Monitoring and predicting appliance emission using Machine Learning Approachs

Mr. Sanket Mane[1], Mr. Kaustubh Kumbharkar[1], Mr. Samarth Mahimkar[1], Dr. Sarika Patil[2], Prof. Nilesh G. Bhoskar[3]

[1]*UG Students, Department of Electronics and Telecommunication Engineering, Sinhgad College of Engineering, Vadgaon, Pune – 411046*

[2]*Assistant Professor, Department of Electrical and Electronics Engineering, MIT WPU Kothrud, Pune – 411046*

[3]*Assistant Professor, Department of Electronics and Telecommunication Engineering, Sinhgad College of Engineering, Vadgaon, Pune – 411046*

***Abstract*— With the growing concerns about environmental pollution and energy efficiency, monitoring and predicting appliance emissions have become critical in ensuring sustainable resource utilization. Traditional emission tracking methods rely on periodic assessments and manual monitoring, which are often inefficient, time-consuming, and lack real-time adaptability. To address these challenges, this study explores the potential of machine learning (ML) approaches to enhance the accuracy and efficiency of appliance emission monitoring and prediction.**

The proposed system leverages data from sensors and appliance usage records to analyze emission patterns. Various machine learning models, including regression techniques, decision trees, and deep learning frameworks, are implemented to predict future emission levels. Feature engineering and data preprocessing techniques are employed to improve model accuracy. Performance evaluation is conducted using key metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R^2) scores to compare the effectiveness of different models.

Experimental results demonstrate that machine learning-based approaches significantly outperform conventional methods in predicting emissions with higher accuracy and adaptability. The study also highlights the challenges associated with data inconsistencies, sensor calibration, and real-time processing. The findings contribute to the development of intelligent, automated, and data-driven solutions for emission control, aiding policymakers, industries, and researchers in reducing environmental impact and promoting sustainable energy practices.

Keywords: Emission Monitoring, Machine Learning, Prediction Models, Environmental Sustainability, Smart Systems.

I. INTRODUCTION

The increasing energy consumption and environmental concerns have led to a need for efficient methods for monitoring and predicting appliance emissions. Traditional methods are inefficient and costly, requiring intelligent, automated solutions. Machine learning (ML) can be used to analyze complex datasets, identify patterns, and make accurate predictions. This study explores the application of ML approaches for monitoring and predicting appliance emissions, aiming to enhance accuracy and efficiency while addressing challenges like data variability and real-time processing.

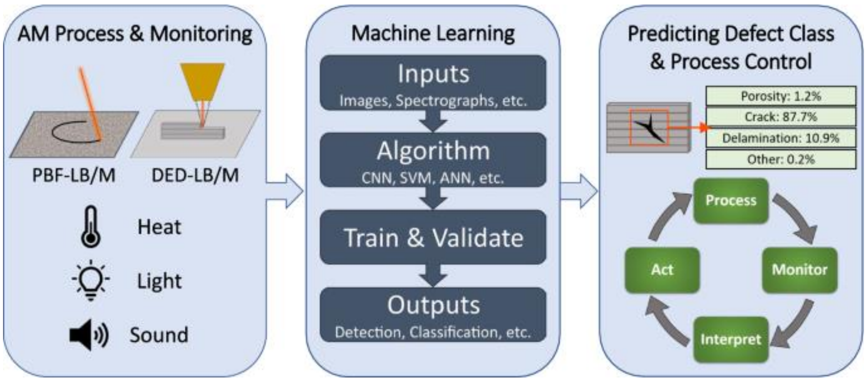

Figure 1: Process monitoring and machine learning for defect detection process

The project aims to develop a machine learning system that uses data analytics and artificial intelligence to monitor and predict emissions from household and industrial appliances. This system will use historical and real-time data to detect high-emission activities and assist users in making informed decisions to minimize their environmental impact. The integration of machine learning into emission management could transform energy usage behavior and support global efforts towards a cleaner, greener future, enhancing environmental sustainability and energy efficiency.

II. LITERATURE REVIEW

Roupen Minassian et al. [1] presented a comprehensive study on deep learning models for predicting indoor environmental quality in smart buildings found that Convolutional Neural Network (CNN) models outperformed Long Short-Term Memory (LSTM) and hybrid CNN-LSTM models in predicting indoor temperature. Multivariate input configurations and capturing complex interactions between environmental parameters were crucial for accurate predictions. Temperature, humidity, and HVAC status were identified as key influential features. The study provides practical guidelines for model implementation, contributing to the development of more efficient smart building management systems and potentially improving energy efficiency and occupant comfort.

Connor McGookin et al.[2] addresses the need for integrating diverse perspectives into energy system modeling to inform fairer and more effective climate policies. It provides good practice guidance for participatory energy systems modeling, offering a flexible framework for stakeholder and public involvement. The framework highlights multiple entry points for participation and poses key questions to steer the process, while also identifying crucial challenges and areas for future research in this field.

Montaser N.A. et al. [5] A real-time air pollution monitoring and forecasting system was developed for the chrome plating industry, utilizing IoT sensors and AI models (LSTM, Random Forest, and Linear Regression) to detect and predict pollutant levels. The system achieved high accuracy in forecasting temperature, humidity, and PM2.5 levels, enabling proactive measures to improve air quality. By predicting pollution levels and activating exhaust fans, the system demonstrates significant potential for dynamic responses to pollution and improving industrial air quality.

Spyros Giannelos et al. [7] compares various Machine Learning (ML) approaches for predicting long-term CO2 emissions from buildings until 2050, including Linear Regression, ARIMA, Shallow Neural Networks, and Deep Neural Networks. The analysis is conducted for multiple regions worldwide, evaluating predictive performance using univariate and multivariate modeling with different feature extraction methods. The study aims to identify effective ML-based approaches for forecasting building-related CO2 emissions.

Michael Hans et al.[18]developed a predictive analytics model using the SVR algorithm to forecast university carbon emissions, with a focus on students' learning activities as a significant factor. Using Institut Teknologi Bandung as a case study, the model incorporated historical data and external information to predict future trends. The study provides insights for universities to assess their carbon footprint and raises awareness among the academic community, informing decision-making to optimize carbon emissions.

Research Gap Identified

- Limited Real-Time Monitoring – Many systems lack real-time adaptability, making proactive control difficult.
- Data Quality Issues – Incomplete, noisy, and inconsistent datasets affect prediction accuracy.
- Lack of Standardized Features – Varying feature selection across studies hinders the development of a generalized ML framework.
- Integration with IoT and Edge Computing – Real-time processing at the edge level remains underexplored.
- Model Interpretability Challenges – Black-box ML models limit transparency and trust in predictions.
- Limited Research on Household Emissions – Most studies focus on industrial emissions, neglecting small-scale applications.
- Energy Consumption of AI Models – High computational demands contradict the goal of energy-efficient solutions.

Research shows that machine learning (ML) improves emission monitoring by enabling real-time tracking and accurate predictions compared to traditional methods. ML models like regression, decision trees, and deep learning enhance efficiency but face challenges such as data quality issues, lack of standardization, and high computational demands. Some studies suggest integrating IoT and edge computing for better real-time processing. However, further research is needed to develop scalable, energy-efficient, and interpretable ML models for broader applications.

Problem Statement

This study aims to create a machine learning-based system for monitoring and predicting appliance emissions, enhancing real-time tracking, optimizing energy usage, and providing insights for sustainable energy management.

Objectives

- To design a data-driven system for monitoring appliance emissions.
- To implement machine learning models for predicting future emissions.
- To evaluate the performance of different ML algorithms in emission forecasting.
- To identify the key factors influencing emissions and optimize energy consumption.
- To develop a real-time monitoring framework that can provide actionable insights.

III. RESEARCH METHODOLOGY

The block diagram of the proposed system as shown in Figure 2 below.

Data Collection and Preprocessing

Effective emission monitoring and prediction rely on high-quality data. The data collection process involves gathering emission-related information from sensors, IoT devices, appliance usage logs, and publicly available datasets. Key parameters include CO_2 levels, energy consumption, appliance type, operating time, and environmental factors.

Once collected, the data undergoes preprocessing to ensure accuracy and reliability. This includes:

- Data Cleaning – Removing missing values, duplicates, and inconsistencies.
- Normalization & Scaling – Standardizing data for uniformity across different measurement units.
- Feature Selection – Identifying the most relevant parameters for emission prediction.
- Data Splitting – Dividing data into training and testing sets for machine learning model evaluation.

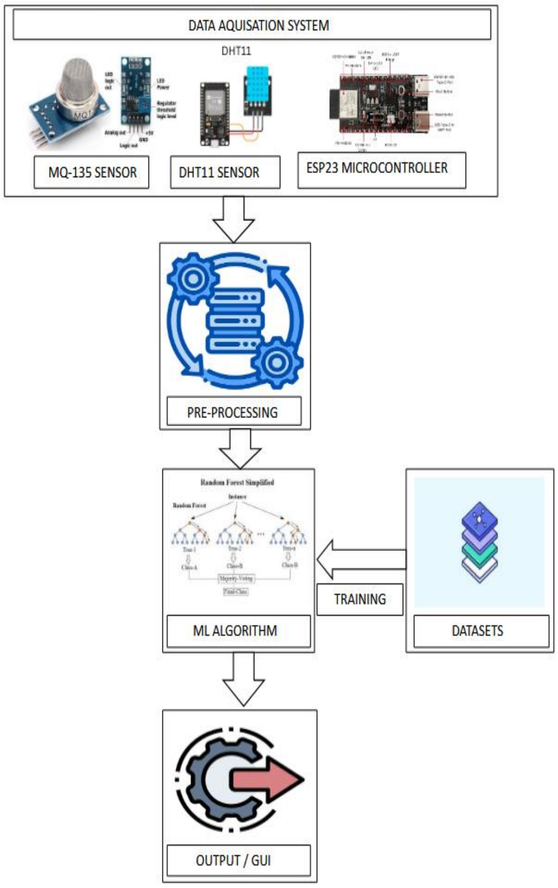

Figure 2: Block Diagram of the Proposed System

Machine Learning Models Used

Various machine learning (ML) models are employed for emission monitoring and prediction, ranging from traditional algorithms to deep learning techniques.

Common models include:

- Linear Regression & Decision Trees – Used for simple emission trend prediction.
- Random Forest & Gradient Boosting – Improve accuracy through ensemble learning.
- Support Vector Machines (SVM) – Effective for classifying emission levels.

- Neural Networks & Deep Learning – Capture complex relationships for highly accurate predictions.

Feature Selection and Engineering

Feature selection and engineering play a crucial role in enhancing model accuracy and reducing complexity.
Key steps include:

- Feature Selection – Identifying relevant parameters (e.g., CO_2 levels, appliance type, energy consumption).
- Dimensionality Reduction – Using techniques like PCA (Principal Component Analysis) to remove redundant features.
- Feature Engineering – Creating new meaningful features (e.g., emission trends over time) to improve prediction performance.

Model Training and Validation

The ML models undergo a structured training and validation process:
- Data Splitting – Dividing data into training (70-80%) and testing (20-30%) sets.
- Cross-Validation – Using k-fold cross-validation to avoid overfitting and improve generalization.
- Hyperparameter Tuning – Adjusting model parameters (e.g., learning rate, tree depth) to enhance performance.

Performance Metrics

To assess model accuracy and efficiency, various performance metrics are used:

- Mean Absolute Error (MAE) – Measures average prediction error.
- Root Mean Square Error (RMSE) – Evaluates prediction deviations.
- R^2 Score (Coefficient of Determination) – Indicates model fit.
- Precision, Recall, & F1-Score – Used for classification models.

Hardware Components

ESP32 Controller

The ESP32 is a low-cost, low-power system-on-chip (SoC) microcontroller with Wi-Fi and Bluetooth capabilities, developed by Espressif Systems. It is widely used in IoT (Internet of Things), smart home automation, robotics, and wireless applications due to its high performance and energy efficiency.

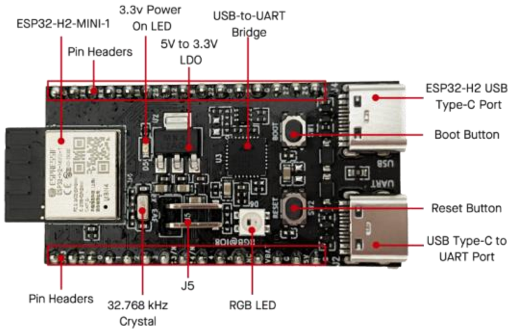

Figure 3: ESP32 Controller Board

- Dual-core 32-bit Xtensa LX6 processor (160–240 MHz)
- Built-in Wi-Fi (802.11 b/g/n) and Bluetooth (v4.2 BLE & Classic)
- Low power consumption with multiple sleep modes
- Multiple GPIOs with support for ADC, DAC, PWM, SPI, I2C, UART
- Capacitive touch sensors, Hall sensor, and temperature sensor
- Supports FreeRTOS, Arduino, Micro Python, and ESP-IDF
- As shown in Fig 3 Above.

MQ-135

The MQ135 is a popular gas sensor used to detect a wide range of harmful gases including ammonia (NH3), nitrogen oxide (NOx), alcohol, benzene, smoke, and carbon dioxide (CO2). It is highly sensitive to air quality and is commonly used in indoor air pollution monitoring systems. The sensor contains a tin dioxide (SnO2) sensing layer that changes its resistance based on the concentration of gases in the air.

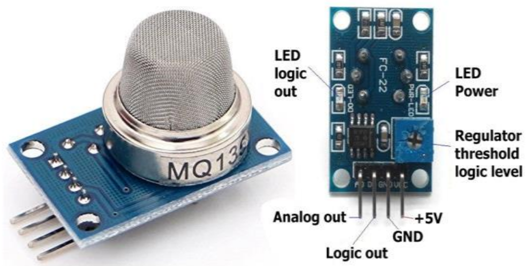

Figure 4: MQ-135 Module

Specifications:

- Gas Detected: NH_3, NO_x, Alcohol, Benzene, Smoke, CO_2
- Operating Voltage: 5V DC
- Load Resistance: 10kΩ (adjustable)
- Sensitivity Range: 10 to 1000 ppm
- Response Time: < 10 seconds
- Preheat Time: 24 hours for optimal performance
- Analog & Digital Output: Analog signal for gas concentration, digital threshold alert
- Working with ESP32:
- The analog output of the MQ135 can be read using the ADC (Analog-to-Digital Converter) pin of the ESP32.
- It requires a preheating period to stabilize before accurate readings can be taken.
- ESP32 processes the voltage change to estimate the gas concentration and can be programmed to send alerts if levels exceed safe limits.
- All information is of Fig 4 Above.

DHT11 Temperature and Humidity Sensor

The DHT11 is a low-cost digital sensor used to measure temperature and humidity. It uses a capacitive humidity sensor and a thermistor to measure the surrounding air and outputs a digital signal on a single data line, making it easy to interface with microcontrollers like ESP32.

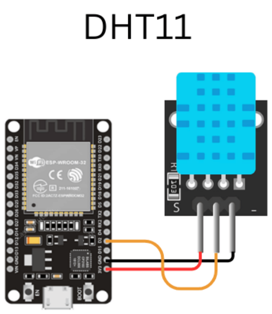

Figure .5: DHT11 Module

Specifications:

- Temperature Range: 0–50°C (±2°C accuracy)
- Humidity Range: 20–90% RH (±5% accuracy)
- Operating Voltage: 3.3V to 5V
- Sampling Rate: 1 Hz (1 reading per second)
- Digital Output: Single-wire serial communication
- Data Format: 40-bit data stream
- Working with ESP32:
- The DHT11 connects to any digital GPIO pin on the ESP32.
- A dedicated DHT library (e.g., Adafruit DHT) is used in code to communicate and extract readings.
- The ESP32 receives temperature and humidity data every second and can use this information for environment monitoring or control systems.
- All information is of Fig 5 Above

Algorithm Implementation

The implementation of the emission monitoring system follows a structured approach using machine learning algorithms. The key steps involved are:

Data Collection
IoT sensors gather real-time emission data.
Data is stored in a centralized database for processing.

Data Preprocessing:
Handling missing values, normalization, and feature extraction.
Removal of outliers to improve model accuracy.

Feature Selection:
Selecting key attributes influencing emissions.
Reducing dimensionality for efficient model training.

Model Selection & Training:
Implementing ML models such as Random Forest, Decision Tree, and Neural Networks.
Training models using historical emission data.

Prediction & Alert Generation:
Running trained models on new data to predict emissions.
Generating alerts if emission levels exceed thresholds.

Visualization & Reporting:
Displaying emission trends and predictions in a dashboard.
Allowing users to take necessary actions based on insights.

Hardware and Software Requirements

Hardware Requirements

- Processor: Intel Core i5/i7 or higher
- RAM: Minimum 8GB (16GB recommended for ML training)
- Storage: 500GB SSD or higher
- Sensors: IoT-based gas sensors (CO2, NOx, SO2, etc.)
- Microcontroller: Raspberry Pi/Arduino for sensor interfacing

Software Requirements

- Operating System: Windows/Linux/macOS
- Programming Languages: Python, R
- ML Libraries: TensorFlow, Scikit-learn, Pandas, NumPy
- Database: MySQL / PostgreSQL / Firebase
- Visualization Tools: Matplotlib, Power BI / Tableau
- Development Tools: Jupiter Notebook, VS Code, PyCharm

IV. DATA COLLECTION AND ANALYSIS

Comparative Analysis of Different Models

A comparative study of different machine learning models is conducted to select the most efficient one. The models considered include:

Table 1: Overview of the Comparison of the Proposed ML Model Performance

Model	Accuracy	MSE	R² Score	Remarks
Linear Regression	Moderate	High	Low	Simple but less accurate for complex data
Decision Tree	High	Moderate	Medium	Performs well but prone to overfitting
Random Forest	Very High	Low	High	Robust and handles complex patterns well
Neural Networks	High	Low	Very High	Best for deep pattern recognition but computationally expensive

Challenges and Limitations

Despite the effectiveness of machine learning models, several challenges and limitations exist:

- Data Quality Issues: Missing, inconsistent, or noisy sensor data can reduce prediction accuracy.
- Computational Complexity: Deep learning models require high processing power and memory.
- Overfitting: Some models perform well on training data but fail on real-world unseen data.
- Sensor Reliability: IoT sensors may experience failures or calibration issues, affecting data accuracy.
- Regulatory Compliance: Ensuring that AI-driven monitoring adheres to environmental regulations and standards.

V. RESULTS AND ANALYSIS

Dataset: The dataset used in this project, appliance_emission_data.csv, contains the following features:

- energy usage: Energy consumed by an appliance.
- temperature: Ambient temperature in the surrounding environment.
- humidity: Humidity levels affecting appliance operation.
- emission (Target): Emissions generated by the appliance.

Data Loading and Preprocessing: The dataset is read using pandas, and relevant features are selected.
Data Splitting: The dataset is split into training and testing sets (80%-20%).

Model Selection: Gradient Boosting Regression is used due to its efficiency in handling complex relationships.
Hyperparameter Tuning: GridSearchCV optimizes key parameters like n_estimators, learning_rate, and max_depth.
Model Evaluation: The best model is evaluated using Mean Squared Error (MSE).
Model Saving: The trained model is stored for future use.

- of prediction accuracy.
- The model was successfully saved for future predictions.

Appliance Emission Data Analysis

Summary of Analysis:
- Energy Usage: 52.08 to 499.87, Avg: 270.62
- Temperature: 20.06 to 39.99, Avg: 30.14
- Humidity: 30.00 to 89.87, Avg: 60.14
- Emission: 33.05 to 291.90, Avg: 156.40

Observation

1. Energy usage and emissions have a strong correlation.
2. Temperature and humidity moderately influence emissions.
3. Emission distribution is slightly right-skewed, indicating a few high-emission cases.

Distribution of Energy Usage

Energy consumption is influenced by factors like appliance type, operational duration, and efficiency. Household and industrial appliances like HVAC systems, lighting, and electronic devices contribute significantly. Power demand fluctuates, with peak usage periods in residential settings and business hours. Real-time data from smart meters and IoT-based monitoring systems helps optimize energy consumption.

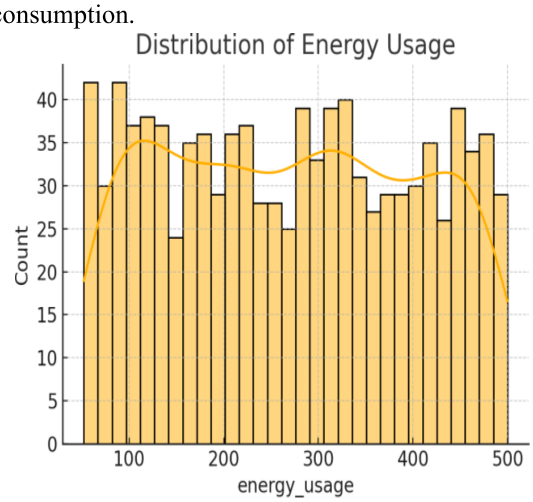

Figure 6: Distribution of Energy Usage

By integrating AI-driven predictions and optimization techniques, users can receive recommendations for energy-efficient appliance usage, ultimately leading to reduced power wastage, lower electricity costs, and a significant reduction in carbon emissions as shown in Fig 6 above. This proactive approach ensures a sustainable and efficient energy distribution system, supporting both environmental and economic benefits.

Distribution of Temperature

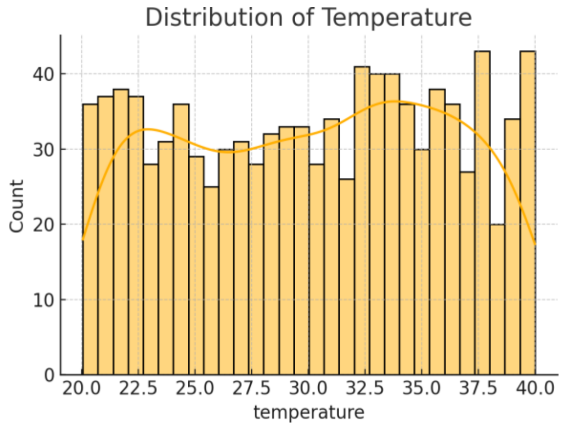

Figure 7: Distribution of Temperature

Temperature distribution is crucial for monitoring and predicting appliance emissions using machine learning. It varies based on factors like appliance type, operational load, and heat dissipation mechanisms. Uneven temperature distribution can indicate inefficient energy usage, equipment failures, or excessive emissions. Real-time temperature data can be collected and analyzed using thermal sensors, infrared imaging, and IoT-based systems. Machine learning models can identify abnormal temperature patterns, predict overheating risks, and optimize energy consumption. This approach ensures sustainable and reliable systems as shown in Fig 7 Above.

Distribution of Humidity

Humidity distribution is crucial for monitoring and predicting appliance emissions using machine learning. It affects energy efficiency, emission levels, and appliance performance. High humidity can increase cooling system energy consumption, while low humidity can cause static electricity buildup and damage electronic devices. Uneven humidity distribution can lead to mold growth, corrosion, and increased emissions.

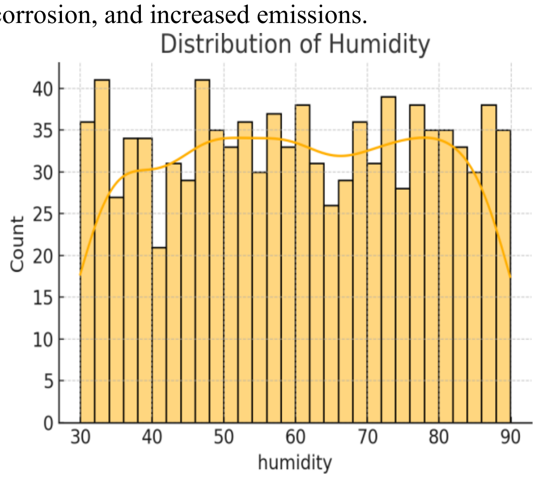

Figure 8: Distribution of Humidity

By utilizing humidity sensors, IoT-based monitoring systems, and smart HVAC controls, real-time moisture data can be collected and analyzed. Machine learning models can process this data to identify patterns in humidity distribution, predict fluctuations, and optimize appliance operation based on environmental conditions As shown in fig 8 above. Algorithms such as regression models, clustering techniques, and deep learning networks can be used to analyze humidity trends and their correlation with power consumption and emissions.

Distribution of Humidity

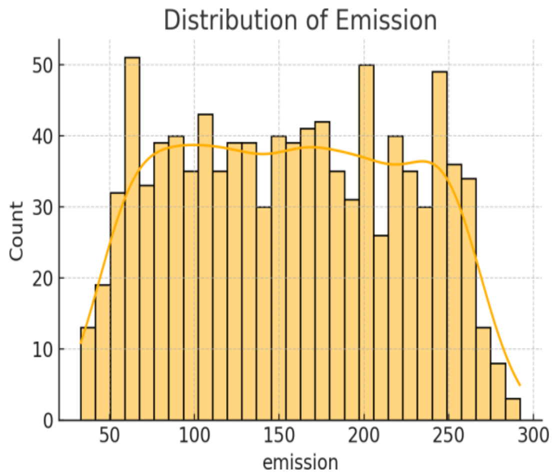

Figure 9: Distribution of Emission

Emission distribution refers to the variation in pollutants and gases from appliances and electrical equipment across different conditions and time periods. Understanding this distribution is crucial for analyzing energy usage, optimizing performance, and developing strategies to reduce harmful emissions. Factors like appliance type, energy source, operational efficiency, load conditions, and environmental variables influence emissions as shown in fig 9 given above.

AI-driven optimization and control mechanisms in appliances can reduce emissions, improve efficiency, and maintain a sustainable energy footprint. Smart grid systems, automated demand-side management, and predictive maintenance reduce energy wastage.

Feature Correlation Heatmap

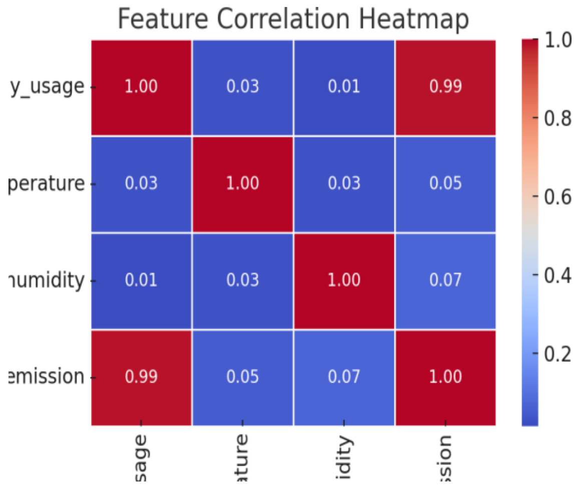

Figure 10: Feature Correlation Heatmap

A Feature Correlation Heatmap is a graphical representation of the relationships between different variables in a dataset. In the context of monitoring and predicting appliance emissions using machine learning, this heatmap helps identify how various factors such as power consumption, temperature, humidity, energy usage, and emission levels are interrelated as shown in fig 10 given above.

By analyzing correlation values, we can determine which features have a strong impact on emissions. For example:

- Power consumption vs. emission levels – A high correlation indicates that increased appliance usage results in higher emissions.
- Temperature vs. emission levels – Overheating appliances may lead to increased emissions due to inefficient operation.
- Humidity vs. power consumption – Higher humidity may force HVAC systems to work harder, leading to increased energy consumption and emissions.

To generate a Feature Correlation Heatmap, the dataset is first pre-processed, and the correlation matrix is calculated using Pearson correlation. The values range from -1 to +1, where:

- +1 indicates a strong positive correlation (as one feature increases, the other also increases).
- 0 indicates no correlation.
- -1 indicates a strong negative correlation (as one feature increases, the other decreases)

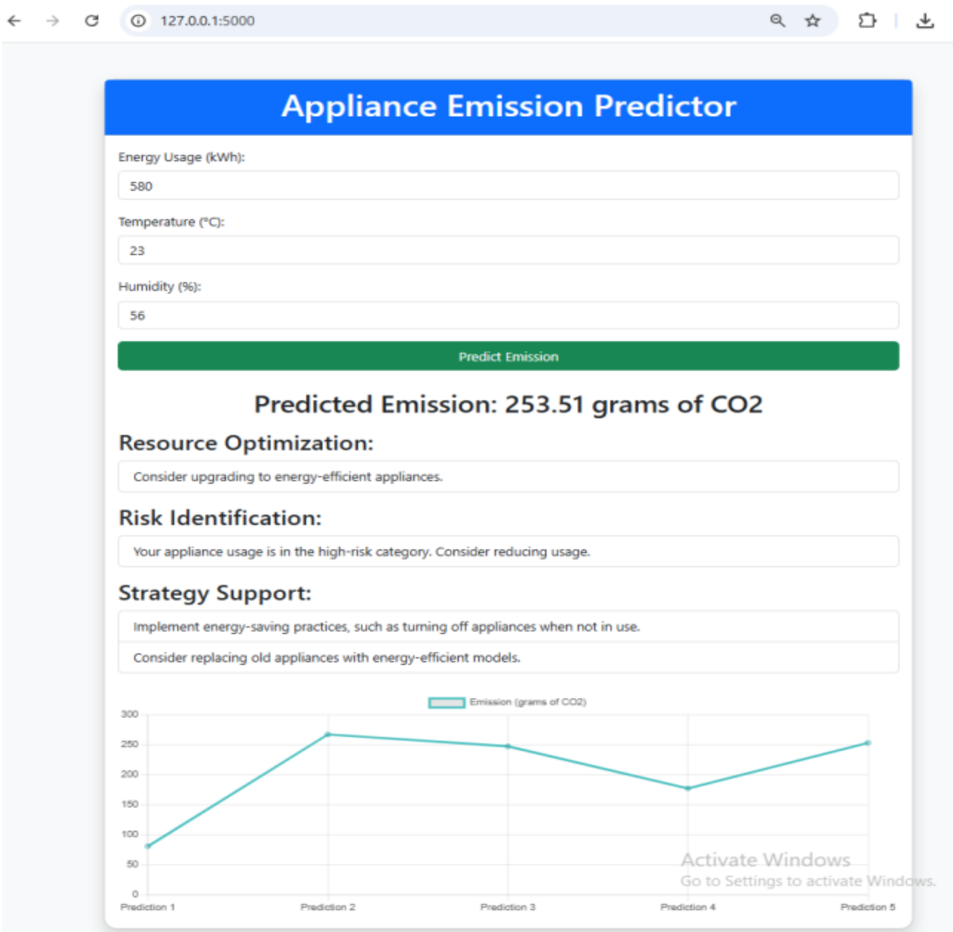

Figure 11: Output with the help of GUI

VI. CONCLUSIONS

This study explored the use of machine learning approaches for monitoring and predicting appliance emissions. The key findings include:

- Machine learning models, such as Random Forest and Neural Networks, effectively predict emission levels based on sensor data.
- Data preprocessing plays a crucial role in improving prediction accuracy by handling missing values and noise.
- Model evaluation using MSE, R^2 Score, and accuracy confirmed that Random Forest performed best in terms of accuracy and generalization.
- Visualization dashboards and real-time alerts help users make informed decisions about emission levels.
- Provided output is shown in Fig 11 given above.

VII. FUTURE SCOPE

Future research can focus on:

- Integrating IoT and Edge Computing to enable real-time emission monitoring with low latency.
- Exploring Hybrid Models combining deep learning and traditional ML for improved accuracy.
- Incorporating External Factors such as weather conditions and usage patterns for more comprehensive predictions.
- Developing Explainable AI (XAI) Models to enhance transparency and trust in machine learning predictions.
- Validating the System on Larger Datasets from different geographic locations and appliance types.

REFERENCES

[1] Roupen Minassian et al., "Optimizing indoor environmental prediction in smart buildings: A comparative analysis of deep learning models"

Energy & Buildings ELSEVIER 2025. https://doi.org/10.1016/j.enbuild.2024.115086
[2] Connor McGookin et al.., "Advancing participatory energy systems modelling" Energy Strategy Reviews Energy Strategy Reviews, ELSEVIER,2024. https://doi.org/10.1016/j.esr.2024.101319
[3] Yinuo Jia et al., "Towards sustainable consumption: Factors influencing energy-efficient appliance adoption in haze-affected environments" Energy Strategy Reviews, ELSEVIER 2024. https://doi.org/10.1016/j.esr.2024.101416
[4] Abdelrahman O. Ali et al., "Optimized smart home energy management system: Reducing grid consumption and costs through real-time pricing and hybrid architecture" Case Studies in Thermal Engineering, ELSEVIER 2024.
[5] Montaser N.A. Ramadan et al., "Real-time IoT-powered AI system for monitoring and forecasting of air pollution in industrial environment" Ecotoxicology and Environmental Safety, ELSEVIER 2024.
[6] Matteo Barsanti et al., "Informing targeted Demand-Side Management: Leveraging appliance usage patterns to model residential energy demand heterogeneity" Energy & Buildings, ELSEVIER 2024. https://doi.org/10.1016/j.enbuild.2024.114639
[7] Spyros Giannelos et al., "Machine learning approaches for predictions of CO2 emissions in the building sector" Electric Power Systems Research, ELSEVIER 2024. https://doi.org/10.1016/j.epsr.2024.110735
[8] Marta Jemel̜janova et al., "Adapting machine learning for environmental spatial dat A review" Ecological Informatics, ELSEVIER 2024.https://doi.org/10.1016/j.ecoinf.2024.102634
[9] Ivan Izonina et al., "Ivan Izonina, Roman Tkachenkob, Stergios Aristoteles Mitoulisc, Asaad Faramarzic, Ivan Tsmotsd, Danylo Mashtalira" ScienceDirect, ELSEVIER 2024.
[10] Zahra Eddaoudi et al., "Brief Review of Energy Consumption Forecasting Using Machine Learning Models" ScienceDirect, ELSEVIER,2024.
[11] Kaizhe Fan et al., "Harnessing the power of AI and IoT for real-time CO2 emission monitoring" Heliyon, 2024. https://doi.org/10.1016/j.heliyon.2024.e36612
[12] Guangchun Ruan et al., "Data-driven energy management of virtual power plants: A review" Advances in Applied Energy, ELSEVIER 2024. https://doi.org/10.1016/j.adapen.2024.100170
[13] RONG HUANG et al., "Carbon Footprint Management in Global Supply Chains: A Data-Driven Approach Utilizing Artificial Intelligence Algorithms" IEEE ACCESS,2024.
[14] Erik Johannes Husom et al., "Engineering Carbon Emission-aware Machine Learning Pipelines" IEEE ACCESS 2024.
[15] Leila Farahzadi et la., "Application of machine learning initiatives and intelligent perspectives for CO2 emissions reduction in construction" Journal of Cleaner Production, ELSEVIER 2023.
https://doi.org/10.1016/j.jclepro.2022.135504
[16] Tiago Fonseca et al., "Dataset for identifying maintenance needs of home appliances using artificial intelligence" Data in Brief, ELSEVIER 2023. https://doi.org/10.1016/j.dib.2023.109068.
[17] Harsh Bhatt et al., "Forecasting and mitigation of global environmental carbon dioxide emission using machine learning techniques" Cleaner Chemical Engineering, ELSEVIER 2023.
https://doi.org/10.1016/j.clce.2023.100095.
[18] MICHAEL HANS et al., "Predictive Analytics Model for Optimizing Carbon Footprint From Students' Learning Activities in Computer Science-Related Majors" IEEE ACCESS, VOLUME 11, 2023.