Çukurova Üniversitesi
Mühendislik Fakültesi Dergisi
Çukurova University
Journal of the Faculty of Engineering
ISSN: 2757-9255 CİLT/VOLUME: 40 SAYI/ISSUE: 2 HAZİRAN/JUNE 2025

Comparative Study of Emission Prediction Using Deep Learning Models

İhsan ULUOCAK [1,a]

[1]*Çukurova University, Faculty of Ceyhan Engineering, Department of Mechanical Engineering, Adana, Türkiye*

[a]***ORCID**: 0000-0001-5014-9514*

Article Info
Received : 28.02.2025
Accepted : 23.05.2025
DOI: 10.21605/cukurovaumfd.1648164
Corresponding Author
İhsan ULUOCAK
iuluocak@cu.edu.tr
Keywords
Optimization
Deep learning
CNG
Biodiesel
***How to cite:** ULUOCAK, İ., (2025). Comparative Study of Emission Prediction Using Deep Learning Models. Cukurova University, Journal of the Faculty of Engineering, 40(2), 337-346.*

ABSTRACT

This study investigates the prediction of exhaust emissions (CO, CO_2, and NO_x) from a diesel engine fueled with biodiesel-diesel blends and compressed natural gas (CNG) using deep learning models. Biodiesel derived from canola, sunflower, and corn oils was blended with conventional, while CNG was introduced at flow rates of 0, 5, 10, and 15 liters per minute (lt/min). Two deep learning architectures, Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM), were employed to predict emissions. The models' performance was evaluated using metrics such as R^2, RMSE, and Kling-Gupta Efficiency (KGE). The results demonstrated that both models achieved high accuracy, with R^2 and KGE values exceeding 0.93 for all emission types. The GRU model showed superior performance in predicting CO and NO_x emissions, while the LSTM model excelled in predicting CO_2 emissions. The study highlights the potential of deep learning models in accurately predicting exhaust emissions and optimizing fuel blends for reduced environmental impact.

Derin Öğrenme Modelleri Kullanılarak Emisyon Tahmininin Karşılaştırmalı Çalışması

Makale Bilgileri
Geliş : 28.02.2025
Kabul : 23.05.2025
DOI: 10.21605/cukurovaumfd.1648164
Sorumlu Yazar
İhsan ULUOCAK
iuluocak@cu.edu.tr
Anahtar Kelimeler
Optimizasyon
Derin öğrenme
CNG
Biyodizel
***Atıf şekli:** ULUOCAK, İ., (2025). Derin Öğrenme Modelleri Kullanılarak Emisyon Tahmininin Karşılaştırmalı Çalışması. Çukurova Üniversitesi, Mühendislik Fakültesi Dergisi, 40(2), 337-346.*

ÖZ

Bu çalışmada, 5G, 6G ve radar uygulamaları için 38 GHz ve 55 GHz frekanslarında çalışan ışıma örüntüsü yeniden yapılandırılabilir bir anten önerilmektedir. Anten, ışıma örüntüsü ve frekans yeniden yapılandırılabilirliğini sağlamak için tek bir BAR64-02V RF anahtarı kullanmakta olup, karmaşık besleme sistemlerine ihtiyaç duymadan etkin hüzme yönlendirmesi sunmaktadır. 38 GHz frekansında, Phi = 90° (180°'den 50°'ye) ve Theta = 90° (–50°'den –131°'ye) açılarında belirgin hüzme kaymaları gözlemlenmiştir. 55 GHz'de ise Phi = 0° yönünde –120°'den –108°'e bir kayma meydana gelmiştir. ON durumunda anten üç bantlı, OFF durumunda ise çift bantlı çalışma sergilemektedir. 38 GHz'de %93, 55 GHz'de %90,6 radyasyon verimi elde edilmiştir. Simülasyon sonuçları, VSWR ve S_{11} parametrelerinin tüm durumlarda empedans uyumu sağladığını göstermektedir. Tasarım, mm-Dalga uygulamaları için bir aday olup, hüzme şekillendirme, girişim azaltma ve sinyal kalitesini artırma gibi avantajlar sunarak yeni nesil kablosuz ve radar sistemleri için önemli bir potansiyel göstermektedir.

1. INTRODUCTION

The increasing global demand for energy and the growing concerns over environmental pollution have driven the automotive industry to explore alternative fuels and advanced emission control technologies [1]. Diesel engines, widely used in transportation and industrial applications, are significant contributors to air pollution due to their emissions of carbon monoxide (CO), carbon dioxide (CO_2), and nitrogen oxides (NO_x) [2]. These pollutants have severe impacts on human health: CO can cause oxygen deprivation leading to headaches, dizziness, and even death, while NO_x contributes to respiratory diseases and cardiovascular problems. Additionally, CO_2, a major greenhouse gas, exacerbates climate change, which indirectly threatens public health through extreme weather events [3].

One promising approach to addressing these challenges is the use of alternative fuels, such as biodiesel and compressed natural gas (CNG), either as pure fuels or in blends with conventional diesel. Biodiesel, derived from renewable sources, is known for its lower emissions compared to diesel. On the other hand, CNG, which primarily consists of methane, offers the advantage of reduced CO_2 and NO_x emissions due to its cleaner combustion characteristics [4]. However, the combustion dynamics and emission profiles of engines operating on biodiesel-CNG-diesel blends are complex and require detailed analysis.

In recent years, deep learning (DL) techniques have emerged as powerful tools for modeling complex systems, offering high accuracy in predicting nonlinear and multivariate relationships. Among these techniques, GRU and LSTM networks have shown great potential in time-series prediction tasks, making them particularly suitable for nonlinear system modeling [5,6]. GRUs and LSTMs are capable of capturing temporal dependencies in sequential data, which is essential for accurately predicting emissions under varying engine operating conditions.

There are many studies modeling and optimization of CNG-diesel internal combustion engine emission. Ramachandran et al used gradient Boosting Regressor (GBR) and LASSO (Least Absolute Shrinkage and Selection Operator) Regression for modeling microalgae biodiesel as a high-reactive fuel directly injected along with various CNG [7]. Sahoo et al. founded an ANN and SVM model to predict performance, combustion, and emission characteristics of a compressed natural gas (CNG) spark ignition (SI) engines [8]. Niknam et al predicted NO_x emission of a dual-fuel (Diesel + CNG) Compression Ignition Engine Using the DCNN model [9]. Despite the growing interest in alternative fuels and machine learning applications, there is a lack of comprehensive studies focusing on the prediction of CO, CO_2, and NO_x emissions from engines fueled with biodiesel-CNG-diesel blends using deep learning techniques. This study aims to address this gap by developing a deep learning-based model using GRU and LSTM architectures to predict the emission characteristics of a diesel engine operating with varying blends of biodiesel, CNG, and diesel. The proposed model leverages experimental data collected under different engine loads and speeds to accurately estimate emissions, providing a valuable tool for optimizing fuel blends and engine parameters to minimize environmental impact.

The primary contributions of this study are;
- The development of a robust deep learning model based on GRU and LSTM for predicting CO, CO_2, and NO_x emissions from biodiesel-CNG-diesel blends,
- Insights into the emission reduction potential of alternative fuel blends. By combining experimental data with advanced deep learning techniques, this research aims to contribute to the development of cleaner and more efficient combustion technologies, ultimately supporting global efforts to reduce harmful emissions and improve air quality.

2. MATERIAL AND METHOD

2.1. Fuel Blending and Data Collection

In this study, three types of biodiesels derived from canola oil, sunflower oil, and corn oil were used as alternative fuels. These biodiesels were blended with conventional diesel fuel at varying ratios to investigate their effects on engine performance and emissions. The blends included 20% biodiesel (B20) and 40% biodiesel (B40), with the remaining percentage being conventional diesel. The selected blending ratios of 20% and 40% biodiesel (B20 and B40) were based on commonly accepted standards in the literature for

evaluating partial replacement of diesel fuel. These ratios strike a balance between maintaining fuel properties similar to diesel while enabling a noticeable reduction in emissions, making them suitable for performance and emission assessments in diesel engines. Neat diesel (100% diesel) was also used as a baseline for comparison. The biodiesel fuels used in this study were produced via the transesterification process, where vegetable oils (canola, sunflower, and corn oils) were reacted with methanol in the presence of a potassium hydroxide (KOH) catalyst. This process converts triglycerides in the oils into fatty acid methyl esters (FAME), which constitute biodiesel. The resulting biodiesels were filtered and purified before being blended with diesel fuel. Additionally, compressed natural gas (CNG) was introduced into the combustion chamber at flow rates of 0, 5, 10, and 15 liters per minute (lt/min) to analyze its impact on emissions.
The experiments were designed to cover a comprehensive range of variables, including biodiesel blends (B20, B40, and neat diesel), CNG flow rates (0, 5, 10, and 15 lt/min), and engine speeds (1200, 1500, 1800, 2100, and 2400 rpm). For each combination of these variables, emission data (CO, CO_2, and NO_x) were collected using the MRU Delta 1600-V gas analyzer. A total of 140 data points were collected, with each condition repeated three times to ensure reproducibility. The average values of these repetitions were used for analysis. The data is derived from Ph.D. thesis of Erinc ULUDAMAR [10].

2.2. Experimental Setup

The experiments were conducted on a Mitsubishi Canter 4D34-2A engine, a four-stroke, four-cylinder diesel engine with a maximum power output of 91 HP at 3500 rpm and a maximum torque of 223 Nm at 2200 rpm. It was connected to an MP 160-4B electrical dynamometer, which was used to measure engine speed and load. The MP 160-4B dynamometer is a high-precision device capable of providing accurate and reliable measurements under various operating conditions. Emission measurements were carried out using an MRU Delta 1600-V gas analyzer, a high-precision device capable of detecting and quantifying CO, CO_2, and NO_x concentrations in the exhaust gases. The MRU Delta 1600-V offers high accuracy and reliability, making it suitable for detailed emission analysis. The engine was operated under no-load conditions at five different speeds: 1200, 1500, 1800, 2100, and 2400 rpm. This range of speeds was selected to simulate various operating conditions and to evaluate the engine's emission characteristics across a wide spectrum. Figure 1 shows the layout of the experimental test setup used to collect engine emission data under different fuel and CNG conditions.

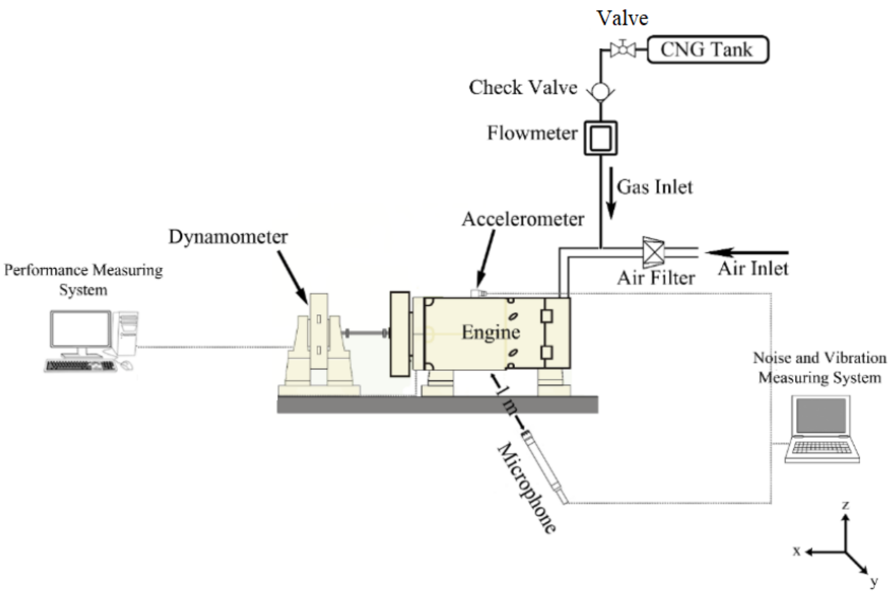

Figure 1. Schematic diagram of the experimental setup [10]

2.3. Input Features and Data Preprocessing

The input features for the deep learning models included cetane number, density, viscosity, engine speed, and CNG flow rate. These features were selected based on their significant influence on combustion dynamics and emission characteristics. The cetane number, density, and viscosity of the fuel blends were measured using standard laboratory techniques, while engine speed and CNG flow rate were controlled and recorded during the experiments. The collected data were preprocessed to ensure consistency and improve model performance. Normalization was applied to scale the input features and output targets (CO, CO_2, and NO_x emissions) to a range of 0 to 1. This step was crucial to ensure that all features contributed equally to the model training process and to prevent any bias due to differences in the scale of the data. The preprocessed data were then split into training and testing sets, with 80% of the data (112 data points) used for training the models and the remaining 20% (28 data points) reserved for testing. This split ensured that the models were evaluated on unseen data, providing a reliable measure of their generalization performance.

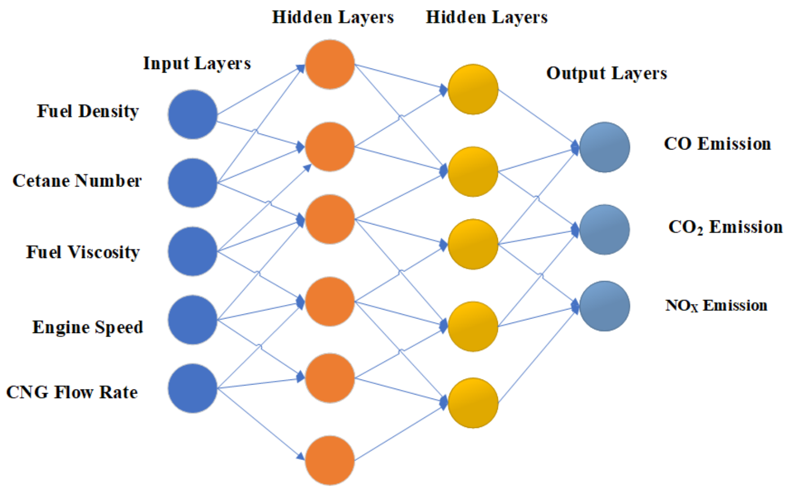

Figure 2. Deep neural network architecture

2.3. Model Development

The deep learning models were developed using the MATLAB Deep Learning Toolbox, which provides a comprehensive set of tools for designing, training, and evaluating deep learning models. Two architectures, LSTM and GRU, were employed to predict the emission characteristics of the engine. Both models were trained using the experimental data, with cetane number, density, viscosity, engine speed, and CNG flow rate as input features. The output targets were the emission levels of CO, CO_2, and NO_x. The models were implemented using MATLAB's deep learning framework, with hyperparameters optimized through Bayesian optimization. Details are visualized on Figure 3.

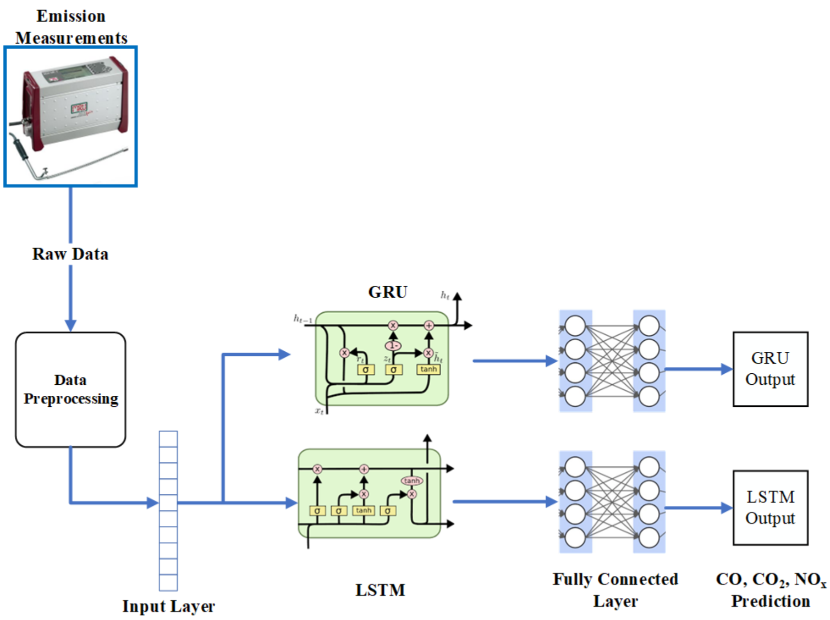

Figure 3. Developed models flowchart

2.3.1. Long Short-Term Memory (LSTM)

LSTM is a type of recurrent neural network (RNN) specifically designed to capture long-term dependencies in sequential data. It addresses the vanishing gradient problem commonly encountered in traditional RNNs by introducing memory cells and gating mechanisms [11,12]. These gates (input, forget, and output gates) regulate the flow of information, allowing the model to retain important information over long sequences while discarding irrelevant data. In this study, the LSTM model was used to predict emissions based on the temporal relationships between engine operating conditions and fuel properties. The gates and working flowchart is depicted on Figure 4.

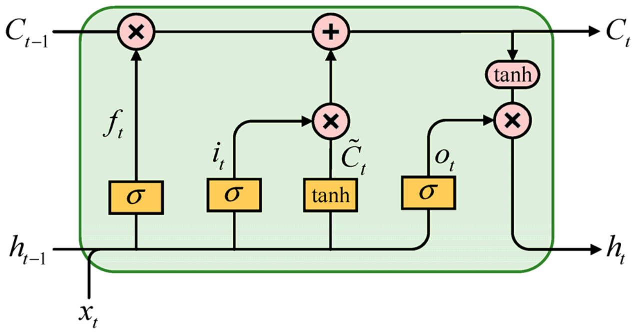

Figure 4. LSTM structure [13]

2.3.2. Gated Recurrent Units (GRU)

GRU is another variant of RNNs that simplifies the architecture of LSTM by combining the input and forget gates into a single update gate. This reduction in complexity often results in faster training times while maintaining comparable performance to LSTM. GRUs are particularly effective in modeling sequential data with shorter dependencies [14,15]. In this study, the GRU model was employed to predict emissions, leveraging its efficiency and ability to capture temporal patterns in the data. The model's performance was evaluated alongside LSTM to determine its suitability for emission prediction tasks. GRU details can be seen on Figure 5.

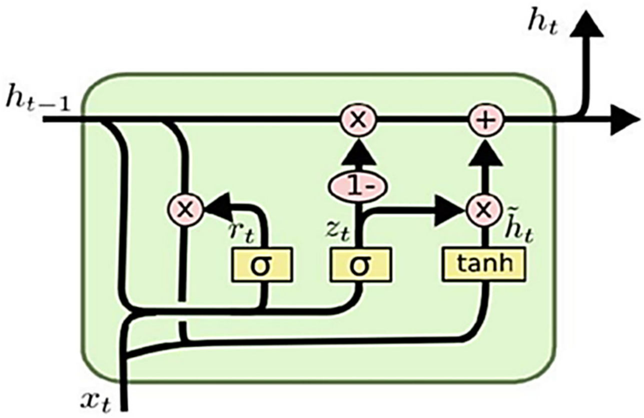

Figure 5. GRU structure [16]

2.4. Performance Evaluation

The performance of the GRU and LSTM models was evaluated using the Kling-Gupta Efficiency (KGE) metric, which combines correlation, variability ratio, and mean bias to provide a comprehensive assessment of model accuracy. KGE is particularly useful for evaluating the performance of models in hydrological and environmental studies [17]. In addition to KGE, Root Mean Square Error (RMSE) and Coefficient of Determination (R^2) were also used to assess the models' performance [18].

3. RESULTS AND DISCUSSION

The performance of LSTM and GRU models in predicting CO, CO_2, and NO_x emissions is summarized in Table 2, while the optimized hyperparameters for each model are presented in Table 1.

Table 1. Optimized hyperparameters

Parameter	Model	Initial learning rate	Hidden neuron unit
CO	LSTM	0.0028	10
	GRU	0.0021	152
CO_2	LSTM	0.099	197
	GRU	0.01	106
NO_X	LSTM	0.0042	11
	GRU	0.001	119

For CO emissions, the LSTM and GRU models showed nearly identical performance, with R^2 values of 0.968 and RMSE values of 0.047 ppm and 0.048, respectively. However, the GRU model achieved a slightly higher KGE value (0.951) compared to the LSTM model (0.939), suggesting better overall efficiency in capturing the variability and bias in CO emission predictions. The optimized hyperparameters for CO emissions reveal that the LSTM model used a learning rate of 0.0028 and 10 hidden neuron units, while the GRU model utilized a learning rate of 0.0021 and 152 hidden neuron units. The higher number of hidden units in the GRU model may have contributed to its marginally better performance in predicting CO emissions.

For CO_2 emissions, the LSTM model outperformed the GRU model, achieving an R^2 of 0.994 and a lower RMSE of 0.022 ppm, compared to the GRU model's R^2 of 0.987 and RMSE of 0.035 ppm. The KGE values for CO_2 predictions were also higher for the LSTM model (0.986) than for the GRU model (0.974), indicating superior performance in capturing the dynamics of CO_2 emissions. The optimized hyperparameters for CO_2 emissions show that the LSTM model used a learning rate of 0.099 and 197 hidden neuron units, while the GRU model employed a learning rate of 0.01 and 106 hidden neuron units. The significantly higher learning rate and hidden neuron count in the LSTM model likely contributed to its superior performance in predicting CO_2 emissions.

In the case of NO_x emissions, the GRU model showed a slight advantage over the LSTM model, with an R^2 of 0.988 and an RMSE of 0.028 ppm, compared to the LSTM model's R^2 of 0.984 and RMSE of 0.034 ppm. However, the LSTM model achieved a higher KGE value (0.983) than the GRU model (0.974), suggesting better alignment between predicted and observed NO_x emissions in terms of correlation, variability, and bias. The optimized hyperparameters for NO_x emissions indicate that the LSTM model used a learning rate of 0.0042 and 11 hidden neuron units, while the GRU model utilized a learning rate of 0.001 and 119 hidden neuron units. The lower learning rate and higher number of hidden units in the GRU model may have contributed to its slightly better RMSE performance, while the LSTM model's higher KGE value suggests better overall efficiency.

Table 2. Accuracy metrics results

Parameter	Model	R^2	RMSE (ppm)	KGE
CO	LSTM	0.968	0.047	0.939
	GRU	0.968	0.048	0.951
CO_2	LSTM	0.994	0.022	0.986
	GRU	0.987	0.035	0.974
NO_X	LSTM	0.984	0.034	0.983
	GRU	0.988	0.028	0.974

The results demonstrate that the proposed methodologies perform efficiently with test datasets. Figures 6-8 present magnified views of the cluster, enabling a comprehensive analysis of the predicted results. The x-axis represents instances, extending up to 28, while the y-axis corresponds to CO, CO_2, and NO_x levels, respectively. Notably, the data trend exhibits significant variations in CO, CO_2, and NO_x concentrations at different intervals.

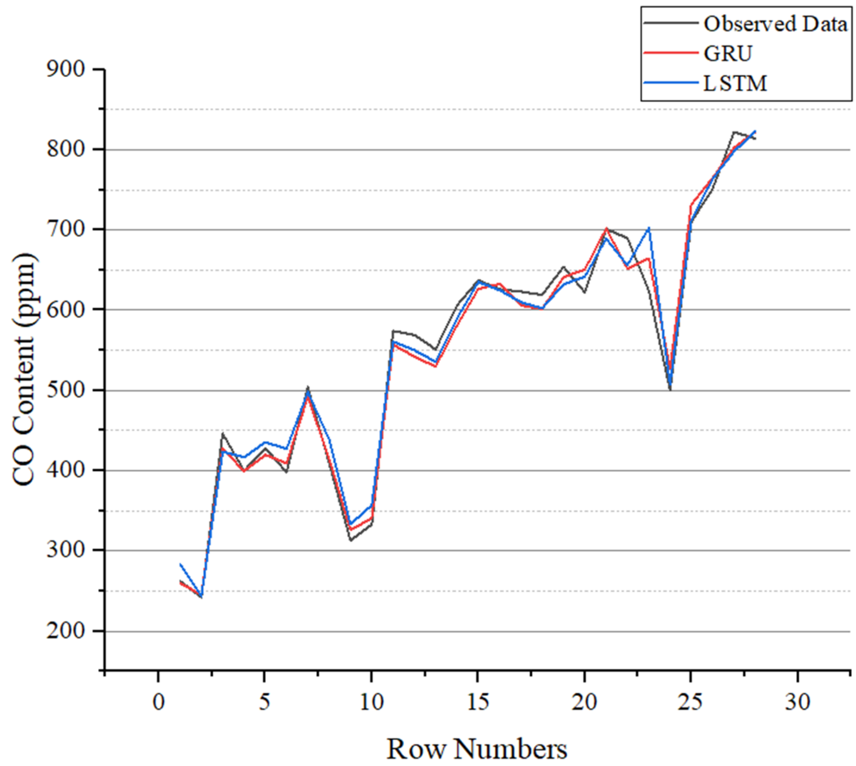

Figure 6. Data clusters observed data vs. model output for CO

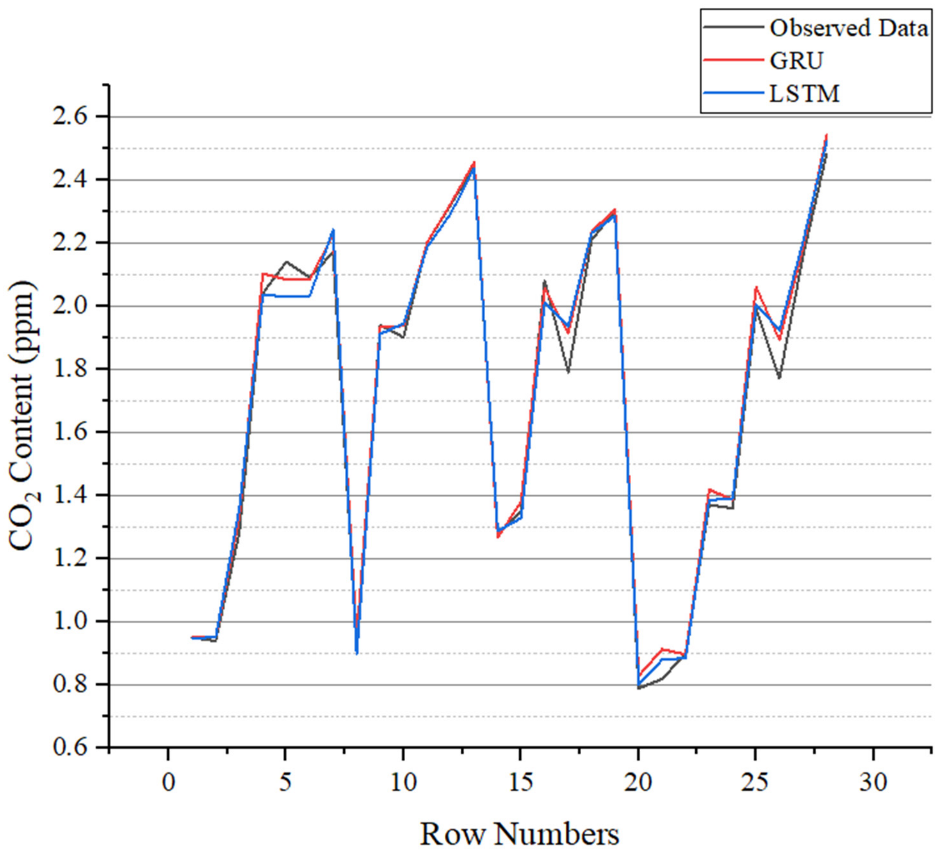

Figure 7. Data clusters observed data vs. model output for CO_2

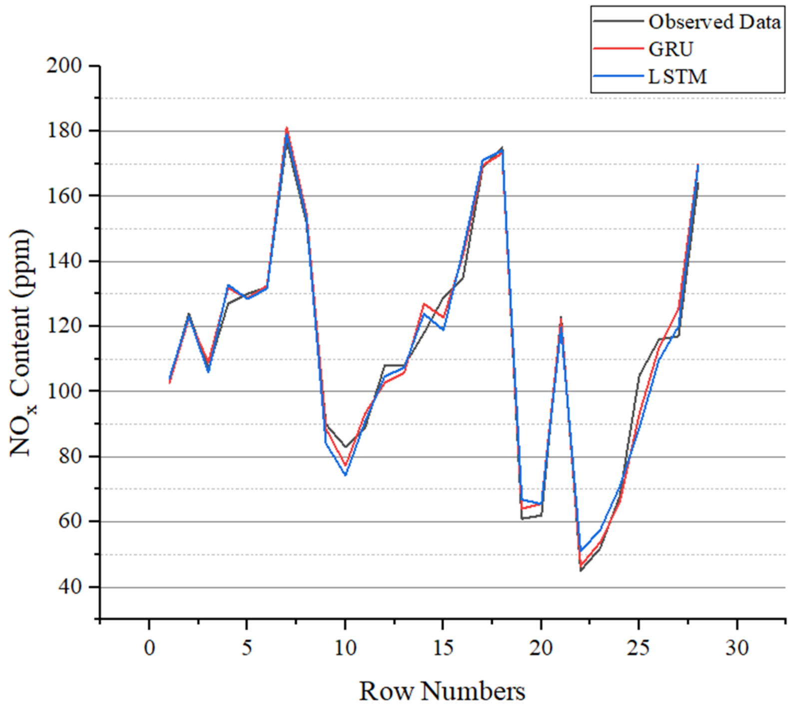

Figure 8. Data clusters observed data vs. model output for NO_x

A box plot is an effective tool for visualizing the distribution of the predicted data. Each box contains a central marker representing the median, while the lower and upper edges correspond to the 25th and 75th percentiles, respectively. The boundary lines extend to the most extreme data points, identified as outliers beyond 1.5 times the interquartile range (IQR). Additionally, the mean value is separately represented using the symbol '□'. Figures 9-11 present box plots of the measured and predicted CO, CO_2, and NO_x data. Based on the results for CO, the mean values for the observed data and CO predictions were 551, 548.8, and 553.2 ppm, respectively. Similarly, for CO_2, the mean values for the observed data, GRU, and LSTM models were determined as 1.71, 1.74, and 1.72 ppm, respectively. Lastly, for NO_x, the mean values of the observed data, GRU, and LSTM predictions were 113.17, 113.76, and 113.2, respectively.

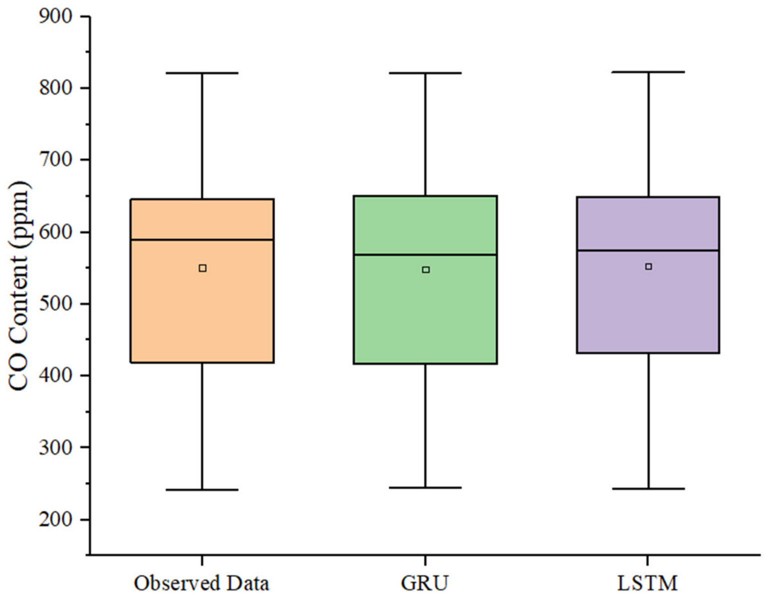

Figure 9. Boxplot illustration observed data vs. model output for CO

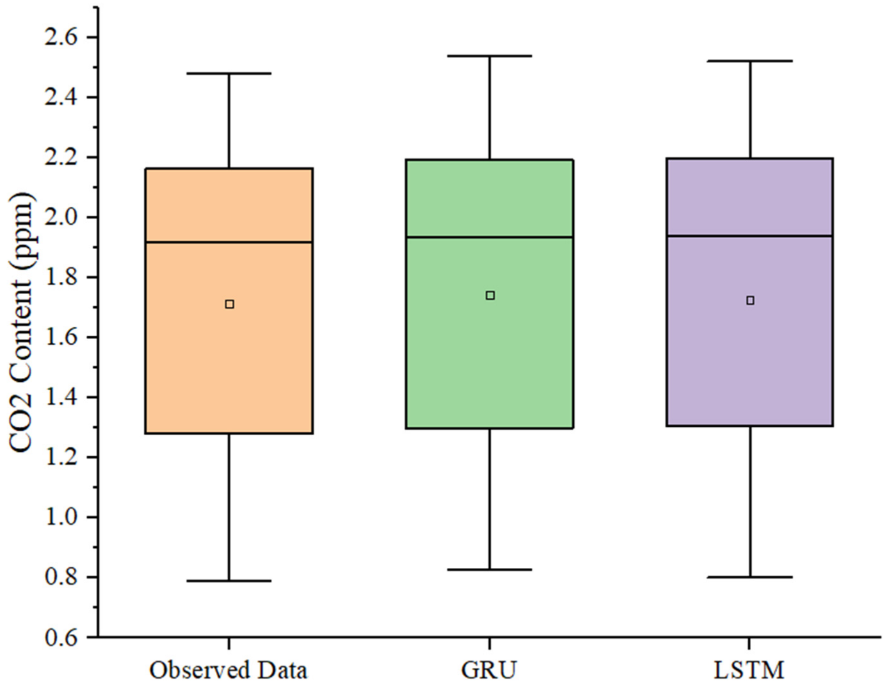

Figure 10. Boxplot illustration observed data vs. model output for CO_2

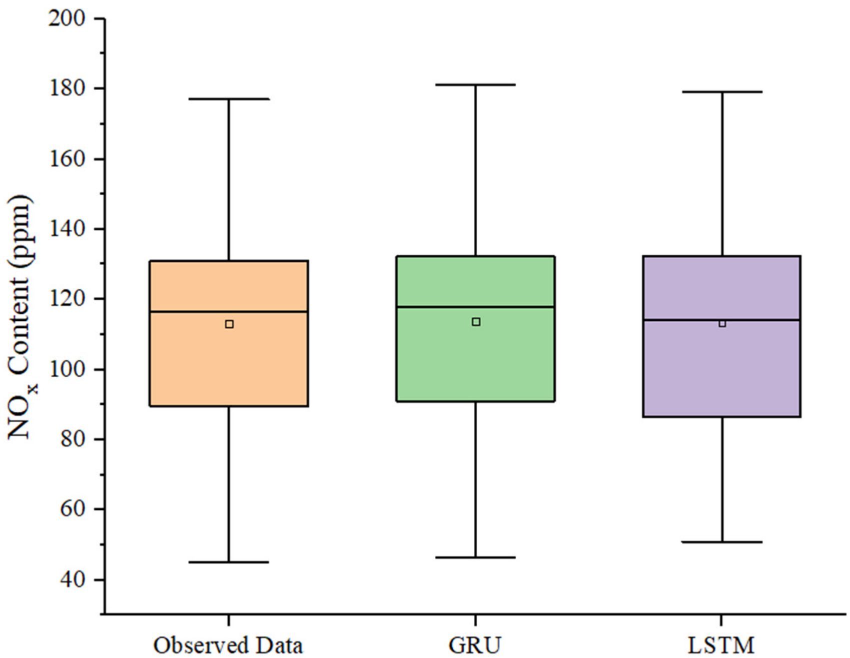

Figure 11. Boxplot illustration observed data vs. model output for NO_x

4. CONCLUSION

This study investigated the prediction of exhaust emissions from a diesel engine fueled with biodiesel-diesel blends and compressed natural gas (CNG) using deep learning models. Biodiesel derived from canola, sunflower, and corn oils was blended with conventional diesel, while CNG was introduced at varying flow rates. Two deep learning architectures, GRU and LSTM, were employed to predict emissions under different engine operating conditions.

The results demonstrated that both GRU and LSTM models achieved high accuracy in predicting CO, CO_2, and NO_x emissions, with R^2 and KGE values exceeding 0.93 for all emission types. The GRU model showed superior performance in predicting CO and NO_x emissions, while the LSTM model excelled in predicting CO_2 emissions. These findings highlight the potential of deep learning models in accurately predicting exhaust emissions and optimizing fuel blends to reduce environmental impact.

Future research could explore the application of these models to a wider range of alternative fuels and engine configurations, as well as the integration of real-time emission monitoring systems. This would further enhance the accuracy and applicability of deep learning models in emission prediction and control.

5. REFERENCES

1. Elgohary, M.M., Seddiek, I.S. & Salem, A.M. (2015). Overview of alternative fuels with emphasis on the potential of liquefied natural gas as future marine fuel. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment, 229*(4), 365-375.
2. Zheng, F., Zhang, H., Yin, H., Fu, M., Jiang, H., Li, J. & Ding, Y. (2022). Evaluation of real-world emissions of China V heavy-duty vehicles fueled by diesel, CNG and LNG on various road types. *Chemosphere, 303,* 135137.
3. Kumari, S. & Singh, S.K. (2023). Machine learning-based time series models for effective CO2 emission prediction in India. *Environmental Science and Pollution Research, 30*(55), 116601-116616.
4. Pathak, S.K., Nayyar, A. & Goel, V. (2021). Optimization of EGR effects on performance and emission parameters of a dual fuel (Diesel+ CNG) CI engine: an experimental investigation. *Fuel, 291,* 120183.
5. Sahoo, B.B., Jha, R., Singh, A. & Kumar, D. (2019). Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophysica, 67*(5), 1471-1481.
6. Uluocak, I. & Bilgili, M. (2024). Daily air temperature forecasting using LSTM-CNN and GRU-CNN models. *Acta Geophysica*, *72*(3), 2107-2126.
7. Ramachandran, E., Krishnaiah, R., Venkatesan, E.P., Parida, S., Dwarshala, S.K.R., Khan, S.A., ... & Linul, E. (2023). Prediction of RCCI combustion fueled with CNG and algal biodiesel to sustain efficient diesel engines using machine learning techniques. *Case Studies in Thermal Engineering, 51,* 103630.
8. Sahoo, S., Kumar, V.N.S.P. & Srivastava, D.K. (2022). Quantitative analysis of engine parameters of a variable compression ratio CNG engine using machine learning. *Fuel*, *311*, 122587.
9. Niknam, Y. (2024). NO_X emission prediction of a dual-fuel (Diesel+ CNG) compression ignition engine using the DCNN model. *Power System Technology*, *48*(1), 1034-1053.
10. Uludamar, E. (2016). Vibration measurement based analysses of internl Combustion Engines. *Doktora Tezi.* Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Makine Mühendisliği Anabilim Dalı, Adana.
11. Bilgili, M., Durhasan, T. & Pinar, E. (2024). Time series analysis of sea surface temperature change in the coastal seas of Türkiye. *Journal of Atmospheric and Solar-Terrestrial Physics, 263,* 106339.
12. Korkmaz, C. ve Kacar, İ. (2024). Zaman serisinin kestirimi için uzun-kısa süreli bellek ağı yaklaşımı. *Çukurova Üniversitesi Mühendislik Fakültesi Dergisi,* 39(4), 1053-1066.
13. Liu, Y., Zhang, H., Wu, C., Shao, M., Zhou, L. & Fu, W. (2024). A short-term wind speed forecasting framework coupling a maximum information coefficient, complete ensemble empirical mode decomposition with adaptive noise, shared weight gated memory network with improved northern goshawk optimization for numerical weather prediction correction. *Sustainability, 16*(16), 6782.
14. Altıparmak, Z. & Aksu, İ.Ö. (2024). Time series installed capacity forecasting with deep learning approach for Türkiye. *Cukurova University, Journal of the Faculty of Engineering, 39(3),* 709-718.
15. Bilgili, M., Pinar, E. & Durhasan, T. (2025). Global monthly sea surface temperature forecasting using the SARIMA, LSTM, and GRU models. *Earth Science Informatics*, *18*(1), 1-17.
16. Chang, V., Xu, Q.A., Chidozie, A., Wang, H. & Marino, S. (2024). Predicting economic trends and stock market prices with deep learning and advanced machine learning techniques. *Electronics*, *13*(17), 3396.
17. Araya, D., Mendoza, P.A., Muñoz-Castro, E. & McPhee, J. (2023). Towards robust seasonal streamflow forecasts in mountainous catchments: impact of calibration metric selection in hydrological modeling. *Hydrology and Earth System Sciences*, *27*(24), 4385-4408.
18. Bilgili, M. & Pinar, E. (2023). Gross electricity consumption forecasting using LSTM and SARIMA approaches: A case study of Türkiye. *Energy*, *284*, 128575.