Data and Metadata. 2025; 4:467
doi: 10.56294/dm2025467

ORIGINAL

Predictive Energy Demand and Optimization in Metro Systems Using AI and IoT Technologies

Demanda de energía predictiva y optimización en sistemas de metro mediante tecnologías de IA e IoT

Mohammed Hatim Rziki[1], Abdelaaziz Hessane[2], Mohamed Khalifa Boutahir[2,3], Hamid Bourray[4], Moulay Driss El Ouadghiri[1], Rita Belkadi[5]

[1]Laboratory of AI, Faculty of Sciences, Moulay Ismail University of Meknes, Morocco.
[2]Engineering science and technology laboratory, IDMS Team, Faculty of Sciences and Tech niques, Moulay Ismail University of Meknes, Errachidia, Morocco.
[3]National School of Artificial Intelligence and Digital, Mohammed First University of Oujda, Berkane, Morocco.
[4]Systems Theory and Computer Science Team, Faculty of Sciences, Moulay Ismail University of Meknes, Morocco.
[5]Laboratory of PMS, Faculty of Sciences, Ibn Tofail University of Kenitra, Morocco.

Cite as: Hatim Rziki M, Hessane A, Khalifa Boutahir M, Bourray H, El Ouadghiri MD, Belkadi R. Predictive Energy Demand and Optimization in Metro Systems Using AI and IoT Technologies. Data and Metadata. 2025; 4:467. https://doi.org/10.56294/dm2025467

Submitted: 24-02-2024 **Revised:** 18-06-2024 **Accepted:** 28-10-2024 **Published:** 01-01-2025

Editor: Adrián Alejandro Vitón-Castillo

Corresponding author: Mohammed Hatim Rziki

ABSTRACT

Introduction: with the rapid urbanization of modern cities, metro systems have become indispensable for efficient mobility. However, the increasing demand for public transportation has led to rising energy consumption, posing significant challenges for operational sustainability. Current energy management strategies in metro networks rely on static models and centralized systems, which often fail to adapt to real-time fluctuations in energy demand, leading to inefficiencies and wasted resources.
Method: this paper proposes an innovative approach to optimizing energy demand in metro systems by integrating Artificial Intelligence (AI) and the Internet of Things (IoT). By leveraging real-time data collected from IoT sensors deployed throughout the metro network, we apply machine learning algorithms such as Random Forests and Neural Networks to dynamically predict energy demand. These predictions enable metro operators to adjust energy consumption in real-time, thus improving overall system efficiency and reducing operational waste. Our approach was validated using data from the Parisian metro system through extensive simulations.
Results: the results of simulations demonstrate significant improvements in energy efficiency. Optimized energy demand management led to a reduction in wasted energy during metro operations, particularly through the utilization of regenerative braking systems.
Conclusions: our findings suggest that integrating AI and IoT technologies into metro systems significantly improves energy efficiency by enabling dynamic energy demand prediction and real-time adjustment of energy consumption. The proposed system is scalable and adaptable, making it suitable for application in metro networks globally, thereby enhancing energy efficiency and supporting sustainable transport initiatives.

Keywords: Energy Demand Optimization; Artificial Intelligence (AI); Internet of Things (IoT); Metro Systems; Machine Learning.

RESUMEN

Introducción: con la rápida urbanización de las ciudades modernas, los sistemas de metro se han hecho indispensables para una movilidad eficiente. Sin embargo, la creciente demanda de transporte público ha

provocado un aumento del consumo de energía, lo que plantea importantes retos para la sostenibilidad operativa. Las estrategias actuales de gestión de la energía en las redes de metro se basan en modelos estáticos y sistemas centralizados, que a menudo no se adaptan a las fluctuaciones en tiempo real de la demanda de energía, lo que provoca ineficiencias y desperdicio de recursos.
Método: este documento propone un enfoque innovador para optimizar la demanda de energía en los sistemas de metro mediante la integración de la Inteligencia Artificial (IA) y el Internet de las Cosas (IoT). Aprovechando los datos en tiempo real recogidos de los sensores IoT desplegados en toda la red de metro, aplicamos algoritmos de aprendizaje automático como Random Forests y Redes Neuronales para predecir dinámicamente la demanda de energía. Estas predicciones permiten a los operadores de metro ajustar el consumo de energía en tiempo real, mejorando así la eficiencia general del sistema y reduciendo los residuos operativos. Nuestro planteamiento se validó con datos del metro de París mediante extensas simulaciones.
Resultados: los resultados de las simulaciones demuestran mejoras significativas en la eficiencia energética. La gestión optimizada de la demanda de energía condujo a una reducción de la energía desperdiciada durante las operaciones de metro, en particular mediante la utilización de sistemas de frenado regenerativo.
Conclusiones: nuestros hallazgos sugieren que la integración de las tecnologías de IA e IoT en los sistemas de metro mejora significativamente la eficiencia energética al permitir la predicción dinámica de la demanda de energía y el ajuste en tiempo real del consumo de energía. El sistema propuesto es escalable y adaptable, por lo que es adecuado para su aplicación en redes de metro a nivel mundial, mejorando así la eficiencia energética y apoyando iniciativas de transporte sostenible.

Palabras clave: Optimización de la Demanda Energética; Inteligencia Artificial (IA); Internet de las Cosas (IoT); Sistemas de Metro; Aprendizaje Automático.

INTRODUCTION

As cities continue to expand, the demand for efficient, sustainable, and scalable public transportation systems has intensified. Metro systems, playing a crucial role in urban mobility, confront the challenge of enhancing their operational capacity while simultaneously lowering their energy usage.[1] In recent years, the growing emphasis on sustainability has spotlighted the importance of energy-efficient metro systems, especially considering the substantial energy these networks consume daily. According to the International Transport Forum (ITF),[2] urban transportation is responsible for nearly 40 % of all CO2 emissions from the transport sector, with metro systems playing a significant role in energy consumption due to their continuous operation and high passenger volumes. In large cities, public transit networks, including metros, consume up to 30 % of the total energy used for transportation, highlighting the urgent need for energy-efficient solutions and sustainable technologies to reduce this environmental impact.[3]

Because of global competition, smart cities have become a necessity, particularly in terms of energy efficiency and reducing transportation and logistics costs, which account for a significant portion of urban financial resources. In this context, advanced technologies have become indispensable for improving metro systems' operational efficiency, energy conservation, and sustainability. This study focuses on integrating Artificial Intelligence (AI) and the Internet of Things (IoT) to optimize energy management in metro systems, especially through energy recovery from regenerative braking systems.[4,5]

The Paris metro system,[6] one of the largest and most complex globally, serves as a model for this research. Established in 1900, it now carries over 1,5 billion passengers annually and consists of 16 lines serving 303 stations, including two fully automated lines. The intricate nature of this metro system, which operates from 5:30 AM to 1:15 AM on weekdays, makes it an ideal case study for testing innovative energy management solutions. With its extensive network of 225 km of track, mostly underground but with some aerial segments, the Paris metro is a critical transportation artery that offers substantial potential for energy recovery and optimization.[6]

Energy demand in metro networks is highly variable, driven by passenger volumes, train schedules, and real-time operational conditions. Traditional energy management systems in metros, while functional, often fail to account for these dynamic variables, leading to inefficient energy use and increased operational costs.[7] For instance, while regenerative braking, which converts kinetic energy from decelerating trains into electrical energy, presents a significant opportunity for energy recovery, many systems struggle to effectively manage and maximize this energy. This gap between potential and performance highlights the need for a more responsive and adaptive approach to energy management in metro systems.[8,9]

Current methods for managing energy in metro systems typically rely on static models and centralized control systems. These approaches frequently use historical data to forecast energy demand and schedule energy recovery processes, but they are unable to adjust in real-time to the unpredictable fluctuations in passenger demand and operational conditions.[10] Energy storage is a commonly used method that captures regenerative energy from braking in batteries or supercapacitors for later use. However, capacity constraints and degradation over time limit these storage solutions, reducing their long-term effectiveness. Another traditional approach is demand-side management (DSM), which focuses on adjusting energy consumption patterns through predefined schedules. While DSM helps in reducing peak demand, it struggles to adapt to real-time operational changes, thus limiting its effectiveness in a dynamic metro environment.[11,12,13]

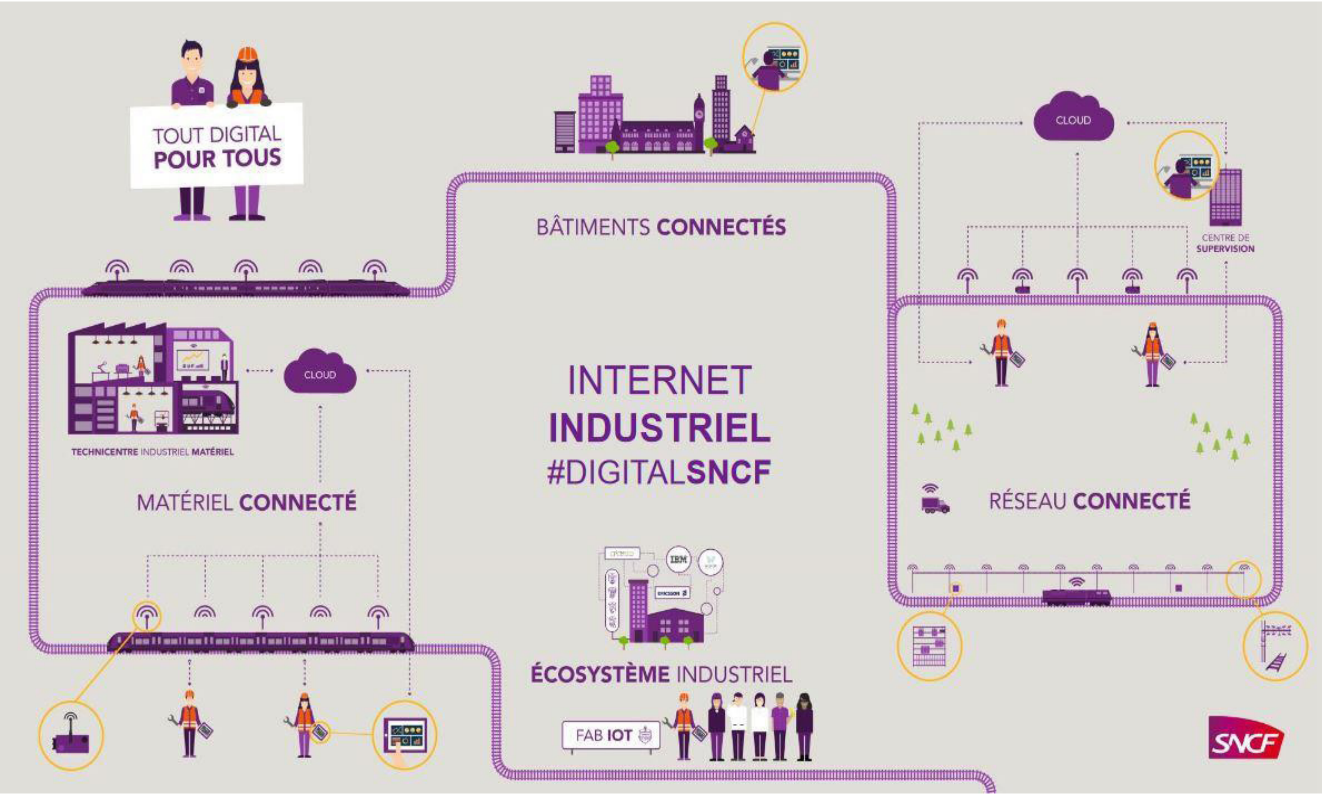

Figure 1. IoT Integration Across SNCF's Transportation Network [14]

To address these challenges, advancements in Artificial Intelligence (AI) and the Internet of Things (IoT) have opened new avenues for optimizing energy management in metro systems. AI's predictive capabilities, combined with IoT's real-time data collection, present a promising solution for adapting energy consumption based on real-time demand.[15] IoT sensors, embedded throughout the metro system, provide granular data on operational variables such as passenger volumes, braking events, and environmental conditions.[16] AI algorithms process this data, predicting future energy demand with high accuracy. Figure 1 illustrates the application of IoT technology in various domains within the SNCF network, emphasizing the interconnectedness of digital systems for optimizing energy and operations. This showcases the potential of IoT-driven solutions to revolutionize metro energy management.

AI's role in this context is to bridge the gap between energy generation and consumption. Machine learning models, trained on historical data from metro operations, are capable of forecasting energy demand based on variables like the number of passengers, speed, and braking frequency.[17] When coupled with IoT data, these models can provide real-time insights, enabling metro systems to optimize energy recovery and consumption dynamically. Moreover, the integration of AI and IoT can facilitate predictive maintenance of metro systems by identifying potential issues before they lead to failures, further enhancing operational efficiency. This approach not only helps to maximize the use of regenerative energy, but it also reduces the overall carbon footprint of metro operations.

To provide a comprehensive understanding of the current state of research and the context for this study, this section reviews key works related to the intersection of artificial intelligence (AI), the Internet of Things (IoT), and energy management in metro systems. Recent advances in smart city technologies have emphasized the role of AI and IoT in optimizing urban transportation networks. As cities grow and public transportation systems expand, ensuring energy efficiency has become a critical goal. Studies have delved into the integration of AI and IoT to monitor, predict, and optimize energy consumption in metro systems. This section will delve into three primary areas of focus: the application of AI and IoT in smart city transportation, machine learning techniques for energy demand prediction, and innovative energy recovery methods such as regenerative braking in metro systems. By analyzing these topics, we aim to highlight the gaps in existing literature and position our proposed model within the broader context of energy-efficient public transport solutions.

AI and IoT in Smart City Transportation Systems

The integration of Artificial Intelligence (AI) and the Internet of Things (IoT) in smart city transportation systems has revolutionized urban mobility by enhancing efficiency, safety, and sustainability. Smart cities, which rely on interconnected networks of sensors and devices, use real-time data to optimize traffic management, reduce energy consumption, and improve overall urban mobility experiences. AI-driven predictive models, combined with IoT's vast data collection capabilities, allow cities to better manage transportation systems by providing real-time insights into traffic flows, energy consumption, and service demands. This subsection explores recent studies that delve into the integration of AI and IoT in smart transportation, highlighting various applications and innovations that have emerged in the field.

Alanazi et al.[18] present a study that examines the development of smart mobility infrastructure in Saudi Arabia, using a benchmarking approach based on smart mobility systems in three Asia-Pacific countries: South Korea, Singapore, and Japan. Their research identifies sixty key indicators, such as IoT, 5G, drones, and self-driving technologies, that contribute to the success of smart mobility systems. The authors highlight Saudi Arabia's potential to replicate these countries' strategies by focusing on inclusive development, policy-driven strategies, and the integration of AI and IoT. The study emphasizes the need for cybersecurity measures due to the vulnerability of IoT technologies to hacking, which could impact the safety and reliability of the transportation system. Li et al.[19] explore a practical application of IoT in public transportation by using CO2 sensors to count passengers in public buses. Urban transit systems can potentially use low-cost IoT devices in their case study, where CO2 sensors can estimate passenger headcounts based on carbon dioxide levels inside buses. The authors employed machine learning models to account for non-linear relationships between passenger numbers and CO2 concentration, achieving promising accuracy despite challenges such as varying air flows and environmental pollution. This study showcases how IoT, when combined with AI, can enhance the operational efficiency of public transportation.

Bachechi et al.[20] focus on anomaly detection in IoT sensor networks for smart cities, particularly in the context of traffic flows and public transport. Their study presents a Python implementation of two algorithms—ST-BDBCAN and ST-BOF—that detect anomalies in spatial time series data. Their approach, tested on Modena traffic data, demonstrates the effectiveness of IoT and AI in real-time traffic monitoring and anomaly detection, which can significantly enhance smart city operations by ensuring data reliability and optimizing transport services. França et al.[21] provide a broader analysis of the smart city ecosystem, emphasizing the role of AI, IoT, and other advanced technologies in urban mobility and infrastructure optimization. Their study highlights the importance of these technologies in improving public services, promoting sustainability, and enhancing the overall quality of life in urban environments. They discuss the interconnected functioning of public transport systems, traffic management, and public services via advanced ICT, demonstrating how smart city ecosystems can drive efficient urban operations.

Machine Learning for Energy Demand Prediction in Metro Systems

Machine learning has emerged as a critical tool in optimizing energy consumption within metro systems, offering the potential to predict energy demand with a high degree of accuracy based on real-time and historical data. By leveraging vast datasets provided by IoT sensors and other monitoring tools, machine learning algorithms can model complex relationships between operational variables such as passenger loads, train schedules, and environmental conditions. Machine learning in energy demand prediction not only allows for real-time adjustments to energy usage, but also improves the efficiency and sustainability of metro systems by optimizing the use of regenerative energy. This subsection reviews recent studies that explore the application of machine learning in urban mobility and energy management, highlighting their relevance to metro systems.

Zhang et al.[22] provide a comprehensive review of machine learning approaches for modeling energy consumption in electric vehicles (EVs). Although focused on EVs, their study is highly relevant to metro systems as it explores how neural networks can enhance prediction accuracy by extracting key features and identifying patterns in energy consumption data. The review highlights the significance of interpretable machine learning models in enhancing the transparency and practicality of these predictions, a concept that could also apply to the prediction of energy demand in metro networks. Ji et al.[23] analyze urban transport emissions and energy consumption through machine learning and deep learning techniques. While the study primarily focuses on forecasting transport-related CO2 emissions in China, the application of machine learning models like Random Forest and Neural Networks extends to the prediction of energy demand in metro systems. The study's emphasis on predicting future energy needs based on variables such as population, car kilometers, and GDP per capita demonstrates the potential of machine learning to make long-term forecasts that inform energy management strategies in transportation systems.

Justin et al.[24] present a model for automated climate control in metro systems using an optimal ensemble learning technique. Their method uses long short-term memory (LSTM), gated recurrent units (GRU), and recurrent neural networks (RNN) to predict indoor air quality variables like temperature and humidity in order to make metro trains and stations use less energy. Despite the study's focus on climate control, the ensemble learning approach offers valuable insights into the application of complex machine learning models for energy demand prediction in metro systems, particularly in optimizing operational conditions for maximum energy efficiency. Chen and Zhang[25] introduce a framework for intelligent transportation systems (ITS) that integrates machine learning to optimize urban mobility. Their research highlights the use of a multilayer objective function and hybrid algorithms, such as ANN-RNN, to model the interaction costs between transportation modes, energy consumption, and environmental impact. While this study covers urban mobility broadly, the machine learning techniques discussed are highly applicable to metro systems, where real-time optimization of energy use can significantly improve sustainability and efficiency.

Energy Efficiency and Regenerative Braking in Rail Transport

Energy efficiency is a pivotal concern in modern rail transport systems, particularly given the substantial energy consumption and associated operational costs. Trains can recover kinetic energy during deceleration and reuse it within the rail network or store it for future use, thanks to the emerging key technology of regenerative braking. This not only reduces energy consumption but also contributes to lowering the carbon footprint of rail systems, aligning with global sustainability goals. The following studies explore various strategies and technologies to enhance energy efficiency in rail transport, with a particular focus on the role of regenerative braking.

Yildiz et al.[26] propose a timetable optimization model for the Istanbul metro network that incorporates a regenerative braking energy model. Their integrated optimization method aligns the speed trajectories of braking and accelerating trains to increase the usage of regenerative energy while simultaneously optimizing the train timetable. The study uses real operational data from the Istanbul M3 subway system to show how genetic and simulated annealing algorithms can improve schedule efficiency, which leads to a big drop in the amount of traction energy used. Sun et al.[27] present an optimal control strategy for metro trains to maximize regenerative braking energy (RBE) absorption by adjusting the speed profile of adjacent trains. Their approach improves the overlap of traction and braking processes among trains, ensuring that regenerative energy is used from one train to the next. The model, tested on the Beijing Metro Line 4, demonstrated a 13 % improvement in energy efficiency, proving the robustness of their method even when random train delays were considered. Sharma et al.[28] focus on the environmental sustainability of rail operations by assessing the impact of energy-efficient technologies such as regenerative braking and lightweight materials. Their case study-based methodology evaluates the energy usage and emissions produced by rail systems, providing valuable insights into how these innovations can enhance sustainability. The study emphasizes the broader implications of adopting such technologies, contributing to the U.N. Sustainable Development Goals 2030.

Župan et al.[29] introduce an energy flow control algorithm for trams based on Pontryagin's Minimum Principle, designed to optimize energy savings during regenerative braking. Their algorithm controls the interaction between a supercapacitor energy storage system and the power grid, maximizing energy savings while minimizing the impact on the grid. Real-time simulations demonstrated the feasibility of this approach, validating the algorithm in practical tram operations using a real-time laboratory emulation setup. Zhu et al.[30] propose an energy-efficient timetabling approach that takes into account varying train loads and realistic speed profiles for bi-directional metro lines. Their model aims to improve regenerative energy utilization by optimizing dwell times and non-uniform headways, which have a significant impact on energy consumption. By using an adaptive large neighborhood search (ALNS) algorithm, the study is able to account for the complexities of energy-efficient timetabling in real-world scenarios. Numerical experiments, conducted on Shanghai Metro Line 17, show that their approach can effectively reduce energy consumption by up to 11,9 % during off-peak hours, making it a viable solution for improving the overall energy efficiency of metro systems. The model's ability to adjust to varying operational conditions highlights its potential for broader application in dynamic metro networks.

Overview and Research Gap

The literature review that was done on three main areas—AI and IoT in smart city transportation systems, machine learning for predicting energy demand in metro systems, and energy efficiency through regenerative braking—shows that modern technologies are being used in more advanced ways to make urban transportation systems use less energy and run more efficiently. However, several gaps remain that underscore the need for further research and innovation.

In AI and IoT integration, while smart cities have seen increasing adoption of IoT sensors and AI-driven models to enhance urban mobility, many existing studies focus on isolated applications, such as monitoring traffic flows or improving user experience. The potential for AI and IoT to provide a more holistic, dynamic solution for energy management, particularly in metro systems, is underexplored. Few works focus on the specific role of AI in conjunction with IoT to enable real-time optimization of energy recovery processes, especially from regenerative braking events, which is a core focus of our research.

While machine learning techniques have successfully predicted energy demand in various contexts, their application in metro networks is still in its infancy. Instead of concentrating on the intricate patterns of energy consumption in metro systems, often influenced by highly variable factors such as passenger load, braking events, and operational schedules, most research focuses on broader urban mobility issues. This gap underscores the necessity for machine learning models specifically designed to navigate the intricacies of metro networks and their dynamic energy requirements. Additionally, current models often rely on historical data and static approaches, limiting their ability to adapt to sudden changes in operational conditions. Despite significant efforts to optimize energy recovery in metro systems for regenerative braking, the limitations of centralized energy storage and control systems continue to constrain many of these approaches. These models fail to account for the dynamic nature of metro operations and the real-time decision-making needed to maximize energy recovery efficiency. Moreover, the focus on traditional methods like supercapacitors and batteries to store energy leaves room for more innovative solutions, such as blockchain-based decentralized systems, which can offer more secure, scalable, and transparent energy management.

Our proposed methodology addresses these gaps by integrating AI, IoT, and blockchain technology into a unified framework for real-time energy optimization in metro systems. By leveraging IoT sensors to collect real-time data, machine learning algorithms to predict energy demand, and a blockchain-based Proof-of-Work (PoW) system to validate energy recovery, our approach offers a more adaptive, efficient, and secure solution to energy management. In addition to enhancing the accuracy of energy predictions, this dynamic, decentralized system guarantees transparent and tamper-resistant validation of energy recovery events, thereby establishing a new benchmark for sustainability and operational efficiency in public transportation networks.

The key research questions driving this study are as follows:

- How can we leverage real-time data from IoT sensors to optimize energy management in metro systems?
- What role can AI play in predicting energy demand and reducing energy consumption during metro operations?
- How effective are machine learning algorithms like Random Forests and Neural Networks in modeling energy demand for metro networks?
- What potential energy savings and operational improvements can AI-driven energy optimization achieve?

This paper aims to:

- Develop an AI-driven model for predicting energy demand in metro systems using real-time IoT data.
- Integrate IoT sensors and machine learning algorithms to dynamically optimize energy consumption and recovery.
- Evaluate the energy savings and improvements in system efficiency achieved by the proposed model using real-world data from the Paris metro system.
- Analyze the scalability of the proposed system for application in larger metro networks globally.

This study's contributions include the development and testing of an AI-IoT integrated approach that dynamically adapts to the energy needs of metro systems in real-time. By validating this system through extensive simulations on real-world metro data, we demonstrate the potential of AI and IoT to revolutionize energy management in public transportation, offering a pathway to more sustainable and efficient urban mobility systems.

We structure the remainder of this paper as follows: Section 2 details the methodology used in this research, including the data collection process, the machine learning models employed for energy demand prediction, and the integration of IoT sensors. The results of the simulations using real-world data from the Parisian metro system are shown in Section 3. These results show how well the proposed AI-driven approach works at improving energy recovery. Finally, Section 4 concludes the paper, summarizing the key findings and proposing directions for future research.

METHOD

The methodology employed in this study aims to develop a comprehensive predictive model for energy demand in metro systems, leveraging Artificial Intelligence (AI) and Internet of Things (IoT) technologies. This is a quantitative, predictive modeling study designed to estimate energy demand using real-time and historical data from metro systems. Given the significant variability in metro system operations and the dynamic nature of passenger demand, the methodology focuses on using real-time data from the transport network to accurately predict energy needs and optimize resource utilization. This approach emphasizes an adaptive and data-driven model to enhance energy efficiency and reduce operational costs in urban rail systems.

Universe and Sample

The universe of this study comprises operational data from metro systems worldwide, with a focus on the Parisian metro network as a representative case study. The sample consists of over 2 710 092 records collected from the Parisian metro's General Transit Feed Specification (GTFS) data, covering detailed information on routes, stops, schedules, and passenger volumes over a one-year period.

We structure our methodology around three core phases: data preparation and feature engineering, demand estimation through machine learning, and model training and validation.

Data collection

The data preparation process began by collecting and integrating operational data from metro systems, particularly the Parisian metro network. The data sources included detailed information on routes, stops, schedules, and passenger volumes, allowing the development of a rich dataset that encapsulates the operational aspects of the metro system. Feature engineering involved extracting relevant variables such as temporal factors (e.g., time of day, day of the week) and demand-related features that play a crucial role in influencing energy consumption.

Statistical Treatment

For demand estimation, the methodology employed clustering techniques to identify patterns in passenger demand and segment the data into high-demand periods. Clustering, specifically using K-means, helped distinguish different operational states in the metro network, capturing the temporal variations in passenger flow and enabling the model to be sensitive to real-world fluctuations in demand. We applied machine learning algorithms after clustering to predict energy consumption based on these patterns. We evaluated a variety of regression models, including Random Forest, Decision Trees, and XGBoost, for their ability to handle the complex relationships between input features and energy demand, focusing on models that offer robustness, accuracy, and interpretability.

The statistical treatment involved data preprocessing steps such as handling missing values, normalization, and encoding of categorical variables. Clustering was performed using the K-means algorithm, with the optimal number of clusters determined by the Elbow method and silhouette analysis. For predictive modeling, supervised machine learning techniques were applied, and models were evaluated using cross-validation and performance metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score.

The final phase involved model training and validation, which ensured the model's reliability and ability to generalize across various operational scenarios. We trained the model using a subset of the collected data and reserved the remaining portion for validation. Performance was assessed using metrics like the coefficient of determination (R^2) and Mean Absolute Percentage Error (MAPE), providing a quantitative measure of the model's predictive accuracy. Additionally, we conducted feature importance analysis to identify the most influential variables affecting energy demand, enhancing the model's explainability and potential use in real-world metro system operations.

Replication Details

To facilitate replication of the study, all data processing and modeling steps were conducted using Python, utilizing libraries such as pandas for data manipulation, scikit-learn for machine learning algorithms, and matplotlib for data visualization. The code and scripts used for data preprocessing, feature engineering, clustering, and model training are documented and available upon request. Hyperparameters for machine learning models were tuned using GridSearchCV, and a fixed random state was set to ensure reproducibility of results.

This multi-phase methodology, which integrates real-world operational data, machine learning, and clustering techniques, aims to provide a scalable and adaptive framework for energy demand prediction in metro systems. By incorporating both predictive and optimization capabilities, the methodology offers a pathway toward more efficient energy management in public transportation, aligning with the goals of sustainable urban mobility.

Data Preparation and Feature Engineering

The data utilized for this study comes from the General Transit Feed Specification (GTFS), which provides comprehensive details on public transit systems. The dataset is composed of various core files, including temporal, spatial, and trip-specific information, crucial for predicting high-demand periods for energy optimization in urban metro systems. To achieve a unified and enriched dataset, these files were systematically merged based on unique identifiers. Table 1 summarizes the core GTFS files and their roles in the dataset.

Table 1. Overview of Core GTFS Data Files

File Name	Description	Key Columns
stop_times.txt	Details each stop along a trip, including arrival and departure times.	trip_id, arrival_time, stop_id
trips.txt	Provides information on trip schedules and associated routes.	trip_id, route_id
stops.txt	Contains geographical data for each stop, such as coordinates and names.	stop_id, stop_lat, stop_lon
calendar.txt & calendar_dates.txt	Defines service dates and exceptions for trips.	service_id, date, exception_type
routes.txt	Captures route details, including types and names.	route_id, route_name

The merging process combined these files using primary keys (trip_id, stop_id, route_id) to create a holistic dataset, which integrates temporal details, trip data, and stop information. The goal was to enable effective prediction of energy demand in real-time by utilizing AI models.

To enhance the predictive power, extensive feature engineering was carried out on the unified dataset. Features were categorized into temporal and non-temporal attributes:

Temporal Features

- arrival_hour: The hour at which the stop occurs.
- day_of_week: Day of the week encoded as (0 = Monday, ..., 6 = Sunday).
- hour_of_day: Specific hour for deeper temporal analysis.
- is_weekend: Binary indicator to distinguish weekend (0 = weekday, 1 = weekend).
- is_peak_hour: Binary value indicating whether the stop time falls within peak hours.
- week_of_year: Week number to track seasonal patterns.
- part_of_day_Afternoon, part_of_day_Evening, part_of_day_Morning: Categorical binary indicators denoting different periods of the day.

Non-Temporal Features

- stop_sequence: Sequence of the stop in its respective trip.
- pickup_type & drop_off_type: Mode of passenger pickup and drop-off (scheduled or on-demand).
- route_type: Type of service provided (bus, tram, metro).
- stop_lat & stop_lon: Latitude and longitude for geographical positioning.
- wheelchair_boarding: Accessibility indicator for wheelchair access.
- exception_type: Specifies exceptional cases (e.g., holidays).
- demand_score: A calculated numeric score representing demand, which serves as the target variable for predictive modeling.

Figure 2 presents the positioning of IoT sensors in the metro system (KPVA, KFS, and RPS operating diagram), which is critical for real-time data collection and predictive modeling.

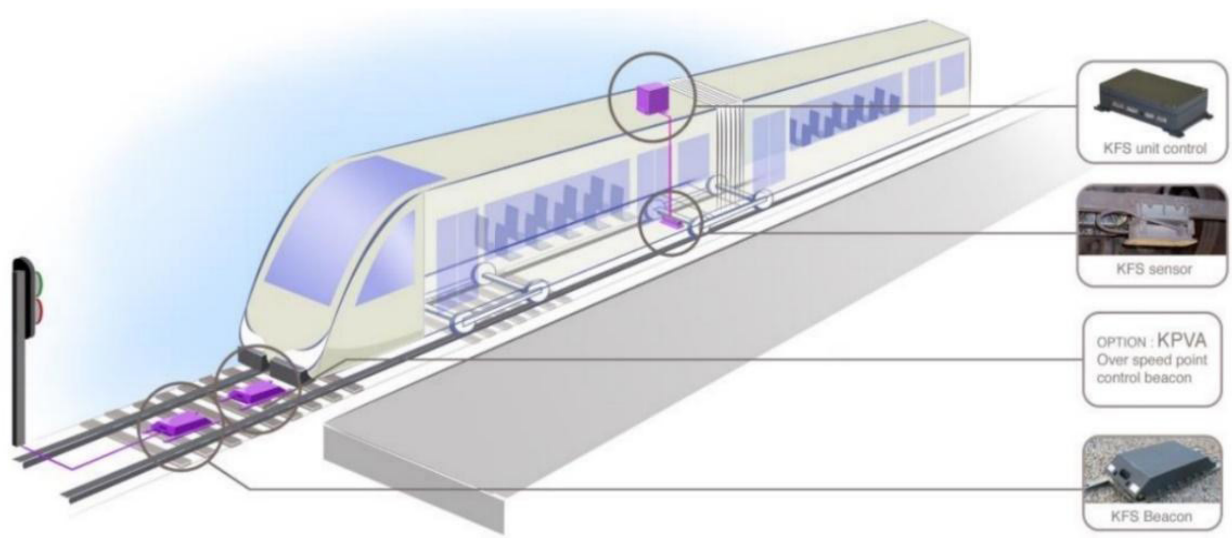

Figure 2. KPVA, KFS and RPS operating diagram

https://doi.org/10.56294/dm2025467

This enriched dataset, with its robust temporal and spatial features, serves as the foundation for building machine learning models aimed at optimizing energy demand and usage across urban metro systems. By incorporating these engineered features, the predictive model can capture complex patterns in demand, ultimately contributing to more sustainable energy management.

Demand Estimation and Clustering

The primary objective of this study is to estimate periods of high energy demand within metro systems by leveraging public transit data. Accurate demand estimation is crucial for optimizing energy consumption, particularly during peak periods when efficient energy management can lead to significant cost savings and reduced environmental impact. To achieve this, we constructed several demand-related features designed to capture the complex patterns and behaviors of trips across different times and stops.

Demand Clusters

To identify and group trips with similar demand characteristics, clustering techniques were employed. We applied k-means clustering, a widely used machine learning algorithm for segmentation, to categorize trips based on a variety of temporal and stop-specific features. The features used for clustering included:

- Temporal features such as arrival_time and day_of_week to capture when trips occur and their temporal distribution.
- Stop-specific features like stop_frequency (the number of times a stop is visited within a specific period) and trip_frequency (the number of trips occurring on a particular route).

By combining these temporal and spatial attributes, the k-means algorithm segmented the data into clusters that reflect varying demand levels across trips and stops. The resulting clusters provide a hierarchical representation of demand within the metro network, offering valuable insights into peak travel periods, regularity of trips, and the dynamics of stop usage over time. The optimal number of clusters (k) was determined based on performance metrics such as the silhouette score and within-cluster sum of squares (WCSS) to ensure well-separated and meaningful clusters.

High Demand Classification

Following the clustering process, each cluster was analyzed to determine whether it represented a period of high demand. A binary feature called High_Demand was introduced to distinguish between high and low demand clusters. The threshold for categorizing a cluster as “high demand” was based on the number of trips within that cluster. Specifically:

Clusters containing more than 25 % of the total trips were labeled as high demand (1).

Clusters representing the remaining 75 % or fewer of the trips were classified as low demand (0).

This threshold was established to effectively capture peak periods, such as rush hours, where energy usage tends to spike. By defining a clear high-demand threshold, we created a feature that assists in quickly identifying periods requiring optimized energy management.

Demand Score

To quantify the demand intensity for each trip more precisely, a continuous demand_score was calculated. This score provides a nuanced measurement of demand by accounting for both trip frequency and stop frequency. The formula used to derive the demand score is illustrated in formula 1.

$$demand\_score = \frac{trip\_frequency \times stop\_frequency}{total\_frequency} \tag{1}$$

where:

- trip_frequency: The number of trips occurring on a specific route or within a particular timeframe.
- stop_frequency: The number of times a stop is visited during the same period.
- total_frequency: The sum of all trip and stop frequencies within a designated time window.

The demand_score offers a relative ranking of trips based on their overall demand within the metro system. By integrating both spatial (stop-level) and temporal (trip-level) features, this metric captures variations in demand across different routes and times, providing a robust indicator for predicting energy needs.

After incorporating the clustering outcomes and demand estimation features, the final dataset was constructed, combining both temporal and non-temporal features alongside demand metrics. This comprehensive dataset comprises over 2 710 092 records and contains 14 features after data cleaning and preprocessing. The features include temporal information (e.g., hour, day, and part of the day), spatial attributes (e.g., latitude,

https://doi.org/10.56294/dm2025467

longitude), trip-specific features (e.g., sequence, frequency), and demand estimation metrics (High_Demand, demand_score).

Model Training and Validation

The core of this research lies in developing predictive models capable of accurately estimating energy demand in urban metro systems. These models, designed to forecast demand based on real-time and historical data, utilize both temporal and non-temporal features to identify patterns that influence energy consumption. This section outlines the training process, model selection, feature engineering, and validation techniques used to ensure robust and accurate predictions.

The model training process involved selecting suitable machine learning algorithms that could effectively capture the complex patterns present in the data. A range of supervised learning models were considered, including:

Random Forests (RF)

Random Forests are ensemble learning methods that utilize multiple decision trees to improve predictive performance. Each tree is trained on a subset of the data, and the final prediction is determined by aggregating the outputs of all the trees, which reduces overfitting and variance. RF is particularly known for its ability to handle large datasets and capture non-linear relationships between features and the target variable, the model provides feature importance, enabling interpretability in understanding which variables most influence the predictions.[31]

Gradient Boosting Machines (GBM)

GBMs are another form of ensemble learning that focuses on sequentially creating and optimizing decision trees. Each new tree attempts to correct the errors made by the previous ones, improving overall model accuracy. This prosting allows GBMs to model complex patterns effectively, especially in tabular data settings. The method's flexibility enables the tuning of hyperparameters such as the learning rate, the number of boosting stages, and the depth of each tree, resulting in high performance when predicting energy demand in transport systems.[32]

Neural Networks (Multi-layer Perceptron - MLP)

Neural networks, particularly the Multi-layer Perceptron (MLP) model, are a type of deep learning algorithm that consists of multiple layers of nodes (neurons) that process input data. These networks are highly effective in learning complex, non-linear relationships in the data and have been widely used in applications where intricate patterns need to be captured. For energy demand, MLPs are capable of processing both temporal and non-temporal features by adjusting the weights and biases through backpropagation. Various hyperparameters such as the number of hidden layers, activation functions, and neurons per layer were tuned to maximize prediction accuracy.[33]

Support Vector Machines (SVM)

SVMs are supervised learning models that perform well in high-dimensional spaces and are used for both classification and regression tasks. The core idea of SVMs is to find the optimal hyperplane that separates different classes or predicts the target variable with maximum margin. Their capacity to utilize drnel functions (linear, polynomial, RBF, etc.) enables them to effectively handle complex, non-linear relationships, making them valuable for exploring the energy demand data and identifying high-dimensional patterns that affect prediction accuracy.[34]

Each of these models was trained to predict the demand_score as a continuous target variable, providing a forecast of energy demand across different routes, times, and conditions. The choice of models was informed by their respective strengths in handling temporal data, categorical variables, and non-linear relationships within the dataset.

Featur Engineering process

To enhance the predictive power of the models, comprehensive feature engineering was performed on the dataset. Given the diversity of the data, the features used for training spanned various categories:

Temporal Features

These included attributes like arrival_hour, day_of_week, and hour_of_day, which provided context on when trips occurred. Additional temporal features like is_weekend, is_peak_hour, and parts of the day (e.g., morning, afternoon, evening) were engineered to capture cyclic patterns and variations across different time windows.

Spatial Features

Location-specific variables like stop_lat, stop_lon, and stop_sequence were included to represent the geographical and sequential nature of the metro trips. This enabled the models to learn how spatial distribution affects demand.

Categorical and Other Contextual Features

These included pickup_type, drop_off_type, route_type, and wheelchair_boarding, which provided information on the trip characteristics and accessibility.

To transform categorical variables into a usable format for model training, one-hot encoding and label encoding were applied. Continuous features were normalized to ensure uniform scaling, particularly for algorithms sensitive to feature ranges (e.g., neural networks).

Model’s training process

The model training process was conducted as follows:

Data Splitting

The dataset was split into training (80 %) and testing (20 %) sets. The training set was used for model training and hyperparameter tuning, while the testing set was reserved for model validation and performance evaluation.

Cross-Validation

K-fold cross-validation (with k = 5) was applied to prevent overfitting and ensure model generalizability. This technique allowed the models to be trained on different subsets of the data, providing a reliable estimate of their predictive performance.

Hyperparameter Tuning

For each model, a grid search was performed over key hyperparameters to find the optimal configuration. For Random Forests, parameters like the number of trees, maximum depth, and minimum samples per leaf were tuned. In the case of GBMs, the learning rate, number of boosting stages, and maximum depth were explored. Neural Networks underwent tuning for the number of hidden layers, neurons per layer, and activation functions.

Model’s evaluation

To evaluate model performance, multiple metrics were utilized, ensuring a comprehensive assessment of prediction accuracy and robustness:

Mean Absolute Error (MAE)

This metric provided a straightforward measure of the average absolute difference between predicted and actual demand scores as depicted in equation 2.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \qquad (2)$$

Where:

- is the actual demand score.
- is the predicted demand score.
- is the number of observations.

Root Mean Squared Error (RMSE)

RMSE was employed to penalize larger errors, giving a more sensitive measure of prediction accuracy when deviations were significant. Equation 3 illustrates how the RMSE is calculated.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \qquad (3)$$

Where:

- is the actual demand score.
- is the predicted demand score.
- is the number of observations.

R² Score (Coefficient of Determination)

R^2 was used to evaluate how well the model predictions fit the actual demand data, with values closer to 1 indicating a better fit. Equation 4 shows how this coefficient is computed.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (4)$$

Where:

- is the actual demand score.
- is the predicted demand score.
- is the mean of the actual demand scores.
- is the number of observations.

Comparisons of these metrics across models allowed for the selection of the best-performing algorithm, ensuring that the chosen model was both accurate and generalizable to new, unseen data.

After training and tuning the models, their performance was validated on the test set, which contained unseen data to simulate real-world demand forecasting. By comparing model predictions to actual demand scores, the final model was selected based on its ability to balance prediction accuracy (as indicated by MAE and RMSE) and robustness (as indicated by R^2). To further validate the model's generalizability, it was also tested across different periods (e.g., weekdays vs. weekends, peak vs. non-peak hours), confirming its ability to handle varied demand scenarios within the metro system.

RESULTS AND DISCUSSION

The goal of this section is to present the results derived from the dataset analysis and predictive modeling techniques used to estimate energy demand in metro systems. Our results provide a comprehensive view of temporal, spatial, and operational factors that affect public transit demand, as well as highlight the predictive capabilities of several machine learning models. The analysis covers patterns in demand scores across different timeframes, model performances in predicting demand, and the relative importance of features that drive these patterns.

We structure the results to address key objectives: first, to understand the relationships between temporal features and demand through correlation matrices and visualizations; second, to explore how the demand score varies over time and geographical locations; third, to evaluate the performance of various machine learning models in predicting demand; and finally, to identify which features most significantly contribute to the prediction of transit demand. These insights will support the optimization of energy consumption in metro systems, fostering a more efficient and sustainable urban transportation network.

Correlation Matrix of Temporal Features

The analysis begins by examining the **correlation matrix** of the temporal features as presented in figure 3, which provides a comprehensive view of the relationships among key variables in the dataset.

The heatmap reveals how features such as arrival_hour, day_of_week, and is_weekend interact with each other and how they potentially influence the demand_score. Key findings include:

Strong correlations are observed between arrival_hour and part_of_day_Morning, Afternoon, and Evening, indicating clear associations between time of day and demand patterns.

The feature is_peak_hour correlates positively with Morning and Afternoon periods, confirming that demand peaks are mostly centered around these times. Interestingly, the correlation with Evening is negative, suggesting a lower likelihood of high demand in the evening.

The weak correlation between demand_score and temporal variables implies that while time-based patterns exist, they are not the sole determinants of transit demand. This finding supports the need for integrating non-temporal features in the predictive model to enhance accuracy.

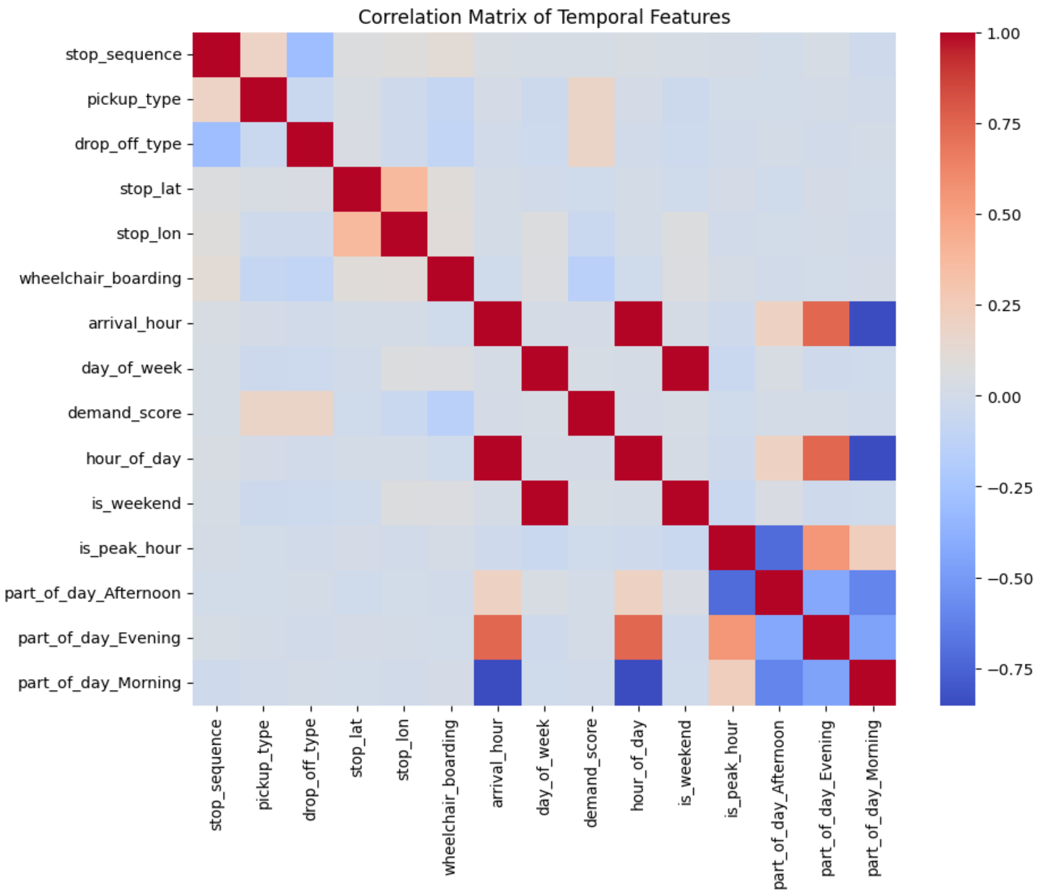

Figure 3. Correlation matrix of Temporal features

Analysis of Demand Scores

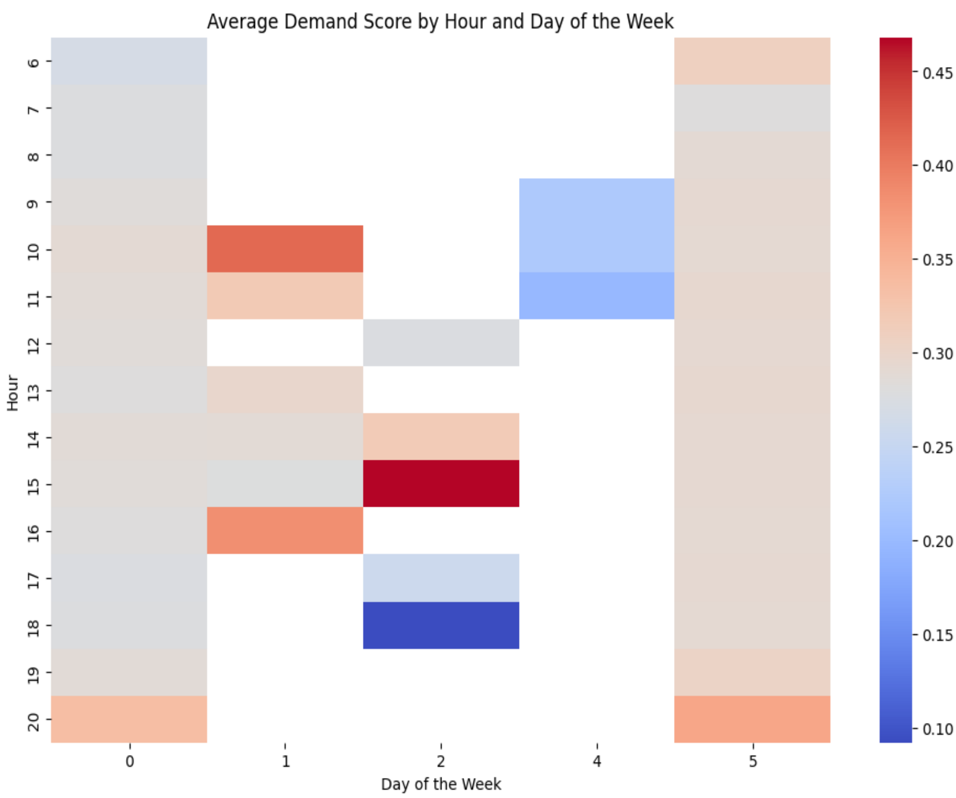

Figure 4a. Average Demand Score by Hour and Day of the Week

https://doi.org/10.56294/dm2025467

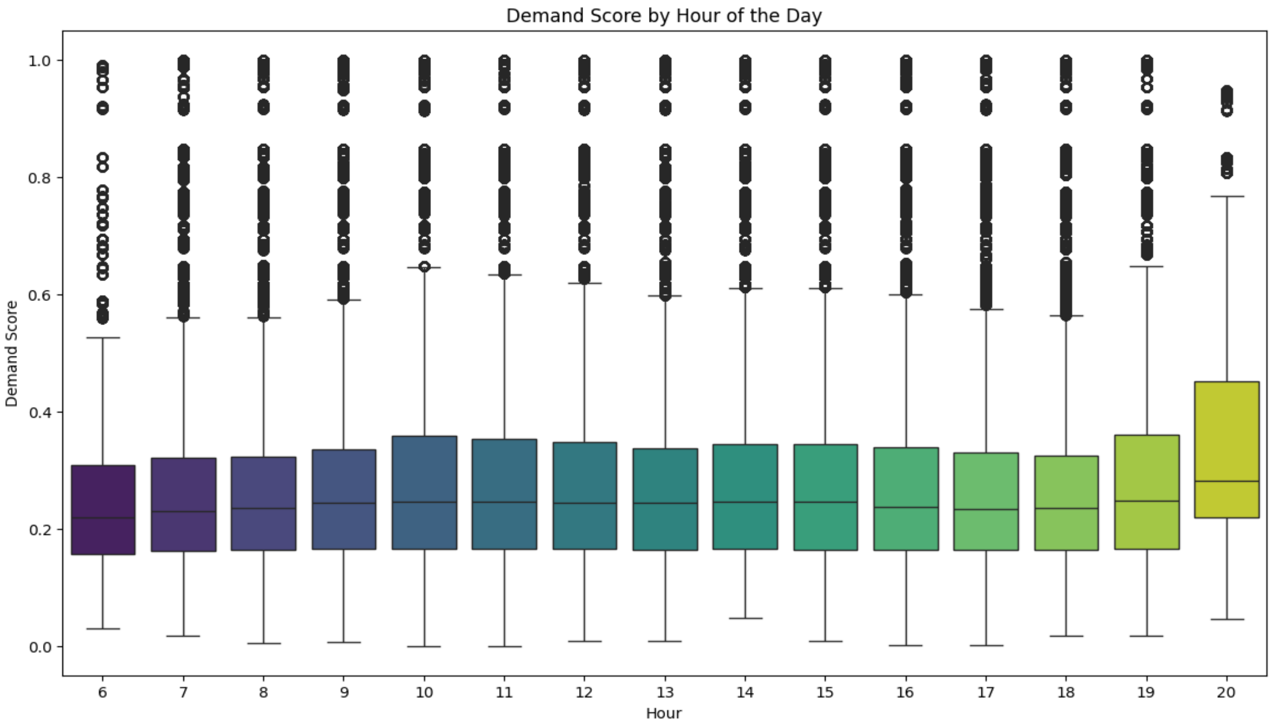

Figure 4b. Demand Score by Hour of the Day

Average Demand Score by Hour and Day of the Week

To further explore the temporal dynamics of demand, a heatmap (Figure 4-a) is used to visualize the average demand score across different hours of the day and days of the week. Key insights include:

Friday midday (10:00–11:00) and Sunday evening (around 20:00) are identified as periods of high demand, likely reflecting both commuter patterns and leisure activities.

The demand remains relatively low during early weekday mornings, with a noticeable increase during mid-afternoon (14:00–16:00), particularly on Fridays, which are often characterized by early departures from work and increased social activities.

Demand Score by Hour of the Day

The boxplot (figure 4-b) illustrates the distribution of demand scores across different hours of the day. Observations include:

Demand remains fairly stable throughout the day but shows an upward trend in the evening (19:00–20:00), indicating a rise in passenger volume during these hours.

The variability in demand is particularly high in the early morning (6:00–9:00) and evening periods (20:00), with a spread of outliers. These findings suggest significant fluctuations during these hours, possibly due to unpredictable factors like traffic, weather, or social events.

Distribution of Demand Scores

Figure 5 presents the distribution of demand scores throughout the dataset.

A clear skew towards lower demand scores, indicating that most trips fall into low-to-moderate demand categories.

Peaks around certain scores suggest that some levels of demand are more common, possibly reflecting routine patterns in transit usage. The notable frequency spikes may correspond to recurring travel behaviors like work commutes or school runs.

Model Performance and Feature Importance

Performance of Regressor Models

To predict the demand scores, multiple regression models were evaluated, and their performance metrics are illustrated in figure 6. The models were assessed based on R^2 scores and Mean Absolute Percentage Error (MAPE):

Tree-based models, such as Decision Tree, Random Forest, Extra Trees, and Extreme Gradient Boosting (XGBoost), show near-perfect R^2 scores (close to 1), demonstrating high accuracy in capturing the complexities of the demand data.

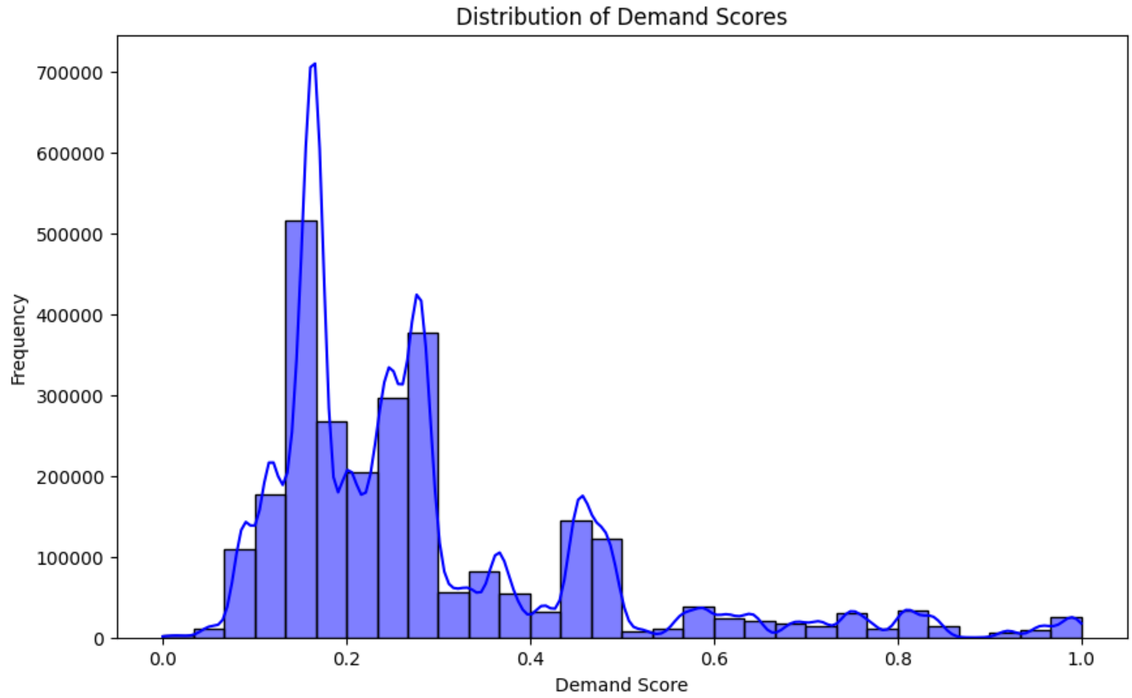

Figure 5. Distribution of Demand Scores

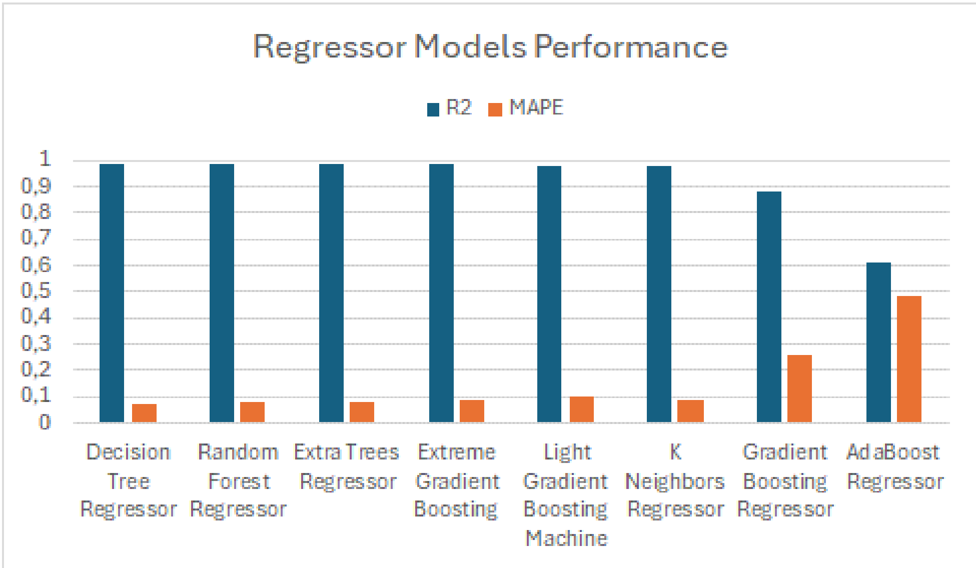

Figure 6. The performance of Regressor Models

Ensemble techniques like Gradient Boosting and AdaBoost also perform well but with slightly lower R^2 scores and higher MAPE. The models show strength in generalization, which is crucial for handling the variability in demand data.

Classical models like Linear Regression, Lasso, and Ridge show poor performance, with R^2 scores below 0,09, underscoring their limited ability to model non-linear relationships present in the dataset. Their limitations are particularly evident in a context with complex interactions between temporal and non-temporal factors affecting demand.

The superior performance of tree-based and ensemble models highlights their capability to handle the non-linear and multifaceted nature of demand data effectively. These models' inherent ability to capture interactions between variables, including spatial-temporal features and operational dynamics, makes them well-suited for this application.

Feature Importance

Figure 7 presents a feature importance plot, derived from the Random Forest model, which identifies the key features contributing to the demand prediction:

stop_lat and stop_lon are the most significant features, highlighting the importance of spatial location in determining transit demand. This finding suggests that certain geographic regions experience higher or lower demand due to factors such as population density, commercial zones, or transit connectivity.

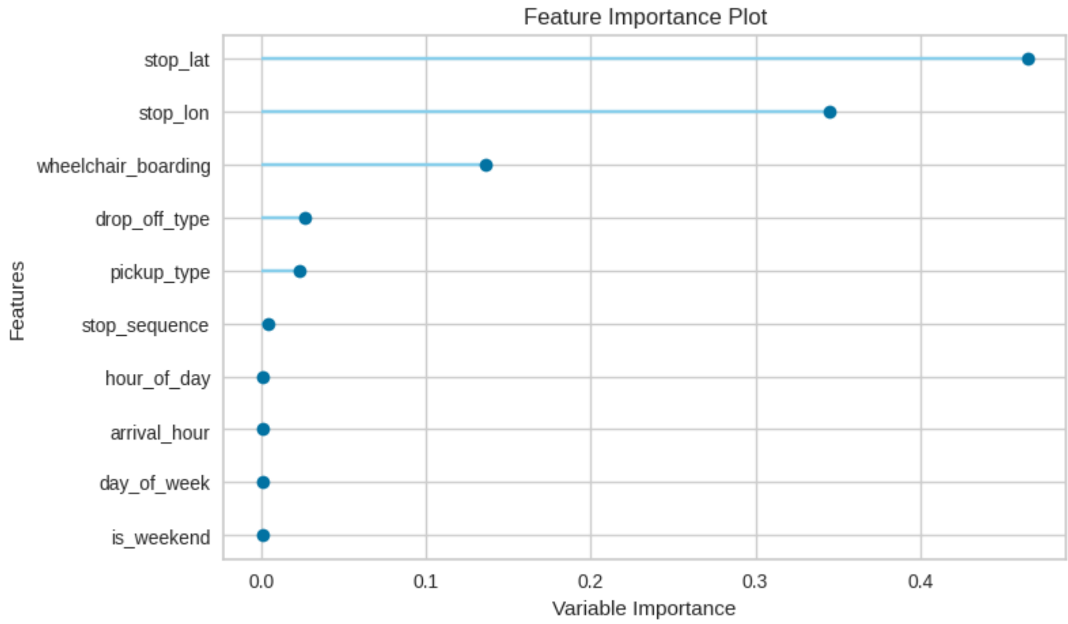

Figure 7. Feature Importance Plot

wheelchair_boarding also plays a prominent role, indicating that stops with accessibility features attract higher demand, possibly due to improved accessibility and user comfort.

Operational features like drop_off_type, pickup_type, and stop_sequence are moderately important, suggesting that the sequencing and type of service at stops influence passenger flow.

Temporal features such as hour_of_day, arrival_hour, and day_of_week have less individual significance but are nonetheless integral to understanding the temporal patterns and variability in demand.

The plot reinforces the conclusion that spatial features primarily drive demand, while operational characteristics and temporal patterns serve as supplementary factors. This highlights the need for optimizing stop locations, route planning, and service schedules to maximize efficiency.

DISCUSSION

The results demonstrate the effectiveness of combining AI and IoT technologies to enhance the predictive modeling of energy demand in metro systems. The findings reveal that spatial features, specifically latitude and longitude of stops, play the most crucial role in determining demand, indicating that geographic location and spatial coverage are significant drivers of transit usage. By understanding how spatial distribution correlates with demand, transit authorities can optimize stop locations, route planning, and scheduling to improve system efficiency and user satisfaction.

Additionally, the results underline the utility of tree-based models such as Random Forest and XGBoost, which outperform classical linear models in predicting demand scores. This is largely due to their ability to capture complex, non-linear relationships and interactions between spatial, temporal, and operational features. These models also show strong robustness against variability and outliers in the data, which is critical for reliable real-time prediction in dynamic metro environments. Moreover, the superior performance of ensemble models indicates that leveraging multiple algorithms together to minimize prediction errors can be a highly effective approach for demand estimation.

The temporal features (like hour_of_day and day_of_week) and operational characteristics (pickup_type, drop_off_type) contribute to the variability in demand, suggesting that while spatial factors are key drivers, the time of travel and type of service offered also significantly impact transit usage. For instance, the higher demand during peak hours on weekdays reflects the work commute patterns, while evening peaks on weekends likely correspond to leisure activities. Such insights can be used to allocate resources effectively, ensuring that energy supply aligns with peak and off-peak demand periods.

Furthermore, the analysis of demand clusters allows for a nuanced understanding of periods of high and low energy usage, supporting targeted interventions to manage these peaks effectively. By segmenting the data into high-demand and low-demand periods, it becomes possible to optimize energy distribution, enabling metro systems to allocate more power during peak times and conserve energy during off-peak hours.

A key contribution of this work is the integration of IoT-based real-time data collection with AI-driven predictive modeling. This combined approach provides a scalable and flexible system that adapts to changing conditions in metro operations. IoT sensors facilitate continuous data gathering on variables like passenger volume, environmental conditions, and trip frequency, providing a rich dataset for AI models to learn from and improve over time. As a result, metro systems can dynamically respond to fluctuations in demand, optimizing energy use for both operational efficiency and environmental sustainability.

Finally, the demand prediction model offers practical advantages for energy recovery and optimization. By accurately forecasting periods of high and low demand, transit operators can better manage energy storage systems and regenerative braking technologies to capture and reuse energy efficiently. This predictive capability contributes to a reduction in operational costs and environmental impact, paving the way for more sustainable and adaptive public transit networks.

In conclusion, the integration of AI and IoT technologies in demand prediction and energy optimization holds immense potential for improving the efficiency of metro systems. The ability to forecast demand based on spatial, temporal, and operational features empowers transit operators to make informed decisions that enhance service delivery while promoting sustainable energy use. Future work may expand this model to incorporate additional data sources and alternative predictive algorithms, further advancing the potential for AI-driven solutions in smart transportation systems.

CONCLUSION

This study demonstrates the significant potential of integrating Artificial Intelligence (AI) and Internet of Things (IoT) technologies for optimizing energy demand in metro systems. By leveraging comprehensive datasets and advanced machine learning techniques, we have developed a predictive framework that captures the complexities of urban transport demand. This approach provides valuable insights for transit authorities to anticipate periods of high usage and align energy resources more effectively.

Implementing this predictive model can lead to improved operational efficiency, cost savings, and environmental benefits by enabling real-time adjustments to energy consumption based on anticipated demand. The integration of AI and IoT offers a scalable and adaptable solution applicable to metro networks worldwide, supporting the development of more sustainable and responsive public transportation systems.

Future research may explore the integration of additional IoT-based data sources, alternative machine learning algorithms, and real-world implementation to further enhance predictive capabilities. Such advancements could significantly impact the evolution of smart, energy-efficient urban transportation.

REFERENCES

1. Elassy M, Al-Hattab M, Takruri M, Badawi S. Intelligent transportation systems for sustainable smart cities. Transportation Engineering. 2024;16:100252.

2. International Transport Forum [Internet]. [cited 2024 Oct 4]. Available from: https://www.itf-oecd.org/

3. Jovanovic MM. Belgrade's Urban Transport CO2 Emissions from an International Perspective. Polish Journal of Environmental Studies. 2016;25(2):635-646.

4. Esfandi S, Tayebi S, Byrne J, Taminiau J, Giyahchi G, Alavi SA. Smart Cities and Urban Energy Planning: An Advanced Review of Promises and Challenges. Smart Cities. 2024;7(1):414-444.

5. Kuo Y, Leung JM, Yan Y. Public transport for smart cities: Recent innovations and future challenges. European Journal of Operational Research. 2023;306(3):1001-1026.

6. Bonjour RATP. Lignes Metro [Internet]. 2024 Sep [cited 2024 Oct 4]. Available from: https://www.bonjour-ratp.fr/en/lignes-metro/

7. Liu X, Dabiri A, Xun J, De Schutter B. Bi-level model predictive control for metro networks: Integration of timetables, passenger flows, and train speed profiles. Transportation Research Part E: Logistics and Transportation Review. 2023;180:103339.

8. Hamada AT, Orhan MF. An overview of regenerative braking systems. Journal of Energy Storage. 2022;52:105033.

https://doi.org/10.56294/dm2025467

9. Ceraolo M, Lutzemberger G, Meli E, Pugi L, Rindi A, Pancari G. Energy storage systems to exploit regenerative braking in DC railway systems: Different approaches to improve efficiency of modern high-speed trains. Journal of Energy Storage. 2018;16:269-279.

10. Liu Y, Yan G, Settanni A. Forecasting the transportation energy demand with the help of optimization artificial neural network using an improved red fox optimizer (IRFO). Heliyon. 2023;9(11):e21599.

11. Cruz C, Tostado-Véliz M, Palomar E, Bravo I. Pattern-driven behaviour for demand-side management: An analysis of appliance use. Energy and Buildings. 2024;308:113988.

12. Tsaousoglou G, Steriotis K, Efthymiopoulos N, Smpoukis K, Varvarigos E. Near-optimal demand side management for retail electricity markets with strategic users and coupling constraints. Sustainable Energy, Grids and Networks. 2019;19:100236.

13. Bakare MS, Abdulkarim A, Zeeshan M, Shuaibu AN. A comprehensive overview on demand side energy management towards smart grids: Challenges, solutions, and future direction. Energy Informatics. 2023;6(1):1-59.

14. Silicon FRANCE. SNCF : Le train connecté à l'IoT pour la grande vitesse [Internet]. [cited 2024 Oct 4]. Available from: https://www.silicon.fr/sncf-train-iot-grande-vitesse-144642.html

15. Bibri SE, Krogstie J, Kaboli A, Alahi A. Smarter eco-cities and their leading-edge artificial intelligence of things solutions for environmental sustainability: A comprehensive systematic review. Environmental Science and Ecotechnology. 2024;19:100330.

16. Zeng F, Pang C, Tang H. Sensors on Internet of Things Systems for the Sustainable Development of Smart Cities: A Systematic Literature Review. Sensors. 2023;24(7):2074.

17. Liu Z, Gao Y, Liu B. An artificial intelligence-based electric multiple units using a smart power grid system. Energy Reports. 2022;8:13376-13388.

18. Alanazi F. Development of Smart Mobility Infrastructure in Saudi Arabia: A Benchmarking Approach. Sustainability. 2022;15(4):3158.

19. Li T, Fong S, Yang L. Counting Passengers in Public Buses by Sensing Carbon Dioxide Concentration. In: Proceedings of the 2018 2nd International Conference on Big Data and Internet of Things. ACM; 2018.

20. Bachechi C, Rollo F, Po L, Quattrini F. Anomaly Detection in Multivariate Spatial Time Series: A Ready-to-Use Implementation. In: Proceedings of the 17th International Conference on Web Information Systems and Technologies - WEBIST. SciTePress; 2021. p. 509-517.

21. França RP, Monteiro ACB, Arthur R, Iano Y. Smart Cities Ecosystem in the Modern Digital Age: An Introduction. In: Chakraborty C, Lin JCW, Alazab M, editors. Data-Driven Mining, Learning and Analytics for Secured Smart Cities. Cham: Springer; 2021.

22. Zhang X, Zhang Z, Liu Y, Xu Z, Qu X. A review of machine learning approaches for electric vehicle energy consumption modelling in urban transportation. Renewable Energy. 2024;234:121243.

23. Ji T, Li K, Sun Q, Duan Z. Urban transport emission prediction analysis through machine learning and deep learning techniques. Transportation Research Part D: Transport and Environment. 2024;135:104389.

24. Justin S, Saleh W, Lashin MM, Albalawi HM. Modeling of Artificial Intelligence-Based Automated Climate Control with Energy Consumption Using Optimal Ensemble Learning on a Pixel Non-Uniformity Metro System. Sustainability. 2022;15(18):13302.

25. Chen G, Zhang JW. Intelligent transportation systems: Machine learning approaches for urban mobility in smart cities. Sustainable Cities and Society. 2024;107:105369.

26. Yildiz A, Arikan O, Cakiroglu H. A Timetable Optimization Model for the Istanbul, Turkey, Metro Network

Considering a Novel Regenerative Braking Energy Model. Journal of Transportation Engineering, Part A: Systems. 2024;150(2).

27. Sun X, Yao Z, Dong C, Clarke D. Optimal Control Strategies for Metro Trains to Use the Regenerative Braking Energy: A Speed Profile Adjustment Approach. IEEE Transactions on Intelligent Transportation Systems. 2023;24(6):5883-5894.

28. Sharma S, Kandpal V, Choudhury T, Gonzalez EDRS, Agarwal N. Assessment of the implications of energy-efficient technologies on the environmental sustainability of rail operation. AIMS Environmental Science. 2023;10(5):709-731.

29. Župan I, Šunde V, Ban Ž, Novoselnik B. An Energy Flow Control Algorithm of Regenerative Braking for Trams Based on Pontryagin's Minimum Principle. Energies. 2022;16(21):7346.

30. Zhu Z, Wang F, Yang R, Jiang Z, Xu R, Vansteenwegen P. Energy-Efficient Timetabling Approach Considering Varying Train Loads and Realistic Speed Profiles. Journal of Transportation Engineering, Part A: Systems. 2024;150(7).

31. Breiman L. Random forests. Machine Learning. 2001;45(1):5-32.

32. Friedman JH. Greedy function approximation: A gradient boosting machine. Annals of Statistics. 2001;29(5):1189-1232.

33. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;323(6088):533-536.

34. Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995;20(3):273-297.

FINANCING
No financing.

CONFLICT OF INTEREST
None.

AUTHORSHIP CONTRIBUTION
Conceptualization: Hatim RZIKI, Abdelaaziz HESSANE, Rita BELKADI.
Data Curation: Mohamed Khalifa BOUTAHIR, Hamid BOURRAY.
Formal Analysis: Hatim RZIKI, Rita BELKADI.
Research: Abdelaaziz HESSANE, Mohamed Khalifa BOUTAHIR, Hamid BOURRAY.
Methodology: Hatim RZIKI, Abdelaaziz HESSANE, Moulay Driss El Ouadghiri.
Project Management: Abdelaaziz HESSANE, Rita BELKADI.
Resources: Moulay Driss El Ouadghiri.
Software: Mohamed Khalifa BOUTAHIR, Hamid BOURRAY.
Supervision: Moulay Driss El Ouadghiri.
Validation: Hatim RZIKI, Rita BELKADI.
Display: Mohamed Khalifa BOUTAHIR, Hamid BOURRAY.
Drafting - Original Draft: Hatim RZIKI, Abdelaaziz HESSANE.
Writing - Proofreading and Editing: Rita BELKADI, Moulay Driss El Ouadghiri.

https://doi.org/10.56294/dm2025467