International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4583 - 4590
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0601

Advanced Carbon Footprint Prediction Using Hybrid Machine Learning and Ai-Assisted Recommendations

Inchara R[1], Dr. Madhu M Nayak[2]
[1]PG Scholar, Dept. of CSE, GSSS Institute of Engg. & Tech. for Women, Mysuru, Karnataka, India
[2]Associate Professor, Dept. of CSE, GSSS Institute of Engg. & Tech. for Women, Mysuru, Karnataka, India
Emails: *incharar20@gmail.com[1], madhu.m@gsss.edu.in[2]*

Abstract
Rising levels of carbon emissions have emerged as a key factor to climate change requiring smart mechanisms of monitoring and mitigation. In this paper, CarbonIQ, a machine learning-based, generative AI-based, and IoT-based data collection integrated carbon footprint prediction and recommendation system will be introduced. The system uses user activity logs and real-time sensor values received by an ESP32 which has been pre-configured with the sensor tags. Gassed to the carbon emission estimation and prediction is connected to gas sensors. The proposed approach uses the Random Forest regression model with a hybrid sensor-fusion mechanism. Besides prediction, a generative AI module allows giving personalized carbon reduction recommendations based on users' actions and their past actions. This system has also been optimized with ranking-based feedback, which will increase the engagement with the system and encourage users to move forward. sustainable practices. That the proposed strategy is useful can be validated with experimental results that demonstrate important outcomes in emissions prediction and taking actionable steps. Incorporating prediction and recommendation into a single structure, one would be able to make better decision. and develop more data- driven solutions to sustainability.

Keywords: *Carbon Footprint Prediction; Generative AI; IoT; Machine Learning; MQTT; Random Forest; Sensor Fusion.*

1. Introduction

An increase in Green House Gases (GHGs), particularly CO_2 (Carbon Di Oxide), is a major contributor Efficient measurements and mitigation efforts are needed in facing the adverse effects of climate change. Due to the growths of urbanization, industrialization and changes in lifestyle, people's carbon of an individual has been growing. footprints. The performance of each activity has a substantial contribution to global carbon footprint, thus individual-level carbon footprint measure. Estimations are important to giving a push to sustainability. Current approaches to carbon footprints calculations are mostly based on static/rules based approaches, and the following problems arise with these approaches: offer approximate outcomes. These are not scalable and fail to take into account temporal effects, or take historical data into account for forecasting. Moreover, they offer few opportunities for behavior change as they don't have adaptive and personalized. feedback. In computational efficiency, they do not utilize any data-driven models which can exploit the nonlinear Relationships in the lifestyles attributes and emissions. In recent years, the development of machine learning (ML) technologies has led to the emergence of smart systems to precisely estimate and predict carbon emissions based on behavioral and environmental information. Moreover, the incorporation of generative AI enables translation of the prediction results to practical recommendations to reduce emissions. However, existing systems often focus on prediction and/or monitoring but not provide an integrated approach that Data collection, prediction & intelligent recommendations. To overcome these challenges, this study introduces an integrated AI-based system for carbon footprint prediction, and intelligent recommendation, which incorporates machine learning models and generative AI recommendation. predicts current emissions and prediction of future ones for the system using data inputted by the user and real-time sensor data. future carbon footprints. The model adopts an ensemble based regression approach in combination with real-time sensor analysis for higher operating efficiency of prediction. While providing personalized recommendations on how to reduce carbon emissions using generative AI, forecasting remains one of its key functions. (based on predictions and past records). Main achievements resulting from this research are:

- A new approach to personalized modelling of carbon emissions that incorporates sensor data and user inputs based on features was introduced. A feature-based approach to personalized modelling of carbon emissions was introduced, with integration of sensor data and user inputs.
- The proposed model forecast the carbon emission according to machine learning prediction and IoT based on environmental sensing.
- An AI-based recommendation engine that is integrated and drives carbon reduction.
- A comprehensive data collection, forecasting and AI-driven decision making system.

2. Literature Survey

2.1. Carbon Footprint Estimation Techniques

Tables should Carbon footprint estimation has progressed from conventional analytical approaches to advanced data-driven systems. Early methods relied on predefined parameters such as energy consumption and fuel usage, while recent approaches integrate IoT and machine learning for improved accuracy. An IoT-enabled framework demonstrates precise estimation of both operational and embodied emissions in IoT-based systems [1]. Similarly, carbon footprint monitoring systems utilize machine learning and deep learning methods to enable continuous tracking and analysis of emissions [2]. In addition, data-driven models incorporating lifestyle parameters such as transportation, energy consumption, and waste generation provide highly accurate predictions along with interpretability using explainable AI techniques [3]. Domain-specific approaches, such as CO_2 estimation in diesel engines, utilize parameters like engine load and fuel consumption to generate accurate predictions [4]. Other studies estimate emissions from industrial energy usage using machine learning models such as Random Forest and SVM [5]. These works highlight a shift toward dynamic and intelligent estimation systems; however, most are limited to specific datasets or domains and lack a unified estimation framework.

2.2. Machine Learning- Based Carbon Prediction

Machine learning algorithms have been extensively used for carbon emission prediction due to their ability to model complex nonlinear relationships. Comparative studies show that there were models that performed better than traditional statistical models, such as that of machine learning and deep learning in predicting CO_2 emissions, particularly when evaluated using metrics such as RMSE and R^2 [15]. Ensemble methods, including Random Forest and Gradient Boosting, further enhance prediction performance. Hybrid approaches combining statistical and deep learning models have also demonstrated improved accuracy. For instance, hybrid ARIMA–Transformer models applied to IoT-based multivariate datasets provide superior forecasting performance in smart city environments [16]. Similarly, a two-stage machine learning framework integrating models such as SVR, ANN, and Random Forest significantly reduces prediction errors compared to single-stage approaches [6]. Application-specific studies further validate the effectiveness of machine learning. These include emission prediction in diesel engines [4], industrial emission modeling [5], and vehicle emission prediction using ensemble learning techniques [13]. Further, machine learning models have been applied to moderate emissions in smart cities by providing predictive insights for policy-making [14]. In spite of

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4583 - 4590
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0601

these enhancements, many models require large datasets and computational resources, limiting their deployment in real-time IoT environments.

2.3.AI-Based Recommendation and Behavioral Nudging Systems

While prediction models are well-developed, fewer studies focus on influencing user behavior. AI-driven recommendation systems have been introduced to bridge this gap by combining carbon footprint estimation with behavioral nudging. One such system utilizes a mobile application that analyzes user activities and provides personalized recommendations, along with chatbot support and reward-based engagement mechanisms [7]. These systems emphasize real-time feedback and user engagement to promote sustainable behavior. However, most existing approaches operate independently of real-time IoT data and predictive models. This lack of integration limits their ability to provide dynamic and context-aware recommendations, highlighting the necessity for systems that merge prediction, monitoring, and behavioral guidance.

2.4.IoT-Based Environmental Monitoring Systems

IoT technologies play a vital role in real-time environmental monitoring and data acquisition. Various researches have proposed cost-effective IoT-based systems for tracking air quality and carbon emissions. For example, sensor-based systems using microcontrollers and gas sensors enable real-time measurement and cloud-based visualization of environmental parameters [8]. Similarly, low-cost IoT mesh networks provide scalable and energy-efficient resolutions for indoor air quality monitoring [9]. Integration of IoT with machine learning has further improved system capabilities. Transfer learning-enabled IoT systems enable continuous CO_2 prediction on embedded devices using edge computing techniques [10]. Additionally, IoT-based climate prediction systems using LSTM models demonstrate the efficacy of combining sensor data with prognostic insights [11]. Survey studies also emphasize the growing importance of amalgamating IoT with machine learning for environmental monitoring and prediction [12]. Regardless of these breakthroughs, obstacles such as scalability, energy efficiency, and seamless integration with predictive models remain unresolved.

2.5.Research Gaps and Problem Motivation

Even though notable growth has been made in carbon footprint estimation, machine learning prediction, and IoT-based monitoring, several key challenges remain. First, many systems focus either on prediction or monitoring, with limited integration between real-time IoT data and advanced machine learning models [10]. Second, most ML models rely on static datasets, reducing their effectiveness in dynamic real-time environments [15]. Third, existing studies are often domain-specific, focusing on vehicles, industrial emissions, or smart cities, without providing a cohesive bodywork that integrates innumerable emission origins such as lifestyle, energy depletion, and environmental data [3]. Additionally, AI-based recommendation systems are rarely integrated with real-time predictive models, limiting their practical impact [7]. Finally, IoT-based systems face challenges related to scalability, cost, and energy efficiency, particularly in large-scale deployments [9]. These limitations highlight the requirement for an integrated system that unites real-time IoT data acquisition, efficient machine learning models, and intelligent recommendation mechanisms. Therefore, this task is inspired by the need to develop a unified framework that integrates carbon footprint estimation, real-time prediction, and personalized recommendation systems within an IoT-based architecture.

3. Results Methodlogy

3.1.System Architecture

In the System Architecture(Figure 1) for the inputs there are 2 layers. One is the user input layer and other is the IoT sensor layer. The user input layers has many fields such as grocery expenses, vehicle travel distance, waste generation, electronic usage hours, clothing purchases, and internet usage that mainly reflects how much a person is contributing for the carbon emission. ESP32 microcontroller has MQ3, MQ5, and MQ7 sensors. These are for monitoring gas concentration in the vehicle exhaust. These sensor values are transmitted via MQTT in real time.As shown in FIGURE 1. System Architecture.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4583 - 4590
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0601

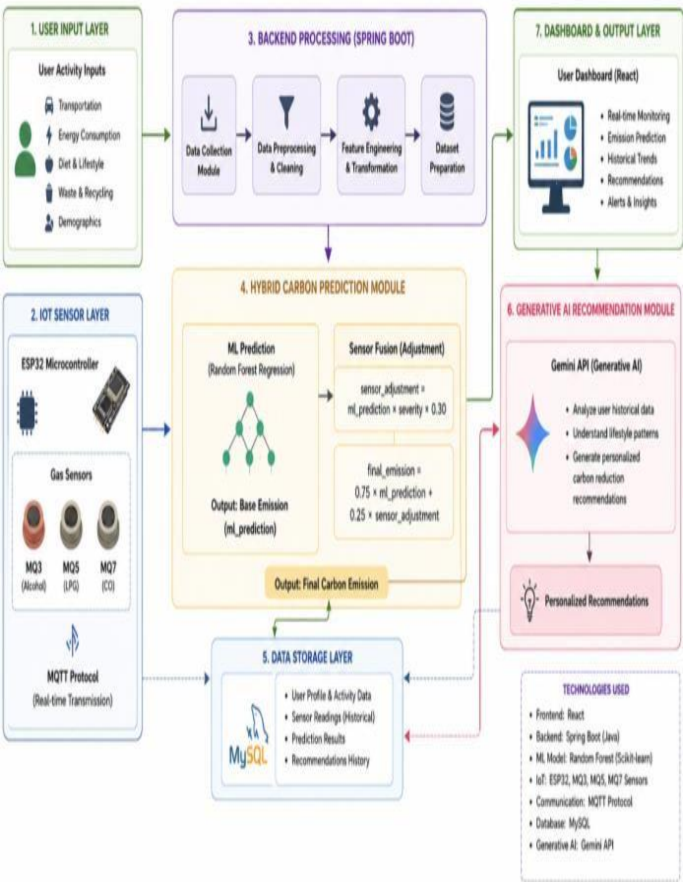

Figure 1 System Architecture

Data collection, pre-processing, handling missing data, normalization, encoding categorical data and one-hot encoding are done in the backend processing layer. Preparing dataset for communicating with machine learning prediction module. The hybrid model used for carbon emission forecasting is Random Forest regression along with sensor severity scoring and hybrid fusion mechanism. The data storage layer is used for storing user profiles, activity data, sensor data, prediction results, historical carbon data, recommendation data for retrieval and monitoring on the dashboard. The dashboard and output layer is presented with real time monitoring data, interactive visualization of emissions, historic trends, and1 future emissions predictions, to enable a user to analyze and make decisions. The generative AI recommendation module processes the historical emissions data and the predictions allied to it to render personalized tips for lowering CO2.

3.2.Data Collection and Preprocessing

The existing dataset is pre-processed to improve consistency and data quality before model training. Due to inconsistent usage of user or not using any particular inputs, for e.g. the person didn't even travel for a particular day, in such cases there can be missing user inputs and sensor values. These are handled by imputation technique whereas categorization attributes are converted using One Hot Encoding. After all these processing, Feature Extraction and Feature Integration are performed before being fed to hybrid model. If all 3 MQ sensors are in range, i.e., $90 \leq MQ3$ value≤ 160, $3900 \leq MQ5$ value≤ 4500 and $200 \leq MQ7$ value≤ 250, then sensor boost factor is applied which is 10% by using AND gate as shown in Figure 2 Sensor Boost Decision Logic.

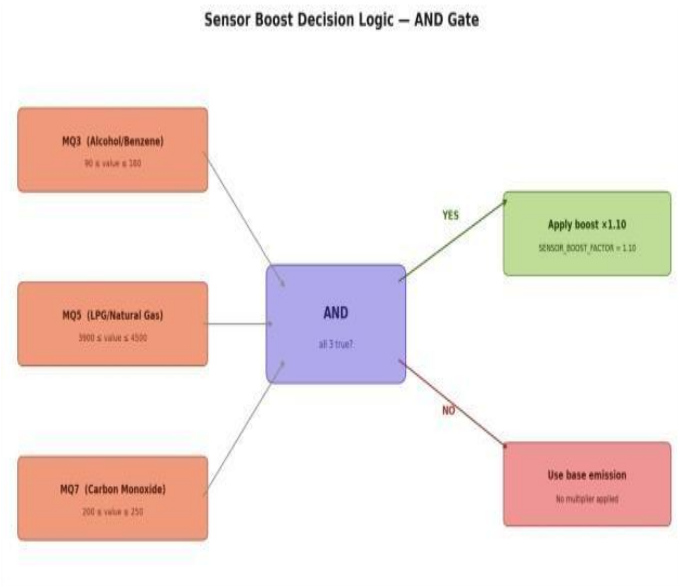

Figure 2 Sensor Boost Decision Logic

The collected and processed data are stored in the MySQL database for dashboard visualization, prediction and historical analysis.

3.3.Machine Learning Model

The proposed processing system is based on a hybrid machine learning approach, which combines the Random Forest (RF) regression model, a rule-based sensor scoring mechanism and a weighted fusion scheme for carbon emission prediction. Random Forest Regressor makes use of 200 decision trees (Figure 3) to know all the lifestyle related features

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4583 - 4590
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0601

like transportation pattern, energy consumption, diet, and travel habits. The predictions by each of the individual trees are averaged to give the machine learning output, which allows this to avoid the overfitting problem and means that it performs very well for the nonlinear relationships.As shown in Figure 3 Code snippet for Random Forest Regressor Model, Figure 4 Sensor Severity.

```
CREATE RandomForestRegressor model
    - n_estimators = 200
    - random_state = 42
    - n_jobs = -1

CREATE pipeline
    Step 1: preprocessing
    Step 2: model training
```

Figure 3 Code snippet for Random Forest Regressor Model

For the incorporation of the real-time environmental conditions, the sensor readings from MQ3, MQ5 and MQ7 are processed by a graduated severity scoring procedure. The value of each sensor is scaled to a range between 0 and 1 by employing a scoring function based on clamps, with the scale of 0 minimizing pollution and 1 the worst pollution (Figure 4). The scores from individual sensors are added and a combined environmental severity index is calculated. Carbon emission can be calculated by a hybrid sensor fusion model which is weighed, in the way as follows:

$$sensor_{\{adjustment\}} = ml_{\{prediction\}} \times severity \times 0.30$$

$$final_{\{emission\}} = 0.75 \times ml_{\{prediction\}} + 0.25 \times sensor_{\{adjustment\}}$$

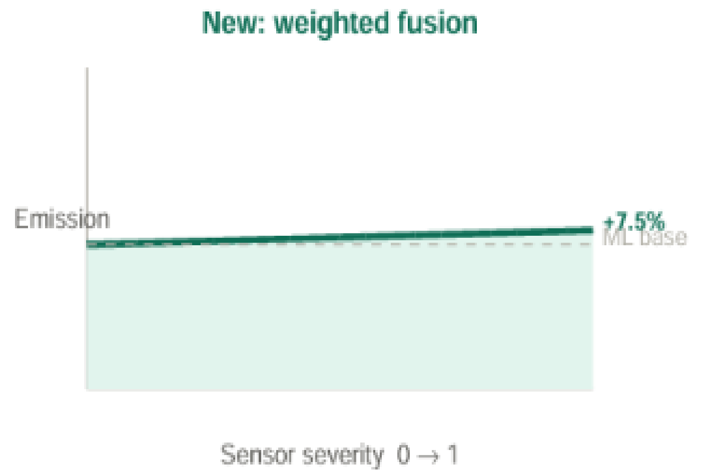

Figure 4 Sensor Severity

The machine learning module accounts for 75% of the prediction, while the sensor fusion based adjustment kicks in to refine the prediction using the sensor data obtained in real-time. The fusion approach allows for avoiding over dependence on the temporal dynamics of the sensor measurements while taking the environmental funnel effect into account in the final prediction.

3.4.GenAI-Based Recommendation System

The proposed system integrates a Generative AI (GenAI)-based module to deliver personalized carbon reduction recommendations. This module relies on "improvisation and rules" methodology, with rules that relate to the prompt to generate contextually relevant Proposals sourced from data on the individual user's emissions. The historical carbon records enable analyses of high emission factors and the recommendations are created based on these analyses. This includes things such as transportation, electricity consumption and amount of travelling. These results, combined with the machine learning predicted carbon footprint in the next month, are then posted. The structured prompts are then embedded in it and can be passed to a GenAI model through API to generate human-readable. suggestions. Before generating and after generating, rules are applied to pre and post filter. Pre- and post-processing accentuates the emission contributors and outputs to the sustainability. guidelines. Moreover, by adding prediction results you can make recommendations ahead of time, thus saving the user the trouble of reducing. to not only just respond to historical data, but also forecast future

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4583 - 4590
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0601

emissions. To create and encourage sustainable behavior, the GenAI considers cumulative emission data and provide the recommendation.

4. Implementation Details

The proposed system is implemented with full stack front end and backend, machine learning, database and IoT. The UI is created with React, which enables the user to input data on to the platform, and visualizes the carbon emissions data through the dashboard. The backend services that include user sign-ups/ins, data validation, communications with APIs, interactions with other pieces of the system, etc. are developed using Spring Boot. Data on users' activity, measurements taken by various sensors, historical emissions record and prediction outputs are stored in the MySQL Database for monitoring and analysis. The machine learning model is written in Python programming language with random forest regression model using a hybrid sensor fusion mechanism to predict the carbon emission. The backend forwards the data received from the user interface to the prediction module, and returns the prediction results from the prediction module to the UI. The ESP32 Module (Figure 5) is programmed and flashed with the Arduino IDE to acquire sensor data from the board and communicate via MQTT.

Figure 5 **Hardware Setup of ESP32**

Random Forest regression model is serialized using Joblib for training and 80:20 train-test split is utilized for the evaluation of this model and deployed into the prediction API embedded within the backend platform for real-time carbon emissions prediction. Furthermore, a Generative AI recommendation module is integrated using the Gemini API to process prediction outputs and historical user records for recommendation generation.

5. Results And Discussion

The results recovered (Table 1) show that the proposed hybrid approach also guarantees stability and reliability in the prediction capacity, while maintaining good generalization ability. The model is well validated with the R2 score reaching 0.918, capturing the model's capability in learning the relationship between user activities and carbon emission patterns, and the cross-validation score assures a less overfitted model and greater stability in the process of multiple trainings. The mean absolute error (MAE) around 420kg CO_2e/year shows a reasonable accuracy in the predictions of carbon emissions for practical application. The suggested model has a high 5-Fold Cross Validation R^2 of 0.913 for predicting carbon emission. Table 1. Performance metrics of proposed hybrid models, Figure 6 Carbon emission monitoring and prediction dashboard

Table 1 **Performance metrics of proposed hybrid models**

Metric	Value
Validation R^2 Score	0.918
5-Fold CV R^2 Score	0.913
MAE	≈420 kg CO_2e/year

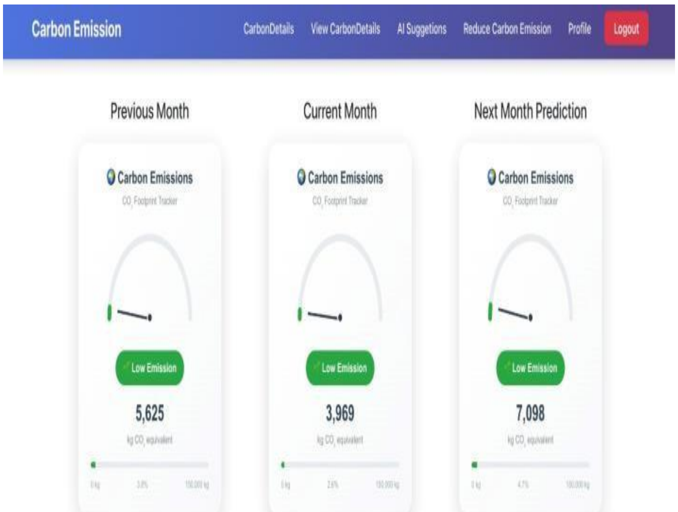

Figure 6 **Carbon emission monitoring and prediction dashboard**

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4583 - 4590
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0601

Figure 6 shows the system dashboard with the display of previous, current and predicted carbon emission. The visualization also allows the user to comprehend the emission trend across time and evaluate the effect they have on the environment.

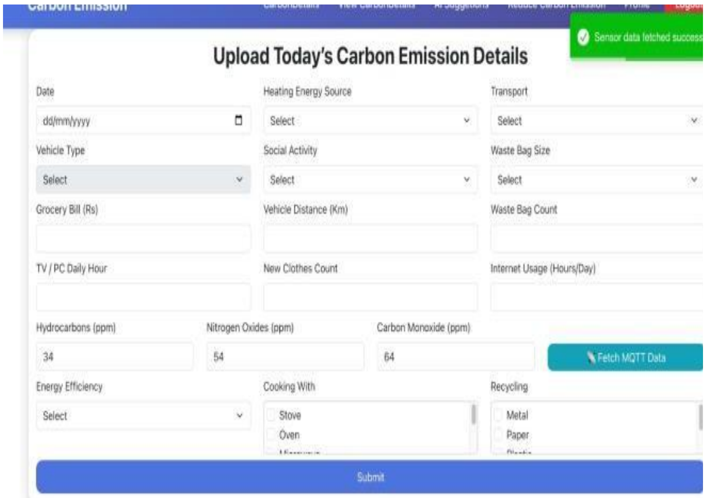

Figure 7 Real-time IoT sensor data collection and user input interface

Figure 7 depicts the user input interface implemented on-top of real-time sensor data capture, here gas sensor values are being successfully sent via MQTT protocol. This shows the successful application of the IoT-based data collection in the software system.

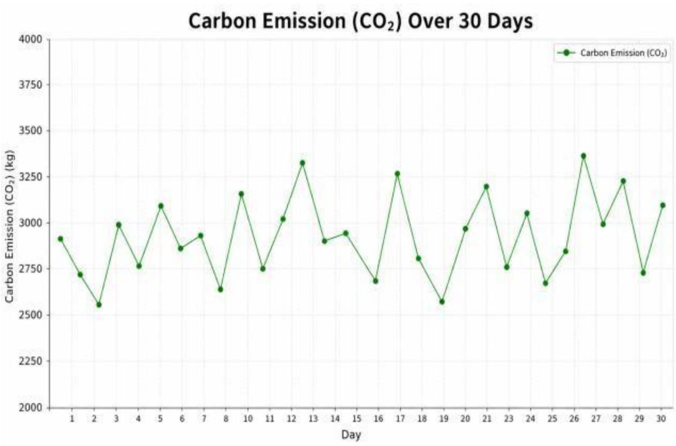

Figure 8 Carbon Emission for the period of 30days

Furthermore, the system has a Generative AI recommendation feature that offers users customized suggestions for reducing carbon emissions that are derived from the prediction results of user historical data.

Conclusion And Future Works

In this paper, an AI-based carbon footprint prediction and recommendation system that combines the advantages of using IoT sensor data, machine learning prediction and Generative AI for recommendation generation is proposed. The proposed framework fuses user activity inputs with real-time measurements from the sensors like MQ3, MQ5 and MQ7, which measure the CO2 concentrations and feed inputs to the hybrid prediction model based on Random Forest. The algorithms of pre-processing and sensor fusion are included to enhance reliability of prediction in various environments. The system proved to be capable of a 5-Fold Cross Validation R^2 of 0.913 and offers intelligent carbon monitoring, prediction and personalized sustainability recommendations. Potential future work includes further incorporation of sensors, cloud based monitoring, and further development of models based on deep learning, enhancing the predictive capability and scalability of the system. The system can be extended to mobile applications and smart city environmental management platforms to support large-scale sustainability management. More can be added to this, such as implementing carbon emission ranking systems to compare and rank users according to their lower carbon emission levels and carbon reduction performance. There is also scope to merge adaptive recommendation systems and real-time behavior analysis into the system for better personalizing carbon reduction recommendations and user involvement.

References

[1]. "IoTCO2: End-to-End Carbon Footprint Assessment of IoT-Enabled Deep Learning Systems," IEEE, 2023.

[2]. "Carbon Footprint Monitoring System Using Machine Learning and Deep Learning Techniques," IEEE Access, 2022.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4583 - 4590
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0601

[3]. "A Multi-Model Approach to Carbon Footprint Prediction Using Explainable AI," IEEE Trans. Sustain. Comput., 2023.
[4]. "CO2 Emissions Prediction Using Machine Learning in Diesel Products," Int. J. Adv. Res. Comput. Commun. Eng., 2021.
[5]. "Predicting CO2 Emission Using Machine Learning," Int. J. Res. Eng.Appl. Manage., 2021.
[6]. C. Wang, M. Li, and J. Yan, "Forecasting CO2 Emissions Using a Two-Stage Machine Learning Model," J. Water Climate Change, vol. 14, no. 2,pp. 456–468, 2023.
[7]. A. J. Jasmy, H. Ismail, and N. Aljneibi, "AI-Driven Carbon Footprint Assessment and Recommendation System for Sustainable Behavior," Discover Sustainability, 2024.
[8]. "Designing a Cost-Effective IoT Air Quality Sensor for Real-Time Monitoring," Proc. IEEE Conf., 2022.
[9]. F. Gül and H. Eroğlu, "Low-Cost IoT Mesh Network for Real-Time Indoor Air Quality Monitoring," IEEE Sensors J., 2021.
[10]. M. AlSelek et al., "Transfer Learning-Enabled IoT System for Continuous CO2 Prediction in Vehicles," IEEE Internet Things J., 2022.
[11]. "IoT-Based Climate Prediction Using Long Short-Term Memory (LSTM)," IEEE Access, 2022.
[12]. S. B. Kasetty and S. Nagini, "IoT-Based Machine Learning Model for Air Pollution Prediction: A Survey," Proc. IEEE Conf., 2021.
[13]. H. Meshref et al., "Towards NEOM: Carbon Emission Prediction Using Machine Learning," IEEE Access, 2024.
[14]. A. H. Alshanbari et al., "Machine Learning Techniques for CO2 Emission Reduction in Smart Cities," IEEE Access, 2023.
[15].
[16]. "Comparative Analysis of Machine Learning, Deep Learning, and Statistical Models for CO2 Prediction," Sustainable Computing, 2022.
[17]. "Hybrid ARIMA–Temporal Fusion Transformer Model for CO2 Forecasting Using IoT Data," IEEE Access, 2023.