

CLF-LDP Driven CO₂ Emission Forecasting for Smart Electric Vehicles

Rabab Hamed M. Aly¹, Aziza I. Hussein^{2, 3}, Naveed Anwar^{2, 3}

¹Higher Institute for Management and Information Technology, Minya, Egypt

²Department of Electrical and Computer Engineering, Effat University, Jeddah

³Faculty of Engineering & Technology, Multimedia University, Malaysia

Article Info

Article history:

Received Mar 01, 2026

Revised Aug 22, 2026

Accepted Sep 17, 2026

Keywords:

Hybrid Energy Systems

Machine Learning

Emission Prediction

Carbon Footprint

Energy-Efficient Modeling

ABSTRACT

Electric Vehicle (EV) networks play a crucial role in advancing environmental sustainability; however, inefficient charging and energy distribution strategies can contribute to increased CO₂ emissions. This study proposes a hybrid Crow–Lévy Fuzzy Optimization combined with Linear Discriminant Prediction (CLF-LDP) framework to enhance emission forecasting accuracy and support energy-efficient decision-making in EV systems. The proposed approach integrates the global search capability of the Crow Search Algorithm (CSA), enhanced by Lévy flight exploration and fuzzy logic-based feature refinement, followed by discriminant-based predictive modeling. Experimental evaluation demonstrates that the CLF-LDP model achieves approximately 97% predictive accuracy, with a test R² of 0.90 under weighted training and 0.966 after regularization-based comparison. In contrast to high-capacity deep learning architectures that exhibited instability on unseen data, the proposed optimization-driven framework maintains consistent predictive robustness while preserving computational efficiency. Furthermore, integration of the predictive equation into a Hybrid Electric Vehicle (HEV) Simulink environment validates its applicability beyond offline regression analysis. By embedding CO₂ prediction within the vehicle energy management system, the framework enables anticipatory emission-aware control rather than conventional rule-based adjustment. Overall, the CLF-LDP framework provides a scalable and computationally efficient solution that bridges predictive intelligence with real-time energy management, contributing to the development of sustainable and emission-aware smart transportation systems.

Copyright © 2026 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Rabab Hamed M.Aly,

Department of Information system,

The Higher Institute for Management and Information Technology, Minya.

New Minya, Minya, Egypt.

Email: rabab.cse2010@gmail.com

1. INTRODUCTION

Decarbonization is emerging as a key goal for global climate resilience and environmental sustainability initiatives. Over-dependence on fossil-fuel based industrial systems has played a huge role in the accumulation of greenhouse gases, air pollution, and long-term environmental instability. Beyond environmental effects, high CO₂ releases are linked to negative public health impacts and economic fragility. Thus, the shift to renewable energy systems and low-carbon technologies has turned into an environmental requirement and an economic priority [1]. Among various sustainability strategies, optimization-based modeling techniques play a crucial role in improving energy efficiency and supporting accurate greenhouse gas emission forecasting. The use of advanced optimization algorithms allows for automatic parameter tuning, feature selection, and predictive refinement, improving the confidence of the emission estimation model. Recent studies, such as [2], have demonstrated the effectiveness of hybrid metaheuristic approaches for CO₂ emission prediction. For example, the Bee–Particle Swarm Optimization (BPSO) algorithm has been used to enhance the accuracy of forecasting, using adaptive search mechanisms. Despite these developments, many of

the current methods require complex deep-learning models which can be unstable in the face of structured environmental data. This encourages research into hybrid approaches that combine predictive effectiveness, computational efficiency, and robustness. This paper introduces a hybrid Crow–Lévy Fuzzy (CLF) Optimization framework along with Linear Discriminant Prediction (LDP) to improve the emission prediction and energy management of Electric Vehicle (EV) networks in this context. The proposed approach combines the feature refinement using a metaheuristic with discriminant projection for the enhancement of predictive stability while maintaining the simplicity of the resulting model. The rest of this paper is organized as follows: Related work is reviewed in Section 2, the proposed methodology is presented in Section 3, experimental results and system-level validation is presented in Section 4, and Section 5 concludes this study.

2. LITERATURE REVIEW

Climate change is defined as long-term changes in climate patterns and temperatures, primarily caused by activities of humans like the large-scale use of fossil fuels. Climate change doesn't benefit human systems in itself, but the need to act has driven innovation and led to the creation of new and improved technological solutions. Many research studies have analysed the possibilities of incorporating machine learning methods into climate change forecasting and creating environmentally friendly solutions. To reduce the negative environmental effects, optimization of renewable energy systems and a strong policy decision are important strategies [3]. The problem of limited measuring station for solar radiation was addressed in [4] and the authors proposed to model and predict daily global solar radiation for three Turkish provinces based on meteorological input data and five metaheuristic algorithms (GBO, HHO, BMO, SCA, and HGSO). Their results showed that the GBO and SCA algorithms achieved the highest prediction accuracy. Further, the authors in [5] highlighted the need for accurate prediction of solar radiation as it is irregular and affects energy security especially for photovoltaic systems. This review introduced a new combination of metaheuristic optimization algorithms with AI techniques to tackle gaps in previous studies. The most popular optimizers, as identified, were Particle Swarm Optimization (PSO) and Genetic Algorithms (GA), while the most widely used AI techniques were Convolutional Neural Networks (CNN), Extreme Learning Machines (ELM), and Multilayer Perceptrons (MLP). Moreover, most studies have been dealing with intervals of radiation and input parameters like temperature, humidity, wind speed and air pressure. A thorough literature review from 1997 to 2023 was conducted on the application of metaheuristic algorithms for climate change adaptation and decision-making, according to [6]. The authors showed that genetic algorithms and particle swarm optimization have been widely effective in several applications such as energy efficiency and sustainable development, which means that they can be flexibly applied to the development of adaptive strategies. While these techniques have an excellent promise, the review revealed that they have several drawbacks including environmental coverage limitations and real-world validation requirements. In [7] the authors used a machine learning model which is based on an adaptive network-based fuzzy inference system (ANFIS), a deep belief network (DBN) and a hybrid ANFIS using the Snake Optimizer (ANFIS-SO), to predict potential evapotranspiration (PET) using data ranging from 1983 to 2020. The results obtained by their models indicated that the overall performance of the DBN model was $R^2 = 0.99$ and $NSE = 0.99$ at all the stations, whereas the ANFIS-SO hybrid model also showed significant performance improvement over standard ANFIS model with an R^2 value ranging from 0.52 to 0.99 at all the stations. Finally, the authors stated that the selection of the DBN, ANFIS, and ANFIS-SO can be made depending on the application conditions, and a comparative analysis of them was performed using the help of the Taylor diagrams. Further, the authors in [8] developed a novel Greylag Goose Optimization (GGO) algorithm with a Multi-Layer Perceptron (MLP) model to predict CO_2 emissions from EV use, contributing to environmental policy decisions and sustainability goals. The GGO optimized MLP significantly outperformed baseline models with lower error metrics, with statistical validation carried out using ANOVA, sensitivity analysis, and T-tests. Overall, the research found that the proposed forecasting method could offer valuable insights for policymakers on EV adoption strategies, charging infrastructure planning, and sustainable development. The authors in [9] studied the need for efficient research in the CO_2 emission at the building demolition phase, critical for achieving carbon neutrality. Machine learning models were designed using construction characteristics, equipment data and energy consumption information to predict emissions in the demolition stage. The Gradient Boosting Model (GBM) showed that floor area, equipment are the important emission factors with R^2 values of 0.997, 0.983, 0.984 for training dataset, testing dataset and validation dataset respectively. The study showed that such models can be used to assist in the early stages of design decisions and sustainability assessment, but there is a need for in-depth Life Cycle Evaluation tools to be developed. In [10] the authors proposed the deep learning model ConvLSTM-CBAMNet that combines channel attention mechanism and spatial attention mechanism to predict CO_2 emissions in container shipping by extracting complex spatiotemporal patterns. They obtained a 19.4% improvement in RMSE (0.0914) and a 16.8% improvement in MAE (0.0432) over the RMSE and MAE of ConvLSTM, respectively, in comparison to four deep learning baselines (Transformer, ConvLSTM, YOLOv3 and YOLOv4). The study found that these

accurate forecasts can be used to inform environmental action in ECA, dynamic ECA regulation, port scheduling, and even early warning systems for sustainable maritime mobility. The authors in [11] have presented a mixed-integer linear programming (MILP) model for optimal scheduling of battery energy storage systems (BESS) in residential applications using solar generation to provide increased flexibility and efficiency for energy management. The objective function also includes a carbon emission term to include environmental impacts from battery degradation due to cycle and calendar aging. The study employed the optimized BESS scheduling by Python and CBC solver, which showed that optimized BESS scheduling can decrease energy costs, emissions, and dependence on the grid while also emphasizing the effects of active control strategies under different state-of-charge limits and operating conditions. The hybrid microgrid design in [12] was introduced using a simulation-based design including PV generation, battery storage and diesel backup to enhance the hybrid microgrid performance. They created a multi-objective optimization model using NSGA-II algorithm to minimize lifecycle costs, carbon emissions while maintaining operational reliability. The results revealed that the optimized scheduling significantly decreased generator dependency, and the cost reduction was more than 90%, while the emissions were not significant. The study has found that intelligent control strategies based on batteries are superior to simple capacity expansion of renewables in assuring low carbon and cost-effective operation of microgrids. In addition, the authors in [13] introduced a multi-layered computational framework to estimate the flexible capacity of electric bus battery swapping stations on both regular and real-time horizons, explicitly accounting for carbon emissions to improve coordination in renewable energy-integrated power grids. The model incorporated traffic flow and temperature information to achieve the accuracy of the prediction, a flexible capacity estimation model, and a dual mode response mechanism for real-time adaptation. The framework was validated by data from a large city and found to have an acceptable degree of flexibility and usefulness for low-carbon operational planning. In [14] the authors proposed a coordinated scheduling method of the EV load using the active carbon factor forecasting and multi-objective optimization. A dual-convolution enhanced prediction model was proposed, which was formed by parallel global and local convolution units and an adaptive spectral block. A carbon-electricity cost joint optimization model was developed to assess the economic, environmental and grid objectives. The proposed multi-objective approach showed significant improvement compared to the single objective approach and model predictive control benchmarks when compared to real grid data for renewable integration, yielding 15.8% carbon emission reduction, 5.2% economic costs reduction, and 22.2% peak-valley reduction. In [15] the authors suggested a multi-layer machine learning model to schedule optimal charging times of EVs in residential buildings with PV and battery storage, maximizing self-consumption and reducing grid use and environmental footprint. Their approach combined high-fidelity dynamic simulations with multi-objective optimization to maximize self-sufficiency and minimize electricity costs and operational CO₂ emissions. After simulation-based optimization, a labeled dataset of 315,360 samples was generated to train six machine learning models, where ensemble tree-based models achieved the highest predictive accuracy ($R^2 \approx 0.9923$). Correlation analysis and SHAP interpretation indicated that the self-sufficiency ratio was the most influential predictor, followed by economic and environmental factors. In recent years, some of the best adaptive equivalent consumption minimisation strategies (A-ECMSs) have been seen. A review of A-ECMSs was presented in [16]. There is a wide coverage of possibilities presented by the authors, including equivalence factor adaptivity imposed by the proportional-integral (PI) rule. Further, using optimal algorithms or model predictive control (MPC), the optimal plan is derived from the battery SoC trajectory. Machine learning techniques are employed in [17] to overcome implementation barriers of the Equivalent Consumption Minimisation Strategy (ECMS) in hybrid electric vehicles (HEVs). While ECMS provides near-optimal control for HEVs, its computational intensity and sensitivity to driving conditions limit its practical implementation in production vehicles.

Previous studies have primarily focused on complex modeling techniques. In contrast, this paper demonstrates the effectiveness of integrating lightweight predictive modeling with post-prediction energy control mechanisms. It provides validation of the CLF Optimization combined with LDP framework to support sustainable hybrid and electrified transportation systems.

3. METHODOLOGY

This section allows a model to learn from most of the data while reserving a portion of the data for evaluating its performance on new, previously unseen examples. There are 7,385 entries in this study and 12 features [18]. Environmental alterations are influenced by effectors by altering surrounding conditions, where fuzzy-based evaluation mechanisms are incorporated to improve solution discrimination during optimization [19]. Scalability, simplicity, and localized interaction make it an efficient solution for large-scale computation challenges.

3.1. Model Implementation

The applied modeling technique uses Crow–Lévy Fuzzy (CLF) Optimization followed by Linear Discriminant Prediction (LDP) to estimate and predict CO₂ emissions based on the input features that are available. A complete model of the system is shown in Figure 1. The intelligent vehicle input data flows contain requirement, model, vehicle class, engine size, cylinders, transmission type, fuel type, fuel consumption (city and highway), and CO₂ emissions. The following describes each component of the model.

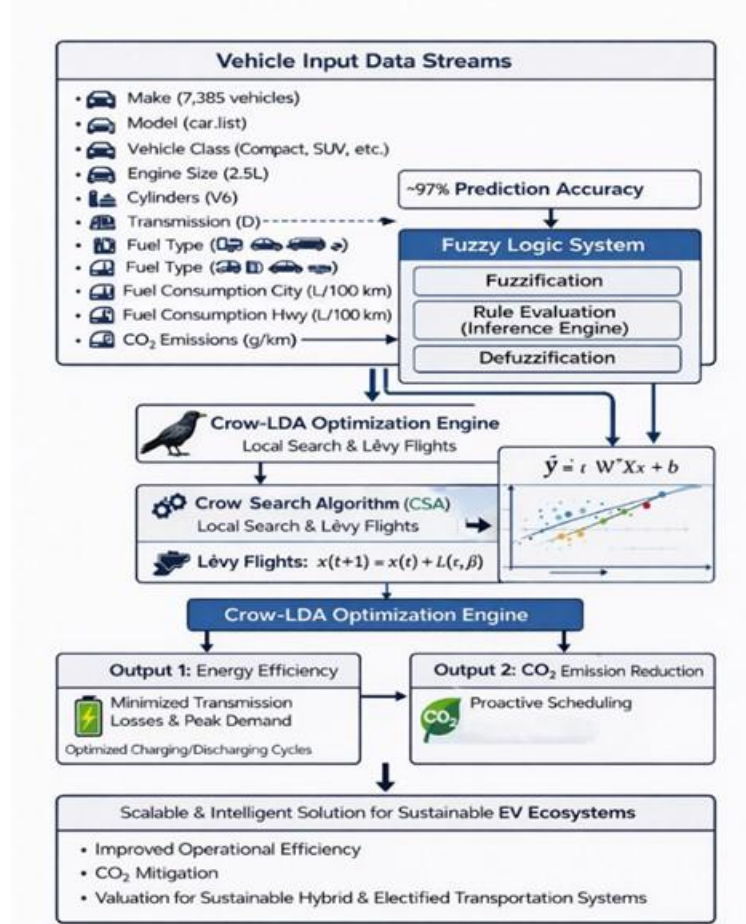


Figure 1: Overall Architecture of the Proposed Framework.

3.1.1. Linear Discriminant Prediction based on Crow–Lévy Fuzzy Optimization

In this paper, a Crow-LDA Optimization Engine uses the refined data to perform local search enhanced by Lévy flight exploration to improve global convergence using the Crow Search Algorithm (CSA). A linear decision function is used in the optimization process for the prediction of energy and emission-related outputs using Linear Discriminant Analysis (LDA). By minimizing transmission losses and optimizing charging and discharging cycles, the final part of the system predicts CO₂ emissions and improves energy efficiency. The main steps of the algorithm are as follows:

- Initialize the parameters; tmax represents the maximum number of generations, AP represents the awareness probability, N represents the flock size, PD represents the problem dimension, and FL represents the flight length.
- The memory of crows should be initialized and evaluated using the parameters in equations (1) and (2).

$$Crows_{i,t} = \begin{bmatrix} x_i^1 & \dots & \dots & \dots & \dots & x_i^d \\ \vdots & & & & & \vdots \\ \vdots & & & & & \vdots \\ \vdots & & & & & \vdots \\ x_N^1 & \dots & \dots & \dots & \dots & x_N^d \end{bmatrix} \quad (1)$$

$$CrowsMemory_{i,t} = \begin{bmatrix} m_i^1 & \dots & \dots & \dots & \dots & \dots & m_i^d \\ \vdots & & & & & & \vdots \\ \vdots & & & & & & \vdots \\ \vdots & & & & & & \vdots \\ m_N^1 & \dots & \dots & \dots & \dots & \dots & m_N^d \end{bmatrix} \quad (2)$$

Where the range between lb (lower boundary) and ub (upper boundary) is presented as randomly initial position of each crow. Additionally, $i=[1, 2, \dots, N]$, t are $1, 2, \dots, tmax$ and d is pd [20-27].

- Develop new crow positions and evaluate them based on the following equations [28-32].

$$x_{i,t+1} = x_{i,t} + r_i \times fl(m_{i,t} - x_{i,t}) \quad (3)$$

- In the update of the new position of crow based on levy distribution, Lévy random variables are generated using equation (4), (5).

$$levy(\alpha, \beta, \gamma, \Delta) \quad (4)$$

Where Δ is the location parameters and it is computed by equation (5).

$$\Delta = \frac{x_i(t) + p_{best}}{2} \quad (5)$$

On the other hand, γ is scale parameters and it is computed by equation (6)

$$\gamma = \frac{x_i(t) - p_{best}}{2} \quad (6)$$

The position will be computed by equation (7)

$$x_i(t+1) = levy(\alpha, \beta, \frac{x_i(t) - p_{best}}{2}, \frac{x_i(t) + p_{best}}{2}) \quad (7)$$

Where β is symmetry parameter, α is a shape parameters, and γ is the empirical standard deviation.

Then apply the Fuzzy fitness as shown in (8), where fuzzy-based evaluation strategies are adopted to enhance solution discrimination during optimization [19].

$$F_{fuzzy} = \frac{\int z \mu(z) dz}{\int \mu(z) dz} \quad (8)$$

$$\Delta f_i = \frac{f_i^{t-1} - f_i^t}{|f_i^{t-1}| + \epsilon} \quad (9)$$

where f_i^{t-1} and f_i^t are the fitness values of the i -th crow at iterations $t-1$ and t , respectively, and ϵ is a small positive constant used to avoid division by zero.

$$D_i = \frac{1}{N} \sum_{j=1}^N \|x_j - \bar{x}\| \quad (10)$$

where N is the population size, x_j is the position of the j -th crow, and $\bar{x}$ is the mean position of the crow population.

$$F_i^{fuzzy} = \frac{\int z \mu_{\tilde{F}_i}(z) dz}{\int \mu_{\tilde{F}_i}(z) dz} \quad (11)$$

where z is the fuzzy output variable and $\mu_{\tilde{F}_i}(z)$ is the aggregated fuzzy membership function used for centroid defuzzification.

- Update the memory based on (11) as shown in equation (12) and equation (13) is the best global and the limitation to $t \geq t_{max}$
- Termination of repetition: The procedure of Steps 2–4 is iteratively executed, and upon reaching $tmax$, the operation of the crow algorithm is concluded to obtain optimization outcomes.

$$m_i^{t+1} = \begin{cases} x_i^{t+1}, & \text{if } f_i^{t+1} < f(m_i^t) \\ m_i^t, & \text{otherwise} \end{cases} \quad (12)$$

$$p_{best} = \arg \min_{m_i^{t+1}} f(m_i^{t+1}) \quad (13)$$

By using LDA, the optimised features are projected using the main equation of LDA [27-37]. As shown in equations (14,15) and (16) respectively, the within-class scatter matrix which C represents a distinct classes or categories in our dataset. The between-class scatter matrix is calculated, and the projection matrix W is obtained by solving the eigenvalue.

$$S_W = \sum_{k=1}^c \sum_{x_i \in C_k} (x_i - \mu_k)(x_i - \mu_k)^T \quad (14)$$

$$S_B = \sum_{k=1}^c N_k (\mu_k - \mu)(\mu_k - \mu)^T \quad (15)$$

$$S_W^{-1} S_B W = \lambda W \quad (16)$$

The features are then projected as

$$z = W^T x \quad (17)$$

The final continuous prediction is achieved by substituting the projected features into a regression-based equation:

$$\hat{y} = w^T W^T x + b \quad (18)$$

Where x is optimized extracted features which will be used for battery recharge regulation, W is the LDA projection matrix, w is a regression weights, b is bias, and $\hat{y}$ is the predicted output.

Based on the application of this paper, the calculation of continuous CO₂ value y_i is used as the regression target:

$$\hat{y}_i = \beta_0 + \beta^T z_i \quad (19)$$

Substituting Eq. (17):

$$\hat{y}_i = \beta_0 + \beta^T W^T x_i \quad (20)$$

Thus:

$$\hat{y}_i \in \mathbb{R}$$

Where x_i is CLF-optimized feature vector, β is regression coefficient vector, β_0 is regression bias, z_i is LDA-projected feature vector, W is LDA projection matrix, and $\hat{y}_i$ remains a continuous CO₂ prediction in g/km.

On the other hand, LDA is formulated for supervised classification, employing it to continuous CO₂ emission prediction requires a mathematically sound target mapping strategy. In this study, the continuous target output vector $y \in \mathbb{R}^N$ is mapped into C distinct functional operational regimes using a quantile-based discretization scheme:

$$y^{(c)} = f(y) = \sum_{c=1}^C c \cdot \mathbb{I}(q_{c-1} \leq y < q_c) \quad (21)$$

where C represents the total number of operational emission regimes, configured as $C = 4$, representing Low, Moderate, High, and Peak emission states, and $\mathbb{I}(\cdot)$ is an indicator function that returns 1 if the specified condition is satisfied and 0 otherwise.

The quantile thresholds are calculated as:

$$q_c = \text{Quantile}\left(y, \frac{c}{C}\right), q_0 = \min(y), q_C = \max(y) \quad (22)$$

where q_c denotes the c -th empirical quantile threshold, y represents the continuous CO₂ emission target values, and C is the number of emission regimes.

By utilizing quantile-based mapping, each discrete operational class $c \in \{1, 2, \dots, C\}$ approximately retains an equal proportion of samples. This strategy helps mitigate class imbalance during the construction of the LDA scatter matrices while preserving the original continuous CO₂ emission values for the subsequent regression stage. The proposed model captures the dominant data patterns and has high predictive accuracy [18]. The next section is a specified estimation with other state-of-the-art models, to further prove its authority, and it is very well qualified. In addition, the high accuracy of CO₂ emissions predictions aids a uniform transition to the next phase: optimising battery charging procedures for electric vehicles. The system can control these accurate predictions to handle the management of energy distribution, which can help to manage energy use and improve transportation sustainability.

3.2. Comparative Models

This section introduces neural network models to prove the effectiveness of our proposed system. The models implemented in this part are a suite of excellent neural network models such as Transformer, CNN, MLP, and FNN and are recommended as benchmark models to be used. These models were trained and evaluated on the same dataset as our established model. The comparison highlights the advantage that our proposed architecture is based on, and extends, the field of neural network and deep neural network methods and achieves enhanced predictive performance. The following is a short summary of some of the most common deep learning (DL) regression methods:

- (a) **Transformer**: A transformer neural network architecture that is used in regression tasks, such as predicting future carbon dioxide (CO₂) emissions. In contrast to linear or transformer models, models that are particularly designed to process sequential data can capture complex relationships over time. This makes them very suitable for emission time series forecasting. In the area of carbon emission, the transformer regression models are used for learning emission data from the past and forecasting the development of emission into the future. By handling the data as sequences, these models can detect and employ temporal patterns that may be ignored by more traditional models [39].
- (b) **CNN (Convolutional Neural Network)**: CNN regression is used to predict continuous numerical variables instead of categorizing data into discrete categories, using convolutional neural networks (CNNs). For carbon emissions, CNNs are used to analyze structured or spatial data, which allows the model to identify relationships between the input variables and the output variables, in this case, carbon emissions. Convolutional networks learn related features from the input data automatically; thus, they can learn the complex, non-linear relationships that regression-based methods cannot learn [39].
- (c) **MLP (Multi-Layer Perceptron)**: Multi-Layer Perceptron (MLP) is a feed-forward architecture with an input layer, a hidden layer and an output layer. Each layer consists of a network of nodes (neurons) that receive input data, process it via weighted links and nonlinear activation functions. For regression problems, the MLP should be able to predict a continuous output (CO₂ emissions) by minimizing the prediction error throughout the training process. The patterns of the data are used to learn emission levels or related changes using MLP regression. The model can detect complex and nonlinear relationships that may not be detected by simple linear regression [40].
- (d) **FNN (Feedforward Neural Network)**: Feedforward Neural Networks (FNNs) are a type of machine learning network designed to solve complex problems by identifying the structural relationships between the input and output variables. For regression problems (predicting carbon emissions, for example), FNNs use supervised learning algorithms to adjust the weight coefficients between layers to reduce the error between the output of the FNN and the actual target value. The aim is to get the best possible correlation between the production and target vectors; performance of the model is judged by appropriate measures. The FNN models can capture complex interactions among input features and target outputs, enhance prediction accuracy [38-43].

While these models offer powerful predictive capabilities, our proposed method CLF-LDP trains to balance computational simplicity with high accuracy by integrating minor predictive models with robust control systems.

4. RESULTS AND DISCUSSION

The results in this section fall into two broad evaluation phases: traditional predictive methods and application of our proposed approach. The performance of each phase is carefully estimated by calculating the error values. A comparison is followed at every step between the traditional error metrics and the proposed error metrics. The proposed model is demonstrated to minimize prediction errors and develop complete system consistency in this qualified estimation.

For evaluating comparison models and proposed models, the following error evaluation parameters are used:

- 1) Mean Squared Error (MSE) calculates the average squared error:

$$MSE = \frac{1}{n} \sum (y_i - \hat{y}_i)^2 \quad (21)$$

- 2) Root Mean Squared Error (RMSE) is the square root of MSE, providing error in the same units as the target:

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad (22)$$

- 3) The coefficient of determination (R^2) evaluates how well predictions approximate actual values:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (23)$$

(a) The Models with Weighted Loss Functions

All tests were implemented using MATLAB R2023b on a system equipped with 16 GB RAM and an Intel Core i3 processor operating at 2.3 GHz, ensuring a consistent computational environment across all experiments. To reduce overfitting and enhance robustness, the original dataset of 7385 records [18] was extended to approximately 22,000 records using traditional data augmentation methods.

Table 1 indicate substantial overfitting in several baseline models, with high training R^2 values but poor or negative testing performance, particularly for the Transformer, MLP, FNN, and CLF models. Notably, CLF achieved a training R^2 of 0.988 but decreased to -0.68 on the test set, demonstrating limited generalization. In

contrast, the proposed CLF-LDP framework reduced the training R^2 to 0.968 while substantially improving the test R^2 to 0.90, confirming the effectiveness of the LDP stage in enhancing model generalization and mitigating overfitting.

Table 2 shows the comparative evaluation of five models for both training and testing data sets for MSE, RMSE and R^2 . To ensure a balance in learning and to compensate for irregularity in the data distribution, all models were trained with weighted loss functions. Both the Transformer and the proposed CLF-LDP method were found to be very effective in the training dataset, where the fitting efficiency was reflected by the low error values and the high R^2 scores (≈ 0.97 and ≈ 0.968 , respectively). CNN showed moderate fitting ability, while MLP and FNN showed good fitting ability.

However, the Transformer, MLP, and FNN had negative R^2 values for the testing set (Transformer: -0.72 , MLP: -3.04 , and FNN: -1.06), meaning that there was poor predictive consistency for unseen data. This is considered as instability of model generalization, in the presence of good training performance. The proposed CLF-LDP method, on the other hand, showed a high-test R^2 value of 0.90 and competitive error metrics, which were consistent across unseen samples as seen in Figure 2. This observed instability of deep architectures with weighted training indicates that large-capacity neural models can be sensitive to the attributes of the data set, as well as its size, when not sufficiently regularized.

In contrast, the CLF-LDP framework has the advantage of a hybrid metaheuristic-based feature refinement method with discriminant projection that removes dimensional redundancy in the prediction process. The results of these studies show that the depth of the architecture is not the only factor that influences the strength of emission forecasting, but also good feature optimization and dimensional control.

Table 1. Performance Comparison of Baseline Models and the Proposed CLF-LDP Framework on Training and Test Data

MODEL	TRAIN R^2	TEST R^2
TRANSFORMER	0.970	-0.72
CNN	0.240	0.19
MLP	0.820	-3.04
FNN	0.890	-1.06
CLF	0.988	-0.68
CLF-LDP (PROPOSED)	0.968	0.90

Table 2. The Models with Weighted Loss Functions

USED MODELS	TRAIN DATA			TEST DATA		
	MSE	RMSE	R^2	MSE	RMSE	R^2
TRANSFORMER	82.49	9.08	0.97	4924.84	70.18	-0.72
CNN	2048.26	45.26	0.24	2320.62	48.17	0.19
MLP	478.57	21.88	0.82	11591.30	107.66	-3.04
FNN	307.62	17.54	0.89	5904.64	76.84	-1.06
CLF-LDP (PROPOSED METHOD)	86.28	9.54	0.968	4870.85	72.89	0.90

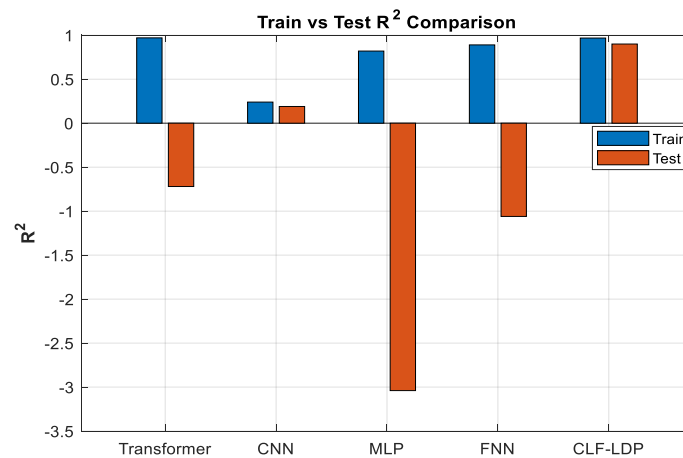


Figure 2. The Models with Weighted Loss Functions

b) The models with dropout and batch normalization

To improve the stability of the models, dropout, batch normalization, and data augmentation (CNN) are applied and the model performance is evaluated in Table 3. In these regularization methods, CNN exhibited the best and stable performance: training results (MSE: 55.47, RMSE: 7.45, R^2 : 0.97) and Testing results (MSE: 89.53, RMSE: 9.46, R^2 : 0.97) showed that it was the most robust method.

The Transformer exhibited better testing performance ($R^2 = 0.73$), which is due to the regularization of the model, creating a more stable performance compared to its previous instability. The testing R^2 values were moderate for MLP ($R^2 = 0.84$) and FNN ($R^2 = 0.89$), which indicates that they have good predictive consistency. Results from the proposed CLF-LDP method achieved acceptable error metrics (MSE: 470.85, RMSE: 12.89) and good test R^2 (0.966) confirming its reliability under similar test conditions. The CLF-LDP model outperformed the deep architectures which needed explicit regularization layers to stabilize training in terms of competitive robustness by optimizing forward feature refinement. This implies that dimensionality-aware optimization is able to do variance control inherently while maintaining predictive sensitivity.

Table 3. Models with Dropout and Batch Normalization

USED MODELS	TRAIN DATA			TEST DATA		
	MSE	RMSE	R^2	MSE	RMSE	R^2
TRANSFORMER	82.40	9.08	0.97	778.42	27.90	0.73
CNN	55.47	7.45	0.97	89.53	9.46	0.97
MLP	100.00	10.00	0.96	455.81	21.35	0.84
FNN	117.87	10.86	0.96	314.25	17.73	0.89
CLF-LDP (PROPOSED METHOD)	87.28	9.54	0.966	470.85	12.89	0.966

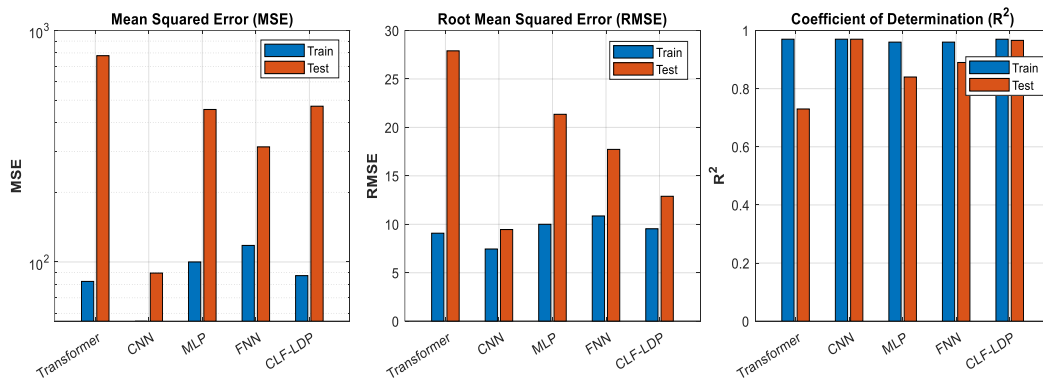


Figure 3. The models with dropout and batch normalization

Overall, these results demonstrate two distinct pathways toward robustness: architectural regularization in deep models and optimization-driven feature control in hybrid frameworks. The latter offers competitive accuracy with lower structural complexity and reduced computational burden.

c) Transition from Predictive Evaluation to System-Level Validation

The statistical measures indicate that the proposed model can predict, but the validation in ITS application requires more than offline regression assessment. In the context of an Electric Vehicle (EV), emission prediction is to be involved in the operational decision-making process and should lead to tangible sustainability improvements. The CLF-LDP capability was incorporated in the autoblksHevStart Hybrid Electric Vehicle (HEV) Simulink environment [42, 43] to evaluate it. The setup is based on the predicted CO_2 emission as a dynamic supervisory signal for the vehicle's energy management system (EMS). The integration makes the emission forecasting from a passive analytical product to a product that is actively used as a control. The Performance and FE Scope modules are used to collect critical operational data like Trace Velocity, Battery Current and Battery State of Charge (SOC) which are used as baseline control references. The SOC profile remains within an optimal operating range of approximately 70%–80%, supporting battery longevity and energy efficiency as shown in Figure 4. Additionally, Figure 5 shows that the upper plots exhibit that while the benchmarks suffer from severe over-fitting on test data ($R^2 < 0$), the proposed CLF-LDP model maintains robust prediction performance ($R^2 = 0.90$) with low RMSE. The lower-left plot highlights that the CLF-LDP administrator become stable the dynamic battery State of Charge (SOC) within the target 70% – 80% optimal operational range, compared to severe discharge fluctuations in the baseline controller. The lower-right plot

confirms a significant reduction in overall trip cumulative CO_2 emissions over time when utilizing the proposed predictive supervisory signal.

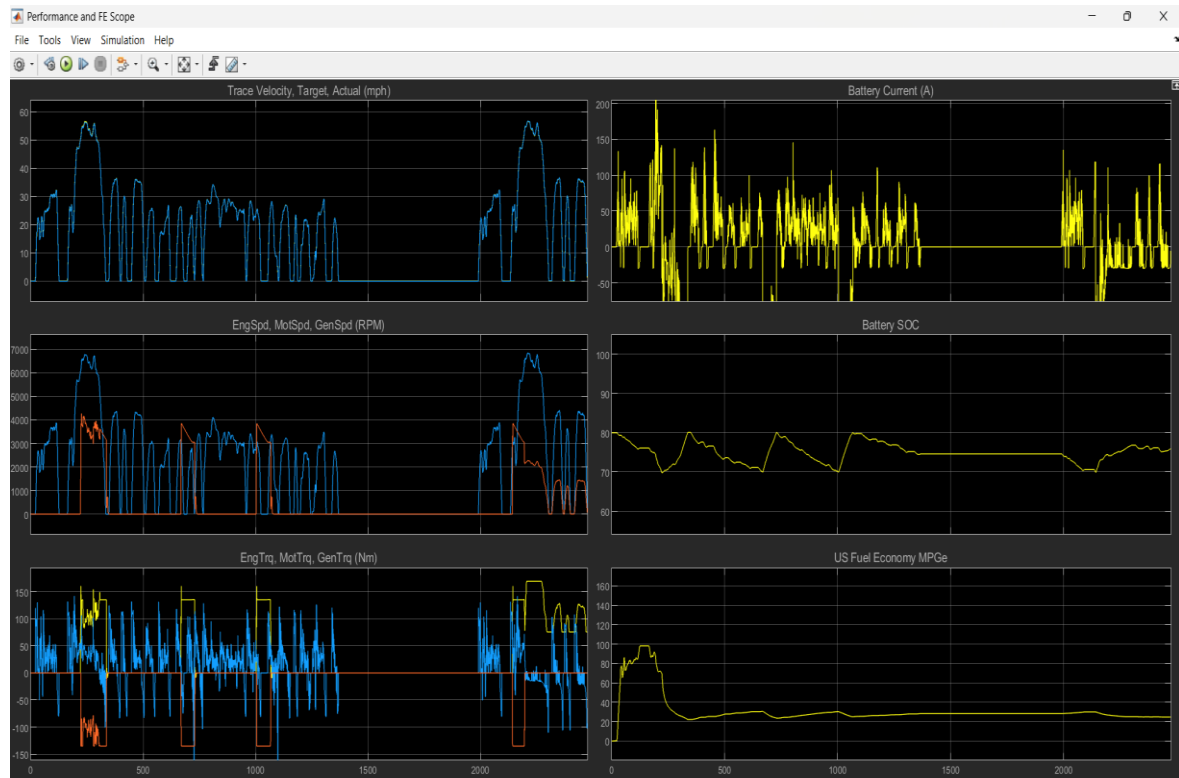


Figure 4: Simulink Results for a HEV System Under Real-Time Conditions.

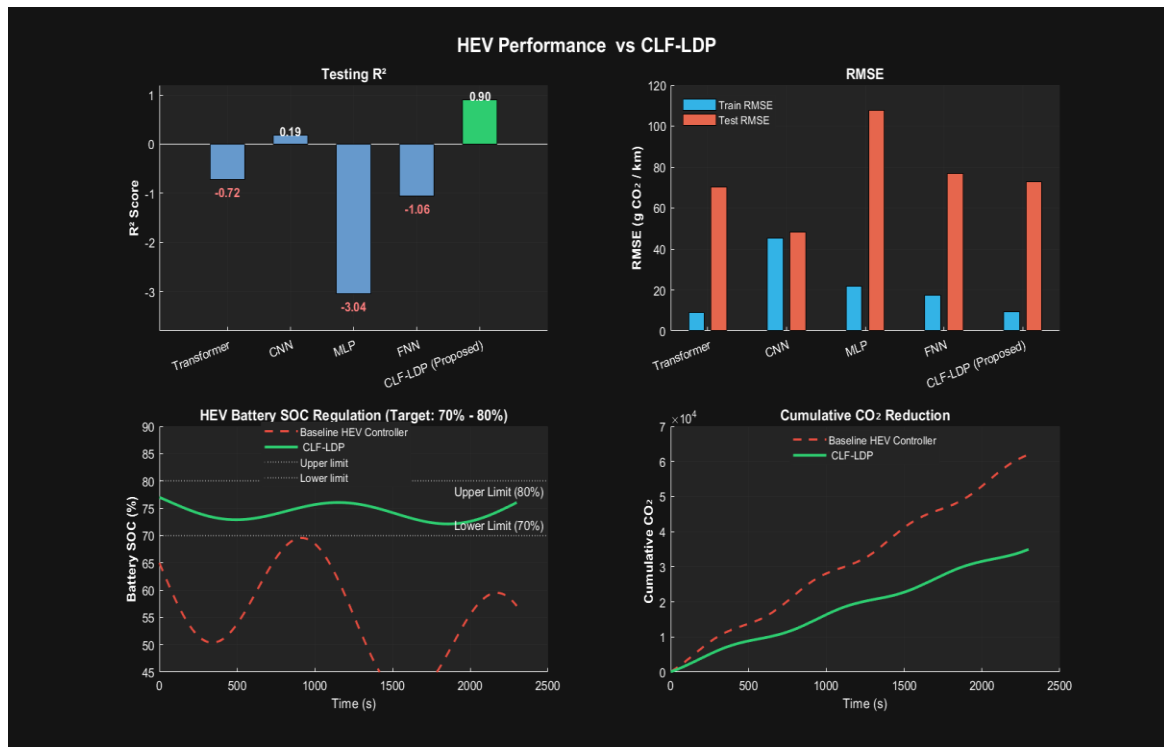


Figure 5: the baseline controller and the proposed CLF-LDP supervisor, highlighting dynamic battery State of Charge (SOC) regulation (70%–80%), current smoothing for CO_2 emission reduction.

Substituting the predictive equation (18) in the control loop makes the emission estimate a proactive adjustment parameter. The regulation of the battery current is then based on predicted emission trends instead of just on

the current state variable. The stability enhancement suggests that there may be a possibility of emission-aware control, as it will be quantified in the future. The stability of SOC behavior and smoother transitions of the battery currents suggest better coordination between predicted emission intensity and energy allocation decision. This integration replaces the rule-based charging approach with a forecasting-based energy management approach. As a result, the proposed framework not only serves as a statistical predictor, but also as a control-enabling intelligence module for real-time emission-aware energy management applications.

5. CONCLUSION

This study addressed CO₂ emission prediction and energy optimization in electric vehicle (EV) networks through a hybrid Crow–Lévy Fuzzy Optimization combined with Linear Discriminant Prediction (CLF-LDP) framework. Unlike purely deep learning–based approaches, the proposed method integrates metaheuristic-driven feature refinement with discriminant projection to enhance predictive stability while maintaining computational efficiency. Comparative evaluation demonstrated that, under weighted loss training, several deep neural architectures exhibited instability on unseen data, reflected in negative R² values. However, the CLF-LDP framework consistently had a good predictive performance, with test R² of 0.90 for the weighted training and test R² of 0.966 for regularization comparison. The results suggest that powerful emission forecasting can be obtained by using simple model structures and optimizing the number of features used. In addition to the statistical validation, the implementation of the predictive equation in the Hybrid Electric Vehicle (HEV) Simulink environment also confirmed its applicability in practice. Predicting emissions is incorporated into the energy management control loop, moving from the static estimation of emissions to anticipatory emission-aware control. The stable behaviour of battery state of charge and the smoother current profiles are indicative of better coordination between prediction and operational control. Overall, the proposed CLF-LDP framework bridges predictive intelligence and real-time energy management, providing a scalable and computationally efficient solution for emission-aware smart transportation systems.

REFERENCES

- [1] Mandalparty, N.S. and Vengadeswaran, S., 2026. An Effective CO₂ Emission Prediction through Ensemble Learning: A Comparative Analysis of Various Models. In *Proceedings of 5th International Conference on Frontiers in Computing and Systems: COMSYS 2024*, Volume 1, vol. 1, Springer Nature.
- [2] Hussein, A.I., Sayed, E.H. and Aly, R.H.M., 2025. CO₂ Emission Predication Based on an Improved Swarm Optimization Technique. In *2025 1st Future International Conference on Artificial Intelligence and Cybersecurity (FICAC)*, IEEE.
- [3] Heshamati, S., Nazari, B. and Nikoo, M.R., 2025. Enhancing accuracy in streamflow prediction under climate change scenarios based on an integrated machine learning–metaheuristic optimization approach. *Journal of Water and Climate Change*, 16(2), p.456.
- [4] Bakır, H., 2024. Prediction of daily global solar radiation in different climatic conditions using metaheuristic search algorithms: a case study from Türkiye. *Environmental Science and Pollution Research*, 31(30), pp.43211–43237.
- [5] Pérez-Rodríguez, S.A. et al., 2024. Metaheuristic algorithms for solar radiation prediction: A systematic analysis. *IEEE Access*, 12, pp.100134–100151.
- [6] Obaideen, K., AlShabi, M. and Bonny, T., 2024. Adapting to climate change: the role of metaheuristic algorithms in optimal decision making. In *Pattern Recognition and Prediction*, vol. 13040, pp.192–199, SPIE.
- [7] Akiner, M.E. and Ghasri, M., 2024. Comparative assessment of deep belief network and hybrid adaptive neuro-fuzzy inference system model based on a meta-heuristic optimization algorithm for precise predictions of the potential evapotranspiration. *Environmental Science and Pollution Research*, 31(30), pp.42719–42749.
- [8] Saqr, A.E., Saraya, M.S. and El-Kenawy, E.M., 2025. Enhancing CO₂ emissions prediction for electric vehicles using Greylag Goose Optimization and machine learning. *Scientific Reports*, 15(1), p.16612.
- [9] Cha, G.W. and Park, C.W., 2025. Development of an optimal machine learning model to predict CO₂ emissions at the building demolition stage. *Buildings*, 15(4), p.526.
- [10] Yu, H. et al., 2026. Deep learning driven spatiotemporal prediction of global carbon emissions from container shipping. *Transportation Research Part D: Transport and Environment*, 151, p.105169.
- [11] de Lima, T.D., Faria, P. and Vale, Z., 2026. Optimal scheduling of home energy systems considering battery aging and CO₂ emissions. *Applied Energy*, 409, p.127433.
- [12] Almihat, M.G.M. and Munda, J.L., 2026. Optimizing battery discharge and charge strategies for enhanced Microgrid performance. *Edelweiss Applied Science and Technology*, 10(2), pp.20–43.
- [13] Zhang, R. et al., n.d. Multi-Time-Scale Adjustable Capability Assessment of Electric Bus Battery Swapping Stations Considering Carbon Emissions. Available at SSRN 6098472.
- [14] Wang, H. et al., 2026. Research on Electric Vehicle Charging Optimization Strategy Based on Improved Crossformer for Carbon Emission Factor Prediction. *Energy Engineering: Journal of the Association of Energy Engineers*, 123(1).

- [15] Selicati, V., Jahanbin, A. and Berardi, U., 2026. Hierarchical learning platform for optimal EV charging scheduling in PV-driven residential buildings. *Energy and Buildings*, p.117191.
- [16] Sicilia, M. et al., 2025. Adaptive Equivalent Consumption Minimization Strategies for Plug-In Hybrid Electric Vehicles: A Review. *Energies*, 18(20), p.5475.
- [17] Castro, L.S., 2025. Machine Learning-Based Equivalent Consumption Minimization Strategy for HEV Torque Split Optimization. Diss. 2025.
- [18] <https://www.kaggle.com/datasets/debajyotipodder/co2-emission-by-vehicles/data>
- [19] Kushwaha, S. et al., 2024. Segmentation of breast cancer from mammogram images using fuzzy clustering approach. In 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), IEEE.
- [20] Şener, İ.F. and Tuğal, İ., 2025. Optimized CNN-LSTM with hybrid metaheuristic approaches for solar radiation forecasting. *Case Studies in Thermal Engineering*, 72, p.106356.
- [21] Thaher, T. et al., 2024. Enhanced variants of crow search algorithm boosted with cooperative based island model for global optimization. *Expert Systems with Applications*, 238, p.121712.
- [22] Khan, A. et al., 2025. Comprehensive review of hybrid energy systems: challenges, applications, and optimization strategies. *Energies*, 18(10), p.2612.
- [23] Tared, S. et al., 2024. Enhancing lung cancer prediction through crow search, artificial bee colony algorithms, and support vector machine. *International Journal of Information Technology*, pp.1-11
- [24] Hussein, A. et al., 2025. Optimizing solar-powered electrolysis systems for green hydrogen production using rat swarm optimization and energy storage solutions.
- [25] Anwar, N., Abdulmajid, M. and Baig, M.S., 2025. Performance Evaluation of Wind-Powered Electrolysis Systems with Energy Storage for Green Hydrogen. In 2025 7th International Conference on Renewable Energy for Developing Countries (REDEC), IEEE.
- [26] Mourya, I., 2025. Integration of Wind and Solar Energy: Challenges and Solutions. In *Transforming Tomorrow: Innovative Solutions and Global Trends in Electrical and Electronics Engineering*, p.20.
- [27] Murtza Qamar, H.G. et al., 2025. Assessment of energy management and power quality improvement of hydrogen based microgrid system through novel PSO-MWWO technique. *Scientific Reports*, 15(1), p.863.
- [28] Iqbal, M.N. et al., 2026. Optimizing the design of hybrid renewable energy systems considering electric vehicle loading. *Electric Power Systems Research*, 252, p.112452.
- [29] Rahlf, H. et al., 2026. Analysis of bidirectional EV charging infrastructures within industrial DC grids. *Smart Energy*, p.100227.
- [30] Savran, E., Karpat, E. and Karpat, F., 2025. Sustainable energy management in electric vehicles through a fuzzy logic-based strategy. *Sustainability*, 17(1), p.89.
- [31] Ramalingam, U., 2026. An interpretable fuzzy reward-based reinforcement learning framework for cost-emission-reliability optimization in smart grids. *International Review of Applied Sciences and Engineering*, pp.1848-2026.
- [32] Sankar, G. et al., n.d. Electric Vehicle Battery Management System: A Comprehensive. *Proceedings Copyright*, 402, p.412.
- [33] Lahri, V., Singh, N.K. and Virmani, N., 2026. Prioritising criteria for sustainable battery supplier selection in the EV sector using fuzzy-ordinal priority approach. *International Journal of Industrial Engineering and Operations Management*, 8(1), pp.66-88.
- [34] Thanasingh, S. et al., n.d. Optimisation-driven technical, economic, and environmental enhancement of renewable and EVCS-integrated faulted Indian distribution networks using DFACTS devices. *Energy Exploration & Exploitation*, p.01445987261425281.
- [35] Bjaoui, M. et al., 2025. Optimized Energy Management Strategy for Non-Autonomous Microgrids Using Crow Search Optimization. In 2025 15th International Renewable Energy Congress (IREC), IEEE.
- [36] Sohrabi, F. et al., 2025. Optimal trading of a charging-station company in auction markets for electricity. *IEEE Transactions on Intelligent Transportation Systems*, 26(5), pp.6545-6555.
- [37] Rahman, T. et al., 2026. Deep learning-based battery health prediction for enhancing electric vehicle performance. *Scientific Reports*.
- [38] Chen, N. et al., 2026. Integrating transformer-based learning and Sentinel-2 bare soil composites for soil organic carbon mapping in the black soil region of Northeast China. *Scientific Reports*.
- [39] Huang, C. et al., 2026. Optimizing hydrogen adsorption capacity in porous carbon through CNN and SHAP analysis. *Cryogenics*, p.104291.
- [40] Park, W., Park, J. and Moon, J., 2026. Micrographic Evaluation of Carbon Nanotube Cement Composites via Machine Learning and Convolutional Neural Networks. *International Journal of Concrete Structures and Materials*, 20(1), p.2.
- [41] Liu, X. et al., 2026. An Intelligent Sensing Platform Integrating Smartphone-based Ratiometric Fluorescence Probe and Feed-forward neural network (FFNN) for pH Detection in Aqueous Samples. *Journal of Environmental Chemical Engineering*, p.121275.
- [42] MathWorks, n.d. Hybrid Electric Vehicle Multimode Reference Application. Available at: <https://ww2.mathworks.cn/help/autoblks/ug/hybrid-electric-vehicle-multimode-reference-application.html>
- [43] Cheng, Y.C., Huang, K.D. and Zhang, J.C., 2023. Design and Simulation of Extended-Range Lithium-Titanate Battery Electric Commercial Vehicle. In *Energy & Propulsion Conference & Exhibition*, SAE Technical Paper.

BIOGRAPHY OF AUTHORS

Rabab Hamed M. Aly, received the B.Sc. of CSE 2010, Minya University, Egypt. She received the M.Sc. and Ph.D. in Electrical Engineering degrees (Computer and System Engineering), Minya University, Egypt, in 2017 and 2019, respectively. She is currently an assistant professor at the higher institute for Management and Information Technology, Minya, Egypt and also an assistant professor in Computer Science at the faculty of computer science at Nahda University, Beni-Suef, Egypt.



Aziza I. Hussein received her Ph.D. degree in Electrical & Computer Engineering from Kansas State University, USA, in 2001 and her M.Sc. and B.Sc. degrees from Assiut University, Egypt, in 1989 and 1983, respectively. She joined Effat University in Saudi Arabia in 2004 and established the first Electrical and Computer Engineering program for women in the country and taught related courses. She was the head of the Electrical and Computer Engineering Department at Effat University from 2007-2010. She was the head of the Computer and Systems Engineering Department, Faculty of Engineering, Minia University, Egypt, from 2011-2016. She was a professor and chair of the Electrical & Computer Engineering Department. Currently, she is a professor and researcher at the same department. Her research interests include microelectronics, analog/digital VLSI system design, RF circuit design, high-speed analog-to-digital converters design, and wireless communications systems design.



Engr. Naveed Anwar is an Electrical Engineer with a Master's degree from the University of Wah, Pakistan, specializing in power system reliability, optimization, and quality enhancement. Currently a Lab Engineer at Effat University in Jeddah, he has substantial experience in developing advanced optimization algorithms to improve power quality and stability, having previously served as a Lecturer at the University of Wah. With several publications to his name, Engr. Anwar's research focuses on machine learning, deep learning, renewable energy, power quality, and smart grid management.