

Keywords: DEA, machine learning, SHAP, AI adoption, XGBoost, random forest, digital transformation, EU

Ewa CHODAKOWSKA ^{1*}

¹ Białystok University of Technology, Poland, e.chodakowska@pb.edu.pl

* Corresponding author: e.chodakowska@pb.edu.pl

A data-driven framework for AI adoption efficiency assessment using hybrid DEA and machine learning methods

Abstract

Despite growing academic interest in artificial intelligence (AI) as a driver of economic transformation, the efficiency with which EU member states translate R&D investment and ICT human capital into measurable AI adoption outcomes remains insufficiently explored. This study proposes a data-driven computational framework that integrates data envelopment analysis (DEA) with machine learning (ML) to assess and explain the efficiency of AI adoption across EU countries during 2021–2024. In the first stage, an output-oriented VRS DEA model is combined with smoothed bootstrap bias correction. In the second stage, extreme gradient boosting (XGBoost) and random forest models, with Shapley additive explanations (SHAP)-based decomposition, are employed to identify key predictors of efficiency, while bootstrapped truncated regression provides complementary statistical inference for robustness validation. Average bias-corrected efficiency increased over time, reflecting the diffusion of AI technologies across the EU. However, the results reveal substantial heterogeneity in the efficiency of AI adoption across EU member states. SHAP analysis identifies GDP per capita, regulatory quality, and the share of the ICT sector as the most important predictors of AI adoption efficiency. Country-grouped cross-validation indicates limited out-of-country predictive generalisability, while sensitivity analyses confirm broad stability of the leading-predictor structure. The truncated regression yields no statistically significant covariate effects, highlighting the inferential uncertainty associated with the small country-year sample. The study demonstrates the usefulness of combining bootstrap-corrected DEA, XGBoost, random forests, and SHAP for interpretable efficiency analysis in rapidly evolving technological environments.

1. INTRODUCTION

Innovation systems have been widely studied across multiple theoretical perspectives, including national innovation systems – NIS (Freeman, 1987; Lundvall, 2010), endogenous growth theory (Romer, 1990), and evolutionary economics (Nelson & Winter, 1982). A common conclusion from the literature is that innovation performance is systemic and depends on complex interactions among institutions, resources, and capabilities – yet it is typically measured by absolute outputs. However, a critical dimension of innovation performance is efficiency – defined as the ability to maximise outputs relative to given inputs (Chodakowska, 2026).

The rapid diffusion of artificial intelligence (AI) technologies across economies has created a new dimension of innovation performance. However, the benefits of technological innovations are unevenly distributed (Orynycz et al., 2026). This paper examines how efficiently European Union (EU) member states transform research and development (R&D) investment and information and communications technology (ICT) human capital into AI adoption outcomes. Although existing datasets reveal significant variation in AI adoption across countries, current approaches focus primarily on implementation levels rather than efficiency. Furthermore, the factors associated with differences in AI adoption efficiency remain largely unexplored.

To address these challenges, this study proposes a data-driven computational framework integrating data envelopment analysis (DEA) with machine learning (ML) techniques. DEA is a well-established non-parametric method for efficiency analysis (Charnes et al., 1978), with bootstrap extensions improving statistical reliability (Simar & Wilson, 1998, 2007). In the second stage, this study employs extreme gradient boosting (XGBoost) (Chen & Guestrin, 2016), and random forest (Breiman, 2001), combined with Shapley additive explanations (SHAP) to ensure model interpretability (Lundberg & Lee, 2017). The choice of ML

methods balances nonlinear modelling capacity, interpretability, and computational efficiency (Cioch et al., 2025).

The proposed framework enables both robust efficiency estimation and transparent exploration of key predictors of AI adoption efficiency across EU member states, using panel data for the period 2021–2024. By integrating efficiency analysis with interpretable ML, the study contributes to the development of data-driven decision support approaches for innovation policy.

The remainder of the paper is organised as follows. Section 2 reviews the literature on DEA in innovation and digital efficiency analysis, hybrid DEA–ML approaches, and AI adoption across EU countries. Section 3 describes the two-stage methodological framework and variable selection. Section 4 presents empirical results, including country-level efficiency scores, SHAP-based predictor analyses, and complementary truncated regression and sensitivity analyses. Section 5 concludes and outlines directions for future research.

2. LITERATURE BACKGROUND

The literature review draws on both individual empirical studies and available systematic reviews. The synthesis of existing reviews follows a meta-review approach, which evaluates and synthesises evidence from existing systematic literature reviews to provide a comprehensive analysis of the current research landscape (Chodakowska et al., 2024; Lunny et al., 2017; Tranfield et al., 2003).

Systematic reviews of the DEA-based literature confirm that it is widely regarded as a powerful non-parametric method for measuring the relative efficiency of national innovation systems, given its ability to handle multiple inputs and outputs simultaneously without requiring parametric assumptions (Narayanan et al., 2022; Zeng et al., 2021). DEA constitutes one of three primary methodological approaches to cross-country efficiency measurement, alongside single indicators and composite indices (Khan et al., 2025).

Considering cross-country and regional comparisons, the DEA-based literature on innovation and digital efficiency – identified through a systematic term frequency-inverse document frequency (TF-IDF) and K-Means clustering approach with principal component analysis (PCA) dimensionality reduction (Chodakowska, Nazarko, et al., 2026) – is organised into three thematic domains: (1) national innovation systems (NIS), (2) green and sustainable innovation, and (3) digital transformation & digital economy. It should be noted that these domains are not mutually exclusive; several publications are cross-coded across multiple categories due to the inherent interdisciplinary nature of the digital and green transition and the overlap between knowledge production and digital governance. Consequently, the cumulative number of articles across clusters exceeds the total sample size (154). Details of the clusters obtained via semi-automated modelling are presented in Table 1.

Tab. 1. Thematic clusters of DEA applications in innovation and digital-economy efficiency research

Research cluster	N	Dominant terms	Representative works (most frequently cited)	Recent studies
National Innovation Systems (NIS)	50	innovation, national innovation, national systems, performance benchmarking	(Guan & Chen, 2012) (Carayannis et al., 2016)	(Ndicu et al., 2024)
Green & Sustainable Innovation	59	green innovation, environmental efficiency, renewable energy, eco-innovation, carbon neutrality	(Tang et al., 2020) (Bresciani et al., 2021)	(Hamid & Wang, 2024)
Digital Transformation & Digital Economy	145	digital economy, e-government, ICT, R&D intensity, human capital, digital skills, institutional quality, AI adoption	(Aristovnik, 2012)	(Bánhidi & Dobos, 2024) (Zhou et al., 2026)

A targeted Scopus search combining DEA with AI, ML, and technology adoption terms returned 64 results, only a few of which directly address AI-related efficiency. Current research remains fragmented, focusing either on regional capabilities in specific economies (Luo et al., 2026) or sector-specific innovation performance (Shi et al., 2024; Wang et al., 2024), leaving a gap in cross-country comparative assessments of AI adoption efficiency. Operational perspectives are provided by, e.g., Sun et al. (2026), who used a bootstrap DEA approach to show that AI patents and digital maturity are significantly associated with efficiency, whereas general AI investments often are not. Despite these advancements, AI adoption is often treated as one of several secondary indicators rather than the central focus of efficiency analysis, and existing work

predominantly relies on firm-level data, leaving a gap in cross-country comparative assessments at the macro level.

Recent publications from 2025–2026, e.g., Zhou et al. (2026) and Furlan et al. (2025), are moving away from pure DEA in favour of hybrid models. These approaches use ML to identify nonlinear relationships between digital maturity and efficiency. An overview of the DEA-ML integration (Kehinde et al., 2026) covers a wide range of ML models, from early approaches such as artificial neural networks (ANN) and support vector machines (SVM), through ensemble methods including random forest (RF) and gradient boosting, to more recent deep learning techniques and emerging frontier-learning models such as EAT, SVF, and DEAM.

Table 2 presents the term frequency analysis of DEA and ML models across various decision-making unit (DMU) contexts. A large share of articles remains unclassified, indicating that AI is often used conceptually rather than methodologically. It should be noted that categories are not mutually exclusive, as individual studies may employ multiple ML techniques. Among identified approaches, unsupervised methods and neural networks dominate, while modern techniques such as tree-based models and explainable AI are gaining prominence but remain comparatively underrepresented.

Tab. 2. Machine-learning methods in DEA-related studies

ML method	Count	Share (%)	Role in the DEA framework	ML/AI function
Unclassified / no specific method detected	726	50.24	general DEA-related study	no specific ML/AI method detected in title/abstract/keywords
Unsupervised ML	307	21.25	DMU grouping/segmentation before or after DEA	clustering, dimensionality reduction, structure discovery
Neural networks / deep learning	205	14.19	efficiency approximation / nonlinear efficiency modelling	nonlinear mapping, prediction, pattern recognition
Tree-based methods	127	8.79	efficiency estimation / second-stage prediction	prediction, classification, feature importance
Fuzzy / ANFIS	87	6.02	handling uncertainty and nonlinear efficiency modelling	fuzzy inference, rule-based reasoning, uncertainty modelling
Support vector machines	68	4.71	efficiency classification/frontier or boundary estimation	classification, regression, margin-based learning
Regression / econometric ML	60	4.15	second-stage DEA / predictor analysis	statistical modelling, inference, explanatory analysis
Sequential deep learning	21	1.45	dynamic efficiency analysis / temporal efficiency modelling	time-series prediction, sequence modelling
Explainable AI	20	1.38	post-DEA interpretation/explanation of efficiency variations	explainability, feature contribution, model interpretation
KNN	13	0.90	similarity-based DMU comparison/classification support	nearest-neighbour classification or regression
Double/causal ML	2	0.14	causal analysis of efficiency relationships	causal inference, treatment-effect estimation

Note: categories are not mutually exclusive; articles may be assigned to multiple categories. Share (%) is calculated relative to the total number of articles (N=1,445).

Figure 1 shows the evolution of ML methods in DEA-based studies. Explainable AI techniques appear only in recent years and still account for a relatively small share of identified approaches. Despite their growing importance in the broader machine learning field, adoption of these methods in DEA-based research remains limited.

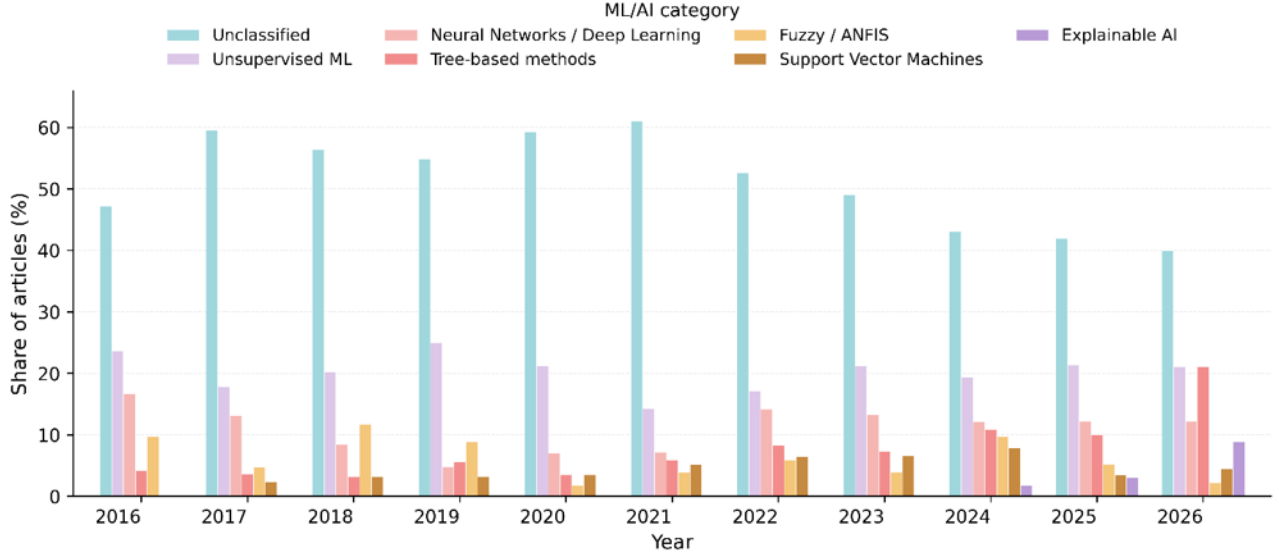


Fig. 1. Evolution of ML methods in DEA literature (categories are not mutually exclusive)

A targeted Scopus query combining DEA, boosting methods (including XGBoost), and explainable AI techniques (such as SHAP, LIME, and XAI) returned only a very small number of studies. This finding highlights a mismatch between the rapid development of advanced ML methods and their actual use in DEA-based research. While hybrid DEA–ML approaches are increasingly reported, they remain largely based on traditional techniques such as neural networks, clustering, and regression models. This study directly addresses this gap by combining bootstrap-corrected DEA with XGBoost and random forest interpreted through SHAP values, alongside bootstrapped truncated regression for complementary statistical inference.

3. ANALYTICAL FRAMEWORK AND MODEL SPECIFICATION

3.1. Analytical process model

The analytical pipeline follows a structured, data-driven approach consistent with established data-mining methodologies such as CRISP-DM, which emphasise business understanding, data preparation, and iterative modelling (Brzozowska et al., 2023; Chodakowska, Danilczuk, et al., 2026). The study employs a two-stage analytical framework (Figure 2) to assess the efficiency of AI adoption across EU member states and to explore the structural factors associated with differences in efficiency.

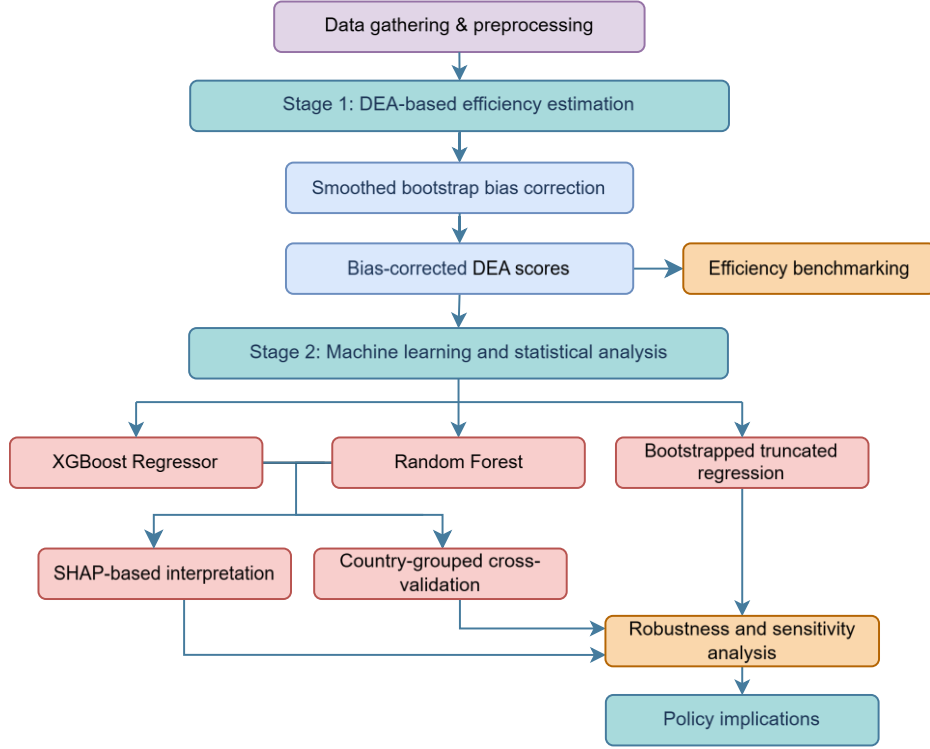


Fig. 2. Two-stage analytical process model

In the first stage, an output-oriented variable returns-to-scale (O-BCC) DEA model (Banker et al., 1984; Charnes et al., 1978) is estimated separately for each year and subsequently subjected to a smoothed bootstrap bias-correction procedure (Simar & Wilson, 1998, 2007). The DEA model estimates the Farrell output-expansion factor $\phi \geq 1$, while the corresponding efficiency score reported in the analysis is $E = 1/\phi$. The efficiency for each country is calculated by solving the linear programming problem:

$$\begin{aligned}
 & \max \phi, \\
 & \sum_{j=1}^n \lambda_j x_{ij} \geq x_{io}, \quad i = 1, \dots, m, \\
 & \sum_{j=1}^n \lambda_j y_{rj} \leq \phi y_{ro}, \quad r = 1, \dots, s, \\
 & \lambda_j \geq 0, \quad j = 1, \dots, n, \\
 & \sum_{j=1}^n \lambda_j = 1, \\
 & \phi \geq 1,
 \end{aligned} \tag{1}$$

where: $\mathbf{x}_j = [x_{1j}, x_{2j}, \dots, x_{ij}, \dots, x_{mj}]$ – input vector,
 $\mathbf{y}_j = [y_{1j}, y_{2j}, \dots, y_{rj}, \dots, y_{sj}]$ – output vector,
 s – the number of outputs,
 m – the number of inputs,
 λ_j – the weights,
 j – index of DMUs in the reference set, $j = 1, \dots, n$,
 o – the DMU under evaluation.

The variable returns to scale (VRS/BCC) specification was selected, given the substantial heterogeneity in country size and economic development across EU member states, for which the assumption of constant returns to scale would be inappropriate. The output-oriented specification was utilised because countries have limited short-term ability to reduce structural innovation inputs such as R&D expenditure or ICT employment, while policy objectives primarily focus on maximising AI-related innovation outputs.

To reduce the finite-sample bias inherent in DEA estimates, a smoothed bootstrap procedure was applied separately within each year. Bootstrap pseudo-samples were generated from a kernel-smoothed distribution of Farrell efficiency estimates with reflection at the boundary ($\phi=1$), and the DEA model was re-estimated for each pseudo-sample. A total of $B=2,000$ bootstrap replications were performed. For each DMU, bootstrap bias was estimated as:

$$\widehat{Bias}(\hat{\phi}) = \frac{1}{B} \sum_{b=1}^B \hat{\phi}^{*(b)} - \hat{\phi}, \quad (2)$$

and the bias-corrected Farrell score was obtained as:

$$\hat{\phi}^{BC} = \hat{\phi} - \widehat{Bias}(\hat{\phi}) \quad (3)$$

For presentation and subsequent ML analysis, the bias-corrected DEA efficiency score was expressed as:

$$E^{BC} = \frac{1}{\hat{\phi}^{BC}} \quad (4)$$

No post-hoc clipping of efficiency scores was applied.

The second stage employs three complementary analytical approaches to explore predictive associations with bias-corrected efficiency and assess their robustness. First, an XGBoost regressor – a flexible ensemble method capable of capturing complex nonlinear relationships and interaction effects among environmental variables – is fitted on the full sample to obtain SHAP values, which decompose each predictor's contribution to the model prediction. Second, a random forest model is estimated as a robustness check. Third, a bootstrapped truncated-normal regression is estimated on the bias-corrected Farrell scores $\hat{\phi}^{BC}$, which are bounded below by one, to provide complementary statistical inference. XGBoost and random forest capture potential nonlinear relationships and interaction effects without imposing a restrictive functional form, whereas the truncated regression provides complementary parametric inference. Consistency of predictor rankings between the two ML models is treated as evidence of the robustness of the feature-importance structure rather than of causality. The regression results are used to assess whether the corresponding linear associations are statistically distinguishable from zero.

Given the small panel ($n=81$) and repeated observations for each country, predictive performance was evaluated using five-fold GroupKFold cross-validation with country as the grouping variable. Consequently, observations from the same country were never simultaneously included in the training and validation sets. Cross-validation was used as a diagnostic of out-of-country generalisability rather than for hyperparameter optimisation.

The XGBoost objective function combines a differentiable loss function with a regularisation component controlling model complexity:

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{t=1}^T \Omega(f_t) \quad (5)$$

where:

- i – the country-year observation,
- n – the total number of observations,
- T – the number of trees,
- f_t – t -th decision tree,
- l – the differentiable convex loss function,
- Ω – the regularisation term.

The final XGBoost specification used 300 trees, a maximum depth of 3, a learning rate of 0.05, a subsample ratio of 0.8, a column subsampling ratio of 0.8, $L_1=0.1$, and $L_2=1.0$.

Random forest aggregates predictions across M independently grown decision trees:

$$\hat{y} = \frac{1}{M} \sum_{b=1}^M T_b(x) \quad (6)$$

where:

- M – the number of trees,

$T_b(x)$ – the b -th tree prediction for observation x .

The random forest model used 500 trees, a maximum depth of 4, a minimum leaf size of 3, and $\sqrt{p}$ predictors considered at each split.

$$\hat{\phi}^{BC} = \mathbf{z}^T \beta + \varepsilon, \varepsilon \sim N(0, \sigma^2), \hat{\phi}^{BC} > 1 \quad (7)$$

$$f(\hat{\phi}^{BC} | \mathbf{z}_i, \hat{\phi}^{BC} > 1) = \frac{\frac{1}{\sigma} \varphi\left(\frac{\hat{\phi}^{BC} - \mathbf{z}_i^T \beta}{\sigma}\right)}{1 - \Phi\left(\frac{1 - \mathbf{z}_i^T \beta}{\sigma}\right)} \quad (8)$$

where $\mathbf{z}_i$ is the vector of second-stage predictors, β denotes the regression coefficients, and σ is the error standard deviation. Statistical uncertainty was assessed using 2,000 parametric bootstrap replications. A country-block bootstrap was additionally used as a sensitivity check for within-country dependence.

SHAP assumes that the contribution of each feature k to the prediction for a specific country is defined as:

$$\varphi_k = \sum_{S \subseteq \{x_1, \dots, x_p\} \setminus \{x_k\}} \frac{|S|!(p-|S|-1)!}{p!} [f(S \cup \{x_k\}) - f(S)] \quad (9)$$

where:

p – the total number of features in the model,

φ_k – the SHAP value (contribution) of feature k to the prediction for a specific observation,

S – the subset of features that does not include the feature x_k ($S \subseteq \{x_1, \dots, x_p\} \setminus \{x_k\}$),

$|S|$ – the number of features in the subset,

$f(S \cup \{x_k\}) - f(S)$ – the marginal contribution of feature x_k when added to the subset S ,

$\frac{|S|!(p-|S|-1)!}{p!}$ – the weighting factor representing the probability of a specific permutation in which feature k is added to subset S .

SHAP values are interpreted as contributions to model predictions and do not represent causal effects of the predictors on DEA efficiency.

The proposed framework offers several methodological advantages. First, smoothed bootstrap bias correction reduces finite-sample bias in the DEA estimates and provides a more statistically defensible efficiency measure. Second, XGBoost enables the exploration of nonlinear relationships and interaction effects without imposing restrictive functional-form assumptions. Third, SHAP enhances interpretability by providing transparent, observation-level estimates of feature contributions to model predictions. Fourth, robustness is assessed through cross-model comparison between XGBoost and random forest, country-grouped cross-validation, bootstrapped truncated regression, and sensitivity analyses of influential observations and constructed predictor values.

3.2. Data source and variable selection

The selection of input and output variables reflects the conceptual distinction between resource endowment and AI adoption outcomes, while second-stage predictors capture structural and institutional factors associated with efficiency differences across countries (Table 3). Table 4 presents descriptive statistics for the analytical sample.

Tab. 3. Variables, definitions, and data availability

Variable		Definition	Availability
Stage 1 – DEA model (BCC, output-oriented, bootstrap)			
Input	R&D expenditure	Gross domestic expenditure on R&D (GERD) as % of GDP; all sectors of performance	2021–2024
	ICT specialists	Persons employed as ICT specialists as % of total employment (Labour Force Survey)	2021–2024
Output	AI adoption rate	% of enterprises (10+ employees, NACE C10–S951 excl. financial sector) using at least one AI technology	2021, 2023, 2024 2022: not collected
	AI adoption depth	% of enterprises using at least three AI technology categories – proxy for intensity of AI integration	2021, 2023, 2024 2022: not collected
Stage 2 – ML and statistical analysis (predictors of efficiency scores)			
Predictor	Digital skills	% of individuals aged 16–74 with basic or above basic overall digital skills (DigComp framework, all five components)	2021, 2023, 2022, 2024: linear interpolation
	ICT sector share	Value added of ICT sector as % of gross value added (GVA); NACE J + related	2021–2023 2024: forward-fill from 2023; Cyprus & Luxembourg: EU27 average
	Government effectiveness	WGI Government Effectiveness estimate: reflects perceptions of quality of public services, civil service and policy formulation (scale –2.5 to +2.5)	2021–2024; Malta: EU27 average
	Regulatory quality	WGI Regulatory Quality estimate: reflects perceptions of the ability of government to formulate and implement sound policies (scale –2.5 to +2.5)	2021–2024
	GDP per capita	GDP per capita expressed as % of EU27 average (PPS-based, current prices) – proxy for economic development level	2021–2024
	Tertiary education	% of population aged 25–34 with tertiary education (ISCED levels 5–8) – proxy for human capital and absorptive capacity for AI	2021–2024

Tab. 4. Descriptive statistics

	Mean	Std. Dev.	Min	Q1	Median	Q3	Max	Skew.	Kurt.
R&D expenditure	1.764	0.912	0.46	1.04	1.5	2.22	3.64	0.573	-0.837
ICT specialists	5.22	1.345	2.4	4.4	5.2	5.7	8.7	0.414	0.619
AI adoption rate	10.072	6.295	1.38	5.19	8.76	13.1	27.58	1.019	0.475
AI adoption depth	2.934	2.374	0.53	1.24	2.45	3.7	12.01	1.806	3.652
Digital skills	57.177	12.33	27.73	49.46	58.89	64.68	82.7	-0.177	0.247
ICT sector share	6.154	5.395	3.14	4.17	4.82	6.11	34.82	4.396	19.607
Government effectiv.	1.133	0.558	-0.078	0.743	1.12	1.579	2.191	-0.084	-0.622
Regulatory quality	1.066	0.543	0.238	0.621	1.1	1.498	2.011	0.17	-1.297
GDP per capita	103.857	42.55	59.4	77	91.3	116.3	261.9	2.292	5.322
Tertiary education	45.122	9.894	22.5	39.2	44.5	51.9	65.2	-0.038	-0.335

The ICT sector share has a very high kurtosis (19.5) – this is due to the outliers of Ireland and Luxembourg (~35% of GVA). These distributional characteristics indicate the presence of structurally atypical country observations; their influence is therefore examined explicitly in the sensitivity analysis. GDP per capita is also right-skewed (2.27), largely due to Luxembourg (~262% of the EU average). The robustness of results with respect to these observations is addressed through dedicated sensitivity analyses in the second stage.

Correlations were calculated using pooled panel observations for the period 2021–2024 (Figure 3).

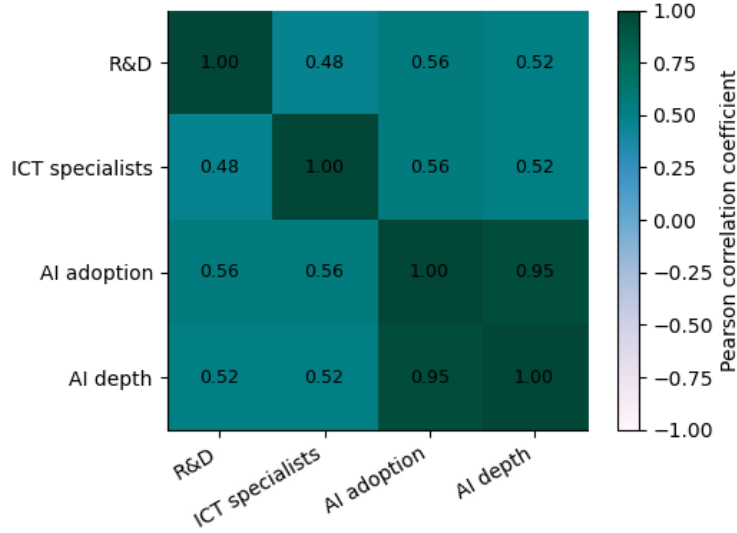


Fig. 3. Pearson correlations (pooled dataset covering 2021–2024)

The correlation matrix indicates moderate positive relationships between both DEA inputs and AI-related outputs, supporting the suitability of the selected model specification. In particular, ICT specialists and R&D expenditure show similar correlations with AI adoption rate ($r = 0.56$) and depth ($r = 0.52$), suggesting that both technological capabilities and digital human capital are associated with AI adoption outcomes. The correlation between the two inputs remains moderate ($r = 0.48$), indicating limited redundancy between the selected resource dimensions. As expected, AI adoption rate and depth are highly correlated ($r = 0.95$), reflecting their complementary nature as measures of AI diffusion and adoption intensity.

4. Results

4.1. DEA-based efficiency estimation

The DEA was conducted using a yearly cross-sectional approach, in which each country was evaluated relative to the efficiency frontier for that year. By estimating a separate frontier for each year, the analysis allows the AI adoption frontier to vary over the period 2021–2024. To improve the robustness of efficiency estimation, the original DEA scores were subjected to a smoothed bootstrap bias-correction procedure. For each year, 2,000 bootstrap replications were generated using the smoothed Farrell-efficiency distribution, and the DEA model was re-estimated for each pseudo-sample. Bias correction was performed on the output-oriented Farrell expansion factors, which were subsequently transformed into efficiency scores as: $E^{BC} = 1/\hat{\phi}^{BC}$. Figure 4 illustrates the bias-corrected DEA efficiency scores across years (overall mean = 0.569, min = 0.133, max = 0.871). The bias correction shifted all originally frontier-efficient observations below unity, reflecting the estimated finite-sample upward bias of the conventional DEA scores.

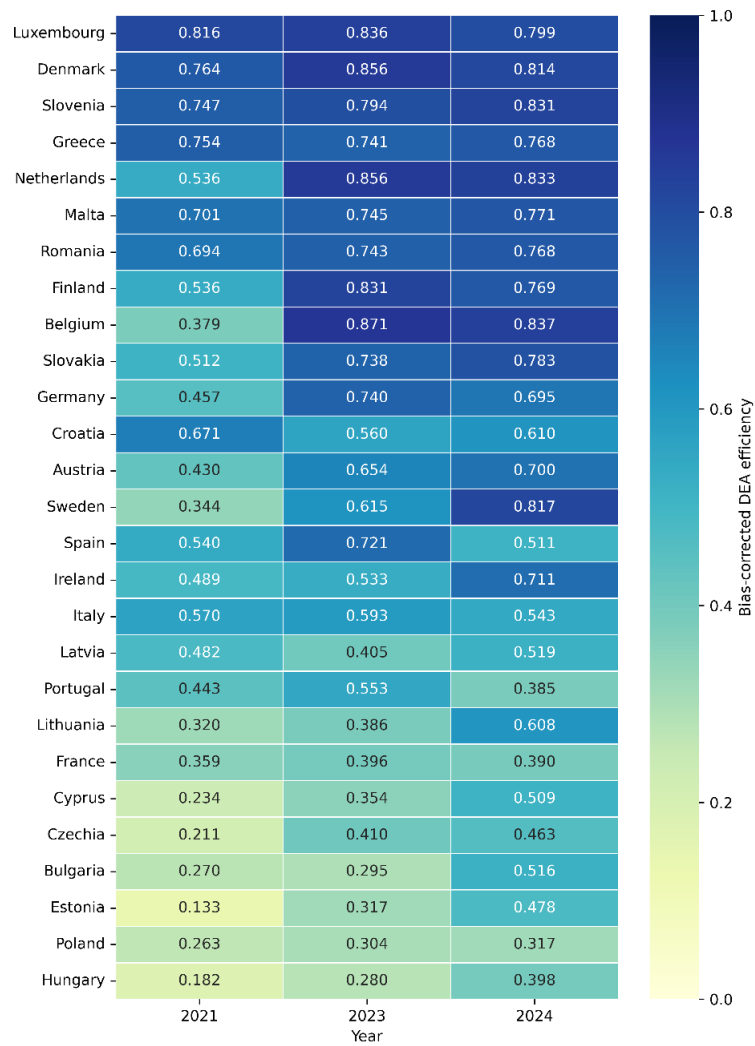


Fig. 4. Bias-corrected DEA efficiency scores across EU member states

The DEA results confirm substantial heterogeneity in the efficiency of AI adoption across EU member states. Denmark, Malta, Romania, and Luxembourg ranked among the relatively most efficient countries across the analysed period. Greece and Slovenia also recorded relatively high efficiency in the later years. In contrast, several Central and Eastern European economies, including Poland and Hungary, remained among the least efficient performers despite gradual improvements.

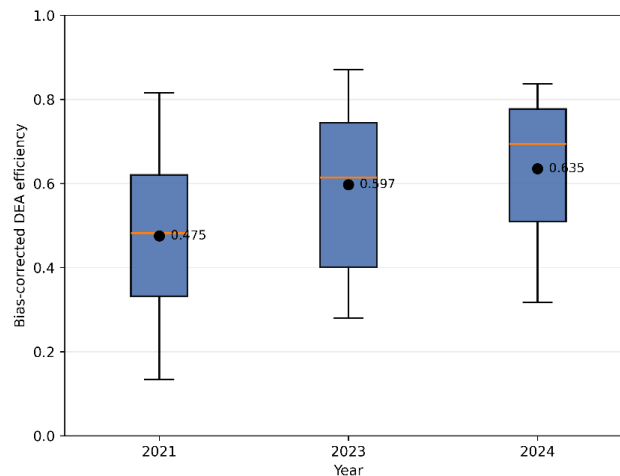


Fig. 5. Distribution of bias-corrected DEA efficiency scores by year

The results in Figure 5 indicate a noticeable increase in average bias-corrected efficiency between 2021 and 2024, suggesting that EU member states, on average, became more efficient relative to their respective annual frontiers. This pattern is consistent with the broader diffusion of AI technologies across the EU, although the DEA results alone do not establish a causal relationship.

4.2. Machine learning analysis

The second-stage analysis examines the predictive structure of bias-corrected DEA efficiency scores using two complementary machine learning approaches: XGBoost with SHAP-based decomposition as the primary model, and random forest with SHAP values as a cross-model robustness check.

SHAP feature importance from the XGBoost model identifies GDP per capita as the leading predictor of AI adoption efficiency (mean $|\text{SHAP}| = 0.066$), followed by regulatory quality (0.052), ICT sector share (0.050), digital skills (0.024), government effectiveness (0.024), and tertiary education (0.019). The random forest model broadly confirms this importance structure: GDP per capita is ranked first in both models, while ICT sector share and regulatory quality consistently occupy the next two positions, although their ordering differs between models (Table 5).

Tab. 5. SHAP feature importance – XGBoost and RF

Variable	XGBoost Mean $ \text{SHAP} $	RF Mean $ \text{SHAP} $
GDP per capita	0.0660	0.0451
Regulatory quality	0.0521	0.0241
ICT sector share	0.0504	0.0254
Digital skills	0.0239	0.0117
Government effectiveness	0.0237	0.0134
Tertiary education	0.0193	0.0074

The SHAP summary plot (Figure 6) reveals distinct patterns of predictive contribution across the three leading predictors. GDP per capita shows a predominantly positive and near-monotonic relationship with predicted efficiency and has the highest overall feature importance (mean $|\text{SHAP}| = 0.066$) – observations with high feature values (pink, high GDP) cluster on the right side of the SHAP axis, while low-income countries are associated with negative SHAP contributions. The dependence plot indicates a broadly positive relationship across the observed income distribution, although the dispersion among middle-income countries suggests that the contribution of GDP varies with the broader structural context.

Regulatory quality displays a more complex, non-monotonic pattern despite being the second-ranked XGBoost predictor (mean $|\text{SHAP}| = 0.052$). Positive and negative SHAP contributions occur across its observed range, indicating that its predictive role is not adequately represented by a simple linear relationship. This counterintuitive pattern may partly reflect sample composition, as low-regulatory-quality observations include several Baltic and Central and Eastern European economies with relatively high AI adoption efficiency. Accordingly, the SHAP pattern should be interpreted as a feature contribution to the model's predictions rather than as evidence of a causal effect of regulatory quality on efficiency.

ICT sector share exhibits a clear negative association at high values. Ireland and Luxembourg are notable cases because of their exceptionally high shares of the ICT sector. However, the sensitivity analysis shows that the importance of the ICT sector share is not driven solely by these two countries: after their exclusion, it remains the second-ranked predictor in both XGBoost and random forest (rank correlations with the full-sample ordering: $\rho = 0.886$ and 0.943 , respectively). This suggests that the observed relationship reflects a broader structural pattern rather than only the influence of the two extreme observations.

Government effectiveness, digital skills, and tertiary education contribute comparatively modest SHAP values across both models, indicating lower predictive importance relative to GDP per capita, regulatory quality, and ICT sector share.

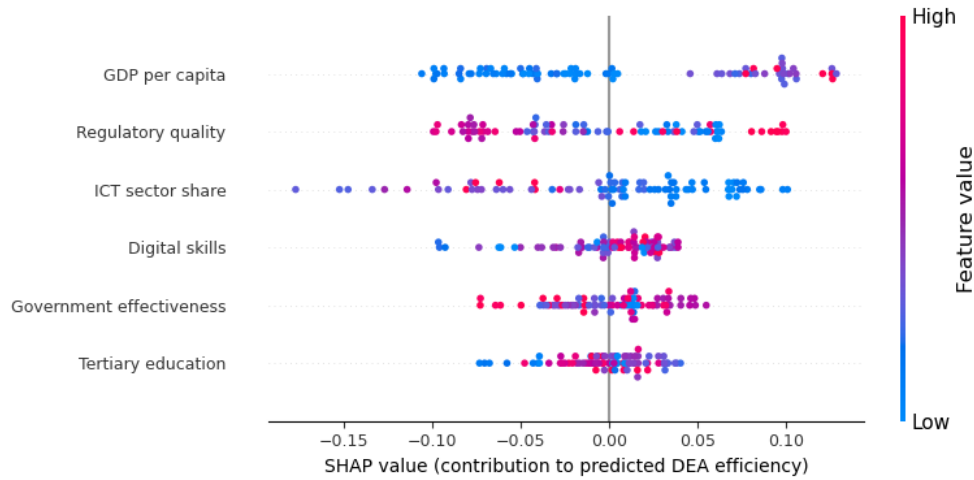


Fig. 6. SHAP summary plot – XGBoost model of bias-corrected DEA efficiency

Overall, the results suggest that AI adoption efficiency in the EU is associated not only with technological resources but also with broader economic and institutional conditions.

4.3. Robustness check

The robustness of SHAP-derived feature-importance structure is assessed using bootstrapped truncated regression, country-grouped cross-validation, and sensitivity analyses. The truncated regression is estimated for the bias-corrected Farrell output-expansion factor, with higher values indicating lower efficiency. Consequently, a negative regression coefficient corresponds to a positive association with DEA efficiency, and vice versa.

The truncated regression results provide no evidence of statistically significant linear associations among the six predictors (Table 6). Although GDP per capita shows a negative coefficient on the Farrell scale, corresponding to a positive association with DEA efficiency, and ICT sector share shows the opposite direction, all bootstrap confidence intervals include zero. Thus, the parametric analysis does not statistically confirm the predictive relationships identified by SHAP. Rather, it highlights substantial inferential uncertainty associated with the small country-year sample. The same conclusion is obtained using a country-block bootstrap that preserves the repeated observations within countries.

Tab. 6. Bootstrapped truncated regression results

Variable	Standardised coefficient	Bootstrap SE	95% bootstrap CI	p value
Intercept	-2.951	16.686	[-6.900; 12.643]	0.860
GDP per capita	-6.852	12.296	[-11.836; 8.434]	0.577
Digital skills	-1.596	3.261	[-3.611; 2.176]	0.625
ICT sector share	3.151	4.953	[-3.464; 6.090]	0.525
Tertiary education	0.312	1.619	[-1.663; 1.888]	0.847
Government effectiveness	2.071	5.108	[-4.509; 5.900]	0.685
Regulatory quality	1.113	3.331	[-2.317; 3.852]	0.738

Coefficients are reported for standardised predictors. The dependent variable is the bias-corrected Farrell output-expansion factor, where higher values indicate lower DEA efficiency. Statistical uncertainty is based on 2,000 parametric bootstrap replications. All 95% bootstrap confidence intervals include zero.

Country-grouped cross-validation further indicates limited out-of-country predictive generalisability. Random forest achieved an out-of-fold MAE of 0.185 and RMSE of 0.213 ($R^2 = -0.161$), whereas XGBoost achieved an MAE of 0.216 and RMSE of 0.246 ($R^2 = -0.554$). These results indicate that the small sample does not support strong claims regarding prediction for previously unseen countries. Accordingly, the ML models are interpreted primarily as exploratory tools for examining the predictive structure of the observed EU sample.

Sensitivity analyses nevertheless indicate that the main feature-importance structure is generally robust. Excluding Ireland and Luxembourg preserves the same three leading predictors—GDP per capita, ICT sector

share, and regulatory quality—with rank correlations of 0.886 for XGBoost and 0.943 for Random Forest relative to the full-sample rankings. Excluding 2024 observations produces similarly stable rankings, while excluding countries affected by EU-average substitutions retains the same three leading variables in both models. Only the most restrictive specification, which reduces the sample from 81 to 48 observations, produces noticeable instability in the XGBoost ordering.

5. CONCLUSIONS

This study proposed a data-driven computational framework integrating smoothed bootstrap-corrected DEA with ML to assess and explore predictive associations with AI adoption efficiency across EU member states over the period 2021–2024. The two-stage design – combining robust efficiency estimation with interpretable predictor analysis – addresses a gap in the literature where AI adoption is frequently treated as a secondary indicator rather than the central focus of macro-level efficiency assessment.

The first-stage results reveal substantial and persistent heterogeneity in the efficiency of AI adoption across EU member states. Bias-corrected efficiency increased on average from 0.475 in 2021 to 0.597 in 2023 and 0.635 in 2024, while substantial cross-country differences remained in each year. The bootstrap correction shifted all originally frontier-efficient observations below unity, indicating that conventional DEA scores were affected by finite-sample upward bias. In contrast, several Central and Eastern European economies, including Poland and Hungary, remained among the least efficient performers throughout the analysed period despite gradual improvements. The observed increase in average efficiency between 2021 and 2024 is consistent with the rapid diffusion of AI technologies across the EU and indicates that countries became, on average, more efficient relative to their respective annual frontiers.

The second-stage analysis identifies GDP per capita as the leading predictor of AI adoption efficiency in both machine-learning models. In the full-sample SHAP analysis, GDP per capita ranks first, while regulatory quality and ICT sector share rank second and third, respectively, although their ordering differs between XGBoost and random forest. This finding suggests that the level of macroeconomic development is strongly associated with variation in predicted AI adoption efficiency, rather than being a demonstrated binding constraint. Regulatory quality and ICT sector share emerge as important additional predictors, with nonlinear SHAP patterns that cannot be interpreted as causal effects. Sensitivity analysis further shows that the importance of the ICT sector share is not driven solely by Ireland and Luxembourg. Government effectiveness, digital skills, and tertiary education display lower predictive importance in the full-sample models. Importantly, the bootstrapped truncated regression yields no statistically significant covariate effects, with all confidence intervals including zero, indicating substantial inferential uncertainty. Country-grouped cross-validation also shows limited out-of-country predictive generalisability, reinforcing the exploratory interpretation of the ML results.

The methodological contribution of this study is threefold. First, integrating DEA with smoothed bootstrap bias correction improves the statistical reliability of efficiency scores compared to standard DEA, which is particularly important in small samples such as the EU-27 panel. Second, the use of XGBoost combined with SHAP enables exploration of nonlinear relationships and interaction effects among efficiency predictors, while providing transparent information about their contributions to model predictions. Third, robustness is assessed through cross-model comparison with random forest, country-grouped cross-validation, bootstrapped truncated regression, and sensitivity analyses addressing influential observations and constructed predictor values. Together, these complementary checks distinguish relatively stable feature-importance patterns from findings that are sensitive to model specification or sample composition.

Several limitations of this study should be acknowledged. The panel covers only three time points (2021, 2023, 2024) due to the availability pattern of AI adoption data by Eurostat. The small sample size ($n=81$) limits the out-of-sample generalisability of ML models, as confirmed by negative out-of-fold R^2 values under country-grouped cross-validation. Consequently, the ML models should be interpreted primarily as exploratory tools for identifying predictive associations within the observed EU sample rather than as models with strong predictive validity for unseen countries. In addition, some predictor observations required interpolation, forward-filling, or substitution with EU27 averages; although dedicated sensitivity analyses indicate broad stability of the main feature-importance structure, these data-construction choices remain a source of uncertainty.

Future research should extend the framework in several directions. As Eurostat's AI adoption data accumulate across more survey waves, dynamic efficiency models – including window DEA or Malmquist productivity indices – will become feasible, enabling the decomposition of efficiency change into technical progress and catch-up components. The integration of causal ML methods, such as double ML or causal forests, would allow the identification of causal effects under appropriate identification assumptions, moving beyond the correlational SHAP framework applied here. Finally, extending the analysis to firm-level microdata would enable the bridging of macro-level efficiency patterns with micro-level adoption decisions, providing a more complete picture of the mechanisms through which economic and institutional conditions are associated with AI adoption efficiency across Europe.

Funding

This research was conducted within the framework of project No. WZ/WIZ-INZ/3/2025 of Bialystok University of Technology, and was financed by a subsidy granted by the Minister of Science and Higher Education of the Republic of Poland.

Conflicts of Interest

The authors declare no conflict of interest.

REFERENCES

- Aristovnik, A. (2012). The relative efficiency of education and R&D expenditures in the new EU member states. *Journal of Business Economics and Management*, 13(5), 832–848. <https://doi.org/10.3846/16111699.2011.620167>
- Bánhidi, Z., & Dobos, I. (2024). Measuring digital development: Ranking using data envelopment analysis (DEA) and network readiness index (NRI). *Central European Journal of Operations Research*, 32, 1089–1108. <https://doi.org/10.1007/s10100-024-00919-y>
- Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078–1092. <https://doi.org/10.1287/mnsc.30.9.1078>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Bresciani, S., Puertas, R., Ferraris, A., & Santoro, G. (2021). Innovation, environmental sustainability and economic development: DEA-Bootstrap and multilevel analysis to compare two regions. *Technological Forecasting and Social Change*, 172, Article 121040. <https://doi.org/10.1016/j.techfore.2021.121040>
- Brzozowska, J., Pizoń, J., Baytikenova, G., Gola, A., Zakimova, A., & Piotrowska, K. (2023). Data engineering in CRISP-DM process: Production data – case study. *Applied Computer Science*, 19(3), 83–95. <https://doi.org/10.35784/acs-2023-26>
- Carayannis, E. G., Grigoroudis, E., & Goletsis, Y. (2016). A multilevel and multistage efficiency evaluation of innovation systems: A multiobjective DEA approach. *Expert Systems with Applications*, 62, 63–80. <https://doi.org/10.1016/j.eswa.2016.06.017>
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444. [https://doi.org/10.1016/0377-2217\(78\)90138-8](https://doi.org/10.1016/0377-2217(78)90138-8)
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>
- Chodakowska, E. (2026). Data envelopment analysis and technologies of the industry of the future: A scoping review. *Management and Production Engineering Review*, 17(2), 1–20. <https://doi.org/10.24425/mper.2026.1318>
- Chodakowska, E., Danilczuk, W., & Nazarko, J. (2026). Artificial intelligence assistance in foresight research: Enhancing technology assessment through data-driven methods. *Advances in Science and Technology Research Journal*, 20(3), 299–317. <https://doi.org/10.12913/22998624/211285>
- Chodakowska, E., Nazarko, J., & Chodakowska, N. (2026). Integrating AI skills with engineering management needs: Evidence from polish management and production engineering curricula and expert SWOT-AHP analysis. *IEEE Engineering Management Review*, 1–11. <https://doi.org/10.1109/EMR.2026.3714927>
- Chodakowska, E., Nazarko, J., Nazarko, Ł., & Rabayah, H. S. (2024). Solar radiation forecasting: A systematic meta-review of current methods and emerging trends. *Energies*, 17(13), Article 3156. <https://doi.org/10.3390/en17133156>
- Cioch, M., Kulisz, M., & Gola, A. (2025). Comparison of machine learning methods in predictive maintenance of machines. *Advances in Science and Technology Research Journal*, 19(11), 33–44. <https://doi.org/10.12913/22998624/208284>
- Freeman, C. (1987). *Technology policy and economic performance: Lessons from Japan*. Pinter Publishers.
- Furlan, M., Lima, P. A. B., Paão Junior, G. D., Mariano, E. B., & Pires, S. M. M. (2025). Proposing a composite index and maturity model for urban sustainability in the Brazilian context: A machine learning and data envelopment analysis approach. *Sustainable Development*, 33(1), 251–269. <https://doi.org/10.1002/sd.3120>
- Guan, J., & Chen, K. (2012). Modeling the relative efficiency of national innovation systems. *Research Policy*, 41(1), 102–115. <https://doi.org/10.1016/j.respol.2011.07.001>
- Hamid, S., & Wang, K. (2024). Are emerging BRICST economies greening? An empirical analysis from green innovation efficiency perspective. *Clean Technologies and Environmental Policy*, 26(2), 533–550. <https://doi.org/10.1007/s10098-023-02622-z>

- Kehinde, T. O., Akpan, J., Orisaremi, K. K., Olanrewaju, O. A., Anyebe, D. I., & Kareem, M. K. (2026). Three decades of DEA-machine learning integration (1996–2025): A bibliometric analysis, science mapping, and state-of-the-art review. *Machine Learning with Applications*, 24, Article 100868. <https://doi.org/10.1016/j.mlwa.2026.100868>
- Khan, R. Z., Razak, L. A., & Premaratne, G. (2025). Green growth and sustainability: A systematic literature review on theories, measures and future directions. *Cleaner and Responsible Consumption*, 17, Article 100274. <https://doi.org/10.1016/j.clrc.2025.100274>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
- Lundvall, B.-Å. (Ed.). (2010). *National systems of innovation: Toward a theory of innovation and interactive learning* (Rev. ed.). Anthem Press. <https://doi.org/10.7135/UPO9781843318903>
- Lunny, C., Brennan, S. E., McDonald, S., & McKenzie, J. E. (2017). Toward a comprehensive evidence map of overview of systematic review methods: Paper 1—purpose, eligibility, search and data extraction. *Systematic Reviews*, 6(1), Article 231. <https://doi.org/10.1186/s13643-017-0617-1>
- Luo, H., Kamarudin, F., Soh, W., & Shan, Z. (2026). Fulfilment efficiency, AI capability, and cross-border e-commerce development in China: Complementarities, regional heterogeneity, and resource-saving potential. *Sustainability*, 18(3), Article 1202. <https://doi.org/10.3390/su18031202>
- Narayanan, E., Binti Ismail, W. R., & Bin Mustafa, Z. (2022). A data-envelopment analysis-based systematic review of the literature on innovation performance. *Heliyon*, 8(12), Article e11925. <https://doi.org/10.1016/j.heliyon.2022.e11925>
- Ndicu, S., Ngui, D., & Barasa, L. (2024). Technological catch-up, innovation, and productivity analysis of national innovation systems in developing countries in Africa 2010–2018. *Journal of the Knowledge Economy*, 15(2), 7941–7967. <https://doi.org/10.1007/s13132-023-01327-4>
- Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Harvard University Press.
- Orynych, O., Matijošius, J., Raymundo, H., Dos Reis, J. G. M., Ruchała, P., & Świć, A. (2026). Energy and emission disutilities of transport modes under transport innovation in the European Union. *Energies*, 19(5), Article 1346. <https://doi.org/10.3390/en19051346>
- Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5, Part 2), S71–S102. <https://doi.org/10.1086/261725>
- Shi, J., Mei, J., Zhu, L., & Wang, Y. (2024). Estimating the innovation efficiency of the artificial intelligence industry in China based on the three-stage DEA model. *IEEE Transactions on Engineering Management*, 71, 9217–9228. <https://doi.org/10.1109/TEM.2023.3323292>
- Simar, L., & Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, 44(1), 49–61. <https://doi.org/10.1287/mnsc.44.1.49>
- Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1), 31–64. <https://doi.org/10.1016/j.jeconom.2005.07.009>
- Sun, L., Wang, L., Jiang, Q., & Zhao, Z. (2026). AI-driven efficiency: Artificial intelligence patents and operational performance in China's fisheries industry. *Aquaculture*, 621, Article 743995. <https://doi.org/10.1016/j.aquaculture.2026.743995>
- Tang, K., Qiu, Y., & Zhou, D. (2020). Does command-and-control regulation promote green innovation performance? Evidence from China's industrial enterprises. *Science of The Total Environment*, 712, Article 136362. <https://doi.org/10.1016/j.scitotenv.2019.136362>
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3), 207–222. <https://doi.org/10.1111/1467-8551.00375>
- Wang, S., Huang, X., Xia, M., & Shi, X. (2024). Does artificial intelligence promote firms' innovation efficiency: Evidence from the robot application. *Journal of the Knowledge Economy*, 15(4), 16373–16394. <https://doi.org/10.1007/s13132-023-01707-w>
- Zeng, J., Ribeiro-Soriano, D., & Ren, J. (2021). Innovation efficiency: A bibliometric review and future research agenda. *Asia Pacific Business Review*, 27(2), 209–228. <https://doi.org/10.1080/13602381.2021.1858591>
- Zhou, M., Cao, Y., & Huang, L. (2026). Digital economy, industrial structure optimization, and agricultural green development efficiency: A double machine learning causal analysis. *Australian Economic Papers*, 65(1), 49–58. <https://doi.org/10.1111/1467-8454.70009>