.100833.

[8] G. Droj, L. Droj, and A.-C. Badea, "GIS-Based Survey over the Public Transport Strategy: An Instrument for Economic and Sustainable Urban Traffic Planning," ISPRS Int. J. Geo-Information, vol. 11, no. 1, p. 16, Dec. 2021, doi: 10.3390/ijgi11010016.

[9] Rita Uchenna Attah, Baalah Matthew Patrick Garba, Ifechukwu Gil-Ozoudeh, and Obinna Iwuanyanwu, "Leveraging geographic information systems and data analytics for enhanced public sector decision-making and urban planning," Magna Sci. Adv. Res. Rev., vol. 12, no. 2, pp. 152–163, 2024, doi: 10.30574/msarr.2024.12.2.0191.

[10] F. Toque, M. Khouadjia, E. Come, M. Trepanier, and L. Oukhellou, "Short & long term forecasting of multimodal transport passenger flows with machine learning methods," in 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), IEEE, Oct. 2017, pp. 560–566. doi: 10.1109/ITSC.2017.8317939.

[11] X. Yan, X. Liu, and X. Zhao, "Using machine learning for direct demand modeling of ridesourcing services in Chicago," J. Transp. Geogr., vol. 83, p. 102661, Feb. 2020, doi: 10.1016/j.jtrangeo.2020.102661.

[12] K. L.-M. Ang, J. K. P. Seng, E. Ngharamike, and G. K. Ijemaru, "Emerging Technologies for Smart Cities' Transportation: Geo-Information, Data Analytics and Machine Learning Approaches," ISPRS Int. J. Geo-Information, vol. 11, no. 2, p. 85, Jan. 2022, doi: 10.3390/ijgi11020085.

[13] K. E. Zannat and C. F. Choudhury, "Emerging Big Data Sources for Public Transport Planning: A Systematic Review on Current State of Art and Future Research Directions," J. Indian Inst. Sci., vol. 99, no. 4, pp. 601–619, Dec. 2019, doi: 10.1007/s41745-019-00125-9.

[14] M. D. Rossetti and T. Turitto, "Comparing static and dynamic threshold based control strategies," Transp. Res. Part A Policy Pract., vol. 32A, no. 8, pp. 607–620, 1998, doi: 10.1016/S0965-8564(98)00019-6.

[15] G. Gentile and K. Noekel, Eds., Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems. in Springer Tracts on Transportation and Traffic. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-25082-3.

[16] E. MacEdo et al., "Real-time information systems for public transport: User perspective," Transp. Res. Procedia, vol. 52, no. 2020, pp. 732–739, 2021, doi: 10.1016/j.trpro.2021.01.088.

[17] N. Akalanka, N. Kankanamge, J. Munasinghe, and T. Yigitcanlar, "Urban Big Data Analytics: A Novel Approach for Tracking Urbanization Trends in Sri Lanka," Land, vol. 13, no. 6, p. 888, Jun. 2024, doi: 10.3390/land13060888.

[18] B. G. T. L. Aberathne, H. Nandalal, and I. M. S. Satthyaprasad, "Application of School Bus Routing Problem in GIS to Organize School Buses in Kandy City, Sri Lanka," J. Geospatial Surv., vol. 3, no. 2, pp. 10–42, 2023, doi: 10.4038/jgs.v3i2.44.

[19] J. R. Enslin, "Traffic Woes, Metro Manila, and Collaborative Problem-Solving: A Case Study of Computer-mediated, Collaboratively Built Information Infrastructure in the Field of Transportation in the Philippines," IKAT Indones. J. Southeast Asian Stud., vol. 1, no. 1, p. 145, Jan. 2018, doi: 10.22146/ikat.v1i2.32357.

[20] K. Bahbouh, J. R. Wagner, C. Morency, and C. Berdier, "Travel demand corridors: Modelling approach and relevance in the planning process," J. Transp. Geogr., vol. 58, pp. 196–208, Jan. 2017, doi: 10.1016/j.jtrangeo.2016.12.007.

[21] Y. Ji and H. Zheng, "Configuration of public transportation stations in Hong Kong based on population density prediction by machine learning," Int. J. Appl. Earth Obs. Geoinf., vol. 136, p. 104339, Feb. 2025, doi: 10.1016/j.jag.2024.104339.

[22] R. H. M. Pereira, "Future accessibility impacts of transport policy scenarios: Equity and sensitivity to travel time thresholds for Bus Rapid Transit expansion in Rio de Janeiro," J. Transp. Geogr., vol. 74, pp. 321–332, Jan. 2019, doi: 10.1016/j.jtrangeo.2018.12.005.

[23] M. Wei, T. Liu, and B. Sun, "Optimal Routing Design of Feeder Transit With Stop Selection Using Aggregated Cell Phone Data and Open Source GIS Tool," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 4, pp. 2452–2463, Apr. 2021, doi: 10.1109/TITS.2020.3042014.

[24] Y. Guo, Y. Zhang, Y. Boulaksil, and N. Tian, "Multi-dimensional spatiotemporal demand forecasting and service vehicle dispatching for online car-hailing platforms," Int. J. Prod. Res., vol. 60, no. 6, pp. 1832–1853, Mar. 2022, doi: 10.1080/00207543.2021.1871675.

[25] C. T. Lawson, A. Muro, and E. Krans, "Forecasting bus ridership using a 'Blended Approach,'" Transportation (Amst)., vol. 48, no. 2, pp. 617–641, Apr. 2021, doi: 10.1007/s11116-019-10073-z.

[26] A. Amicosante, A. Avenali, T. D'Alfonso, M. Giagnorio, A. Manno, and G. Matteucci, "Predicting costs of local public bus transport services through machine learning methods," Expert Syst. Appl., vol. 260, p. 125396, Jan. 2025, doi: 10.1016/j.eswa.2024.125396.

[27] C. McKenna, S. Clarke, and F. Golpayegani, "Floating Buses: Dynamic Route planning and Passenger Allocation based on Real-time Demand," in 2019 IEEE 5th International Conference on Computer and Communications (ICCC), IEEE, Dec. 2019, pp. 2203–2207. doi: 10.1109/ICCC47050.2019.9064471.

[28] J. Zhang et al., "A Real-Time Passenger Flow Estimation and Prediction Method for Urban Bus Transit Systems," IEEE Trans. Intell. Transp. Syst., vol. 18, no. 11, pp. 3168–3178, Nov. 2017, doi: 10.1109/TITS.2017.2686877.

[29] K. P. Seng, L.-M. Ang, E. Ngharamike, and E. Peter, "Ridesharing and Crowdsourcing for Smart Cities: Technologies, Paradigms and Use Cases," IEEE Access, vol. 11, pp. 18038–18081, 2023, doi: 10.1109/ACCESS.2023.3243264.

[30] W. Daamen, Modelling Passenger Flows in Public Transport Facilities. 2004. [Online]. Available: http://www.narcis.nl/publication/RecordID/oai:tudelft.nl:uuid:e65fb66c-1e55-4e63-8c49-5199d40f60e1

[31] M. Pei, P. Lin, and J. Ou, "Real-Time Optimal Scheduling Model for Transit System with Flexible Bus Line Length," Transp. Res. Rec., vol. 2673, no. 4, pp. 800–810, 2019, doi: 10.1177/0361198119837502.

[32] Y. Lu, A. Misra, W. Sun, and H. Wu, "Smartphone Sensing Meets Transport Data: A Collaborative Framework for Transportation Service Analytics," IEEE Trans. Mob. Comput., vol. 17, no. 4, pp. 945–960, Apr. 2018, doi: 10.1109/TMC.2017.2743176.

ISSN: 2961-5992

[33] R. Liu and N. Wang, “Data-Driven Bus Route Optimization Algorithm Under Sudden Interruption of Public Transport,” IEEE Access, vol. 10, pp. 5250–5263, 2022, doi: 10.1109/ACCESS.2022.3140947.

[34] G. Calabrò, G. Inturri, M. Le Pira, A. Pluchino, and M. Ignaccolo, “Bridging the gap between weak-demand areas and public transport using an ant-colony simulation-based optimization,” Transp. Res. Procedia, vol. 45, no. 2019, pp. 234–241, 2020, doi: 10.1016/j.trpro.2020.03.012.

[35] Z. Huang, Q. Li, F. Li, and J. Xia, “A Novel Bus-Dispatching Model Based on Passenger Flow and Arrival Time Prediction,” IEEE Access, vol. 7, pp. 106453–106465, 2019, doi: 10.1109/ACCESS.2019.2932801.

[36] J. Zhang, “Agent-Based Optimizing Match Between Passenger Demand and Service Supply for Urban Rail Transit Network With NetLogo,” IEEE Access, vol. 9, pp. 32064–32080, 2021, doi: 10.1109/ACCESS.2021.3060816.

[37] C. Brakewood, G. S. Macfarlane, and K. Watkins, “The impact of real-time information on bus ridership in New York City,” Transp. Res. Part C Emerg. Technol., vol. 53, pp. 59–75, Apr. 2015, doi: 10.1016/j.trc.2015.01.021.

[38] B. Steenwinckel et al., “FLAGS: A methodology for adaptive anomaly detection and root cause analysis on sensor data streams by fusing expert knowledge with machine learning,” Futur. Gener. Comput. Syst., vol. 116, pp. 30–48, Mar. 2021, doi: 10.1016/j.future.2020.10.015.

[39] M. Driss Laanaoui, M. Lachgar, H. Mohamed, H. Hamid, S. Gracia Villar, and I. Ashraf, “Enhancing Urban Traffic Management Through Real-Time Anomaly Detection and Load Balancing,” IEEE Access, vol. 12, pp. 63683–63700, 2024, doi: 10.1109/ACCESS.2024.3393981.

[40] R. Wolniak and W. Grebski, “Smart mobility in smart city – Singapore and Tokyo comparison,” Sci. Pap. Silesian Univ. Technol. Organ. Manag. Ser., vol. 2023, no. 176, 2023, doi: 10.29119/1641-3466.2023.176.44.

ISSN: 2961-5992