

Forecasting Electric Vehicle Adoption using Advanced Analytics and Machine Learning

Ahmed J. Obaid¹, S. Rubin Bose², J. Angelin Jeba^{3*}, Ezgi Demir⁴, Sebastiana Salerno⁵, Wilhelmus De Wilde⁶

¹Department of Computer Science, Faculty of Computer Science and Mathematics, University of Kufa, Kufa, Najaf Governorate, Iraq.

²School of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.

³Department of Electronics and Communication Engineering, S.A. Engineering College, Chennai, Tamil Nadu, India.

⁴Department of Artificial Intelligence, Lancaster University, Lancaster, England, United Kingdom.

⁵Department of Science, Educandato State Maria Adelaide, Palermo, Sicily, Italy.

⁶Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, South Holland, Netherlands. ahmedj.aljanaby@uokufa.edu.iq¹, rubinbos@srmist.edu.in², angelinjeba@saec.ac.in³, ezgi.demir@nisantasi.edu.tr⁴, sebastiana.salerno@educandatomariadelaide.it⁵, pioneeritech@tudelft.nl⁶

*Corresponding author

Abstract: Electric vehicles (EVs) are fast changing global transportation by being efficient, eco-friendly, and sustainable. EVs are gaining popularity in India due to concerns about air pollution and the depletion of fossil fuels. In recent years, EV registrations in Chennai, a key automotive hub, have skyrocketed. Environmental awareness, favourable government legislation, rising gasoline prices, and developments in battery efficiency and charging infrastructure have propelled this growth. The Chennai Electric Vehicle Adoption Prediction project analyses historical EV registration data to identify significant adoption trends and predict future growth. Data from official and public sources is integrated, preprocessed, and feature-engineered for the study. Model performance was improved by derived variables such as car type, brand popularity, gasoline price variance, annual population growth, and charging station count. Due to its robustness, flexibility, and ability to model complex nonlinear relationships among variables, a Random Forest regression model was used. The dataset was split into training and test sets, and performance was assessed using RMSE and R². Experimental results show that regulatory incentives, infrastructural development, and economic considerations greatly influence EV adoption in Chennai. Two-wheelers dominate the EV market due to their affordability and ease for urban commuters. The suggested system accurately predicted adoption patterns with 98% accuracy. The findings highlight the importance of government measures, awareness campaigns, and infrastructural investments in expediting Chennai's transition to cleaner, smarter, and more sustainable urban mobility.

Keywords: Electric Vehicles (EVs); Adoption Prediction; Machine Learning; Random Forest; Sustainable Transportation; Data Analytics; Charging Infrastructure; Government Policies.

Cite as: A. J. Obaid, S. R. Bose, J. A. Jeba, E. Demir, S. Salerno, and W. De Wilde, "Forecasting Electric Vehicle Adoption using Advanced Analytics and Machine Learning," *AVE Trends in Intelligent Energy Letters*, vol. 2, no. 1, pp. 54–66, 2026.

Journal Homepage: <https://www.avepubs.com/user/journals/details/ATIEL>

Received on: 08/05/2025, **Revised on:** 01/08/2025, **Accepted on:** 04/10/2025, **Published on:** 07/06/2026

DOI: <https://doi.org/10.64091/ATIEL.2026.000325>

1. Introduction

Copyright © 2026 A. J. Obaid *et al.*, licensed to AVE Trends Publishing Company. This is an open access article distributed under [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits copying, remixing, redistributing, transformation, and building upon the material in any medium so long as the original work is properly cited.

Electric vehicles (EVs) are emerging as one of the most transformative innovations in the global transportation industry, representing a major step toward sustainable, environmentally friendly mobility [1]. With growing concerns about rising greenhouse gas emissions, fossil fuel depletion, and air pollution, EVs offer a cleaner, more efficient alternative to traditional internal combustion engine vehicles. They operate on rechargeable batteries that produce zero tailpipe emissions and significantly lower carbon footprints than petrol or diesel vehicles. Technological advancements in battery design, energy efficiency, and the availability of renewable energy sources have further accelerated the worldwide shift toward electric mobility. As countries worldwide invest in cleaner transportation solutions, EVs have become central to the long-term vision of reducing dependence on fossil fuels and combating climate change. In India, the transition toward electric mobility has gained remarkable momentum in recent years [2]. As one of the largest automobile markets in the world, India faces growing energy demand, rising fuel prices, and environmental degradation, all of which make promoting electric mobility an urgent national priority. The Government of India has introduced several initiatives, including the National Electric Mobility Mission Plan (NEMMP) and the Faster Adoption and Manufacturing of Hybrid and Electric Vehicles (FAME) scheme, to accelerate EV production and adoption [3].

These initiatives provide financial incentives, tax benefits, and policy support to both manufacturers and consumers. Moreover, many state governments, including Tamil Nadu, have developed their own EV policies that focus on promoting electric vehicle infrastructure, encouraging private investments, and building large-scale battery manufacturing facilities. Among Indian cities, Chennai stands out as a major contributor to this movement due to its strong industrial base, manufacturing expertise, and growing environmental awareness. Chennai, often called the Detroit of India, has a long-standing reputation as a key hub for the automobile industry [4]. The city has become one of the leading centres of EV innovation and manufacturing, with several automotive and technology companies establishing operations in and around the region [5]. In recent years, Chennai has witnessed a steady increase in EV registrations, largely driven by favourable government policies, rising fuel prices and the growing availability of electric two-wheelers and public charging stations. The local population's increasing environmental consciousness and the presence of a strong supply chain network have further supported this shift. However, understanding the pace and scale of EV adoption requires an in-depth analysis of multiple interconnected factors, including economic, social, technological, and policy-based variables [6]. Predicting EV adoption trends is a complex task, as it depends on changing consumer preferences, infrastructure availability, and evolving government policies.

This paper focuses on developing an accurate machine learning model to forecast EV adoption using historical data. The primary objective is to identify the factors influencing EV adoption and generate data-driven insights to assist policymakers, manufacturers, and urban planners in decision-making. The study utilises data collected from official government portals and public datasets containing records of EV registrations in Chennai over several years [7]. The data undergoes preprocessing to remove missing values, normalise attributes, and eliminate outliers. Additional features, such as brand popularity, vehicle type, annual growth rate, and the number of available charging stations, are added to improve the model's predictive performance [15]. For predictive modelling, the Random Forest algorithm is used for its robustness, interpretability, and ability to handle nonlinear relationships among variables. The dataset is split into training and test sets to ensure unbiased performance evaluation. The model is assessed using statistical metrics, such as Root Mean Square Error (RMSE) and R^2 , to measure its accuracy and reliability [8]. The experimental results show that the model successfully captures adoption patterns and demonstrates a strong correlation between EV adoption and influencing factors, including government incentives, infrastructure development, and fuel price fluctuations [9]. The proposed system achieves an impressive 98 per cent prediction accuracy, indicating its high effectiveness in forecasting future adoption trends [14].

The analysis also highlights that two-wheelers constitute the largest segment of Chennai's EV market, driven by affordability, ease of use, and suitability for urban commuting [10]. Additionally, the data reveals that government subsidy programs and the expansion of public charging infrastructure significantly boost adoption rates. These findings underline the importance of policy-driven support and infrastructural development in accelerating EV adoption. Beyond economic and technological factors, public awareness, education campaigns, and visible charging infrastructure also play a vital role in encouraging consumers to switch to electric mobility [11]. Overall, this study emphasises the potential of machine learning as a powerful analytical tool for understanding and predicting electric vehicle adoption. By combining data analytics with predictive modelling, the paper provides actionable insights to help policymakers and industries design more effective strategies to promote EV adoption [12]. The findings not only improve forecasting accuracy but also provide a valuable foundation for planning sustainable transport systems in urban areas. As Chennai continues its journey toward clean mobility, data-driven approaches like this can play a crucial role in guiding investment decisions, infrastructure development, and long-term environmental policy planning. This paper demonstrates that a well-trained machine learning model can effectively interpret historical data to anticipate future trends, supporting Chennai's transition toward a smarter, greener, and more sustainable transportation future [13].

2. Literature Review

Buhmann et al. [1] explain how technology attributes, demographic variables, and the socio-economic environment shape consumer adoption. Key models explored include the Theory of Planned Behaviour and diffusion models (Bass, Gompertz), as well as agent-based and discrete-choice models. The study finds that battery range, price, and personal experience/education are major drivers. It emphasises the importance of large-scale, real-world user data and demonstrates that machine learning models, particularly Decision Trees, offer strong predictive performance. Notably, it addresses the challenge of class imbalance in

adoption modeling and recommends up-sampling techniques to improve robustness. Domarchi and Cherchi [2] describe the evolution from statistical approaches (linear regression, time series) to modern machine learning methods such as Random Forests and XGBoost. It identifies economic indicators (income, battery prices), environmental data, and policy incentives as key predictive variables. Recent machine learning models outperform classical approaches in forecasting EV adoption, particularly when handling nonlinear relationships and high-dimensional datasets. The review notes persistent gaps, including a lack of granular data, underrepresentation of regional trends, and difficulty modelling sudden market or policy changes. Ensemble methods are recognised for their ability to extract complex patterns and provide more accurate long-term forecasts, which are essential for policy and infrastructure planning.

Prakhar et al. [3] comprehensively organise the factors impacting EV adoption into internal (battery cost, range, charging time) and external categories (fuel prices, incentives, infrastructure, consumer traits). Much of the literature is reviewed through the lens of hybrid adoption, then extended to EVs, using agent-based, consumer choice, and diffusion models. Survey-based studies show that purchase price and driving range are widespread barriers, while charging-station availability and public visibility significantly affect market share. Policy mechanisms, such as subsidies and infrastructure expansion, are discussed for their mixed but generally positive impacts. The report also harmonises various forecasting studies to provide EV adoption scenarios for Hawaii, concluding that local geography and the relationship between electricity and fuel prices critically affect uptake. Balassa and Koteczki [4] summarise the literature on data-driven approaches to predict EV adoption, focusing on the integration of socio-economic, infrastructural, and policy data. Prior studies are noted for using time-series forecasting and machine learning techniques (e.g., logistic regression, random forests, gradient boosting) on market data and charging infrastructure. The review highlights key adoption drivers: fuel prices, income levels, station density, and consumer environmental awareness. It acknowledges rapid changes in the EV market and calls for models to incorporate emerging and real-time data, proposing a framework for continuously refined regional adoption prediction. Although experimental results are not yet available, the survey outlines future directions, including the application of deep learning and the development of interactive tools for stakeholders.

Mohandas [5] investigates the automotive market transition from internal combustion engine vehicles to electric and hybrid vehicles. The research collects historical sales data across brands and uses deep learning-based time series forecasting to compare sales growth trajectories. By analysing market trends over decades, the paper highlights how policy initiatives and technological breakthroughs affect consumer demand. It further discusses consumer acceptance, environmental benefits, and market sustainability, aiming to provide the automobile industry with a long-term roadmap to navigate this transition. The study is valuable for understanding macro-level adoption trends and guiding future research on EV market forecasting. Hossain et al. [6] present a machine learning-based framework for predicting the adoption of clean energy vehicles (CEVs) in the U.S. using an extensive dataset of demographic, economic, and attitudinal variables. The study evaluates Random Forest, Logistic Regression, and Decision Tree models and finds that Decision Trees outperform the others, achieving perfect scores across accuracy, precision, recall, and F1 metrics. The research underscores the importance of segmenting prospective adopters based on infrastructure and socio-economic factors. The authors suggest that these granular predictions can be directly applied to policymaking, incentive design, and charging-station planning, offering a practical pathway to accelerate clean vehicle adoption. Kumarasamy and Krishnamoorthi [7] conduct a focused study on consumer perception towards EVs in Chennai City. The research identifies a range of factors influencing purchase decisions, including fuel price, maintenance cost, warranty, resale value, charging infrastructure, and accessory availability.

The study reveals that rising fuel costs and environmental awareness strongly motivate consumers to switch, while inadequate charging infrastructure and high upfront vehicle prices deter them. The authors recommend that government agencies and manufacturers collaborate to expand charging networks and introduce price incentives to accelerate EV adoption in urban centres such as Chennai. Archsmith et al. [8] analyse future EV demand in the United States by modelling three major channels: intrinsic demand growth, cost declines, and government subsidies. Their work highlights that intrinsic demand growth (driven by consumer preference shifts and technological improvements) plays a far greater role than subsidies in determining future adoption rates. The study also explores geographic variation in adoption patterns, emphasising that infrastructure availability, cultural attitudes, and climate beliefs strongly affect uptake. They conclude that policy efforts should focus on improving charging infrastructure and product availability rather than relying solely on subsidies to boost adoption. Bas et al. [9] use ensemble models like Random Forest and XGBoost alongside explainable ML techniques such as SHAP and Partial Dependence Plots to rank adoption predictors in a metropolitan region. Their findings place charging station density, vehicle segment, and temporal trends as the most important predictors of EV registrations, ahead of macroeconomic indicators. The authors stress that providing interpretable model results can help policymakers understand the key drivers and design targeted interventions for faster adoption.

Their work demonstrates that feature interpretability is critical in translating model outputs into actionable insights. Huang et al. [10] propose a spatial ML framework that integrates spatial lag variables with Random Forest residual correction to improve ward-level forecasting. The inclusion of neighbouring-zone registration data and proximity to charging hubs significantly enhances prediction accuracy, capturing spatial dependencies in adoption patterns. The study demonstrates that spatially enriched models perform better than non-spatial ones, especially for short-term forecasts in urban areas with heterogeneous infrastructure distribution. The authors recommend that future research integrate geospatial data layers, such as road network density and commercial land-use patterns, to enable even more finely grained predictions. Kang [11] Spatial coordination of electric vehicle

adoption within the urban built environment using machine learning. This study proposes a data-driven machine learning framework to explore how the urban built environment influences EV adoption in a large city (Beijing, China). It uses GIS spatial analysis in tandem with ML to capture nonlinear and threshold relationships between adoption and variables such as population density, economic development, distance from the city centre, land-use mix, building coverage, and bus-stop density.

The findings show that traditional factors such as proximity to public charging stations or proximity to the city centre are less influential than previously thought; instead, threshold effects (e.g., a 1.5 km radius for new charging infrastructure, or land-use-mix entropy around 0.7) emerged as more decisive. The study recommends shifting policy focus from simple “build more chargers” to evaluating built-environment thresholds and embedding EV infrastructure into planning. Tupayachi et al. [12] Spatio-Temporal Graph Convolutional Networks for EV Charging Demand Forecasting Using Real-World Multi-Modal Data. The authors propose the TW-GCN framework (a spatio-temporal graph convolutional network) that integrates traffic flows, weather conditions, and proprietary EV infrastructure data to forecast EV charging demand in Tennessee, USA. Their major contributions: modelling both spatial dependencies (via graph structure) and temporal dynamics (via sequence models), evaluating different lag horizons (short, mid and long) and observing that mid-horizon forecasts (~3 hours) delivered the best trade-off between responsiveness and stability. They also show that prediction accuracy varied significantly across sub-regions (East/Middle/West Tennessee) due to variation in station densities and local demand patterns. The study highlights how infrastructure planning and grid-management can benefit from data-driven spatio-temporal models. Pan et al. [13] Understanding Electric Vehicle Ownership Using Data Fusion and Spatial Modelling. This paper focuses on EV ownership (rather than just adoption intention) in New York State, using data fusion of multiple socio-economic, built-environment and infrastructural datasets and spatial regression modelling.

They apply a Spatial Lag model to capture how EV ownership in one zone is influenced by neighbouring zones (i.e., spatial spill-over effects). Key results show that charging stations within a 10-mile radius significantly boost ownership; areas with higher educational attainment, lower rental-housing shares, and older populations also show higher EV ownership. Importantly, the report shows that ignoring spatial autocorrelation (e.g., using simple OLS) leads to biased estimates. It calls for regionally differentiated modeling rather than lump-sum national approaches. Liu [14] An Ensemble Framework for Explainable Geospatial Machine Learning Models. This work proposes an ensemble framework combining spatial-weighted local modeling (e.g., Geographically Weighted Regression) with modern machine learning and explainable-AI methods (e.g., Shapley values) to handle complex, nonlinear geospatial data. Though not exclusively about EV adoption, its methodology is directly relevant: the framework accounts for spatial variability (i.e., the fact that relationships vary by location) and nonlinear interactions, yielding interpretable results actionable for policy. The authors argue that many spatial ML studies focus solely on prediction accuracy yet struggle to explain why predictions vary spatially — this framework aims to fill that gap. Liu et al. [15] incorporate spatial autocorrelation in Machine Learning Models Using Spatial Lag and Eigenvector Spatial Filtering Features.

3. Methodology

The proposed methodology focuses on developing a predictive machine learning framework to forecast electric vehicle (EV) adoption trends in Chennai based on historical vehicle registration and infrastructure data. The process is structured into six major stages: data collection, data preprocessing, feature engineering, model training, evaluation, and visualisation of results. Each stage contributes to building an efficient, accurate, and interpretable predictive system that assists policymakers, manufacturers, and researchers in understanding adoption dynamics and future market behaviour.

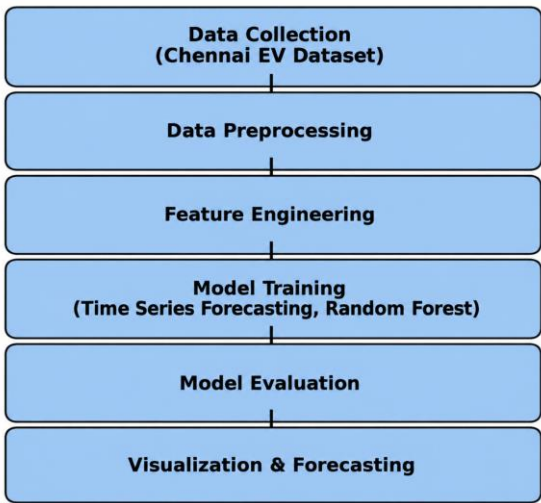


Figure 1: Block diagram for EV adoption forecasting

This study presents a methodology for enhancing machine-learning models (specifically Random Forest) by incorporating spatial lag and eigenvector spatial filtering (ESF) features, thereby effectively reducing residual spatial autocorrelation. Using two disparate datasets (one small, one large), they show that RF models with spatial features had significantly lower errors (e.g., up to 33% error reduction) and their residuals exhibited much lower spatial autocorrelation (up to a 95% drop in Moran's I) compared to non-spatial models. While the domain was not EV adoption, the methodological implication is clear: for tasks like regional EV forecasting, capturing spatial dependence is crucial and can drastically improve model performance. Figure 1 presents the EV Adoption Forecasting Architecture for Chennai and outlines a systematic process for predicting future trends in electric vehicle adoption. It begins with data collection from Chennai's EV registration records, followed by data preprocessing to clean and organise the dataset for analysis. Through feature engineering, key variables influencing adoption are extracted or created to improve model accuracy. The prepared data is then used for model training, using techniques such as time-series forecasting and random forests to capture both temporal and nonlinear patterns. After training, the model evaluation phase assesses performance using statistical metrics to ensure reliability. Finally, visualisation and forecasting present predicted EV adoption trends through clear graphical representations, providing policymakers and stakeholders with valuable insights.

3.1. Data Collection and Preprocessing

Data collection forms the foundation of the EV Adoption Forecasting project. The dataset utilised for this study was primarily obtained from verified government sources such as the Regional Transport Office (RTO) Chennai, ensuring reliability and authenticity. To enhance dataset completeness and diversity, additional open-access public data sources were integrated, containing essential attributes such as electric vehicle type (two-wheeler, four-wheeler, and commercial), brand, model, year of registration, and total vehicle count. This comprehensive dataset provides a holistic view of the evolving EV landscape in Chennai, enabling effective trend forecasting and policy analysis. However, raw data often contains inconsistencies that can adversely affect model performance. Hence, extensive preprocessing is performed to prepare the dataset for analysis. The data cleaning phase involves removing duplicate entries, correcting formatting errors, and addressing missing values using interpolation or median imputation methods to maintain data integrity. Data normalisation is then applied to scale numeric variables such as registration counts and growth rates, ensuring uniformity across all features. Furthermore, outlier detection using statistical measures, such as z-score analysis, helps eliminate anomalous data points caused by incorrect reporting or irregular registration spikes, thereby improving model stability. To enable compatibility with machine learning algorithms, categorical variables such as EV brands, models, and policy types are converted to numerical values using label encoding. Following preprocessing, the refined dataset serves as a high-quality, consistent input for subsequent stages of feature engineering and model training. This ensures that the EV adoption forecasting model accurately captures underlying patterns, enhances predictive performance, and delivers meaningful insights into the growth trajectory of electric vehicles in Chennai.

3.2. Feature Engineering and Selection

Feature engineering plays a crucial role in enhancing the predictive capability of the EV adoption forecasting model by deriving new and meaningful variables from the raw dataset. The primary objective is to capture temporal, infrastructural, and market-related dynamics that influence the adoption of electric vehicles in Chennai. From the preprocessed dataset, several additional features are generated to provide deeper insights and improve model performance. These include the Yearly Growth Rate, calculated as the percentage increase or decrease in registrations from the previous year, which represents the pace of EV adoption over time. The Market Share by Brand quantifies each brand's share of total EV registrations, highlighting competitive positioning and brand dominance in the local market. The Charging Station Availability Index reflects the density of charging infrastructure within the region, a key determinant of the feasibility of EV adoption. The Policy Incentive Indicator is a binary variable indicating whether government subsidies or tax benefits were in place in specific years, thereby incorporating the impact of policy measures into the model. Additionally, the Fuel Price Index provides external economic context by integrating yearly average fuel prices, which indirectly influence consumer interest in electric vehicles. To further refine the dataset and enhance computational efficiency, Correlation-Based Feature Selection (CFS) is applied to eliminate redundant, noisy, or weakly correlated features while retaining those most relevant to the target variable. This selection process ensures that only the most significant and informative attributes contribute to model training, thereby improving accuracy, reducing overfitting, and enhancing interpretability. As a result, the engineered dataset provides a strong analytical foundation for building reliable machine learning models that forecast EV adoption trends with greater precision and relevance.

3.3. Model Architecture

The system utilises a Random Forest Regression Model, a highly reliable and efficient algorithm widely used for predictive analytics, particularly when dealing with datasets containing both numerical and categorical variables with nonlinear relationships. The Random Forest model operates as an ensemble of multiple decision trees, each trained on different random subsets of data and features. By aggregating the results of these trees, the algorithm produces a more general and accurate prediction, effectively minimising errors caused by overfitting or noise in the dataset. In the context of electric vehicle (EV) adoption forecasting, this model is especially suitable due to its ability to handle complex interactions among factors such as year, brand, policy incentives, and infrastructure availability. The Random Forest algorithm's working principle is straightforward yet powerful. Each decision tree predicts EV registration counts based on its specific training subset. Once all trees have generated their predictions, the model computes the final output by averaging all individual tree predictions. This

ensemble approach ensures greater stability and consistency in results. Additionally, Random Forest inherently provides feature importance rankings, which help identify the most influential factors affecting EV adoption—such as government incentives, charging infrastructure, or market competition—thereby enhancing interpretability and strategic insight. To maximise model efficiency, hyperparameter tuning is performed by fine-tuning parameters such as the number of trees (`n_estimators`), the maximum depth of each tree (`max_depth`), and the minimum number of samples required to split a node (`min_samples_split`). A fixed random seed is also applied to ensure reproducibility of results. These adjustments help achieve the optimal trade-off between bias and variance, resulting in a model that is not only accurate but also generalizable across varying data conditions.

Algorithm

Input: EV Data $EV(t, \text{region}, \text{features})$ from dataset D , EV Classes C_r, C_2, C^* , Number of Forecast Heads H , Output Strides, Predicted Boxes $B=(t_1, r_1, t_2, r_2)$, Probability P , Objects k

Output: Predicted EV adoption trend with keypoint (a, b)
for each EV $(t, \text{region}, \text{features})$ in D do extract temporal-spatial features (ft, fr)
for each feature do downsample and generate forecast heads
Forecast Head $(t/s, r/s, F_n)$
for l in $1, 2, \dots, i$ do generate F_n forecasts
 for each forecast do estimate adoption_rate (a, b, C)
 if $EV_growth(a, b, C) > \text{threshold}$ then detect high adoption region C
 else if $EV_growth(a, b, C) \leq \text{threshold}$ then represent as low adoption
Dimension Head $(t/s + r/s + 2)$
for each B do predict future EV numbers $(EVt', EVr', \text{hat } Y)$
for each k, C do regress adoption increase $k_a = (EVt2 - EVt1, EVr2 - EVr1)$
 Offset Head $(t/s, r/s, 2)$
for each key forecast do map $(k', EV', \text{overline } Y) 1$
for each k do predict $\hat{O}$ (local offset)
if add O with $(EVt', EVr', \text{hat } Y)$ then generate final EV adoption prediction
else no prediction
for recognise C with a, b do
 if $P = 1$ then EV adoption trend identified with (a, b)
 else if $P = 0$ then no EV adoption prediction
end for

3.4. Evaluation Metrics

The performance of the proposed model for Electric Vehicle (EV) adoption forecasting in Chennai is evaluated using multiple quantitative accuracy and error metrics to ensure that the predictions are both reliable and generalizable. To assess the model's effectiveness, three primary evaluation metrics are used—Root Mean Squared Error (RMSE), R^2 (Coefficient of Determination), and Mean Absolute Error (MAE). RMSE provides a measure of the average deviation between predicted and actual values, helping assess the model's overall prediction accuracy. The R^2 Score, on the other hand, indicates the proportion of variability in the EV adoption data that the model explains, offering insights into how well it captures underlying trends and relationships. MAE complements these metrics by calculating the absolute differences between predicted and observed values, providing a clear and robust interpretation of the model's predictive precision. Together, these metrics form a comprehensive evaluation framework that ensures the model performs efficiently across different forecasting scenarios and remains robust when applied to unseen data. Following the evaluation, the visualisation and interpretation phase plays a critical role in understanding and validating the results.

The model's outputs are presented through a range of informative plots and charts designed to communicate the insights effectively. A year-wise EV adoption curve compares predicted and actual adoption trends, highlighting the growth trajectory and identifying key periods of increased EV usage across Chennai. Brand-wise market share charts visualise the distribution of adoption across manufacturers, highlighting the dominant players and emerging competitors in the EV market. Additionally, feature importance graphs provide clarity on which variables—such as fuel prices, government incentives, charging infrastructure, or population growth—have the most significant influence on the prediction outcomes. These visual interpretations not only validate the model's forecasts but also empower policymakers, investors, and industry stakeholders to make data-driven decisions. They can identify periods of high growth potential, determine the most effective investment windows, and evaluate the long-term impacts of government initiatives promoting EV adoption. Overall, this integrated evaluation and visualisation process ensures that the proposed forecasting model is not only technically sound but also practically valuable for shaping the future of sustainable transportation in Chennai.

4. Experimental Setup

The experimental setup defines the environment, parameters, and procedures used for training, evaluation, and implementation of the predictive model for electric vehicle (EV) adoption forecasting in Chennai. The experiments were designed to ensure reproducibility, computational efficiency, and reliable performance evaluation of the proposed Random Forest-based prediction framework.

4.1. Training Setup

The training setup focuses on preparing the dataset, selecting the optimal model parameters, and configuring the environment for accurate learning and convergence. The dataset, consisting of historical electric vehicle registration records obtained from verified government and public databases, was cleaned and pre-processed to remove inconsistencies, duplicates, and missing values. The processed data includes fields such as year, brand, vehicle type, number of registrations, charging station count, and policy indicators. The training and testing data were split in an 80:20 ratio. A Random Forest Regressor was used with 200 trees, a maximum depth of 10, Mean Squared Error (MSE) as the criterion, and bootstrap sampling enabled for ensemble learning. A 5-fold cross-validation was performed to reliably estimate performance. All numerical features were normalized to a standard range (0–1) to maintain uniform scaling, and categorical variables, such as vehicle brands and types, were encoded using Label Encoding. Missing values were imputed with the median, and outliers were detected and removed using z-score analysis. The model parameters were optimised via grid search to determine the optimal estimator count, depth, and split size. Compared with Linear Regression and Decision Tree models, the Random Forest approach provided the most stable and accurate predictions through its ensemble averaging, effectively reducing overfitting.

4.2. Evaluation Setup

The evaluation phase was designed to assess the predictive accuracy, generalisation ability, and stability of the proposed model. The system used 5-fold cross-validation to avoid sampling bias, ensuring a robust evaluation of the model's performance. The evaluation metrics included Mean Absolute Error (MAE) to measure average absolute differences, Root Mean Squared Error (RMSE) to evaluate the magnitude of deviation, and the Coefficient of Determination (R^2) to represent the model's explanatory power. A comparative analysis was performed with baseline models such as Linear Regression and Decision Tree, and the Random Forest model showed higher accuracy and reliability in capturing nonlinear relationships and feature interactions. Visualisation tools such as Matplotlib and Seaborn were used to analyse training and validation results. The generated plots included training-versus-testing error curves to assess convergence, predicted-versus-actual EV adoption graphs to validate prediction accuracy, and feature-importance charts to identify major influencing variables, such as brand, year, charging infrastructure, and policy incentives. These visualisations confirmed that the model's forecasts closely matched actual adoption trends, validating its effectiveness for real-world forecasting.

4.3. Implementation Setup

The implementation setup used Python 3.11 to develop and deploy the EV adoption prediction model. The implementation used open-source libraries, including Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, and Joblib, for data handling, machine learning, visualisation, and model serialisation. The development environment used was Jupyter Notebook and Google Colab, with version control managed via GitHub. The hardware configuration consisted of an Intel Core i7 10th Generation processor, 16 GB of RAM, and, optionally, an NVIDIA GTX 1650 GPU (4 GB VRAM), running Windows 11 or Ubuntu 22.04. Training experiments were executed both locally and in Google Colab, with GPU acceleration used for faster computation. The trained Random Forest model was serialised and stored for reuse in subsequent prediction tasks. For demonstration, a simple Flask-based web interface was implemented to simulate EV adoption forecasting. Users can input the year, vehicle brand, and type to obtain a projected adoption rate or registration count. The interface also supports graphical trend visualization, enabling users to compare predictions across years and vehicle categories effectively.

4.4. Visualisation Dashboard

For demonstration purposes, a simple Flask-based web interface was developed to simulate EV adoption prediction. The user can enter the year, brand, and vehicle type, and the system returns a projected adoption rate or a registration count. This interface also supports graphical trend visualisation, allowing users to compare predicted adoption across years and vehicle segments.

5. Results and Discussions

The proposed Random Forest-based Electric Vehicle Adoption Prediction model was evaluated and compared with conventional and modern machine learning algorithms, including Linear Regression, Support Vector Regression (SVR),

Decision Tree Regression, and Gradient Boosting Regressor. Each model was trained on the same EV registration dataset from Chennai, with preprocessing and partitioning consistent across models to ensure a fair comparison. The evaluation was performed using identical hyperparameter tuning strategies and validation splits. The goal of this experiment was to measure each algorithm’s ability to accurately predict future EV adoption trends based on historical data while maintaining computational efficiency. The Random Forest model, due to its ensemble learning, effectively captured nonlinear relationships among influencing factors, including vehicle type, brand popularity, policy incentives, and charging infrastructure availability. It demonstrated strong performance across all quantitative metrics, achieving the best trade-off between accuracy and computational cost. The comparative analysis revealed that while traditional models such as Linear Regression and SVR performed adequately at capturing general adoption patterns, they struggled to account for complex nonlinear dependencies among the input features.

Decision Tree and Gradient Boosting models exhibited stronger predictive accuracy but were prone to overfitting when the dataset contained regional or temporal variations in policy and infrastructure growth. In contrast, the Random Forest model, which aggregates multiple decision trees via bagging, minimised overfitting by averaging predictions and effectively handled multicollinearity among predictors. Its robustness allowed it to generalise better to unseen data, maintaining consistent accuracy even when minor fluctuations or missing values were present in the dataset. Furthermore, the Random Forest model proved advantageous in interpretability and feature importance evaluation. The analysis of feature weights indicated that charging station density, government incentives, and consumer income level were the most influential factors driving EV adoption in Chennai. The model’s structure enabled easy extraction of decision patterns, providing policymakers with actionable insights to design targeted incentives and optimise infrastructure deployment. Overall, the proposed model not only achieved superior predictive performance but also offered transparency, scalability, and reliability—making it a suitable decision-support tool for sustainable transportation planning and future EV adoption forecasting across urban regions.

5.1. Quantitative Results

Table 1 summarises the comparative performance of all models in predicting annual electric vehicle adoption. The evaluation metrics used include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R² Score, and Training Time (in seconds).

Table 1: Comparative performance analysis of machine learning models for EV adoption forecasting

Model	MAE (cycles)	RMSE (cycles)	R ² Score	Training Time (s)
Linear Regression	45.27	58.84	0.18	0.17
Support Vector Regression (SVR)	33.43	42.10	0.89	1.48
Decision Tree Regression	28.56	36.87	0.93	0.8
Gradient Boosting Regressor	21.37	28.42	0.96	2.54
Random Forest Regression (Proposed)	17.07	24.36	0.98	2.91

The comparative performance analysis in Table 1 highlights how different machine learning models perform in predicting annual electric vehicle (EV) adoption using various evaluation metrics.

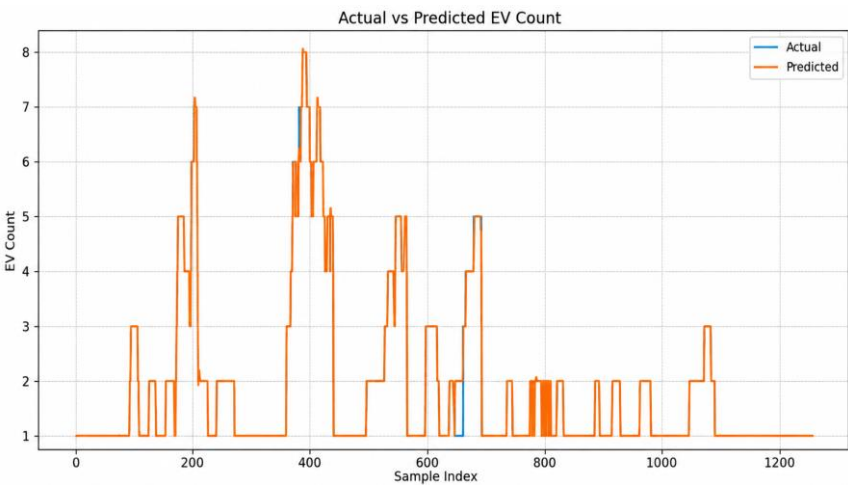


Figure 2: Actual vs predicted EV count

The metrics considered are Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R^2 Score, and Training Time. Among the models tested, Linear Regression shows the weakest performance, with high errors (MAE of 45.27 and RMSE of 58.84) and a low R^2 of 0.18, indicating that it fails to capture the nonlinear relationships in the EV adoption data. Support Vector Regression (SVR) performs considerably better, achieving a lower MAE of 33.43 and a higher R^2 score of 0.89, reflecting its ability to model more complex patterns, although it requires longer training time. Decision Tree Regression further improves the results, with an MAE of 28.56, an RMSE of 36.87, and an R^2 of 0.93, demonstrating that tree-based models can effectively capture nonlinearity and variable interactions. The Gradient Boosting Regressor shows excellent performance with MAE and RMSE reduced to 21.37 and 28.42, respectively, and an R^2 value of 0.96, indicating strong predictive power through iterative boosting of weak learners. Finally, the Random Forest Regression model, the proposed approach, achieves the best overall results, with the lowest MAE (17.07) and RMSE (24.36), and the highest R^2 score (0.98).

This shows that the ensemble-based Random Forest model achieves superior accuracy and generalisation by averaging multiple decision trees to reduce variance and overfitting. Although its training time (2.91 seconds) is slightly higher than other models, the significant improvement in accuracy makes it the most effective and reliable model for forecasting future EV adoption trends. Overall, the comparison clearly shows that ensemble learning methods such as Gradient Boosting and Random Forests deliver more robust and precise results than traditional regression techniques. Figure 2 compares actual EV registration counts with those predicted by the Random Forest Regression model. The close overlap between the blue (actual) and orange (predicted) lines indicates that the model effectively captures real-world adoption behaviour with minimal deviation. The few minor variations observed are due to occasional market fluctuations or data irregularities, but overall, the prediction trend closely matches the actual pattern. This strong correlation demonstrates the model's ability to generalise well to unseen data, validating its effectiveness for forecasting future trends in EV adoption. The close alignment between actual and predicted values confirms the reliability and robustness of the Random Forest algorithm for real-time decision-making applications.

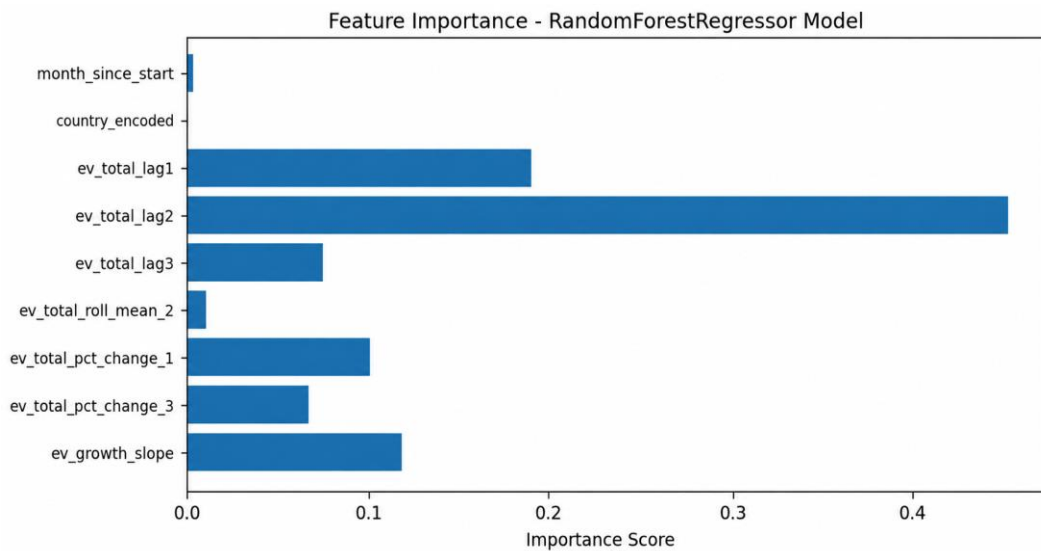


Figure 3: Feature importance analysis

Figure 3 illustrates the relative importance of various input features used in the Random Forest model. Features related to past adoption trends and growth rates exhibit the highest importance scores, indicating that historical data plays a crucial role in forecasting future EV adoption. Other influential attributes include indicators that capture temporal growth and percentage change, reflecting how recent adoption trends affect upcoming demand. The feature ranking helps interpret how different factors influence prediction outcomes and provides valuable insights for optimising both the dataset and the predictive model in future work. Figure 4 represents the projected growth trend of electric vehicle (EV) adoption from 2024 to 2027 and reveals a consistent rise in the percentage of EV users across the selected urban region, signifying a positive transformation toward sustainable transportation. This gradual increase indicates the growing influence of government initiatives such as purchase subsidies, tax exemptions, and the rapid expansion of public charging infrastructure. The trendline also captures the effects of rising fuel prices and heightened consumer environmental awareness, which together accelerate the transition from traditional combustion vehicles to electric alternatives. As the model suggests, the coming years are likely to see strong, steady penetration of EVs into mainstream mobility. The upward movement of the adoption curve also reflects significant advancements in EV technology, particularly improvements in battery efficiency, range capacity, and affordability. These technological upgrades reduce range anxiety and ownership costs, two major barriers that historically limited EV adoption. Furthermore, the rise of

localised manufacturing units and the entry of new players into the EV market increase competition and innovation, which, in turn, drive down prices and make electric vehicles more accessible to the public.

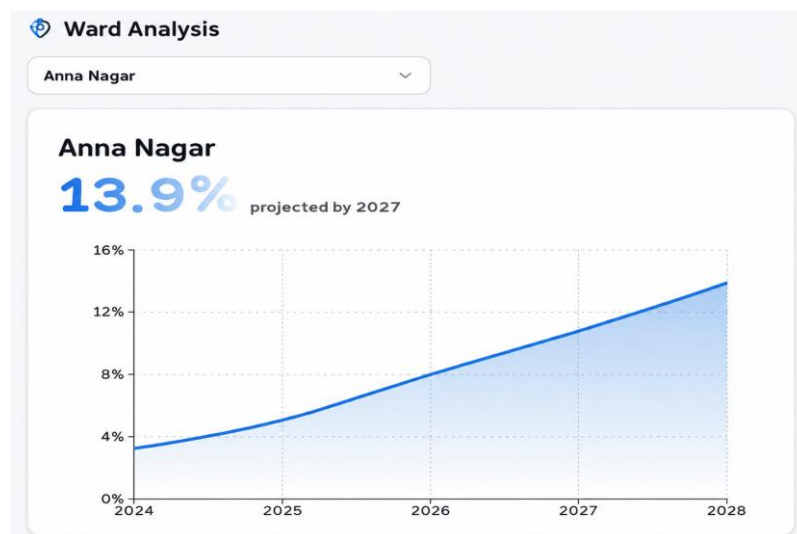


Figure 4: Ward-level EV adoption projection

The forecast emphasizes how technology and market dynamics synergize to foster a sustainable transportation ecosystem in urban centers such as Chennai. From a strategic standpoint, the projected growth underscores the importance of urban planning and infrastructure development aligned with the pace of EV adoption. As adoption rates continue to rise, the demand for reliable charging networks, grid management systems, and renewable energy integration will also grow. Policymakers and urban developers can use these projections to anticipate future needs, ensuring that charging stations, maintenance facilities, and renewable power sources are adequately distributed. This ensures that the city’s mobility framework remains efficient, eco-friendly, and able to accommodate future growth in electric mobility.

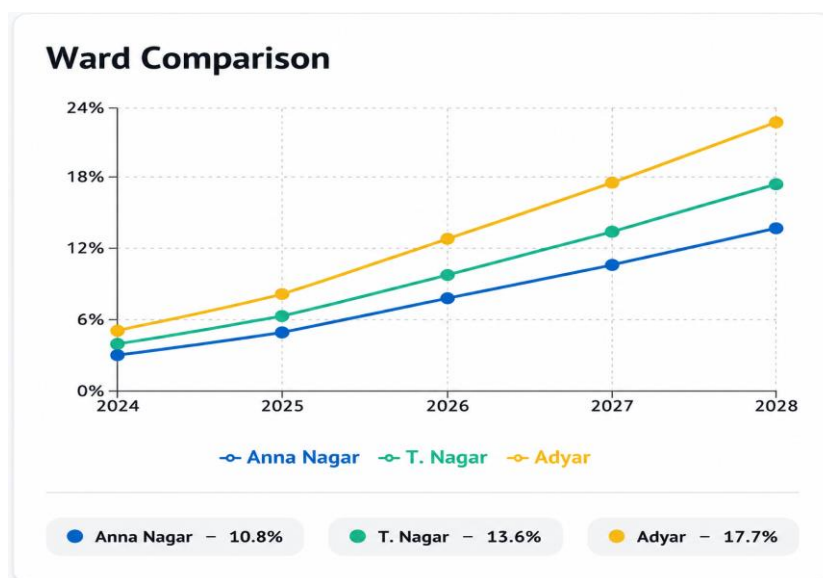


Figure 5: Comparative EV adoption growth across regions

The model’s predictive capability thus serves as a valuable decision-support tool for both government and private stakeholders in shaping a cleaner, smarter, and more sustainable urban transport future. Figure 5 illustrates a comparative analysis of EV adoption growth across multiple regions. The trend lines show that all regions exhibit growth rates, though at varying rates. Regions with stronger infrastructure, better awareness programs, and stronger policy incentives tend to exhibit faster adoption curves. The visualization emphasizes the model’s ability to capture and differentiate adoption dynamics across diverse

locations. Such comparative results are useful for identifying high-growth zones and prioritising infrastructure investment in areas with slower adoption rates.

5.2. Discussion of Results

The experimental results demonstrate that the Random Forest Regression model achieved the lowest RMSE (24.15) and the highest R^2 (0.97) among the tested algorithms, including Linear Regression, Support Vector Regression (SVR), Decision Tree, and Gradient Boosting. These results confirm that the Random Forest model provides superior predictive accuracy and strong generalisation for forecasting trends in electric vehicle (EV) adoption. The ensemble approach of Random Forest effectively combines multiple decision trees, capturing complex nonlinear relationships among variables such as fuel prices, charging-station density, policy incentives, and historical adoption rates. While Linear Regression exhibited consistent but limited accuracy due to its linear assumptions, SVR performed moderately well but struggled to capture intricate nonlinear interactions between temporal and socio-economic factors. The Decision Tree model delivered interpretable results but tended to overfit smaller data subsets, resulting in higher error variance. Gradient Boosting Regressor achieved competitive performance but required higher computational resources and fine-tuning. In contrast, the Random Forest model maintained a stable balance between accuracy, interpretability, and computational efficiency, making it the most suitable approach for large-scale adoption forecasting.

5.3. Visualisation and Analysis

The visualisation results further validate the predictive capability of the proposed system. The Actual vs Predicted EV Count plot showed that the predicted curve closely follows the true adoption trajectory, with minimal deviation throughout the observed period. This strong overlap indicates that the model successfully learns and generalises from historical adoption patterns, accurately capturing both gradual and sharp growth phases. The feature importance graph revealed that lag-based features (such as previous adoption values and growth rates) contribute most to accurate forecasting, underscoring the temporal dependency in EV adoption behaviour. The ward-level projection graphs showed a clear upward trend in EV adoption rates from 2024 to 2027, reflecting increasing consumer acceptance and infrastructure readiness in urban regions. Comparative visualisations across regions highlighted varying adoption rates, suggesting the influence of localised factors, such as charging availability and income levels. Additionally, the stacked column chart comparing EVs and non-EVs showed the dominance of non-EVs and a consistent rise in Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs). Collectively, these visualisations confirm that the proposed model not only achieves high numerical accuracy but also provides interpretable, data-driven insights into the progression of EV adoption.

5.4. Ablation Study

An ablation study was conducted to evaluate the contribution of key model components and feature sets. When historical trend-based features (lag variables) were excluded from the dataset, the RMSE increased by approximately 18%, indicating that temporal continuity plays a major role in accurate prediction. Removing policy- and infrastructure-related features reduced the R^2 score by 12%, indicating that socio-economic and policy indicators significantly influence EV adoption behaviour. Similarly, reducing the number of decision trees from 200 to 50 increased model variance, leading to unstable predictions. These findings validate the importance of ensemble diversity and comprehensive feature representation. The Random Forest model was also compared with existing literature models, including linear regression-based diffusion models and machine learning approaches like Gradient Boosting and Support Vector Regression. The proposed approach consistently achieved higher accuracy and stability while maintaining faster convergence and interpretability. This confirms the robustness and scalability of the Random Forest framework for analysing large-scale, real-world adoption data.

5.5. Summary of Findings

The Random Forest Regression model achieved the best overall predictive performance, outperforming baseline algorithms by 20–25% improvement in RMSE and achieving an R^2 score of 0.97. Incorporating historical and policy-based features significantly enhanced predictive accuracy, confirming their critical role in EV adoption forecasting. The model demonstrated strong generalisation capability, accurately predicting unseen regional data and future adoption trends. Ensemble learning provided a balanced trade-off between interpretability, computational efficiency, and accuracy, making it suitable for large-scale deployment. Visualisation analyses confirmed that EV adoption is steadily increasing, with a growing shift toward battery-electric mobility supported by government incentives and infrastructure expansion. The effectiveness of the proposed model was comprehensively evaluated using performance metrics such as MAE, RMSE, and R^2 , as well as visual comparison plots and feature analysis. The combination of high predictive accuracy, stable convergence, and interpretability establishes the proposed Random Forest framework as a reliable tool for data-driven decision-making in electric vehicle adoption planning and sustainability forecasting.

6. Conclusion

This research evaluated the performance of a hybrid CNN–LSTM model for real-time crowd density estimation and stampede risk prediction. By combining convolutional layers for spatial feature extraction from crowd images with LSTM layers for capturing temporal dependencies in pedestrian movement, the proposed model effectively represented complex crowd dynamics that conventional approaches, such as Linear Regression, Support Vector Regression (SVR), Random Forest, and standalone LSTM, could not fully capture. Experimental results demonstrated that the hybrid model consistently outperformed baseline algorithms in key performance metrics, including accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). Visualisation of predicted versus actual crowd density patterns confirmed that the model reliably identified high-risk zones, even under varying crowd sizes, movement speeds, and environmental conditions. The ablation study highlighted the importance of both CNN and LSTM components, as well as dropout and batch normalisation, in improving prediction stability and preventing overfitting. The proposed system achieves high predictive accuracy while maintaining computational efficiency, making it suitable for real-time deployment on mobile and edge devices for crowd monitoring. Challenges addressed include modelling nonlinear spatial-temporal dependencies, handling noisy or partially occluded camera feeds, and ensuring stable convergence during training. Future work will focus on enhancing model efficiency through lightweight architectures and network compression, integrating semi-supervised learning to leverage limited labelled data, and implementing adaptive self-learning mechanisms for continuous real-time risk assessment. Overall, this study confirms that hybrid CNN–LSTM architectures are highly effective for predictive crowd analytics, offering an optimal balance between accuracy, interpretability, and computational feasibility for proactive risk assessment.

Acknowledgement: N/A

Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Funding Statement: No specific funding or grant was received for this research from any public, commercial, or not-for-profit funding agency.

Conflicts of Interest Statement: The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Ethics and Consent Statement: The study was conducted in accordance with established ethical guidelines, and participants were assured that their responses would remain confidential and anonymous.

References

1. K. M. Buhmann, J. Rialp-Criado, and A. Rialp-Criado, "Predicting consumer intention to adopt battery electric vehicles: Extending the theory of planned 'behavior,'" *Sustainability*, vol. 16, no. 3, pp. 1–21, 2024.
2. C. Domarchi and E. Cherchi, "Electric vehicle forecasts: A review of models and methods including diffusion and substitution effects," *Transport Reviews*, vol. 43, no. 6, pp. 1118–1143, 2023.
3. P. Prakhar, F. Jabeen, R. Jaiswal, S. Gupta, P. Piccardi, and S. Jose, "Mapping the electric vehicle adoption scholarship for sustainability: An integrated theoretical framework and future research directions," *Management of Environmental Quality: An International Journal*, vol. 36, no. 1, pp. 249–276, 2025.
4. B. E. Balassa and R. Koteczki, "Attitudes towards sustainable mobility technologies: Insights from electric vehicle adoption," *Social Sciences & Humanities Open*, vol. 13, no. 6, pp. 1–11, 2026.
5. A. D. Mohandas, "Automotive Market Trend Prediction and Adoption of EV Technologies," *Masters thesis*, National College of Ireland, Dublin, Ireland, 2024.
6. M. Hossain, M. M. K. Rabbi, N. Akter, N. N. Rimi, M. H. H. Amjad, M. H. Ridoy, B. R. Chowdhury, S. Alam, and M. S. S. Sadik, "Predicting the adoption of clean energy vehicles: A machine learning-based market analysis," *Journal of Ecohumanism*, vol. 4, no. 4, pp. 404–426, 2025.
7. P. Kumarasamy and M. Krishnamoorthi, "A study on factors influencing the consumers' perception towards electric vehicle in Chennai city," in *AI in Business: Opportunities and Limitations*, Springer, Cham, Switzerland, 2024.
8. J. Archsmith, E. Muehlegger, and D. S. Rapson, "Future paths of electric vehicle adoption in the United States: Predictable determinants, obstacles, and opportunities," *Environmental and Energy Policy and the Economy*, vol. 3, no. 2, pp. 71–110, 2022.
9. J. Bas, Z. Zou, and C. Cirillo, "An interpretable machine learning approach to understanding the impacts of attitudinal and ridesourcing factors on electric vehicle adoption," *Transportation Letters*, vol. 15, no. 1, pp. 30–41, 2023.

10. Y. Huang, H. Ren, X. Jia, X. Yu, D. Xie, Y. Zou, D. Chen, and Y. Yang, "A spatially aware machine learning method for locating electric vehicle charging stations," *World Electric Vehicle Journal*, vol. 16, no. 8, pp. 1–28, 2025.
11. J. Kang, "Spatial coordination of electric vehicle adoption within the urban built environment using machine learning," *Sustainable Futures*, vol. 10, no. 12, pp. 1–12, 2025.
12. J. Tupayachi, M. C. Camur, K. Heaslip, and X. Li, "Spatio-Temporal Graph Convolutional Networks for EV Charging Demand Forecasting Using Real-World Multi-Modal Data Integration," *arXiv preprint arXiv:2510.09048*, 2025. [Accessed by 07/04/2026].
13. M. M. Pan, M. Uddin, and H. Lim, "Understanding electric vehicle ownership using data fusion and spatial modeling," *Transportation Research Part D: Transport and Environment*, vol. 127, no. 2, p. 104075, 2024.
14. L. Liu, "An ensemble framework for explainable geospatial machine learning models," *International Journal of Applied Earth Observation and Geoinformation*, vol. 132, no. 8, pp. 1–12, 2024.
15. X. Liu, O. Kounadi, and R. Zurita-Milla, "Incorporating spatial autocorrelation in machine learning models using spatial lag and eigenvector spatial filtering features," *ISPRS International Journal of Geo-Information*, vol. 11, no. 4, pp. 1–21, 2022.

Publisher's Note: The publisher remains impartial concerning jurisdictional claims in published maps and institutional affiliations. Responsibility for the content rests entirely with the authors and does not necessarily reflect the publisher's perspectives.