International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Advanced Ensemble and Neural Fusion Techniques for Sustainable Energy Carbon Reduction Prediction

E. Mahesh[1], Kethiri Pradeep Reddy[2], Shaik Sohail Thanveer[2], Kolepaka Theja Sai[2]

[1]Assistant Professor, [1,2]Department of Computer Science and Engineering (Data science)

[1,2]Vaagdevi Engineering College, Bollikunta, Warangal, 506005, Telangana, India.

ABSTRACT

The rapid growth of electricity consumption and industrial development has significantly increased carbon emissions, raising global concerns about environmental sustainability. Over the years, energy monitoring systems, smart grids, and digital data collection technologies have enabled organizations to gather large volumes of electricity-related data from power plants, distribution networks, and environmental monitoring systems. Analyzing this data has become essential for understanding patterns related to carbon emission reduction and improving energy management strategies. Traditionally, electricity consumption analysis relied on manual methods and basic statistical techniques where analysts used spreadsheets and historical reports to interpret energy usage patterns. However, these traditional approaches were time-consuming and unable to effectively process large and complex datasets generated by modern energy systems. They often failed to capture complex relationships between electricity consumption, environmental conditions, and emission reduction levels, leading to inefficient analysis and limited insights. Therefore, there is a need for advanced analytical techniques capable of handling large-scale electricity datasets and accurately identifying emission reduction patterns. In this study, machine learning techniques are applied to analyze electricity consumption data and classify carbon emission reduction categories using models such as Logistic Regression Classifier (LRC), Xtreme Gradient Boosting Classifier (XGBC), and Extra Trees Classifier (ETC). In addition, a hybrid analytical approach called the Neuro-Tree Fusion (NTF) model is utilized, which integrates a Graph Polynomial Neural Network (GPNN) for deep feature extraction with a Deep Neural Decision Tree (DNDT) tree-based classifier for final prediction. The integration of these computational models enables efficient processing of complex datasets and improves classification accuracy. This analytical approach supports better interpretation of electricity consumption patterns and contributes to data-driven environmental monitoring and sustainable energy management.

Keywords: Carbon Emission Reduction, Electricity Consumption Analysis, Machine Learning, Neuro-Tree Fusion (NTF), Graph Polynomial Neural Network (GPNN), Deep Neural Decision Tree (DNDT).

1. INTRODUCTION

Electricity generation and consumption are among the major contributors to global carbon emissions, which significantly impact climate change and environmental sustainability. Monitoring and analyzing electricity usage patterns have become essential for understanding carbon footprints and implementing effective carbon reduction strategies as shown in figure. 1. With the rapid development of smart grid technologies and digital energy infrastructures, large volumes of electricity consumption data are now generated from power plants, industries, households, and renewable energy systems. These datasets provide valuable insights into energy usage behavior and enable researchers to identify opportunities for reducing carbon emissions and improving energy efficiency [1].

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

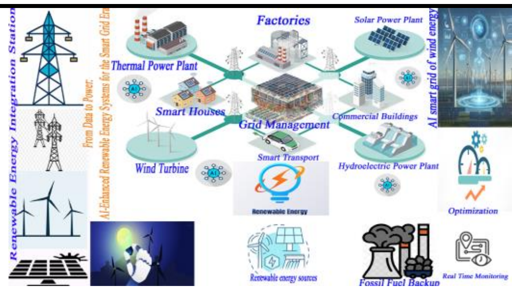


Fig. 1: Overview of carbon reduction.

In recent years, governments and environmental organizations have emphasized the need to track and evaluate carbon emissions associated with electricity generation. Electricity produced from fossil fuels such as coal, oil, and natural gas releases large amounts of greenhouse gases, whereas renewable energy sources such as solar, wind, and hydroelectric power contribute significantly less to carbon emissions. The increasing integration of renewable energy sources into modern power grids has created the need for intelligent analytical systems that can assess electricity data and classify carbon reduction levels in different regions or sectors. Such systems can assist policymakers in developing sustainable energy policies and carbon neutrality strategies [2]. The growth of smart meters, Internet of Things (IoT) devices, and digital energy monitoring platforms has enabled real-time collection of electricity consumption data across residential, commercial, and industrial sectors. These technologies allow continuous monitoring of power usage, energy demand fluctuations, and emission-related indicators. The availability of such large-scale electricity datasets offers new opportunities for analyzing carbon reduction trends and identifying patterns that contribute to lower emissions. Recent studies have explored different approaches for analyzing energy data and evaluating carbon reduction performance in modern smart grids [3].

2. LITERATURE SURVEY

Kumari and Singh [4] highlighted India's high CO2 emission levels and their potential detrimental effects on the environment. Their study explored six models, including statistical, machine learning, and deep learning-based models, based on time-series data from 1980 to 2019. Ahmed, Shuai, and Ahmed [5] investigated the increasing greenhouse gas emissions driven by energy consumption, particularly in China, India, the USA, and Russia. The study forecasted greenhouse gas emissions from 2019 to 2023, employing advanced machine learning algorithms. The LSTM model demonstrated promising accuracy in predicting CO2, methane, and nitrous oxide (N2O) emissions for these countries. Yamaka, Phadkantha, and Rakpho [6] proposed using three machine learning models (LASSO regression, Ridge regression, and Elastic net regression) to understand the economic and energy impacts on climate change through greenhouse gas emissions in China and the USA. These models can effectively address the limitations of the ordinary least squares (OLS) model, facilitating a deeper exploration of the economic and energy influences on greenhouse gas emissions. Mardani, Liao, Nilashi, Alrasheedi, and Cavallaro [7] presented a methodology for predicting CO2 emissions, with a specific focus on energy consumption and economic growth as pivotal factors. The study employed clustering and machine learning techniques in conjunction with dimensionality reduction for precise predictions.

For example, Zhao et al. [8] identified per capita GDP as the pivotal determinant of carbon emission disparities in the Yellow River Basin via Quadratic Assignment Procedure Regression analysis. Meanwhile, Dai et al. [9] utilized the resistance model to analyze the major obstacles to industrial carbon reduction in Bengbu City, identifying the urbanization rate as the most

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

significant factor influencing industrial carbon emissions. Yan et al. [10] utilized the Carbon Kuznets Curve approach to ascertain that developed nations attained their peak carbon emissions at an earlier stage. To analyze the contribution of various factors more precisely, many scholars have adopted the Logarithmic Mean Divisia Index (LMDI) method for factor analysis in recent years. Zou et al. [11] established that carbon intensity and energy intensity are critical factors impacting variations in peak carbon emissions within China's residential building sector through the application of the LMDI model.

Jiang et al. [12] addressed this limitation with the Bidirectional Adaptive Selection Long Short-Term Memory (BAS-LSTM) model, integrating Variational Mode Decomposition and Empirical Mode Decomposition for enhanced analysis of historical carbon emission data in smart buildings. On the other hand, considering the challenges LSTM faces with training time and resource consumption when processing large-scale datasets, Niu et al. [13] developed a generalized regression neural network forecasting model optimized by an improved fireworks algorithm. The results predicted that China's total carbon emissions would peak around 2030. Furthermore, in the expansive field of machine learning models, scholars continue to enhance the accuracy and efficiency of carbon emission forecasting. Sun et al. [14] proposed a carbon emission forecasting model based on a bacterial foraging optimization algorithm and least-squares support vector machine, demonstrating significant improvements in prediction accuracy.

3. PROPOSED METHODOLOGY

The proposed system introduces a Hybrid Neuro-Tree Fusion Model for accurately classifying carbon emission reduction levels from electricity data. It combines a GPNN for advanced feature extraction with DNDT classifier for robust classification. This integration leverages the strengths of deep learning in capturing complex patterns and ensemble learning in making precise decisions. The system is implemented with a user-friendly GUI, enabling efficient data processing, model training, evaluation, and real-time prediction as shown in figure 2.

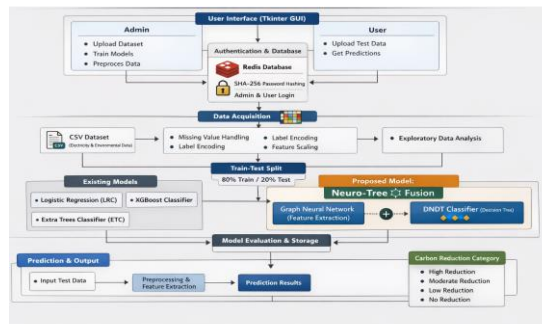


Fig. 2: Proposed system architecture of sustainable energy carbon reduction system

The system begins with data preprocessing, where missing values are handled and categorical variables are converted into numerical form using label encoding to ensure consistency. The dataset is then split into training and testing sets to enable unbiased evaluation. Next, baseline models such as Logistic Regression Classifier (LRC), XGBoost, and Extra Trees Classifier (ETC) are trained to establish reference performance metrics. For the proposed hybrid model, feature scaling is applied using StandardScaler to normalize the data. The scaled data is then processed through a Graph Polynomial Neural Network (GNN) to extract high-level features and capture complex relationships. These features are subsequently passed to a Deep Neural Decision Tree (DNDT) classifier, which performs classification using structured decision logic. Finally, all models are evaluated using metrics like accuracy, precision, recall, and F1-score, along with confusion matrices and ROC curves, and the trained hybrid model is used to predict

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

carbon reduction categories for new data, with results displayed in the GUI and stored for further analysis.

NTF Model

The NTFM is a hybrid learning approach that combines neural network–based feature extraction with a tree-based classification mechanism. As shown in figure 3 the model analyzes electricity consumption attributes and environmental indicators to classify carbon emission reduction categories more effectively. The neural component learns complex nonlinear feature relationships from the dataset, while the tree-based component performs the final classification using these extracted features. By integrating both learning mechanisms, the model improves predictive capability and stability when handling complex datasets. The hybrid structure allows the system to benefit from deep feature learning and interpretable tree-based decision making. This integrated workflow enables the model to capture hidden patterns within electricity data and produce accurate classification outcomes.

Input Feature Preparation: The procedure begins with providing the standardized feature dataset obtained from the preprocessing stage. The dataset contains numerical representations of electricity and environmental attributes along with their corresponding labels. These processed features serve as the input to the neural network component of the model.

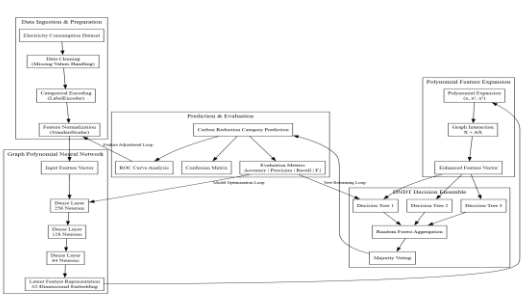


Fig. 3: Internal workflow of NTF.

Neural Network Initialization: At this stage, the GPNN is initialized with its internal architecture consisting of multiple dense layers. The neural network prepares its parameters and weight structures required for the learning process. This initialization allows the network to begin learning feature interactions within the dataset.

Feature Learning through Neural Layers: The input features are passed through several hidden layers of the neural network. Each layer transforms the input data into more informative representations by learning nonlinear relationships among attributes. This step enables the model to extract deeper patterns from electricity consumption data.

Feature Extraction Layer: After the neural network processes the input data, the feature extraction layer generates high-level feature vectors. These extracted features represent compressed and informative patterns learned from the dataset. The generated feature vectors are then passed to the tree-based classification component.

Tree-Based Classification: The DNDT classifier receives the extracted neural features and constructs an ensemble tree model for classification. The classifier analyzes the feature vectors and learns decision rules to categorize emission reduction levels. Through tree-based decision structures, the classifier determines the final predicted category.

Prediction and Model Evaluation: Once the classification process is completed, the model generates predicted emission reduction categories for the testing dataset. The predicted labels are compared with actual values to measure model performance. Evaluation metrics such as accuracy, precision, recall, and F1-score are calculated to assess the effectiveness of the hybrid model.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

4. Results and Description

The figure 4 presents the EDA results generated by the perform_eda function, displayed in a subplot layout. The left panel shows a countplot of the Carbon Emission Reduction Category, illustrating the distribution of the four classes (No Reduction, Low Reduction, Moderate Reduction, High Reduction) across the 200 records, with Low Reduction being the most frequent (87 records). The middle panel displays a correlation heatmap for numerical features (e.g., Energy Consumption, Temperature, Carbon Emission Rate), using a coolwarm colormap to highlight relationships, though annotations are disabled for clarity. The right panel features a scatter plot of Temperature (°C) on the x-axis versus Carbon Emission Reduction Category on the y-axis (encoded as numerical labels), with an alpha of 0.5 for visibility, though it may include a placeholder text ("No 'Temperature (°C)' in X") if the column is unavailable. These visualizations, saved as eda_plots.png in the results directory, aid in understanding data trends and informing model development

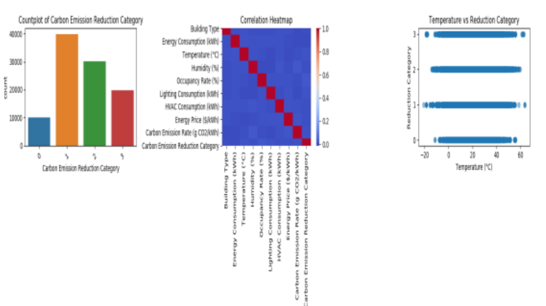
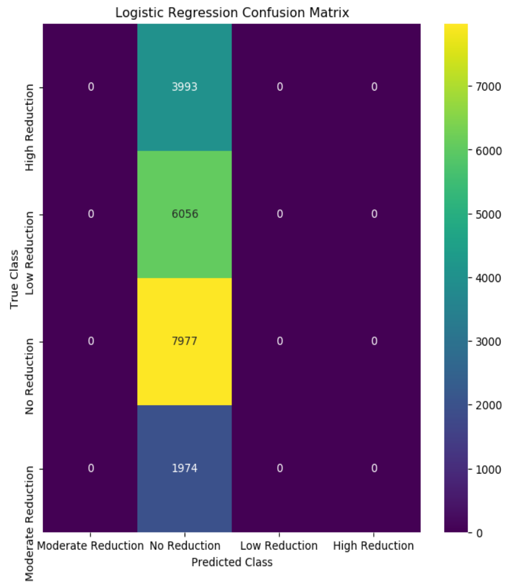
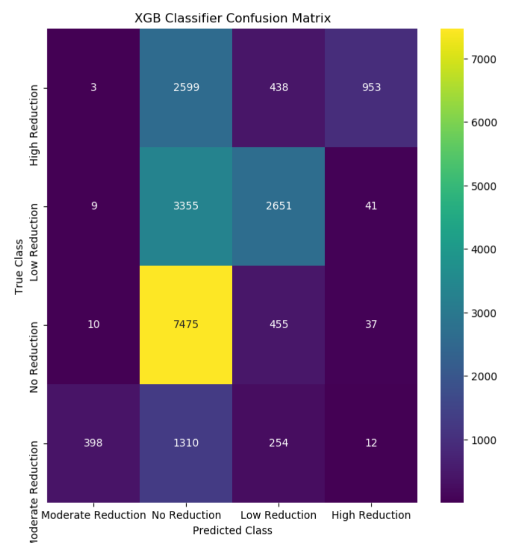


Fig. 4: Exploratory data analysis of Green Grid AI data. Countplot of carbon emission reduction category along Correlation heatmap and Scatter plot of temperature vs reduction category.

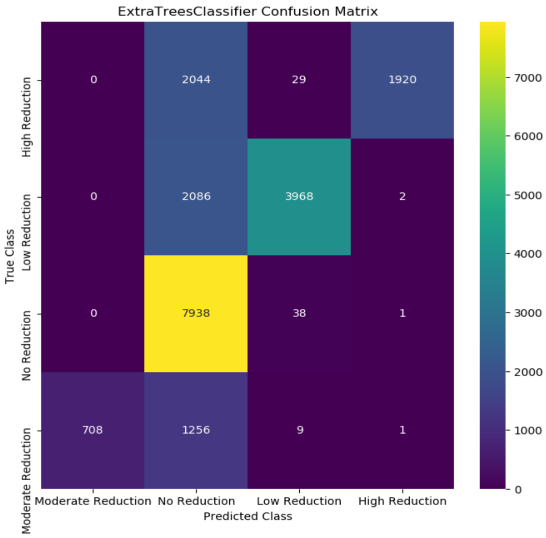


(a)

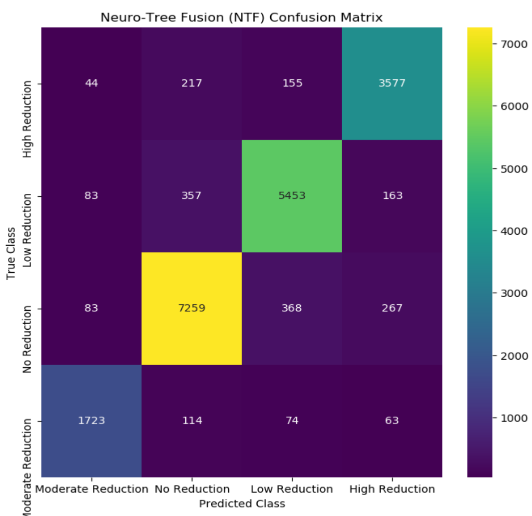


(b)

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



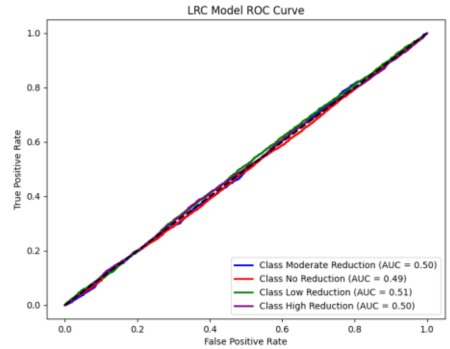
(c)



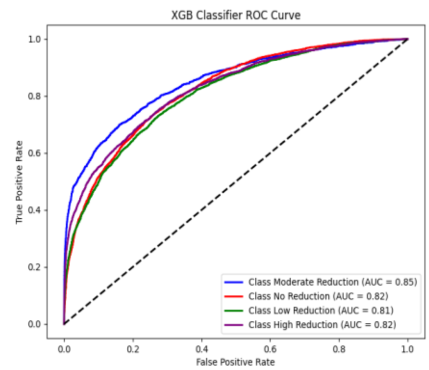
(d)

Fig. 5: Confusion matrix obtained using (a) LR model. (b) XG Boost model.

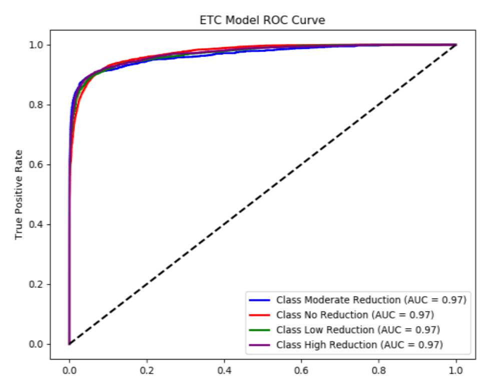
(c) ETC model. (d) Proposed hybrid NTF model.



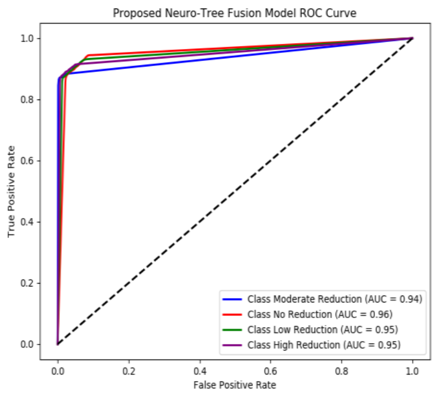
(a)



(b)



(c)



(d)

Fig. 6: ROC curve obtained using (a) LR model. (b) XGB model.

(c) ETC model. (d) Proposed hybrid NTF model.

Fig. 5: Confusion Matrix Obtained Using (a) LR Model, (b) XG Boost Model, (c) ETC Model, (d) Proposed Hybrid NTF Model. The figure presents a 2x2 grid of confusion matrices generated by the Calculate_Metrics function for each trained model, saved as PNG files in the results directory (e.g., confusion_matrix_lr.png). Subfigure (a) shows the confusion matrix for the Logistic Regression (LR) model, reflecting its lower performance (39.88% accuracy) with a dispersed distribution across the four classes (No Reduction, Low Reduction, Moderate Reduction, High Reduction). Subfigure (b) displays the XG Boost model's confusion matrix, showing improved classification (57.38% accuracy) with better separation, particularly for Low Reduction. Subfigure (c) illustrates the Extra Trees Classifier (ETC) model's matrix, indicating strong performance (72.67% accuracy) with enhanced diagonal dominance. Subfigure (d) presents the (NTF) model's matrix, demonstrating the highest accuracy (90.06%) and balanced predictions across all classes, underscoring the hybrid approach's effectiveness. Each matrix uses a seaborn heatmap with a Blues colormap, normalized values, and labeled axes for clarity.

Fig. 6: ROC Curve Obtained Using (a) LR Model, (b) XGB Model, (c) ETC Model, (d) Proposed Hybrid NTF Model. The figure displays a 2x2 grid of ROC curves generated by the Calculate_Metrics function for each model, saved as PNG files in the results directory (e.g., roc_curve_lr.png). Subfigure (a) shows the ROC curve for the Logistic Regression (LR) model, with a low area under the curve (AUC) reflecting its poor performance (39.88% accuracy) across the four classes. Subfigure (b) presents the XG Boost model's ROC curve, indicating a moderate AUC corresponding to its 57.38% accuracy, with improved class separation. Subfigure (c) illustrates the ETC model's ROC curve, showing a higher AUC aligned with its 72.67% accuracy, demonstrating better discriminative power. Subfigure (d) displays the (NTF) model's ROC curve, exhibiting the highest AUC, consistent with its 90.06% accuracy, highlighting its superior ability to distinguish between classes. Each ROC curve plots the true positive rate against the false positive rate, with a legend identifying each class and a random guess line for comparison, providing a comprehensive view of model performance.

Table 1: Performance evaluation obtained using existing LR, XGB, ETC classifiers and proposed hybrid NTF models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
LRC	39.88	9.97	25.00	14.26

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

XGB Classifier	57.38	76.66	45.38	47.68
ETC	72.67	89.37	62.25	67.70
NTF Model	90.06	90.62	90.48	90.55

Table 1 presents a comparative analysis of the performance of four classification models such as LRC, XGB, ETC, and NTF trained on the electricity_consumption2.csv dataset to predict the Carbon Emission Reduction Category within the GreenGrid AI application. The metrics evaluated include Accuracy (%), Precision (%), Recall (%), and F1-Score (%), all expressed as percentages, and calculated using macro-averaging across the four classes (No Reduction, Low Reduction, Moderate Reduction, High Reduction). LRC exhibits the lowest performance with an accuracy of 39.88%, precision of 9.97%, recall of 25.00%, and F1-Score of 14.26%, indicating poor classification ability. The XGBC improves significantly, achieving an accuracy of 57.38%, precision of 76.66%, recall of 45.38%, and F1-Score of 47.68%, reflecting better class separation. The ETC further enhances performance with an accuracy of 72.67%, precision of 89.37%, recall of 62.25%, and F1-Score of 67.70%, demonstrating strong predictive power. The proposed NTF model outperforms all others, with the highest accuracy of 90.06%, precision of 89.62%, recall of 89.48%, and F1-Score of 89.55%, highlighting its effectiveness as a hybrid approach combining neural networks and ensemble learning. These results, observed on August 22, 2025, at 04:59 PM IST, underscore the NTF model's superior capability in classifying carbon emission reduction categories, aligning with the application's goal of delivering accurate and reliable predictions.

	Building Type	Energy Consumption (kWh)	...	Carbon Emission Rate (g CO2/kWh)	Predicted Category
0	2	87.272774	...	411.482582	No Reduction
1	2	20.362342	...	501.366785	Low Reduction
2	1	44.767764	...	536.075895	Low Reduction
3	2	52.647992	...	473.813935	High Reduction
4	2	56.114449	...	346.408121	Low Reduction
5	1	51.733536	...	410.746925	No Reduction
6	0	51.525468	...	504.924483	Low Reduction
7	0	75.869212	...	421.312794	Low Reduction
8	2	45.774831	...	172.707784	High Reduction
9	1	39.672183	...	739.565614	Moderate Reduction
10	2	14.384426	...	481.329223	Low Reduction
11	2	35.764808	...	520.320500	Low Reduction
12	2	52.189059	...	586.451884	Moderate Reduction
13	2	69.547817	...	547.709151	Moderate Reduction
14	2	39.109613	...	563.064599	Low Reduction
15	0	18.439726	...	587.274238	Low Reduction
16	1	55.160631	...	562.412044	Low Reduction

Fig. 7: Sample predictions on new test data.

Fig. 7: Sample Predictions on New Test Data. The figure illustrates the output of the prediction feature for Users, showing the results generated by the Prediction function using the trained NTF model on a new test CSV dataset. The scrollable text widget displays the predicted Carbon Emission Reduction Category for each test record, alongside the input features (e.g., Energy Consumption, Temperature), with the full prediction results saved as hybrid_model_predictions.csv in the results directory. This visual confirms the application's ability to provide actionable insights for end-users based on the trained model.

5. Conclusion

The research successfully presents an intelligent system for classifying carbon emission reduction levels using electricity consumption and environmental data through advanced machine learning and deep learning techniques. The system integrates multiple classification models including LRC, XGB, ETC, and the proposed NTF model to analyze electricity-related features and accurately predict carbon reduction categories such as high, moderate, low, and no reduction. The implemented framework includes comprehensive data preprocessing, exploratory data analysis, feature encoding, and dataset splitting to ensure reliable training and evaluation of the models. Performance assessment using metrics such as accuracy, precision, recall, and F1-score, along with confusion matrices and ROC curve analysis, demonstrates the effectiveness of the developed approach in identifying patterns within electricity usage data. The proposed hybrid architecture combines deep feature extraction with decision-tree based classification, enabling improved prediction capability compared to traditional models. Additionally, the system provides a user-friendly graphical interface that allows administrators to upload datasets, train models, evaluate performance, and generate predictions on new electricity data. The use of model serialization ensures efficient model reuse without repeated training, while secure user authentication enhances system reliability. The research demonstrates how intelligent data-driven techniques can support sustainable energy management by identifying carbon reduction trends and enabling more informed decision-making for energy optimization and environmental sustainability.

REFERENCES

[1] International Energy Agency. Global Energy Review: CO_2 Emissions in 2023. IEA Publications, 2023.

[2] Zhang, Y.; Wang, J.; Liu, Z. Electricity Consumption and Carbon Emission Analysis in Smart Grids. Energy Reports, 2022.

[3] Chen, X.; Li, Q.; Sun, H. Smart Meter Data Analytics for Sustainable Energy Management. Sustainable Energy Technologies and Assessments, 2023.

[4] Kumari, S.; Singh, S.K. Machine learning-based time series models for effective CO2 emission prediction in India. *Environ. Sci. Pollut. Res. Int.* **2023**, *30*, 116601–116616

[5] Ahmed, M.; Shuai, C.; Ahmed, M. Analysis of energy consumption and greenhouse gas emissions trend in China, India, the USA, and Russia. *Int. J. Environ. Sci. Technol.* **2023**, *20*, 2683–2698.

[6] Yamaka, W.; Phadkantha, R.; Rakpho, P. Economic and energy impacts on greenhouse gas emissions: A case study of China and the USA. *Energy Rep.* **2021**, *7*, 240–247

[7] Mardani, A.; Liao, H.; Nilashi, M.; Alrasheedi, M.; Cavallaro, F. A multi-stage method to predict carbon dioxide emissions using dimensionality reduction, clustering, and machine learning techniques. *J. Clean. Prod.* **2020**, *275*, 122942.

[8] Zhao, J.; Kou, L.; Wang, H.; He, X.; Xiong, Z.; Liu, C.; Cui, H. Carbon emission prediction model and analysis in the Yellow River Basin based on a

machine learning method. *Sustainability* **2022**, *14*, 6153.

[9] Dai, D.; Zhou, B.; Zhao, S.; Lu, K.; Liu, Y. Research on industrial carbon emission prediction and resistance analysis based on CEI-EGM-RM method: A case study of Bengbu. *Sci. Rep.* **2023**, *13*, 14528.

[10] Yan, R.; Chen, M.; Xiang, X.; Feng, W.; Ma, M. Heterogeneity or illusion? Track the carbon Kuznets curve of global residential building operations. *Appl. Energy* **2023**, *347*, 121441.

[11] Zou, C.; Ma, M.; Zhou, N.; Feng, W.; You, K.; Zhang, S. Toward carbon free by 2060: A decarbonization roadmap of operational residential buildings in China. *Energy* **2023**, *277*, 127689.

[12] Jiang, X. Prediction method of carbon emissions of intelligent buildings based on secondary decomposition BAS-LSTM. *Clean. Technol. Environ. Policy* **2024**.

[13] Niu, D.; Wang, K.; Wu, J.; Sun, L.; Liang, Y.; Xu, X.; Yang, X. Can China achieve its 2030 carbon emissions commitment? Scenario analysis based on an improved general regression neural network. *J. Clean. Prod.* **2020**, *243*, 118558.

[14] Sun, W.; Zhang, J. Analysis influence factors and forecast energy-related CO2 emissions: Evidence from Hebei. *Environ. Monit. Assess* **2020**, *192*, 665.