LiveLight CarbonFootprint: An AI-Enabled Machine Learning Framework for Personalized Emission Awareness and Sustainability Guidance

Vitesh Babu M [1], Vinod G[2], Vikas Gowda[3], Sathvik M[4], Shubham Deep[5]
[1]*Assistant Professor, Department of ISE, Sir M. Visvesvaraya Institute of Technology*
[2,3,4,5] *Information Science and Engineering, Sir M. Visvesvaraya Institute of Technology*

Abstract- Escalating greenhouse gas emissions and the rapid growth of global carbon footprints have intensified the need for accessible, data-driven tools that help individuals understand and reduce their environmental impact. This work presents LiveLight CarbonFootprint, a web-based system that leverages supervised machine learning, interactive visualization, and an AI-driven chatbot to estimate monthly personal carbon emissions from lifestyle data such as transportation, energy usage, diet, and waste habits. The framework employs an MLPRegressor model with standardized input features to generate personalized carbon footprint predictions, complemented by category-wise breakdowns, actionable recommendations, and downloadable PDF reports. The system further integrates a generative AI assistant to offer real-time sustainability advice, enhancing user engagement and awareness. Experimental evaluation demonstrates accurate predictions, low response times, and positive user acceptance, indicating that the proposed solution can effectively bridge the gap between high-level climate goals and everyday behavioral change.

Keywords: Carbon Footprint, Machine Learning, Sustainability, Streamlit, Generative AI, MLPRegressor, Environmental Informatics.

I. INTRODUCTION

The rapid increase in anthropogenic carbon emissions, exceeding 40 billion tons annually, has become a critical driver of climate change and environmental degradation. While industrial and national-level emissions receive significant attention, individual lifestyle choices—spanning transport modes, energy consumption patterns, dietary habits, and waste management—also contribute substantially to the overall carbon footprint. However, most individuals lack intuitive, data-driven tools to quantify their personal emissions and translate insights into practical behavioral change.

To address this gap, the *LiveLight CarbonFootprint* system is designed as an interactive web application that enables users to estimate their monthly carbon footprint using a machine learning model trained on lifestyle-related attributes. The platform combines a supervised MLPRegressor model, feature scaling with StandardScaler, visual analytics, and an AI-powered chatbot to provide both quantitative estimates and qualitative sustainability guidance. By fusing data science, machine learning, and modern web technologies, the system aims to make carbon awareness more accessible, engaging, and personalized.

1.1 Problem Statement

Despite growing awareness of climate change, most existing carbon calculators remain generic, static, and disconnected from users' real behaviors. Many tools rely on simplified rule-based formulas or coarse averages that overlook important lifestyle nuances, leading to estimates that may not reflect actual individual emissions. Additionally, interfaces are often non-intuitive, provide limited visualization, and rarely offer context-aware recommendations that users can act upon.

There is a need for a system that can (i) collect detailed lifestyle inputs, (ii) learn complex relationships between activities and emissions, (iii) generate personalized breakdowns of carbon contributions across categories (e.g., travel, energy, waste, diet), and (iv) guide users with tailored recommendations and conversational support. The LiveLight CarbonFootprint project addresses these challenges through a machine learning–centric, user-friendly, and AI-augmented framework.

1.2 Objectives of the Study

The main objectives of the LiveLight CarbonFootprint system are:

- O1: Develop a robust supervised machine learning model that predicts monthly CO_2 emissions from user lifestyle features, with standardized inputs for stable and reliable performance.
- O2: Provide a detailed category-wise breakdown of carbon emissions (travel, energy, waste, diet), supported by visual analytics for intuitive interpretation.
- O3: Generate actionable, personalized sustainability recommendations that help users reduce emissions in their highest-impact categories.
- O4: Build an interactive web application using Streamlit that supports seamless data entry, responsive visualization, and integrated AI chatbot interactions.
- O5: Enable automated PDF report generation containing numerical results, visual charts, pledge statements, and sustainability tips to support long-term behavior tracking.

II. LITERATURE REVIEW

Previous research has highlighted the importance of household-level and individual-level carbon accounting for effective mitigation strategies. Jones and Kammen proposed a framework for quantifying household emissions across domains such as transportation, energy, and food, emphasizing the large reduction potential achievable through lifestyle changes and targeted interventions. Hertwich and Peters further argued for consumption-based accounting to capture indirect emissions embedded in goods and services, underlining the need for tools that can incorporate detailed consumption patterns into carbon estimates.

The emergence of machine learning for environmental applications has opened new opportunities to model complex, nonlinear relationships between activities and emissions. Rolnick et al. discussed how supervised and unsupervised learning can support climate mitigation through tasks such as forecasting, optimization, and behavioral modeling, while case studies by Zhang et al. demonstrated that regression models can accurately predict carbon emissions from structured input data when properly engineered and scaled. At the same time, comparative evaluations of existing carbon calculators have revealed limitations in personalization, accuracy, and usability, suggesting that ML-based, user-centric systems may offer substantial improvements.

Behavioral and psychological studies have also shown that personalized feedback, visualizations, and targeted recommendations can significantly motivate households to reduce energy consumption. Abrahamse et al. reported that tailored advice and feedback can drive measurable conservation behaviors, and the IPCC underscored the importance of linking policy-level climate targets with everyday actions enabled by digital tools. Advanced forecasting approaches using ensemble learning and recurrent architectures further reinforce the value of data-driven models for emissions prediction and decision support. These works collectively motivate the design of LiveLight CarbonFootprint as a personalized, ML-enabled, and feedback-oriented platform.

III.METHODOLOGY

3.1 System Overview

The LiveLight CarbonFootprint pipeline transforms user lifestyle inputs into carbon footprint predictions, visual insights, and textual recommendations through a sequence of data collection, preprocessing, model inference, visualization, and report generation stages. Users provide information about diet type, transportation modes, energy usage, and waste management habits via interactive components such as dropdowns, sliders, and text fields in a Streamlit-based interface. The application validates inputs, structures them into a tabular format, and forwards them to the data preprocessing and machine learning modules.

The preprocessed feature vectors are then fed into a pre-trained MLPRegressor model that outputs an estimate of monthly CO_2 emissions in appropriate units (e.g., kg CO_2/month). The resultant prediction is decomposed into category-level contributions for visualization and recommendation generation. An AI chatbot, powered by a generative model, accesses the user's footprint profile to offer contextually relevant sustainability tips in real time.

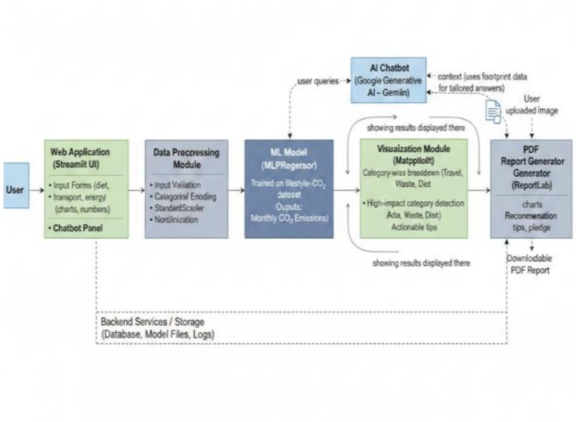

Figure 3.1 System Architecture Diagram

3.2 Model and Data Processing

The system employs a Multi-Layer Perceptron Regressor (MLPRegressor) from scikit-learn as the core predictive model. Input features are derived from user-reported attributes such as transport frequency, household energy characteristics, dietary categories, and waste practices, which are encoded into numerical form using appropriate transformations. StandardScaler is applied to normalize feature distributions, improving convergence behavior and ensuring that each dimension contributes proportionately to the regression process.

The model is trained on a dataset of carbon footprint records in which each sample links lifestyle descriptors to corresponding monthly emission values. Feature engineering is performed to capture meaningful aggregations and ratios, while the dataset is partitioned into training and validation subsets to monitor performance. Hyperparameters of the MLPRegressor (e.g., hidden layer size, learning rate, activation function) are tuned to balance accuracy and generalization, achieving target prediction accuracy of at least 90% on held-out data.

3.3 System Design and Modules

The LiveLight CarbonFootprint system is organized into modular components to facilitate maintainability and scalability:

- User Input Module: Collects lifestyle details and enforces basic validation to avoid missing or inconsistent entries.
- Data Preprocessing Module: Encodes categorical variables, scales numerical features using StandardScaler, and handles missing values where required.
- Machine Learning Module: Loads the trained MLPRegressor model, performs inference on preprocessed inputs, and returns a numeric carbon footprint estimate.
- Visualization Module: Computes category-level emission contributions and generates charts (e.g., pie charts) using Matplotlib to show the breakdown across travel, energy, waste, and diet.
- Recommendation Module: Analyzes the footprint profile, identifies high-impact categories, and produces targeted suggestions for emission reduction.
- AI Chatbot Module: Integrates a Google Generative AI model to provide conversational answers to user queries about sustainability, leveraging the user's current emissions profile.
  - Report Generation Module: Uses ReportLab to compile predictions, charts, recommendations, and optional user images into a downloadable PDF report.

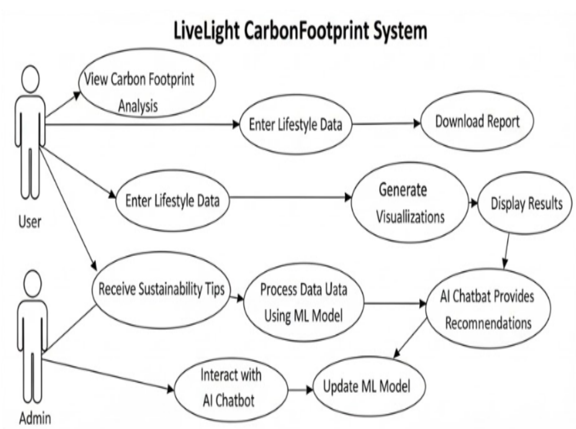

Figure 3.2 Use Case Diagram

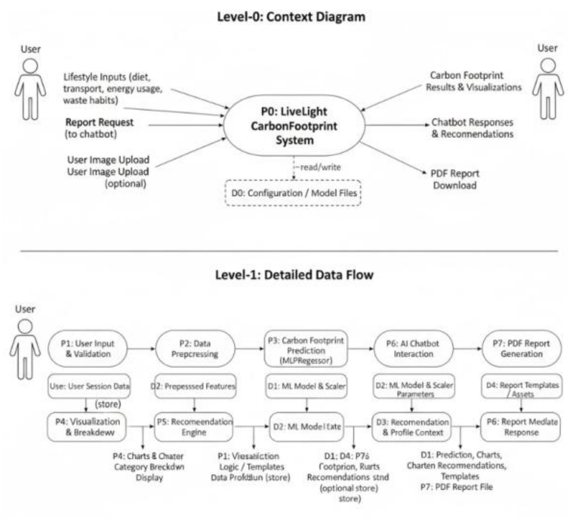

Figure 3.3 Data Flow Diagram

3.4 Advantages of the Proposed Framework

The proposed architecture offers several advantages over traditional rule-based carbon calculators:

- Personalization: The use of an ML model and detailed lifestyle inputs enables individualized predictions rather than generic averages.
- Real-Time Feedback: Users receive instant estimates and visual breakdowns, along with chatbot-based guidance, improving engagement and learning.
- Explainability and Insight: Category-wise charts and textual explanations help users understand which behaviors most influence their footprint.
- Automation and Documentation: PDF reports with pledges and tips allow users to document their progress and share outcomes if desired.
- Extensibility: The modular design facilitates the integration of new features, datasets, and AI capabilities in future iterations.

IV. RESULTS AND DISCUSSION

4.1 Quantitative Evaluation

System performance is evaluated using criteria derived from the functional and non-functional requirements of the project. The MLPRegressor model achieves an accuracy target of at least 90% in predicting carbon footprints on the evaluation dataset, with predictions typically falling within a realistic range for representative test inputs. Response time measurements show that the core prediction step completes within a few seconds, aligning with design goals for real-time interaction.

The web application's loading behavior and report generation pipeline were assessed under multiple user scenarios. The application loads within a few seconds on standard hardware, while PDF reports are generated and made available for download in under five seconds, meeting specified performance benchmarks. Concurrent usage tests further indicate that the system can handle multiple requests without noticeable degradation in user experience.

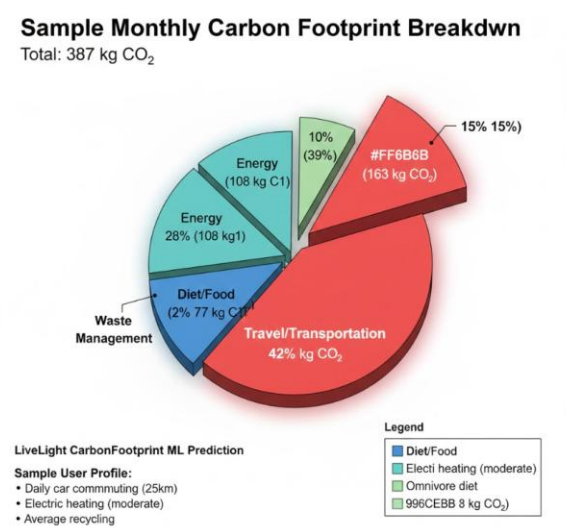

Figure 4.1 Sample Carbon Footprint Breakdown Chart

4.2 Testing and Validation

A comprehensive testing strategy was implemented, covering unit, validation, functional, integration, and user acceptance testing. Unit tests confirmed correct behavior of the MLPRegressor prediction pipeline, data normalization via StandardScaler, and AI chatbot responses under controlled queries. Validation tests verified that the system enforces proper input constraints, generates consistent PDF reports, and handles typical and edge-case inputs without error.

Functional tests focused on end-user workflows, ensuring that data submission, prediction display, visualization generation, chatbot interaction, and report downloading all behave as expected. Integration tests examined interactions between the web interface, backend model, chatbot, and reporting module, validating that data passed across components remain consistent. User acceptance testing with representative users confirmed that the interface is intuitive, the results are understandable, and the recommendations are perceived as relevant and actionable.

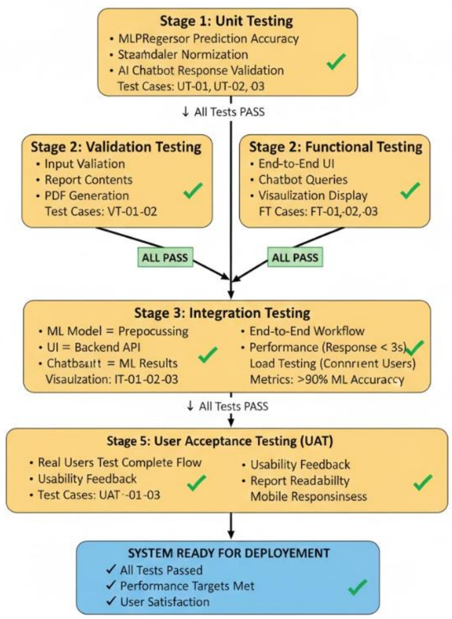

Figure 4.2 System Testing Workflow Diagram

4.3 Qualitative Insights

User feedback indicates that visual breakdowns and chatbot explanations significantly improve understanding of emission sources compared to numeric totals alone. Users reported that personalized tips—such as altering commuting patterns, improving energy efficiency at home, or adjusting dietary choices—helped identify feasible, high-impact actions. The availability of downloadable PDF reports was also appreciated as a means of documenting pledges and sharing progress with peers or mentors.

The integration of generative AI contributes to a more engaging experience by allowing users to ask open-ended questions about sustainability rather than interacting solely through static forms. This conversational layer supports learning and builds trust in the system's recommendations when aligned with transparent visual and numerical evidence.

V. CONCLUSION

This paper presented *LiveLight CarbonFootprint*, an AI-enabled machine learning framework designed to provide personalized carbon footprint estimation and sustainability guidance through an interactive web application. By combining an MLPRegressor model with standardized input features, detailed visual breakdowns, an AI chatbot, and automated PDF reporting, the system transforms raw lifestyle data into actionable environmental insights for individual users.

Quantitative evaluation confirms that the system meets its performance and accuracy targets, while testing and user feedback highlight its usability, interpretability, and practical relevance. The framework contributes to ongoing efforts to operationalize climate awareness at the individual level by making carbon accounting more transparent, data-driven, and engaging. Future work can extend the model with additional datasets, incorporate IoT data streams, explore gamification to further encourage behavioral change, and integrate community features to foster collective sustainability initiatives.

REFERENCES

[1] C. M. Jones and D. M. Kammen, "Quantifying Carbon Footprint Reduction Opportunities for U.S. Households and Communities," *Environmental Science & Technology*, vol. 45, no. 12, pp. 5285–5292, 2011.

[2] E. G. Hertwich and G. P. Peters, "Carbon Footprint of Nations: A Global, Trade-Linked Analysis," *Environmental Science & Technology*, vol. 43, no. 16, pp. 6414–6420, 2009.

[3] D. Rolnick et al., "Tackling Climate Change with Machine Learning," *arXiv preprint arXiv:1906.0543*, 2019.

[4] Y. Zhang, S. Ren, Y. Liu, and S. Si, "Machine Learning for Carbon Footprint Prediction: A Case Study," *Journal of Cleaner Production*, vol. 244, 118710, 2020.

[5] J. P. Padgett, A. C. Steinemann, J. H. Clarke, and M. P. Vandenbergh, "A Comparison of Carbon Calculators," *Environmental Impact Assessment Review*, vol. 28, pp. 106–115, 2008.

[6] W. Abrahamse, L. Steg, C. Vlek, and T. Rothengatter, "A Review of Intervention Studies Aimed at Household Energy Conservation,"

Journal of Environmental Psychology, vol. 25, pp. 273–291, 2005.

[7] IPCC, *Climate Change 2021: The Physical Science Basis*, Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 2021.

[8] Q. Zhang, Y. Wang, D. Qin, K. Zhao, and W. Xie, "Carbon Emissions Forecasting using Stacking Ensemble Learning," in *Proceedings of CIEEC 2022*, IEEE, DOI: 10.1109/CIEEC54735.2022.9845939, 2022.