IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14357

CO_2 EMISSIONS PREDICTION USING MACHINE LEARNING IN DIESEL PRODUCTS

Dr Sireesha K[1], Sri Lakshmi Harshitha S[2], Harshitha P[3], Lakshmi Harika S[4], Bhavya V[5]

Associate Professor, CSE, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India[1]

Student IV CSE, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India[2,3,4,5]

Abstract: Diesel-powered engines are major contributors to CO_2 emissions, driving environmental pollution and climate change. Predicting and mitigating these emissions is essential for improving fuel efficiency and minimizing environmental impact. Over a 10-month period (October 2022 to December 2023), this project aims to deliver actionable insights. This study applies machine learning models to analyse and predict CO_2 emissions using historical data. The models, including Linear Regression, Random Forest, and XGBoost, are trained on engine parameters, fuel traits, and operational data to generate accurate predictions. Key input variables, such as engine load, fuel consumption, and temperature, are processed to provide real-time emission estimates, categorized as low, moderate, or high. This approach enhances diesel engine efficiency and enables industries, researchers, and policymakers to make informed, data-driven decisions for reducing carbon footprints. Through AI-driven methods, the project advances sustainability by offering precise, actionable guidance for emission control and regulatory compliance. This initiative fosters decarbonization, balancing environmental responsibility with operational efficiency.

Keywords: CO_2 emissions, diesel engines, machine learning, emission prediction, fuel consumption, engine load, temperature, Linear Regression, Random Forest, XGBoost, sustainability, emission control, regulatory compliance.

I. INTRODUCTION

Diesel engines are a significant source of carbon dioxide (CO_2) emissions, contributing to environmental damage and accelerating climate change. Addressing this issue is essential for enhancing fuel efficiency, lowering emissions, and achieving global sustainability targets. This project utilizes machine learning (ML) techniques to effectively predict and reduce CO_2 emissions. By evaluating engine-related and environmental factors, it provides industries, researchers, and policymakers with actionable insights to support sustainable operations and emission management. Advanced ML models, including Linear Regression, Random Forest, and XGBoost, are applied to historical emissions data to identify patterns between key variables such as fuel consumption, engine load, and operating conditions. Once trained, these models generate real-time predictions, enabling industries to implement data-driven strategies for reducing emissions and limiting environmental impacts.

Objectives of the Study

The primary objective of this research is to develop a machine learning-based predictive model for CO_2 emissions in diesel-powered products. This involves:

1. Create a highly accurate machine learning model to predict CO_2 emissions based on engine performance, fuel usage, and environmental conditions.

2. Improve fuel efficiency and lower emissions to reduce environmental impact and support sustainable practices.

3. Deliver real-time emission insights to help industries comply with environmental regulations efficiently and effectively.

4. Replace manual testing methods with automated, sensor-driven systems for enhanced precision and streamlined operations.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025
DOI: 10.17148/IJARCCE.2025.14357

II. LITERATURE SURVEY

A literature survey examines prior research on using machine learning techniques to predict CO_2 emissions from diesel engines. It provides an understanding of various approaches, technologies, and their effectiveness in addressing emission-related challenges. Recent advancements highlight the importance of real-time monitoring, predictive analytics, data-driven optimization, and automated systems to enhance accuracy and meet environmental regulations. Below are the key studies reviewed:

1. Title: Predictive Modelling of CO_2 Emissions in Diesel Engines Using Machine Learning [1] Authors**:** John D. Wilson, Sarah T. Roberts Year: 2023 Publisher: Journal of Environmental Analytics: This research explores the use of machine learning algorithms like Random Forest and Gradient Boosting to predict CO_2 emissions in diesel engines. The study uses datasets incorporating variables such as engine load, fuel type, and operating conditions. While these models achieve notable accuracy, their limitations include the inability to process real-time data, high computational demands, and the need for extensive preprocessing. In contrast, our approach addresses these gaps by utilizing real-time sensor data, lightweight machine learning algorithms designed for edge devices, and automated preprocessing systems, ensuring accurate and efficient predictions.
2. Title: Artificial Intelligence for Emission Control in Industrial Diesel Engines [2] Authors: Priya Sharma, Ravi Kumar Year: 2023 Publisher: International Journal of Machine Learning Applications, Lda. Description: This study presents an AI-based framework using Support Vector Machines and Decision Trees to predict CO_2 emissions from industrial diesel engines. The system identifies critical factors, including engine RPM and fuel injection rate, to categorize emission levels. Despite its potential, the framework faces challenges with scalability for large datasets and practical implementation in real-world scenarios. Our project overcomes these issues by leveraging scalable cloud-based systems, employing advanced algorithms like XGBoost for in-depth analysis, and incorporating automated feature engineering to enhance adaptability and efficiency in industrial use cases.
3. Title: Deep Learning Approaches for CO_2 Emission Prediction [3] Authors: Muhammad Asif, Chen Wei Year: 2023 Publisher: Energy and Environment Journal Description: This research investigates the application of deep learning models, particularly Long Short-Term Memory (LSTM) networks, to predict CO_2 emissions under dynamic operating conditions. While these models show improved accuracy, they face challenges such as overfitting and high computational requirements. To address these limitations, our project combines deep learning for analysing time-series data with tree-based models for feature importance analysis. Additionally, we use advanced regularization techniques and real-time hardware integration to maintain a balance between accuracy and computational efficiency.

These studies highlight the potential of machine learning in CO_2 emission prediction, emphasizing the importance of real-time data processing, automation, and advanced modelling techniques. Building on these findings, our project integrates real-time sensor data, predictive analytics, and optimized machine learning algorithms to provide scalable, efficient, and practical solutions for managing CO_2 emissions in diesel engines.

III. PROPOSED METHODOLOGY

1. System Architecture:

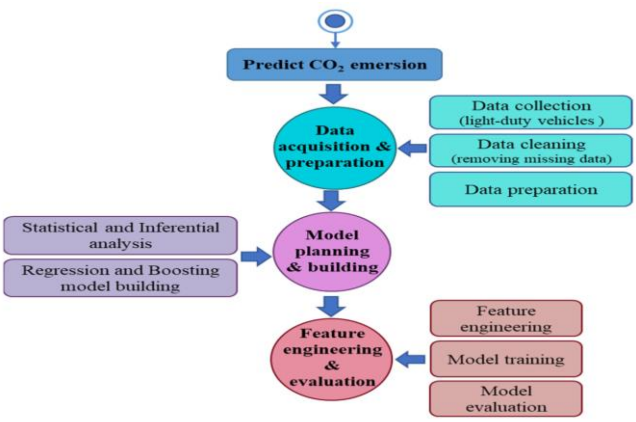

Fig. 1. System Architecture

DOI: 10.17148/IJARCCE.2025.14357

The system architecture for CO_2 emission prediction follows a structured approach involving data acquisition, model development, and evaluation. Initially, relevant data from light-duty vehicles or industrial sources is collected, cleaned, and prepared to ensure consistency. Statistical analysis is conducted to identify key patterns, followed by the implementation of machine learning models such as regression and boosting algorithms for accurate predictions. Feature engineering enhances model performance by selecting significant attributes like fuel consumption and vehicle type. The model undergoes training and evaluation using metrics such as RMSE and R^2 to ensure reliability. Once optimized, the best-performing model is deployed for real-time CO_2 emission forecasting.

2. Technology Stack:

The CO_2 emissions prediction system is built using Python, leveraging machine learning frameworks like Scikit-learn and TensorFlow for model training. Data processing is handled using Pandas and NumPy, while visualization tools like Matplotlib aid in trend analysis. Databases such as MySQL or PostgreSQL store structured data, with Flask or Django enabling web-based deployment. Google Colab provides a cloud-based environment for model training, Version control with Git and cloud storage solutions like Google Drive streamline data management and collaboration.

3. Key Technological Components:

The key technology components of the CO_2 emissions prediction system include Machine Learning Algorithms**,** such as Random Forest and Neural Networks, for accurate emissions forecasting**.** Data Processing Tools**,** like Pandas and NumPy, facilitate data cleaning and transformation. Model Training Platforms**,** such as TensorFlow and Scikit-learn, support the development of predictive models. Database Management Systems**,** including MySQL or PostgreSQL, ensure efficient storage and retrieval of emissions data. Visualization Libraries, like Matplotlib and Seaborn, aid in interpreting trends. Cloud Computing Services**,** such as Google Colab, provide scalable model training environments. Web Frameworks**,** like Flask or Django, enable seamless model deployment and user interaction.

4. System Workflow:

The CO_2 emissions prediction system follows a structured workflow, starting with data acquisition and preprocessing**,** where emission data from diesel-powered sources is collected, cleaned, and standardized for analysis. Next, feature engineering and selection identify key factors such as fuel consumption and engine load that significantly impact emissions. The model development **and** training phase involves applying machine learning techniques, including regression models and deep learning algorithms, to learn patterns in the data. Once trained, the models undergo evaluation and optimization using performance metrics like RMSE and R-squared, with hyperparameter tuning to improve accuracy. Finally, the prediction and deployment phase integrates the best-performing model into a real-time system, enabling continuous monitoring and visualization of CO_2 emissions, assisting industries and policymakers in making informed environmental decisions.

IV.SECURITY IMPLEMENTATION

1. Authentication Mechanisms: Secure Access System:

The CO_2 emissions prediction platform ensures secure access by utilizing advanced authentication tools such as AWS Cognito or Firebase Authentication. These systems make sure only authorized users can access the platform. Users can log in through multiple methods, including OAuth, which enables secure sign-ins with trusted services like Google or Microsoft accounts. Additionally, the platform assigns specific permissions to users based on their roles—whether they are administrators, data engineers, or analysts—using role-based access control (RBAC). To further protect sensitive information like engine data and emissions predictions, two-factor authentication (2FA) is included as an extra layer of security.

2. Bot Protection: Intelligent Rate Limiting:

To prevent misuse and block automated attacks, the system incorporates effective bot protection techniques. Tools such as Cloudflare are used to implement rate limiting, which restricts the number of predictions or data uploads each user can perform within a set period. For instance, users might have a daily limit on requests to maintain system performance and stability. CAPTCHA is also used to distinguish between genuine users and bots, preventing automated systems from overwhelming the platform or submitting fake data. These measures ensure the platform remains secure, reliable, and efficient for all users.

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025
DOI: 10.17148/IJARCCE.2025.14357

3. Input Validation: Reliable Data Validation:

The system emphasizes the importance of working with valid and accurate data by implementing robust input validation processes. Libraries like Zod are used to define strict rules for data entry, ensuring that only clean and properly formatted data is processed. For example, fields such as engine load or fuel consumption are checked to ensure they contain numerical values within a specified range. Invalid or malicious inputs, such as SQL injection attempts or out-of-range values, are automatically

rejected. By validating data thoroughly, the system minimizes security risks, improves reliability, and delivers precise CO_2 emissions predictions.

V.EXPERIMENTAL RESULTS AND ANALYSIS

1. Screenshots of the Application:

Fig.1.Index Page

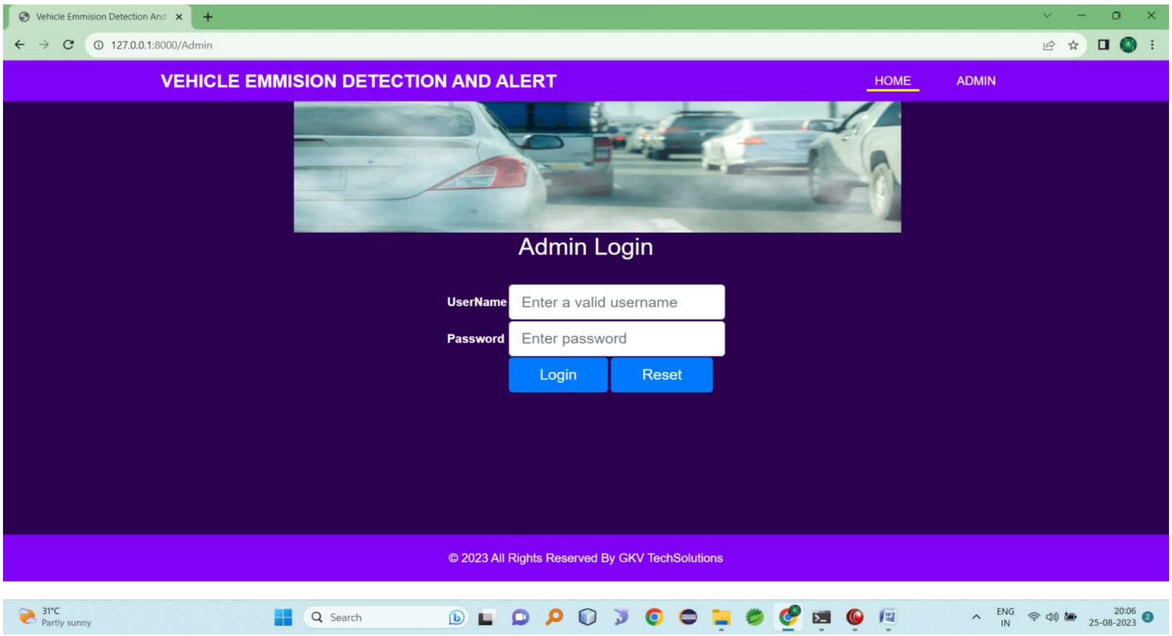

Fig. 2. Admin Login Page

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14357

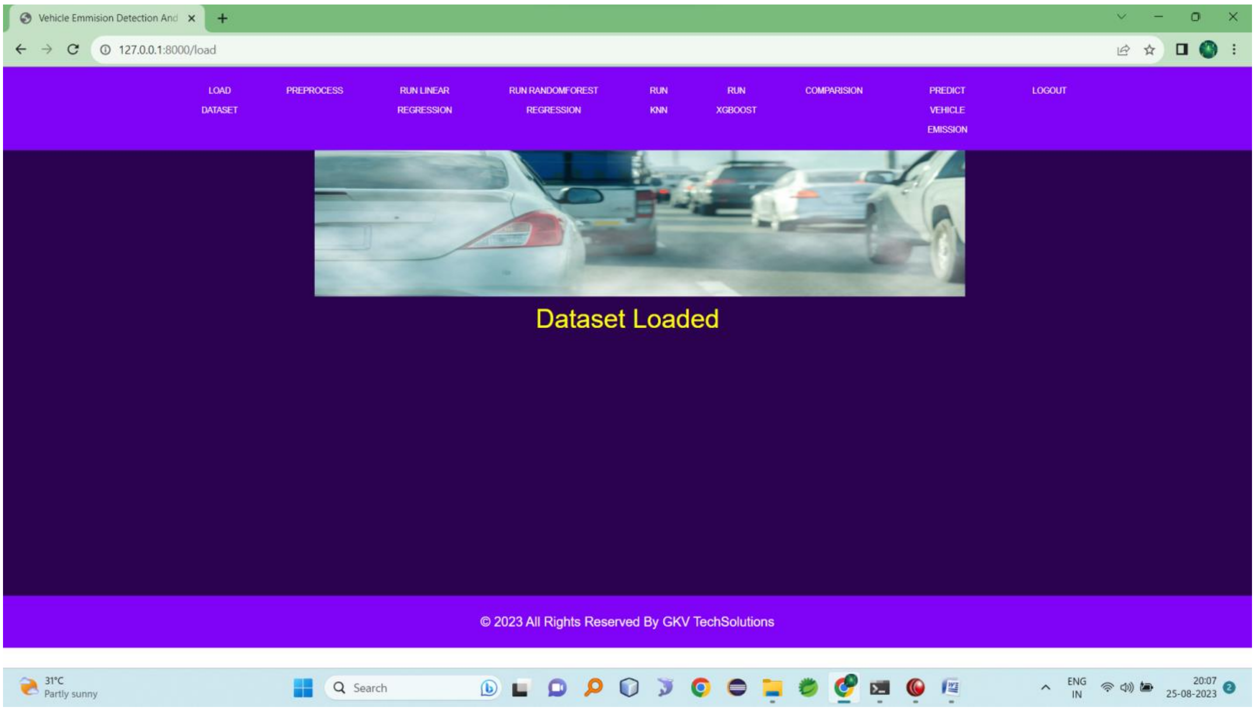

Fig. 3. Dataset Loaded Page

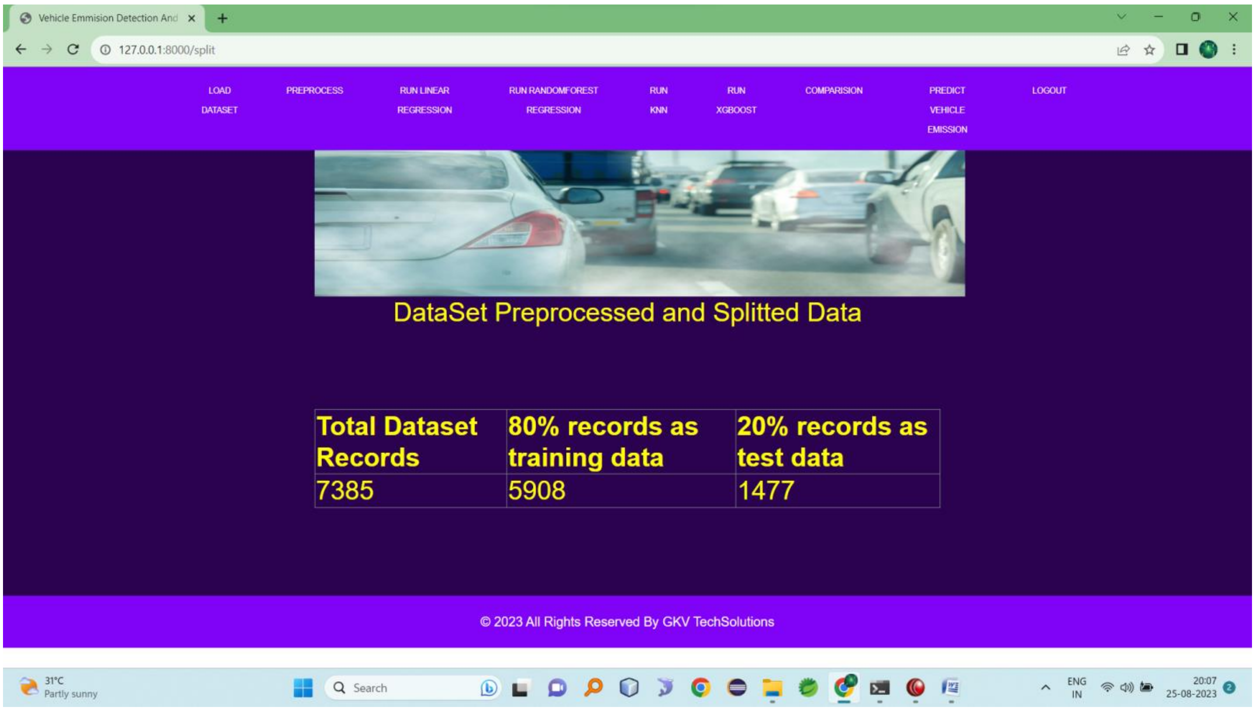

Fig. 4 Preprocess Completes Page

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025
DOI: 10.17148/IJARCCE.2025.14357

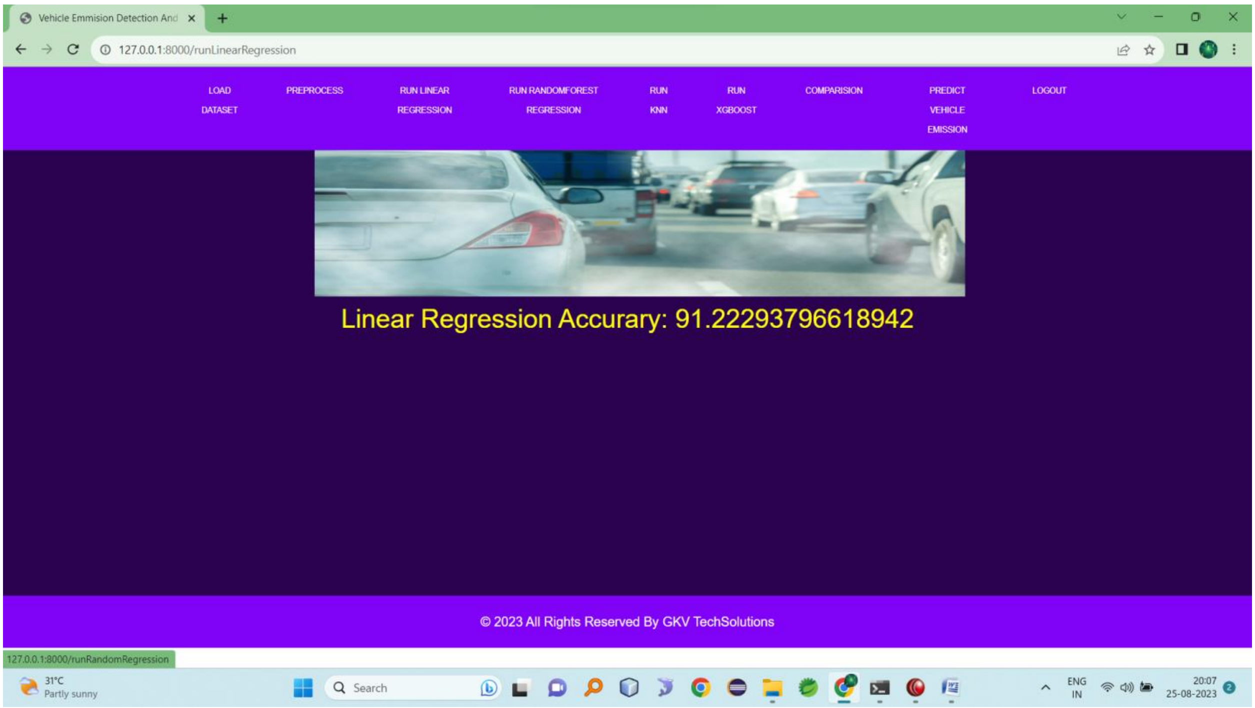

Fig. 5. Linear Regression page

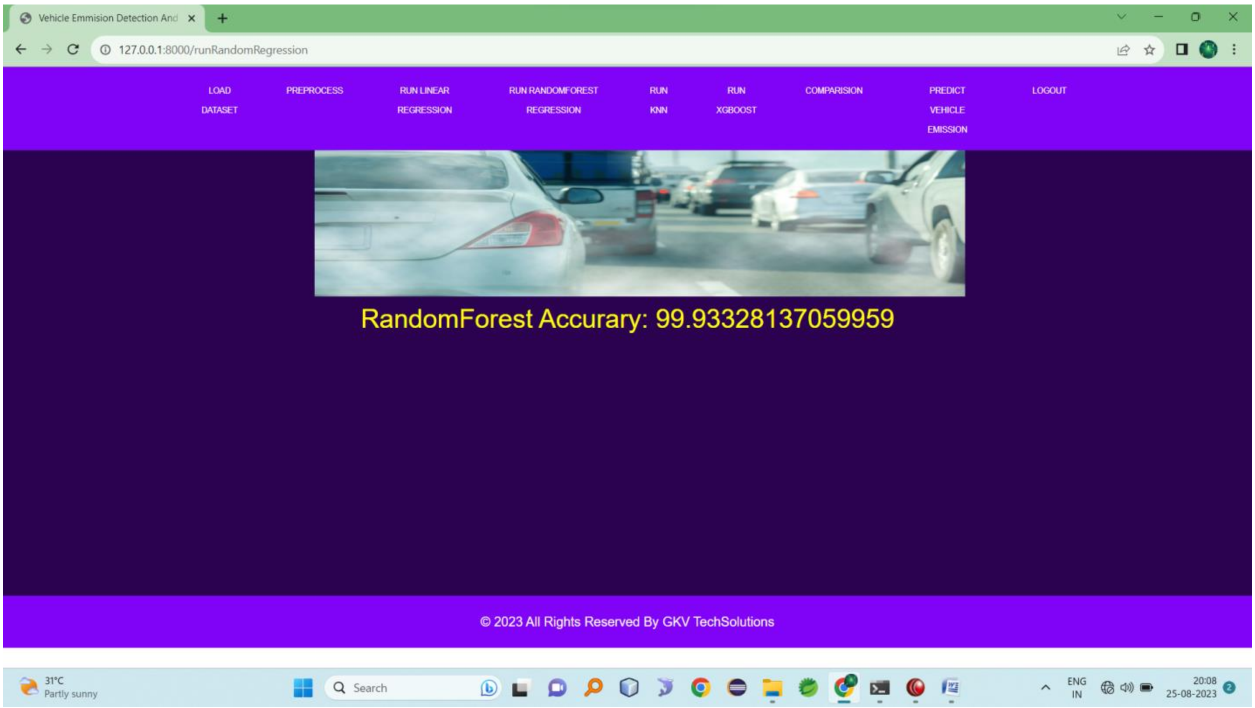

Fig. 6. Random Forest Page

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 3, March 2025
DOI: 10.17148/IJARCCE.2025.14357

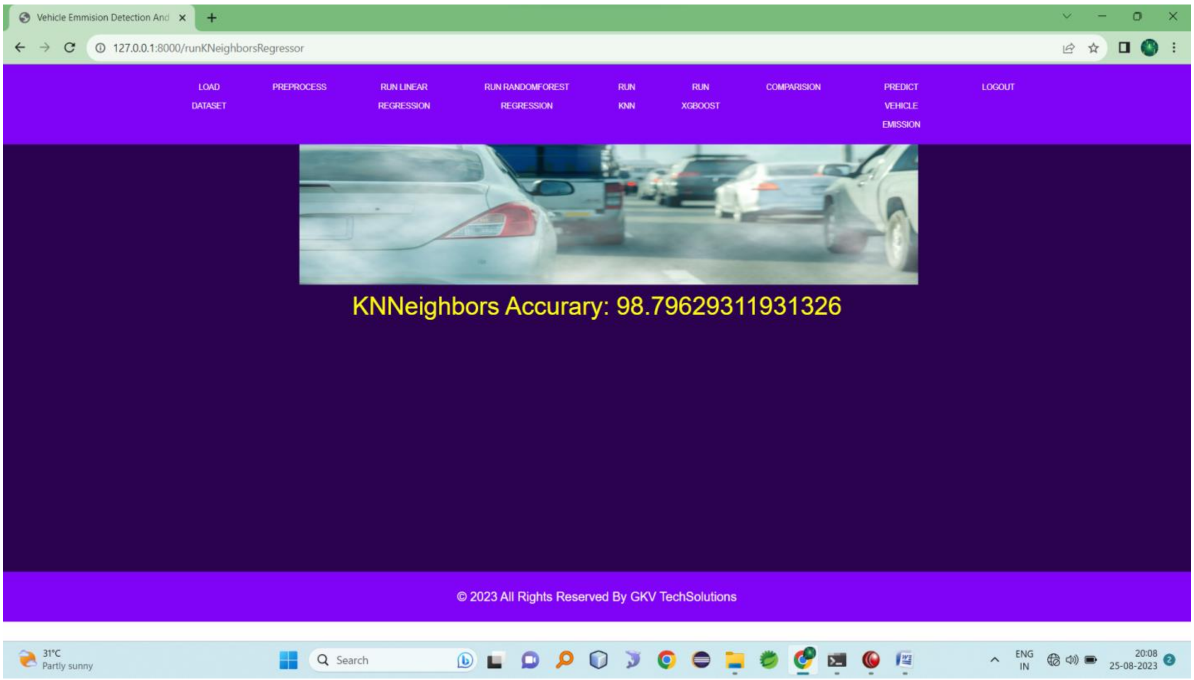

Fig.7. KN Neighbors Accuracy Page

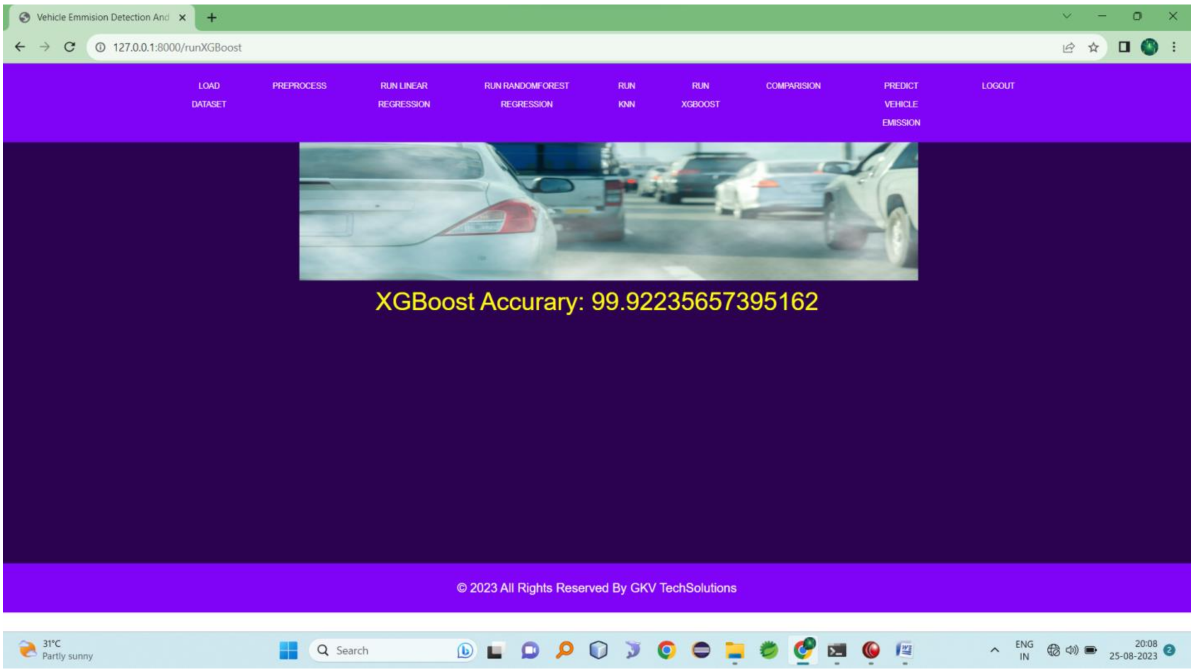

Fig.8. XGBoost Accuracy Page

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 3, March 2025
DOI: 10.17148/IJARCCE.2025.14357

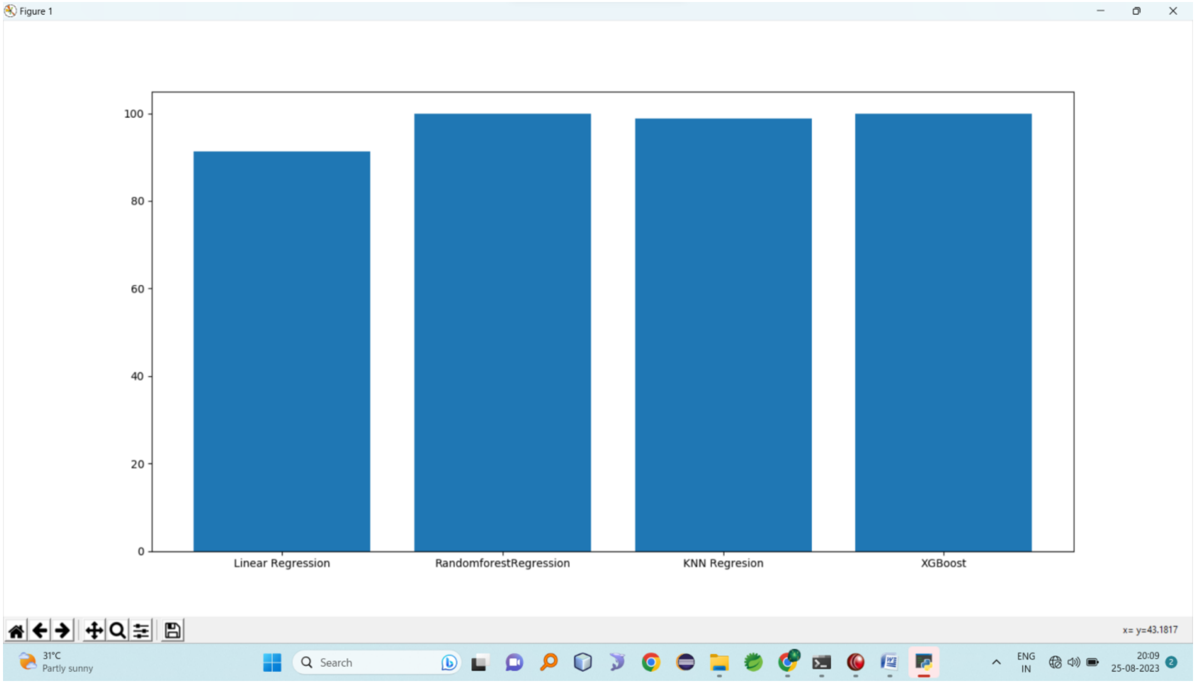

Fig.9. Comparision Graph Page

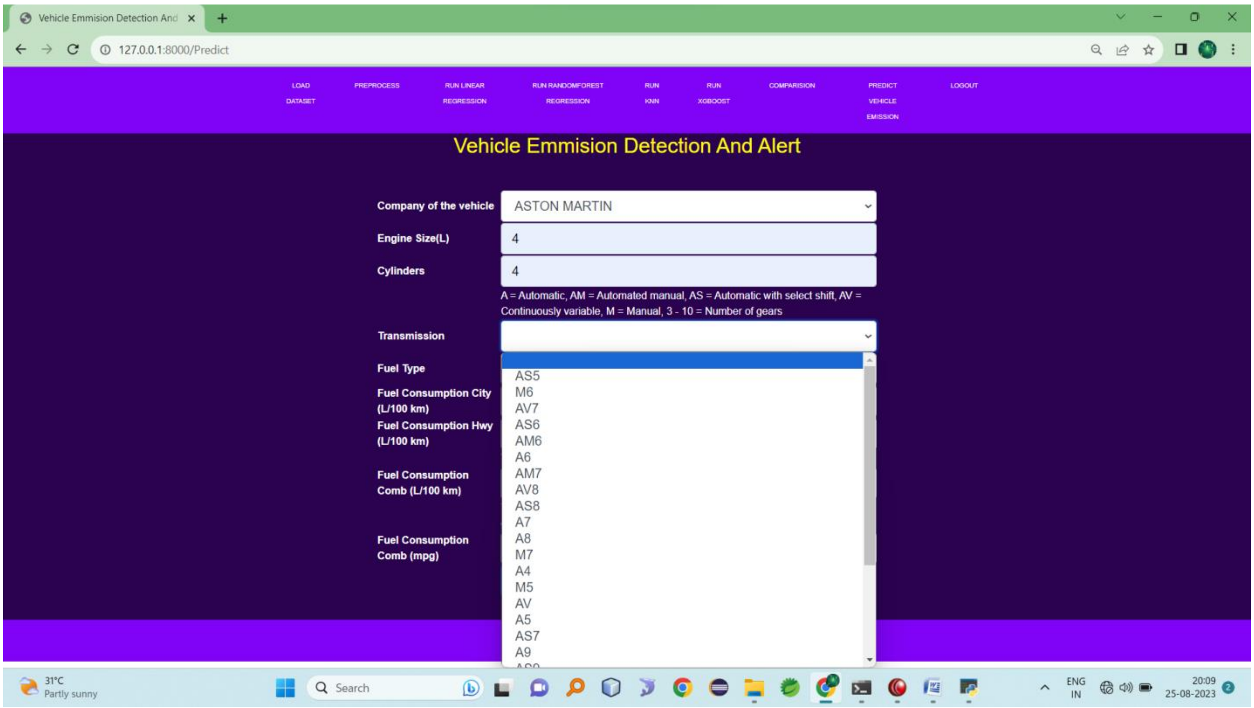

Fig.10.Input page

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025
DOI: 10.17148/IJARCCE.2025.14357

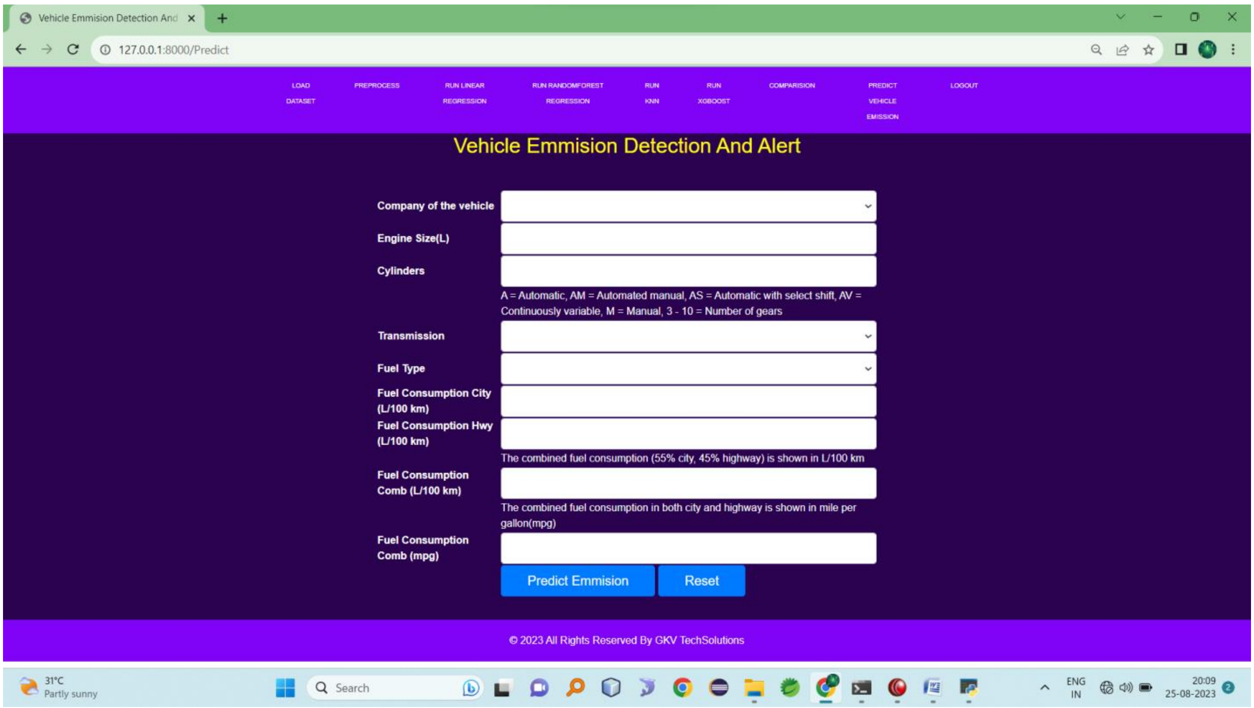

Fig.11.Input Page

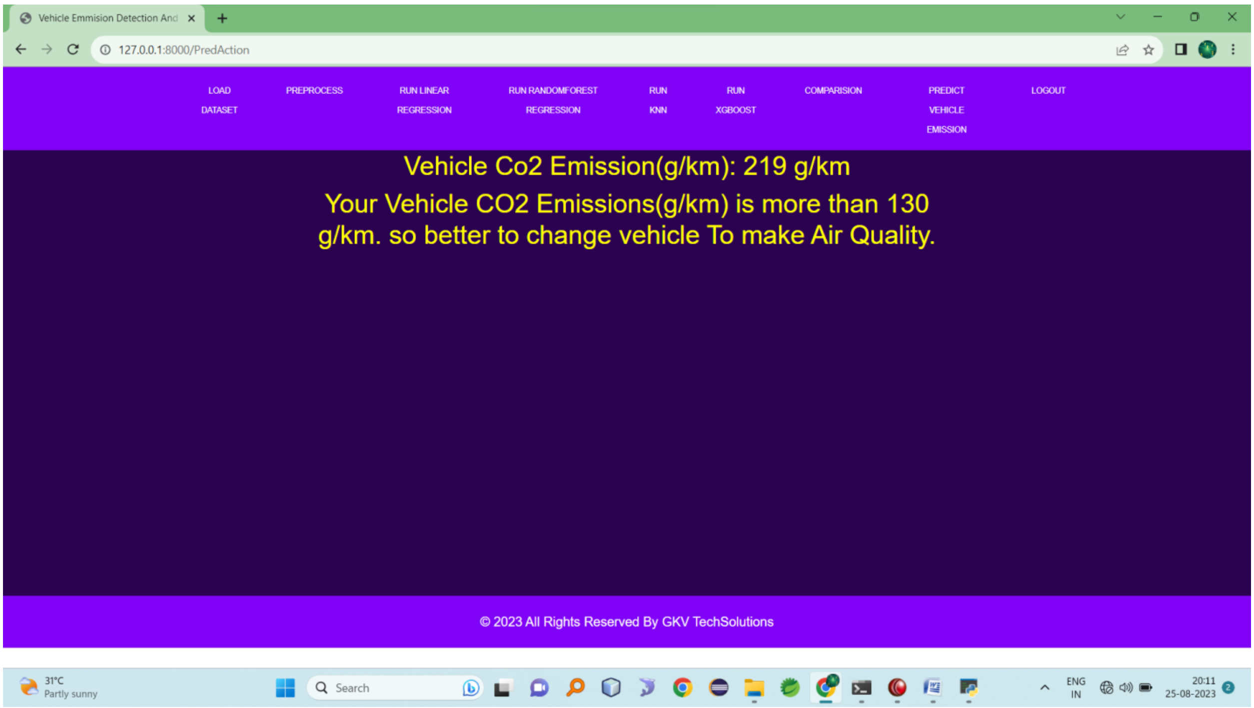

Fig.12. Output Page

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 3, March 2025
DOI: 10.17148/IJARCCE.2025.14357

2.Comparative Analysis

Aspect	Existing Systems (Limitations)	Related Work (Previous Research & Findings)	How Our Project Improves
Real-Time Prediction	Not available; relies on manual or lab-based testing.	Liu et al, (2022) integrated LOT & ML for real-time monitoring.	Uses ML for dynamic CO2 tracking.
Accuracy	Empirical models struggle with non-linearity and engine variations.	Zhang st al, (2019) used Decision Trees, but overfitting was a concern.	ML models like XGBoost improves accuracy with optimized feature selection.
Cost & Time Efficiency	Expensive due to laboratory-based emission testing.	Heywood (1998) introduced EFA, which reduced costs but lacked adaptability.	Automated prediction reducing costs & time for industries.

VI. CONCLUSION

Environmental protection has gained prominence as a key concern in both academic research and industrial practices. This study tackles the task of predicting critical vehicle details from telemetry data to enable effective emission monitoring. Using a variety of data mining techniques, accurate predictions were achieved for fuel types and the registration periods of gasoline vehicles, based on telemetry data provided by an environmental protection agency in a specific urban area.

However, predictions for the registration periods of diesel vehicles reached an accuracy of only around 70%. This is largely due to factors such as the manual categorization of registration timelines and the varying usage patterns of vehicles by different users. Future efforts will focus on refining these predictions by integrating more comprehensive datasets and applying improved algorithms to enhance the accuracy of emissions-related insights.

VII. FUTURE SCOPE

The future of CO_2 emission prediction systems for diesel engines focuses on using advanced artificial intelligence and machine learning to make predictions more accurate and efficient. By employing deep learning techniques and real-time data from vehicle sensors connected to the Internet of Things (IoT), these systems can monitor emissions more precisely, even in changing driving conditions. Additionally, federated learning can be used to work with data from multiple sources while keeping that data private and secure.

To make the system more reliable, blockchain technology could be used to securely record emission data, ensuring the information is tamper-proof and meets regulatory requirements. The system can also be developed to support various types of vehicles and environmental conditions, making it useful in different regions with diverse diesel technologies and emission standards.

Edge computing, which processes data directly on the vehicle, can reduce delays and provide immediate feedback to drivers, helping them adjust their driving behaviour to reduce emissions. Future versions of the system could include features like predictive maintenance, where engine and emission data trends are analysed to alert users about potential issues before they occur, reducing emissions through proactive action.

DOI: 10.17148/IJARCCE.2025.14357

By combining these innovations with user-friendly interfaces and compliance tools, such systems could play a major role in cutting emissions and promoting a greener, more sustainable future.

REFERENCES

[1] Yashnik, S. A., Denisov, S. P., Danchenko, N. M., & Ismagilov, Z. R. (2016). The synergetic impact of palladium addition on the catalytic performance of monolithic platinum-manganese-alumina catalysts designed for controlling diesel vehicle emissions. Applied Catalysis B: Environmental, 185(15), 322–336.

[2] Fujita, E. M., Campbell, D. E., Zielinska, B., et al. (2012). A comparative evaluation of the MOVES2010a, MOBILE6.2, and EMFAC2007 models for mobile source emissions using on-road tunnel and remote sensing data. Journal of the Air & Waste Management Association, 62(10), 1134–1149.

[3] Fu, J. S., Dong, X., Gao, Y., Wong, D. C., & Lam, Y. F. (2012). Sensitivity and linearity analysis of ozone in East Asia, focusing on the impacts of domestic emissions and intercontinental transport. Journal of the Air & Waste Management Association, 62(9), 1102–1114.

[4] Kfoury, A., Ledoux, F., Roche, C., Delmaire, G., Roussel, G., & Courcot, D. (2016). PM2.5 source apportionment influenced by steelworks emissions in a French urban coastal area using a constrained non-negative matrix factorization model. Journal of Environmental Sciences, 40, 114–128.

[5] Cheng, Y. (2011). A comparative analysis of vehicle exhaust pollution control systems in China and the United States. China University of Geosciences, Beijing, China.

[6] Wu, X. (2012). Detection and revision of exhaust pollutant standards using acceleration simulation modes for gasoline vehicles in use. Chang'an University, Xi'an, China.

[7] Bishop, G. A., Starkey, J. R., Ihlenfeldt, A., Williams, W. J., & Stedman, D. H. (1989). Infrared long-path photometry as a remote sensing tool for assessing automobile emissions. Analytical Chemistry, 61(10), 671A–677A.

[8] Stephens, R. D., & Cadle, S. H. (1991). Measuring carbon monoxide emissions from vehicles on the road using remote sensing technologies. Journal of the Air & Waste Management Association, 41(1), 39–46.

[9] Stedman, D. H., Bishop, G., & McLaren, S. (1995). U.S. Patent No. 5,401,967. Alexandria, VA, USA: U.S. Patent and Trademark Office.

[10] Guenther, P. L., Stedman, D. H., Bishop, G. A., Beaton, S. P., Bean, J. H., & Quine, R. W. (1995). A hydrocarbon detection system for remote sensing of vehicle exhaust emissions. Review of Scientific Instruments, 66(4), 3024–3029.

