EnPress Publisher

Journal of Infrastructure, Policy and Development 2024, 8(16), 9931.
https://doi.org/10.24294/jipd9931

Article

The frontiers of ecological art research: A multidimensional analysis based on bibliometrics and machine learning predictions

Zexi Liu[1], Jin Ho Im[2], Yujia Chen[3], Xiaoyang Qiao[4], Wenliang Ye[2,*]

[1] Chonbuk National University, Jeonju 561756, Republic of Korea
[2] Chodang University, Muan 58530, Republic of Korea
[3] Haikou University of Economics, Haikou 570100, China
[4] Kangwon National University, Chuncheon 24341, Republic of Korea
*** Corresponding author:** Wenliang Ye, youseesy67@gmail.com

CITATION

Liu Z, Im JH, Chen Y, et al. (2024). The frontiers of ecological art research: A multidimensional analysis based on bibliometrics and machine learning predictions. Journal of Infrastructure, Policy and Development. 8(16): 9931. https://doi.org/10.24294/jipd9931

ARTICLE INFO

Received: 28 October 2024
Accepted: 4 December 2024
Available online: 24 December 2024

Abstract: With the advancement of technology, human impact on the natural world has progressively intensified, leading to an increased frequency and expanded scope of environmental issues. Ecological art integrates natural and human environments, aiming to restore damaged ecosystems through artistic expression while reflecting artists' understanding and concerns about nature. As an emerging interdisciplinary art form, ecological art has garnered significant attention; however, systematic research in this field remains insufficient. This study combines bibliometric analysis with machine learning techniques to examine ecological art literature indexed in the WOS Core Collection between 2004 and 2024, uncovering research trends, core themes, and future directions. In the bibliometric analysis, tools such as VOSviewer, CiteSpace, Bibliometrix, Pajek, and HistCite were utilized to systematically analyze co-occurrence networks, research hotspots, and their evolutionary pathways. For machine learning analysis, comparative experiments were conducted using Random Forest, Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Extreme Gradient Boosting (XGBoost), and LightGBM models. Ultimately, Random Forest was selected for predicting citation rates and interpreting influential factors. The findings indicate that machine learning techniques quantitatively reveal the centrality of "Ecological Design" and "Sustainability" as core themes in ecological art while highlighting the emergence of new topics such as "Urban Greenspace" and "Eco-Art Education" in recent years. The publication year was identified as the most critical factor influencing citation rates, followed by the number of authors and keywords. The first author's affiliation country also significantly impacted citation performance. This study, through the integration of bibliometric and machine learning approaches, provides a novel methodological perspective for identifying research frontiers in ecological art, predicting academic influence, and guiding future scholarly and practical applications.

Keywords: ecological art; bibliometrics; machine learning; sustainability

1. Introduction

Ecological art is an art practice centered on ecology, exploring the relationship between humans and nature through various artistic forms while advocating ecological awareness and sustainable development. Ecological art has emerged as an effective tool for addressing environmental challenges, playing a key role in transforming urban spaces and advancing sustainability (Guy et al., 2015). This art form not only examines aesthetic structures, forms, and evolution within ecological environments, but also seeks to explore the interconnectedness between humans and other life forms on Earth. Artists collaborate with nature as a creative partner, which is helping to foster new aesthetic perspectives and understandings of art (Tsevreni, 2022). Today, ecological art has evolved beyond traditional expressions by integrating digital technologies, creating new forms of interaction that further engage the public in environmental conservation (Brunner et al., 2013). This innovative approach has significant potential to reshape existing social and environmental systems (Simon, 2006). Its interdisciplinary nature positions it as a valuable tool for comprehending complex ecological issues, offering new methods and insights for solving environmental problems (Song, 2009). Another key aspect of ecological art is its focus on public engagement, using artistic expression to highlight environmental issues, encourage community involvement, and enhance environmental awareness (Cucuzzella, 2021). Research has shown that ecological art is effective in raising environmental consciousness, inspiring behavioral change, and fostering systemic reform (Weder and Voci, 2021). In the realm of education, ecological art is becoming a powerful tool for environmental education, helping students develop ecological awareness and encouraging them to reflect on the relationship between humanity and nature through artistic expression (Prášek, 2024).

Despite the significant potential of ecological art, comprehensive reviews and forward-looking analyses of this field are still lacking in academic circles. In this paper, we use bibliometric tools like VOSviewer, CiteSpace, Bibliometrix, Pajek, and HistCite to systematically analyze ecological art literature from the past two decades in the Web of Science Core Collection. Additionally, machine learning models are applied to predict citation rates, with Shapley Additive Explanations (SHAP) used to interpret the impact of various factors on the predictions, aiming to uncover key trends and emerging issues in this field.

Our findings highlight the promising role of ecological art in fostering environmental awareness and social transformation. The study maps the global distribution of research and the evolution of international collaboration networks, identifying key research themes and future directions, particularly in areas such as ecological design, sustainability, urban green spaces, and ecological art education. By predicting the academic impact of research papers, this study offers a new approach for evaluating scholarly influence in the ecological art field, providing valuable insights for advancing both research and practice.

2. Methodology

2.1. Data sources and search strategy

The Web of Science Core Collection serves as the primary global database for literature retrieval. To ensure data comprehensiveness and to minimize errors caused by daily updates, we selected publications from 1 January 2004, to 15 August 2024. The search was conducted using key terms such as “ecological art”, “eco-art”, and “sustainable art”.

By entering the search query shown in **Figure 1** into the WOS core collection, 820 articles were retrieved. To ensure the specificity and quality of the study, we conducted an initial screening of these articles, excluding those unrelated or only weakly related to ecological art. After removing duplicates, 433 articles were finally included. We confirmed that the dataset included is representative over time, but caution is still needed when interpreting the distribution of documents during low-density periods to avoid time bias impacting the study results.

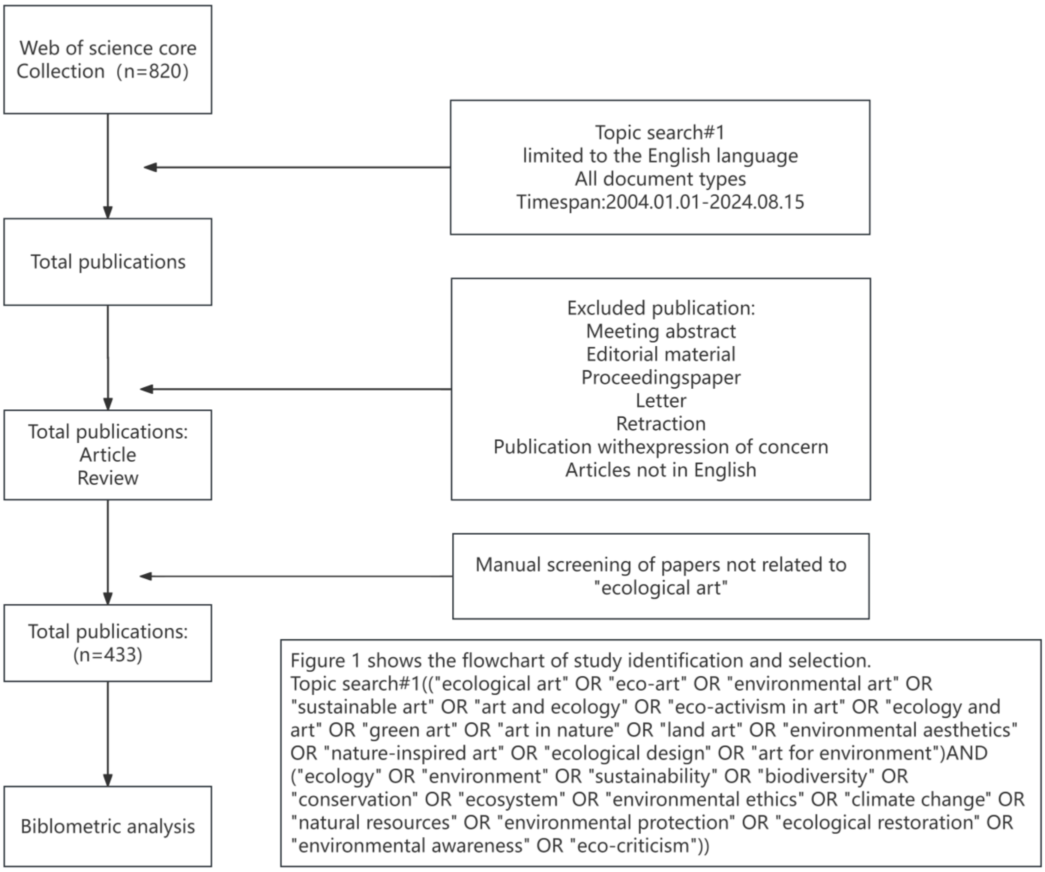

Figure 1. Flowchart of ecological art research literature screening and analysis.

2.2. Bibliometric tools and techniques

To systematically analyze research trends and frontier issues in the field of ecological art, this study employed a variety of bibliometric tools and techniques. The screened 433 articles were exported in both plain text file and tab-delimited file formats, including full records and cited references, and then imported into VOSviewer (version 1.6.19), CiteSpace (version 6.2.R4), Bibliometrix (R package version 4.4.1), Scimago Graphica (version 1.0.35), Pajek (version 5.19), and HistCite Pro (version 2.1). By leveraging the strengths of different tools, a multi-dimensional and comprehensive perspective was achieved to enhance the validity and reliability of the results.

When handling large-scale bibliometric data, VOSviewer excels in visualization capabilities (Chen, 2006; Van Eck and Waltman, 2010). CiteSpace integrates network visualization, spectral clustering, automatic cluster labeling, and text summarization, showing particular strengths in identifying emerging trends and transient patterns (Chen et al., 2010). Bibliometrix, an open-source R package, offers flexible and easily extendable functionalities, enabling quick integration with other statistical R packages to address the ever-evolving demands of scientometric analysis (Aria and Cuccurullo, 2017). Bibliometrix covers a wide range of techniques and provides a user-friendly interface via Biblioshiny, making complex science mapping analysis easier to perform (Moral-Muñoz et al., 2020). Bibliometrix has demonstrated strong capabilities in bibliometric analysis and science mapping, effectively enabling data mining and quantitative analysis to reveal the current state and future directions of research in the field (Xie et al., 2020). Pajek, based on graph theory and social network analysis, is effective for various types of co-occurrence matrix analysis, and has proven to be highly effective in applications within information science (Leydesdorff and Vaughan, 2006). SCImago Graphica is not only suitable for visualizing data but also for exploratory data analysis (Hassan-Montero et al., 2022). HistCite, with its powerful analytical functions, can systematically visualize large volumes of literature, helping to quickly identify key literature and research hotspots in the field (Pan et al., 2018).

2.3. Shapley additive explanations–SHAP

To explain the machine learning models, the SHAP method was introduced to interpret the relationship between features and outcomes, providing support for the reliability of the model's results. SHAP is a unified framework for interpreting model predictions, based on the concept of Shapley values (Lundberg, 2017). Rooted in game theory, SHAP is the only feature attribution method that ensures consistency and local accuracy based on expectations. This method treats each feature as a contributor, and by calculating the contribution value of each feature, the sum of these values leads to the model's final prediction (Parsa et al., 2020). For ensemble tree models, when performing classification tasks, the model outputs a probability value. SHAP calculates the Shapley value for each feature, which measures the contribution of each feature to the final prediction. Assuming f represents the explanatory model, k represents the total number of features, and $\varnothing$ represents the Shapley value for each feature, the formula is given as:

$$\mathrm{f}(n) = \emptyset_0 + \sum_{i=0}^{n} \emptyset_i$$

Here, $\mathrm{f}(n)$ represents the model's predicted value, $\varnothing_0$ is the average prediction across all training samples, and $\varnothing_\mathrm{i}$ is the attribution value corresponding to each feature. The specific formula for $\varnothing_\mathrm{i}$ is:

$$\varnothing(u, i) = \sum_{S \subseteq N \setminus (i)} \frac{|S|!\,(N - |S| - 1)!}{N!} (u(S \cup \{i\} - u(S))$$

In the formula: N represents the entire set of features, S is an ordered subset of N, $\mathrm{fx}(S \cup \{i\} - \mathrm{fx}(S)$ is the marginal contribution of feature i when added to subset S.M is the total number of input features. When the SHAP value corresponding to a feature is positive, it indicates that the feature has a positive impact on the prediction result; conversely, a negative SHAP value implies that the feature decreases the model's prediction accuracy. One major advantage of SHAP is its ability to reflect the contribution of each feature to the prediction result.

3. Bibliometrics

3.1. Analysis of disciplinary intersections and publication trends in ecological art research

Figure 2 systematically illustrates the interdisciplinary characteristics, publication trends, and document coupling clusters within the field of ecological art research, shedding light on the structural framework and dynamic evolution of this research domain. **Figure 2A**, through a disciplinary network map, highlights the interdisciplinary nature of ecological art studies, identifying environmental studies, physical geography, and multidisciplinary sciences as central disciplines in this field. Among the keywords, “Ecological Design” and “Sustainability” emerge as high-frequency terms occupying prominent central positions, underscoring their pivotal role within the domain. The dense connections between nodes further reveal deep interdisciplinary integration and collaborative research, particularly among environmental sciences, sociology, and design studies.

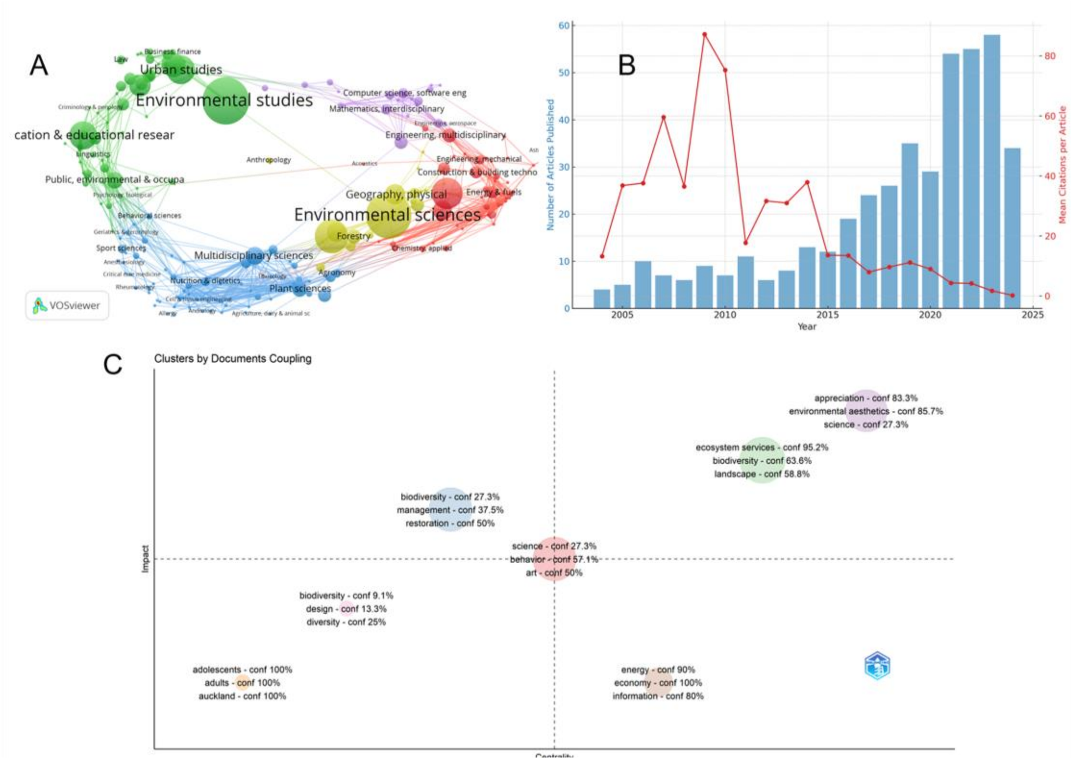

Figure 2. Interdisciplinary characteristics, publication trends, and document coupling clusters in ecological art research, **(A)** disciplinary overlay visualization in ecological art research; **(B)** publication trends and citation impact in ecological art research (2004–2024); **(C)** clustering analysis of document coupling in ecological art research.

Figure 2B, through the analysis of publication trends (represented by blue bars) and citation impact (depicted by the red line), delineates the temporal evolution of ecological art research. The results indicate a rapid increase in the number of publications from 2004 to 2024, with a notable surge in interest after 2010. **Figure 2C** employs document coupling analysis to uncover the thematic distribution and clustering characteristics within the field. Core clusters are predominantly concentrated on “Ecosystem Services” and “Biodiversity,” located in regions of high centrality and influence, signifying their status as key research areas. Furthermore, topics such as “Design” and “Restoration,” although exhibiting lower centrality, demonstrate significant connectivity within the research network, highlighting their critical role in fostering interdisciplinary collaboration.

3.2. Countries/regions, institutions, and authors

In ecological art research, the global research efforts are mainly concentrated in a few countries (**Figure 3A,B**). According to the data shown in the tables, the United States leads with 175 articles, followed closely by China with 170 articles, while the United Kingdom and Australia contribute 55 and 50 articles, respectively (**Table 1**). Research institutions in these countries play central roles in the output and international collaboration of ecological art research. The University of California System leads with 13 articles, while the University of Arizona and the University of Michigan also perform prominently in collaboration and research output (**Figure 3D**)

Figure 3C presents the collaboration network among key research institutions. The University of California has reinforced its dominant position in ecological art research through collaborations with numerous institutions worldwide. Australia’s University of Sydney and University of Melbourne, as well as China’s Peking University, also hold significant positions in international cooperation.

Table 1. Major contributing countries, leading institutions, and authors in ecological art research.

Country	Freq	Affiliation	Articles	Authors	Articles Fractionalized
USA	175	UNIVERSITY OF CALIFORNIA SYSTEM	13	BRADY E	3.50
CHINA	170	UNIVERSITY OF ARIZONA	8	NASSAUER JI	3.08
UK	55	UNIVERSITY OF MICHIGAN	8	WANG Y	0.89
AUSTRALIA	50	UNIVERSITY OF MICHIGAN SYSTEM	8	WANG ZF	1.33
SPAIN	38	UNIVERSITY OF TORONTO	7	FELSON AJ	0.82

The contribution of core authors in this field is equally noteworthy. BRADY E and NASSAUER JI have not only published a considerable number of articles but also achieved significant influence through their work. BRADY E, from the University of California, has contributed 3.50 articles (proportionally allocated), while NASSAUER JI, from the University of Arizona, has contributed 3.08 articles (see **Table 1**). The high output and citation rates of these authors indicate their profound impact on ecological art research.

The Lotka’s Law analysis (**Figure 3F**) reveals the dominance of a few prolific authors in ecological art research. This phenomenon shows that a small group of leading scholars produces the majority of research output, while most authors publish fewer papers. The visualization in **Figure 3E** highlights how these high-output authors, through continuous research activity, have driven the ongoing development of the ecological art field.

Overall, the global landscape of ecological art research exhibits a high degree of concentration, primarily dominated by a few countries, institutions, and core authors. As global attention on ecological issues continues to rise, these research forces will likely continue to shape the future direction of ecological art research and have a lasting impact on global environmental policy and social practices

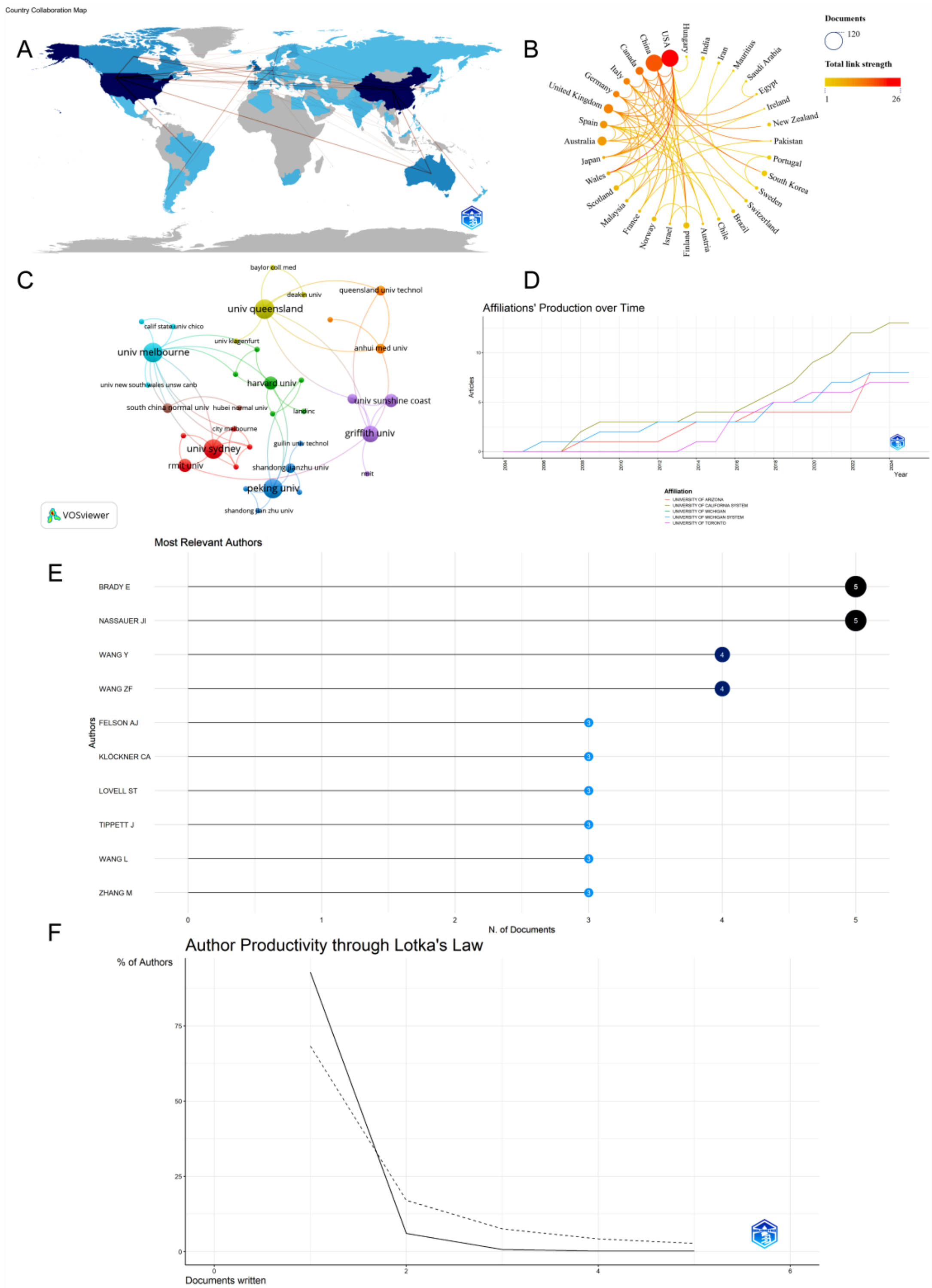

Figure 3. Collaboration and productivity analysis of countries, institutions, and authors in ecological art research, **(A)** global collaboration geographic distribution in ecological art research; **(B)** international collaboration network in ecological art research (based on document quantity and connection strength); **(C)** institutional collaboration network in ecological art research; **(D,E)** most influential authors in ecological art research; **(F)** author productivity analysis in ecological art research based on Lotka's Law.

3.3. Highly cited papers

Analyzing highly cited papers is key to understanding the core and frontier dynamics of academic research. As shown in **Table 2**, the paper by Nassauer JI, What will the neighbors think? Cultural norms and ecological design, published in 2009, ranks first with 260 citations, averaging 16.25 citations per year. This paper explores how cultural norms and neighborhood aesthetics influence individuals' preferences for residential landscape design (Nassauer et al., 2009). Similarly, Freilich S's 2010 paper, The large-scale organization of the bacterial network of ecological co-occurrence interactions, has also garnered a high citation frequency. This paper discusses the ecological design principles followed by microbial communities, reflecting how organisms in nature coexist and thrive through competition and cooperation in limited resources (Freilich et al., 2010). For ecological artists, these interactions in micro-ecosystems can serve as inspiration, guiding them in considering material choices, placement of artworks, and the dynamic relationship between the artwork and its environment during the creative process.

Table 2. Highly cited papers in ecological art research and their key metrics.

Paper	DOI	Total Citations	TC per Year	Normalized TC
NASSAUER JI, 2009, LANDSCAPE URBAN PLAN	10.1016/j.landurbplan.2009.05.010	260	16.25	2.98
FREILICH S, 2010, NUCLEIC ACIDS RES	10.1093/nar/gkq118	223	14.87	2.96
GROSS M, 2007, CURR SOCIOL	10.1177/0011392107079928	168	9.33	2.82
FAEH D, 2009, CIRCULATION	10.1161/CIRCULATIONAHA.108.819250	166	10.38	1.90
STEFANAKIS AI, 2019, SUSTAINABILITY-BASEL	10.3390/su11246981	143	23.83	12.80
PALMER MA, 2014, ECOL ENG	10.1016/j.ecoleng.2013.07.059	138	12.55	3.64
LOVELL ST, 2009, FRONT ECOL ENVIRON	10.1890/070178	130	8.13	1.49
GARCÍA-SERNA J, 2007, CHEM ENG J	10.1016/j.cej.2007.02.028	127	7.06	2.13
LOVELL ST, 2010, AGR SYST	10.1016/j.agsy.2010.03.003	123	8.20	1.63
NASSAUER JI, 2012, LANDSCAPE URBAN PLAN	10.1016/j.landurbplan.2012.03.014	123	9.46	3.88

In **Figure 4A**, the co-citation network of highly cited papers reveals the close connections between these core documents. Brady (2003) and Carlson (2000) are linked by strong association lines, highlighting their importance in the research theme of "environmental aesthetics". These papers not only form the foundational theory of ecological art but also contribute to the ongoing development of the field.

Gross M's 2007 paper, The Unknown in Process: Dynamic Connections of Ignorance, Non-Knowledge, and Related Concepts, explores the complex relationship between knowledge and ignorance in ecological art and design from a sociological perspective. The paper groundbreakingly introduces the concepts of "non-knowledge" and "ignorance" into ecological art research, revealing how these ideas shape public understanding of environmental design and art during social transformation and reflections on modernity. Gross's work has prompted a reevaluation of the role of ecological art in social change, expanding the theoretical foundation of art in responding to environmental uncertainty and reflecting on modernity (Goralnik et al., 2017).

Figure 4C shows the 20 papers with the strongest citation bursts from 2004 to 2024, reflecting the rapid rise of these key issues over specific periods. Brady E.'s research significantly influenced the fields of ecological aesthetics and environmental ethics between 2022 and 2024, emphasizing the complex aesthetic relationship between humans and the natural environment (Brady and Prior, 2020). Carlson A.'s research further advanced this discussion during 2020 to 2021, particularly playing a crucial role in exploring the ethical dimensions of environmental aesthetics (Shapshay et al., 2018). Wu DS's research gained widespread attention between 2021 and 2022, highlighting the impact of environmental pollution on public health and the importance of potential solutions (Wu and Ning, 2018).

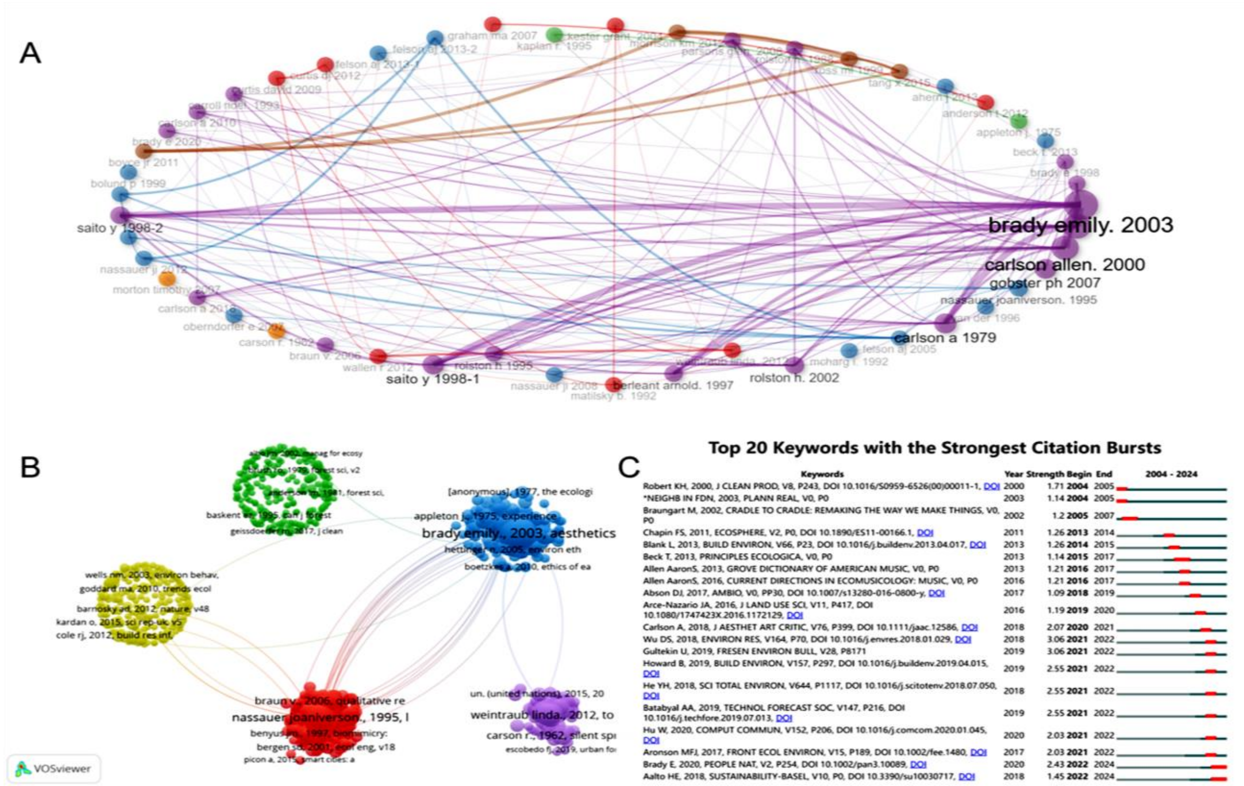

Figure 4. Citation network and clustering analysis in ecological art research, **(A)** co-citation network of influential publications in ecological art research; **(B)** thematic clustering in ecological art research based on co-citation analysis; **(C)** the top 20 papers with the strongest citation bursts in ecological art research.

3.4. Keyword analysis

In the keyword analysis of ecological art research, based on the analysis results from VOSviewer and CiteSpace tools, we can explore the research hotspots, development trends, and the relationships between keywords in this field.

Figure 5A presents the keyword co-occurrence network in ecological art research. In the figure, "ecological design" occupies a central position in the network, closely linked to other keywords such as “sustainability”, “biodiversity”, and “environmental aesthetics”. These keywords reflect the main focus areas of current ecological art research. The dense connections between keywords demonstrate the close relationships between different topics, especially highlighting the interdisciplinary intersections among ecology, environmental science, and design studies.

The keyword clustering (**Figures 5A,B**) reveals several research themes within ecological art research and categorizes them into different research directions. The most prominent cluster is “ecological design,” which is not only an independent research topic but is also closely related to areas such as ecosystem services and green infrastructure. This may be strongly linked to the growing global focus on sustainable development. **Figure 5C** illustrates the temporal development of research themes in ecological art. Since 2000, “environmental art” and “ecological design” have maintained a high level of research attention, demonstrating their lasting influence in the field. In recent years, emerging themes such as “urban greenspace” and “eco-art education” have surfaced, indicating that ecological art research is evolving toward a more diversified and systematic approach.

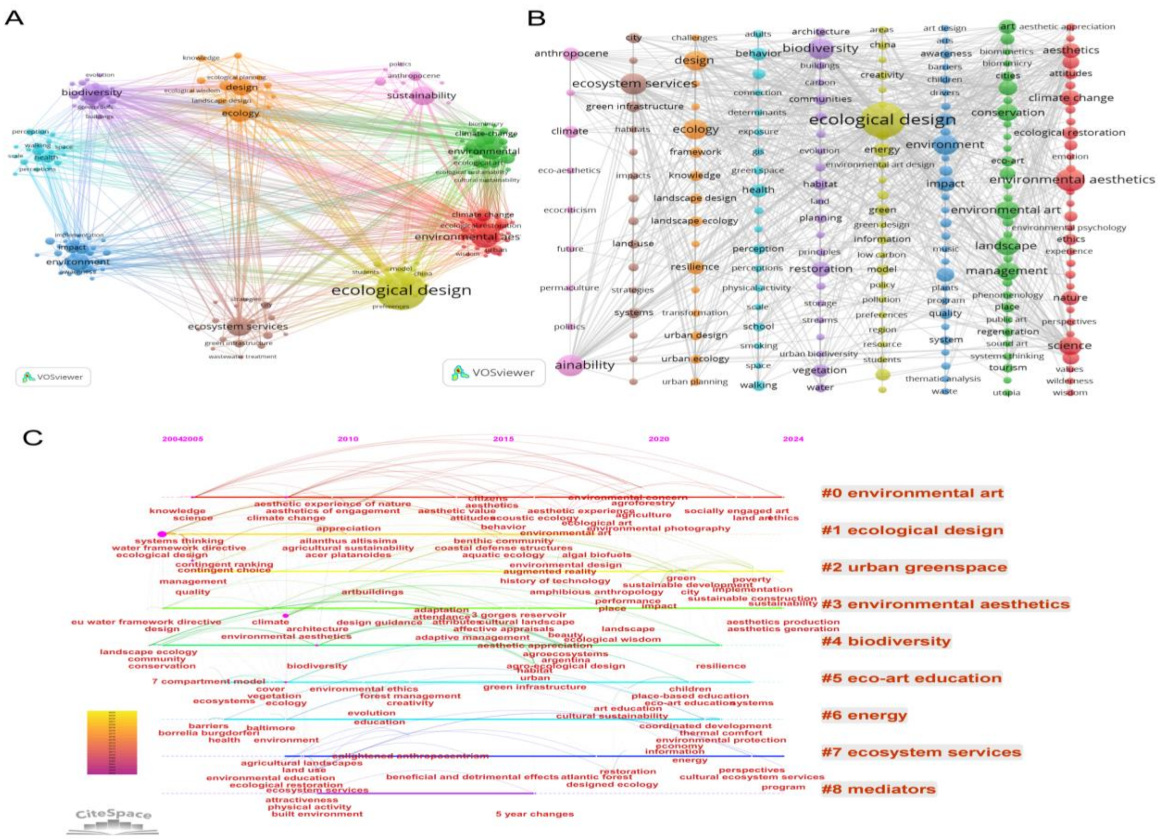

Figure 5. Keyword clustering and thematic evolution analysis in ecological art research, **(A)** keyword co-occurrence network in ecological art research; **(B)** keyword thematic clustering in ecological art research; **(C)** temporal evolution of research themes in ecological art (2004–2024).

4. Predicting citation probability of ecological art literature

4.1. Selection of machine learning methods

We conducted multiple simulations using supervised regression models to select the optimal machine learning method. The prediction models were based on a range of metadata related to the literature, including the year of publication, type of authorship (single or multiple authors), number of authors, country of the first author, total number of keywords, and the use of specific keywords.

The following techniques were used in our simulations: Linear regression, elastic net (which is a linear regression combined with Lasso and Ridge regularizers) (Zou and Hastie, 2005), random forest (Breiman, 2001), support vector machines with an RBF kernel (Cortes and Vapnik, 1995), k-nearest-neighbors (Fix, 1985), extreme gradient boosting trees (XGBoost), and light gradient boosting machine (LightGBM) (Ke et al., 2017). Based on existing research, we did not use deep learning techniques (Silva et al., 2024), as our dataset is relatively small, and the complexity of these methods may lead to overfitting. To train and validate the machine learning models, this study divided the dataset of 433 articles into training and validation sets with a ratio of 80:20. Specifically, the training set consisted of 346 articles for model training, while the validation set included 87 articles for evaluating model performance. The division process was based on random splitting to ensure the representativeness and randomness of the data. Additionally, to avoid data bias, the validation set was designed to closely match the training set in terms of key features such as publication year, number of keywords, and number of authors, ensuring the fairness and reliability of the model evaluation.

Figure 6 shows the simulation results using a cross-validation procedure. The points represent the average RMSE (Root Mean Squared Error) obtained in the ten independent folds that were not used for training during the cross-validation process. Cross-validation was performed using the optimal hyperparameters for each technique.

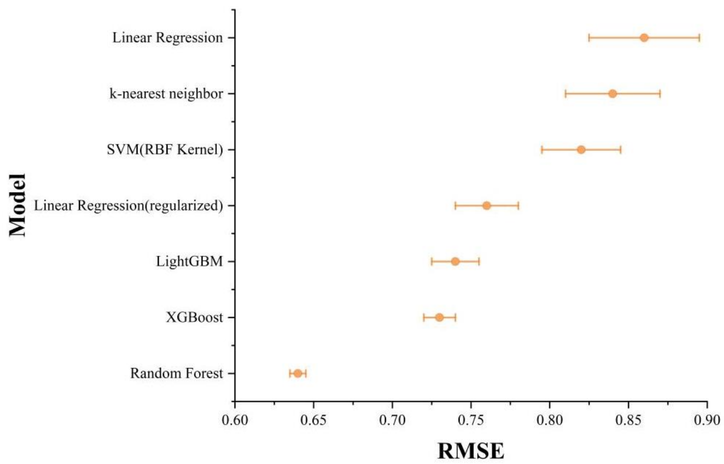

Figure 6. Comparison of RMSE across different machine learning models in predicting citation rates of ecological art literature.

During the model-building process, we executed a series of data preprocessing steps. For each value in the categorical variables, we created dummy variables to represent different keywords and the five major countries to which the first author belonged. For papers without keywords, we specifically set a “no keyword” dummy variable. As these operations might generate a large number of attributes, we further streamlined the feature set to avoid excessive data complexity.

We removed attributes with near-zero variance, meaning those where the most common value occurred more than 20 times as frequently as the second most common value. Additionally, we deleted features highly correlated with other variables, with the minimum absolute correlation threshold set at 0.80, to reduce the issue of multicollinearity. This process resulted in a final set of 20 key attributes along with the target variable.

To fill in missing values in the dataset, we used the k-nearest-neighbors algorithm to impute 25 missing entries in the publication year variable.

In the model selection and optimization process, we aimed to minimize the root mean square error (RMSE) by adjusting and optimizing specific hyperparameters for each model. We employed 10-fold cross-validation (K = 10) across the entire dataset to determine the optimal parameter settings, rather than reserving a small subset of data for validation, as we were dealing with a relatively small sample size (433 papers). Given the strong right skewness in the distribution of annual citation counts, we used the logarithm of annual citation counts as the target variable and avoided computational anomalies by adding 1 to the citation count before dividing by the number of years.

We optimized the hyperparameters of each model technique through a grid search with 100 different combinations. For the elastic net model, we tuned the “mixture” parameter, which controls the blend between Lasso and Ridge regularization (where a value of 1 indicates Lasso only, a value of 0 indicates Ridge only, and intermediate values represent a mix of the two), as well as the regularization weight “penalty” to balance the model’s loss function and regularization term. For the random forest model, we optimized the number of features randomly selected for each tree (“mtry”) and the minimum number of observations per leaf node. In the support vector machine (SVM) model, we fine-tuned the regularization weight (“cost”) and the RBF kernel’s parameter (“rbf_sigma”). For the k-nearest neighbors algorithm, we adjusted the number of neighbors (“neighbors”) and the type of kernel function used for weighting the distance between samples. In the extreme gradient boosting (XGBoost) and light gradient boosting machine (LightGBM) models, we optimized the learning rate (“learn_rate”), tree depth (“tree_depth”), and the number of trees.

The final optimized hyperparameters are detailed in **Table 3**. Cross-validation results (**Figure 6**) showed that tree-based ensemble models performed best on the dataset. Given that RMSE was our primary performance indicator, the random forest model stood out with the lowest RMSE, making it the best-performing model.

During model fitting, we also observed that RMSE was highly affected by outliers, which were primarily concentrated in papers with extremely high citation counts (e.g., the 2009 paper by Joan Iverson Nassauer with 260 citations). To address this, we considered mean absolute error (MAE) as a secondary evaluation metric. In the remaining folds of the cross-validation, the model's average MAE was 0.60, indicating that the prediction error in annual citation counts ranged between 0.41 and 1.10 citations per year. Given the challenge of identifying highly cited papers solely based on metadata, predicting the average annual citation count remains a complex issue.

Table 3. Presents the optimal hyperparameters achieved for each regression technique.

Technique	Parameters	Description	Value
Elastic Net	penalty	Regularization weight relative to the loss function.	0.1466
	mixture	Mixture between Lasso (= 1) and Ridge (= 0) regularization.	0.2193
SVM (RBF Kernel)	cost	Regularization weight relative to the loss function	1.3776
	rbf_sigma	Controls the bandwidth of the Gaussian kernel.	0.0016
k-nearest neighbor	neighbors	Number of neighbors to consider	12
	weight_func	Kernel function used to weight distances between samples	optimal
	dist_power	Exponent of the Minkowski distance metric.	0.3618
Random Forest	mtry	Number of random features considered in each tree	61
	trees	Number of random trees	1201
	min_n	Minimum number of observations needed to split a node.	6
XGBoost	learn_rate	Learning rate in the boosting process. 0.0036 tree	0.0038
	tree_depth	Maximum depth of each tree	4
	trees	Number of random (and sequential) trees.	1763
LightGBM	learn_rate	Learning rate in the boosting process	0.0024
	tree_depth	Maximum depth of each tree.	6
	trees	Number of random (and sequential) trees.	1470

Note: K-nearest neighbor classifier depends asymptotically only on the dimension d of the feature vectors and the value is close to optimal when d is large (Samworth, 2012).

4.2. Using SHAP values to interpret predictions of nonlinear methods

In this section, we aim to gain a deeper understanding of how each attribute contributes to the target variable, specifically the number of citations per year. To achieve this, we focus on the optimal configuration of the random forest model (i.e., our selected best model). By utilizing the SHAP (Shapley Additive Explanations) method along with the interpretability features of the random forest model itself, we were able to uncover the extent to which each feature influences the prediction results.

As seen in **Figure 7**, the year of publication is the most important factor influencing citation counts, followed by the number of authors and the number of keywords. The country of the first author also significantly impacts the citation count. These findings suggest that the publication time and author characteristics play critical roles in predicting citation rates, further emphasizing the importance of these variables in analyzing academic impact.

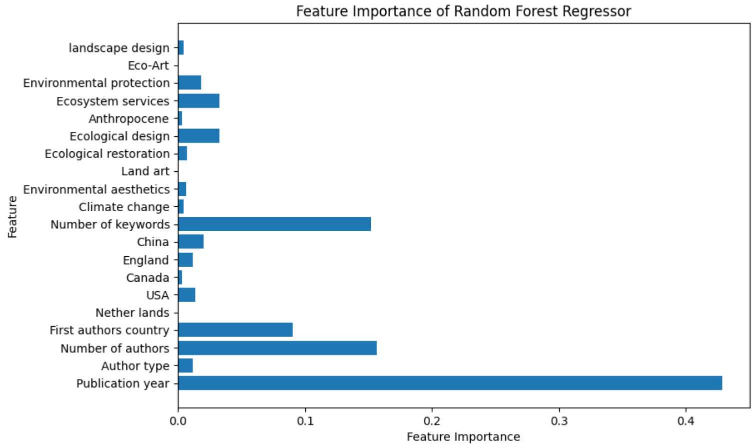

Figure 7. Importance of tree structures in random forest (key nodes used in random forest).

The target variable is the log of the number of citations per year plus one. We use the Random Forest with the optimal hyperparameter configuration. The horizontal axis represents the SHAP value, and the vertical axis contains the predictive attributes. Positive (negative) SHAP values indicate that the attribute increases (decreases) the predicted target variable. For each attribute, each dot represents the SHAP value for a specific instance, i.e., how that attribute contributes to the overall prediction of that instance. The dot color is proportional to the attribute's value, with lighter colors indicating higher values for that attribute. The number on the right side of the vertical axis is the average absolute SHAP value of each attribute (a proxy for the attribute's importance).

SHAP to Explain the Contribution of Each Attribute to the Overall Prediction of Instances

The SHAP method breaks down an instance's prediction into the individual contributions of each attribute, summed with a constant corresponding to the average prediction. Shapley values are derived from game theory literature and represent the marginal contribution of each feature (Shapley, 1953). The higher the absolute SHAP value, the greater its contribution to the prediction of a specific instance.

Figure 8 visualizes the SHAP values of all attributes in the instance using a beeswarm plot, a commonly used tool in interpretable machine learning methods based on Shapley values. The horizontal axis represents the SHAP value, while the vertical axis lists the attributes in the predictive model. A positive (or negative) SHAP value indicates that the attribute has a positive (or negative) contribution to the prediction of the target variable (i.e., annual citation count). Each point within an attribute represents the SHAP value of that specific instance, illustrating the individual contribution of that attribute to the overall prediction. Since each prediction is based on all attributes, the number of points shown for each attribute in the plot is the same. The color of the points is proportional to the attribute value, with lighter colors indicating higher attribute values. This visualization provides an intuitive understanding of whether a particular attribute generally has a positive or negative contribution to the target variable.

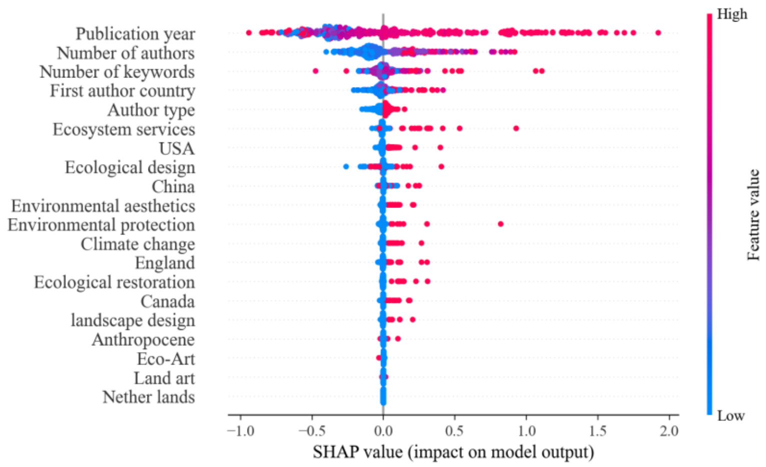

Figure 8. SHapley additive explanations—SHAP estimated for all attributes across all examples in the dataset.

As shown in **Figure 8**, the most critical attribute in predicting annual citation counts is the publication year. This result is quite intuitive: Earlier published papers tend to receive more citations than recently published ones, likely due to their longer exposure time in the academic community, which increases their visibility and potential impact. The number of authors is also an important predictive factor. The beeswarm plot further reveals a non-monotonic relationship between the number of keywords and the log-transformed annual citation count. This finding underscores the advantage of nonlinear methods in handling complex data distributions, as these methods better accommodate data diversity than traditional linear approaches.

We further explored the relationship between the number of keywords and the target variable using SHAP dependence plots. **Figure 9** shows the SHAP dependence plots for two attributes: (1) the number of authors and (2) the number of keywords in the paper. These plots are especially useful for analyzing numerical attributes such as publication year. The SHAP dependence plot clarifies the impact of changes in the number of keywords on model predictions, revealing its nonlinear relationship. For papers with fewer than two authors, this relationship shows negative growth, but after exceeding two authors, this negative relationship becomes less pronounced, although SHAP values begin to decline again once the number of keywords surpasses seven.

Typically, articles with more keywords may cover broader topics and research areas. Although most journals limit authors to selecting three to six keywords, we observed that a higher number of keywords can increase the visibility of a paper in search engines, and an increase in the number of keywords is often associated with a higher citation count. Papers with multiple authors tend to receive more citations than single-author papers, reflecting the importance of interdisciplinary collaboration in enhancing research impact. Large collaborative teams involving researchers with diverse expertise may drive innovation in the field and improve the likelihood of academic success.

The presence of keywords like “Ecosystem Services” and “Ecological Design” also increases the likelihood of further citations. These findings highlight the growing attention on topics related to ecosystem services and ecological design in academic research. Our results also indicate a rising interest in emerging research themes alongside traditional research areas.

The SHAP analysis aligns with the results of the random forest model, confirming that the top four influencing factors are the year of publication, the number of authors, the number of keywords, and the type of authorship. This further validates the importance of these factors in predicting paper citations.

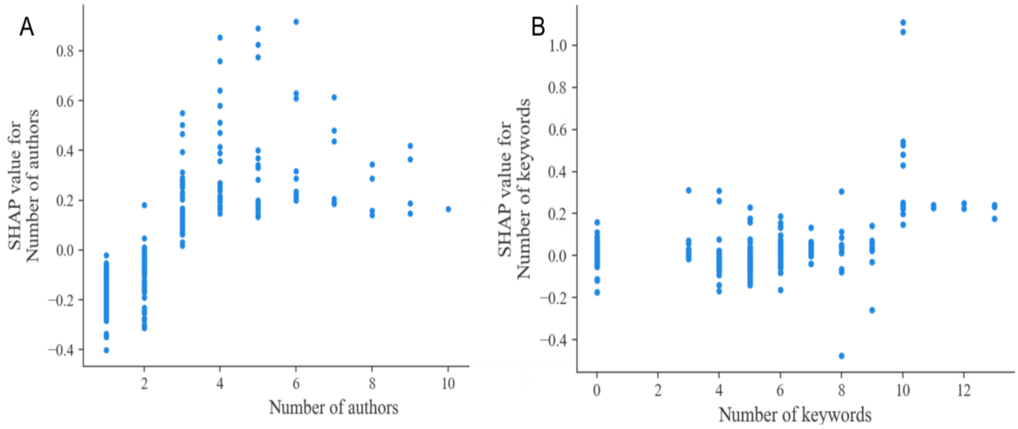

Figure 9. Partial dependence plots for the two most relevant numerical features in explaining predictions (in terms of SHAP values), **(a)** number of authors in the paper; **(b)** number of keywords in the paper.

5. Discussion

With the growing global concern over environmental issues (Lazăr et al., 2023), the influence of the field of ecological art has steadily increased. The rise of research topics such as "urban greenspace" highlights changes in the access and inequalities surrounding green spaces in the urbanization process (Wu et al., 2023). The role of ecological art in environmental education has also gained increasing attention, with artistic interventions proving effective in enhancing students' environmental awareness and behavior (Ison and Bramwell-Lalor, 2023). These findings not only showcase the potential of ecological art in education but also underscore the importance of interdisciplinary integration (Walshe et al., 2023). The diversity and interdisciplinary nature of ecological art research not only drive theoretical development but also provide new perspectives for solving practical environmental problems (Jiang et al., 2023).

Table 4 illustrates the ongoing attention and thematic evolution within ecological art research. In 2013, the theme of "management" saw a significant increase, while "health" and "science" have gradually become hotspots since 2015, expanding the scope of research to broader social and scientific fields. Terms like "urban greenspace" and "ecosystem" have risen in frequency since 2010, underscoring the potential of ecological art in urban planning and ecosystem conservation. The growth of terms such as "community" and "perception" indicates the importance of ecological art in promoting public engagement and community building.

Table 4. Time distribution and frequency analysis of common terms in ecological art research.

Term	Frequency	Year (Q1)	Year (Median)	Year (Q3)
land-use	6	2009	2010	2019
health	12	2008	2014	2022
vegetation	10	2010	2014	2018
quality	7	2008	2015	2023
management	19	2013	2016	2020
communities	7	2012	2016	2019
model	6	2012	2016	2021
ecosystems	7	2007	2017	2019
perception	5	2015	2017	2020
science	23	2016	2018	2021

Ecological art improves human-nature interactions through artistic forms, contributing to the sustainable management of ecosystems (Yuan and Kim, 2024). Through the analysis of highly cited papers, we identified influential research topics and authors in the field of ecological art. These studies not only advance theoretical development but also enhance public environmental awareness through practical applications.

The application of machine learning models in predicting citation rates of ecological art literature revealed key factors influencing academic impact, such as publication year, number of authors, number of keywords, and the strength of international collaboration networks. These findings provide researchers with a new perspective to evaluate the impact of their work and offer new strategies for the academic community to enhance research quality and influence. However, the predictive power of the models is limited by dataset size and feature selection, failing to capture all influential factors, thus presenting certain limitations.

With the continued development of interdisciplinary research, ecological art is demonstrating its significant potential in social innovation, environmental protection, and public engagement. Studies have shown that collaborations between art and science can evoke emotional resonance and foster a deeper understanding of the natural world, thereby enhancing environmental awareness and encouraging broader ecological action (Goralnik et al., 2017). Meanwhile, the rapid advancement of digital technology offers new possibilities for innovative expressions in ecological art, such as using visualization tools and interactive media to enhance public understanding and engagement with ecological issues. In the future, by deepening interdisciplinary collaboration and expanding research methods, ecological art is likely to play a greater role in addressing global environmental challenges, serving as a bridge between scientific communication and environmental conservation.

6. Conclusion

This study employed multidimensional bibliometric analysis and advanced machine learning techniques to systematically analyze the research evolution trends and frontier explorations in the field of ecological art. Over the past two decades, ecological design and sustainability have remained core topics, reflecting the growing global awareness of environmental crises. These findings not only enrich the theoretical framework of ecological art but also reveal its potential value in promoting social innovation and environmental protection practices. Emerging interdisciplinary trends, such as urban greenspace planning and ecological art education, indicate that ecological art is gradually permeating diverse application scenarios. Leveraging the predictive capabilities of machine learning models, this study further elucidated key factors influencing academic paper citation rates, including publication year, author size, keyword density, and the tightness of international collaboration networks. These findings deepen the understanding of the mechanisms behind academic impact and provide a scientific basis for optimizing future research planning and implementation strategies.

This study also acknowledges the geographic and cultural limitations of the dataset, which may affect the generalizability of the conclusions. Future research should expand to a broader range of geographic and cultural contexts to comprehensively assess the effectiveness of ecological art practices in different global social and cultural settings.

In summary, this study provides valuable academic insights into the field of ecological art, advancing its theoretical development and practical exploration, and pointing to future research directions. By strengthening interdisciplinary collaboration and innovative practices, ecological art will play an increasingly critical role in the global processes of environmental protection and social innovation.

Author contributions: Conceptualization, ZL and JHI; methodology, ZL and WY; software, YC; validation, YC and JHI; formal analysis, ZL; investigation, XQ; resources, XQ; data curation, WY; writing—original draft preparation, WY; writing—review and editing, JHI and WY; visualization, WY; supervision, WY; project administration, JHI; funding acquisition, ZL. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest

References

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of informetrics, 11(4), 959-975.

Brady, E., & Prior, J. (2020). Environmental aesthetics: A synthetic review. People and Nature, 2(2), 254-266.

Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

Brunner, C., Nigro, R., & Raunig, G. (2013). Post-media activism, social ecology and eco-art. Third Text, 27(1), 10-16.

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. Journal of the American Society for information Science and Technology, 57(3), 359-377.

Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. Journal of the American Society for information Science and Technology, 61(7), 1386-1409.

Cortes, C., & Vapnik, V. (1995). Support vector machine. Machine learning, 20(3), 273-297.

Cucuzzella, C. (2021). Making the invisible visible: Eco-art and design against the Anthropocene. Sustainability, 13(7), 3747.

Fix, E. (1985). Discriminatory analysis: nonparametric discrimination, consistency properties (Vol. 1). USAF school of Aviation Medicine.

Freilich, S., Kreimer, A., Meilijson, I., Gophna, U., Sharan, R., & Ruppin, E. (2010). The large-scale organization of the bacterial network of ecological co-occurrence interactions. Nucleic acids research, 38(12), 3857-3868.

Goralnik, L., Nelson, M. P., Gosnell, H., & Leigh, M. B. (2017). Arts and humanities inquiry in the long-term ecological research network: Empathy, relationships, and interdisciplinary collaborations. Journal of Environmental Studies and Sciences, 7, 361-373.

Guy, S., Henshaw, V., & Heidrich, O. (2015). Climate change, adaptation and eco-art in Singapore. Journal of Environmental Planning and Management, 58(1), 39-54.

Hassan-Montero, Y., De-Moya-Anegón, F., & Guerrero-Bote, V. P. (2022). SCImago Graphica: a new tool for exploring and visually communicating data. Profesional de la información, 31(5).

Ison, M., & Bramwell-Lalor, S. (2023). The arts in environmental education: connecting learners with their talents and nature. Environmental Education Research, 29(7), 964-979.

Jiang, Q., Wang, Z., Yu, K., Dou, Y., Fu, H., & Liang, X. (2023). The influence of urbanization on local perception of the effect of traditional landscapes on human wellbeing: A case study of a pondscape in Chongqing, China. Ecosystem Services, 60, 101521.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.

Lazăr, S., Dobrotă, D., Breaz, R. E., & Racz, S. G. (2023). Eco-Design of Polymer Matrix Composite Parts: A Review. Polymers, 15(17), 3634.

Leydesdorff, L., & Vaughan, L. (2006). Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment. Journal of the American Society for Information Science and Technology, 57(12), 1616-1628.

Lundberg, S. (2017). A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874.

Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. Profesional de la Información, 29(1).

Nassauer, J. I., Wang, Z., & Dayrell, E. (2009). What will the neighbors think? Cultural norms and ecological design. Landscape and Urban Planning, 92(3-4), 282-292.

Pan, X., Yan, E., Cui, M., & Hua, W. (2018). Examining the usage, citation, and diffusion patterns of bibliometric mapping software: A comparative study of three tools. Journal of informetrics, 12(2), 481-493.

Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S., & Mohammadian, A. K. (2020). Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. Accident Analysis & Prevention, 136, 105405.

Prášek, P. (2024). In What Sense Can Art Be Ecological? Art as an Event Flash of Nature According to Henri Maldiney. Filozofia, 79(4).

Samworth, R. J. (2012). Optimal weighted nearest neighbour classifiers.

Shapley, L. S. (1953). A value for n-person games. Contribution to the Theory of Games, 2.

Shapshay, S., Tenen, L., & Carlson, A. (2018). Environmental aesthetics, ethics, and ecoaesthetics. The Journal of Aesthetics and Art Criticism, 76(4), 399-410.

Silva, T. C., Braz, T., & Tabak, B. M. (2024). Mapping the landscape of energy markets research: A bibliometric analysis and predictive assessment using machine learning. Energy Economics, 136, 107698.

Simon, S. (2006). Systemic educational approaches to environmental issues: the contribution of ecological art. Systemic Practice and Action Research, 19(2), 143-157.

Song, Y. I. K. (2009). Community participatory ecological art and education. International Journal of Art & Design Education, 28(1), 4-13.

Tsevreni, I. (2022). Encounters among Environmental Education and Eco-Art in the Anthropocene. International Journal of Art & Design Education, 41(3), 482-495.

Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523-538.

Walshe, N., Bungay, H., & Dadswell, A. (2023). Sustainable Outdoor Education: Organisations Connecting Children and Young People with Nature through the Arts. Sustainability, 15, 3941.

Weder, F., & Voci, D. (2021). From ignorance to resonance: analysis of the transformative potential of dissensus and agonistic deliberation in sustainability communication. International Journal of Communication, 15, 24.

Wu, D., & Ning, S. (2018). Dynamic assessment of urban economy-environment-energy system using system dynamics model: A case study in Beijing. Environmental research, 164, 70-84.

Wu, S., Chen, B., Webster, C., Xu, B., & Gong, P. (2023). Improved human greenspace exposure equality during 21st century urbanization. Nature Communications, 14(1), 6460.

Xie, H., Zhang, Y., Wu, Z., & Lv, T. (2020). A bibliometric analysis on land degradation: Current status, development, and future directions. Land, 9(1), 28.

Yuan, J., & Kim, C. S. (2024). The Ecological Design of Marine Urban Green Space Plant Landscaping Based on the Concept of Sustainability. Plants, 13(7), 923.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology, 67(2), 301-320.