Research Article

Vinh V. Le, Nguyen Quang Lich, Thi Hai Le Nguyen and Hong Giang Nguyen*

An Integrated Quantitative and Machine Learning Approach to Examining Scholarly Publishing Barriers in Vietnam

https://doi.org/10.1515/opis-2025-0024
Received May 27, 2025; accepted August 8, 2025; published online May 26, 2026

Abstract: Getting published in globally indexed journals is still a major hurdle, especially for those scholars working in the still-developing academic systems. This study seeks to understand these barriers at the underlying level by surveying 120 researchers from Hue University and Times Higher Education (THE) and Quacquarelli Symonds (QS)-ranked universities. Deploying quantitative analysis and machine learning (ML) techniques, such as Random Forest (RF), Decision Tree (DT), and Gradient Boosting (GB), showed that financial burdens, institutional constraints, and limited collaboration opportunities were the three biggest obstacles. RF was found to have the highest classification accuracy among all models in predicting role and discipline barrier levels. Main recommendations highlight the calls for targeted interventions: subsidy for publication costs and enforce transparent article processing charges (APCs) waiver practices (*Q1. High Publication Fees*) Encourage transnational collaboration through exchange programs and conference sponsorship (*Q7. Language/Cultural barriers*) Role-sensitive funding models to address inequities within and between disciplines and academic ranks (*Q12. Lack of Metadata Knowledge*) Improving access to centralized data repositories and offering more technical/analytical training (*Q6. Insufficient Funding*, *Q10. Lack of Access to Trending Data*), as well as improving mentorship frameworks for early-career researchers (*Q5. Lack of International Networks*, *Q12. Lack of Metadata Knowledge*), are key structural fixes to inequitable disadvantages. Together, these strategies lower barriers while encouraging larger and more diverse participation in global higher education. The main purpose of the present study is to identify the problems encountered by the researchers in publishing their articles in the Web of Science (WoS) and Scopus-indexed journals. Above all, this study is to lay out how the major barriers, their intermingling and working opposition to one another, and what institutional action is needed to realize improved rates of successful publication. To do this, a combination of quantitative analysis and ML is used to help inform and tell a more robust narrative around the effect of the researchers' published work in high-impact publications. These quantitative methods and ML techniques are developed based on an empirical survey of 120 scholars from Hue University and globally ranked institutions. To collect data, a structured questionnaire on 12 factors affecting academic performance was used. Descriptive statistics, correlation matrix, ANOVA, and Structural Equation Modeling (SEM) were used to analyze the data. At the same time, RF, DT, and GB models were trained to classify and predict barrier levels. The biggest obstacles noted were the high cost of publishing in science, technology, engineering, and math (STEM) fields, the lack of global coordination and collaboration, and unequal access to resources. The RF model performed best in classification accuracy compared to the DT and GB models. Recommendations included subsidizing conference fees, developing mentorship programs, expanding access to industry data, and role-based funding. Together, these efforts would eliminate existing publication inequities.

***Corresponding author: Hong Giang Nguyen**, School of Engineering and Technology – Hue University, 49000 Hue City, Vietnam, E-mail: giangnh@hueuni.edu.vn. https://orcid.org/0009-0001-0862-9941
Vinh V. Le, PhD, Posts and Telecommunications Institute of Technology, 11300 Hanoi City, Vietnam, E-mail: vinhlv@ptit.edu.vn. https://orcid.org/0009-0003-9758-3408
Nguyen Quang Lich, PhD, Associate Professor **and Thi Hai Le Nguyen,** MBA, School of Engineering and Technology – Hue University, 49000 Hue City, Vietnam, E-mail: nguyenquanglich@hueuni.edu.vn (N.Q. Lich), nguyenthihaile@hueuni.edu.vn (T.H.L. Nguyen). https://orcid.org/0000-0001-5469-7227 (N.Q. Lich)

Novelty: This study was groundbreaking in ML with quantitative analysis methods of combination to proactively identify and address these publication barriers. These findings provide additional ammunition for our policy-makers, research administrators, and scholars, who all need to be armed with enough knowledge to work toward an equitable scholarly publishing environment.

Keywords: barrier factors; quantitative analysis; WoS/Scopus indexed; machine learning

1 Introduction

Technical publishing remains a key part of the scientific ecosystem, connecting the scientific community to society and gaining recognition for the contributions scholars make to humanity. However, despite its focus on disseminating knowledge and validating research, the publication and citation system remains subject to class bias, regional disparities, and herd instincts. Furthermore, these barriers particularly negatively impact on researchers from developing countries and lesser-known engineering schools, making their work appear to be overlooked, even when it is valuable. In response to these limitations, strategic mechanisms like metadata optimization have emerged as essential tools for increasing citation visibility and navigating the rigid standards of indexing platforms such as WoS and Scopus (Bornmann and Haunschild 2023).

Meanwhile, scholars working within resource-constrained settings often experience less access to funding, collaborative networks, and high-impact journals. Moreover, indexing systems that reward citation impact, international co-authorship, and journal prestige only serve to worsen these inequities (Samala et al. 2024). Consequently, this has the disastrous effect of excluding valuable research born from the Global South and other underrepresented contexts from academic publishing (Arunachalam et al. 2024; Buagayan et al. 2024).

In order to tackle these systemic barriers, careful and strategic solutions need to be implemented. Specifically, these solutions include promoting policies within academia, such as supporting publication, building equitable partnerships, improving research, training, and practices around academic principles, and sharing and developing a research culture rooted in integrity and critical thinking (Abayeva et al. 2024). As a result, by proactively identifying and addressing these barriers, the academic community can gradually make strides towards a more equitable and representative international publishing environment (Ahmad et al. 2022).

In the case of Vietnam, these barriers are often interconnected and interact with each other, forming a complex network of factors that affect the ability to publish internationally. Nevertheless, despite many efforts to improve research policies and investment in science and technology, the lack of systematic quantitative analysis to clearly identify the relative role of each barrier makes it difficult to develop effective support policies.

Within these contexts, barriers to academic research according to WoS and Scopus criteria consist of several factors as follows:

Q1 High Publication Fees:

Despite open-access publishing making research much more visible, this visibility comes at a cost, as APCs place a heavy financial burden on researchers. As a result, most high-impact or even moderately high-impact journals indexed in WoS and Scopus become inaccessible, with APCs often exceeding $3,000 per article, thereby excluding underfunded researchers from publishing in them. To illustrate this further, depending on the relative quality of the journals, Elsevier's APCs range from approximately $200 to $10,400. Moreover, a recent review found that the average APC across open-access journals was $906, with significant variation by discipline (Borrego 2023). Consequently, these prohibitive costs further widen inequities between campuses in affluent and less affluent regions (Jain et al. 2021).

Q2 Meeting Journal Standards:

The editorial policies of both Scopus and WoS indexed journals place a strong emphasis on methodological rigor, originality or novelty, and the potential for a high citation count (Piran and Tran 2024). Consequently, for the majority of early-career researchers, this leads to a very discouraging rejection rate for reasons including insufficient detailed statistical analysis, inadequate thoroughness in the literature review, or lack of alignment with the journal's scope. In fact, the top reasons for manuscript rejection are often a lack of originality, weak methodological quality, or incompatibility with the journal's aims and scope. In addition, not good study design, not being transparent about findings, and not following journals' standards are common rejection reasons (Bhende et al. 2024).

Q3 Limited Collaboration:

International and interinstitutional collaboration raises the visibility of WoS/Scopus-indexed journals (Miao et al. 2024). However, many public health and transportation professionals still point to long-standing barriers to successful and meaningful collaboration (Deli et al. 2021). Moreover, scholars from resource-barriered developing nations may face even greater obstacles in building such international networks (Gaye et al. 2024). So it is in this context that international partnerships matter. Now universally assumed to be off the charts in terms of its power to increase citations and scholarly impact, research collaboration has become the indisputable if not unimpeachable research variable. While efforts are growing to bridge global partnerships, nonetheless, structural inequities still act as one of the largest barriers to equitable, collaborative spaces. Consequently, strengthening inclusive collaborative infrastructures is key not only for building more impactful research but also for tackling global inequities in resources and capacity through funded global partnerships (Faure et al. 2021).

Q4 Lack of Access to Data:

Transparency and reproducibility are crucial in Scopus-indexed journals. However, the inability to access proprietary datasets makes it difficult to conduct vital validation and replication studies. In response, open science movements are working to change these practices. Nevertheless, data-sharing practices vary drastically across industries (Adams et al. 2018; Fassnacht et al. 2024). As a result, the cumulative inequities regarding access to data create significant barriers to reproducibility, and data-sharing practices are certainly not standardized across the field. Consequently, without standardized data-sharing protocols, reproducibility is further hampered, thereby hindering greater research and innovation advancement particularly in historically underfunded fields. Ultimately, unclear data-sharing policies continue to be among the biggest obstacles to improving scientific rigor and collaboration across and within disciplines (Cunha-Oliveira et al. 2024).

Q5 Lack of International Networks:

Creative output developed in collaboration with international authors greatly enhances visibility and reach. Nonetheless, many early-career researchers remain excluded from such global partnerships due to persistent geographic inequalities, visa restrictions, and limited funding opportunities (Yao 2021). Additionally, co-authorship networks, whether seen as evidence of fragmentation or growing inequality, tend to reflect a hierarchical structure dominated by a central, high-prestige core, while marginalizing numerous lower-prestige or peripheral groups. This dynamic risks reinforcing existing disparities within the research landscape (Khan et al. 2024). Confronting these structural obstacles demands urgent policy changes and the construction of inclusive infrastructure to facilitate equitable and persistent global research collaboration.

Q6 Insufficient Funding:

Research funding is a critical factor in determining both the academic impact and public visibility of scientific work (Hussinger and Carvalho 2022). Arguably, the most immediate danger to equity in global science is the glaring disparity in funding opportunities between countries and regions. This ongoing gap restricts the scope of researchers, especially from low- and middle-income countries, to participate in international efforts and be acknowledged by the global community (Petersen 2021). Indeed, small research grants meaningfully increase the long-term research output of university faculty (Ou et al. 2024). As a result, for many scholars in low- and middle-income countries, such funding imbalances present substantial obstacles to participating in global collaborations and gaining broader academic and intellectual recognition (Ou et al. 2024).

Q7 Language/Cultural barriers:

English language supremacy of WoS- and Scopus-indexed journals spells huge obstacles for non-native speakers. Linguistic bias in peer review, combined with the high cost of professional language editing services, greatly limits the chances of otherwise high-quality research being accepted for publication (Silver et al. 2023). Many non-Anglophone researchers experience repeated rejections not due to scientific flaws, but because of perceived deficiencies in language use (Mirhosseini et al. 2024). These barriers often result in delayed publication, extensive revision cycles, or outright rejection based on language proficiency rather than scholarly merit (Tenzer et al. 2021). Researchers from under-resourced regions are doubly disadvantaged, though, as academic publishing grows more and more English-centric, the lingua franca of the field. This cycle compounds inequality in the international research community. Moreover, recent studies show that language barriers affect not only submission and peer review but also citation practices and the overall visibility of published work (Marsden and Morgan-Short 2023). To combat these inequities, inclusive editorial policies, accessible or subsidized language support, and increased awareness of implicit language bias among editors and reviewers are urgently needed.

Q8 Limited Institutional Support:

Universities with old research offices have a full suite of support services–grant-writing, statisticians, publication gurus–that make their faculty research more successful. By contrast, without this infrastructure, organizations rarely support an environment for long-term research productivity in the organization. Experience and research strongly suggest that healthy institutional research ecosystems correlate strongly with high levels of output (Owan et al. 2024). Although such support structures impact individual faculty success directly, they are just as crucial for cultivating and maintaining research-rich environments that can maintain long-term academic productivity (Bashir et al. 2022). Institutional characteristics, like open research repositories and mentorship programs, also buffer some of the typical barriers to research engagement, increasing research involvement. At the end of the day, universities willing to make serious investments in both human capital and intangible research infrastructure, along with a developmental approach to academic culture that emphasizes scholarly excellence, will be best equipped to increase their global research presence and push open science forward.

Q9 Limited Funding for Research Topics:

Research relying on WoS data, particularly from those fields left out by its author-centric lens, inequitable funding distribution, such as the over 80 % of funding going to STEM fields versus the humanities and social sciences, which still are fields of study deserving of infinitely more funding. The ideological dispositions and preferences of fund arbiters can further bias research priorities by forcing researchers to shape their work around the aims of top funders, usually at the fiscal expense of a much broader intellectual plurality. This disparity, in turn, limits the advancement and societal impact of the humanities and social sciences by continually undervaluing them (Heyard and Hottenrott 2021). And just as funding landscapes have historically shaped the geography of innovation, the compounding and cumulative nature of binding inequities in funding present also as inequalities in institutional research capacity across geographies, resulting in persistent and embedded inequalities in research capacity across geographic boundaries (Graves Jr et al. 2022).

Q10 Lack of Access to Trending Data:

As with all academic pursuits, research's disruptive potential is a matter of timing, especially in young, fast-moving sectors that are extremely dynamic and innovative. However, timely access to knowledge continues to be an issue. In much of the Global South; however, paywalled databases and access-restricted journals make it difficult for researchers to keep up with recent work. While open-access efforts have done a lot to close this gap by granting free access to the scholarly literature, great inequality remains. Multiple studies have emphasized the institutional and financial obstacles that prevent academics from reading recent work or publishing in high-impact venues (Cortés-Sánchez et al. 2024). The irritation numerous academics feel when these paywalls get in their way is compounded by the divisive rise of Sci-Hub-like sites that provide illicit entry to millions of articles. Although digital tools like Digital Object Identifiers (DOIs) have greatly increased the discoverability of work, the benefits are still unevenly distributed across disciplines and regions, deepening existing global disparities in research participation and impact (Turki et al. 2023).

Q11 Insufficient Collaboration with Experts:

Engaging with more experienced, senior researchers who are later in their careers is very important for junior researchers, particularly those interested in fast-tracking the improvement and expedited delivery of high-caliber science publications. Early career scientists who want to broaden and hone their research should actively pursue opportunities to be connected to and mentored by these more experienced scholars, for these relationships can provide a rich resource to help young scholars navigate the sometimes turbulent seas of academic writing, the sometimes harsh peer review process, and the inscrutable world of scientific research itself (Silver et al. 2023). Cultivating this next generation of early career researchers requires deep, integrated mentorship that helps them to imagine new and different ways of conceptualizing innovative, novel ideas, producing better research designs, and disseminating their findings in a broader, more impactful way. In addition, co-authors are almost always well-known, well-respected, and widely published researchers from across the continent; their very presence on papers helps these emerging scholars build thick academic portfolios, which then catapults them into the highly sought-after professional networks, along with greater fidelity and deference from academic gatekeepers.

Q12 Lack of Metadata Knowledge:

And for research outputs to be discoverable and accessible, strong metadata practices are key. Metadata, which includes titles, abstracts, keywords, and citation formats, is the digital fingerprint of scholarly work, connecting it to worldwide indexing services such as WoS and Scopus. When these components are carefully and strategically crafted, they significantly improve a publication's discoverability, meaning its chances of being found by other researchers and, therefore, cited and making an academic impression. Yet its importance for research publishing notwithstanding, effective metadata creation is a skill gap for many academics–especially those at under-resourced institutions or at the beginning of their careers. The absence of adequate training or institutional support in this area often leads to suboptimal metadata quality, which not only diminishes the reach of individual research outputs but also contributes to systemic inequities in global knowledge exchange (Aydın Çolak and Eroğlu 2025).

Due to the above-mentioned substantial obstacles, it is necessary to process the accumulated data computationally. In particular, one must integrate quantitative analysis tools with ML approaches in order to investigate and simulate elements influencing Vietnamese researchers' international publishability. Moreover, the ML approach not only supports processing large and complex data sets but also helps to uncover nonlinear relationship patterns between input and output variables that are frequently missed by traditional statistical methods. Accordingly, the main ML models utilized in this study are DT, RF, and GB. The choice of these models is grounded in three basic reasons. First, these models are proficient at modeling nonlinear relationships and effectively managing complex structured data. Specifically, DT supplies a simple and intuitive logic platform for classification and causal explanation, which aligns with the original goal of discovering the key factors leading to publication or non-publication. Second, RF enhances the accuracy and stability of DT by combining multiple trees and using majority voting techniques. This, in turn, helps reduce overfitting and increase the robustness of the model on real-world datasets. Third, GB is a computationally efficient model, which is especially suitable for imbalanced datasets (e.g., the number of internationally published researchers is much smaller than the unpublished group). Furthermore, this model is often highly regarded for its accuracy in prediction and ability to learn deeply from small data samples. In particular, all three models are well integrated with the SHAP (SHapley Additive exPlanations) method, a vital tool in interpreting and justifying the outputs of ML models.

This study intends to use an integrated approach of ML and quantitative analysis to identify and evaluate the key obstacles faced by Vietnamese researchers in publishing in high-impact scientific journals such as WoS and Scopus. By interpreting survey data with the assistance of ML models, the study doesn't just pinpoint barriers; it elucidates their consequences, connections, and interplay. The findings attempt to offer actionable insights in order to suggest institutional fixes to increase the rate of successful international publications and Vietnam's academic prestige on the global stage.

The study steps include as bellows: Section 2 mentions literature review. Section 3 presents the methodology, providing the experimental design and modeling techniques. Section 4 displays the results and discusses the experimental performance and key impact factors. Finally, Section 5 provides concluding remarks, summarizing the main results and their implications for scholars.

2 Literature Review

Numerous initiatives have been first to recalibrate the difficulty that researchers face when attempting to get papers published in WoS and Scopus-indexed journals. As a result, these inquiries leverage a diverse range of ML techniques in tandem with quantitative analysis to resolve the barriers, such as below:

Hazra et al. (2020) took a data-driven approach to forecasting journal submission outcomes by training classifiers of authors based on their probability of having their work accepted using authors' hierarchical structure with the performance of ML based classification techniques such as RF and GB algorithms. Their results show that authors with higher acceptance rates are indeed more successful, more so because they have better academic profiles, such as a larger collaborative network, larger citation rate, and h-index. Additionally, authors from these backgrounds are further still forced into reviews that are less rigorous and/or critical through the peer review system. The predictive models performed well in classifying manuscripts into the accepted and rejected classes. Though these models were originally developed to explain model predictions, these feature importance explanations were still very useful in explaining the distinguishing structural and academic-related variables that increased the likelihood of a manuscript being accepted.

Alohali et al. (2022) used DT Regression to predict the eventual total cumulative citation counts for research articles across the thesis segment of otology. The barrier provided a baseline of 10 ML models, from a basic Support Vector Regression (SVR) through to an Artificial Intelligence (AI) enabled Artificial Neural Network, that have each been shown to effectively model human choices, against a dataset that encompassed journal impact scores, author reputations, and publication metadata. The highest R^2 determination coefficient value (R^2 = 0.87) indicates an extremely high predictive capability; hence, among others, the DT model performed better than other models.

Zhang and Wu (2024) proposed early citation data as predictors to estimate the future paper's citation impact, and applied this work to non-English papers from all research fields using different ML models, RF, GB, and SVR. Because of the strength of ML models, the study generated the best-performing, domain-specific ensemble models, which outperformed general-purpose models, yielding the greatest (86 %) overall forecasting accuracy in each study area. This all greatly illustrates the gist of our argument here, reiterating the broader need for deeper contextual and field-specific modeling to make citation impact predictions as precise, accurate, and useful.

Hoang and Dang (2021) called for the development of an analytical framework to identify the barriers that local researchers need to navigate in order to engage with the international academic community as a whole. Three major barriers include financial constraints, lack of institutional support/guidance, and lack of English language proficiency. Ultimately, 65 % of respondents to our survey perceived language to be the biggest hurdle, followed by 54 % not having qualified research mentors, and 49 % not being able to afford attending international conferences or paying publication fees. These burdens show to large effect on international publishing rates of early-career scholars in Vietnam.

Nicholas et al. (2024) examined how early-career researchers (ECRs) remain burdened by economic barriers, especially when facing the increasing push for open access publishing. ECRs, broadly defined as those researchers within 10 years of their doctor of philosophy, are suffering a triple whammy of financial precarity, institutional disillusionment, and a crisis of available advancement opportunities. The research illustrates the ways in which APC-related concerns and pressures affect ECRs' publishing practices, creating a two-tiered academic environment where the ability to pay positions scholars to share their research more widely.

These studies focus on publication success and citation impact but underscore the numerous hurdles to publishing in elite venues. Furthermore, these problems should give researchers a sense of the tactics they can employ to satisfy journals.

3 Methodology

3.1 Quantitative Analysis

3.1.1 Descriptive Statistics (Exploratory Data Analysis – EDA)

This task draws on the principle of interpreting and understanding what has been gathered through the creation of wide-reaching, overarching key stats to identify broad trends and trends. Measurement of central tendency (mean and standard deviation), and frequency distributions were deployed for exploratory data analysis to describe the dataset (Tennant 2018). Univariate and bivariate statistical techniques supply a first look at central tendency, variability, and distribution, laying the groundwork for more complex inferential analyses to next steps.

3.1.2 Pearson Correlation Regression Analysis

Applying Pearson's correlation coefficient, a statistical measure that expresses the strength and direction of linear relationships between pairs of variables. By analyzing correlation coefficients, this method supports determining whether specific research barriers are interrelated, thereby offering insights into potential underlying patterns (Dufera et al. 2023).

3.1.3 One-Way ANOVA (Comparative Statistical Test/Analysis)

This analysis skill makes it possible to an examination of mean values between three or more groups. One-way ANOVA tests whether the differences that we see between our groups are statistically significant using information on variance within groups and between groups. The F-statistic and associated p-values are calculated to evaluate if observed differences are the result of random variation or represent significant differences (Aksnes and Sivertsen 2023). This methodological approach proves an effective strategy for engaging new voices, and it shines an important light on how varying professional and disciplinary backgrounds shape perspectives on institutional and systemic barriers to research publication.

3.1.4 Structural Equation Modeling

SEM is used to analyze complex interrelationships between sets of variables related to research publication obstacles. While needing additional practical and accessible implementation, this approach enables the elucidation of causal pathways and latent interactions, providing greater insight into how various barriers impact one another. To establish the validity of the model, indices of model goodness of fit were inspected, including the Chi-Square Test, Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA) (Whittaker and Schumacker 2022). This application improves the understanding of structural dependencies in research obstacles, creating a strong foundation for predictive modeling.

3.2 Structural Machine Learning Models

3.2.1 Random Forest

RF is a bagging ensemble learning algorithm that builds many DTs on training sets constructed through bootstrapping and combines their outcomes through majority voting for classification or averaging for regression. Its robustness is great, because it does not overfit and generalizes much better due to the aggregation of uncorrelated DTs. RF has become one of the most popular ML algorithms in applied fields, from finance to education and artificial intelligence, because of its high predictive accuracy and high interpretability (Oluchukwu Njoku et al. 2023).

$$\hat{y} = \frac{1}{T} \sum_{1}^{t} f_t (x) \tag{1}$$

where, $\hat{y}$, T, and $f_t(x)$ are final forecasting, total number of DTs, and forecasting from the t^{th} DT for input x, respectively.

3.2.2 Decision Tree

DTs are powerful, yet intuitive and interpretable ML models commonly used for classification and regression. They work by recursively splitting the dataset into subsets based on input feature values, with the goal of minimizing a cost function like Gini impurity. Its simplicity and transparency make it especially useful for structured data analysis and initial exploratory modeling (Chen et al. 2025).

$$\text{Gini} = 1 - \sum p_i^2 \tag{2}$$

where, p_i is proportion of samples of class i at a node

3.2.3 Gradient Boosting

GB sequentially builds models, each new model trying to fix the issues of the models before it. By applying gradient descent optimization, it iteratively minimizes a chosen loss function, which makes it very powerful for classification and regression problems. Its strength as a data wrapper for structured datasets has continued to spread its application across myriad domains, including geosciences and broad-scale predictive modeling, fueled by the eventual release of public domain API keyless access (Rizkallah 2025).

$$f(x) = f_0 (x) + \eta \sum_{m=1}^{M} f_m (x) \tag{3}$$

where, $f(x)$ is final model prediction, $f_0(x)$ is initial prediction (commonly the mean of the target variable), η is learning rate, M is total number of boosting iterations $f_0(x)$: Initial, and $f_m(x)$ is output of the mth weak learner.

Overall, in this supervised ML process, such as first import and preprocess the original dataset. Since categorical features and target labels are used, Label encoder is used to encode the categorical features and target labels. This dataset is subsequently organized into training and testing sets, which allows for evaluation on how well the model will perform. Then training RF, DT, and GB models. Subsequent to training, each model is then evaluated with metrics of accuracy, a confusion matrix, and a classification report. The output of the variable importance of all models combined, used to determine important variables. Lastly, policy recommendations are suggested, informed by the most powerful features as measured by feature importance. At the same time, the data in Table 1 illustrates an overview of three commonly used ML models in terms of structure, operation, configuration, and symmetry. It discusses how each model creates decision logic.

Table 1: Optimum performance summary table.

Aspect	Decision tree classifier	Random forest classifier	Gradient boosting classifier
Model type	Single decision tree	Ensemble of multiple decision trees (bagging)	Ensemble of trees in sequential boosting manner
Operating principle	Splits data at each node using optimal criteria (gini) to reduce impurity	Trains each tree on a random subset of data, aggregates results via majority voting	Each tree learns from the residual errors of the previous one, minimizing a loss function step by step
Code configuration	Decision tree classifier (random_state = 42)	Random forest classifier (random_-state = 42, *n*_estimators = 100)	Gradient boosting classifier (random_-state = 42, *n*_estimators = 100)
Structural symmetry	Root node (decision) → Sub-decisions → Leaves	Parallel trees → Individual predictions → Majority vote	Tree1 → Tree2 (fix Tree1) → Tree3 (fix Tree2) → … → Tree100

3.3 Evaluation Metrics

In survey analysis, especially when using classification models (e.g., predicting a respondent's primary research area or behavior category), evaluation metrics like Kappa, Precision, Recall, and F1-score help performing in assigning the correct labels to survey responses.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

where, TP (True Positives), and FN (False Negatives) call actual cracks correctly predicted as cracks, and actual cracks incorrectly predicted as non-cracks.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

where, FP (False Positives) is Non-cracks incorrectly predicted as cracks.

$$\text{F1 - score} = 2x\,\frac{\text{Precision} + \text{Recall}}{\text{Precision} \times \text{Recall}} \tag{6}$$

where, F1-score is the harmonic mean of precision and recall.

$$\text{Kappa} = \frac{P_0 + P_e}{1 - P_e} \tag{7}$$

where, P_0: Observed accuracy or agreement between predictions and true labels, P_e: Expected agreement by chance.

3.4 Structure of Survey Questionnaire

The final survey questionnaire sought to determine what primary barriers scholars at Hue University and other Vietnamese universities ranked within QS and THE university rankings were facing. It was aimed specifically at these 120 researchers (whose at least two of papers were published on WoS or Scopus), and they were divided into three parts: Position/Role included lecturers, researchers, and associate professors; Years of Experience was divided less than 5 years, 6–10 years, 11–20 years, and more than 20 years; Primary Research Area (PRA) consisted of Agriculture, Biology & Environmental Sciences (ABES), Artificial Intelligence and Data Science (AIDS), Engineering, Computing & Technology (ECT), Environmental and Climate Studies (ECS), Social And Behavioral Sciences (SBS), Electronics & Telecommunications Collection (ETC). 12 factors are mentioned in the introduction section. The factors were measured using a five-point Likert scale, ranging from 'strongly disagree' to 'strongly agree'.

3.5 Sampler Selection

Determining a proper sample size is an important part of designing a survey to make sure the results reflect the larger population. Beyond the sample size, if the sample is small, there is a risk that the findings won't be dependable, painting a picture in which conclusions are drawn from unreliable data. Too large a sample is both unnecessary and expensive, taking up more time and resources than necessary. To find the minimum size of sample size, researchers usually borrow Slovin's formula method, which is a widely practiced method that helps them determine how many observations are needed, depending on the total population and the acceptable margin of error.

Slovin's formula can be stated mathematically as:

$$n = \frac{N}{1 + N^{*}e^{2}} \tag{8}$$

where, n explains the sampling design, i.e., the number of individuals drawn for the survey; N denotes the observed population size; e is the margin of error, or how much error the study can tolerate, or how precise the study needs to be.

Furthermore, with respect to the ML sample, the collected data is separated into two stages: the training and testing phases use 96 samples (equivalent 80 %), 24 samples (equivalent to 20 %), respectively. In addition, a construct for in-depth analysis serves as the foundation for the usage of the PRA variable.

This formula offers researchers the freedom to set their own sample size by deciding how much uncertainty they're willing to accept in their findings. An increased margin of error will lead to a smaller sample size, but the trade-off is the higher uncertainty level of the surfaced results, and vice versa.

4 Results and Discussions

Publishing research in WoS and Scopus-indexed journals entails a number of formidable barriers. To examine these barriers systematically, a simple random sampling procedure was employed to select participants from a larger population of 600 scholars who had published at least two journal articles indexed in WoS or Scopus over the past two years. The final sample was evenly divided, with half of the participants affiliated with Hue University and the other half drawn from universities listed in the 2025 Times Higher Education (THE) and QS World University Rankings.

To determine an appropriate sample size, Slovin's formula was applied. Based on this calculation, a target of 120 respondents yielded a margin of error of 8.17 % $\left(e = 8.17\,\% = \sqrt{\frac{600-120}{600^{*}120}}\right)$. This margin of error was deemed acceptable for ensuring broad representativeness across the original pool. Importantly, the stratified sampling approach ensured proportional representation between scholars from Hue University and those affiliated with globally ranked institutions.

As depicted in Figure 1, the reliability of survey items was assessed using Cronbach's Alpha, with an overall value of 0.68 indicating moderate internal consistency. In Figure 2, twelve histograms display survey responses regarding the most pressing barriers in academic publishing. Prominent barriers included excessive publication fees, limited funding availability, and insufficient interinstitutional collaboration. The observed bimodal distributions suggest varying levels of respondent awareness, emphasizing the need for targeted capacity-building interventions. These visualizations not only highlight the scale and nature of key barriers but also serve as a foundation for subsequent policy development.

Furthermore, Figure 3 presents descriptive statistics (minimum, maximum, mean, and standard deviation) for each of the twelve survey items (Q1–Q12). Notably, the minimum scores (denoted by red dots) cluster around 3.0, while maximum scores (green dots) consistently reach the upper limit of 5.0. Mean scores range from 3.76 (Q12) to 3.94 (Q2), reflecting generally favorable responses. Questions such as Q1, Q3, Q5, Q6, and Q9 all exhibit mean values above 3.85, indicating strong agreement among respondents. The standard deviations, ranging from

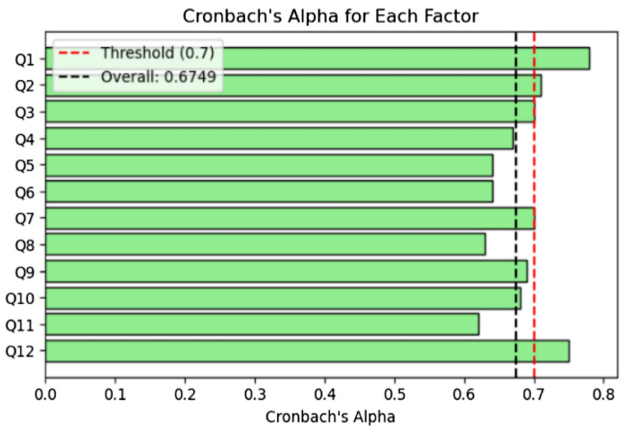

Figure 1: Cronbach's Alpha of 12 factors.

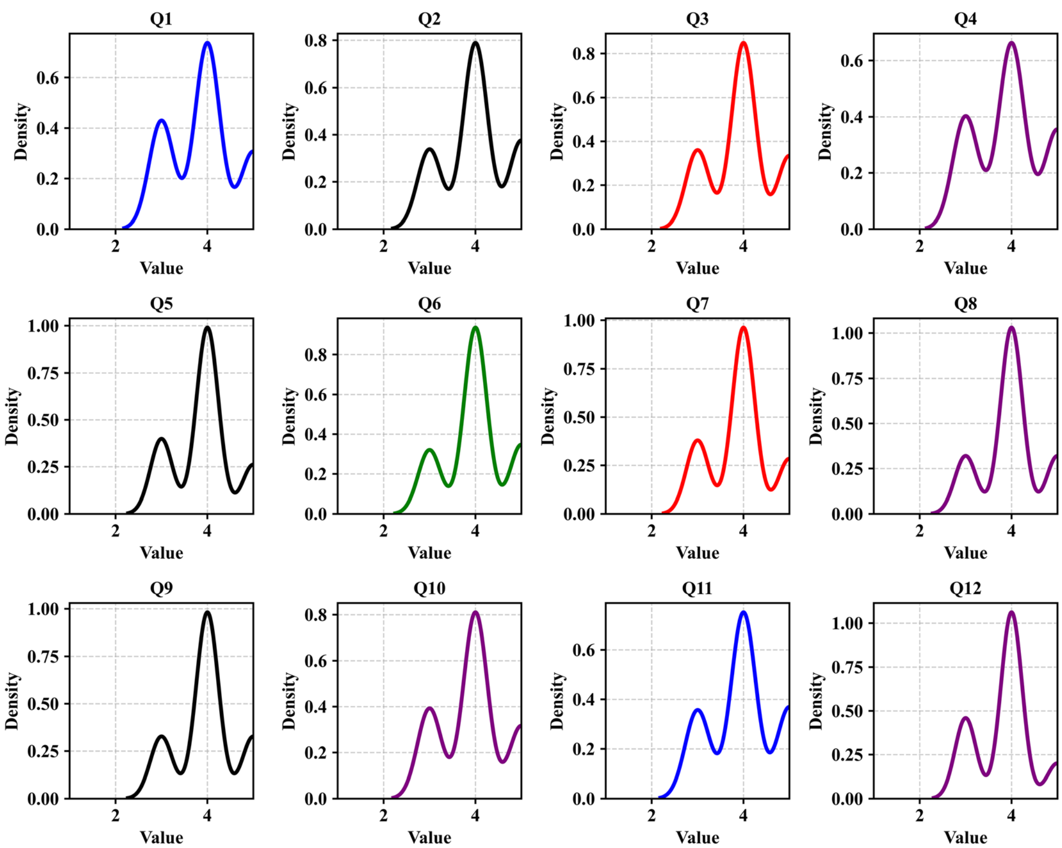

Figure 2: Histograms for 12 factors.

±0.42 (Q6) to ±0.60 (Q4), suggest relatively low variability overall, although higher dispersion in Q4 and Q10 points to more diverse respondent perspectives on those items.

In relation to inter-item relationships, Figure 4 displays the correlation matrix for Q1–Q12. Here, the intensity of red shading represents the strength of the correlation, with darker hues denoting stronger positive associations. While most correlations were modest (ranging from 0.2 to 0.6), certain item pairs exhibited notable associations. For instance, Q2 and Q7 showed Figure 4 highest correlation (r = 0.59), suggesting that those who reported barriers in collaboration also lacked global research networks. Moderate correlations between Q5 and Q6 (r = 0.49), and between Q6 and Q7 (r = 0.45), further underscore thematic interlinkages in the barriers

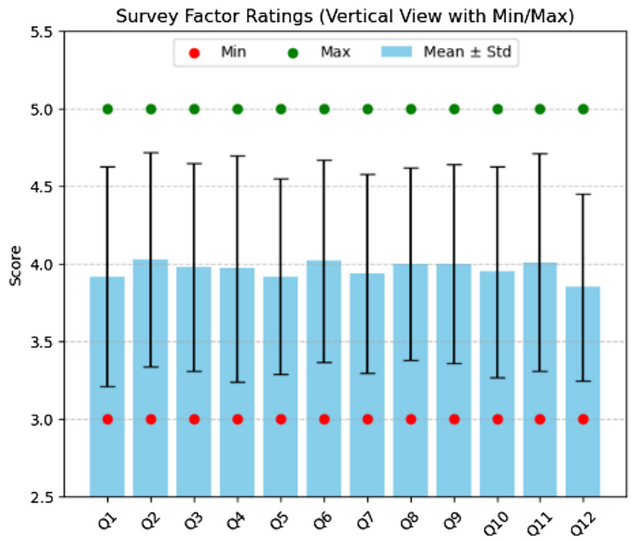

Figure 3: Descriptive statistic.

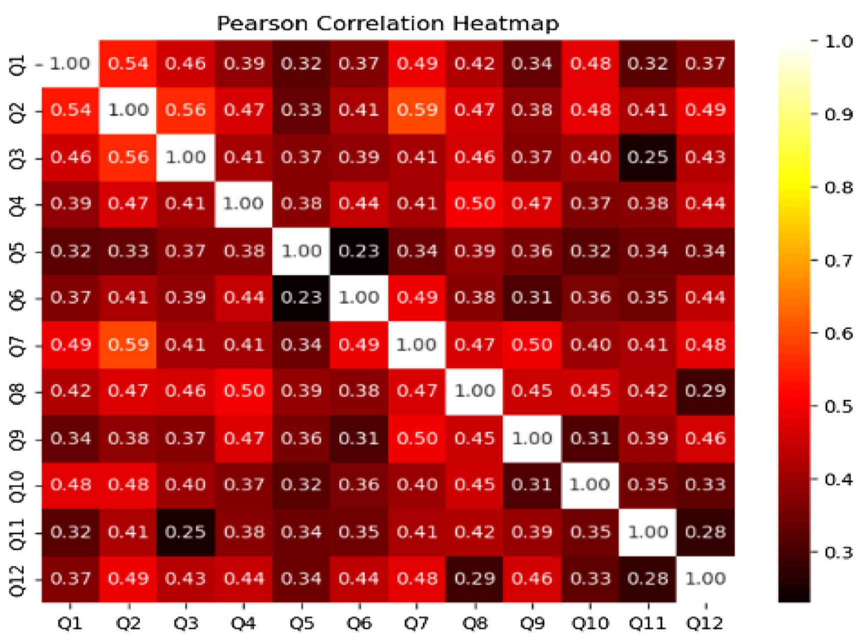

Figure 4: Pearson correlation among the factors.

experienced. Taken together, these findings justify the need for more integrated support mechanisms, such as enhanced networking opportunities and collaborative funding platforms.

Transitioning to structural analysis, Figure 5 visualizes the SEM that maps the interdependencies among key barriers. The model's goodness-of-fit indices, which include Chi-Square (40.7), *p*-value (0.0), Comparative Fit Index (CFI = 0.95), and Root Mean Square Error of Approximation (RMSEA = 0.01), demonstrate a strong, statistically valid model. Key relationships include the strong links between Q5 and both Q4 (0.69) and Q6 (0.66), indicating that poor collaboration undermines funding access and data availability. Similarly, Q2 (0.94) and Q3 (0.84) are tightly linked to Q1, highlighting the cost burdens associated with compliance and international visibility. Other critical pathways involve Q9 influencing Q4 (0.62) and Q10 influencing Q8 (0.88), while Q7 connects to Q1 (0.61), possibly reflecting costs related to translation and editing services. These insights point to the necessity of multifaceted institutional responses, such as promoting collaboration (Q5, Q8), supporting international outreach (Q2), and improving access to data and technical tools (Q6, Q10).

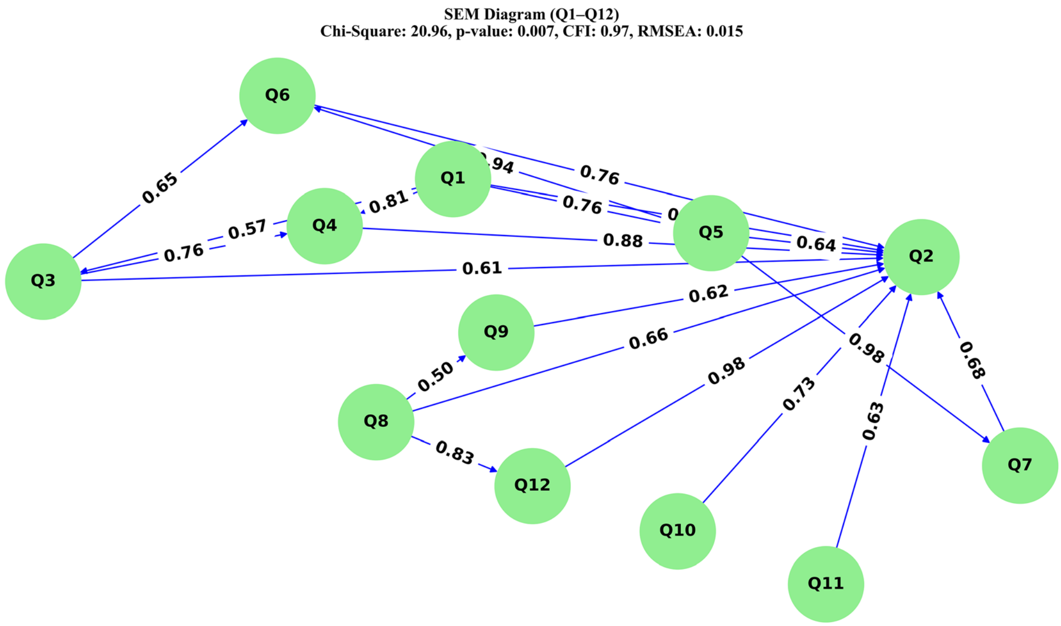

Figure 5: SEM.

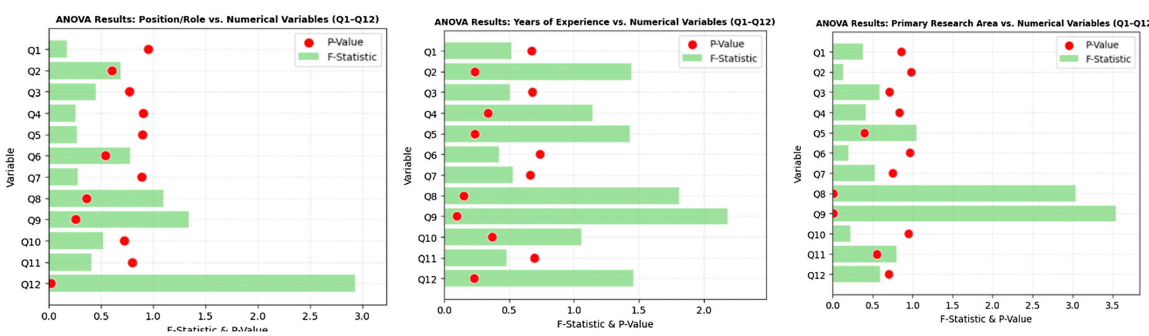

Figure 6: ANOVA analyze a different categorical factors with position/role, years of experience, and primary research area.

Expanding further, Figure 6 provides ANOVA results that investigate the effects of Position/Role, Years of Experience, and PRA on perceived barriers. In the first panel, Position/Role reveals two statistically significant differences: Q6 ($p = 0.02$) and Q12 ($p = 0.05$), suggesting that roles influence perceptions of data access and funding distribution. In the second panel, Years of Experience affects Q5 ($p = 0.04$) and again Q12 ($p = 0.04$), indicating that perceptions of collaboration and funding shift with career stage. In the final panel, PRA yields significant effects for three barriers, suggesting disciplinary disparities in how institutional and funding constraints are perceived. These findings underscore the importance of nuanced, role-specific, and discipline-sensitive policies to mitigate inequities in the publishing landscape.

Although traditional statistical approaches such as SEM and ANOVA provided insight into relationships and group-level differences, they lacked predictive depth. Therefore, to address this limitation and anticipate future barriers, the study also employed ML models to forecast and assess key predictors.

Figure 7 presents the test-phase predictions across 402 sample indices. RF predictions (green line) closely align with actual class values (black line), particularly around indices 5, 17, and 34. In contrast, DT (orange dashed line) and GB (blue dashed line) display larger deviations, with DT misclassifying several labels between indices 10–15 and GB performing erratically between indices 30–40. These results suggest RF's superior consistency in label prediction.

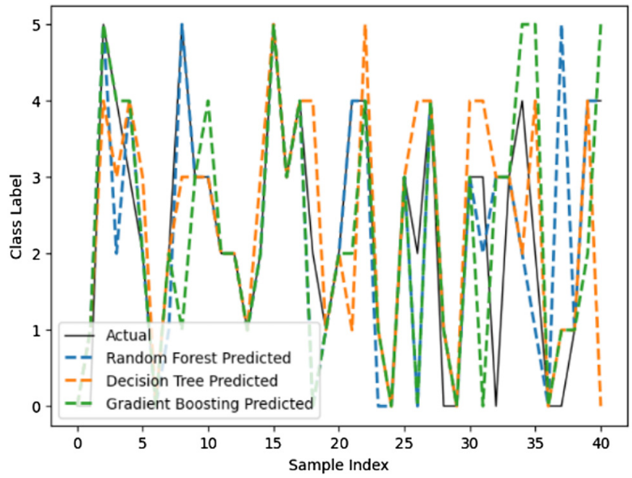

Figure 7: Testing phase: actual versus predicted.

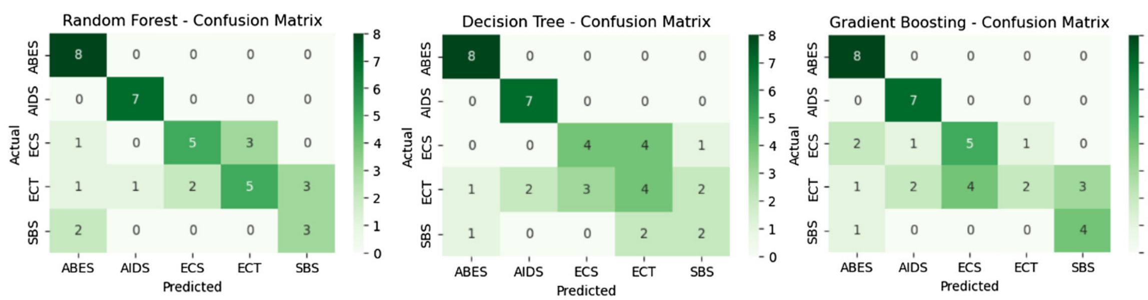

Figure 8: Confusion matrix.

Figure 8 displays the confusion matrices across five academic categories (ABES, AIDS, ECS, ECT, and SBS). RF exhibited the best performance, achieving perfect classification for ABES (8/8) and AIDS (7/7), and correctly identifying 5 of 6 ECS cases. However, it struggled with ECT (2/6 correct) and SBS (3/6 correct). DT also achieved perfect accuracy in ABES and AIDS but dropped to 4/6 in ECS. GB matched RF's accuracy in ABES and AIDS but underperformed in ECT and SBS. Misclassification patterns, particularly in ECT and SBS, may stem from limited training data and overlapping feature semantics. These insights highlight the need for better feature engineering and more balanced datasets. In terms of overall performance, Figure 9 summarizes five evaluation metrics across the models. RF achieved the highest accuracy (0.68), outperforming DT (0.61) and GB (0.60). RF also led in Kappa (0.60), Precision (0.68), Recall (0.68), and F1-Score (0.67), demonstrating its consistent superiority across all measures.

Figure 10 further examines feature importance across the models. Q1 and Q7 consistently emerged as the most influential predictors. RF ranked Q1 (0.11), Q7 (0.10), and Q12 (0.10) as its top contributors. DT emphasized Q1

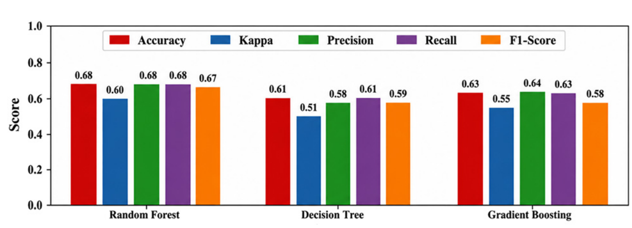

Figure 9: Model evaluation metrics.

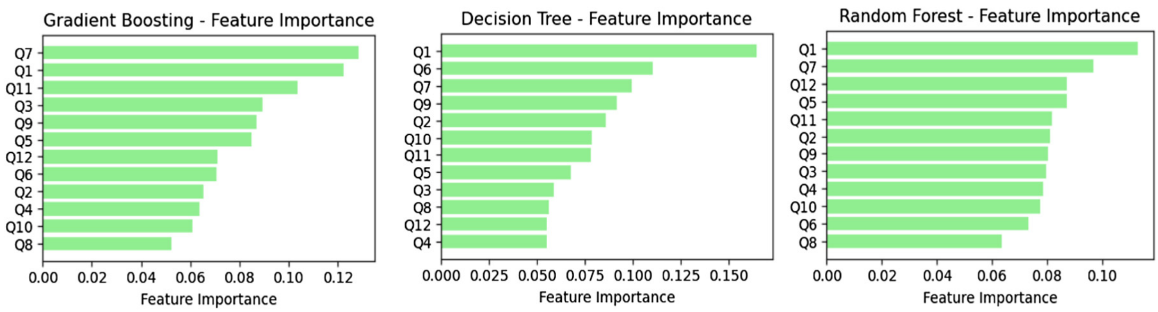

Figure 10: Feature importances.

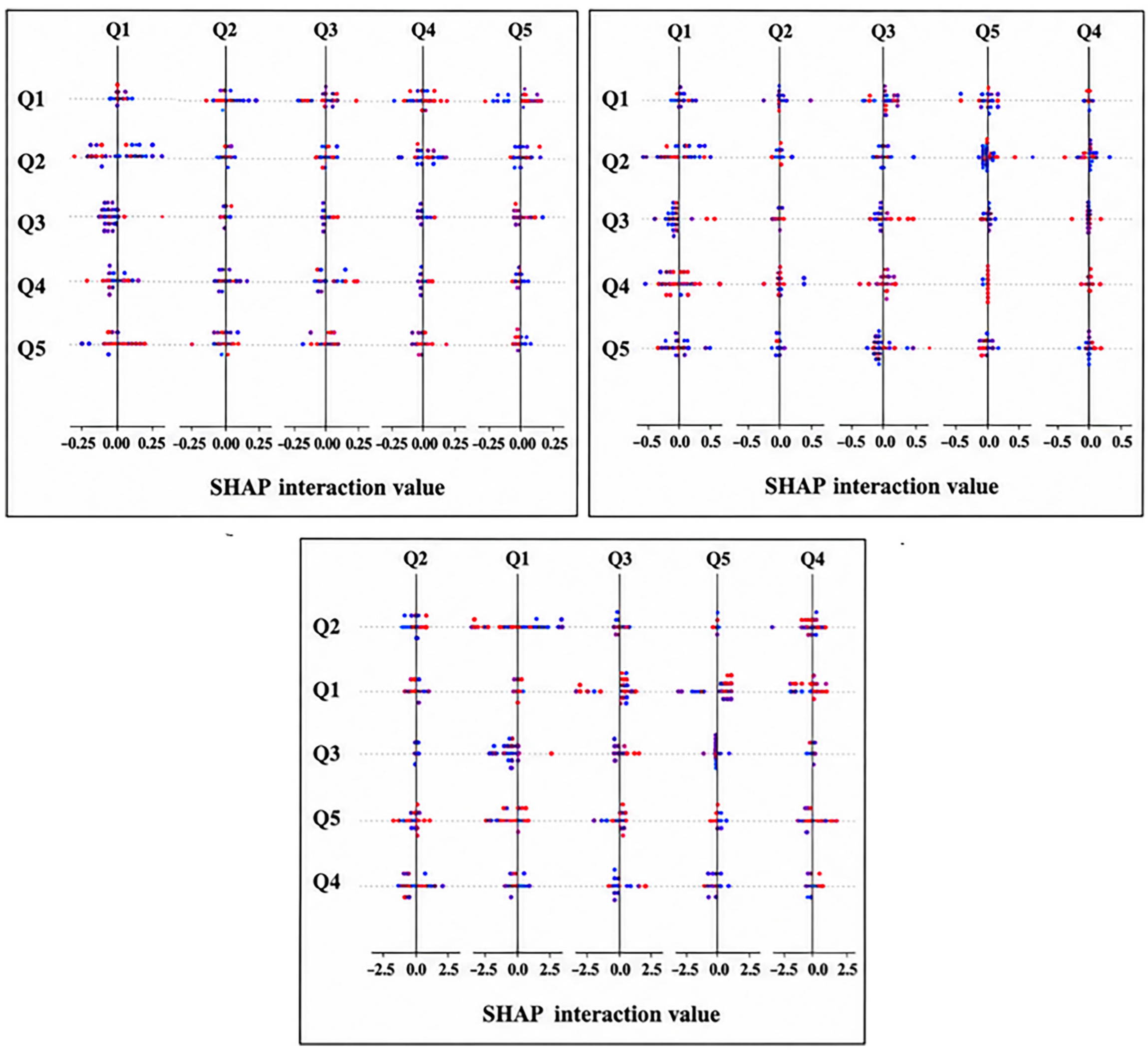

Figure 11: SHAP for three models.

(0.17), Q6 (0.11), and Q7 (0.10), while GB ranked Q7 (0.13), Q1 (0.12), and Q11 (0.11) highest. The divergence in third-ranking features suggests different learning behaviors among models.

Figure 11 presents SHAP interaction values among Q1–Q5 features. RF exhibited high variation (±2.5), indicating complex feature interactions. GB showed tighter bounds (±0.25), suggesting more stable relationships. DT oscillated within ±0.5, indicating nominal feature interaction effects. The strongest SHAP interactions occurred between Q1–Q2, Q3–Q5, and Q4–Q1, underlining how RF excels in capturing nonlinear dynamics.

To truly address these barriers, institutions and policymakers must implement equity-driven strategies. First, subsidizing publication fees (Q1) and standardizing APC waiver policies can expand participation among scholars in under-resourced contexts. Second, fostering global research connectivity (Q7) through international exchange, collaborative digital platforms, and conference sponsorships enhances impactful academic partnerships. Third, funding formulas should be adjusted to prioritize underrepresented disciplines and positions (Q12). Fourth, improving access to research tools and datasets (Q6, Q10) via centralized repositories and technical training will empower scholars to meet rigorous publishing requirements. Finally, mentorship for early-career researchers can help navigate institutional structures and develop collaborative competencies (Q5, Q12).

Although traditional statistical methods effectively established relationship patterns, ML provided superior predictive power and highlighted nuanced trends. By integrating quantitative rigor with predictive modeling, this study offers a forward-looking perspective on barriers to academic publishing. It not only identifies existing barriers but also anticipates future ones, providing a robust foundation for institutional reform and more inclusive research ecosystems.

5 Conclusions and Recommendations

This study offers a nuanced insight into the systemic, institutional, and disciplinary barriers that prevent researchers, especially those from new research settings such as Vietnam, from publishing in WoS and Scopus-indexed journals. Through the fusion of quantitative and ML, the research not only found the leading obstacles, such as lack of institutional support, high publication costs, and insufficient collaboration, but also uncovered deeper structural inequity patterns and emergent threats, such as increasing compliance and data access restrictions.

Machine learning models, however, notably improved the study's prediction, especially emphasizing problems in complicated and interdisciplinary fields such as ECT and SBS. Although RF was the most consistently predictive model, the findings also highlighted the importance of enhanced feature engineering and better balanced datasets in order to reduce misclassification. SHAP analysis, in addition, uncovered these subtle interactions between variables, providing insight into exactly how particular barriers exacerbate one another.

Crucially, the research suggests practical interventions to lower these hurdles. These are practices such as cross-institutional and international collaboration, targeted funding mechanisms, research training, and the equitable allocation of resources across disciplines. While not unique to Vietnam, the findings speak to wider trends across low and middle-income countries, and provide a worldwide roadmap for inclusive academic publishing reform.

To expand on these insights, future work should include qualitative work on researchers' lived experiences, as well as longitudinal studies to evaluate the long-term efficacy of policy interventions. Partnerships between institutions, publishers, and national education agencies will be crucial in making these suggestions a reality. In the end, cultivating a fair, open, and nurturing academic ecosystem will be key to empowering worldwide research prominence, output, and influence, particularly for academics operating in resource-limited contexts.

Acknowledgment: The research number of DHH2025-23-01 of Hue University has supported the collection of field data.
Funding information: Authors state no funding involved.
Author contribution: Acknowledgement that all authors accepted the responsibility for the content of the manuscript and consented to its submission, reviewed all the results, and approved the final version of the manuscript. The authors confirm contribution to the paper as follows: Hong Giang Nguyen, Nguyen Quang Lich: writing – original draft, visualization, resources, Methodology, Investigation, software, resources, data curation.

Vinh V. Le, Nguyen Thi Hai Le: writing – review & editing, writing – original draft, supervision, resources, methodology, conceptualization.

Conflict of interest: Authors state no conflict of interest.

Data availability statement: The data that support the findings of this study are available from the corresponding author, upon reasonable request.

References

Abayeva, N., Mustafina, L., Zhurov, V., Yerakhtina, I., and Mustafina, B. (2024). Leveraging mathematics to enhance critical thinking in technical universities. *Asian J. Univ. Educ.* 20: 566–581.

Adams, D., Sumintono, B., Mohamed, A., & Noor, N.S.M. (2018). E-learning readiness among students of diverse backgrounds in a leading Malaysian higher education institution. *MJLI*, 15, 227-256.

Ahmad, T.S.A.S., Abedin, N.F.Z., Ahmad, I., Paramasivam, S., and Mustapha, W.Z.W. (2022). Awareness and attitudes of undergraduate students towards plagiarism: are there any differences between genders? *Asian J. Univ. Educ.* 18: 597–605.

Aksnes, D.W. and Sivertsen, G. (2023). Global trends in international research collaboration, 1980-2021①. *J. Data Inf. Sci.* 8: 26–42.

Alohali, Y.A., Fayed, M.S., Mesallam, T., Abdelsamad, Y., Almuhawas, F., and Hagr, A. (2022). A machine learning model to predict citation counts of scientific papers in otology field. *Biomed Res. Int.* 2022: 2239152.

Arunachalam, A., Subbarayalu, A.V., Ramalingam, B., Prabaharan, S., Samiappan, D., Marimuthu, P.R., Chinnusamy, S., Shanmugavelu, G., Neethidoss, P., and Ganesan, B. (2024). Quality of work life among the faculty members of higher education institutions in India. *Asian J. Univ. Educ. (AJUE)*, 20, 582–597. https://ir.uitm.edu.my/id/eprint/111512

Aydın Çolak, F. and Eroğlu, Ş. (2025). Challenges in ensuring metadata quality in institutional academic repositories: a qualitative study in Turkish research universities. *Electr. Libr.* 43: 387–404.

Bashir, S., Gul, S., Bashir, S., Nisa, N.T., and Ganaie, S.A. (2022). Evolution of institutional repositories: managing institutional research output to remove the gap of academic elitism. *J. Libr. Inf. Sci.* 54: 518–531.

Bhende, V.V., Sharma, T.S., Krishnakumar, M., Ramaswamy, A.S., Bilgi, K., and Mankad, S.P. (2024). Decoding the rejection code: understanding why articles get axed. *Cureus* 16, https://doi.org/10.7759/cureus.56920.

Bornmann, L., and Haunschild, R. (2023). Reference publication year spectroscopy (RPYS) of papers published by loet leydesdorff: a giant in the field of scientometrics passed away. *Profesional de la información*, 32. e320701

Borrego, Á. (2023). Article processing charges for open access journal publishing: a review. *Learn. Publ.* 36: 359–378.

Buagayan, J.C., Cabuquin, J.C., Avila, N.B., and Gravoso, C.S. (2024). Prevalence and underlying factors influencing academic dishonesty in mathematics among students at a state university in the Philippines. *Asian J. Univ. Educ.* 20: 540–565.

Chen, S., Wang, W., Chen, X., Zhang, M., Lu, P., Li, X., and Du, Y. (2025). Enhancing Chinese comprehension and reasoning for large language models: an efficient LoRA fine-tuning and tree of thoughts framework. *J. Supercomput.* 81: 50.

Cortés-Sánchez, J.D., Robinson-García, N., Chinchilla-Rodríguez, Z., and Ramírez-Cajiao, M.C. (2024). The scholarly age—beyond the academic age using techno-scientific, knowledge appropriation and mentoring outputs, pp. 321–330. https://doi.org/10.1007/978-3-031-57850-2_24.

Cunha-Oliveira, T., Ioannidis, J.P., and Oliveira, P.J. (2024). Best practices for data management and sharing in experimental biomedical research. *Physiol. Rev.* 104: 1387–1408.

Deli, W., Kaur, A., and Awang Hashim, R. (2021). Who delivers it and how it is delivered: effects of social-emotional learning interventions on learning anxiety and dropout intention. *MJLI* 18, https://doi.org/10.32890/mjli2021.18.1.1, https://repo.uum.edu.my/id/eprint/27978.

Dufera, A.G., Liu, T., and Xu, J. (2023). Regression models of pearson correlation coefficient. *Stat. Theory Relat. Fields* 7: 97–106.

Fassnacht, M., Leimstoll, J., Benz, C., Heinz, D., and Satzger, G. (2024). Data sharing practices: the interplay of data, organizational structures, and network dynamics. *Electron. Mark.* 34: 47.

Faure, M.C., Munung, N.S., Ntusi, N.A., Pratt, B., and de Vries, J. (2021). Considering equity in global health collaborations: a qualitative study on experiences of equity. *PLoS one* 16: e0258286.

Gaye, B., Isiozor, N.M., Singh, G., Gaye, N.D., Ka, M.M., Seck, D., Gueye, K., Kitara, D.L., Lassale, C., Malick, A., et al (2024). Barriers to global engagement for African researchers: a position paper from the alliance for medical research in Africa (AMedRA). *J. Glob. Health* 14: 03042.

Graves Jr, J.L., Kearney, M., Barabino, G., and Malcom, S. (2022). Inequality in science and the case for a new agenda. *Proc. Natl. Acad. Sci. U. S. A.*. 119(10): e2117831119.

Hazra, R., Aryan, Aggarwal, H., Marsili, M., and Mukherjee, A. (2020). Characterising authors on the extent of their paper acceptance: a case study of the journal of high energy physics. In: *Proceedings of the ACM/IEEE joint conference on digital libraries in 2020*, pp. 7–16

Heyard, R. and Hottenrott, H. (2021). The value of research funding for knowledge creation and dissemination: a study of SNSF research grants. *Humanit. Soc. Sci. Commun.* 8: 1–16.

Hoang, C.H. and Dang, T.T.D. (2021). Developing a framework to explore local researchers” engagement with global academia: the case of Vietnamese social sciences scholars. *Asian Soc. Sci.* 17: 1.

Hussinger, K. and Carvalho, J.N. (2022). The long-term effect of research grants on the scientific output of university professors. *Ind. Innov.* 29: 463–487.

Jain, V.K., Iyengar, K.P., and Vaishya, R. (2021). Article processing charge may be a barrier to publishing. *J. Clin. Orthop. Trauma* 14: 14–16.

Khan, Z., Wu, J., Khan, H., Amankwah-Amoah, J., Czinkota, M., and Zahoor, N. (2024). Exploring periphery–core dynamics in international management: a review, characterizations, and future research agenda, 42: 653–657.

Marsden, E. and Morgan-Short, K. (2023). (Why) are open research practices the future for the study of language learning? *Language Learn.* 73: 344–387.

Miao, L., Larivière, V., Lee, B., Ahn, Y.-Y., and Sugimoto, C.R. (2024). Persistent hierarchy in contemporary international collaboration, arXiv preprint arXiv:2410.13020 https://doi.org/10.48550/arXiv.2410.13020.

Mirhosseini, S.-A., Rashed, F., and Shirazizadeh, M. (2024). We have become protectors of English: revisiting policies of publishing in English in non-anglophone academia. *CILP* 25: 266–284.

Nicholas, D., Revez, J., Abrizah, A., Rodríguez-Bravo, B., Boukacem-Zeghmouri, C., Clark, D., Xu, J., Swigon, M., Watkinson, A., Jamali, H.R., et al (2024). Purchase and publish: early career researchers and open access publishing costs. *Learn. Publ.* 37: e1617.

Oluchukwu Njoku, A., Nyunga Mpinda, B., and Olawale Awe, O. (2023). Improving the accuracy of financial bankruptcy prediction using ensemble learning techniques, pp. 3–29. Pan African Conference on Artificial Intelligence

Ou, G., Zhao, K., Zuo, R., and Wu, J. (2024). Effects of research funding on the academic impact and societal visibility of scientific research. *J. Informetr.* 18: 101592.

Owan, V.J., Ameh, E., and Anam, E.G. (2024). Collaboration and institutional culture as mediators linking mentorship and institutional support to academics' research productivity. *ERPP* 23: 19–44.

Petersen, O.H. (2021). Inequality of research funding between different countries and regions is a serious problem for global science. *Function* 2: zqab060.

Piran, M.J. and Tran, N.H. (2024). Enhancing research methodology and academic publishing: a structured framework for quality and integrity, arXiv preprint arXiv:2412.05683 https://doi.org/10.48550/arXiv.2412.05683.

Rizkallah, L.W. (2025). Enhancing the performance of gradient boosting trees on regression problems. *J. Big Data* 12: 35.

Samala, A.D., Rawas, S., Criollo-C, S., Fortuna, A., Feng, X., Prasetya, F., Uluçay, N.Ö., Jaya, P., and Hidayat, R. (2024). Social media in education: trends, roles, challenges, and opportunities for digital-native generations–a systematic literature review. *Asian J. Univ. Educ.* 20: 524–539, https://ir.uitm.edu.my/id/eprint/111509.

Silver, R.E., Lin, E., and Sun, B. (2023). Applied linguistics journal editor perspectives: research ethics and academic publishing. *Res. Methods Appl. Linguist.* 2: 100069.

Tennant, J.P. (2018). The state of the art in peer review. *FEMS Microbiol. Lett.* 365: fny204.

Tenzer, H., Pudelko, M., and Zellmer-Bruhn, M. (2021). The impact of language barriers on knowledge processing in multinational teams. *J. World Bus.* 56: 101184.

Turki, H., Fraumann, G., Hadj Taieb, M.A., and Ben Aouicha, M. (2023). Global visibility of publications through digital object identifiers. *Front. Res. Metr. Anal.* 8: 1207980.

Whittaker, T.A. and Schumacker, R.E. (2022). *A beginner's guide to structural equation modeling*. Routledge.

Yao, B. (2021). International research collaboration: challenges and opportunities. *J. Diagn. Med. Sonogr.* 37: 107–108.

Zhang, F. and Wu, S. (2024). Predicting citation impact of academic papers across research areas using multiple models and early citations. *Scientometrics* 129: 4137–4166.