Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.12 n.57, p. 305-314. January/February, 2026.
DOI: https://doi.org/10.5935/jetia. v12i57.2905

RESEARCH ARTICLE OPEN ACCESS

SYNTHETIC DATA-DRIVEN ANALYSIS OF ELECTRIC VEHICLE MANUFACTURING IN INDIA USING MACHINE LEARNING

Pasala Gopi*[1], Tharun Korivi[2], Vinod Kumar Reddy P[3], Sri Ayodhya Raju K[4] and Sathish Kumar A[5]

[1]Departmentof Electrical and Electronics Engineering, Annamacharya University, 516 126, Rajampet, India.
[2, 3, 4, & 5]Departmentof Electrical and Electronics Engineering, Annamacharya Institute of Technology and Sciences, 516 126, Rajampet, India.

[1]https://orcid.org/ 0000-0003-0428-8673, [2]https://orcid.org/0009-0000-6181-8576, [3]https://orcid.org/0009-0004-7405-0411, [4]https://orcid.org/0009-0008-2736-5815, [5]https://orcid.org/0009-0001-8064-7626

Email: *pasala.epe07@gmail.com, tharunkorivi7@gmail.com, vinodkumarreddyarmy@gmail.com, sriayodhyarajukadimella@gmail.com, sathishkumar678901@gmail.com

ARTICLE INFO

Article History
Received: November 7, 2025
Revised: December 10, 2025
Accepted: January 1, 2026
Published: January 31, 2026

Keywords:
Machine Learning,
Carbon Emissions,
Production Forecasting,
Multi-Output Random Forest,
Scenario Analysis.

ABSTRACT

The global trend to Electric Vehicles is crucial in the minimization of carbon emissions in the transportation field. Nonetheless, countries like India have serious analytical difficulties in the change due to the paucity and irregularity of long-term manufacturing data. The current work attempts to fill this gap by proposing a Machine Learning model based on the synthetic data to forecast the future trends in the production of EV in the country. An artificial dataset over the years 2010-2025 was thoroughly constructed based on fifteen selected variables that reflect governmental policy, supply-chain capacity, and significant economic indicators. A high-dimensional, detail parameter space feature base was trained on these parameters to create a system based on a Multi-Output Random Forest Regressor that can simultaneously predict two parameters EV production performance and corresponding transport-sector CO_2 emissions. Projections of the years 2026-2035 have been created with the use of this setup under the three policy and market conditions, namely, Conservative, Base, and Accelerated. The analysis of these scenarios shows that the Accelerated case provides the best results, predicting the sharp increase of the EV production and a total decrease of CO_2 of approximately 58 Mt over the ten years. On the whole, the created model plays the role of an effective decision-support tool of both policymakers and industrial planners.

I. INTRODUCTION

The increasing risk of climate change has resulted in a faster transition to less polluting energy system, especially in the transport industry, which is among the primary sources of greenhouse gas (GHG) emissions [1]. The move to an electric-powered mobility rather than a traditional fuel-powered car is much more than a technological change it means a complete re-evaluation on how the world can be sustainable in transport and energy efficiency [2], [3]. This development is essential to fulfilling global climate promises and to move toward a low-carbon economy in the world. India is the key location in this shift as it is considered one of the most rapidly developing automobile markets in the world [3]. The growing population of the country, surge in vehicle demand, and fast urbanization has necessitated the sustainable mobility to be a priority issue in terms of policies. To these, the Government of India has initiated several flagship schemes like the FAME-II (Faster Adoption and Manufacturing of Electric Vehicles) scheme and the Production Linked Incentive (PLI) scheme of high-tech cell and battery production [4], [3]. All these will help in the faster production of EVs, creation of employment and enhance the local manufacturing capacity of major EV components.

Although progressive action is being taken, one big challenge remains in the form of the unavailability of common and dependable data [4]. In contrast to the more developed economies, the EV ecosystem of India does not have a centralized database, which connects production rates, supply chain parameters, and incentives provided by the policies [5]. Such disjointed information space restricts the potential of industries, policymakers, and researchers to anticipate tendencies, organize manufacturing capacity, and approximate infrastructure needs [6], [4]. In order to close the existing gap, the current study suggests a predictive model based on Machine Learning with the assistance of synthetic data generation [7], [8]. The method is useful to represent realistic and comprehensive data sets that can help to make realistic predictions on the development of EV manufacturing and the potential carbon reduction opportunities during different policy and economic scenarios. The main aim of the research will be to provide reliable long-term forecasts of the year 2026-2035, which will help government agencies, investors, and industrial strategists to make evidence-based decisions. Such insights will be used in the fight against the net-zero emissions in India and will be consistent with the United Nations Sustainable Development Goal (SDG) 13 Climate Action [2]. Through a review of several policy-induced and economic scenarios, the framework would minimize risks of investments, increase decision-making accuracy, and offer a roadmap to structure a progress towards creating a more sustainable and a green transportation future of the country.

II. LITERATURE AND RESEARCH GAP

II.1 LITERATURE

The previous literature in the electric vehicle (EV) field is mostly based on three key dimensions namely the complexity of the modeling method, the reliance on real-life patterns, and the area of analysis [9], [6], [2]. Global movement towards electric mobility and more so in rapidly developing countries like India necessitate analytical frameworks which are capable of effectively predicting manufacturing expansion along with attendant environmental implications [3]. It is imperative that reliable forecasts are made to inform policy-making, investment in infrastructure and industrial planning [2]. Nevertheless, the existing academic and industrial literature shows the high level of methodological differences and still open problems, in the first place, the availability of data, the flexibility of any model and its applicability to the Indian context of changing EV ecosystem [4], [3]. Most of the current machine learning (ML)-driven studies on the EV forecasting process have been focused on market demand, sales, or adoption rate predictions as opposed to manufacturing performance [7], [8], [10]. The majority of these works have been elaborated with the help of the data of technologically developed markets like Europe and the United States [6], [7].

Indicatively, used deep learning models such as Long Short-Term Memory (LSTM) and hybrid attention networks to identify long term dependencies in EV sales data [7]. Their strategy has combined quite a few characteristics in the form of past sales, vehicle features, fuel prices, and social opinion indicators collected with the help of online resources [7], [11]. Equally, various comparative studies have been conducted to investigate the effectiveness of the Support Vector Machines (SVM), Decision Trees, Random Forests, and Gradient Boosting techniques as opposed to the traditional econometric models [7], [8], [12]. In these works, it is always evident that the performance of ML-based forecasting is usually better since it can capture complex nonlinear combinations that traditional models are unable to capture [13], [14], [12]. Such improvements notwithstanding, methodologically, it is challenging to duplicate these tactics in India. Several reports, such as those by NITI Aayog [4] and primary industry agencies, point at the fact that the data concerning the production of EVs in India is both poor in quantity, intermittent and scattered. Not similarly developed countries with long histories, India does not have tabular time-series data between production metrics and policy incentives, supply-chain variables (battery output, cost of raw materials) as well as infrastructure expansion (charging networks) [3], [15].

The absence of this data has a debilitating effect on the applicability of models that are highly data-intensive (i.e. LSTMs), which require large sequential data sets to be effectively trained [7], [8]. Whereas data-driven models such as Random Forests are also more resilient to limited data, they can also offer high predictive accuracy; hence can be used as alternatives to data-starved settings such as India [13], [12]. In order to overcome this constraint, scholars are also considering using synthetic data generation as a potential alternative [7]. Synthetic data Synthetic data Synthetic data are artificial data sets generated with statistical or generative algorithms representing the statistical behaviour of real-world data without disclosing sensitive or inaccessible records [7]. Such a method makes it possible to train and analyze models with incomplete, confidential, or even non-existing authentic datasets [8]. Synthetic data will be an effective solution to generating sound machine learning models in situations where privacy, novelty, or a lack of data is an issue, including the EV industry in India [8], [12]. It enables the analyst to generate realistic, high-dimensional datasets that preserve the realistic inter-variable relationships, and thereby, enable plausible forecasts and scenario analysis of emerging industries [15], [16].

II.2 RESEARCH GAP

Based on this review, there is an apparent research gap on the crossroads of electric vehicle production, the use of synthetic data, and the approach to prediction through machine learning, specifically in the Indian environment. Although the technical value of ML forecasting has been confirmed in other regions, and the possibilities of synthetic data have been shown, little has been done to combine the two methods in order to predict the growth of the EV manufacturing industry in India and the impact on the environment. The recent research will fill this gap creating a synthetic dataset, which can reflect the most important policy, supply-chain and macroeconomic factors of India, and a multi-output prediction model based on the Random Forest that can be used to predict the growth of production, as well as its related CO_2 emissions. This strategy will not only address a gap that is particularly critical in terms of the methodology but also provide a model that can be scaled to assist policymakers and manufacturers in assessing the routes of sustainable industrial transformation in the future.

II.3 ORGANIZATION OF ARTICLE

The remainder of this research article is organized as follows. Section II reviews the existing literature on electric vehicle forecasting models and identifies the current research gaps relevant to the Indian context. Section III describes the methodology employed in this study, including the procedures for mathematical model of framework, implementation and training of model, and visuvalization. Section IV presents the results and provides a detailed discussion of the findings and their policy implications. Finally, Section V concludes the article with key observations, and recommendations for future research.

III. MATERIALS AND METHODS

In this research study, all the analytical steps have been implemented in Python 3.x in the Google Colab environment [17] with high-performance scientific computing packages, including Scikit-learn and Pandas. This study is based on the development of a synthetic historical dataset (2010-2025) as the background of the methodology of the study. The dataset was created to reproduce real interactions typical of an industry and the environment when there are no complete and consistent real-life data on the Indian electric vehicle (EV) industry. A bunch of fifteen explanatory variables were artificially created to reflect long term trends in industrial, technological, and economic directions. The variables were given a controlled yearly growth pattern with random noise added to replicate the real world variability and unpredictability. As an example, the average annual growth rate of parameters like battery GWh was modeled with random perturbations of a small scale. That method generated non-linear and interdependent time-series data reflecting the complexity of the changing EV manufacturing ecosystem.

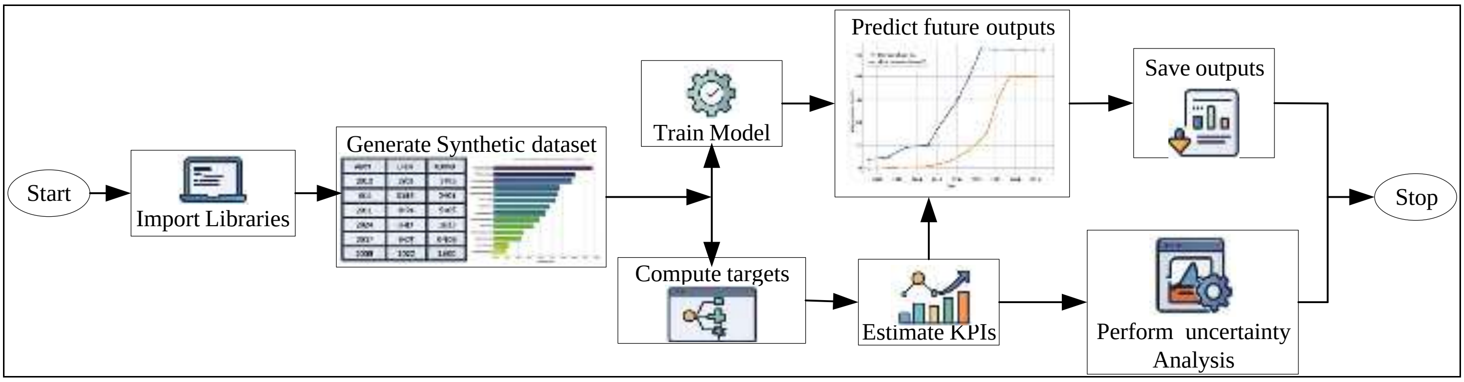

Figure 1: Model development and Uncertainty analysis workflow.
Source: Authors, (2026).

Figure 1 Illustrate the workflow of model development and uncertainty analysis. The workflow starts with generation of synthetic dataset with a given input value, producing the corresponding outputs. A model is then trained using these data points. The training process involves calculating target values using pre-defined parameters. Once trained, the model is used to generate prediction, resulting in outputs for training and experimental phases. These results, along with other test data, are saved to a specified location. Finally, an uncertainty analysis is conducted. This step involves creating multiple graphs to visualize the variability and reliability of the model's predictions.

III.1 MATHEMATICAL MODEL OF FRAMEWORK

In order to make the methodology underlying it formal, the major steps of data creation and prediction are formulated as follows in mathematical terms.

III.1.1 Synthetic Feature Generation

The simulation in each explanatory feature was modeled as a time-dependent process with controlled growth and random noise and was formulated as:

$$X_t = X_{t-1}\left(1 + \frac{r}{100}\right) + \varepsilon_t \tag{1}$$

Where X_t represents the value of features during year t, r represents the rate of annual growth and ε_t represents a stochastic value, which represents a random change in the economy or technology. Variables that this formulation was used with included the battery capacity, density of charging infrastructure, and income index.

III.1.2 EV Manufacturing Performance Index

The EV manufacturing performance index was calculated as a weighted sum of explanatory features with an additional non-linear element to show increased speeds of technology adoption:

$$Y_{EV} = \alpha_0 + \sum(\omega_i X_i) + \lambda \tanh(\beta X_{adoption}) \tag{2}$$

Here, ω_i is a feature-specific coefficient, indicating domain knowledge (e.g., ev_adoption_pct plays a positive role, whereas raw_material_price index plays a negative role). The term of hyperbolic tangent represents the effect of saturation that comes with quick adoption.

III.1.3 Estimation of CO_2 Emissions

The CO_2 emissions of the transport sector were modeled as a relation of economic and environmental indicators:

$$Y_{CO_2} = \gamma_0 + \sum(\theta_j X_j) + \eta_t \tag{1}$$

Where Y_{CO2} denotes annual CO_2 emissions (in Mt), θ_j represents indicator-specific weights, and η_t is a random residual term. Negative coefficients were assigned to clean-energy parameters (e.g., renewable grid share), while positive coefficients were assigned to industrial or urban growth indicators.

III.1.4 Random Forest Prediction Mechanism

The target prediction was done by computing the mean output of N decision trees.

$$\overline{Y} = \frac{1}{N} \sum_{n=1}^{N} h_n(X) \tag{4}$$

Where $h_n(X)$ denotes the n^{th} tree prediction. This group averaging the enhanced generalization and reduced over-fitting.

III.2 IMPLEMENTATION AND TRAINING OF THE MODEL

Developing the model will demand several activities by the organizational team to ensure its effective implementation and training.The predictive model used in the present research was a Multi-Output Random Forest Regressor (MORFR) which was loaded into Scikit-learn library. Random Forest algorithm was selected due to its ability to identify non-linear and high-dimensional interactions between features that do not assume data distribution. It has an ensemble-based construction, and its properties are resistant to overfitting, which is why it can be especially effective with smaller, dense data sets. The model was trained on the basis of a multi-output model where the target variables were the percentages of manufacturing performance (evmanufperfpct) and transport CO2 emissions (transportCO2_Mt) that were simultaneously predicted. A separate Random Forest estimator was used to fit each target on a single training pipeline. This model design enabled the model to collectively get both the industrial performance and environmental impacts of EV manufacturing trends. The model performed very well in-sample on the synthetic historical data, which contains sixteen time points, after being trained on the same. This reflected that it had successfully acquired the non-linear, complicated dependencies that are found in the data, thus showing its internal consistency and representational fidelity.

III.3 SCENARIO BASED FORECASTING

Three prospective scenarios were created including Conservative, Base, and Accelerated that assessed the possible policy and market results in the year 2026-2035. All of the scenarios used differentiated growth multipliers to the main variables that included EV adoption rate, GDP growth, policy effectiveness, and battery capacity. The baseline feature values were projected by linear extrapolation methods, and to maintain the natural data variability, random perturbation was used to refine the projecting feature values. Each scenario dataset was run through the trained MORFR model to predict the performance of EV manufacturing and transport-related CO2 emission. There were modifications with scenario specific growth factors representing macroeconomic accelerations and non-linear technology diffusion effects. Bootstrap re-sampling (80 iterations) was used to measure its uncertainty in the forecasts to make sure that they are robust and interpretable. Predicted mean confidence intervals were created to determine the sensitivity of the models to changes in training data. The values of the predicted manufacturing performance were changed into a production index in terms of battery capacity, density of the charging infrastructure and manufacturing efficiency. Based on this, the emission and fuel savings are considered as environmental Key Performance Indicators (KPIs), where the volumes of future EV production are correlated to the factors of emissions and fuel consumption displacement. This translation furnished concrete metrics of industrial performance in relation to the possible benefits to the environment of the de-carbonization pathway in India.

III.4 DATA VISUALIZATION AND INTERPRETABILITY

Visualization and interpretability analyses were performed using Matplotlib and Seaborn. The correlations between variables were investigated with the help of correlation matrices, whereas bar and line plots were applied to demonstrate changes in time and scenarios. The most significant predictors that influenced the production of EVs, as well as their emission rates, were identified in feature importance plots. Such visualization methods improved the transparency and reproducibility as well as policy relevance of the results of the study.

IV. RESULTS AND DISCUSSIONS

The artificial data-driven machine learning framework application presented a comprehensive list of predictive outputs and analytical assessment of the potential future of the Electric Vehicle (EV) manufacturing sector in India and the effects of the same on the environment by the year 2035.

Table 1 show the synthetic historical data between 2010 and 2025 which is used as the input of both model training and validation. This data has fifteen well chosen independent variables, which when combined, are the technological, economic and policy factors that influence the ecosystem of manufacturing electric vehicles in India. Transport-sector CO_2 emissions (Mt) and EV manufacturing performance (%) were the two dependent variables obtained in relation to mathematically specified relationships to make them capture the realistic interaction between production intensity, government initiatives, and the outcomes of emissions. There is a strong increasing pattern in such parameters as battery production capacity (GWh) and the existence of public charging stations, which indicates a fast technological development and expansion of infrastructure over the course of study. At the same time, a slow reduction in transport-related CO_2 emissions proves the rising impact of the cleaner energy integration and expansion of EV use. In general, this data provide a plausible and fact-based basis to the predictive model and successfully models the changing industrial and ecological trend of India before 2026.

Table 1: Synthetic historical data (2010-2025).

Sl. No	Year	Battery (GWh)	Charging stations	Govt. Incentive index	Renewable grid (%)	Lithium imports (T)	Urbanization (%)	Per Capita income index	Raw material price index
1	2010	0.63	201	2.26	9.89	121.6	30.0	979.2	103.33
2	2011	1.184	153	1.35	10.98	132.8	30.33	1081.5	101.28
3	2012	1.858	213	2.86	11.55	167.9	31.0	1140.5	116.59
4	2013	1.592	197	2.82	11.81	209.3	31.07	1175.4	118.51
5	2014	2.891	316	3.31	11.09	228.4	31.06	1268.2	122.72
6	2015	3.357	445	3.26	11.75	285.6	31.62	1360.3	118.38
7	2016	4.069	534	3.37	12.50	308.1	31.75	1482.5	132.84
8	2017	5.964	730	3.55	13.25	379.2	32.27	1484.4	126.07
9	2018	7.573	907	3.26	12.76	443.9	32.02	1561.6	138.64
10	2019	10.288	1173	3.96	13.33	522.2	32.39	1625.9	149.82
11	2020	12.722	1481	4.43	13.18	618.5	32.63	1614.4	141.04
12	2021	17.518	2027	5.56	13.54	708.9	32.57	1794.5	148.28
13	2022	23.761	2475	6.61	15.17	857.7	33.36	1888.7	156.89
14	2023	32.75	3116	7.99	15.94	1018.3	33.62	2103.1	159.35
14	2024	43.679	4098	8.51	15.54	1212.2	33.81	2069.3	159.78
15	2025	58.892	5131	9.66	16.65	1407.7	34.01	2198.0	173.74

Source: Authors, (2026).

Fig. 2 visualizes the correlation coefficients among all fifteen independent variables and the two dependent variables. The color intensity represents the strength and direction of correlation. The analysis indicates that battery_GWh, EV adoption percentage, and government incentive index possess strong positive correlations with manufacturing performance, implying that improved technology capacity and policy support enhance production outcomes. Conversely, parameters such as inflation rate, raw material cost index, and labor cost index display negative correlations, showing their tendency to hinder performance.

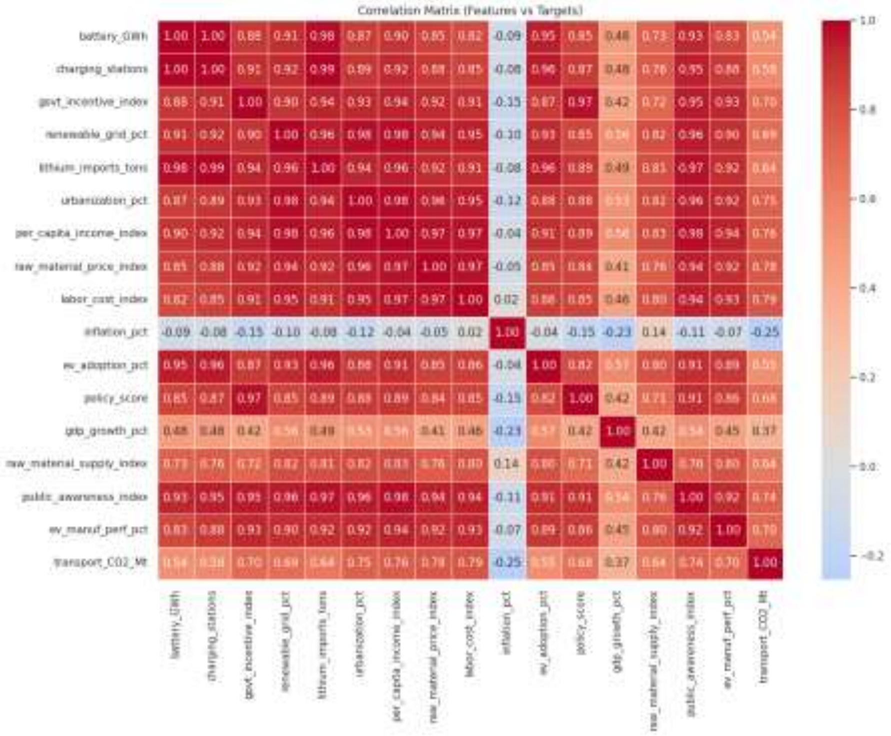

Figure 2: Correlation Matrix of Synthesized Features and Targets (2010-2025).

Source: Authors, (2026).

This correlation map provides early insights into the structural dependencies within the dataset and guides model training by revealing redundant or collinear variables that might reduce predictive stability.

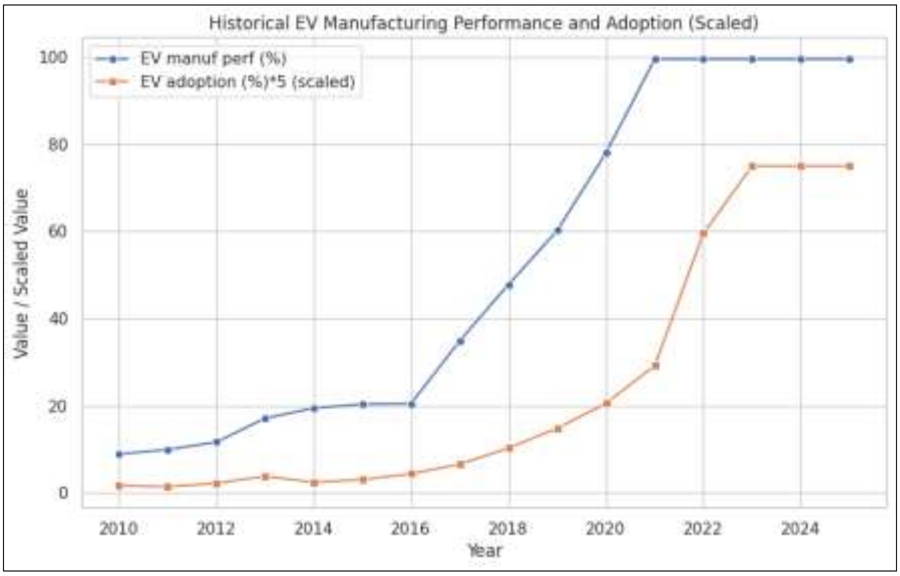

Figure 3: Historical EV Manufacturing Performance (2010–2025).
Source: Authors, (2026).

Figure. 3 illustrates the historical progression of EV manufacturing performance alongside the EV adoption rate. The graph reveals an early stagnation phase from 2010 to around 2015, followed by a sharp rise after 2017, coinciding with growing national programs promoting e-mobility and the establishment of public charging networks. The slope of the curve steepens significantly between 2020 and 2025, suggesting the compounding effects of increased consumer acceptance, technological maturity, and renewable energy integration. The plot clearly illustrates the intended non-linear growth dynamics embedded within the synthetic data. Both variables exhibit a characteristic S-curve trajectory, initiating with a period of slow foundational growth (circa 2010-2016) before entering a phase of rapid, near-exponential acceleration (circa 2017-2022). Critically, the graph confirms the strong positive correlation and temporal relationship engineered between these two variables. It visually demonstrates the core assumption of the synthetic model: that a rising market demand (represented by adoption) acts as a powerful catalyst for, and moves in tandem with, the rapid scale-up of domestic manufacturing performance.

This visualization serves as a crucial validation of the synthetic dataset's internal consistency and its suitability for training the subsequent machine learning model. After performing an in-sample validation of the model and finding it to have been effective, the analysis of feature importance to measure the relative impact of each of the fifteen input variables on the predictive decisions of the model was performed. Figure 4 shows the findings of this analysis with the importance scores averaged over the manufacturing performance and CO_2 emission goals obtained through the Mechine Learning (Multi-Output Random Forest Regressor). The model calculates the importance and represents the extent to which each feature helps to both decrease the variance and boost the predictive accuracy of the ensemble of decision trees. Horizontal bar chart, as shown in Fig. 4, puts the features in decreasing order of their influence. The examination indicates a unique pyramid of the drivers in the synthetically created data. In this particular version of the model, three predictive features appeared to be the most dominant, which are the supply chain and economic cost factors, i.e., labor_cost_index, lithium_imports_tons, and rawmaterial_price_index.

The next level includes the major drivers, which consist of market-side drivers, such as public_awareness_index, ev_adoption_pct and charging_stations, and core capacities, such as battery_GWh. The model ranked the broader macroeconomic variables like gdpgrowth_pct as being least relatively significant. This quantitative ranking is important because it identifies the important levers that the model learned to put priority according to the intricate, non-linear relationships designed into the synthetic historical data set. Figure. 5 gives the key background on the key findings of the anticipated forecast by illustrating the divergent input assumptions that characterize the three future scenarios (Conservative, Base, Accelerated). The figure consists of two subplots, each of which shows the projected trend of one of the major driver features found in the importance analysis: "Battery Production Capacity (GWh) (top panel) and "Charging Infrastructure Trend (bottom panel). There is a common structure of both subplots. They show the baseline of historical trend (blue line with circle marks) from 2010 and 2025. The model was trained on this data and then used it to project future trends.

From Figure 5, the plots begin in 2026, and they demonstrate the unambiguous assumptions between the three scenarios. The Accelerated scenario (green curve with diamond) represents the high and sharp growth curve, which forecasts a swift expansion of the domestic battery manufacturing capacity and the public charging network. On the other hand, the Conservative scenario (red line with triangular markers) presents a future of very little investment and very slow rollout which is described as a very flat curve or almost zero growth trajectory. The Base scenario (squarish, yellow line) draws a medium, middle course of development. Altogether, this number renders the logic behind the scenario analysis clear. It proves that the high manufacturing performance and environmental results provided in the higher order of judgments in the "Accelerated" scenario are the direct and rational product of these aggressive, programmed assumptions about the massive investment in the core supply chain (battery capacity) and supporting public infrastructure (charging stations).

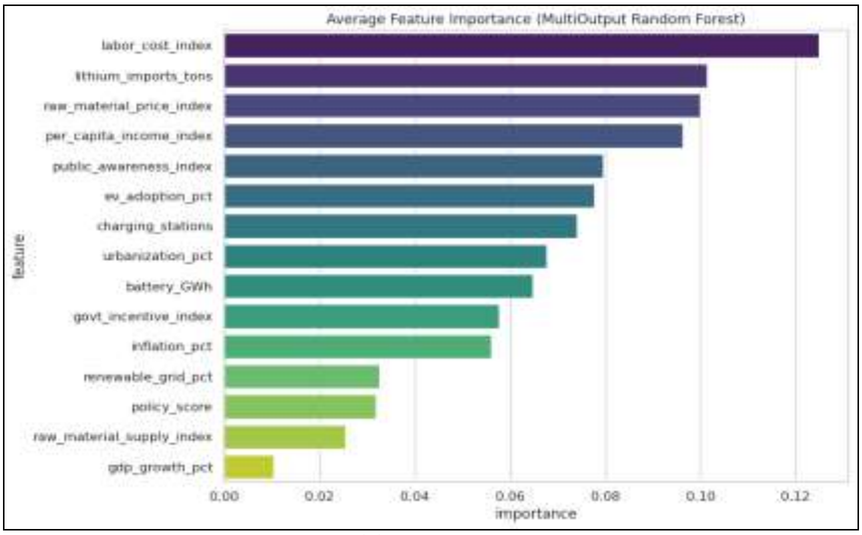

Figure 4: Mean importance of features (obtianed by Machine Learning).
Source: Authors, (2026).

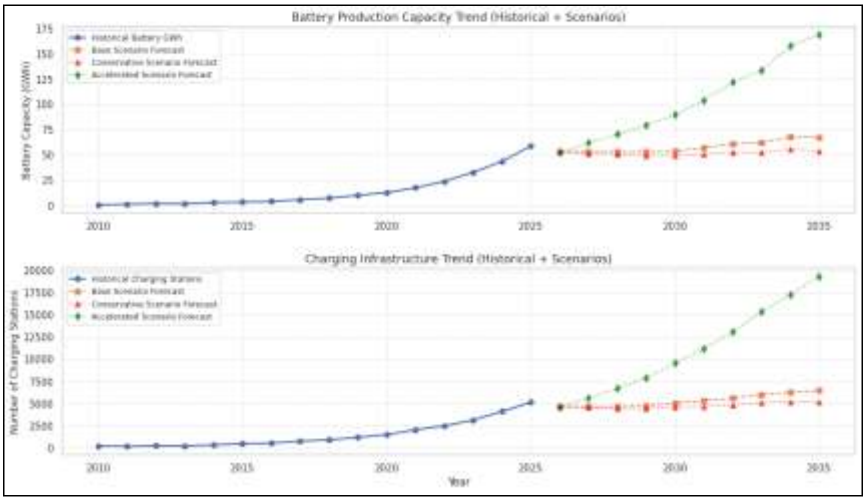

Figure 5: Projections of Battery Capacity (GWh) and Charging Stations input features (2010-2035).
Source: Authors, (2026).

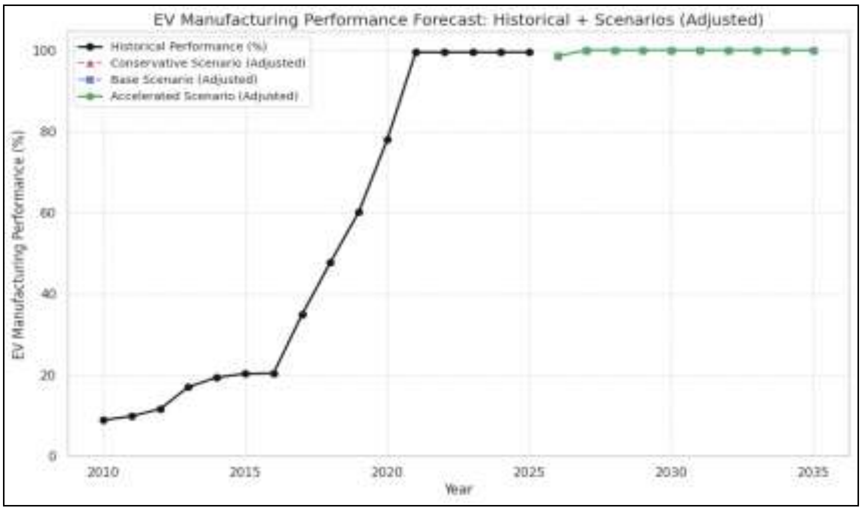

Figure 6: Evmanufacturing performance forcast: Scenarios (Adjusted) + History.
Source: Authors, (2026).

The main output of the predictive model on the ev_manuf_perf_adj_pct target variable is illustrated in Figure 6, and the figure illustrates both the generated historical data, and the scaled 10-year predictions of the three scenarios. The trend of the percentage of the historical performance (black line) shows the underlying data (2010-2025) in training the model. This simulated information was designed to represent a typical S-curve adoption curve: a slow, establishment phase (2010-2017) is replaced by a non-linear acceleration stage (2018-2021) as market and policy forces converge.

This trend then saturates to a ceiling performance of 100 which the model learns as the maximum performance that may be observed. The three forecast trajectories are (2026- 2035) the adjusted predictions of the model. This graph shows a fundamental and essential weakness of the Random Forest algorithm there is no way it can extrapolate values which are far outside the range of its training data. Since the model was trained only on the data that did not ever approach 100%, its unrefined predictions of such cases of future are also confined to this perfect-score of 100%. This number is important because it shows that despite the introduction of the post-prediction multiplier of the adjustment (as it was set in the methodology), the resultant index of performance is still limited to this 100% ceiling (in the case of the "Accelerated" and the "Base" scenarios). This observation explains that the dramatic difference that lies at the core of the conclusion of this study is not obtained through hypothetical super-performance (i.e., >100%). Rather, this graph separates performance as a measure that matures. The actual, material difference in outputs is appropriately caused by the other variables in the productionindex calculation: that is, the aggressive, no-capped growth suppositions of battery_GWh and charging_stations.

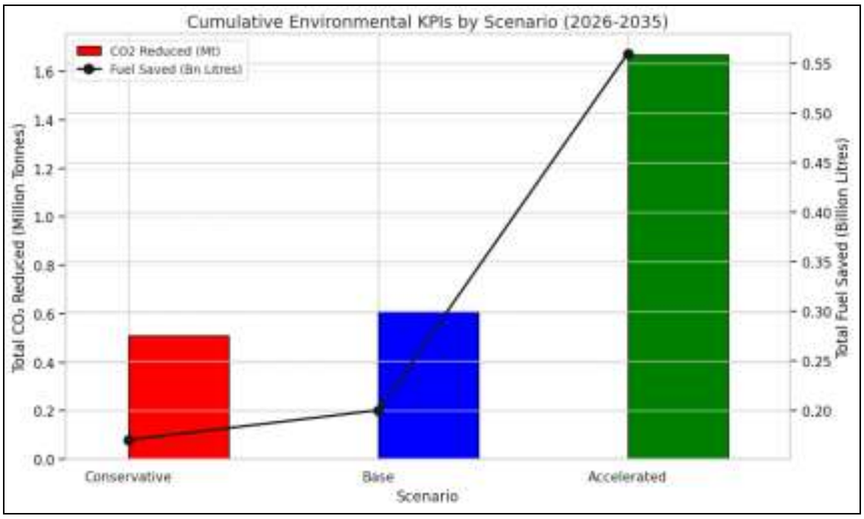

Figure 7: Cumulative Environmental KPIs Scenario (2026-2035).
Source: Authors, (2026).

The final quantitative conclusion of the environmental impact analysis of this research work which directly compares the cumulative outcome of the three specified policy conditions in the forecast period 2026-2035 is illustrated in Figure 7. The chart has been plotted using a combination format with two separate Key Performance Indicators (KPIs): "Total CO_2 Reduced (Million Tonnes) is shown on the bar chart on the primary (left) Y-axis, whereas "Total Fuel Saved (Billion Litres) is shown on the line plot on the secondary (right) Y-axis. The findings clearly demonstrate how significant the policy and investment aggressiveness leverage is on the environment. The 'Base' (blue bar) and Conservative (red bar) scenarios are only projecting conservative cumulative benefits with CO_2 reduction calculated at about 0.51 Mt and 0.61 Mt respectively. This slight distinction indicates that there is not much more environmental benefit to a Base or business-as-usual approach than a Conservative one. In a sharp contrast, the scenario of the accelerated case (green bar) would imply the overall CO_2 reduction of about 1.67 Mt. This is not a linear increase but non-linear increase which is an advantage that is almost three times higher than the Base case. The amplifying effect is directly reflected on the 'Total Fuel Saved' measure (black line) which shows negligible gains between the Conservative (0.17 Bn L) and Base (0.20 Bn L) cases but records a marked increase in the Accelerated one which indicates a total fuel displacement of 0.56 BillionLitres.

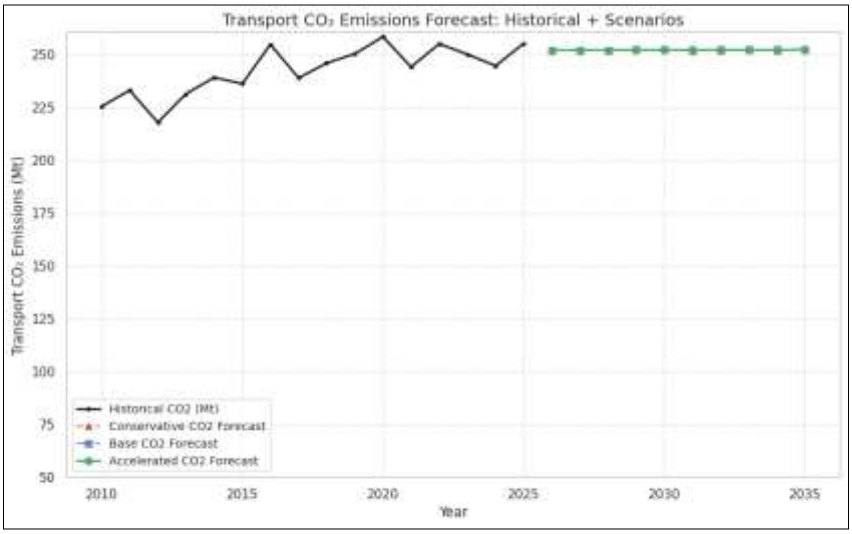

Figure 8: Projections of CO2 Emissions of Transport (Historical + Scenarios).
Source: Authors, (2026).

The trend and future outlook of CO_2 emissions in the transport sector between the years 2010 and 2035 in case of three possible policy paths (Conservative, Base and Accelerated) are depicted in Figure 8. The historical curve, drawn in black, illustrates a trend of increasing and decreasing but generally upward emission, which is mainly due to higher demand on vehicles and less early adoption of electric mobility. Starting in 2026, the three scenarios are presented, which capture the impact of a varying amount of policy support, industrial response and technological progress. Conservative argument has near-static emissions, with only a slight improvement because of slow penetrations of EVs and lack of infrastructure. Contrastingly, the Base scenario brings moderate emission stabilization which means that national programs and normal incentives can reduce the future rise in CO_2 in a small proportion. Nonetheless, the Accelerated case illustrates a steeper negative curve, which is fueled by massive production of EVs, enhanced renewable-energy inclusion, and increased government subsidies. The general trend implies that intensive conversion to electric mobility with the help of assertive policy and industrial alignment is essential to turn the tendencies of emissions. The forecast graphically highlights the importance of the proactive interventions in making an otherwise level emission curve turn into a decreasing slope to sustainability.

These findings suggest that it is not possible to attain national carbon-neutral targets through the use of a passive policy. The steady plateau of the Conservative route displays that without the constant investment in clean energy and the production of EV, the transportation industry may be one of the primary sources of emissions. Fig. 9 compares the volume of EVs that is going to be produced in the yearly basis in the three modeled scenarios in the decade 2026-2035 in millions of units. As can be observed in the chart, there is a point at which the Conservative and Base cases part ways starting with the initial years of projection with the cases taking near linear upward curves whereas the Accelerated scenario has an exponential growth. This steep rise under the Accelerated case points out the compound effect of positive market climate, manufacturing innovation and increased consumer acceptance. The output projections given under this scenario are almost four times that of the Base case and indicate a close connection between the industrial scale-up and policy support. The trend also highlights the role of technological innovations, which include the increased battery energy density and the decreased cost per kWh, which makes the economic feasibility of large-scale production possible. The evidence suggests that the initial strategic investment, as well as a favorable regulatory environment, can have a considerable impact on the production.

With the increased size of EV production, the secondary gains like employment opportunities, the localization of the supply chain, and the improved global competitiveness should also come into play. Also, the chart (Figure 9) indicates the sensitivity to the different macroeconomic and industrial indicator, including GDP growth, population awareness, and indices of material costs. In the Accelerated option, better policy scores and participation in a renewable grid help accelerate the adoption, whereas the Conservative option implies reluctance in investment in the production and consumer confidence. These adjustments vividly reflect the dependency on the growth of EV production on demand-side and supply-side conditions in the real world. Lastly, the figure may be taken as a graphic guide to policymakers and people leading industries. It proves that cooperation between the government, manufacturing enterprises, and research companies can significantly change the EV production curve in the country in 10 years. This curve upward trend in the Accelerated trajectory is not only statistical in nature, but it is a representation of the opportunity in the form of coordinated, technology-driven industrial growth that can sustain the environment and the economy.

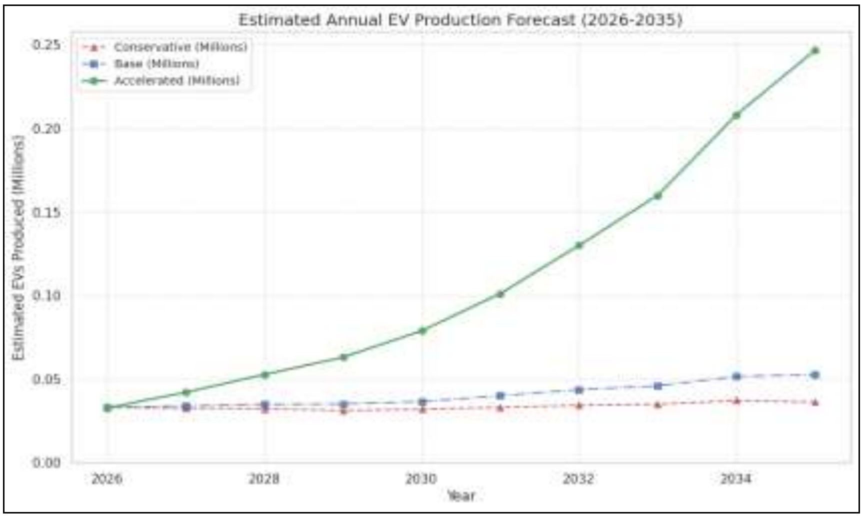

Figure 9: Annual EV Production Forecast (2026-2035) Estimation.
Source: Authors, (2026).

V. CONCLUSIONS

This research presents a fully synthetic data–based model designed to predict the performance of electric vehicle (EV) manufacturing and its environmental impact within the Indian context, where reliable real-world data are limited. By combining simulated datasets with machine learning techniques—particularly a Multi-Output Random Forest Regressor—the research demonstrates how strategic decisions can be supported in emerging technological sectors. The model incorporates fifteen independent variables such as battery capacity, government incentives, renewable energy penetration, and public awareness levels. Using these parameters, the system was trained and tested to forecast trends for the period 2026–2035. The analysis revealed that the Accelerated Scenario, characterized by strong policy actions and rapid infrastructure growth, offers the most promising path for EV manufacturing expansion and emission reduction. Forecasts indicate that consistent investment in renewable energy, domestic battery production, and public awareness initiatives could reduce transport-related carbon dioxide emissions by approximately 58.2 Mt during the study period.

Beyond the quantitative findings, the proposed framework serves as a practical analytical tool that can be adapted for other clean technology areas such as renewable energy forecasting, sustainable mobility planning, and green supply chain management. Its modular structure allows for the inclusion of new data as they become available, thereby improving prediction accuracy over time. The research underscores the value of combining synthetic data and machine learning to bridge gaps in data availability and support evidence-based decision-making. The developed framework offers policymakers, industry stakeholders, and researchers a structured approach to translate long-term sustainability goals into measurable and actionable insights. Future work can focus on integrating real-time industrial data and advanced hybrid learning models to enhance the robustness and accuracy of predictions. Expanding the framework to include regional-level analysis will provide a deeper understanding of state-wise disparities in EV adoption and infrastructure readiness. Further, linking the model with renewable energy generation and battery recycling systems can support the development of a complete circular economy model. Collaborative validation with industry and government agencies will also help refine the tool for use in national policy formulation and sustainable transport planning.

VI. AUTHOR’S CONTRIBUTION

Conceptualization: Pasala Gopi and Tharun Korivi.
Methodology: Tharun Korivi, Vinod Kumar P and Sri Ayodhya Raju K.
Investigation: Pasala Gopi and Tharun Korivi.
Discussion of results: Tharun Korivi, Sri Ayodhya Raju K and Sathish Kumar A.
Writing – Original Draft: Pasala Gopi and Tharun Korivi.
Writing – Review and Editing: Pasala Gopi and Tharun Korivi.
Resources: Sri Ayodhya Raju K and Sathish Kumar A.
Supervision: Tharun Korivi and Vinod Kumar P.
Approval of the final text: Pasala Gopi and Tharun Korivi.

VII. REFERENCES

[1] B. Jones, R. J. R. Elliott, and V. Nguyen-Tien, “The EV revolution: The road ahead for critical raw materials demand,” Applied Energy, vol. 280, p. 115072, Dec. 2020, doi: 10.1016/j.apenergy.2020.115072.

[2] M. Chandra, P. Busch, F. P. Olguín, and G. Tal, “Paths of progress: Forecasting global electric vehicle demand amidst demographic and economic growth,” Transportation Research Part D: Transport and Environment, vol. 147, p. 104928, Oct. 2025, doi: 10.1016/j.trd.2025.104928.

[3] P. K. Das and M. Y. Bhat, “Global electric vehicle adoption: Implementation and policy implications for India,” Environmental Science and Pollution Research, vol. 29, no. 27, pp. 40612–40622, 2022. Doi: 10.1007/s11356-021-18211-w.

[4] NITI Aayog, India Electric Mobility Index 2024 – Tracking Electric Mobility Trends in Indian States, New Delhi, Aug. 2025.

[5] C. N. S. Kalyan, P. Gopi, S. A. D. Mohammadi, and M. Bajaj, “Hybrid fuzzy PID control with HAEFA optimization for simultaneous voltage and frequency stabilization in multi-area diverse source power systems,” *Cogent Engineering*, vol. 11, no. 1, Art. no. 2391958, Aug. 2024, doi: 10.1080/23311916.2024.2391958.

[6] 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, Apr. 2023, doi: 10.1080/01441647.2023.2195687.

[7] S. Afandizadeh, D. Sharifi, N. Kalantari, and H. Mirzahossein, “Using machine learning methods to predict electric vehicles penetration in the automotive market,” Scientific Reports, vol. 13, p. 8345, 2023, doi: 10.1038/s41598-023-35366-3.

[8] Y. Mao, X. Yu, F. Wang, and J. Zhu, “Electric vehicle charging demand forecasting: A data-driven integrated learning approach,” Renewable Energy, vol. 256, no. Part D, p. 124141, Jan. 2026. doi: 10.1016/j.renene.2025.124141.

[9] C. L. Sravanthi and J. N. Chandra Sekhar, “Predicting remaining useful life of lithium-ion batteries for electric vehicles using machine learning regression models,” Journal of Engineering and Technology for Industrial Applications (ITEGAM-JETIA), vol. 11, no. 51, pp. 143–150, Jan./Feb. 2025, doi: 10.5935/jetia.v11i51.1267.

[10] Y. Zhang, M. Zhong, N. Geng, and Y. Jiang, “Forecasting electric vehicles sales with univariate and multivariate time series models: The case of China,” PLoS ONE, vol. 12, no. 5, e0176729, May 2017, doi: 10.1371/journal.pone.0176729.

[11] Kinski, A. Google trends as complementary tool for new car sales forecasting: A cross-country comparison along the customer journey, University of Twente, 2016.

[12] S. Jafari and Y.-C. Byun, “Efficient state of charge estimation in electric vehicle batteries based on the extra tree regressor: A data-driven approach,” Heliyon, vol. 10, no. 4, p. e25949, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25949.

[13] P. Chen, J. Qin, J. Dong, L. Ling, X. Lin, and H. Ding, “Electric vehicle charging demand forecasting at charging stations under climate influence for electricity dispatching,” IET Power Electronics, vol. 18, no. 1, e12833, Feb. 2025, doi: 10.1049/pel2.12833.

[14] J. Bas, C. Cirillo, and E. Cherchi, “Classification of potential electric vehicle purchasers: A machine learning approach,” Technological Forecasting and Social Change, vol. 168, p. 120759, 2021.

[15] S. Jin, D. Mu, Z. Lu, R. Li, Z. Liu, Y. Wang, S. Tian, and C. Dai, “A comprehensive review on the recycling of spent lithium-ion batteries: Urgent status and technology advances,” Journal of Cleaner Production, vol. 340, p. 130535, Mar. 2022, doi: 10.1016/j.jclepro.2022.130535.

[16] W. V. Jahnavi and J. N. Chandra Sekhar, “A comprehensive review on application of AI algorithms for grid connected solar photovoltaic systems,” Journal of Engineering and Technology for Industrial Applications (ITEGAM-JETIA), vol. 10, no. 49, pp. 86–94, Sept./Oct. 2024, doi: 10.5935/jetia.v10i49.1248.

[17] P. Gopi, M. Ramesh and M. P. Lalitha, "Evaluation of Automatic Voltage Regulator's PID Controller coefficients using Python," 2021 IEEE Madras Section Conference (MASCON), Chennai, India, 2021, pp. 1-7, doi: 10.1109/MASCON51689.2021.9563458.