Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

p-ISSN: 0972-6268
(Print copies up to 2016)
e-ISSN: 2395-3454

Vol. 24 | No. 4 | Article ID B4314 | 2025

Original Research Paper

https://doi.org/10.46488/NEPT.2025.v24i04.B4314

Open Access Journal

ConForMiSt: A Multi-model Dual-phase Framework Utilizing Machine Learning for Carbon Footprint Prediction and Reinforcement Learning for Decision Optimization

Sumukh R. Kashi, Yashas Surendra†, H. R Vineeth and K. S. Sowmya
Department of Information Science and Engineering, BMS College of Engineering, Bangalore, India
†Corresponding author: Yashas Surendra; yashassurendra.is20@bmsce.ac.in

Abbreviation: Nat. Env. & Poll. Technol.
Website: www.neptjournal.com

Received: 01-03-2025
Revised: 28-04-2025
Accepted: 30-04-2025

Key Words:

Carbon footprint
Emissions
Machine learning
Reinforcement learning
Q-learning
Touch sensor
User interface
Transport vehicles

Citation for the Paper:

Kashi, S.R., Surendra, Y., Vineeth, H.R. and Sowmya, K.S., 2025. ConForMiSt: A multi-model dual-phase framework utilizing machine learning for carbon footprint prediction and reinforcement learning for decision optimization. *Nature Environment and Pollution Technology*, 24(4), B4314. https://doi.org/10.46488/NEPT.2025.v24i04.B4314

Note: From 2025, the journal has adopted the use of Article IDs in citations instead of traditional consecutive page numbers. Each article is now given individual page ranges starting from page 1.

ABSTRACT

Over the past decade, there has been a significant surge in harmful waste emissions of greenhouse gases, namely carbon dioxide, methane, and fluorinated gases, in the atmosphere. Two major categories of activities can be broadly identified as contributing to this condition. The first is the proliferation of worldwide industrial activity, as accounted for by industrial plants across all major continents. Second, human activity contributes to carbon emissions produced as a result of wide-ranging everyday activities that involve the use of electricity, transportation, food consumption, and other consumer-mindset-driven activities. This study focuses on the second category to build a dual-stage framework that will assess, evaluate, and recommend suitable mitigation measures to regulate usage patterns. The dual-stage approach is a novelty based on sound engineering principles. Carbon emission data gathered by the system were analyzed to detect footprint generation patterns using mathematical models. After the analysis, machine learning models selected from rigorous performance metrics (MAE, RMSE) were leveraged to predict the carbon footprint in the first stage. The second stage employs a reinforcement learning framework that captures several aspects of emissions in a 'state' and is used to analyze predictions and generate recommendations considering user preferences. The ability to absorb user goals for emission data is a strength. This unique finer engineering of state representation exemplifies experimental data that show minimal variation in state goal values within 2000 steps. A web application was developed to visualize various aspects, such as usage patterns and predictions. The user interface provides interventional, specific, and personalized recommendations. These aspects are then utilized to provide insights at the aggregated level in the context of a group of individuals, which is yet another strength of this framework. The extensibility of the proposed methodology for carbon emission mitigation for higher aggregated levels is demonstrated by an exemplar 'location statistic' radar chart in the context of the vehicle and electrical appliances categories.

INTRODUCTION

Over the past decade, there has been a meteoric rise in the number of industrial plants and the subsequent waste generated by them. Pollution resulting from such plants can be mainly attributed to the harmful gases they discharge. A significant increase in the presence of gases such as carbon dioxide, methane, nitrous oxides, and fluorinated gases in the atmosphere is due to increasing air pollution (Ulfat & Noori 2024). These gases trap large quantities of heat, constricting their circulation in the atmosphere. A direct consequence of this is the increased carbon presence on a global scale.

Statistical data on greenhouse gas emissions in the Indian subcontinent provide valuable insights into the various sources of carbon emissions across the country. India accounted for ~7.3% (3.9 billion metric tons (GtCO2) of the global greenhouse gas emissions in 2022, making it the third largest emitter after China (29.2%) and the United States of America (~11.2%) (Chandel 2022). As of 2022, the G20 countries have been responsible for three-quarters of global warming (Jones et al. 2023).

The natural recovery process of the Earth system to capture and store CO2 will be exceeded unless an anthropogenic mitigation component is added to restore balance.

This study aimed to address the mitigation component by assessing the amount of carbon footprint contributions by human activities on a personal scale and appropriately provide behavioral modifications to mitigate the carbon footprint.

In the context of human activity, a personalized carbon footprint describes the impact of any given activity of a person or group of people that results in the emission of greenhouse gases into the atmosphere (Schwenkenbecher 2014). This is an indicator of the amount of strain imposed on the environment. A personalized carbon footprint can result from a broad variety of activities, including electricity usage, personal vehicle usage for commuting, the purchase of clothes, electronic products, and various other activities, such as food consumption. Furthermore, buying food items and paper-based products adds to the personal carbon footprint. The domestic sector is a major contributor to the carbon footprint of India (Jain & Nagpal 2021). Various sources of footprints include electricity consumption from home appliances such as tube lights, ceiling fans, air conditioners, and washing machines. Additionally, fuel consumption from personal vehicles contributes considerably to the footprint (Onat et al. 2015). The public transportation sector is another source of footprint emissions in a country. Footprint emissions are also contributed by livestock management and agriculture (Ramachandra et al. 2015). Detailed assessments of the life cycles of carbon-emitting substances provide a good understanding of their overall environmental impact. Huang et al. (2023) provided a recent study of various components of the Household Carbon Footprint (HCF) in India is provided by Huang et al. (2023). According to this study, 39% was contributed by energy consumption, 20% by travel, and 14% by food. Therefore, if these contributory components can be reduced, the HCF will significantly improve.

Increased emissions of greenhouse gases result in increased atmospheric temperatures (Khan et al. 2024) and altered weather patterns, causing ice caps to melt, rising sea levels, and potentially catastrophic events such as hurricanes, floods, and droughts (Kiehbadroudinezhad et al. 2024). Another notable consequence of increased carbon emissions is the rise in respiratory problems, cardiovascular diseases, and premature death of flora and fauna. This can be mainly attributed to pollutants such as nitrogen oxides, sulfur dioxide, and particulate matter (PM) (Kumar et al. 2023).

Although the major contribution to the global carbon emissions problem stems from macro-industrial activities, it is important and cogent to address carbon emission contributions and problems from individual entities arising from their activities and behaviors.

MATERIALS AND METHODS

Problem Statement and Solution Strategy

The primary objective of this study was to analyze carbon footprint generation patterns from various carbon-emitting devices and provide a mitigation solution (Goswami et al. 2024).

Summarized problem statement: Collect carbon footprint data from various categories of carbon emitters. The data was analyzed, and future emissions for various activities in these categories were predicted. Using this information, provide recommendations to reduce carbon footprint emissions from various activities.

A workable solution strategy: Our solution framework is ConForMiSt, an acronym for Carbon Footprint Mitigation System, and consists of the following elements:

1. Collect raw consumption data from various categories of carbon emitters
2. Analyze carbon footprint generation patterns
3. Conceptualize machine learning models that will capture essential aspects of these patterns to formulate a carbon footprint estimation and mitigation model.
4. Use reinforcement learning models to provide recommendations to reduce carbon from various human activities – A framework.
5. Provide a web-based and cloud-centric architecture for scalability and extensibility, so that this application may be used in a macro-level setting like an institution or organization.

Related Work

We now describe the literature that describes carbon footprint measurement and the attendant mitigation mechanisms. To perform this study, we captured and categorized the work into various categories driven by human activities that lead to the generation of carbon footprints. This prior study provides a starting point for mitigating the carbon footprint.

Mobile phone usage: A study that describes the usage of devices with networking and communication capabilities has been succinctly described in (Lövehagen et al. 2023).

Experiments were performed on smartphones, feature phones, tablets, laptops, and personal computers in this study, revealing varied emissions of carbon on any given day. The equivalent carbon footprint (CO_2e) of phone use is 57 g CO_2e per minute of use. The study findings revealed that the respondents spent the most on texting, with an average of 17.95 kg CO_2e per day. A general recommendation for reducing electronic usage is also described in this study.

Forest activity: A categorization of the carbon emissions produced from different types of forests, namely primary, naturally regenerated, and planted forests, has been provided by Mancini et al. (2016). The effects of forest wildfires and harvested wood products were factored into the calculation of the carbon footprint. A detailed account of forest carbon sequestration as a contributing factor to the carbon footprint has been provided.

Organizational activities: The concept of operational boundaries that define the physical and geographical limits of an organization's activities for measuring and reporting its greenhouse gas emissions has been detailed in (Gao et al. 2014). A study on carbon emissions generated in the corporate sector was conducted by leveraging several machine learning models (Musa et al. 2024).

Food consumption and behaviour: The analysis of changes in consumer behavior in the context of food products that contain labels detailing the carbon footprint output of food processing has been summarized in (Rondoni et al. 2021). Labels enable food manufacturers to indicate the impact of their food production processes on the environment. Consequently, this allows consumers to make informed choices when purchasing processed food products. The findings of this study reflect the positive attitude of countries such as Egypt and China towards carbon footprint information, urging other emerging countries to develop a similar outlook on processed food products.

Approaches and frameworks for the estimation of carbon footprint: An approach revolving around statistical modelling has been employed to predict carbon footprints from a corporate standpoint for climate finance risk analysis (Nguyen et al. 2021). Models such as linear regression, k-nearest neighbors, and decision tree ensembles have been leveraged to carry out predictions. The model performance was evaluated using the mean absolute error (MAE) and multiple resampling techniques. A framework that targets systemic reporting, energy promotion, and carbon efficiency using machine learning techniques was described by Henderson et al. (2020). The use of green defaults and effective component integration to achieve pro-environmental behavior has been discussed. An empirical framework comprising elastic network regression and machine learning models to arrive at effective carbon footprint mitigation strategies was provided by Dong and Zhang (2023). Energy consumption dynamics have been examined, and the correlation between carbon emissions and their causative factors has been analyzed to provide suggestions for carbon emission mitigation measures (Liu et al. 2023).

A comparison of the ConForMist framework with existing works is provided in Table 1 to highlight the key features and differences of the proposed system.

Research gap in existing implementations of Home Energy Management Systems (HEMS): Existing frameworks in HEM systems primarily focus on energy savings, electricity, and money. The research gap addressed in ConForMist is the carbon emission focus-centric approach in homes. The proposed system will help stakeholders track and achieve reduced carbon emission goals. There is a need to develop sustainability solutions that can be easily integrated into existing architectures. The architecture functions without considering smart meters, smart appliances, and smart grids, and hence can be deployed faster, cheaper, and provides a scalable solution in less developed electrical grids and homes.

We used carbon emission data directly to train the regression models. The outputs of carbon emissions are offset

Table 1: Comparison report of ConForMist with existing literature frameworks.

Research Paper	Methodology	Key Results	Comparison
Lissa et al. (2021)	Deep Reinforcement Learning	Unique comfort factor v/s energy savings metric.	User's preferences are taken as input, and use of ML models are used with historical data to provide statistical data.
(Lee & Choi 2019)	Deep Q-Network, Deep Learning	Forecast of indoor temperatures and ToU prices lead to real-time decision making.	Different models for different appliances, hence providing more granularity.
(Mahmoud & Ben Slama 2025)	Deep Learning, Reinforcement Learning	Vehicle to Home bilateral communication	Usage of carbon emission data over energy units to focus on home emission data.
(Solatidehkordi et al. 2023)	LSTM, IOT	Real-time classification of appliance data usage	The microcontroller with timestamp recording, UI Dashboard provides yearly, location-based data and emission data statistics.

by a percentage (determined by the end-user) and fed as input to the reinforcement model as goal states. The novelty of the reinforcement model lies in its learning, which is based on the concept of having four phases in a day, along with optimization of the usage of devices. Another research gap being addressed is the user flexibility on the input to their solutions and trade-offs on comfort versus energy savings.

Top Level Systems Approach System Design and Implementation

The first stage handles tasks ranging from the collection of user consumption data to its cleaning and pre-processing. As a starting point, six different carbon-emitting sources were considered for measurement. Data related to carbon emissions resulting from human activities were collected. These sources were chosen to demonstrate proof of concept, considering their consumption priority over low-usage devices in the Indian subcontinent. Two broad categories of carbon emissions are considered (Transport and Electricity) to demonstrate the concept of operations, and this can be easily extended to other categories as well. The transport component comprises Four-wheeler and Two-wheeler vehicles that run on two types of fuel sources: petrol and diesel. The usage patterns of vehicles running on both types of fuel were considered during the dataset formulation process.

Data collection and pre-processing: The top portion (a) of Fig. 1 shows the data collection and pre-processing steps. The various sources of carbon emissions across various individuals in each of these two broad categories of carbon emitters were collected, aggregated, and analyzed. The carbon emission factors of these appliances were used to compute their carbon footprint output for a specific duration of usage. The devices are completely dependent on their corresponding carbon emission factors. The method of data collection involved recording the device readings during the four phases of each day over a course of 12 months, with the following apportionment. The results are presented in Table 2.

The usage (in hours) was recorded during each phase of the day, and the carbon footprint was computed. The usage-footprint pair values were stored in a dataset (CSV file) for further cleaning and processing. Subsequently, the values in the dataset were checked for their logical correctness to clean the dataset from inaccurate measurements recorded by the sensors. If a user-input value exceeds the range provided in Table 2, it is perceived as an erroneous input and subsequently discarded.

Table 2: Apportionment of Phases of a Day for Data Collection.

Phase of Day	Time (24-hour Clock Format)
Morning	00:00 - 05:59
Afternoon	06:00 - 11:59
Evening	12:00 - 17:59
Night	18:00 - 23:59

Initial dataset preparation: The bottom left portion (b) of Fig. 1 shows the data preparation steps. In the second stage, the cleaned dataset is fed to machine learning models for training in the next stage of implementation. The selection of a particular machine learning model for predicting the carbon footprint output of a given device was made after examining the input-output variable relationships. A linear regression model was trained for the input-output data columns that exhibited a linear relationship (Maulud et al. 2020). Random Forest regression (Ali et al. 2012) and Gradient Boosting models were trained for non-linear data points (Otchere et al. 2022), and the choice between these two models was made by comparing their mean absolute errors on the training data.

Prediction of carbon footprint from activities monitored: Part (c) of Fig. 1 shows the prediction steps. The third and final stage of the implementation involves the generation of personalized recommendation reports based on the device usage history of the individual. A reinforcement learning model customized for each type of vehicle or electrical appliance drives the recommendation generation engine. Additionally, data insights such as the average and total usage of appliances and major contributors by the location of the users at an aggregated level were generated by leveraging NumPy, Pandas, and Seaborn libraries. The stakeholders of the aggregated data insights are government organizations whose objective is to reduce carbon footprint emissions in areas under their jurisdiction by levying taxes and fines.

Generation of recommendations, strategies and on the dashboard: Part (d) of Fig. 1 identifies the recommendation step. A Web-based tool with role-based access was developed to enable analysis, interpretation, and key strategies/actions. A key driver for the user interface is to enable individuals to access personalized recommendation reports and even larger entities, such as government organizations, to access aggregated data insights from different regions of the state and city.

System Design And Implementation

Engagement of machine learning models: The structured dataset offers a rich training dataset for the machine learning models. In the case of the electrical appliance activity category, by analyzing the correlations between time-dependent appliance usage, the models can learn to estimate the carbon footprint with increasing accuracy. A similar case is observed in the transport activity category, which involves

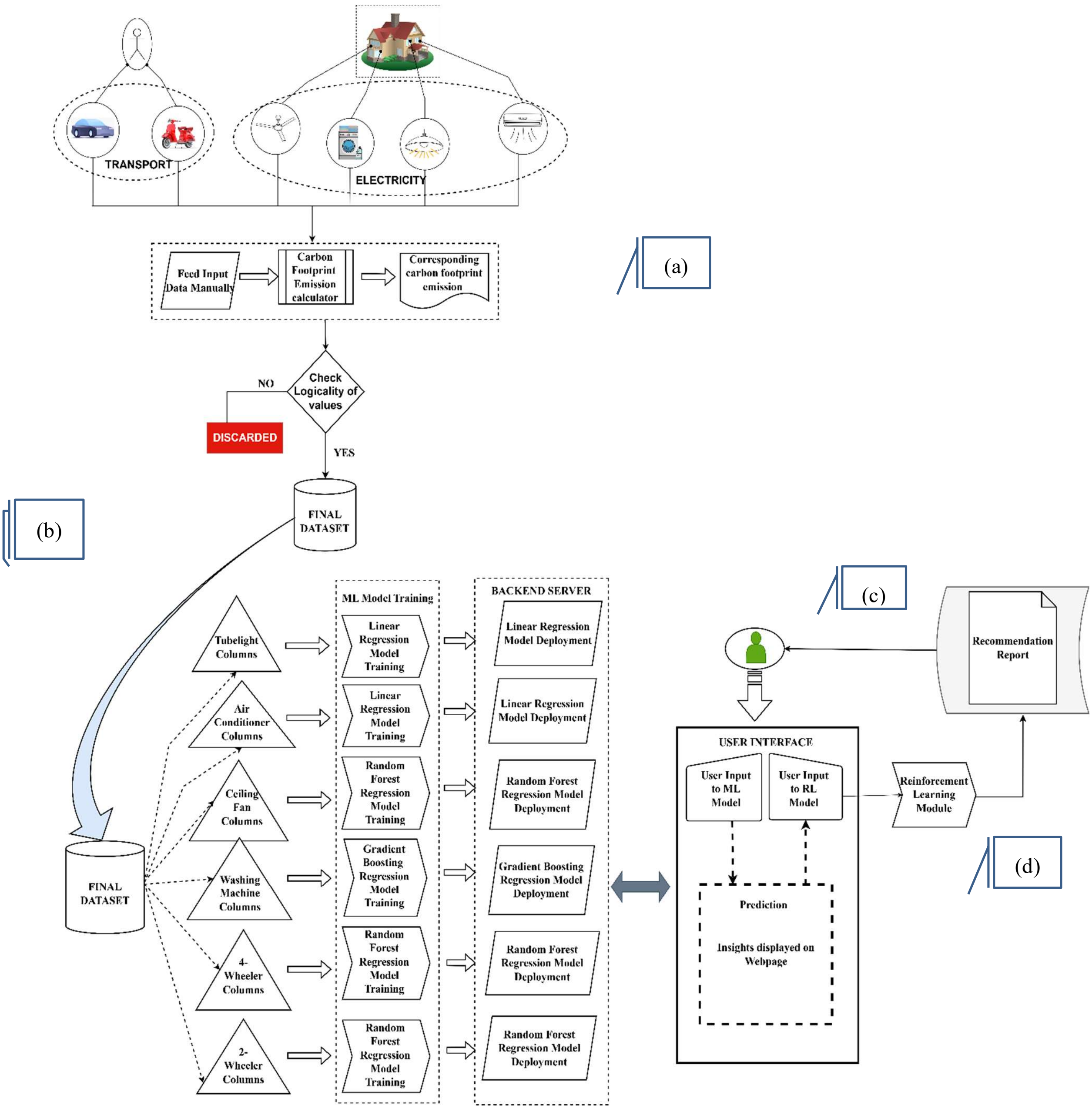

Fig. 1: A Bird's Eye View of ConForMiSt Process.

analyzing the correlation between vehicle travel patterns, fuel types, and other characteristics. This is a data-driven approach that focuses on capturing diurnal variations in user behavior. This approach is not reliant on manual data input and paves the way for an automated and efficient system for carbon footprint assessment.

Several machine learning models were trained using the dataset produced after preprocessing. The approach begins by feeding the structured data of a specific appliance or vehicle to a machine learning model. The selection of a machine learning model for a specific appliance or vehicle is determined by empirical findings obtained from experiments conducted in a set of investigative trials (Varoquaux & Colliot 2023). The models were chosen after comparing each model individually based on the mean absolute error. The machine Learning models considered for extracting empirical evidence were Linear Regression, Random Forest Regression and Gradient Boost Regression (Table 3).

The dataset used to feed these models was divided into training and testing sets using various trial counts (Ranjan et al. 2019). Four training trials, each with a distinct training and test split of the dataset, were used to train the machine learning models. In the initial experiment, 30% of the data was used for testing and 70% for training. The second trial used an 80%–20% split for testing and training. The third trial used a 90%–10% dataset split for the same, while the fourth trial had a 75%–25% split for the test-train dataset.

A grid-search approach to identify the right combination of hyperparameters was employed to methodically build and evaluate a model for each combination of algorithm parameters specified in a grid (Bhatti & Iqbal 2000).

Comparing the metrics of the various models for the transport activities and electricity usage from Table 3, the following conclusions were drawn: The Linear Regression model was found to provide accurate results for data recorded by tube lights and air conditioners for the electricity usage category. Similarly Random Forest Regression model was found to provide accurate results for inputs provided by ceiling fans for the electricity usage category, while the same model was also the model of choice for two- and four-wheeler usage in the transport category. Finally, the Gradient Boosting Regression model provided accurate results for the inputs from the washing machine.

***Output of trained models*:** The aforementioned machine learning models predict emissions from carbon emitter categories based on input usage data.

***Post processing*:** These trained models were serialized into a pickle file and later deployed on a web server. Pickle files are generated by employing the technique of serialization. Serialization is the process of decomposing complex data structures into a sequence of primitive data parts, which can be saved directly in a file or transferred over a network (Bellman 1957). On the receiver end, the pickled files are deserialized and fed with inputs from users to obtain a prediction of the amount of carbon emissions with respect to the amount of usage in hours or kilometers. The obtained prediction value is displayed to the user using a responsive

Table 3: Experimental ML Models for Third Stage (Prediction).

Design Rationale	Carbon emitter Category/ Human Activity/	ML Model chosen for Analysis	Perf. Metric	Metric Value	Metric Value - RMSE
Relationship between input-output variables	Transport (2Wheeler)	Linear Regression	MAE	19.566	24.53
To fit non-linear data points	Transport (2Wheeler)	**Random Forest Regression**	MAE	2. 499	3.13
To reduce the mean absolute error	Transport (2Wheeler)	Gradient Boosting Regression	MAE	2. 614	3.28
Relationship between input-output variables	Transport (Four-Wheeler)	Linear Regression	MAE	29.077	36.45
To fit non-linear data points and	Transport (Four-Wheeler)	**Random Forest Regression**	MAE	5.349	6.7
To reduce the mean absolute error	Transport (Four-Wheeler)	Gradient Boosting Regression	MAE	4.735	5.94
Relationship between input-output variables	Electricity (Tube Lights)	**Linear Regression**	MAE	0.218	0.27
To fit non-linear data points	Electricity (Tube Lights)	Random Forest Regression	MAE	2.757	3.46
To reduce the mean absolute error	Electricity (Tube Lights)	Gradient Boosting Regression	MAE	8.583	10.76
Relationship between input-output variables	Electricity (Air conditioner)	**Linear Regression**	MAE	0.175	0.22
To fit non-linear data points	Electricity (Air conditioner)	Random Forest Regression	MAE	0.879	1.1
To reduce the mean absolute error	Electricity (Air conditioner)	Gradient Boosting Regression	MAE	1.901	2.38
Relationship between input-output variables	Electricity (Ceiling Fan)	Linear Regression	MAE	9.163	11.48
To fit non-linear data points	Electricity (Ceiling Fan)	**Random Forest Regression**	MAE	0.200	0.25
To reduce the mean absolute error	Electricity (Ceiling Fan)	Gradient Boosting Regression	MAE	0.631	0.79
Relationship between input-output variables	Electricity (Washing Machine)	Linear Regression	MAE	3.304	4.14
To fit non-linear data points	Electricity (Washing Machine)	Random Forest Regression	MAE	0.336	0.42
To reduce the mean absolute error	Electricity (Washing Machine)	**Gradient Boosting Regression**	MAE	0.189	0.24

user interface. The user is offered a choice to offset their prediction by 5%, 10%, or 15%. The reduced value of carbon emissions represents the user's preferred amount of emissions or a target that must be achieved within a specific duration. The offset value, along with preferences containing the amount of usage expected in all four time slots of a day, namely morning, afternoon, evening, and night, is given as input to the corresponding reinforcement learning model deployed on the web server and tuned to the particular appliance. The goal consumption input, preferences in a day, and fuel type in the case of vehicles are provided, and personalized recommendations of the amount of usage in the hour-minute format in the case of electrical appliances and kilometer format in the case of vehicles are provided to achieve the goal consumption.

Engagement of reinforcement learning models: The predicted carbon emissions from the use of carbon emitters (devices in various categories) were used as the base data for further processing after user interactions.

User interaction involves the following steps:

Present a summary of the carbon emissions from the individual's 12-month recorded data

Three different reduction level choices are provided to the user.

Accept the goal quantity for reduction in terms of percentage over the base data.

This target is used to process information through a reinforcement learning model to develop a suitable strategy for reducing the user's carbon footprint. The reduced carbon emissions from the offset value are the goal state of the reinforcement model.

The reinforcement learning model is mechanized as follows.

Each state of the model is defined as a tuple indicative of

Phase of the day

Device Operational Parameter

Energy Consumption Level

An action set consisting of action values represented as an enumerative list

Action_1Action2, ……

A value function (q-value) that evaluates the value of a state and provides a value for the best state that generates the lowest consumption value. The reward value generation is driven by the simple principle of providing rewards for high carbon generation (e.g., a higher temperature setting in the case of an air conditioner).

The underlying principle of using reinforcement learning is to optimize and simulate the usage of devices during 24 h based on a percentage reduction from the previous machine learning algorithm prediction values. The output provided to the user will be the apportionment of the optimal usage of

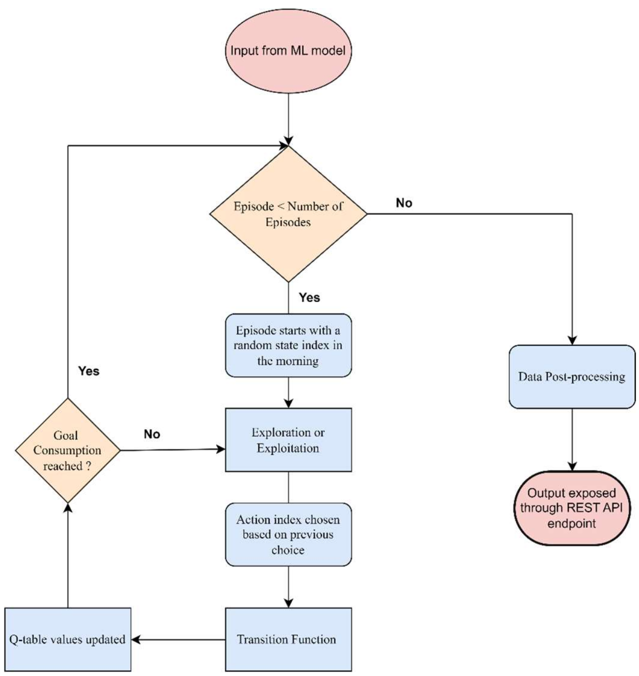

Fig. 2: Phase 2 Agent Mechanization using RL.

the devices among different phases of the day. The agent is given a goal consumption value as the input and learns until the consumption for an episode reaches the goal value. The agent mechanization is shown in Fig. 2.

The reinforcement learning algorithm was devised for six different models corresponding to tube lights, fans, washing machines, air conditioners, two-wheeler, and four-wheeler vehicles. Each of them has the same underlying principle, which consists of a Markov decision process (Watkins & Dayan 1992) with a policy of q-learning (Babiuch & Postulka 2020). The idea behind considering the states in this process stems from the dataset structure, which is the division of the 24 h into four phases, namely, Morning, Afternoon, Evening and Night. Any action taken by the devices considered leads to a change in the carbon emission output from the corresponding device. The relation between the states and the actions must be established and defined in a manner to construct the Q-table.

a. Actions

The actions define the interaction of a user who uses the different functionalities of the devices under consideration. These actions are independent; that is, one action has no effect on how the other action is performed. A brief outline of the actions for various device categories is provided below.

Light: [“Off”, “Low Brightness”, “Medium Brightness”, “High Brightness”]

The above actions considered modern, present-day LED’s which have controls to change their brightness using a regulator component.

Fan: [“Off”, “Low Speed”, “Medium Speed”, “High Speed”]

The different speeds of the fan, which can be set by rotating the dial, are mapped to the actions above.

Air Conditioner: [“Off”, “Moderate Cooling”, “Cold Cooling”, “Coldest Cooling”, “Eco Mode”, “Turbo Mode”]

The actions for an air conditioner consider the choices that an appliance user generally selects, ranging from a completely switched-off state and degree of cooling (three levels) to specific modes (Eco and Turbo). Eco Mode refers to ideal cooling, which will lead to reduced consumption but achieve the required temperature at a slower pace and fan speed. Turbo Mode is quite the opposite, having higher consumption while reaching the required temperature faster, along with a faster fan speed.

Washing Machine: [“Off”, “Quick Wash”, “Normal Wash”, “Heavy Duty”]

Actions mapped for a Washing Machine are those that represent general modes of the different types of washing in a general washing machine. If the gradation of these modes is considered, the order of heavy, normal, and quick washes indicates the descending order of consumption.

Transport: [“Off”, “Accelerate”, “Cruising”, “Speeding”, “Engine Braking”]

The actions for the transport category device are indicated above. Although the actions are intuitive, we differentiate between accelerating and speeding as follows: the accelerating action is a mediated increase in speed, while the speeding action is going at a rapid pace at a higher gear. Cruising and Engine Braking are ideal techniques for reducing fuel consumption and emissions.

b. States

An accurate model of the state description is required. Therefore, a state is represented as a tuple consisting of ({phase of the day}, {Device Operational Parameter}, {Energy Consumption Level}). State modelling is driven by three primary factors that determine a state: (a) the phase of the day influences the consumption profile; (b) the device under consideration may operate in various modes (e.g., a vehicle operating in low and high gears, an air conditioner operating in eco mode or moderate-cooling or turbo cooling mode); and (c) consumption level. The device operating parameters are then amenable to directly mapping the user choice to the operating mode of the device. The Zero factor is equivalent to the OFF state, and the device does not emit a footprint, as provided in Table 5. The correlation between the device factor and consumption level may be either positive or negative, depending on the action taken.

The various states are defined using Table 4 by choosing a discrete value from each column of the device.

Table 4: State “Tuple” Components for defining State Representation.

Phase of the day	Device	Device Operational Parameter	Consumption
Morning, Afternoon, Evening, Night	Light	Zero, Low, Medium and High Brightness	High, Moderate, Low
	Fan	Zero, Low, Medium and High Speed	
	Air Conditioner	Zero, Low, Moderate and Coldest Temperature	
	Washing Machine	Zero, Quick, Normal and Heavy Wash	
	Transport	Zero, 1 ,3 and 5 gears	

c. Transition Functions

The recommender system uses the Q-Learning category of reinforcement learning and therefore always chooses an action based on the Q-value state at a state, according to Equation (1).

$$TD(s_t, a_t) = r_t + \gamma \cdot \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \quad \ldots(1)$$

where,

s_t: Current state at time step t

a_t: Action taken at time step t

r_t: Reward received after taking action a_t in state s_t

γ: Discount factor ($0 \leq \gamma \leq 1$) that determines the importance of future rewards

$Q(s_{t+1}, a)$: Estimated future reward for taking action a in the next state s_{t+1}

$\max_a Q(s_{t+1}, a)$: Maximum Q-value over all possible actions in the next state s_{t+1}

$TD(s_t, a_t)$: Temporal Difference (TD) error, which measures the difference between the current estimate and the updated estimate of the Q-value.

A minor augmentation to this transition (which is based on the Q-value) is the check for an "End State" at the end of every action. This is because the user is provided with a choice to select the offset for total carbon footprint reduction, which is translated to a "maximum consumption limit" check. If this limit is reached in any state, the learning process is stopped. An exemplar encoding of these end states for various device and transport vehicle categories is provided in Table 5.

The transitions are based on a weight factor, the inputs of which are provided by the user. The structure of the weights is represented as [a, b, c, d], which represent the weights of usage for Morning, Afternoon, Evening and Night. Based on these weights, the reinforcement-learning agent transitions to different phases. For example, if a user decides that the need for an LED would be higher in the evening and night, he could increase the relative weightage of these two phases. A random number generator generates the next-phase index based on the relative weights. The second part of the transition function is the action transition function. This function defines the logic for the device factor variation based on the action. The transitions vary depending on the device.

The "Engine Braking" state can transition to three possible states, making it a stochastic process. When considering the braking process, the vehicle slows down to a lower gear. Hence, the three states are justified. Another factor considered in the transport model is the traffic probability. The higher the factor, the higher the probability of the agent transitioning to the state of the first gear. Under higher traffic conditions, the cars are slower and would be in the lower gears, that is, either the first or the second gear. The second-gear state is not considered in our model for simplicity and to avoid redundancy.

Another key factor to be considered is the agent learning beyond the maximum carbon consumption. Hence, the maximum consumption limit was considered.

$$\text{MaxConsmpCheck} = 6*(\text{device consmp factor kgCo2.h}^{-1}) \quad \ldots(2)$$

System implementation – Proof of concept: The second category of carbon emitters in Table 3, 'Electricity'

Table 5: End state encodings for devices and vehicles.

Device: Light	
Action	Next State
"Off"	(Zero, Zero)
"Low Brightness"	(Low Brightness, Low Consumption)
"Medium Brightness"	(Medium Brightness, Medium Consumption)
"High Brightness"	(High Brightness, High Consumption)
Device: FAN	
Action	Next State
"Off"	(Zero, Zero)
"Low Speed"	(Low Speed, Low Consumption)
"Medium Speed"	(Medium Speed, Medium Consumption)
"High Speed"	(High Speed, High Consumption)
Device: Air Conditioner	
Action	End State
"Off"	(Zero, Zero)
"Moderate Cooling"	(Moderate Temperature, Low Consumption)
"Cold Cooling"	(Low Temperature, Moderate Consumption)
"Coldest Cooling"	(Coldest Temperature, High Consumption)
Device: Washing Machine	
Action	End State
"Off"	(Zero, Zero)
"Quick Wash"	(Quick Wash, Low Consumption)
"Normal Wash"	(Normal Wash, Moderate Consumption)
"Heavy Duty"	(Heavy Duty, High Consumption)
Transport: VEHICLE	
Action	End State
"Off"	(Zero, Zero)
"Accelerate"	(3, Moderate)
"Cruising"	(5, Low)
"Speeding"	(5, High)

Algorithm 1: Recommendation Algorithm

```
Require:
– reward_function (state)
– action_function (choice, action index)
– transition (cons, action, weights, max_consump_check)
– num_episodes
– epsilon
– Q_table
– actions
– states
– learning_rate
– discount_factor
– goal_consumption
Ensure:
– Q_table update
– Iteration for num_episodes
Begin
for each episode in range (num_episodes):
   if max_consump_check is true:
      Print "Max" and Break
   Initialize weights: [a, b, c, d]
   Initialize total_consumption: 0
   Choose random state index: state_index
   Initialize cons_values for times of day
   while true:
      if random number < epsilon:
         Choose random action_index
      else:
         Choose action_index with the highest Q–value
      Simulate environment:
         – Get a reward
         – Compute next_state_index, max_consump_check
         – Update Q–value in Q_table
      Update total_consumption
      Update consumption values
      Update state_index
      if total_consumption >= goal_consumption:
         Break
End
```

consumption driven by Tube/LED Light bulbs, was chosen as a representative example for experimentation and demonstration of our solution. The experimental setup shown in Fig. 3 provides a test bed for collecting the usage data of individuals using the light appliance. The hardware circuit architecture consists of an ESP-32 WROOM 32D microcontroller, TTP-223 touch sensor, 10KΩ resistor, and LED bulb. The pin configurations for the microcontroller are shown in Fig. 3 to clarify the interfaces. The choice of the ESP32 module was driven by its Wi-Fi connectivity feature, which eases integration with a web server.

a. Experimental setup of data gathering

The touch sensor is placed on the switch, which is used to control the appliance. The touch sensor interfaced with the ESP32 module was used to record the electricity consumption of the individual. The implementation (code) that supports the touch sensor functionality was configured as follows: the bulb was initially in the OFF state. When the sensor is pressed (stimuli), the bulb transitions to the ON state and continues to remain in the same state until the sensor is pressed again. This subsequent action results in the transition of the bulb to the OFF state. The timestamp at which the bulb transitioned to the ON and OFF states was recorded (Arduino IDE interface), from which the duration for which the bulb was utilized was computed.

b. Data Collection and Processing from Sensor equipment

Fig. 4 provides an overview of the data processing pipeline, showing how the information is gathered from devices using touch sensors and a microcontroller (ESP32) (Hiraman 2018). The recorded data included device usage data in milliseconds. The usage data obtained from the

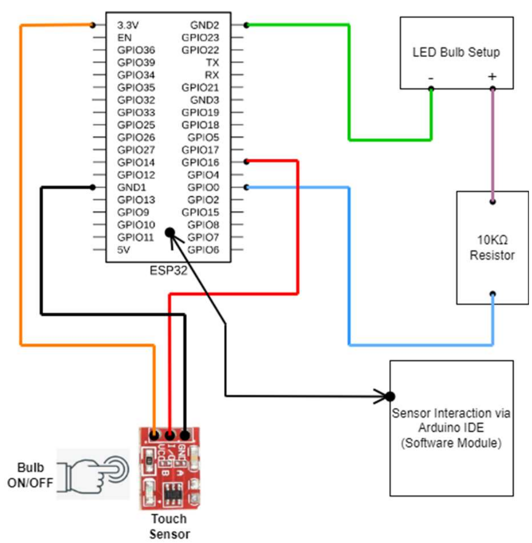

Fig. 3: Experimental setup for Lighting Consumption Data Gathering.

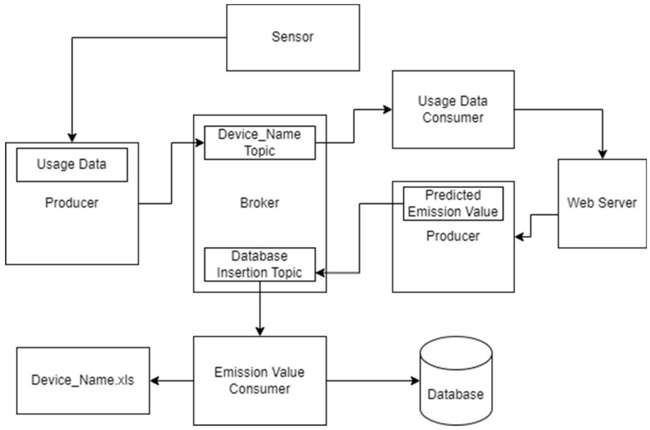

Fig. 4: Data Processing Pipeline in ConForMiSt.

sensors were stored in a text file, with each device usage interval duration (each on-off condition) stored on a new line.

The device usage data is sent to an Apache Kafka platform for further processing. Apache Kafka is a platform for processing real-time streaming data, using a distributed publisher-subscriber messaging system. A significant advantage is its ability to handle large volumes of data (concurrent requests from users) and reduce latency in providing predictions to users (Thein 2014). As noted earlier, device usage data are recorded with files, and these files need to be uniquely identified. A file's unique hash value enables the identification process, called a message digest. This is created using a hashing algorithm, such as MD5, which is of fixed length and uniquely identifies the contents of the file.

An update operation to the file replaces the pre-existing message digest with a new hash value, which uniquely identifies the modified content of the file. A Python script that indefinitely monitors the state of the message digests triggers an event when the file is modified with a new sensor reading. The latest data obtained is passed to the Kafka broker with the corresponding device name (e.g., "Tube-light") as the topic through a Kafka Producer via a message after serialization. Another Python script, which runs the Kafka Consumer indefinitely, fetches the message from the broker and deserializes it to obtain the usage data. Producers publish messages to Kafka topics, and consumers subscribe to and consume these topics (Bahrami et al. 2018).

The obtained usage data were sent to a web server running the trained Machine Learning models using an HTTP request to the API endpoints exposed by the web server (Raptis et al. 2022). The web server returns the predicted emission value, which is again sent to the broker using a Kafka Producer that is configured to a topic (Wu et al. 2019) named "Database Insertion" after serialization of the dictionary containing the user details, device name, sensor type, emission, and usage values. The Python script, which indefinitely runs the Kafka consumer, fetches the message (Narayanan 2024) under the "Database Insertion" topic in the broker and deserializes it to obtain the dictionary. The dictionary is saved to a database and a CSV file along with timestamps, thus serving as a dataset to retrain the existing models.

RESULTS AND DISCUSSION

The following subsection describes the inferences drawn from the machine learning models of each type of device and vehicle belonging to the carbon emitter categories of electricity and transport. The consumption parameters were examined using graphical plots.

Results from Machine Learning Models

Structured datasets offer a rich training ground for machine learning models. By analyzing the correlations between time-dependent appliance usage, vehicle travel patterns, fuel types, and the calculated carbon footprint, the models can learn to estimate the carbon footprint with increasing accuracy. This data-driven approach, which focuses on capturing diurnal variations in user behavior, has the potential to reduce reliance on manual data input in the future, paving the way for a more automated and efficient system for carbon footprint assessment. The results were derived from models trained on a 12-month dataset. However, the models can be trained beyond the 12-month dataset.

Inferences Drawn from Two-Wheeler and Four-Wheeler Plots

The partial dependence plot in Fig. 5 exhibits an overall positive correlation between two-wheeler usage in kilometers and emission values, with a nonlinear relationship characterized by multiple local maxima, minima, and periodic fluctuations that are not significant. The plot

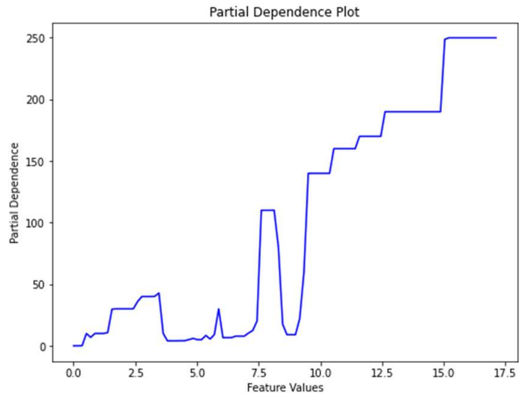

Fig. 5: Partial Dependence Plot - 2-Wheeler.

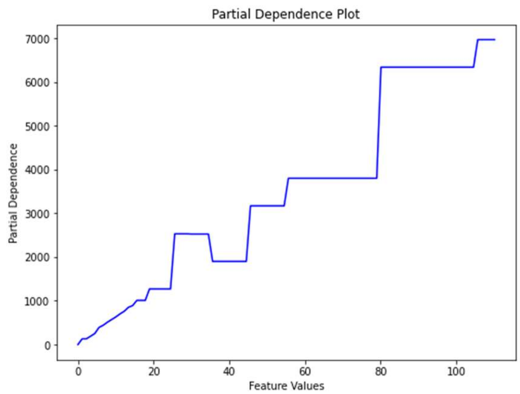

Fig. 6: Partial Dependence Plot - 4-Wheeler.

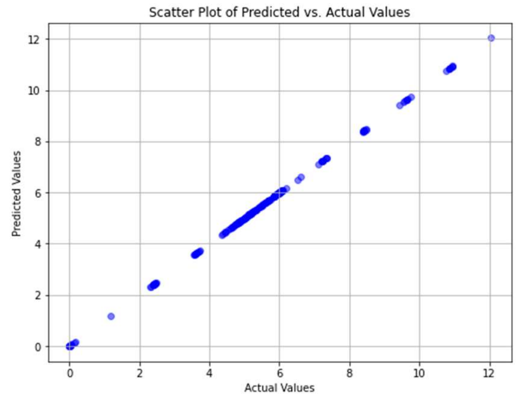

Fig. 7: Scatter Plot - Tube Lights.

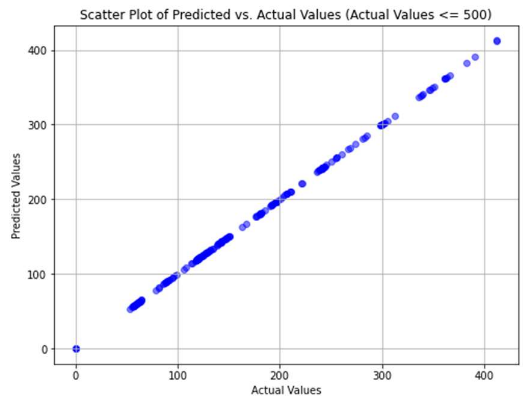

Fig. 8: Scatter Plot - Air-Conditioner.

describes the effect of an increasing number of kilometers on the emissions produced by the two-wheeler. While the intricate shape indicates the significance of two-wheeler usage as an influential feature for predicting emissions, it also highlights the trade-off between capturing complex nonlinear relationships and the interpretability of the model's behavior.

The partial dependence plot in Fig. 6 represents a non-linear relationship between four-wheeler usage in kilometers and emission values, with an overall positive correlation and distinct patterns or regimes based on the fuel type (petrol or diesel), where the curves for petrol and diesel vehicles diverge at certain points, suggesting interactions between the features of four-wheeler usage and fuel type in influencing emission values. The plot captures complex nonlinear relationships and feature interactions, whereas the varying slopes and curvatures highlight the model's ability to capture these interactions.

Inferences Drawn From Tube-Light and Air Conditioner Plots

The plot in Fig. 7 is a scatter plot showcasing the differences between the actual and predicted emission values obtained from training the Linear Regression model on tube-light usage in hours and its corresponding emissions in kgCO2e. A scatter plot was chosen to analyze the input-output relationship of the variables because it provides a clear representation of the data points and their alignment with the predicted output data points. The simplicity of the scatter plot in terms of visualization provides a clear rationale for employing it over a partial dependence plot. The scatter points follow a linear trend, indicating that the Linear Regression model reasonably captures the relationship between the input features and target variables, indicating the overall good fit of the model. Points closer to the diagonal indicate that the model is more accurate in its predictions.

The plot in Fig. 8 is a scatter plot showcasing the differences between the actual and predicted emission values obtained by training the Linear Regression model on Air Conditioner usage data in hours and its corresponding emissions in kgCO2e. The scatter points follow a linear trend, indicating the Linear Regression model's ability to capture the relationship between the input and target features, inferring an overall good fit. The scatter points are

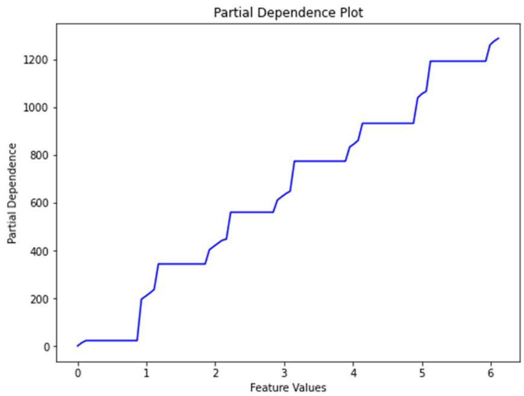

Fig. 9: Partial Dependency Plot - Washing Machine.

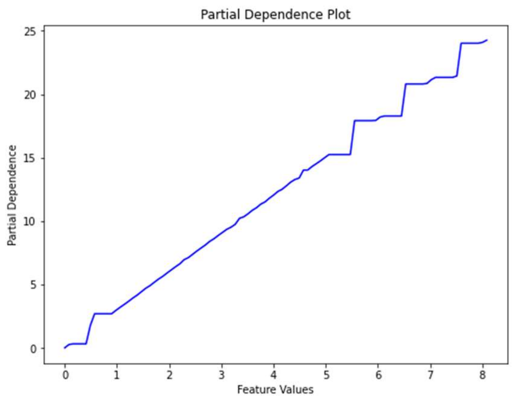

Fig. 10: Partial Dependency Plot - Ceiling Fan.

uniformly distributed, indicating that the model was trained and tested using a large variety of data points. Points closer to the diagonal indicate that the model makes more accurate forecasts.

Inferences Drawn from the Washing Machine and Ceiling Fan Plots

The plot in Fig. 9 represents a partial dependency plot obtained from Gradient Boost Regression after training it on usage data in hours and the corresponding emission values of a Washing Machine. The plot displays an overall increasing trend, indicating a positive correlation between the amount of usage and emission values. The curve shows an increase in carbon emissions at higher fan usage hours, suggesting that longer washing machine usage times have a significantly greater impact on emissions. The plot displays a step-wise pattern, which is characteristic of tree-based models, such as Gradient Boost Regression.

This pattern suggests that the model captures different regimes or thresholds in the relationship between washing machine usage hours and emissions.

The plot in Fig. 10 represents a partial dependency plot obtained from Random Forest Regression after training it on usage data in hours and the corresponding emission values of a ceiling fan. The plot displays a monotonically increasing trend, indicating a positive correlation between the number of hours of use and the emission values. The curve shows a steeper increase in carbon emissions at higher fan usage hours, suggesting that longer fan usage times have a disproportionately large impact on emissions.

Results from Reinforcement Learning Models

The following plots contain three distinct parameters of the reinforcement learning model that drive the performance of the model during real-time deployment.

The following metrics were graphically plotted to examine the performance of the reinforcement learning model:

Steps

Variation

Consumption Dictionary Values

These metrics are defined as follows:

Steps: The steps are represented by the number of iterations performed by the agent to reach the goal consumption in each episode.

Variation: The precision of the findings with regard to the deviation from the intended state and the calculated overall consumption was described by the variation between the final and the goal consumption. The lower the variation value, the better the model performs, indicating a more successful implementation of the proposed system.

Consumption Dictionary Values: Consumption dictionary values refer to the values for each consumption state listed in the consumption dictionary. These values provide a

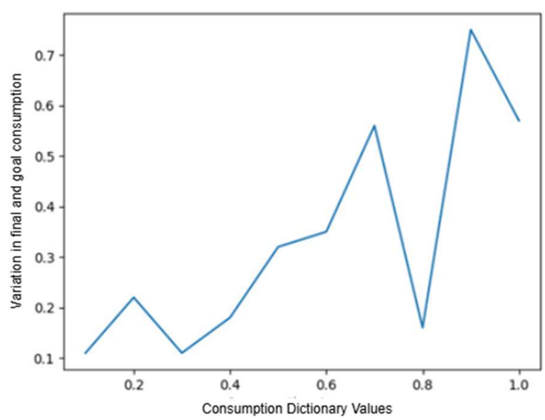

Fig. 11: Line Plot - Steps v/s Consumption Dictionary Values.

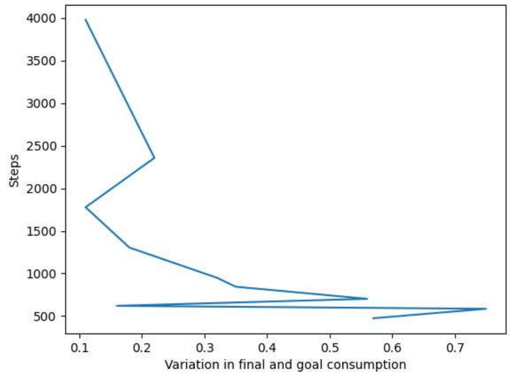

Fig. 12: Line Plot - Steps v/s Variation in final and goal state consumption.

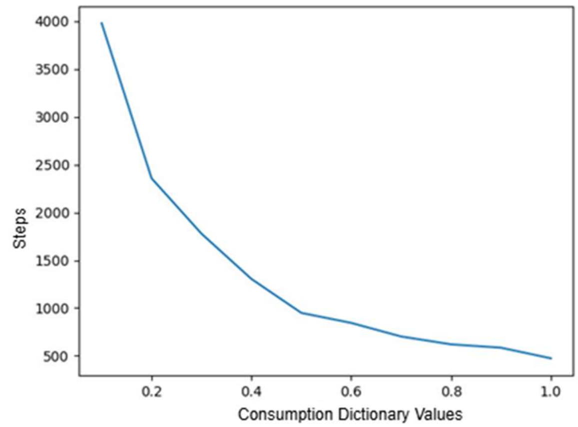

Fig. 13: Line Plot - Consumption Dictionary Values v/s Variation in final and goal state consumption.

relative factor of consumption between the existing states and represent a unit of consumption consumed by the agent in that state in one iteration. The consumption dictionary has been represented as follows: {"Zero": 0, "Low Temperature": 1, "Moderate Temperature": 2, "Coldest Temperature": 5}

Steps vs Consumption Dictionary Values Plot

The horizontal axis represents a subset of the possible values of the consumption dictionary, and the vertical axis represents the number of steps taken to achieve the values in the matrix shown in Fig. 11. A decreasing curve describes the relationship between the consumption dictionary values and the number of steps taken to achieve them. Lower consumption values in the dictionary lead to lower units of consumption per iteration of learning. Therefore, a greater number of steps were taken to achieve the final goal state.

Steps vs Variation in the Final and Goal Consumption Plot

The graph in Fig. 12 shows the number of steps taken versus the variation in the final and goal state consumptions. This variation results from the absolute difference between the final and goal state consumption. A higher number of steps indicates a lower variation between the final and goal consumption. However, there is a trade-off between the accuracy and the time taken for learning in a given episode. A relatively more accurate result is achieved when the algorithm spends more time reaching the goal state and vice versa. The sudden steepness and bends in the graph indicate the randomness of the algorithm, which can be attributed to the exploration-exploitation trade-off in Q-learning.

Variation in Final and Goal Consumption vs Consumption Dictionary Values Plot

Fig. 13 shows the variation in the final and goal state consumptions with respect to the consumption dictionary values. Smaller values of in the consumption dictionary

Table 6: Key Results obtained using the ConForMist framework.

Users	Device	Input [$km.h^{-1}$]	ML model output [kgCO2e]	Offset[%]	Phase of the day	User Preference [%]	RL model output
User 1	Tube light	5	6.02	5	Morning, Night	53, 74	2h 04m, 2h 40m
User 2	Ceiling fan	8	24.08	10	Afternoon, Evening	27, 80	1h 50m, 5h 22m
User 3	Washing machine	3	632.22	15	Evening, Night	90, 54	1h 25m, 0h 59m
User 4	Air Conditioner	9	541.89	10	Morning, Night	67, 61	4h 22m, 3h 38m
User 5	2-wheeler	20	249.56	15	Morning, Evening	42, 63	9.26km, 6.25km
User 6	4-wheeler	50	3169	5	Morning, Afternoon	78, 13	39.54km, 8.14km
User 7	4-wheeler	30	2249.97	10	Afternoon, Night	24, 70	8.33km, 22.6km
User 8	Air Conditioner	7	421.47	15	Morning, Evening	80, 49	3h 21m, 2h 37m
User 9	Washing Machine	2	421.48	5	Afternoon, Evening	55, 51	1h 02m, 0h 59m
User 10	Ceiling fan	10	1288.41	10	Evening, Night	63, 25	3h 30m, 1h 51m
User 11	Tubelight	6	7.22	15	Afternoon, Night	69, 49	2h 58m, 2h 07m
User 12	2-wheeler	15	189.81	10	Morning, Afternoon	50, 45	6.67km, 5.33km

Table 7: Average Monthly Electricity Consumption for a household in Bangalore city.

Appliance Type	Consensus Monthly Consumption (Units)	Consensus Monthly Carbon Emission [kgCo2e].
Lighting	11	9.02
Heating	0	0
Cooling	67	54.94
Appliance	45	36.9

Table 8: ConForMist Monthly Carbon Emissions Benchmarked Against Consensus Table Values.

Appliance Type	User Monthly Aggregated Carbon Emission [KgCo2e]	Variance Between Consensus data v/s ConForMist data
Lighting	10	0.98
Heating	0	0
Cooling	55	0.06
Appliance	38	1.1

will lead to smaller differences between the final and goal consumptions. The choice of consumption values in the dictionary is critical to the use case presented in this study. The objective is to reach a balanced value that is appropriate for the goal state, the device being used, and the level of accuracy required. The deployment of this model requires the values in the consumption dictionary to be optimized to achieve the desired performance.

The key results after user interaction with the ConForMist framework are presented in Table 6.

Benchmarking and Validation with Real-World Data

Data from the ConForMist framework has been compared with real-world data provided in a survey report (Hernandez et al. 2022). Table 7 provides a benchmark for validating real-world data.

The data in Table 8 provides the user with monthly aggregated carbon emissions and the variance between the consensus and ConForMist data. The benchmarked results show that the variance is minimal, indicating the efficacy of the proposed system.

The User Interface

GreenStride is a web-based tool developed to comprehensively record usage data across various carbon-emitter categories and elicit recommendations provided by the underlying machine learning data gathered. The developed front-end user interface allows two types of users to register and log in.

The functionalities implemented in the web-based tool are shown in Figs. 14–18. The key features of the web-based tool are listed in Table 9.

CONCLUSIONS

The proposed framework contributes to several areas.

- Data Gathering: The proposed system presents a simple approach to gathering data from primary carbon emitter sources (exemplar ESP32-based hardware as the foundation, chosen for its computational and IoT

Table 9: Website UI Functionality Description Table.

Key Features	Description
User Roles and Analytics Features in the Emissions Monitoring Application	The application supports two distinct user types. The first comprises organizational or governmental entities that can monitor emissions and resource usage across regions such as constituencies, cities, or districts. These users gain insights through visual representations of emission and usage patterns over time. The second type includes individual users who can log the emissions and consumption associated with their appliances and vehicles. They receive personalized visual analytics, including charts and graphs, depicting their usage trends over defined time periods.
Emission Tracking and Optimization Features for Individual Users	Individual users can specify the type of vehicle or appliance they use, along with the usage windows (in km or hours) and the date of usage to calculate the corresponding emissions in kilograms of CO_2 equivalent ($KgCO_2e$). The latter feature is represented in Fig. 14. Users also have the option to reduce their estimated emissions by 5%, 10%, or 15%, upon which the system provides time-of-day–specific recommendations (morning, afternoon, evening, and night) for optimized usage aimed at lowering emissions, which are represented in Fig. 15 and Fig. 16.
Interactive Dashboard and Data Visualization for Emission Insights	Users are provided with a dashboard that displays comprehensive statistics, including average, minimum, maximum, and total emissions and usage data across all recorded vehicles and appliances. The platform enables visualization of usage and emission patterns through various filters such as month, appliance type, and vehicle category. Product-specific statistics through radar charts, as shown in Fig. 17, monthly trends via area charts as shown in Fig. 18 and daily statistics are presented using line charts as shown in Fig. 19, facilitating a multi-dimensional understanding of consumption and emission behavior.
Advanced Dashboard and Regional Analytics for Governmental and Organizational Users	Governmental and organizational users have access to a dedicated dashboard that presents aggregated emissions and usage patterns. An additional location-based filter enables these users to visualize statistics specific to a selected region. Furthermore, an extended radar chart is provided as shown in Fig. 20 to depict average, minimum, and maximum emissions and usage metrics across all monitored locations, facilitating comparative analysis and informed decision-making.

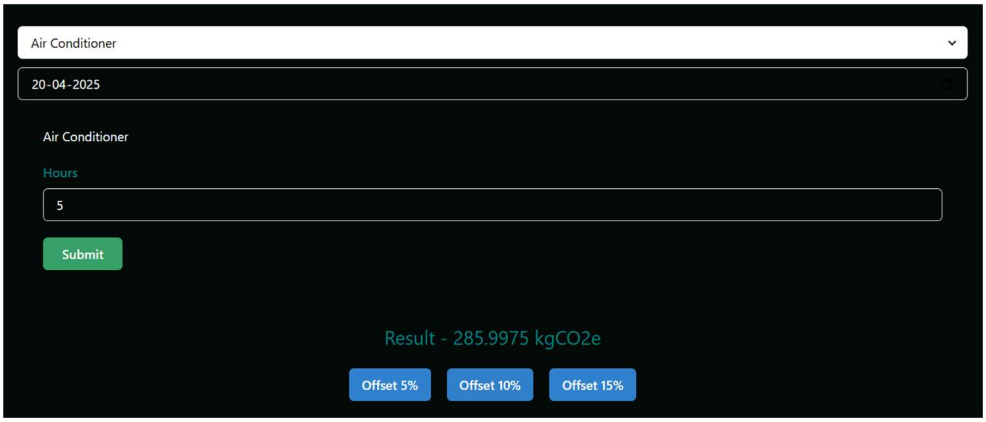

Fig. 14: ConForMiSt Web User Interface (UI)–Emission calculator based on usage.

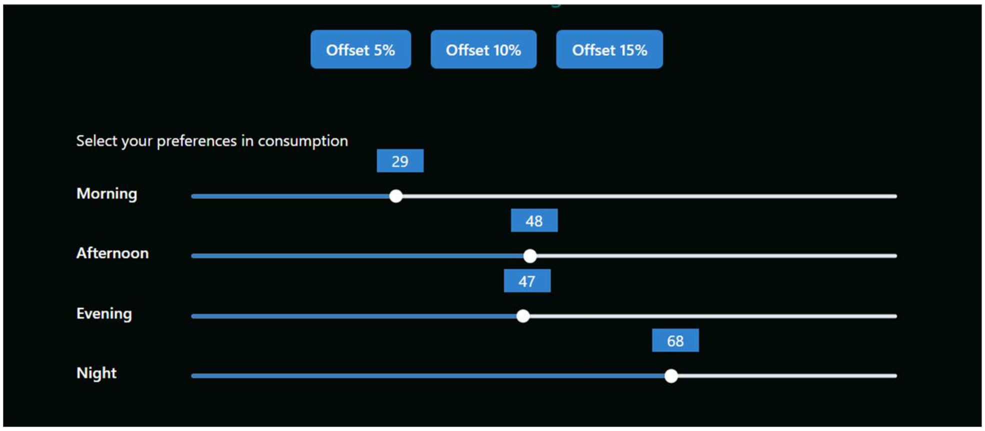

Fig. 15: ConForMiSt UI Dashboard - User preferences for usage recommendations.

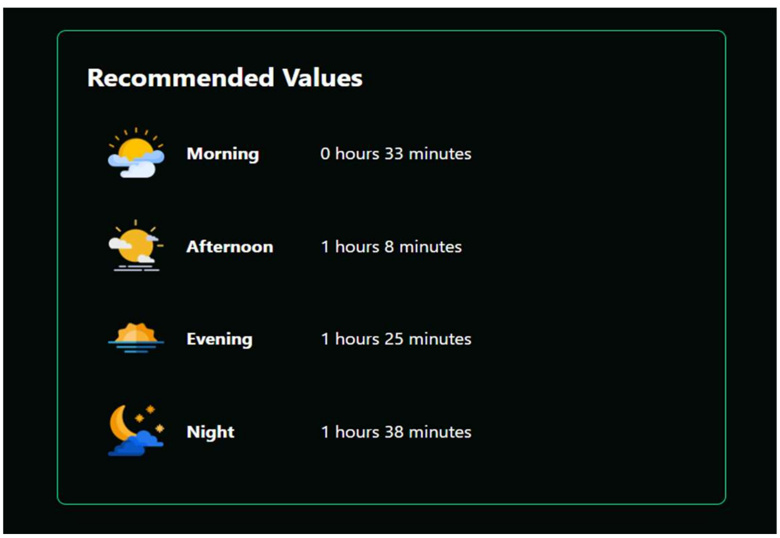

Fig. 16: ConForMiSt UI Dashboard - Usage recommendation for lowering emissions

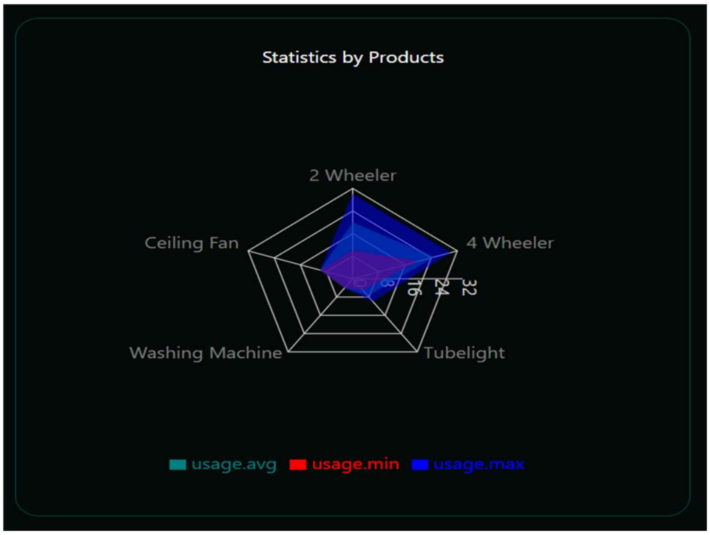

Fig. 17: ConForMiSt UI Dashboard - Carbon Footprint Radar Plots for Appliances

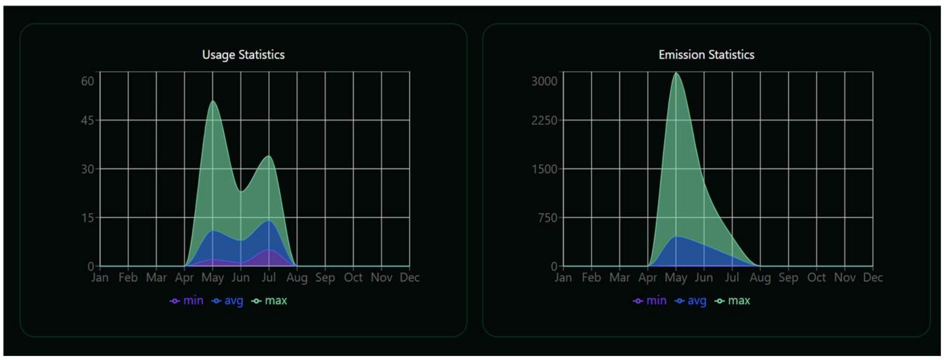

Fig. 18: ConForMiSt UI Dashboard - Recommendations (Appliances)

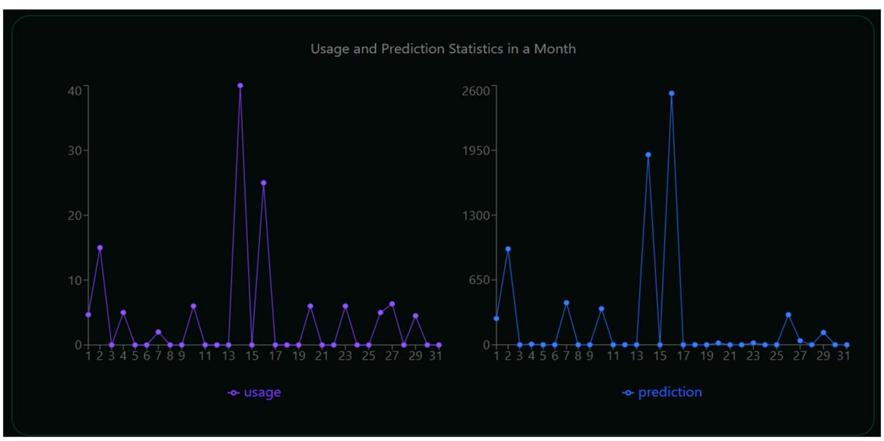

Fig. 19: ConForMiSt UI Dashboard - Line charts for usage and emission patterns for a month

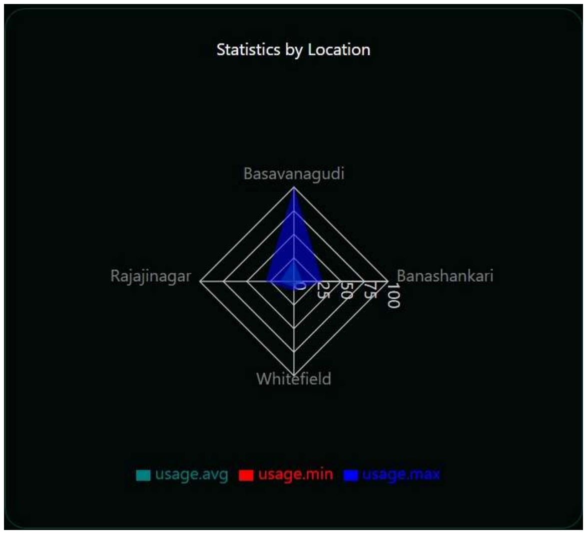

Fig. 20: ConForMiSt UI Dashboard – Radar chart representing location-wise statistics

features) and recording real-time data in a database enabled by Apache Kafka.

- Framework Design: A dual-phase framework: The design provides separation of concerns: the first stage is to understand the data, and the second stage is to drive recommendations with a model-free approach.
- The first stage uses various ML models operating on a structured dataset to elicit insights regarding carbon emission data. The models selected in this stage used the RMSE performance index.
- In the second stage, a reinforcement learning model that uniquely models states and actions for various consumption categories drives the recommendation process. The unique state modeling that captures many aspects is a novelty.
- Experimentation and Result Interpretation: Inferences drawn from two exemplar categories, namely, electrical appliances and transport, using scatter and partial dependence plots.
- User Interface (UI) presentation: These models are used to provide recommendations that will influence future consumption judiciously and also mitigate carbon emissions. Options exist to analyze current consumption and emissions data (Eg, Radar plots, statistical data)
- Comparison with other frameworks: The proposed ConForMiSt was evaluated against models proposed for similar goals using deep learning, RL, deep Q-learning, and LSTM with IoT. However, ConForMiSt has a broader scope for analyzing different categories of data and provides a mechanism for aggregating data for higher organizational levels.
- An important aspect of hyperparameter (consumption) optimization was elucidated.

LIMITATIONS

- Relatively limited data sample size used for training and testing: Machine Learning models operated on consumption and corresponding emission data for the span of a year, which was relatively small.
- A larger sample size spanning multiple years would help reduce the Mean Absolute Errors and increase the accuracy of the models, thus achieving a better fit. Although the sample size considered gathering data in different time slots of a day for a multitude of carbon emission producers, it did not consider demographic factors such as age, gender, and location of the consumers.
- Incorporating parameters such as age, sex, and emission source can greatly help in detecting hidden patterns and improving predictions from Machine Learning models by providing better context.

FUTURE WORK

Smart home systems and designs can incorporate the implementation proposed in this study to encourage consumers to use resources effectively. Smart thermostats, lighting controls, and automated energy-efficient appliance control systems can be implemented to monitor and reduce energy usage and the resulting carbon footprint. Federated machine learning models can be trained and deployed across various devices to make predictions and provide recommendations, thereby eliminating the need for a centralized training and deployment station. Edge Computing technologies can be paired with the sensor implementation proposed in this study for improved real-time data collection and processing.

REFERENCES

Ali, J., Khan, R., Ahmad, N. and Maqsood, I., 2012. Random forests and decision trees. *International Journal of Computer Science Issues (IJCSI)*, 9(5), pp.272-278.

Babiuch, M. and Postulka, J., 2020. Smart home monitoring system using ESP32 microcontrollers. *Internet of Things*, 12, 100314.

Bahrami, M., Singhal, S. and Singhal, M., 2018. API learning: Applying machine learning to manage the rise of the API economy. *Companion Proceedings of the Web Conference 2018. International World Wide Web Conferences Steering Committee*, pp.1591-1598.

Bellman, R., 1957. A Markovian decision process. *Journal of Mathematics and Mechanics*, 6(5), pp.679-684.

Bhatti, N. and Iqbal, M., 2000. Object serialisation and deserialisation using XML. *10th International Conference on Management of Data (COMAD 2000)*.

Chandel, T.A., 2022. Green entrepreneurship and sustainable development. International Perspectives on Value Creation and Sustainability Through Social Entrepreneurship. *IGI Global*, pp.173-208.

Dong, H. and Zhang, L., 2023. Transition towards carbon neutrality: Forecasting Hong Kong's buildings' carbon footprint by 2050 using a machine learning approach. *Sustainable Production and Consumption*, 35, pp.633-642.

Gao, T., Liu, Q. and Wang, J., 2014. A comparative study of carbon footprint and assessment standards. *International Journal of Low-Carbon Technologies*, 9(3), pp.237-243.

Goswami, M., Premkumar, A., Kishan, R. and Nautiyal, S., 2024. A methodological perspective on inclusive assessment of household carbon footprint. *Environment, Development and Sustainability*, pp.1-28.

Henderson, P., Hu, J. and Romoff, J., 2020. Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research*, 21(248), pp.1-43.

Hernandez, M., Bhagavatula, K., Cibi, S., R., K., Krishnan, S., Malaviya, S. and Jairaj, B., 2022. Shifting household energy use in Bangalore, India: Using behaviorally informed energy reports. *Working Paper*. Washington, DC: World Resources Institute.

Hiraman, B.R., 2018. A study of Apache Kafka in big data stream processing. *2018 International Conference on Information, Communication, Engineering and Technology (ICICET)*. IEEE, pp.1-4.

Huang, L., 2023. Sustainable lifestyle: Urban household carbon footprint accounting and policy implications for lifestyle-based decarbonization. *Energy Policy*, 181, p.113696.

Jain, M. and Nagpal, A., 2021. Examining the role of principal contributors to 'sectoral carbon emissions': In the Indian context. *Vision*, 25(4), pp.399-406.

Jones, M.W., Peters, G.P., Gasser, T., Andrew, R.M., Schwingshackl, C., Gütschow, J., Houghton, R.A., Friedlingstein, P., Pongratz, J. and Le Quéré, C., 2023. National contributions to climate change due to historical emissions of carbon dioxide, methane, and nitrous oxide since 1850. *Scientific Data*, 10(1), p.155.

Khan, A.A., 2024. An analytical study on the impact of global warming: Effects on the environment. *Recent Trends in Commerce, Management, Accountancy and Business Economics*, p.54.

Kiehbadroudinezhad, M., Merabet, A. and Hosseinzadeh-Bandbafha, H., 2024. Health impacts of greenhouse gas emissions on humans and the environment. *Advances and Technology Development in Greenhouse Gases: Emission, Capture and Conversion. Elsevier*, pp.265-291.

Kumar, A., Mishra, S., Bakshi, S., Upadhyay, P. and Thakur, T.K., 2023. Response of eutrophication and water quality drivers on greenhouse gas emissions in lakes of China: A critical analysis. *Ecohydrology*, 16(1), e2483.

Lee, S. and Choi, D.H., 2019. Reinforcement learning-based energy management of a smart home with a rooftop solar photovoltaic system, an energy storage system, and home appliances. *Sensors*, 19(3937).

Lissa, P., Deane, C., Schukat, M., Seri, F., Keane, M. and Barrett, E., 2021. Deep reinforcement learning for home energy management system control. *Energy and AI*, 3, p.100043.

Liu, H., Wong, W.-K., Phan, T.C., Nassani, A.A., Haffar, M. and Abu-Rumman, A., 2023. Linkage among urbanization, energy consumption, economic growth and carbon emissions. Panel data analysis for China using the ARDL model. *Fuel*, 332, 126122.

Lövehagen, N., Malmodin, J. and Bergmark, P., 2023. Assessing embodied carbon emissions of communication user devices by combining approaches. *Renewable and Sustainable Energy Reviews*, 183, 113422.

Mahmoud, M. and Ben Slama, S., 2025. Deep learning-based home energy management incorporating vehicle-to-home and home-to-vehicle technologies for renewable integration. *Energies*, 18(129).

Mancini, M.S., Galli, A., Niccolucci, V., Lin, D., Hanscom, L. and Marchettini, N., 2016. Ecological footprint: refining the carbon footprint calculation. *Ecological Indicators*, 61, pp.390-403.

Maulud, D.H. and Abdulazeez, A.M., 2020. A review of linear regression in machine learning. *Journal of Applied Science and Technology Trend*s, 1(2), pp.140-147.

Musa, M., Rahman, M.M., Khan, M.A. and Alam, K., 2024. Sustainable development challenges in Bangladesh: an empirical study of economic growth, industrialization, energy consumption, foreign investment, and carbon emissions—using dynamic ARDL model and frequency domain causality approach. *Clean Technologies and Environmental Policy*, 26(6), pp.1799-1823.

Narayanan, P.K., 2024. Engineering real-time data pipelines using Apache Kafka. *Data Engineering for Machine Learning Pipelines: From Python Libraries to ML Pipelines and Cloud Platforms. Berkeley,* CA: Apress, pp.277-322.

Nguyen, Q., Diaz-Rainey, I. and Kuruppuarachchi, D., 2021. Predicting corporate carbon footprints for climate finance risk analyses: a machine learning approach. *Energy Economics*, 95, p10512

Onat, N.C., Kucukvar, M. and Tatari, O., 2015. Conventional, hybrid, plug-in hybrid or electric vehicles? State-based comparative carbon and energy footprint analyses in the United States. *Applied Energy*, 150, pp.36-49.

Otchere, D.A., Agyekum, E.B. and Adjei, E., 2022. Application of the gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions. *Journal of Petroleum Science and Engineering*, 208, 109244.

Ramachandra, T.V., Aithal, B.H. and Sreejith, K., 2015. GHG footprint of major cities in India. *Renewable and Sustainable Energy Reviews*, 44, pp.473-495.

Ranjan, G.S.K., Verma, A.K. and Radhika, S., 2019. K-nearest neighbors and grid search CV based real-time fault monitoring system for industries. *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*. IEEE, pp.1-5.

Raptis, T.P. and Passarella, A., 2022. On efficiently partitioning a topic in Apache Kafka. *2022 International Conference on Computer, Information and Telecommunication Systems (CITS)*. IEEE, pp.1-5.

Rondoni, A. and Grasso, S., 2021. Consumers' behaviour towards carbon footprint labels on food: A review of the literature and discussion of industry implications. *Journal of Cleaner Production*, 301, p.127031.

Schwenkenbecher, A., 2014. Is there an obligation to reduce one's individual carbon footprint? Critical *Review of International Social and Political Philosophy,* 17(2), pp.168-188.

Solatidehkordi, Z., Ramesh, J., Al-Ali, A.R., Osman, A. and Shaaban, M., 2023. An IoT deep learning-based home appliances management and classification system. *Energy Reports, 9(Supplement 3),* pp.503–509.

Thein, K.M.M., 2014. Apache Kafka: Next-generation distributed messaging system. *International Journal of Scientific Engineering and Technology Research*, 3(47), pp.9478-9483.

Ulfat, W. and Noori, S., 2024. The climate and global warming responses to the changing greenhouse gas concentration in the atmosphere. *Nangarhar University International Journal of Biosciences*, pp.545-548.

Varoquaux, G. and Colliot, O., 2023. Evaluating machine learning models and their diagnostic value. *Machine Learning for Brain Disorders*, pp.601-630.

Watkins, C.J.C.H. and Dayan, P., 1992. Q-learning. *Machine Learning*, 8, pp.279-292.

Wu, H., Shang, Z. and Wolter, K., 2019. Performance prediction for the Apache Kafka messaging system. *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. IEEE, pp.1931-1938.