Proceedings of the 8th IEOM Bangladesh International Conference on Industrial Engineering and Operations Management, Dhaka, Bangladesh, December 20-21, 2025

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/BA08.20250232
Published: December 20, 2025

Multi-Model Machine Learning Analysis of Ecological Load Ratio Using Classification, Regression, and Clustering Approaches in EU Countries

Anik Debnath and Akash Sikdar Sudip
Department of Industrial and Production Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
anikdbnath.mcc@gmail.com, akashsikdarsudip92@gmail.com

Souvik Mandol
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
souvikmt99@gmail.com

.

Abstract

In this research, ecological pressure on 27 European Union countries is analyzed using simulations of two supplementary dimensions: ELR-Class and Ecological Load Ratio (ELR), which are highly dependent on the parameters named GDP growth (GDPG), manufacturing value added (MVA), personal remittances received (PRR), urban population (UP), electricity production from oil, gas, and coal (EPR), forest rents (FR), and population density (PD), extracted by PSO from a large feature space. For the ELR-Class problem, comparison analysis has been done among supervised classification techniques and Gaussian algorithm-backed unsupervised clustering techniques. These supervised models show high accuracy values of more than 0.98 and perform significantly better than the unsupervised clustering models because these models do not work well under the condition of significant data imbalance between the two stages of ecological condition: ecological balance and ecological overshoot. In the regression analysis, done separately for ELR variable using supervised models with an emphasis on ensemble-based models, R^2 value of more than 0.98 was recorded for the models named ABR, and LGBR exhibiting the outstanding ability of capturing higher non-linearity.

Keywords
ELR, ELR-Class, Supervised Model, Unsupervised Model, Gaussian Algorithm, PSO

1. Introduction

Ecological degradation is being increasingly thrust into focus because of fast-paced economic expansion, industrialization, urbanization, and population pressure that raise human pressure on nature. While most research studies assess this pressure using ecological footprint (EF) measures, EF calculates total pressure without factoring ecological capacity. To fill this gap, this study focuses on ecological load ratio (ELR), ratio of ecological footprint and bio-capacity, because it measures ecological footprints in relation to biocapacity and provides a more direct and interpretable assessment of ecological sustainability. Although sound theoretically, with much research left to be explored, ecological load ratio is presently underutilized in predictive studies using conventional methods in statistics and econometrics due to linear assumptions. Due to strong nonlinear relationships among ecological, economic, and demographic factors, machine learning methods seem more suitable in this research work. In this study, a variable named ELR-Class, used as output parameter for classification and clustering, is defined to describe the ecological scenario whether it is sustainable or not based on corresponding ELR value, ELR having values less than or equal to one means a balanced or sustainable ecological condition (Ecological Balance), on the other hand ELR values greater than one describe the imbalanced situation (Ecological Overshoot).

1.1 Research Objectives

1) How well can supervised machine learning models classify ecological states, such as Ecological Balance and Ecological Overshoot, with different energy-economic factors as inputs?
2) Can unsupervised machine learning algorithms detect hidden ecological trends in this dataset, and how well are these algorithms in performance compared to supervised machine learning algorithms used for classification?
3) To what extent can supervised regressive machine learning models be used to make accurate predictions of ELR, and which factors have come forward as being most important using model-generated feature importance?

2. Literature Review

The pace of development in machine learning (ML) algorithms has significantly changed the face of modeling and forecasting in the environmental sector. In contrast, classical econometric models, such as linear regressions, ARDL models, and cointegration, are less appropriate because they demand stationarity, linearity, and limited interaction effects. These models lack flexibility as they are constrained by linearity and limited interaction, which reduces their efficiency and capability to address the complex relationships between eco-social variables and environmental pressure. The increasing reliance on diverse global data in calculating the ecological footprint (EF) and Eco-Load Ratio (ELR) demands a more efficient tool like machine learning.

Supervised machine learning models have been increasingly used to forecast EF values at the geographical and groups of countries level. The advantage of tree-based machine learning models in forecasting EF in the G-20 countries, as evident by the use of various macroeconomic variables, is illustrated by Ahmad Roumiani and Mofidi (2021). The efficiency of machine learning models, including Random Forest, Gradient Boosting, and XGBoost, in distinguishing non-linear relationships among various energy-related variables and EF, is illustrated by Janković et al. (2020). In Turkey (Ahmed et al., 2020) and Afghanistan (Arian et al., 2025), the efficiency of supervised machine learning models over other econometric approaches in explaining the relationship between various socio-economic and development variables and EF is evident. In the general setup of various regions, the contribution of natural resource utilization and demography as determinants of EF in China, as evident by Pata and Isik (2021), is significantly well explained by machine learning models compared to ARDL models.

Besides the use of tree-based models in ensembling, other deep learning models have shown great success in predicting the values of ecological footprint as well. In this case, through research by Cihan (2024), classical, hybrid, and deep learning models in the field of time series are compared, and it is clear that those with neural networks are superior to classical models in understanding longer-term behavior. In affirmation of this fact, Moros-Ochoa et al. (2022) accurately forecast the values of both the actual ecological footprint and the biocapacity through the use of artificial neural networks. The effectiveness of deep learning models is demonstrated by other research by Van der Woude et al. (2024) carried out in Latin America and Caribbean nations.

Although they are highly predictive, supervised machine learning models still have several restrictions. The main issue is interpretability, as most very successful models are now often described as 'black-box models,' and it is difficult to understand processes through which input variables push outputs into certain ecologic outcomes. Feature importance tools are known to resolve this issue to some extent, although all authors agree on insufficiency of 'black-box models' to be fully transparent in ecologic decision-making.

The use cases of Unsupervised Learning approaches by Mostafa (2010) on nation-based clusters through Kohonen's Self-Organizing Maps were successful, and their applications on global ecologies today, as shown by Zhang et al. (expected 2025) on geopolitical risk and ecologic footprint, establish the fact that ecologic patterns, hidden and possibly neglected by Supervised Learning, can also be detected and quantified through Learning approaches. The applicability and potential of Clustering are realized well in cases wherein ecologic balance and ecologic overshoot can be expressed through intricate relationships among GDP growth, urbanization, ecologic density, and forest size, and are possibly non-linear and less correlated.

Unsupervised learning techniques, namely KMeans, Gaussian Mixture Model (GMM), Bayesian GMM, DBSCAN, and HDBSCAN, used in this study, are beneficial for delineating observations in a similarity-driven manner, as opposed to being driven by labels. In GMM and Bayesian GMM, clusters can be assumed to be of varying shapes, thereby accounting for overlapping patterns of ecological states. DBSCAN and HDBSCAN, which are density-based techniques, can identify clusters and outliers of arbitrary shape, allowing the transition zone from ecological overshoot to balance to be identified, which can play a significant role in modeling, as structural changes precede large-scale ecologic damage.

Another emerging trend in EF research is the utilization of dimensionality reduction techniques. Many modern approaches, including UMAP, allow researchers to discover low-dimensional patterns within high-dimensional data on socio-economic phenomena. The utilization of UMAP is now gaining popularity as researchers attempt to discover hidden patterns in EF, patterns that may remain undetected by PCA or correlation matrices.

Cross-country studies also strengthen the relevance of ML to research on the environment. In this case, Kumaran et al. (2024) employed both ML and ARDL approaches to investigate the effect of income inequality, forest areas, and technology on the ecological footprint in Indonesia, illustrating through the findings that forest-related variables play an important role in reducing the level of environmental pressures. In general, research studies on the African continent (Castro-Nieto et al., 2023) have illustrated through the findings, and through various studies, that population growth, resource depletion, and, indeed, various economic factors all play crucial yet highly non-linear roles in causing ecologic degradation.

In a critical void within the current body of literature is the fact that virtually all research is directed toward forecasting EF, with very few studies on ELR, although ELR is clearly the more interpretable metrics compared to BC. In addition, no studies were identified within the body of literature surveyed as predicting all, or any, of the three forecasting tasks by combining supervised regression, supervised classification, and Gaussian-supported unsupervised clustering within one modelling process. This issue is remedied within this research by being comprised of supervised regression (DTR, ETR, RFR, ABR, GBR, LGBR, XGBR) for forecasting continuous ELR, supervised classification (DTC, ETC, RFC, ABC, GBC, LGBC, XGBC) for predicting either EB or EO, and unsupervised clustering (KMeans, SphericalKMeans, GMM, BGMM, DBSCAN, HDBSCAN) for locating hidden patterns within ecological systems and transitions.

In general, there is much improvement in the use of ML in forecasting the value of EF and in recognizing the factors that affect it. The use of regression, classification, and Gaussian-driven clustering on ELR and ELR-Class, as is done in this research, is, in fact, the first contribution to modeling ecologies, as this combines all factors into a whole.

3. Methods

3.1 Methodological Overview

The research process began with data acquisition and preprocessing, involving imputation and scaling, and proceeded with feature selection by means of metaheuristic algorithms (GA and PSO), exploratory data analysis, and the training of models, namely supervised classification, supervised regression, and Gaussian-backed unsupervised models. The evaluation process was carried out by calculating appropriate metrics, and feature analysis was performed as part of the process (Figure 1).

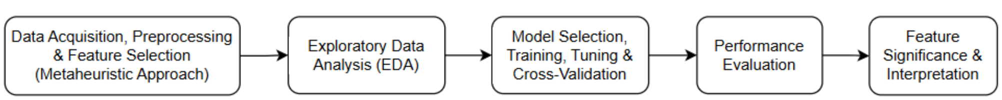

Figure 1. Flowchart of the Process

3.2 Data Acquisition, Preprocessing and Optimal Feature Selection

For analysis, this study employed data collected from 27 European Union countries, providing a cumulative total of 837 observations derived from an original set of close to 120 features. For feature selection, criteria were set based on established research studies with individual variable acquisition done systematically through World Development Indicators (WDI). For target variable identification, ELR and ELR-Class were acquired separately via Global Footprint Network. Data sets were then compiled into a single document and then the features with more than 20% missing values were removed from the dataset to ensure quality of research output. Missing values were then treated using K-Nearest Neighbors (KNN) similarity-based imputation techniques to account for similarity patterns among observations. To correct issues of distributional skewness and stabilized variances among these observations, transformation was done using Yeo-Johnson transformation method with transformation done on the whole set of observations. Lastly, metaheuristic-based Genetic Algorithm and Particle Swarm Optimization were performed, with the optimal feature subset selected based on cross-validated Random Forest performance, where Particle Swarm Optimization outperformed the Genetic Algorithm, to determine dominant features with seven features selected: GDP growth (GDPG), manufacturing value added (MVA), personal remittances received (PRR), urban population (UP), electricity production from oil, gas, and coal (EPR), forest rents (FR), and population density (PD) applied uniformly for classification, clustering, and regression analysis processes of the study (Figure 2).

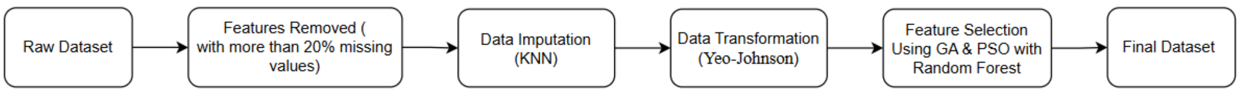

Figure 2. Data acquisition, preprocessing, and optimal feature selection workflow

3.3 Exploratory Data Analysis

Figures 3 and 4 show the Pearson correlation matrices for ELR-Class and ELR, respectively. The Pearson correlation matrices show moderately to very weak correlations between the energy-economic variables and target indicators, which implies that linear relationships are not sufficient in explaining ELR behavior for classification and regression. Yet, this implies that linear correlation, as a method of quantifying the strength of ELR relationships, is inadequate, as likely ELR relationships are non-linear.

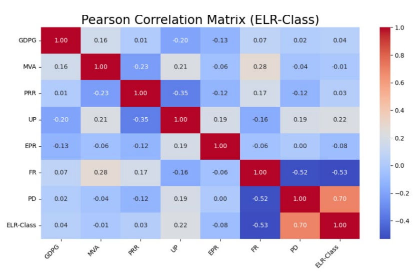

Figure 3. Pearson Correlation for Classification Approach

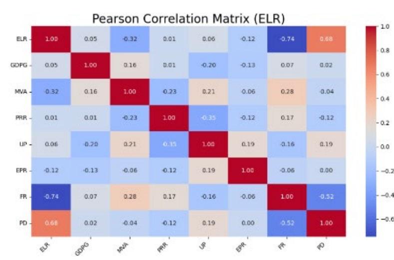

Figure 4. Pearson Correlation for Regression Approach

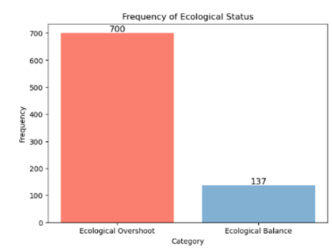

Figure 5. Count Plot

In this case, the diagram shown in Figure 5 represents the classes and their distinct distribution, where Ecological Overshoot clearly exceeds Ecological Balance by 700 against 137 instances, thereby ensuring the need to use sound modelling techniques, as statistics will fail to emphasize the non-linear relationships shown by ecological datasets.

3.4 Model selection, Training, Tuning, and Validation

In the current research, both supervised and unsupervised learning approaches were used for classifying the ecological status (ELR-Class) and the prediction of ELR. In order to identify ELR-Class, being part of the supervised learning task, seven models were trained, namely Decision Tree Classifier (DTC), Extra Trees Classifier (ETC), Random Forest Classifier (RFC), Ada Boost Classifier (ABC), Gradient Boosting Classifier (GBC), LightGBM Classifier (LGBC), and XG Boost Classier (XGBC). The choice of models was guided by their superior non-linear modeling properties, and by the fact that ensemble learning offers significant improvements in model performance. Stratified K-Folds Cross-Validation was adopted as part of the modeling process to avoid issues associated with accurate performance evaluation and overfitting. The choice of hyperparameters in all models was optimized through Bayesian optimization technique called Optuna.

In case of continuous ELR, seven supervised regression models were utilized, namely, Decision Tree Regressor (DTR), Extra Trees Regressor (ETR), Random Forest Regressor (RFR), Ada Boost Regressor (ABR), Gradient Boosting Regressor (GBR), LightGBM Regressor (LGBR), and XG Boost Regressor (XGBR). The evaluation of these models is carried out by employing 'K-Fold Cross-Validation,' and hyper parameters were optimized through Optuna.

In this research, unsupervised learning techniques were used to identify hidden patterns in the ecologies by employing Gaussian and density-based approaches. Clustering algorithms used were KMeans, Spherical, Gaussian Mixture Models (GMM), Bayesian GMM, DBSCAN, and HDBSCAN. Silhouette Score and BIC were used as evaluation metrics. Also, as the clusters produced by the unsupervised learning techniques are unlabeled, the Hungarian algorithm was used to compact the clusters generated by unsupervised algorithms into two clusters

3.5 Supervised ML Classification Algorithms

The performance of the supervised classifiers is presented in Table 1. The best-performing Gradient Boosting algorithm (GBC) obtained the best validation accuracy of 0.9867 ± 0.0082, and this is closely followed by the XGBoost (XGBC) and AdaBoost (ABC) classifiers, all of which generalise with less variance. The lowest-performing algorithm is the Decision Tree (DTC) in terms of accuracy, indicating the algorithm's relative weakness in modeling complex boundaries.

Table 1. Accuracy and F1- Score for Different Supervised Classification Model

Model	Accuracy		F1-Score	
	Train	Val	Train	Val
DTC	0.9860 ± 0.0038	0.9809 ± 0.0082	0.9917 ± 0.0023	0.9887 ± 0.0048
ETC	0.9944 ± 0.0016	0.9854 ± 0.0085	0.9966 ± 0.0010	0.9913 ± 0.0051
RFC	0.9918 ± 0.0024	0.9846 ± 0.0083	0.9951 ± 0.0015	0.9908 ± 0.0049
ABC	0.9984 ± 0.0012	0.9855 ± 0.0105	0.9990 ± 0.0007	0.9914 ± 0.0063
GBC	0.9993 ± 0.0007	0.9867 ± 0.0082	0.9996 ± 0.0004	0.9921 ± 0.0049
LGBC	0.9950 ± 0.0019	0.9839 ± 0.0089	0.9970 ± 0.0011	0.9904 ± 0.0053
XGBC	0.9996 ± 0.0006	0.9859 ± 0.0087	0.9998 ± 0.0004	0.9916 ± 0.0051

The F1-Score values also follow the same trend, and the models with the highest F1-Score are GBC, XGBC, and ABC, whereas the lowest F1-Score is shown by DTC. The ensemble models have been performing well compared to the single-tree classifier.

3.6 Feature Importance and Interpretation from Supervised Models

From Table 2, there are distinct trends in the level of importance of the input attributes. PD is always ranked as the most important contributor for all the supervised models, confirming the crucial role of the concentration of population in deciding ELR-Class. The presence of FR as the second most important contributor indicates the significance of forest-related economic activities. At the other end, both MVA and PRR were consistently ranked as the feature with the lowest level of importance, indicating that the level of manufactured value added and remittance amount are less significant in influencing the ELR variable. The other features, namely GDPG, UP, and EPR, ranked at the middle positions.

Table 2. Feature Importance Ranking of Input Parameters across Supervised Classification Models

Model	Features						
	GDPG	MVA	PRR	UP	EPR	FR	PD
DTC	2	7	6	3	5	4	1
ETC	5	7	4	3	6	2	1
RFC	4	7	6	3	5	2	1
ABC	3	7	6	2	4	5	1
GBC	3	6	7	4	5	2	1
LGBC	3	7	5	4	6	2	1
XGBC	6	7	5	4	3	2	1

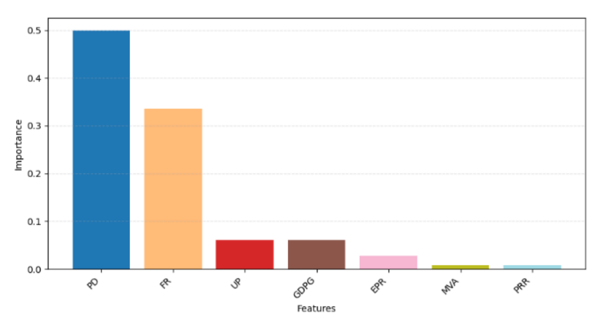

Figure 6. Feature Importance for Gradient Boosting Classifier

The above observations are further supported by Figure 6, which shows the same dominance of PD and FR in the GBC model as those factors play a leading role in recognizing balance versus overshoot ecologies.

3.7 Unsupervised ML Algorithms

The following is the summary of the results obtained from unsupervised clustering models used with the Hungarian algorithm (Table 3). Each model, regardless of the differing number of clusters, has achieved two clusters after the alignment process. The accuracy score obtained by all models was 0.9104, except for the model GMM which obtained 0.92. The accuracy score obtained by GMM was higher than those of other models because of the flexibility and capability of the model to handle mixed distributions. The other models were still capable of obtaining accurate models following the Hungarian algorithm.

Table 3. Summary of Results from Unsupervised Models after Matching

Model	Number of Clusters Before Matching	Number of Clusters after Matching	Accuracy (Matched)
KMeans	3	2	0.9104
Spherical	3	2	0.9104
GMM	10	2	0.92
Bayesian GMM	3	2	0.9104
DBSCAN	11	2	0.9104
HDBSCAN	5	2	0.9104

The result of the GMM clustering analysis is assessed by the use of the confusion matrix (Figure 7) and the UMAP transformation (Figure 8). The confusion matrix indicates that most of the Ecological Overshoot cases (680) and almost all Ecological Balance cases (90) were correctly assigned to their paired clusters, while only few cases were assigned incorrectly (47 and 20 cases, respectively). The UMAP transformation in Figure 8 depicts two well-separated groups, confirming the successful identification of the two distinct ELR patterns by the GMM.

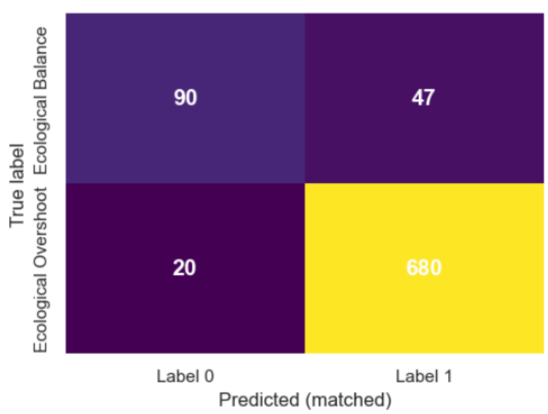

Figure 7. Confusion Matrix of GMM with Hungarian Matching

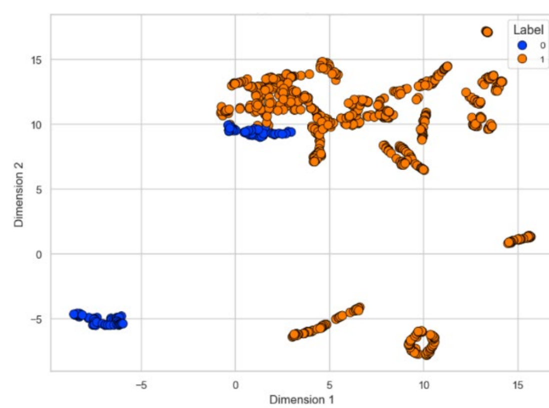

Figure 8. UMAP Visualization of GMM Clusters after Hungarian Matching

3.8 Supervised ML Regression Algorithms

The performance of the regression models for predicting ELR is shown in Table 4. From the validation R^2 values, it is clear that the model with the maximum predicting capability is indeed ABR, followed by LGBR, and ETR is the poorest, as it recorded the lowest value for R^2, thereby having less explaining power. On similar grounds, ETR registered the maximum error, whereas LGBR recorded the lowest, implying high accuracy in calculating ELR. The maximum value for MSE is recorded by ETR, whereas the lowest is by ABR, thereby proving it to be the best model on the whole, combining both low error and high accuracy. Ensemble models, namely ABR and LGBR, are much superior to the non-ensemble models.

Table 4. Performances of ML Regression Models

Model	R^2		MAE	MSE
	Train	Val	Val	Val
DTR	0.9618 ± 0.0065	0.9260 ± 0.0440	0.1665 ± 0.0238	0.2614 ± 0.0706
ETR	0.9194 ± 0.0055	0.9004 ± 0.0136	0.2438 ± 0.0134	0.3137 ± 0.0231
RFR	0.9813 ± 0.0010	0.9617 ± 0.0092	0.1311 ± 0.0119	0.1935 ± 0.0224
ABR	0.9953 ± 0.0003	0.9823 ± 0.0059	0.0866 ± 0.0075	0.1306 ± 0.0216
GBR	0.9888 ± 0.0007	0.9686 ± 0.0062	0.1186 ± 0.0078	0.1757 ± 0.0171
LGBR	0.9997 ± 0.0000	0.9815 ± 0.0057	0.0847± 0.0081	0.1339 ± 0.0216
XGBR	1.0000 ± 0.0000	0.9771 ± 0.0083	0.0934 ± 0.0104	0.1483 ± 0.0272

3.9 Feature Importance and Interpretation from Supervised Regression Models

In regard to the feature importances for the supervised regression models predicting ELR, this is shown in Table 5. From the regressors, it is clear that, on average, FR and PD were always ranked as the most important features, consistent with what was observed in the supervised classification models too. This consistency shows that, regardless of whether it is classification or regression, forest rents and population density are always core drivers of ELR. On the other hand, GDPG and PRR were consistently ranked as the least influential features, whereas the other features, namely MVA, UP, and EPR, ranked in the middle positions, although with differing categorizations compared to those in the classification task. This implies that, although the set of most influential features is unchanged regardless of the modeling methodology, the strength of influence of middle- and low-ranking features shifts depending on whether the model is learning ecological status or ELR output.

Table 5. Feature Importance Ranking of Input Parameters across Supervised Classification Models

Model	Features						
	GDPG	MVA	PRR	UP	EPR	FR	PD
DTR	7	5	6	3	4	1	2
ETR	7	4	5	3	6	1	2
RFR	7	4	5	3	6	2	1
ABR	6	4	7	1	5	3	2
GBR	6	4	5	3	7	2	1
LGBR	7	5	4	2	6	3	1
XGBR	7	5	4	3	6	1	2

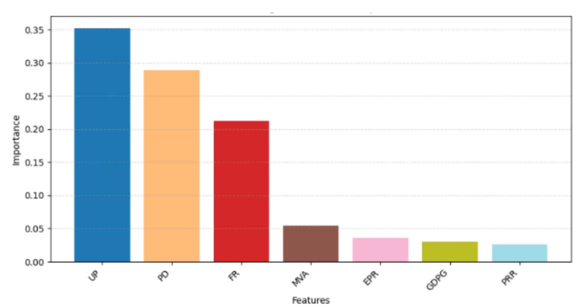

Figure 9. Feature Importance for Ada Boosting Regressor

The trend is also evident in Figure 9, which shows for the ABR case the features with the largest importance values are those already identified as significant, namely FR, PD, and UP.

4. Results

The following is a comparison of the classifiers' accuracy and the accuracy of the various clustering models on the ELR-Class prediction task, as shown in Figure 10. The supervised models, especially the GBC, XGBC, and ABC, were the three best of in terms of accuracy, with all accuracy values above 0.985. Notably, models like GMM and DBSCAN, which are unsupervised models, were less accurate compared to the supervised models as shown by the accuracy distribution, which indicates the level of complexity involved in modeling the ELR-Class without considering the labels.

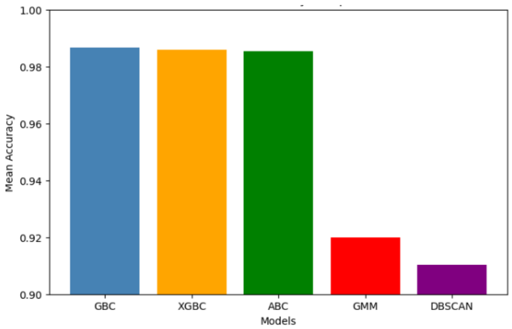

Figure 10. Accuracy Comparison between Different Models

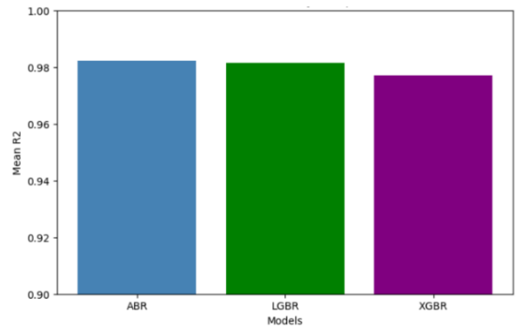

Figure 11. R^2 Score Comparison between Different ML Models to Forecast ELR

In Figure 11, the forecasting capability of the best regression models used to forecast the ELR value is shown. The accuracy level revealed by the results is close to perfect, as all three models, namely ABR, LGBR, and XGBR, have been able to attain a high level of mean R^2 value, all of which are over 0.97. The high level of forecasting accuracy obtained by the models indicates that they were successful in modeling the ELR variation process. Together, Figures 10 and 11 illustrate that, whereas the value of insights obtained through clue-based clustering is significant, the accuracy achieved through classification and regression is significantly superior.

5. Limitations and Future Scope

One key limitation of this work is the non-proportionate distribution of the ecological condition, as Ecological Balance is represented by only 137 cases, whereas Ecological Overshoot is represented by 700 cases. This may have directly affected the performance of the unsupervised learning algorithms used in this research, which are very much dependent on the level of data density.

In spite of the mentioned limitation, the results reveal various opportunities for research on this topic in the future. Firstly, the use of Artificial Neural Networks (ANNS) may potentially improve the accuracy level of ELR prediction and classification, especially on non-linear relationships. Secondly, the combination of deep learning and symbolic regression may be a promising avenue to pursue, which may provide insights on obtaining 'meaningful' mathematical equations for understanding EL dynamics in an interpretable manner.

6. Conclusion

In this research, the proposed machine learning approach was used to evaluate ecological pressure on the basis of both categorical (ELR-Class) and continuous (ELR) variables. The supervised learning classification models performed very well, all of them exhibiting an accuracy greater than 0.98, always beating the unsupervised machine learning models and focusing on the determination of ecologic balance and ecologic overshoot. In the case of ELR forecasting, the regression models, namely, ABR, LGBR, and XGBR, were found to possess high accuracy rates, as evidenced by the high value of R^2, always above 0.97. This study has found that population density and forest rents were the most significant factors across supervised classification and regression models, whereas mid and lower ranking features were different for supervised classification in comparison to regression models. The overall outcome of this research represents a strong data-driven platform in understanding ecological patterns and activities and thereby helping to formulate well-informed policies and sustainability practices.

References

Ahmed, Z., Zhang, B., and Cary, M. Linking economic globalization, economic growth, financial development, and ecological footprint: Evidence from symmetric and asymmetric ARDL. *Ecological Indicators* 121, **2020**.

Arian, A. B., Nazary, M. N., Karimi, A. Z., and Obiad, M. Predicting the impacts of key development indices on the ecological footprint in Afghanistan using deep learning. *International Journal of Environmental Science and Technology*: 1–24, **2025**.

Cihan, P. Comparative performance analysis of deep learning, classical, and hybrid time series models in ecological footprint forecasting. *Applied Sciences* 14(4): 1479, **2024**.

Janković, R., Mihajlović, I., Štrbac, N., and Amelio, A. Machine learning models for ecological footprint prediction based on energy parameters. *Neural Computing and Applications* 33(12): 7073–7087, **2021**.

Kumaran, V. V., Ridzuan, A. R., Senadjki, A., Kanaan, A. M. J., and Esquivias, M. A. The impacts of income inequality, forest area, and technology innovations on ecological footprint in Indonesia: ARDL and ML approach. *Discover Sustainability* 5(1): 373, **2024**

Moros-Ochoa, M. A., Castro-Nieto, G. Y., Quintero-Español, A., and Llorente-Portillo, C. Forecasting biocapacity and ecological footprint at a worldwide level to 2030 using neural networks. *Sustainability* 14(17): 10691, **2022**.

Mostafa, M. M. Clustering the ecological footprint of nations using Kohonen's self-organizing maps. *Expert Systems with Applications* 37(4): 2747–2755, **2010**.

Pata, U. K., and Isik, C. Determinants of the load capacity factor in China: A novel dynamic ARDL approach for ecological footprint accounting. *Resources Policy* 74: 102313, **2021**.

Roumiani, A., and Mofidi, A. Predicting ecological footprint based on global macro indicators in G-20 countries using machine learning approaches. *Environmental Science and Pollution Research* 29(8): 11736–11755, **2022**.

van der Woude, D., Castro Nieto, G. Y., Moros Ochoa, M. A., Llorente Portillo, C., and Quintero, A. Artificial intelligence in biocapacity and ecological footprint prediction in Latin America and the Caribbean. *Environment, Development and Sustainability*: 1–22, **2024**.

Zhang, C., Wang, Q., and Li, R. Revisiting the relationship between geopolitical risk and ecological footprint: A comprehensive analysis based on dual machine learning. *Journal of Environmental Management* 374: 124125, **2025**.

Biographies

Anik Debnath is in his final year of undergraduate studies in Industrial and Production Engineering from the Bangladesh University of Engineering and Technology (BUET). His specializations include artificial intelligence, data science, and statistical forecasting techniques, as well as his field of research interest in operation research, composite materials, and environmental sustainability. He has been involved in research that combine machine learning techniques and optimization frameworks to address challenging industrial and environmental problems. His activities reflect a strong will to create data-informed, sustainable, and innovative solutions that combine analytical accuracy and real-world relevance.

Akash Sikdar Sudip is currently pursuing his BSc in Industrial and Production Engineering at the Bangladesh University of Engineering and Technology (BUET). His field of expertise includes operations research, optimization methods, production planning, supply chain management, and quality control. Besides these, he is adept at carrying out data analysis and econometric model designs, such as Autoregressive Distributed Lag (ARDL) methods. He is passionate about this field and wants to work on application-oriented projects with an impactful contribution.

Souvik Mandol is a final year Computer Science and Engineering student at the Bangladesh University of Engineering and Technology. He is a machine learning and deep learning enthusiast with competence in web development and problem solving. He pursues his research interests in AI and data science related projects.