AI-DRIVEN INNOVATION IN EDUCATION AND RESEARCH: A ROADMAP FOR GLOBAL SOUTH DEVELOPMENT

Basil B. Duwa[1,2*], Natacha Usanase[1,2], Fatimah Adamu Wakil[1,2], Yoshebel F. Zirra[3], Celestine Iwendi[4], Dilber Uzun Ozsahin[1]

[1]Operational Research Center in Healthcare, Near East University, TRNC Mersin 10, Turkey,

[2]Department of Biomedical Engineering, Near East University, TRNC Mersin 10, Turkey,

[3]Health and Life Science Coventry University, United Kingdom

[4]Centre of Intelligence of Things, University of Bolton, United Kingdom

Abstract

With more than 1.2 billion young people, the Global South faces an urgent need to move beyond traditional education systems. Furthermore, this urgent transformation mainly focuses on reimagining education as a tool for comprehensive human development, which is founded in the region's unique cultural values, moral traditions and communal cohesion, rather than the means of dispensing information. Therefore, this study explores the role of Artificial Intelligence (AI) in addressing structural issues in education in low- and middle-income countries. We developed an integrated educational challenge score focusing on the out of school rates, completion rates and proficiency levels using a multi-indicator dataset from 124 countries. AI was adopted to rank the countries. Our results for the developed composite score showed Niger, South Sudan, and Djibouti as the most affected third world nations having the urgent need for educational upgrade. In the regression result, the Random Forest Regressor most accurately predicted unemployment based on education variables ($R^2 = 0.17$), indicating a modest but relevant relationship between disparities in education and labor outcomes. The results guide a strategic roadmap for the development of AI in education in the global south.

Keywords: *AI, Education, Global South, Predictive models, world education, Challenge score*

1. INTRODUCTION

Education is a critical component of both individual and social success. At the individual level, it allows people to live fuller, meaningfully; at the social level, it promotes innovation, economic progress, and the ability to address difficult global issues. In contrast, the United Nations educational, scientific and cultural organizations (UNESCO) has reported over 250 million out-of-school children in 2023, with over 120 million of the data accounting from girls and women[1].

Though, over the last few centuries, the world has witnessed a tremendous educational revolution, from an era when basic education was a privilege for the few to one in which school enrolment and attendance are the global standard, but the global south keeps lacking in this important space. Additionally, having access to any learning portal isn't enough. The standard and quality of the learning remains a major challenge. In many low-income nations, an overwhelming number of children continue to leave primary school with no elementary literacy or numeracy skills[2]. The sub-Saharan Africa accounts for over 30% of all the global record of the out-of-school children. These ongoing learning gaps, are frequently linked to institutional imbalances, perpetuate poverty cycles and exacerbate global income inequalities. Addressing the quality disparity is critical to ensuring equitable growth and genuine educational transformation. This therefore, has shown a need to focus on implementation of modern educational techniques around the global developing regions.[3]

Furthermore, other parts of the world continue to thrive in educational research and development. Figure 1 presents the proportion of R&D (Research and Development) researchers per million inhabitants in 2022, emphasizing the global distinctions in research. Countries in Europe, North America, and East Asia, such as South Korea, Japan, Germany, and the United States, have the largest R&D researcher concentrations, typically exceeding 5,000 or 10,000 per million inhabitants. Many African, South Asian, and Latin American countries, on the other hand, have fewer than 500 R&D researchers per million people, with others without any statistics at all. This imbalance highlights major global disparities in research structures creative ability, and accessibility to scientific advances[4].

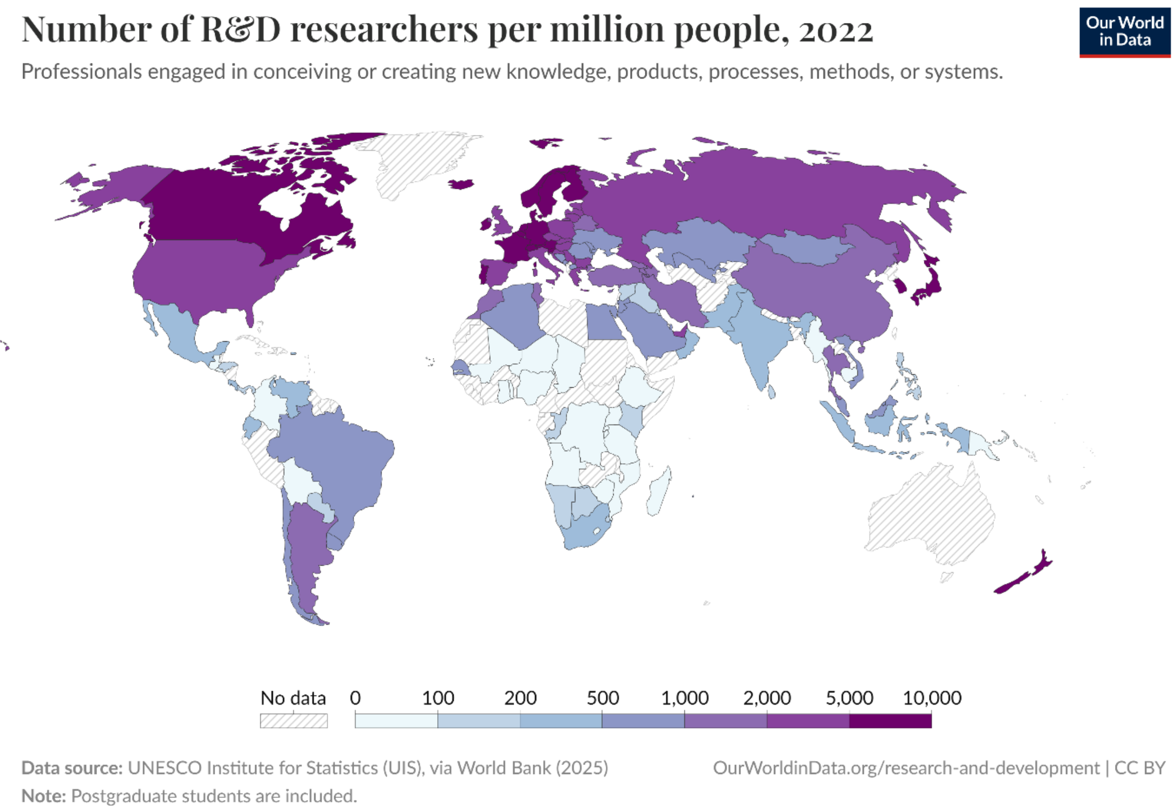

Figure 1: Record of Research and Development

The global growing population has established an increasing necessity for the adoption of AI in the human daily activities. Hence, this has broadened the application of AI in various sectors, such as in economy, education and research, agriculture, health, transportation and many more. Furthermore, AI's adoption and application in education has spread across its digital use, through the enhancement of easy class delivery for instructors and in the smooth assimilation for the students across all educational backgrounds. Thus, AI has enormous potential for solving long-standing educational problems, particularly those related to accessibility, performance, and fairness[5]. In the Global South, where educational institutions face limited resources and systemic inadequacies, AI can provide scalable, adaptable solutions that are suited to the different needs of learners. However, its impact is dependent on strategic implementation that is consistent with local circumstances and inclusive development goals.

Furthermore, in the quest for a better implementation and application of AI in education, the UNESCO has revealed that over $1.7 trillion has been budgeted for the Global research and development across the globe. Therefore, various research studies have reported unique instances of the application of AI in education and research[4]. Mwansa et al. (2025) revealed that digital exclusion is prevalent in rural South Africa, with low prices, infrastructure deficiencies, and digital illiteracy serving as major impediments to connection and engagement. While their research provides useful insights on localized difficulties, it is spatially limited and descriptive[3]. Comparatively, our study offers a wider, data-driven approach that not only rates developing nations by educational need, but also applies ML models to investigate the structural connection between education and unemployment providing extensible discoveries for AI-based strategy and reform. Similarly, Vučković and Sikimić's (2025) article raised ethical and epistemological issues about AI in education, highlighting the potential for global injustice if these systems exacerbate current gaps between the Global North and South. While their work establishes an important ethical structure, it is primarily theoretical[6]. In contrast, our research expands this debate using an applied, data-driven method, assessing educational demand across Global South nations and illustrating how AI may be carefully tailored to overcome structural injustices and lower unemployment using quantitative initiatives[2]. Figure 2 presents a bibliometric report of some reported studies on global education using AI, extracted from 2025 studies on Scopus. The visualization is done by the VOSviewer tool.

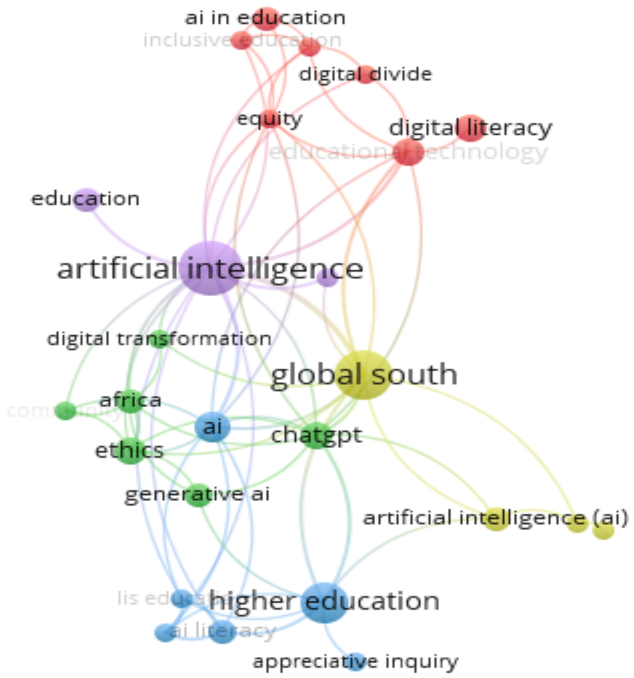

Figure 2: Bibliometric report of AI in Global Education [7]

Despite the growing enthusiasm in AI adoption in education, there is a dearth of data-driven approaches for determining where the need for interventions[8] is most required and how they connect to larger improvements in development in the developing countries. However, current research studies frequently focus on the developed nations, ignoring the complex systemic hurdles in the low-resource nations[5]. This study fills that gap by integrating educational performance indicators with machine learning (ML) techniques to rank nations by demand and assess the predictive relationship between education and unemployment, providing concrete recommendations for selected AI-driven intervention.

2. MATERIALS AND METHODS

2.1 Data collection

The dataset was collected from a public database, Kaggle as a csv file. The data contains 203 instances from 124 countries with 29 features.

2.2 The Experimental workflow

The study adopts an integrated educational challenge score system and a regression ML method in the analysis of the dataset. The study's design started with a global education dataset that included 202 cases from 124 nations as shown in Figure 3. Following initial data import, preparation steps such as data wrangling, cleaning, and normalization were performed to set up the dataset for proper research, such as the study by [9].

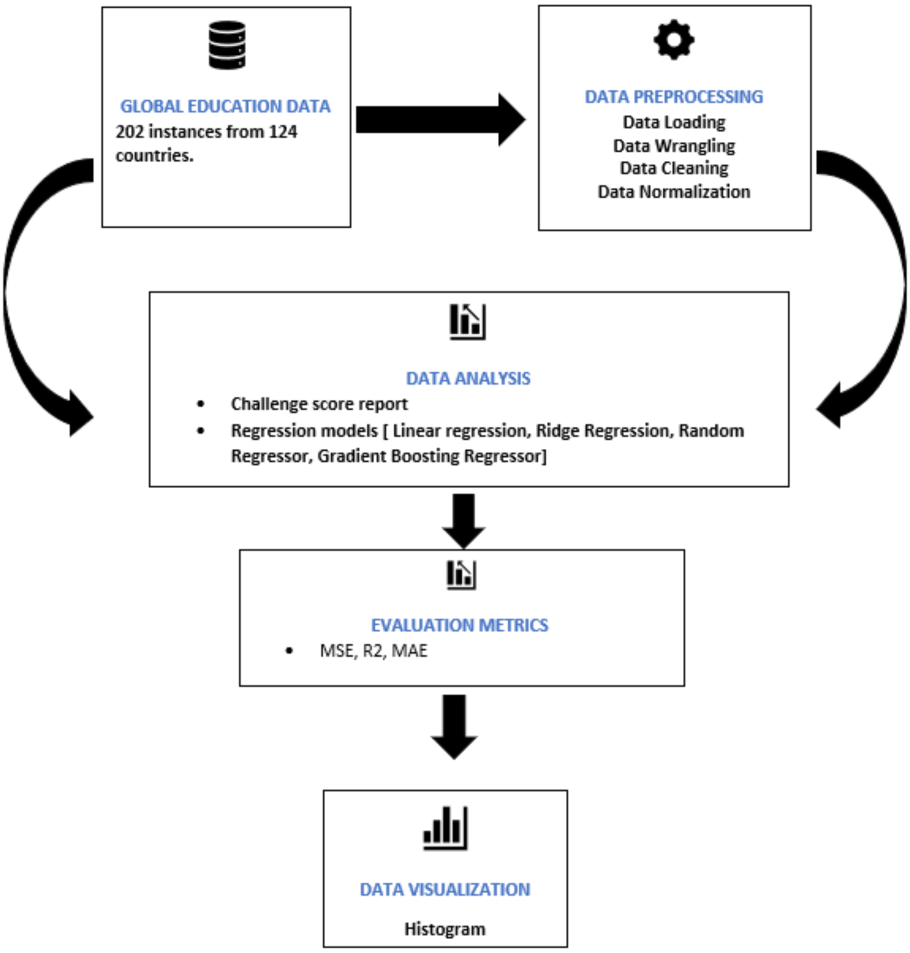

Figure 3: Experimental workflow

The primary analysis entailed calculating an integrated educational challenge score and training various models of regression (Linear Regression, Ridge Regression, Random Forest Regressor, and Gradient Boosting Regressor) for predicting rates of unemployment using educational markers. Model performance was evaluated using conventional measures including MSE, R^2, and MAE [10]. The final results were visualized using histograms to show patterns and aid understanding.

3. RESULTS

3.1 Analysis of the challenge scores across countries.

Table 1 ranks countries based on their generated challenge score, that assesses the number of educational issues encountered by each nation using the Global_Education.csv dataset. The results show that Niger has the highest overall challenge score (1.742), then comes South Sudan (1.705), Djibouti (1.680), Chad (1.660), and Mali (1.534). Similarly, other countries with notable high ratings are Eritrea (1.469), Ethiopia (1.466), the Marshall Islands (1.464), and Tuvalu (1.424). These countries primarily represent sub-Saharan African regions and small island states, implying that deep-rooted educational inequality may be driven by poor infrastructure, socioeconomic instability, and limited access to learning technology. These findings highlight the vital importance of tailored AI-based educational programs in these places.

Table 1: Table of challenge scores and the countries

	Countries and areas	challenge_score
129	Niger	1.742189
167	South Sudan	1.704727
51	Djibouti	1.680873
35	Chad	1.660059
109	Mali	1.533772
58	Eritrea	1.469062
61	Ethiopia	1.466483
111	Marshall Islands	1.464280
187	Tuvalu	1.423671
121	Mozambique	1.394957
28	Burkina Faso	1.394110
73	Guinea	1.390736
57	Equatorial Guinea	1.297735
29	Burundi	1.288047
138	Papua New Guinea	1.283801
112	Mauritania	1.274896
106	Malawi	1.220219
30	Cape Verde	1.206342
100	Liberia	1.197516
160	Sierra Leone	1.182561

Figure 4 presents the distribution of challenge scores across the analyzed nations using a histogram. A large number of countries had scores between 1.2 and 1.5, reflecting modest educational challenges throughout the sample. Additionally, in this spectrum, several clusters emerge, with a few nations achieving exceptionally excellent scores above 1.6, including Niger and South Sudan, as identified in Table 1. This distribution implies that, while educational problems are common, there is a group of nations with much worse educational conditions. This classification can help policymakers and development groups differentiate intervention targets based on intensity levels.

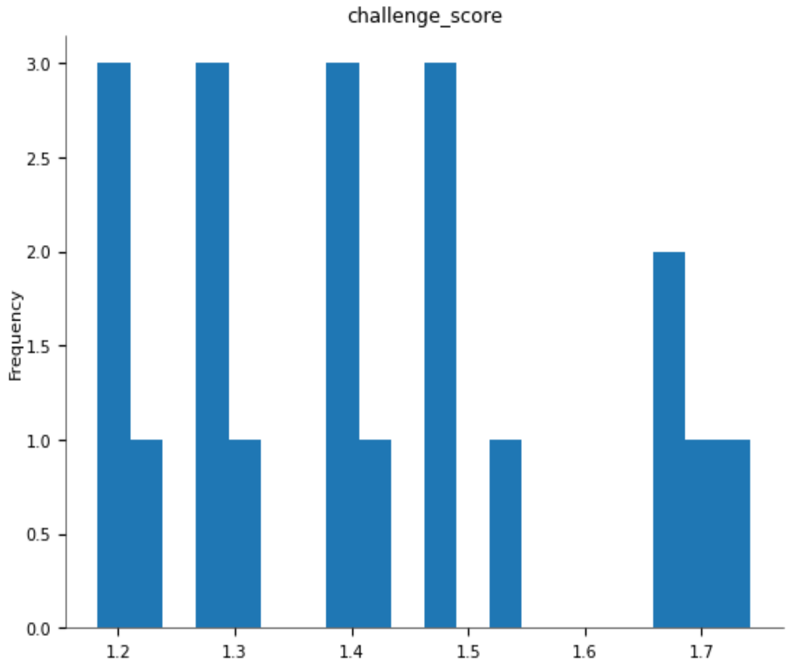

Figure 4: Distribution of challenge score

3.2 Predictive modelling

The collected dataset, with the educational and socioeconomic variables, was successfully loaded and cleaned by fixing a UnicodeDecodeError with the 'latin1' encoding. There were no missing or duplicate numbers discovered. Feature exploration identified connections in the unemployment rate and a number of variables, including educational achievement, enrolment, and literacy. Various studies have adopted similar Machine learning techniques to predict different problems[11]. To enrich the dataset, new characteristics were created, such as Total_OOSR, Total_Completion_Rate, Total_Proficiency_Score, and Youth_Literacy_Gender_Gap as shown in Table 2.

Table 2: Predictive Models

Model	MSE	R^2	MAE
Random forest	29.82	0.172	3.92
Ridge Regression	32.75	0.090	4.25
Gradient Boosting	34.05	0.054	4.19
Linear regression	36.22	-0.006	4.46
Lasso Regression	33.99	0.056	4.39

Following the standardization of numerical values and an 80/20 train-test split, 5 models of regression were trained and assessed. The Random Forest Regressor fared better than all others, as indicated by the model performance table, with the lowest Mean Squared Error (MSE = 29.82), the lowest Mean Absolute Error (MAE = 3.92), and the most significant R-squared value (R2 = 0.172). On the other hand, Lasso Regression and Linear Regression performed worse, with low and negative R2 values and increased mistakes. The Random Forest model's greater capacity to identify intricate, non-linear trends in the data is demonstrated by these outcomes.

4. DISCUSSION

The study's results provide important new information about the correlation between unemployment rates and educational indicators in various nations. Features like literacy levels, proficiency scores, enrolment rates, and completion rates show varied degrees of connection with unemployment, according to preliminary data analysis. The accessibility and quality of education appear to influence labor market outcomes, as evidenced by positive relationships with specific finishing and enrolment rates and negative correlations with out-of-school rates and proficiency scores. Curiously, geographic variables like longitude also showed a negative relationship with unemployment rates, which may reflect regional differences in educational and economic development.

Furthermore, countries were evaluated using a challenge_score, which combined indices of availability, completion, and competence, to give a thorough picture of the world's educational requirements. According to the results, the nations with the worst educational problems were Niger (1.742), South Sudan (1.705), Djibouti (1.680), Chad (1.660), and Mali (1.534). These countries, which are mostly in Sub-Saharan Africa, have structural obstacles like poor educational facilities, gender inequality, and interruptions brought on by conflicts. These results align with earlier research highlighting Africa's susceptibility to global education objectives. Prioritizing interventions, especially AI-based ones that can improve access and learning results in settings with limited resources, depends on identifying these regions.

In this study, feature engineering added composite metrics that reflect multifaceted educational aspects to the dataset, including Youth Literacy Gender Gap, Total Completion Rate, and Total Proficiency Score. By combining similar indicators, these features increased model resilience and may have decreased multicollinearity and noise. Five machine learning models, Linear Regression, Ridge, Lasso, Gradient Boosting, and Random Forest were assessed following the implementation of scaling and an 80/20 data split. With the lowest MSE (29.82), lowest MAE (3.92), and highest R2 (0.172), the Random Forest Regressor outperformed the others in terms of prediction. Its strength is its ability to capture intricate non-linear correlations between predictors, trends that linear approaches like Linear Regression (R2 = -0.006) are unable to adequately represent.

Conclusively, interpretability is still crucial for converting findings into workable policy, even when predictive modelling found patterns in the data. By determining the relative significance of characteristics like literacy gaps or completion rates in forecasting unemployment, explainable AI (XAI) technologies like SHAP may be able to direct focused solutions. Finding the areas that most require interventions and comprehending which educational aspects have the greatest impact on labor market outcomes are two perspectives that can be obtained by integrating insights from challenge_score rankings with model-driven analysis.

5. CONCLUSION

To identify which regions may most benefit from AI driven educational solutions, this study examined the relationship between unemployment rates and educational metrics across countries. The analysis, which used the Global_Education.csv dataset, showed that, though complicated and frequently non-linear, there are quantifiable links between unemployment and literacy levels, proficiency scores, educational access, and completion rates. The prediction architecture was enhanced through feature engineering, which added composite measures such as Youth Literacy Gender Gap, Total Completion Rate, and Total Proficiency Score. The study by Mwansa et al. is descriptive and spatially constrained, but it offers helpful insights on localized challenges. In contrast, our study provides a more comprehensive, data-driven method that ranks developing countries according to their educational needs and uses machine learning (ML) models to examine the structural relationship between education and unemployment, yielding findings that may be expanded upon for AI-based policy and change.

In addition to predictive modelling, the study calculated a challenge_score to rank nations according to their weaknesses in education. The most serious educational issues, namely in the areas of access, proficiency, and gender inequality, were found in Niger, South Sudan, Djibouti, Chad, and Mali. AIpowered educational solutions, such adaptive learning platforms as well remote education tools, and automated tutoring systems may have a revolutionary effect in these high-need areas, according to these findings, which also offer an evidence-based basis for prioritizing initiatives.

The data is publicly accessible at https://doi.org/10.34740/kaggle/dsv/6879147

REFERENCES

[1] "250 million children out-of-school: What you need to know about." Accessed: Jul. 10, 2025. [Online]. Available: https://www.unesco.org/en/articles/250-million-children-out-school-what-you-need-know-about-unescos-latest-education-data

[2] O. S. Falebita and P. J. Kok, "Strategic goals for artificial intelligence integration among STEM academics and undergraduates in African higher education: a systematic review," *Discover Education*, vol. 3, no. 1, pp. 1–22, Dec. 2024, doi: 10.1007/S44217-024-00252-1/METRICS.

[3] G. Mwansa, M. R. Ngandu, and Z. Mkwambi, "Bridging the digital divide: exploring the challenges and solutions for digital exclusion in rural South Africa," *Discover Global Society*, vol. 3, no. 1, pp. 1–24, Dec. 2025, doi: 10.1007/S44282-025-00189-2/TABLES/12.

[4] "How much does your country invest in R&D?" Accessed: Jul. 10, 2025. [Online]. Available: https://uis.unesco.org/apps/visualisations/research-and-development-spending/

[5] H. Ritchie, V. Samborska, N. Ahuja, E. Ortiz-Ospina, and M. Roser, "Global Education," *Our World in Data*, Oct. 2023, Accessed: Jul. 10, 2025. [Online]. Available: https://ourworldindata.org/global-education

[6] A. Vučković and V. Sikimić, "Global justice and the use of AI in education: ethical and epistemic aspects," *AI Soc*, vol. 40, no. 4, pp. 3087–3104, Apr. 2025, doi: 10.1007/S00146-024-02076-X.

[7] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010, doi: 10.1007/S11192-009-0146-3.

[8] B. B. Duwa, N. Usanase, and B. Uzun, "Enhanced Drug Classification for Cancers of the Liver with Multi-Criteria Decision-Making Method-PROMETHEE," *Article in Global Journal of Sciences*, 2025, doi: 10.48165/gjs.2025.2102.

[9] B. B. Duwa *et al.*, "Ensemble Predictive Modeling for Dementia Diagnosis," *Proceedings - International Conference on Developments in eSystems Engineering, DeSE*, pp. 352–357, 2024, doi: 10.1109/DESE63988.2024.10912022.

[10] A. G. Usman *et al.*, "Second-order based ensemble machine learning technique for modelling river water biological oxygen demand (BOD): Insights into improved learning," *J Radiat Res Appl Sci*, vol. 18, no. 2, p. 101439, Jun. 2025, doi: 10.1016/J.JRRAS.2025.101439.

[11] D. Uzun Ozsahin, B. B. Duwa, I. Ozsahin, and B. Uzun, "Quantitative Forecasting of Malaria Parasite Using Machine Learning Models: MLR, ANN, ANFIS and Random Forest," *Diagnostics 2024, Vol. 14, Page 385*, vol. 14, no. 4, p. 385, Feb. 2024, doi: 10.3390/DIAGNOSTICS14040385.