Proceedings of Book Series PBS -Vol 42, pp.14-21 **Advances in Energy Transition and Environmental Impacts**

Machine Learning-Based Prediction of Transport-Related CO_2Emissions Across Heterogeneous National Contexts

Jawher Bouaziz[#1], Najah Kechiche[#2], Khdeijetna Mohamed El Moctar[2] ,Walid Hassen[#3]

[#] *Laboratory of Metrology and Energy Systems, National Engineering School of Monastir, Monastir University. 5019Monastir, Tunisia*

[1]*The Higher National Engineering School of Tunis (ENSIT), Tunis University, 1008 Tunis, Tunisia.* bouazizjawher1997@gmail.com

[2]*Higher Institute of Transport and Logistics of Sousse, Sousse University, 4023 Sousse, Tunisia,* kechiche2000@gmail.com, khadoujm39@gmail.com

[3] *National Engineering School of Monastir, Monastir University. 5019Monastir, Tunisia* hassen.walid@gmail.com

***Abstract*— In this paper, we investigate transport-related CO_2emissions (TCO_2) across 31 countries using a set of socioeconomic, energy, sectoral, and meteorological factors as well as temporal indicators. Correlation analysis reveals a large variability between the studied countries, with socioeconomic indicators and sectors showing consistently positive and negative relationships with TCO_2 emissions. In contrast, renewable energy consumption generally exhibits negative correlations. To estimate TCO_2, we used three machine-learning models, such as Support Vector Regression (SVR), Ridge Regression (RR), and Random Forest (RF). The best model was RF (R^2 = 0.87, MAPE = 3.08%), followed by SVR with moderate accuracy, while the performance of RR was inadequate to predict the TCO2. In terms of national-level predictions, the RF method has good generalization ability, with very high accuracies (<10% rRMSE) obtained in several developed nations (i.e., Germany, Denmark, the United States, France, and Portugal) along with reasonable performance in developing countries (i.e., China, India, and Tunisia, etc.). However, the regression plot shows a strong agreement between observed and estimated values, with only minor deviations, particularly at milder and higher emission levels. These results confirm that the RF approach can be applicable in heterogeneous contexts.**

***Keywords*—Transport-related CO_2 emissions; Machine learning; Random Forest; Renewable energy; Cross-country analysis; Emission prediction; Sustainable transport.**

I. INTRODUCTION

Nowadays, Environmental problems, especially climate change, are one of the most serious environmental dilemmas. Greenhouse gas (GHG) emissions are significant markers of climate change, including carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O), and fluorinated gases. Overall emissions can be measured in terms of CO2 equivalents [1]. Predictive model development and evaluation are critical to long-term climate planning, especially as global warming continues and countries face intense pressure to reduce GHG emissions across all sectors and fulfill their commitments under the Paris Agreement [2]. One of the main forces behind global commerce and economic expansion, the transportation sector is responsible for around 25% of worldwide energy-related CO_2 emissions, placing it in the top three [3] . As a result, reducing carbon emissions in this industry is extremely difficult. Global transport emissions increased 2.1% in 2022, mostly due to growth in industrialized nations; however, the increase would have been greater if low-carbon vehicle adoption had not accelerated [4]. In order to direct mitigation actions and inform adaptation measures, predictive models are essential. As the significance of precisely predicting GHG and CO_2emissions has increased, researchers have used a range of techniques, from gray modeling and conventional statistical methods to machine learning and sophisticated deep learning algorithms [5]. In addition to deep learning architectures, including Long Short-Term Memory (LSTM) networks and Multilayer Perceptrons (MLP), machine learning techniques such as Support Vector Machines (SVM), Extreme Learning Machines (ELM), Gradient Boosting, XGBoost, and Random Forests have been used extensively. These models are highly

Copyright © 2026
ISSN: 2961-6611

Proceedings of Book Series PBS -Vol 42, pp.14-21 **Advances in Energy Transition and Environmental Impacts**

predictive. They successfully capture complex and nonlinear relationships ([6-8]). An overview of several studies that demonstrate the use and efficacy of these prediction methods in environmental and transportation-related CO2 emission forecasting. For instance, Kazancoglu et al. [9], Lu et al. [10] and Huang et al. [11] have applied grey modelling techniques to forecast transport-related energy usage and CO_2emissions. They used several types of data, such as gross domestic product (GDP), population, and energy consumption, as input variables. Qiao et al. [12], included multi-stage forecasting frameworks that encompassed socioeconomic, transportation, and energy indicators, while Ratanavaraha and Jomnonkwao [13] employed regression and time-series methods to predict the emissions of $CO_2$from the transport sector in Thailand. Recently, Çınarer et al. [14], utilized machine learning techniques, such as support vector machines (SVM), extreme gradient boosting (XGBoost) and multilayer perceptrons (MLP), to calculate future emission outcomes. Generally, most previous studies have focused on a single country, a limited number of countries, or specific transport modes. Only a few studies have investigated transport-related CO_2emissions across many countries ([15], [16]). To address this gap, the present study first identifies the major determinants of transport-sector CO_2 emissions across the selected countries. Second, it predicts transport-related CO_2emissions (TCO2) for 31 countries (G31) using socioeconomic, demographic, and environmental indicators as input variables and compares the performance of three machine-learning models. Finally, it determines the model that provides the highest prediction accuracy for each country and identifies the five countries exhibiting the best predictive performance.

II. Methodology

A. Study Framework

The methodology adopted in this study consists of several sequential steps. First, data collection was performed using socioeconomic, energy, meteorological, and environmental variables obtained from different databases. Second, we analyse the relationships between the input variables and transportation CO_2 emissions through a correlation matrix. To improve data quality, the collected data were then preprocessed through cleaning, normalization, and handling of missing values. After preprocessing, feature selection techniques were applied to identify the most relevant variables influencing transportation CO_2emissions. For each model, the algorithm selects the features with the highest correlation for each country, where surface temperature and energy consumption are consistently included along with four additional features showing the strongest correlations. Third, we trained three machine learning techniques, such as Support Vector Regression (SVR), Ridge Regression (RR), and Random Forest (RF), and we evaluated their performances by using statistical metrics. Then, we selected and validated the best-performing approaches to ensure prediction accuracy and reliability. Finally, we identify the five countries with the best predictive performance using the most commonly used evaluation metric in this analysis, as supported by a regression plot.

B. Data Collection

We use annual data from 2002 to 2022 for 31 countries in the current study. The independent and dependent variables were collected from the World Bank Open Data and Our World Data databases. Table 1 presents the variables that we considered in our study, including their abbreviations, descriptions, and units of measurement.

TABLE I

OVERVIEW OF THE PREDICTOR VARIABLES.

Feature	Abbreviation	Domain	Unit/Description
Population	POP	Socioeconomic	Number of inhabitants
Energy Combustion	EC	Energy	TJ
Carbon Dioxide Emissions from Buildings	CO_2_BLD	Sectoral Emissions	$MtCO_2$
Carbon Dioxide Emissions from Industry	CO_2_IND	Sectoral Emissions	$MtCO_2$
Carbon Dioxide Emissions from Land-Use Change and Forestry	CO_2_LUF	Sectoral Emissions	$MtCO_2$

Copyright © 2026
ISSN: 2961-6611

Proceedings of Book Series PBS -Vol 42, pp.14-21 **Advances in Energy Transition and Environmental Impacts**

Carbon Dioxide Emissions from Manufacturing and Construction	CO_2_MAN	Sectoral Emissions	$MtCO_2$
GDP per Capita	GDPpc	Economic	USD per capita
Urbanization Rate	URB	Socioeconomic	% of urban population
Surface Temperature	TEMP	Meteorological	°C
Renewable Energy Share	REN	Energy	% of total energy consumption
Gross Domestic Product	GDP	Economic	Billion USD
Year	YEAR	Temporal	Calendar year

C. Performance Evaluation

To evaluate the performance of the prediction models, this study employed four of the most commonly used metrics, particularly in TCO_2emissions prediction ([15], [16]) such as mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2). In addition, the relative root mean square error (RMSE) is also used in TCO_2emission prediction tasks and provides a normalized measure of prediction error. Its values can be interpreted according to the following classes [6]:

- rRMSE < 10%: Excellent prediction
- 10% ≤ rRMSE < 20%: Good prediction
- 20% ≤ rRMSE < 30%: Fair prediction
- rRMSE ≥ 30%: Poor prediction

III. Results and Discussion

In this section, we present the comparative results of three machine learning algorithms, namely support vector regression (SVR), ridge linear regression (RL), and random forest (RF). The dataset was divided into 80% for training and 20% for testing to evaluate model performance and generalization capability. All analyses were conducted using the Python programming environment.

A. Correlation Analysis of Transport CO_2 Emissions and Explanatory Variables

Figure 1 presents the cross-country correlation matrix of the relationships among $TCO_2$and dependent variables across 31 countries. There are substantial differences in the factors associated with transport-related emissions among countries. POP, GDP, and URB rate generally show strong positive correlations with TCO_2, particularly in North African countries as well as Asian countries. We observed high positive correlations in Morocco ($r \approx 0.99$), India ($0.95 \leq r \leq 0.99$), China ($0.95 \leq r \leq 0.99$), Tunisia ($0.91 \leq r \leq 0.96$), and Singapore ($0.83 \leq r \leq 0.98$). These results indicate that demographic growth, economic development, and urbanization rate are important factors of TCO_2 in these countries.

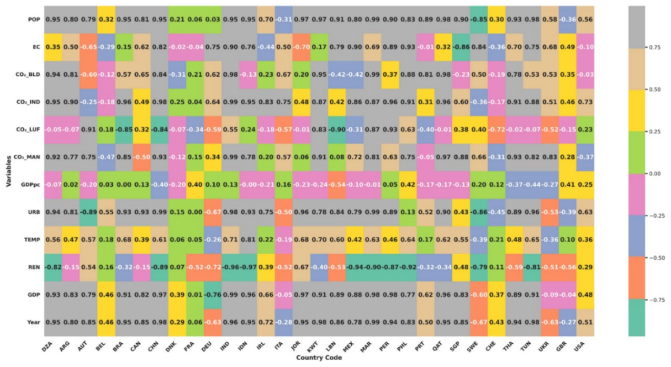


Fig. 1 Correlation of Transport CO_2 Emissions across All Countries.

On the other hand, some developed countries had weak or even negative associations between socioeconomic indicators and TCO_2. For instance, Sweden has negative correlations between socio-economic factors and emission trends ($-0.86 \leq r \leq -0.60$). Likewise, Germany and Austria had significant negative correlations with GDP ($r = -0.76$ in Germany) and even urbanization rate in both cases ($r = -0.67$ in Germany

Copyright © 2026
ISSN: 2961-6611

and r = −0.89 in Austria). This implies that development and urbanization in some developed countries may be largely unassociated with carbon emissions. The Year variable shows a strong positive correlation with transport CO_2 emissions in most countries, peaking in Morocco (MAR, r = 0.99). In contrast, negative correlations are observed in Sweden (SWE, r = -0.67), Germany (DEU, r = -0.63), Ukraine (UKR, r = -0.63), and Italy (ITA, r = -0.28). The surface temperature reaches its highest positive correlation in Indonesia (IDN, r = 0.81), while the strongest negative correlations occur in Sweden (SWE, r = -0.39), Ukraine (UKR, r = -0.36), and Italy (ITA, r = -0.19), highlighting country-specific relationships between temperature and transport emissions.

The use of renewable energy (REN) showed a negative relationship with emission-related factors for many nations. The strongest negative relationship was found for Indonesia (r = -0.97), India (r = -0.96), Mexico (r = -0.94), Morocco (r = -0.90), and China (r = -0.89), suggesting the possible effectiveness of renewable energy in decreasing carbon emissions. Furthermore, sectoral emissions related to buildings (CO_2_BLD), industry (CO_2_IND), and manufacturing (CO_2_MAN) showed strong links with TCO_2 in certain developing nations but weak or negative associations in some developed countries. Overall, it is clear that there are large differences among countries regarding the relevant factors related to CO_2 emissions in transport sectors.

B. Cross-Country Comparison of the Performance of Three Predictive Models

RF exhibits the highest performance out of all analyzed algorithms. In particular, it provides the lowest error values (MAE = 3.63, RMSE = 7.36, and MAPE = 3.08%), suggesting high prediction precision and good stability, as shown in Tab2. The value of R^2 = 0.87 also proves the high capacity to describe the variability in the target variable, while the low relative RMSE (6.76%) confirms the same. While SVR demonstrates relatively good performance when we compare it to RR (MAE = 9.79, RMSE = 13.96, MAPE = 6.41%), its low R^2 = 0.71 proves that it has limited capacity for describing the variance of the target variable. Thus, SVR proves to have reasonable predictive power; however, it might fail to model the nonlinearity of processes within the system. On the contrary, the poor performance is demonstrated by Ridge Linear Regression (RR), which provides relatively high prediction errors (MAE = 11.06, RMSE = 22.58, and MAPE = 10.10%). Moreover, in addition to other metrics, this algorithm exhibits the negative value of R^2 = -40.08. Thus, RR cannot be regarded as a reasonable choice for modeling the processes at hand, especially considering the negative value of R^2 = -40.08. On the other hand, the RR model failed to capture the underlying nonlinear relationship between the independent and dependent variables used in our study. Table 2 summarizes the performance of machine learning models used in our study across all studied countries.

TABLE III

PERFORMANCE COMPARISON OF MODELS FOR TRANSPORT CO_2 EMISSION PREDICTION.

Model	MAE	RMSE	MAPE (%)	R^2	rRMSE (%)
SVR	9.79	13.96	6.41	0.71	10.41
RL	11.06	22.58	10.10	-40.08	18.93
RF	3.63	7.36	3.08	0.87	6.76

C. Country-Level Performance Assessment of the RF Model

As demonstrated in Section III.B, the RF achieved the highest predictive performance among all machine-learning approaches evaluated in this study. To further assess its robustness and transferability across all countries chosen in our study and different national contexts, we evaluated country-level model performance based on rRMSE. Additionally, a regression analysis was carried out to compare the observed and modeled TCO_2-equivalent emissions, using the rRMSE classification, which allowed for the determination of which countries performed best and worst, as well as provided additional information regarding the accuracy and ability of the model to generalize its predictions (Fig.2).

Copyright © 2026
ISSN: 2961-6611

Proceedings of Book Series PBS -Vol 42, pp.14-21 **Advances in Energy Transition and Environmental Impacts**

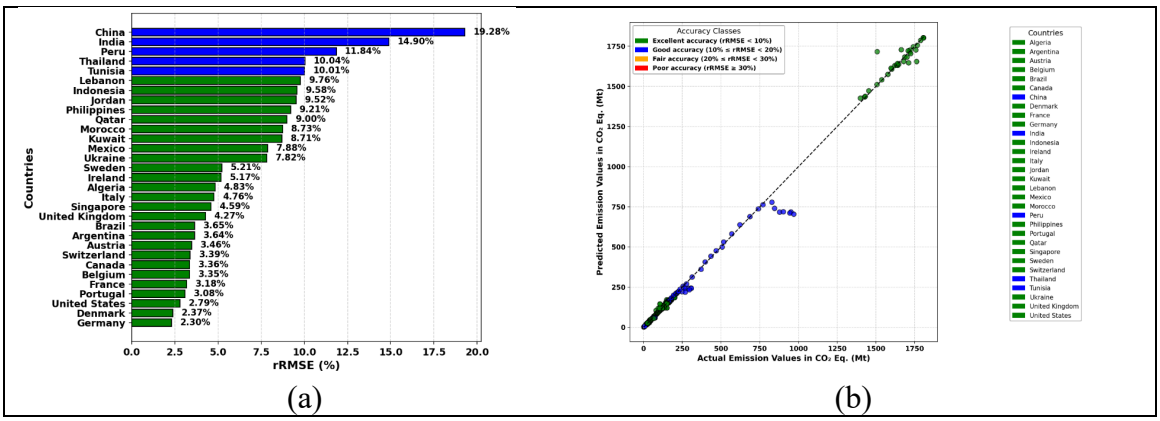


Fig. 2 Machine learning model efficiency for transport-related CO_2 emissions: (a) Country-specific rRMSE (%) values ranked by prediction accuracy. (b) Scatter plot of observed versus predicted TCO_2 emissions (Mt).

As shown in figure 2(a), the bar charts indicate that Germany (2.30%), Denmark (2.37%), the United States (2.79%), Portugal (3.08%), and France (3.18%) all yielded rRMSE values below 10%. According to commonly adopted performance criteria, these results indicate excellent predictive accuracy (green category). In addition, we observe that China (19.28%); India (14.90%); Peru (11.84%); Thailand (10.04%); and Tunisia (10.01%) have rRMSE values ranging between 10% and 20% (blue category), showing a satisfactory prediction. The regression plot demonstrates a strong agreement between observed and predicted CO_2equivalent emissions, with most data points from all countries lying close to the reference line. As shown in figure 2(b), we observe accurate predictions across the studied countries. However, we can see clearly a slight underprediction for countries classified as having good predictive performance (blue), particularly for emission levels above 250 Mt, around 750 Mt, and below 1000 Mt. However, for countries classified as having excellent predictive performance, we observe only minor deviations from the reference line, with limited cases of both underprediction and overprediction at higher emission levels. These findings confirm not only the robustness and effectiveness of the RF model in predicting CO_2emissions from the transport sector but also its capability to provide accurate and reliable predictions across different emission levels among the 31 countries included in this study.

D. Country-Level Performance Assessment of the RF Model

The correlation analysis supports earlier evidence found in the literature and shows a link between socioeconomic factors and TCO_2 emissions. In North African countries as well as Asian countries, transport-related emissions of $CO_2$are positively correlated and strongly related to GDP, population, and urbanisation rate. These findings are consistent with those reported in various studies ([17], [18]). However, among some developed countries (Germany and Sweden), the GDP was found to be either weakly correlated or negatively correlated with their TCO_2 emissions, suggesting a degree of decoupling between these two variables. Similar results have been reported in several studies ([19], [20]).

The predictive results further support these observations. The RF model achieved strong overall performance across the 31 countries studied, with a MAPE of 3.08%, an rRMSE of 6.76%, and an R^2 of 0.87. When we compare our findings to a study focusing on the top 10 African CO_2emitters [16], which reported higher R^2 values (approximately 0.95-0.96). In addition, another study [15] focusing on G20 countries also achieved slightly better predictive performance. However, these studies were conducted on more homogeneous datasets characterized by relatively similar economic, demographic, and energy-consumption patterns. In contrast, the present study encompasses a highly heterogeneous dataset including 31 countries with diverse socio-economic, demographic, climatic, and transport-sector characteristics. Such diversity increases the complexity of the prediction task and naturally reduces the achievable level of explained variance. Therefore, although the RF model shows a slightly lower R^2 than those reported in previous studies, it demonstrates strong predictive capability, robustness, and generalization ability across a broad range of heterogeneous contexts. These results suggest that the RF can be used in more challenging and variable conditions, highlighting its applicability for large-scale international assessments of TCO_2.

Copyright © 2026
ISSN: 2961-6611

Proceedings of Book Series PBS -Vol 42, pp.14-21 **Advances in Energy Transition and Environmental Impacts**

IV. CONCLUSIONS

In this study, we examined, first, TCO_2 emissions across 31 countries using various explanatory variables such as socioeconomic, energy, temporal, and even meteorological variables, and second, we evaluated the performance of three machine-learning models for emission prediction. The correlation analysis revealed significant differences and heterogeneity among G31 countries, with year, population growth, economic development, and urbanization rate being more correlated than the other variables used in our study, particularly in several developing and emerging economies. However, we observed weak or negative correlations in some developed countries, suggesting that economic growth and even sectorial emissions can be decoupled from transport-related carbon emissions. Renewable energy consumption and TCO_2emissions were generally found to be negatively correlated. These results confirm the importance of clean energy transitions in mitigating the transport sector. Among all approaches used in this study, the RF algorithm yielded the highest predictive accuracy when we compared it to SVR and RR, while RR was inadequate for the prediction tasks of heterogeneous contexts. The results also show how important it is to consider the political strategies and the specific situation in each country, especially for reducing and predicting TCO_2emission tasks. Our future research could incorporate additional explanatory variables, longer time series, and scenario-based analyses to further improve prediction accuracy and support evidence-based transport and climate policies.

ACKNOWLEDGMENT

The authors would like to sincerely thank the Laboratory of Metrology and Energy Systems, National Engineering School of Monastir, University of Monastir, 5019 Monastir, Tunisia, for providing the research environment, and financial support.

REFERENCES

[1] Ho, W.T.; Yu, F.W. Optimal selection of predictors for greenhouse gas emissions forecast in Hong Kong. J. Clean. Prod. 2022, 370, 133310.
[2] Agreement, P. (2017). United nations framework convention on climate change (unfccc). (2015). Climate Change Secretariat: Bonn, Germany.
[3] International Energy Agency. CO2 Emissions in 2022; International Energy Agency: Paris, France, 2023.
[4] Ene Yalçın, S. Development of A Forecasting Framework Based on Advanced Machine Learning Algorithms for Greenhouse Gas Emissions. Systems 2024, 12, 528.
[5] Wang, H., Wang, X., Yin, Y., Deng, X., & Umair, M. (2024). Evaluation of urban transportation carbon footprint–Artificial intelligence based solution. Transportation Research Part D: Transport and Environment, 136, 104406.
[6] Ağbulut, Ü. (2022). Forecasting of transportation-related energy demand and CO2 emissions in Turkey with different machine learning algorithms. Sustain Prod Consum 29: 141–157.
[7] Alam, G. M. I., Arfin Tanim, S., Sarker, S. K., Watanobe, Y., Islam, R., Mridha, M. F., & Nur, K. (2025). Deep learning model based prediction of vehicle CO2 emissions with eXplainable AI integration for sustainable environment. Scientific Reports, 15(1), 3655.
[8] Guo, X., Kou, R., & He, X. (2024). Towards Carbon Neutrality: Machine Learning Analysis of Vehicle Emissions in Canada. Sustainability, 16(23), 10526.
[9] Kazancoglu, Y., Ozbiltekin-Pala, M., & Ozkan-Ozen, Y. D. (2021). Prediction and evaluation of greenhouse gas emissions for sustainable road transport within Europe. Sustainable Cities and Society, 70, 102924.
[10] Lu, I. J., Lewis, C., & Lin, S. J. (2009). The forecast of motor vehicle, energy demand and CO2 emission from Taiwan's road transportation sector. Energy policy, 37(8), 2952-2961.
[11] Huang, S., Xiao, X., & Guo, H. (2022). A novel method for carbon emission forecasting based on EKC hypothesis and nonlinear multivariate grey model: Evidence from transportation sector. Environmental Science and Pollution Research, 29(40), 60687-60711.
[12] Qiao, Q., Eskandari, H., Saadatmand, H., & Sahraei, M. A. (2024). An interpretable multi-stage forecasting framework for energy consumption and CO2 emissions for the transportation sector. Energy, 286, 129499.
[13] Ratanavaraha, V., & Jomnonkwao, S. (2015). Trends in Thailand CO2 emissions in the transportation sector and Policy Mitigation. Transport Policy, 41, 136-146.
[14] Çınarer, G., Yeşilyurt, M. K., Ağbulut, Ü., Yılbaşı, Z., & Kılıç, K. (2024). Application of various machine learning algorithms in view of predicting the CO2 emissions in the transportation sector. Science and Technology for Energy Transition, 79, 15.

Copyright © 2026
ISSN: 2961-6611

Proceedings of Book Series PBS -Vol 42, pp.14-21 **Advances in Energy Transition and Environmental Impacts**

[15] Ouni, M., Mrad, S., & Mraihi, R. (2025). Pathways to sustainable transportation in G-20 countries: Unveiling the role of green technology, green energy, green finance and digital economy using panel data and machine learning analyses. Research in Transportation Business & Management, 60, 101355.

[16] Adam, A. M., Yusuf, A. M., Amir, I. Y., Arab, O. A., & Ali, O. A. (2025). Predicting transportation-related greenhouse gas emissions in Africa's top emitters: a machine learning approach. Environmental Sciences Europe, 37(1), 166.

[17] Godil, D. I., Yu, Z., Sharif, A., Usman, R., & Khan, S. A. R. (2021). Investigate the role of technology innovation and renewable energy in reducing transport sector CO2 emission in China: A path toward sustainable development. Sustainable Development, 29(4), 694–707. https://doi.org/10.1002/sd.2167.

[18] Li, Y., Du, Q., Lu, X., Wu, J., & Han, X. (2019). Relationship between the development and CO2 emissions of transport sector in China. Transportation Research Part D: Transport and Environment, 74, 1–14. https://doi.org/10.1016/j.trd.2019.07.011.

[19] Liu, X., Razzaq, A., Shahzad, M., & Irfan, M. (2022). Technological changes, financial development and ecological consequences: A comparative study of developed and developing economies. Technological Forecasting and Social Change, 184, Article 122004. https://doi.org/10.1016/j.techfore.2022.122004.

[20] Raihan, A. (2023). Economy-energy-environment nexus: The role of information and communication technology towards green development in Malaysia. Innovation and Green Development, 2(4), Article 100085. https://doi.org/10.1016/j. igd.2023.100085 .

Copyright © 2026
ISSN: 2961-6611