



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Explainable Machine Learning for Benchmarking National Digital Research Ecosystems: Implications for Higher-Education Strategy

Rissa A. Lasap^{1,2*}, Luigi Carlo M. De Jesus^{2,3,4}

¹ Biological Science Department, Centro Escolar University, Manila, Metro Manila, Philippines. E-mail: rissalasap22@gmail.com, ORCID: 0009-0002-1029-3460

² Naveen Jindal Youth Fellowship Program, O.P. Jindal Global University, Sonapat, Haryana, India.

³ Department of Computer, Electronics, and Electrical Engineering, De La Salle University, Manila, Metro Manila, Philippines.

⁴ School of Engineering, Asia Pacific College, Makati City, Metro Manila, Philippines. E-mail: luigi_dejesus@dlsu.edu.ph, ORCID: 0000-0001-7889-6044

*Corresponding author: Rissa A. Lasap (rissalasap22@gmail.com)

Abstract

National research systems increasingly depend on the joint availability of scientific human capital, sustained research and development (R&D) investment, and trustworthy digital infrastructure. Yet international benchmarking often relies on single indicators or opaque composite rankings that obscure the system conditions associated with research output. This study develops an explainable machine-learning benchmark of national digital research ecosystems and translates the findings into higher-education strategy. Twelve open World Development Indicators were combined in a country-year data architecture. The outcome was 2023 scientific and technical journal articles per million population; predictors were the latest available values from 2019–2023 for R&D intensity, researcher density, tertiary enrollment, economic capacity, high-technology exports, secure servers, public education expenditure, electricity access, mobile subscriptions, and fixed broadband. The merged public dataset contains 197 countries. A primary analytic cohort of 99 countries was defined by observed R&D intensity and researcher density, while limited missingness in the remaining predictors was imputed only within training folds. Four prespecified models were evaluated using five-times repeated five-fold cross-validation. Random forest achieved the lowest held-out error (mean $R^2=0.820$; RMSE=0.613 on the log1p outcome; Spearman $\rho=0.912$), closely followed by extremely randomized trees. Cross-validated permutation analysis identified researcher density as the dominant predictor, followed by GDP per capita, secure Internet servers, R&D intensity, and electricity access. Complete-case sensitivity analysis ($n=79$) produced similar results ($R^2=0.810$; $\rho=0.914$). The results support a capacity-sequencing interpretation: universities cannot treat AI adoption, digital transformation, research workforce development, and basic infrastructure as independent initiatives. The benchmark is descriptive rather than causal and is released with country-level source years, inclusion flags, model diagnostics, a data dictionary, and reproducible analysis files.

Keywords: explainable artificial intelligence; research analytics; higher education; digital infrastructure; national innovation systems; secondary data; random forest

1. Introduction

Artificial intelligence is moving from isolated educational applications toward institution-wide infrastructure for teaching, research, governance, and decision support. Recent work in this journal has framed the AI-native university as an integrated institutional capability rather than a collection of disconnected tools [1], while parallel research has emphasized governance, digital security, data protection, and epistemic responsibility in AI-enabled higher education [2,3]. These arguments share an important premise: sophisticated applications are sustainable only when the surrounding research, information, and infrastructure systems are sufficiently mature.

The difficulty is measurement. Institutional leaders and public agencies are presented with global indices, university rankings, telecommunications indicators, R&D statistics, and publication counts, but the indicators are rarely connected in a transparent analytical model. Single-variable comparisons can be misleading because scientific output is produced by interacting conditions: trained researchers, funding continuity, access to tertiary education, economic resources, secure computing infrastructure, reliable electricity, and connectivity. A benchmark designed for higher-education strategy should therefore be multivariate, empirically testable, and explainable to non-specialist decision makers.

The higher-education literature has documented the rapid expansion of AI applications, but systematic reviews continue to identify limited educator participation, uneven governance, and insufficient attention to human control and trustworthiness [4,5]. Digital transformation research similarly treats technology-enabled change as an organizational and strategic process rather than a simple acquisition decision [6]. At the national level, innovation-capacity theory proposes that research inputs, human capital, institutions, and knowledge infrastructure jointly shape innovation output [7,8]. These streams have rarely been brought together through a reproducible cross-country information-system benchmark that is both predictive and interpretable.

This study addresses that gap by combining twelve openly available secondary-data series into an auditable country-level dataset and comparing four machine-learning models under repeated cross-validation. The objective is not to rank nations normatively or to infer causality. Instead, the study asks three operational questions: (1) how accurately can publication intensity be estimated from national research, education, economic, and digital-system indicators; (2) which indicators contribute most to held-out predictive performance; and (3) how can the resulting pattern inform capacity sequencing for higher-education research and AI-enabled transformation?

The contribution is fourfold. First, the study supplies a downloadable CSV in which each predictor retains its original source year, avoiding the false appearance of perfect temporal synchronization. Second, it uses an explicit inclusion rule that requires observed values for R&D intensity and researcher density, thereby avoiding large-scale imputation of the two constructs most central to the research-capacity argument. Third, it reports repeated out-of-sample performance and held-out permutation importance rather than in-sample fit. Fourth, it translates the model into a non-causal decision framework for universities and public research systems. The analytical logic is summarized in Figure 1.

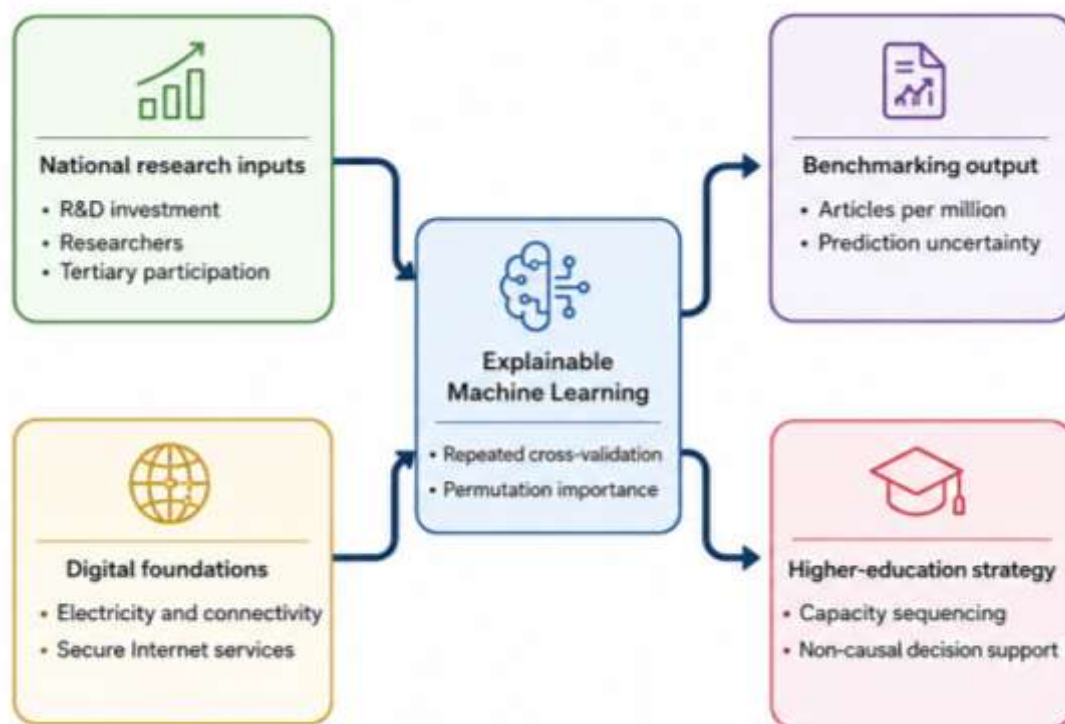


Figure 1. Analytical framing of the national digital research-ecosystem benchmark.

1.1 National research capacity and publication output

Cross-national research performance has long been associated with economic capacity, research expenditure, and the size of the scientific workforce. Early comparative studies demonstrated large differences in national scientific productivity and impact [9,10]. Research funding and English proficiency were associated with publication rates in high-ranked medical journals [11], while GDP, R&D expenditure, university counts, and journal infrastructure were linked to publication output in environmental science [12]. These studies established the relevance of macro-level inputs, but most relied on linear associations, domain-specific samples, or a limited set of predictors.

National innovative-capacity theory provides a broader explanation. Furman, Porter, and Stern conceptualized innovative capacity as the ability of a country to produce and commercialize new-to-the-world knowledge through common innovation infrastructure, specialized clusters, and linkages among institutions [7]. Endogenous growth theory similarly places knowledge creation and human capital at the center of long-run development [8]. For research systems, this implies that publication output should not be interpreted as the result of funding alone: it reflects the interaction of scientific labor, institutions, enabling infrastructure, and accumulated economic capacity.

1.2 Digital infrastructure as research infrastructure

Contemporary research work is digitally mediated. Literature discovery, data collection, high-performance computing, cloud collaboration, secure storage, reproducible software, preprint dissemination, and international co-authorship all depend on reliable digital systems. Secure Internet servers capture only one component of this environment, yet they provide a comparable proxy for the density of authenticated digital services. Fixed and

mobile connectivity indicate access pathways, while electricity access represents a foundational constraint that cannot be substituted by software policy.

For higher-education leaders, the implication is that digital transformation and research development should not be managed as separate portfolios. A university may acquire generative-AI subscriptions or learning-analytics platforms, but the research benefits will remain limited when laboratories lack stable connectivity, computing governance, data stewardship, or trained researchers. This study therefore treats digital indicators as complements to—not replacements for—human and financial research capacity.

1.3 Explainable machine learning for policy-facing analytics

Machine-learning ensembles can represent nonlinearities and interactions that conventional linear models may miss. Random forests average decorrelated decision trees and are robust to nonlinear response surfaces [16]; gradient boosting constructs an additive predictive function through sequential error correction [17]. Nevertheless, predictive accuracy alone is insufficient when results are used for strategy. Explainability methods such as SHAP and LIME have been developed to expose model behavior [19,20], and explainable machine learning has been proposed as a bridge between prediction and scientific insight [21].

The present study uses model-agnostic permutation importance computed on held-out data. A feature is permuted within a validation fold, and the increase in prediction error indicates how much the fitted model depended on that feature for unseen cases. This approach does not establish causal importance and can understate the role of correlated predictors, but it is directly tied to out-of-sample performance and is understandable to policy users.

2. Materials And Methods

2.1 Study design and data provenance

A cross-sectional secondary-data design was implemented using the World Development Indicators (WDI), an open international data system maintained by the World Bank [14]. The original series draw on the UNESCO Institute for Statistics, International Telecommunication Union, World Bank national accounts, United Nations trade data, and other documented providers. R&D concepts follow internationally standardized definitions associated with the Frascati Manual [13]. The data were accessed on 28 July 2026 and retained in long country-year form before merging [15].

The outcome was scientific and technical journal articles in 2023 divided by the 2023 population and multiplied by one million. The per-capita transformation reduces the mechanical advantage of population size, although it does not adjust for field composition or citation impact. Predictors were selected before modelling to represent five system domains: research investment, scientific human capital, tertiary participation, economic/industrial context, and digital foundations. For each predictor, the latest observed value from 2019 through 2023 was used. The source year was preserved in a paired column in the released CSV. Table 1 lists the variables, codes, and units, while Figure 2 summarizes the reproducible data and modelling workflow.

Table 1. Variables in the secondary-data architecture.

Indicator	WDI code and unit
Articles per million	IP.JRN.ARTC.SC; derived outcome
R&D expenditure	GB.XPD.RSDV.GD.ZS; % of GDP
Researchers	SP.POP.SCIE.RD.P6; per million people
Tertiary enrollment	SE.TER.ENRR; gross %
GDP per capita	NY.GDP.PCAP.CD; current US\$
High-technology exports	TX.VAL.TECH.MF.ZS; % of manufactured exports
Secure Internet servers	IT.NET.SECR.P6; per million people
Education expenditure	SE.XPD.TOTL.GB.ZS; % of government expenditure
Electricity access	EG.ELC.ACCS.ZS; % of population
Mobile subscriptions	IT.CEL.SETS.P2; per 100 people
Fixed broadband	IT.NET.BBND.P2; per 100 people

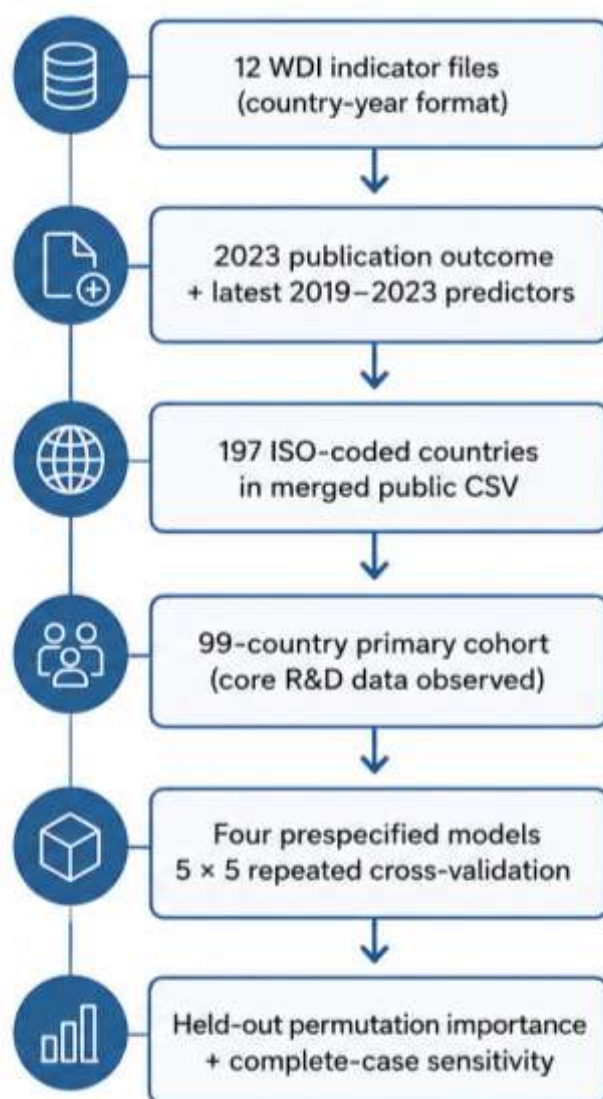


Figure 2. Data integration, cohort definition, validation, and explanation workflow.

2.2 Country cohort and missing-data policy

The merged dataset contains 197 ISO-coded countries with a 2023 article count and population denominator. The primary analytic cohort required non-missing R&D expenditure and researcher density during 2019–2023, yielding 99 countries. This restriction was adopted because those two variables are the central resource inputs in the conceptual model; imputing them for nearly half the global sample would create an analytically impressive but substantively fragile benchmark.

Within the 99-country cohort, tertiary enrollment and public education expenditure each had 10.1% missingness, and high-technology exports had 5.1% missingness. Other predictors were complete. Missing values were replaced by the training-fold median, and missingness indicators were added within each modelling pipeline. Consequently, validation-fold information did not influence imputation. A complete-case sensitivity analysis included the 79 countries observed on all ten predictors. Multiple imputation is useful in many inferential settings [23], but fold-specific median imputation was selected here because the aim was out-of-sample prediction, missingness was limited to three non-core variables, and the procedure could be replicated identically across all models.

2.3 Modelling and validation

The outcome was transformed as $\log(1 + \text{articles per million})$ to reduce right skew and stabilize error variance. Four prespecified regression algorithms were compared: elastic net, random forest, extremely randomized trees, and histogram gradient boosting. Elastic net used a Yeo-Johnson transformation after imputation, with $\alpha=0.03$ and l1 ratio=0.20. The random forest used 250 trees, 70% feature sampling, and a minimum leaf size of two. Extra trees used 250 trees, 80% feature sampling, and a minimum leaf size of two. Histogram gradient boosting used learning rate 0.06, 250 iterations, 15 maximum leaves, minimum leaf size 10, and L2 regularization of 1.0. Hyperparameters were fixed in advance to prevent optimization against a small cross-country sample. Generalization was evaluated with five-fold cross-validation repeated five times using a fixed random seed, producing 25 held-out estimates per model. The principal metrics were coefficient of determination (R^2), root

mean squared error (RMSE), and mean absolute error (MAE) on the transformed outcome, plus Spearman rank correlation on the original publication-intensity scale. Model selection was based primarily on mean held-out RMSE. The use of repeated validation reduces dependence on one arbitrary partition, although folds remain statistically dependent. Therefore, the reported 95% bootstrap intervals across folds are descriptive uncertainty intervals rather than formal confidence intervals for a superpopulation. Cross-validation procedures were implemented in scikit-learn [18]; the design follows the general principle that model assessment must remain separate from model fitting [22].

2.4 Explainability, diagnostics, and robustness

For the best-performing algorithm, permutation importance was computed on each held-out fold of an additional five-fold split. Each predictor was permuted five times per fold, producing 25 error changes per feature. Importance is reported as the increase in held-out RMSE; values near or below zero indicate that the feature did not provide stable incremental information under the fitted model and correlated predictor set.

Out-of-fold predictions were averaged across the five repeats for each country. The difference between observed and predicted log publication intensity was retained as a diagnostic residual. Positive residuals indicate greater observed publication intensity than the model estimated from the measured inputs; negative residuals indicate the reverse. These values are not efficiency scores because omitted factors—including language, disciplinary mix, collaboration networks, research quality, conflict, migration, institutional autonomy, and database coverage—can influence them. Robustness was examined through the complete-case cohort and model-class comparison.

2.5 Ethics, transparency, and reproducibility

The study used public, aggregate national statistics and involved no human participants, personal data, intervention, or automated decision affecting an individual. Ethics-board review was therefore not required. The release package includes the merged CSV, primary analytic sample, data dictionary, descriptive statistics, cross-validation results, permutation importance, diagnostic residual examples, figures, and analysis code. This level of provenance is particularly important because the journal requires authentic data and identification of all sources.

3. Results

3.1 Sample characteristics

The 99-country analytic cohort displayed substantial heterogeneity. Median publication intensity was 408.4 articles per million people (interquartile range [IQR], 130.0–1263.3), with values ranging from 4.3 to 2623.7. Median R&D expenditure was 0.648% of GDP (IQR, 0.230–1.522), and median researcher density was 1312.8 per million people (IQR, 424.1–4535.8). Median GDP per capita was US\$13,830.84, while median secure-server density was 5256.3 per million. Electricity access had a ceiling distribution: the median and both quartiles were 100%, but the minimum was 15.6%.

As summarized in Table 2, the wide ranges justify nonlinear modelling. For example, secure-server density extended from 14.1 to 557,174.3 per million, and researcher density from 1.4 to 18,129.8 per million. Tertiary enrollment also exceeded 100% in some countries because the gross enrollment ratio includes students outside the official age group. These indicator properties were retained rather than winsorized, because tree ensembles are comparatively robust to extreme values and the released data should remain faithful to the source.

Table 2. Selected descriptive statistics for the primary analytic cohort (n=99).

Variable	Median	Interquartile range
Articles per million	408.43	130.04-1263.28
R&D expenditure, % GDP	0.648	0.230-1.522
Researchers per million	1312.75	424.07-4535.83
Tertiary enrollment, %	64.73	44.24-79.85
GDP per capita, US\$	13,830.84	5,229.66-35,444.55
Secure servers per million	5,256.31	427.03-35,502.77
Electricity access, %	100.00	100.00-100.00
Fixed broadband per 100	25.99	11.39-36.96

3.2 Predictive performance

Random forest achieved the lowest mean held-out RMSE (0.613; descriptive 95% interval, 0.557–0.674) and the highest mean R^2 (0.820; 0.788–0.849). Its mean MAE was 0.463 and Spearman correlation was 0.912. Extra trees was statistically and practically close, with RMSE 0.623, R^2 0.814, and the highest rank correlation (0.920). Histogram gradient boosting and elastic net produced lower but still substantial predictive performance. The narrow separation between the two tree ensembles indicates that the main conclusion is not dependent on one specific ensemble construction. Table 3 reports the model comparison, Figure 3 shows the held-out RMSE distributions, and Figure 4 compares observed and repeated out-of-fold predictions.

Table 3. Repeated cross-validation performance (25 held-out folds per model).

Model	R^2	RMSE	Spearman rho
Random forest	0.820	0.613	0.912
Extra trees	0.814	0.623	0.920

Histogram gradient boosting	0.783	0.649	0.895
Elastic net	0.785	0.660	0.903

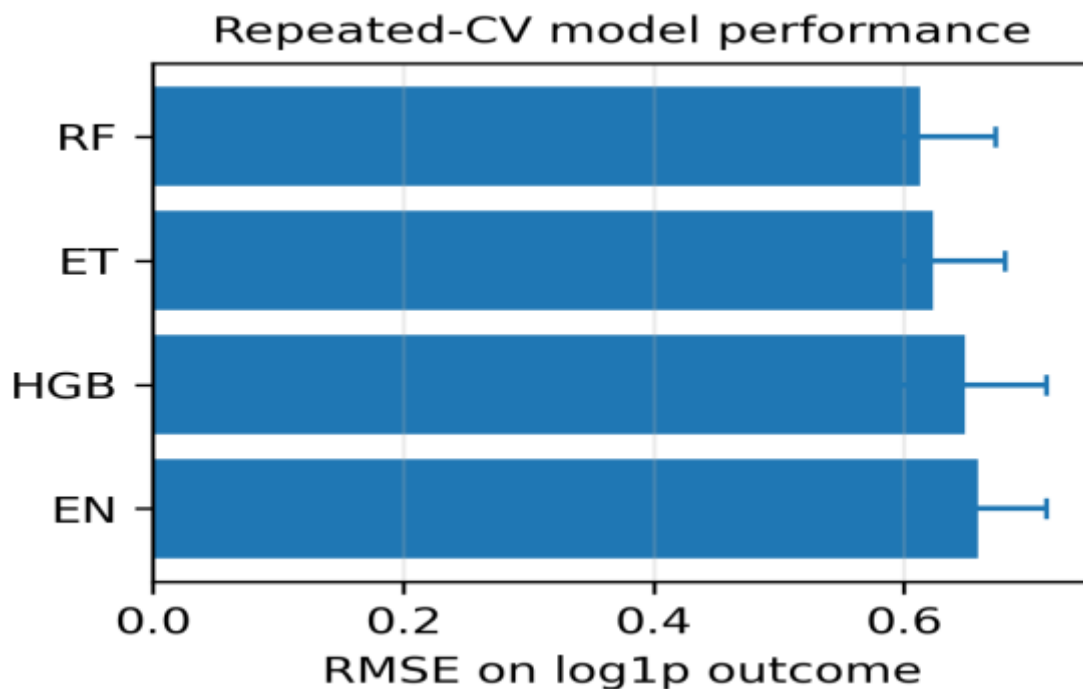


Figure 3. Mean held-out RMSE with descriptive bootstrap intervals across 25 folds.

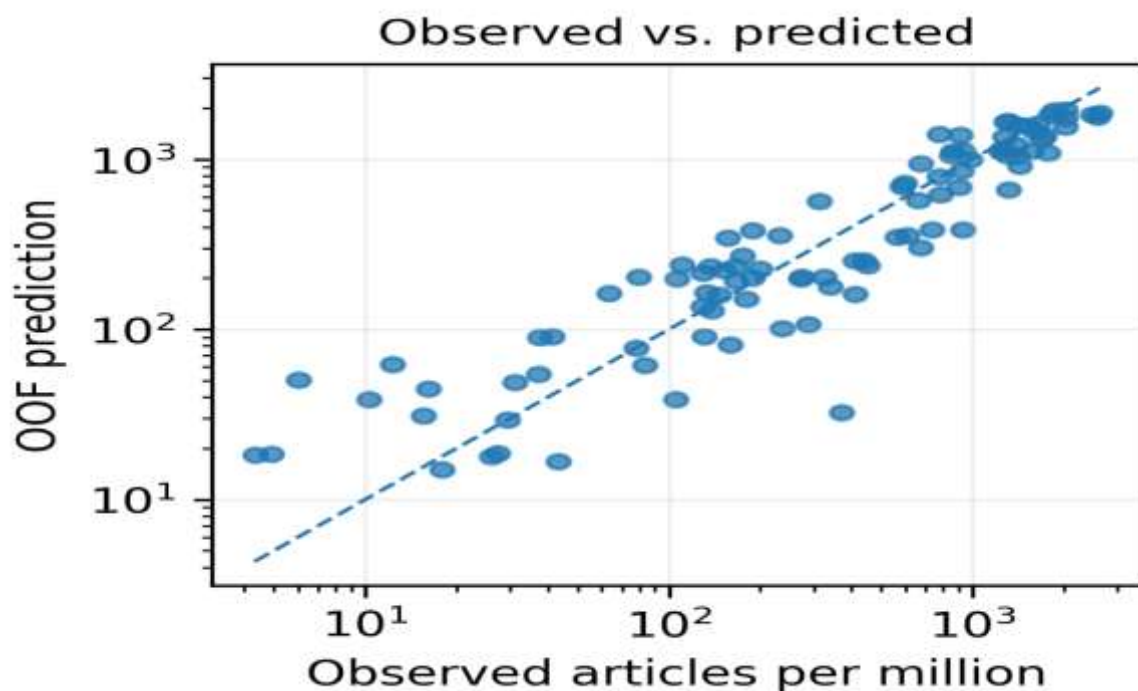


Figure 4. Repeated out-of-fold predictions versus observed publication intensity (log scales).

3.3 Held-out feature importance

Researcher density was the dominant feature: permuting it increased held-out RMSE by 0.413 on average. GDP per capita ranked second (0.154), followed by secure Internet servers (0.071), R&D expenditure as a share of GDP (0.063), and electricity access (0.062). The remaining indicators contributed little stable incremental information after the model had access to these variables. Table 4 reports the numerical permutation results, and Figure 5 visualizes their relative magnitudes.

Low incremental importance should not be interpreted as proof that tertiary enrollment, education expenditure, high-technology exports, broadband, or mobile access are irrelevant. Several are conceptually upstream,

correlated with stronger predictors, measured with error, or subject to ceiling effects. Permutation importance answers a narrow predictive question: how much did the fitted model's error increase when one feature's observed pattern was disrupted while the remaining features stayed intact?

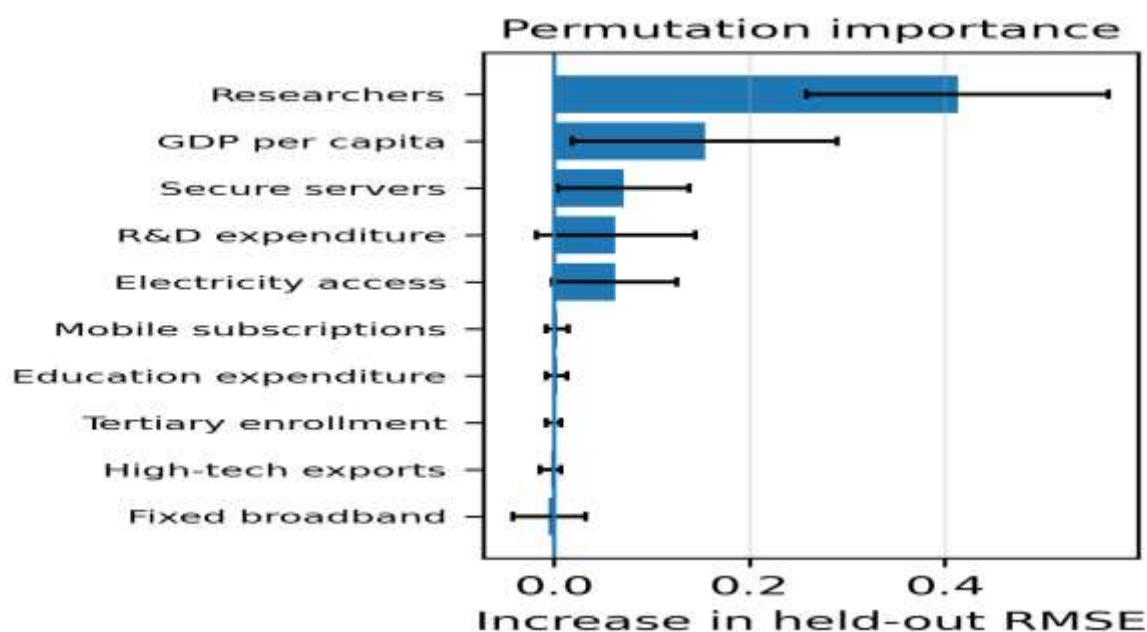


Figure 5. Cross-validated permutation importance for the random-forest model.

Table 4. Held-out permutation importance for the random-forest model.

Predictor	Mean delta RMSE	SD
Researchers per million	0.413	0.155
GDP per capita	0.154	0.136
Secure Internet servers	0.071	0.068
R&D expenditure	0.063	0.082
Electricity access	0.062	0.064
Mobile subscriptions	0.003	0.011
Education expenditure	0.003	0.011
Tertiary enrollment	-0.001	0.008
High-technology exports	-0.004	0.010
Fixed broadband	-0.005	0.037

3.4 Robustness and diagnostic residuals

The complete-case analysis ($n=79$) produced mean $R^2=0.810$, $RMSE=0.570$, $MAE=0.436$, and Spearman $\rho=0.914$. The direction and magnitude were therefore similar to the primary analysis, suggesting that limited fold-specific imputation did not create the main result.

Repeated out-of-fold residuals identified countries whose observed publication intensity differed substantially from the model estimate. Guatemala, Honduras, Myanmar, El Salvador, and Mali were among the largest negative residuals; Iraq, Sri Lanka, Kuwait, Ukraine, and Nigeria were among the largest positive residuals. These cases should be treated as diagnostic prompts rather than league-table findings. A residual may reflect collaboration structure, disciplinary specialization, diaspora networks, language, conflict-related denominator effects, bibliographic coverage, or data-timing differences that are absent from the model.

4. Discussion

4.1 Principal interpretation

The results show that a compact set of national research and digital indicators can estimate cross-country publication intensity with high rank agreement under repeated held-out evaluation. The strongest signal was the density of researchers, not a connectivity variable or a technology-export measure. This finding supports a human-capital-first interpretation: digital platforms and AI tools can amplify research work, but they do not substitute for a sufficiently large, supported scientific workforce.

GDP per capita contributed substantial predictive information beyond researcher density. Economic capacity likely summarizes many omitted system advantages, including institutional stability, laboratory quality, research administration, doctoral education, procurement capacity, library access, international mobility, and the ability

to retain researchers. It should not be read as an immutable determinant. Rather, its importance signals that research policy is embedded in a broader resource environment.

Secure-server density and electricity access were the next digital-system variables. The result is consistent with a layered architecture: reliable power enables computing; secure digital services enable authenticated data exchange and institutional platforms; researchers and R&D investment convert those foundations into knowledge production. The negligible incremental importance of mobile and fixed broadband after other features entered the model may reflect saturation, multicollinearity, and the fact that subscription counts do not measure research-grade bandwidth, latency, affordability, or campus network quality.

4.2 Implications for higher-education strategy

For university systems, the benchmark argues against an AI-first procurement strategy. Institutions with limited research staffing, unstable electricity, weak data governance, and fragmented digital identity systems are unlikely to transform research capacity merely by purchasing generative-AI licenses. A defensible sequence begins with research workforce development and retention, continues through stable R&D financing and doctoral pathways, and builds secure digital services and interoperable information governance around those capabilities.

The model also supports differentiated strategy. In research-intensive systems, AI investments may focus on secure compute, reproducible workflows, research software engineering, and responsible automation. In systems with a small research workforce, scholarships, doctoral supervision, protected research time, and regional research networks may yield larger foundational gains. Where electricity and secure digital services remain constrained, infrastructure reliability and shared national research platforms should precede high-cost AI deployment.

Higher-education dashboards should therefore distinguish inputs, enabling systems, outputs, and outcomes. Publication intensity is an output, not a complete measure of societal impact, teaching quality, research integrity, or inclusion. A mature information system would pair the present indicators with open-science compliance, data-management capacity, collaboration diversity, field-normalized citation impact, societal engagement, and measures of responsible AI governance. NIST's AI Risk Management Framework emphasizes trustworthy and risk-aware deployment [24], while UNESCO's guidance on generative AI calls for human-centered governance, capacity building, and protection of educational and research values [25].

4.3 Contribution to computer information systems research

The study contributes an information-system artifact rather than only a statistical association. The artifact integrates heterogeneous public data, preserves temporal provenance at the cell level, makes analytic exclusions explicit, embeds preprocessing within validation, and exposes model dependence through held-out importance. These design choices address common weaknesses in policy dashboards, where indicators are silently synchronized, missing observations are hidden, and composite scores are presented without predictive validation.

The approach is transferable. The same architecture can be adapted to regional research ecosystems, institutional benchmarking, industrial innovation systems, or domain-specific research capacity. Future implementations could use an automated extract-transform-load pipeline, a versioned data catalog, model cards, uncertainty-aware dashboards, and role-based access for policymakers and university leaders. The model should be treated as decision support, not an automated allocation rule.

4.4 Limitations

First, the study is cross-sectional and cannot establish temporal or causal effects. Research investment may affect publications with multi-year lags, and high-output systems may subsequently increase R&D budgets. Second, the analytic cohort is selected by availability of R&D and researcher data; countries with weaker statistical systems are underrepresented. The complete-case sensitivity analysis addresses imputation but not selection bias.

Third, scientific and technical article counts do not capture books, local-language journals, policy reports, patents, artistic outputs, data sets, software, or research quality. Database coverage can vary by country and discipline. Fourth, national indicators conceal within-country inequality among universities, regions, and fields. Fifth, source values within 2019–2023 are not perfectly synchronized; retaining source years improves transparency but does not eliminate temporal mismatch.

Sixth, permutation importance can distribute or suppress importance among correlated predictors and should not be interpreted causally. Seventh, the model contains only ten predictors and omits governance, academic freedom, collaboration networks, language, doctoral completion, migration, conflict, and institutional research management. Finally, descriptive bootstrap intervals across repeated folds do not have the same inferential interpretation as confidence intervals from independent samples.

5. Conclusions

An explainable machine-learning benchmark built from open international indicators can characterize national digital research ecosystems with strong out-of-sample agreement. Across 99 countries with observed core R&D inputs, researcher density was the most important predictive feature, followed by economic capacity, secure digital services, R&D intensity, and electricity access. The pattern supports a layered strategy for higher education: develop and retain researchers, sustain research investment, secure the digital environment, and ensure foundational infrastructure before expecting AI tools to transform research performance. The benchmark

is not a causal ranking and should not be used mechanically to allocate resources. Its principal value lies in transparent integration, auditable source years, explicit missing-data decisions, repeated validation, and an interpretable link between national information systems and higher-education research strategy.

ACKNOWLEDGMENTS

The authors acknowledge the World Bank and the UNESCO Institute for Statistics for maintaining the public indicator systems used in this study.

FUNDING INFORMATION

The authors received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors for this work.

AUTHOR CONTRIBUTIONS

R.A.L.: Conceptualization, educational interpretation, literature synthesis, writing - original draft, and writing - review and editing. L.C.M.D.J.: Conceptualization, methodology, software, data curation, formal analysis, validation, visualization, writing - original draft, and writing - review and editing. Both authors approved the final manuscript and accept accountability for all aspects of the work.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All secondary data used in the study are redistributed in the accompanying CSV package with indicator codes, source-year columns, inclusion flags, and derived model diagnostics. Original indicators are publicly available through the World Bank World Development Indicators portal. The package also contains a data dictionary, analysis sample, descriptive summaries, model-performance files, feature-importance results, diagnostic residual examples, figures, and reproducible Python code.

GENERATIVE AI DISCLOSURE

Generative AI assisted with language organization and document preparation. The authors independently verified the analysis, factual claims, citations, and final manuscript and remain fully responsible for the submitted work.

References

1. Gawande, V., and Gawande, S. V., 2026, "The AI-Native University: A Conceptual Framework and Maturity Model for AI Transformation in Higher Education," *International Journal of Computer Information Systems and Industrial Management Applications*, 18(6s), pp. 961–982. DOI: <https://doi.org/10.70917/ijcisim-2026-2999>
2. Vásquez Suarez, C. L., Vásquez Suarez, E., and Ospina Figueroa, M., 2026, "Governance of Generative Artificial Intelligence in Higher Education Institutions: An Approach from Digital Security and Data Protection," *International Journal of Computer Information Systems and Industrial Management Applications*, 18(4s), pp. 403–419. Available at: <https://cspub-ijcisim.org/index.php/ijcisim/article/view/2525> (Accessed on date: 28 July 2026).
3. Zhang, T., Dantes, G. R., Wahyuni, D. S., and Santosa, M. H., 2026, "Project-Based Learning in the Era of Generative Artificial Intelligence: Toward a Hybrid Epistemic Governance Framework for Undergraduate Computer Science Education," *International Journal of Computer Information Systems and Industrial Management Applications*, 18. DOI: <https://doi.org/10.70917/ijcisim-2026-1793>
4. Zawacki-Richter, O., Marín, V. I., Bond, M., and Gouverneur, F., 2019, "Systematic Review of Research on Artificial Intelligence Applications in Higher Education—Where Are the Educators?" *International Journal of Educational Technology in Higher Education*, 16, 39. DOI: <https://doi.org/10.1186/s41239-019-0171-0>
5. Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., and Martinez-Maldonado, R., 2024, "Human-Centred Learning Analytics and AI in Education: A Systematic Literature Review," *Computers and Education: Artificial Intelligence*, 6, 100215. DOI: <https://doi.org/10.1016/j.caeai.2024.100215>
6. Vial, G., 2019, "Understanding Digital Transformation: A Review and a Research Agenda," *Journal of Strategic Information Systems*, 28(2), pp. 118–144. DOI: <https://doi.org/10.1016/j.jsis.2019.01.003>
7. Furman, J. L., Porter, M. E., and Stern, S., 2002, "The Determinants of National Innovative Capacity," *Research Policy*, 31(6), pp. 899–933. DOI: [https://doi.org/10.1016/S0048-7333\(01\)00152-4](https://doi.org/10.1016/S0048-7333(01)00152-4)
8. Romer, P. M., 1990, "Endogenous Technological Change," *Journal of Political Economy*, 98(5), pp. S71–S102. DOI: <https://doi.org/10.1086/261725>
9. May, R. M., 1997, "The Scientific Wealth of Nations," *Science*, 275(5301), pp. 793–796. DOI: <https://doi.org/10.1126/science.275.5301.793>

10. King, D. A., 2004, "The Scientific Impact of Nations," *Nature*, 430, pp. 311–316. DOI: <https://doi.org/10.1038/430311a>
11. Man, J. P., Weinkauff, J. G., Tsang, M., and Sin, D. D., 2004, "Why Do Some Countries Publish More Than Others? An International Comparison of Research Funding, English Proficiency and Publication Output in Highly Ranked General Medical Journals," *European Journal of Epidemiology*, 19, pp. 811–817. DOI: <https://doi.org/10.1023/B:EJEP.0000036571.00320.b8>
12. Meo, S. A., Al Masri, A. A., Usmani, A. M., Memon, A. N., and Zaidi, S. Z., 2013, "Impact of GDP, Spending on R&D, Number of Universities and Scientific Journals on Research Publications among Asian Countries," *PLOS ONE*, 8(6), e66449. DOI: <https://doi.org/10.1371/journal.pone.0066449>
13. OECD, 2015, *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development*, OECD Publishing, Paris. DOI: <https://doi.org/10.1787/9789264239012-en>
14. World Bank, 2026, "World Development Indicators," Washington, DC, accessed 28 July 2026. <https://data.worldbank.org/indicator>
15. UNESCO Institute for Statistics, 2026, "2026 R&D Data Release," Montreal, Canada. <https://www.uis.unesco.org/en/2026-rd-data-release>
16. Breiman, L., 2001, "Random Forests," *Machine Learning*, 45, pp. 5–32. DOI: <https://doi.org/10.1023/A:1010933404324>
17. Friedman, J. H., 2001, "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics*, 29(5), pp. 1189–1232. DOI: <https://doi.org/10.1214/aos/1013203451>
18. Pedregosa, F., et al., 2011, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, 12, pp. 2825–2830. <https://jmlr.org/papers/v12/pedregosa11a.html>
19. Lundberg, S. M., and Lee, S.-I., 2017, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems*, 30. <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>
20. Ribeiro, M. T., Singh, S., and Guestrin, C., 2016, "Why Should I Trust You? Explaining the Predictions of Any Classifier," *Proc. NAACL Demonstrations*, pp. 97–101. DOI: <https://doi.org/10.18653/v1/N16-3020>
21. Roscher, R., Bohn, B., Duarte, M. F., and Garcke, J., 2020, "Explainable Machine Learning for Scientific Insights and Discoveries," *IEEE Access*, 8, pp. 42200–42216. DOI: <https://doi.org/10.1109/ACCESS.2020.2976199>
22. Varma, S., and Simon, R., 2006, "Bias in Error Estimation When Using Cross-Validation for Model Selection," *BMC Bioinformatics*, 7, 91. DOI: <https://doi.org/10.1186/1471-2105-7-91>
23. van Buuren, S., and Groothuis-Oudshoorn, K., 2011, "mice: Multivariate Imputation by Chained Equations in R," *Journal of Statistical Software*, 45(3), pp. 1–67. DOI: <https://doi.org/10.18637/jss.v045.i03>
24. National Institute of Standards and Technology, 2023, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1, Gaithersburg, MD. DOI: <https://doi.org/10.6028/NIST.AI.100-1>
25. UNESCO, 2023, *Guidance for Generative AI in Education and Research*, UNESCO, Paris. DOI: <https://doi.org/10.54675/EWZM9535>