Biodiversity Informatics, 19, 2025, pp. 33-45

WELCOME TO THE MACHINE: A PAN-CONTINENTAL OVERVIEW OF MACHINE LEARNING APPLICATIONS IN ECOLOGY AND CONSERVATION

Juliano A. Bogoni[1*], Derick Victor de Souza Campos[2*], Claumir C. Muniz[2], Manoel dos Santos-Filho[1], Jéssica Eloá Poletto[3]

[1] *Universidade do Estado de Mato Grosso, Centro de Pesquisa de Limnologia, Biodiversidade e Etnobiologia do Pantanal, Programa de Pós-Graduação em Ciências Ambientais, Laboratório de Mastozoologia, Cáceres, MT, Brazil.*
[2] *Universidade do Estado de Mato Grosso, Centro de Pesquisa de Limnologia, Biodiversidade e Etnobiologia do Pantanal, Programa de Pós-Graduação em Ciências Ambientais, Laboratório de Investigação Ambiental do Pantanal Norte – LIPAN, Cáceres, MT, Brazil.*
[3] *Faculdade de Ciências Médicas da Universidade Estadual de Campinas, Piracicaba, SP, Brasil.*

Abstract. Machine-learning emerged as an excellent alternative to understanding ecological patterns and processes at different spatiotemporal scales. The study aimed to offer a global overview of the *status quo* on the use of machine-learning in ecology and conservation globally. Using keywords in the Scopus engine, we indexed all publications in ecology and conservation using machine-learning. We employed descriptive statistics and regressions models to provide an overview and predict geopolitical patterns. The majority of manuscripts were condensed in economically affluent countries, such as the United States (USA) and China (CHN) which together amount to 91 (36.8%) studies. There is a spatial aggregation in the authors' affiliations, once 182 (73.7%) studies derived from both Nearctic and Palearctic teams, whereas Tropical teams published 65 (26.3%) manuscripts and the most-cited papers also are concentrated in northern regions. In ecology and conservation, machine-learning first appear in the literature in 2003. Since then, the number of publications has increased exponentially, from 09 manuscripts in 2010, to 120 manuscripts 10 years later. Most studies (N = 173; 70.1%) focused on landscape and vertebrate ecology. The primary aims of the publications were widely variable but strongly adherent to providing the best-information on both landscape-scale classifications and species distribution modelling. The manuscripts encompass different methods, from maximum entropy to boosted regression trees and random forest, sometimes using a range of deep-learning architectures. Finally, the predictive variables (i.e., mammal diversity and *per capita* GDP) do not exert significant influences on the number of studies published. Finally, we recommend a well-structured and collaborative agenda aiming to integrate less-resourced countries into scientific advancements, fostering more equitable and effective responses to global environmental challenges.

Keywords: data analysis, global-scale, informatics, numerical ecology, tropical forest.

*corresponding authors: bogoni.ja@gmail.com | derick@unemat.br

Introduction

Ecology is a relatively young science that fundamentally seeks to understand the causes and consequences (i.e., processes) of diversity patterns and species distributions across global environments (Brown, 1995; Haeckel, 1866). Universal features of ecological processes exhibit mathematical properties that are inherently non-linear and complex, historically addressed only through mathematical approximations (Bogoni et al., 2019; Conway, 1977; May, 1976). Since the 1920s, explicit models as in Volterra (1926), Lotka (1925), Elton (1924) have been employed in ecology to predict and describe synchronization mechanisms in animal behaviour as Araujo et al. (2013), predator-prey relationships and dynamics in Sherratt et al. (1997), Kar et al. (2010), host-parasitoid interactions in Hassell (2000), and various other ecological dynamics, such as seed predation and dispersal (Bogoni et al., 2019) and species distribution (Bogoni and Tagliari, 2021; Tagliari et al., 2023).

As global biodiversity faces unprecedented and widespread declines in modern history (Bogoni et al., 2020; Ceballos et al., 2020), addressing contemporary ecological challenges—such as biodiversity loss, climate change, and the growing demand for ecosystem services—has become a pressing priority for ecologists (Rammer and Seidl, 2019). This issue is particularly critical in the global tropics, where habitat loss is most severe. For example, within just half a decade in the early 21st century, from 2000 to 2005, tropical forests in both wet and dry regions experienced a staggering loss of approximately 475,000 km² of forest cover (around 50%) (Hansen et al., 2010). Thus, understanding how to mitigate natural habitat degradation processes, particularly across tropical landscapes, is imperative for safeguarding biodiversity on a global scale.

Machine-learning methods have emerged as an excellent alternative for predicting and understanding ecological patterns and processes across various spatiotemporal scales. These methods consist of computational algorithms designed to analyse the often non-linear structures of complex data, thereby generating predictive models based on estimated patterns (Rammer and Seidl, 2019; Olden et al., 2008). In comparison to classical statistical approaches (e.g., regression models), machine learning relies on computational power to identify and model complex relationships, prioritizing predictive accuracy over the estimation of parameters and confidence intervals (Breiman, 2001). Positioned at the intersection of statistics and computer science—and driven by advances in artificial intelligence and data science—the application of machine learning in ecology and conservation has emerged as a rapidly expanding field (Rammer and Seidl, 2019; Jordan and Mitchell, 2015).

However, the conducted search reveals a lack of studies examining how machine-learning methods are being applied in ecology and conservation or providing a pan-continental overview of how these methods can assist ecologists in ensuring biodiversity persistence across global landscapes. Additionally, disparities between local and global scales have historically biased scientific advancements (Bogoni et al., 2021; Hughes et al., 2021). Modern ecological methods, such as machine learning, tend to be more widely utilized in regions with more affluent economies, where access to advanced computational resources is readily available (e.g., Bogoni et al., 2021). In contrast, countries where scientific development is still in its early stages—and which often host the highest levels of biodiversity—have likely made less use of these computational advancements. This limitation is particularly concerning when it comes to understanding ecological patterns and processes and addressing the severe biodiversity crises in these regions.

Our primary aim was to provide a global pictorial overview of the use of machine learning methods in ecology and conservation by assembling a comprehensive database through an extensive literature review. Specifically, we sought to: (1) quantitatively and qualitatively evaluate how machine learning methods have been applied in peer-reviewed ecology and conservation publications across a pancontinental scale; (2) assess the research questions, focal areas, and taxonomic groups addressed in studies involving machine learning within the disciplines of ecology and conservation; and (3) discuss how countries with lower investment in science or less frequent use of modern analytical methods can leverage these theoretical and analytical advancements to address their environmental and ecological challenges. We tested the following hypotheses: (1) the vast majority of publications originate from authors affiliated with economically affluent countries; (2) the application of machine learning in ecology and conservation is predominantly focused on species distribution models and landscape ecology; and (3) at a predictive level, a country's per capita income positively correlates with the number of its publications, while mammal diversity (used as a proxy for country-scale biodiversity) is negatively correlated with publication output, as poorer countries often harbor greater biodiversity than wealthier nations. Our goal is to provide a global synthesis of how machine learning has been integrated into ecological and conservation research. By identifying geographic, taxonomic, and methodological patterns in the literature, it discusses existing disparities in scientific output and highlights opportunities for greater inclusion of biodiversity-rich yet underrepresented countries.

Material and Methods

Our study is characterized as data gathering research, with a posteriori application of plans with descriptive statistical analysis method and mapping. Thus, in early May (04th-May-2023), we used the Scopus search engine[1] to index and aggregate all publications (published or in press) in ecology and conservation that used machine-learning methods. To do so, we used a conjunct of keywords allied to AND/OR operators, as follows: "machine learning" AND/OR "deep learning" AND "ecology" AND/OR "conservation (similar to Magioli et al., 2015). This allowed us to compile the number of scientific publications that included these terms in their manuscript titles, abstracts, and keywords.

After capturing the studies, we extracted the following information: (1) the total number of studies; (2) the publication year of each manuscript; (3) the taxonomic group(s) studied or the primary approach (e.g., landscape ecology); (4) the country of primary affiliation of the first author; (5) the city-based geolocation of the study in latitudinal and longitudinal decimal degrees; (6) the set of study keywords; (7) the study title; and (8) the total number of citations for each manuscript. Documents with incomplete information, missing location, author, or keywords, were not consider in our analysis.

Once the database was consolidated, we used descriptive statistics and mapping to provide a pancontinental overview of the state of machine-learning applications in ecology and conservation. We analysed the number of published articles, the annual publication trends, and the distribution of articles by country. Additionally, we generated a word-cloud based on the top-100 keywords provided by the authors. We also explored the proportion of taxonomic groups or approaches studied in a pictorial synthesis of studies. Then, we identified the top-cited manuscripts (i.e., all those with more than 100 citations) and provided detailed insights into their primary aims, methods, and the specific sub-disciplines of ecology and conservation they addressed.

The data analyses were performed in R v.4.0.5 (R Core Team, 2023) using *stats* (R Core Team, 2023), *sp* (Bivand et al., 2013; Pebesma and Bivand, 2005), *maptools* (Bivand and Lewin-Koh, 2021) and the dependent R-package. Additionally, we utilized data from the International Union for Conservation of Nature (IUCN) Red List of Threatened Species[2] to obtain country-level mammal richness as a proxy for national biodiversity. Mammals were chosen for this analysis as they represent one of the most extensively studied and well-documented taxonomic groups worldwide (Bogoni et al., 2021).

We also sourced data from the World Bank[3] to obtain country-level per capita Gross Domestic Product, hereafter, *GDP*. Both variables mentioned above were subsequently used to predict the number of studies published per country. To do so, we used a generalized linear regression model (*glm*), using Poisson distribution under the data variance given that the response variable is a count (non-negative integers) (Dobson, 1990), correcting the predictive data asymmetry by $\log x + 1$. Thus, the *glm* was performed as follows:

$$Studies_{(i\text{th country})} \sim \log(MammalRichness_{(i\text{th country})}) + \log(GDP_{(i\text{th country})})$$

where $Studies_{(i\text{th country})}$ is the number of studies published in the country, $\log(MammalRichness_{(i\text{th country})})$ is the natural log of mammalian species richness in the country, and $\log(\mathrm{GDP}_{(i\text{th country})})$ is the natural log of the Gross Domestic Product for country.

Then, deriving a bivariate effect plot for both predictive variables *vs*. number of studies. The statistics of *glm* model were presented for z- and p-values, estimates and standard errors. The regression analysis and bivariate plots were performed in R v.4.0.5 (R Core Team, 2023) based on the *stats* R-package (R Core Team, 2023).

Results

Our literature search identified 251 documents related to machine learning and/or deep learning with a focus on ecology and conservation. After applying an intuitive filter to exclude documents with incomplete information, our study was able to analyse pancontinental trends based on 247 manuscripts. This overview revealed that the vast majority of published manuscripts were concentrated in economically affluent countries, such as the United States (USA) and China (CHN), which together accounted for 91 studies (36.8%) (Fig. 1A). Furthermore, there was a notable spatial concentration in the primary affiliations of corresponding authors: 182 studies (73.7%) originated from teams in the Nearctic and Palearctic regions, particularly Europe, while Neotropical and Paleotropical teams

[1] https://www.scopus.com/home.uri.

[2] https://www.iucnredlist.org/.

[3] https://www.worldbank.org/en/home.

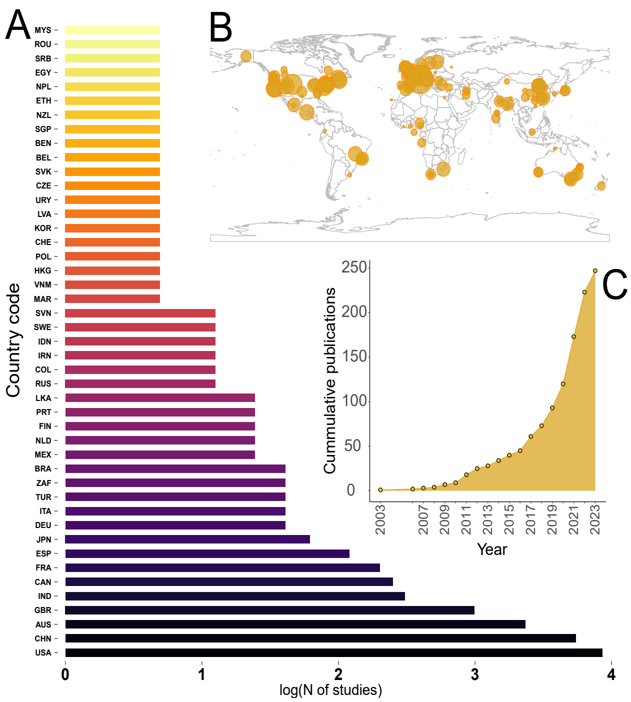

Figure 1. Number of studies (log-scaled) involving machine-learning in ecology and conservation among countries (A), its global geo-distribution and citation number (log-scaled; B), and the cumulative publications over the years (C).
Source: Original search results.

contributed 65 studies (26.3%). Of these, only 11 studies (4.5%) were published in the biodiversity-rich Neotropical realm (Fig. 1B). Similarly, the most-cited papers also predominantly originated from northern countries, mirroring the geographic bias in publication origins (Fig. 1B).

In terms of publication accumulation, the machine-learning publications in ecology and conservation first appear on literature in 2003. Yet, this topic increased exponentially since the 2010s (Fig 1B). In 2010 this overview indicated nine published manuscripts, whereas 10-yrs later this number reached 120 publications (Fig. 1C), an increment of 111 publications. Moreover, in 2021 and 2022, these techniques seem to have become a hot-topic and consolidating more than 100 (>40.5%) publications in ecology and conservation (≥50 manuscripts per year; Fig 1C). Across the countries, most studies (N = 173; 70.1%) are fundamentally focused on landscape ecology (which here includes species distribution), bird ecology and conservation, wildlife ecology with more than one taxonomic group, mammal ecology and conservation, plant ecology and conservation, fish ecology and conservation, and invertebrate ecology and conservation (Fig 2A). Proportionally, these areas respond to 28, 10, 10, 7, 6, 5, and 5%, respectively, of all approaches (Fig. 2B).

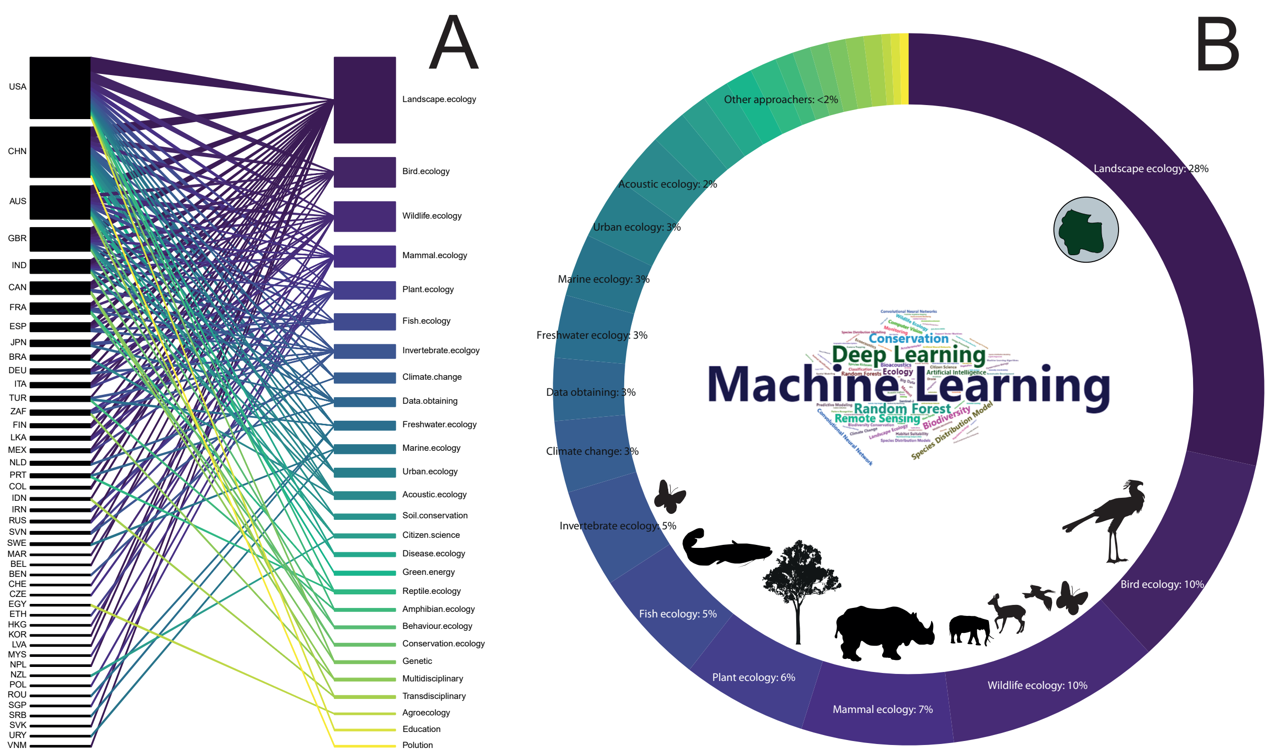

Figure 2. (A) Network between publication-based country vs. major approach of machine-learning studies in ecology and conservation; and (B) Percentage (and main authors keywords) of approaches in ecology and conservation using machine-learning methods.
Source: Original search results

Analysis of the top-cited manuscripts (N = 14; 6%), totaling 9887 citations, reveals that the primary objectives of these publications varied significantly, yet consistently focused on providing high-quality information on both landscape-scale classifications and species distribution modeling (Table 1). Whereas, some top-cited manuscripts provide comparisons between machine-learning algorithms in solving an ecological or conservation issue (Table 1). These 100% northern-based researches, published in high-impact journals, encompasses different methods, from maximum entropy models to boosted regression trees and random forest, and sometimes uses a vast gamma of deep learning architectures (e.g., AlexNet, NiN, VGG, GoogLeNet, and ResNet; see Norouzzadeh et al. (2018); Table 1).

Finally, the mammal diversity (species richness) exerts significant influences on the number of country-scale studies, whereas per capita GDP does not. Mammal diversity had a positive and significant tendency [*z-score* = 2.362; estimates = 0.41; *se* = 0.17; *p* = 0.02] while per capita GDP had a positive but non-significant tendency [*z* = 1.357; *estimate* = 0.14; *se* = 0.10; *p* = 0.175] in influencing the patterns of global distribution of studies (Fig. 3).

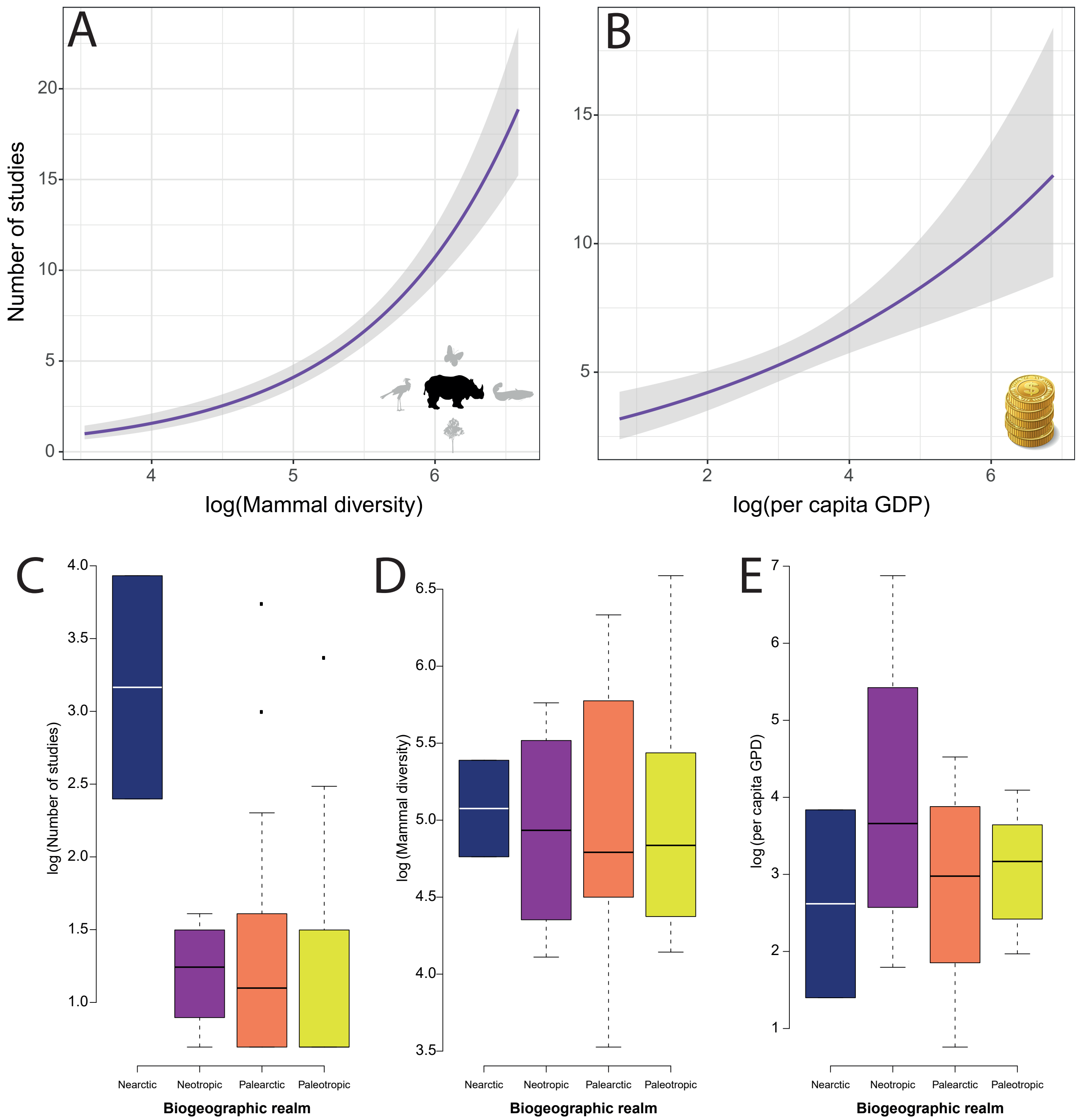

Figure 3. Bivariate plot between the number of studies per country vs. mammal diversity (A) and *per capita* GDP (B), and distribution of per country studies (C), mammal diversity (D), and *per capita* GDP (E) across the major biogeographic realms.
Source: Original search results

This model presented deviance and goodness-of-fit (G-statistic) with $p < 0.001$, low overdispersion (OD = 0.346) and diagnostic residuals plots suggest that the distribution is satisfactory in terms of residuals.

Discussion

The complexities of ecosystems and the challenges of managing and protecting biodiversity require constant computational advances (Chaves, 2013), in which, machine-learning approaches, poses as strong candidate to help ecologists perform sophisticated data analysis. Machine-learning algorithms have a high capacity of processing large datasets, identify patterns, and make predictions that can support in decision-making and resource allocation (Rammer and Seidl, 2019). Our main results showed that machine-learning has been widely used in habitat mapping particularly within landscape ecology, predictive modelling, and other ecological data analysis devoted to understanding patterns and drivers of diversity and species distribution, especially in vertebrate and plant ecology. The integration of satellite imagery and machine-learning provides a powerful tool for mapping and monitoring habitats, tracking deforestation, and assessing land-use changes (McLaren et al., 2018; Kampichler et al., 2010). Moreover, machine-learning can predict species distribution, movement patterns, and migration routes, supporting in conservation planning (Pittman and Brown, 2011). In the context of ecological data analysis, machine-learning approaches were used to process large datasets derived from field observations and sensors (e.g., Christin et al., 2019; Stowell and Plumbley, 2014) as to identify complex relationships and patterns in ecological systems.

These main approaches address important ecological and conservation challenges. Both landscape ecology and predictive modelling are essential tools for understanding and responding to the accelerating changes in habitats and climate. Ecologists are increasingly required to improve their analytical skills to better predict biodiversity and species distribution responses to highly modified landscapes and pervasive climatic changes. Understanding the process that influences the species distributions is a critical issue in ecology and conservation (Franklin, 2009; Hutchinson, 1957), especially for species under threats (Phillips et al., 2004). In this context, machine-learning algorithms have become widely used in biogeography, ecology, and conservation biology to estimate the relationship between species occurrences and environmental variables (Elith et al., 2011; Elith and Leathwick, 2009; Franklin, 2009).

For instance, species distribution models (SDMs) enable researchers to explore key questions in conservation, ecology, and evolution, such as: (1) determining priority areas for conservation and contributing to the protection of species (Faleiro et al., 2013; García, 2006; Chen and Peterson, 2002); (2) understanding invasive process of species (Giovanelli et al., 2008; Ficetola et al., 2007); (3) generating ecocultural niche modelling that reflects ecological influences on past human culture distributions (Banks et al., 2008); (4) proposing past distributions of species (Carnaval and Moritz, 2008; Hugall et al. 2002); and (5) predicting future distributions of species under changing of climatic and environmental characteristics (Tagliari et al., 2023; Bogoni and Tagliari, 2021; Siqueira and Peterson, 2003).

Our results on the use of machine learning methods in ecology and conservation have shown an exponential increase since 2017. In 2010, our search identified nine published manuscripts, whereas ten years later, this number had increased by 111 publications. Furthermore, our findings reveal a strong geographic concentration of ecological studies using machine learning approaches in northern countries. Historically, the vast majority of scientific publications are located in these regions. For instance, only in 2020, the US authors signed about ~755,000 scientific publications, while Brazilian authors reached ~100,000 publications on Scopus (see search logs[4]), representing 86.6% less. These values are reflected in the context of machine learning approaches in ecology, as our results indicate that Brazilian authors published 92.0% fewer manuscripts than their counterparts in the United States.

Brazil is undoubtedly Earth's most biodiverse country, harbouring the largest set of known and unknown species (Moura and Jetz, 2021), but still presenting glaring knowledge gaps such as the multispecies Wallacean shortfalls (Bogoni et al., 2021). Addressing these discrepancies requires substantial investment in scientific research. In contrast, with the exception of 2022 – when investments in science were temporarily resumed –, the Brazilian government has systematically deepened budget cuts for science since 2017 (see Angelo, 2017; Escobar, 2017).

In terms of machine-learning, we advocated that the investments should be addressed to form and solidify human resources and increase the computational power of institutions, especially all those located across the countryside, given that both these factors are critical to performing high-complex analysis. Moreover, our findings based on proxies of biodiversity and financial resources in

[4] https://www.scopus.com/results/results.uri?st1=United+States&st2=&s=AFFILCOUNTRY%28United+States%29&limit=10&origin=resultslist&sort=plf-f&src=s&sot=b&sdt=cl&sessionSearchId=1f06790275-90dc7c447bf72ed3d85980&yearFrom=2020&yearTo=2020 | https://www.scopus.com/results/results.uri?st1=United+States&st2=&s=AFFILCOUNTRY%28Brazil%29&limit=10&origin=searchbasic&sort=plf-f&src=s&sot=b&sdt=b&sessionSearchId=1f0679027590dc7c44-7bf72ed3d85980&yearFrom=2020&yearTo=2020

predicting the pancontinental patterns in the application of machine-learning in ecology showed no significant influence on the number of studies published globally. However, mammal diversity showed a positive tendency whereas *per capita* GDP had a negative tendency in potentially influencing the distribution patterns of published studies. Economically affluent regions — despite having low biodiversity when compared to tropical countries — typically account for the vast majority of scientific projects research fundings, while biodiversity-rich countries tend to have lower per-capita incomes. Furthermore, despite the highly-recognised quality of researches led by global south scientists—often conducted under highly adverse field and financial conditions and whose legacy chips away at much of our knowledge gaps (Bogoni et al., 2021)—rarely is a tropical country a protagonist in terms of publications number.

In this context, for example, Canada and China together harbour a total of 753 mammalian species (195 and 558, respectively; Smith and Xie, 2013; Banfield, 1974), whereas Brazil has 775 mammal species described (Abreu et al., 2022). Therefore, a kind of *per capita* publication in relation to mammalian species reaches 0.07 and 0.006, respectively, representing 13-fold more publications lead by these northern countries weighted by its biodiversity when compared to Brazil. Yet, our results indicate that this is not an exclusivity for Neotropical countries, such as Brazil. Excepting for Australia, Paleotropical countries amount to 25 studies, representing only ~10% of the total. Moreover, the top-10 richest countries in terms of GDP *per capita* (70% of them located in the Nearctic of Palearctic) amount by 38.1% of publications ($N = 94$), while the top-10 poorer countries (60% located in the Paleotropical region) signed only 10.1% of scientific publications ($N = 27$) focused in machine-learning in ecology.

Our initial hypotheses were partially corroborated. Our results indicated that the vast majority of publications derive from authors affiliated in economically affluent countries, and the major focus of machine-learning in ecology and conservation was widely applied to species distribution models and landscape ecology. In predictive terms, although there is a clear tendency for richer countries to publish more than poorer ones, and a negative relationship between biodiversity and publication output, mammal diversity showed a positive and significant trend, while per capita GDP showed a positive but non-significant trend.

Based on our global overview of the use of machine learning methods in ecology and conservation, we conclude that a wide range of ecological and conservation issues has already been addressed using these techniques. Machine learning thus represents a strong promise for the coming years, contingent only on the availability of human and financial resources. This promise can therefore be solidified by a fine-tuned agenda aiming to initiate a discussion concerning countries with lower investments in scientific research and those that employ fewer modern analytical methods. Such an agenda would also allow for an exploration of how these nations can benefit from theoretical and analytical advancements to address their persistent environmental and ecological challenges. Finally, collaboration between scientific communities across different economic levels could enable more equitable and effective responses to shared global challenges.

Declarations

Funding - Not Applicable

Conflicts of interest/Competing interests

The authors have no conflicts of interest to declare.

Availability of data and material

Data used here are widely available at Scopus search engine (https://www.scopus.com/home.uri).

Code availability

See Supporting Information Code 1

Author Contributions Statement

JAB and DVSC: conceptualization, data acquisition, data analysis and figures, writing & revising the original draft; DVSC, CCM, MSF & JEP: conceptualization, review and editing.

References

Abreu, E.F., Casali, D., Costa-Araújo, R., Garbino, G.S.T., Libardi, G.S., Loretto, D., Loss, A.C., Marmontel, M., Moras, L.M., Nascimento, M.C., Oliveira, M.L., Pavan, S.E., Tirelli, F.P., 2022. Lista de Mamíferos do Brasil (2022-1) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.7469767

Angelo, C., 2017. Scientists plead with Brazilian government to restore funding. Nature 550: 166–167. https://doi.org/10.1038/nature.2017.22757

Araujo, S.B.L., Rorato, A.C., Perez, D.M., Pie, M.R., 2013. A spatially explicit model of synchronization in fiddler crab waving displays. PLoS One 8(3): e57362. https://doi.org/10.1371/journal.pone.0057362

Banfield, A.W.F., 1974. The Mammals of Canada. University of Toronto Press.

Banks, W.E., d'Errico, F., Peterson, A.T., Vanhaeren, M., Kageyama, M., Sepulchre, P., Ramstein, G., Jost, A., Lunt, D., 2008. Human ecological niches and ranges during the LGM in Europe derived from an application eco-cultural niche modeling. J. Archaeol. Sci. 35: 481–491. https://doi.org/10.1016/j.jas.2007.05.011

Bogoni, J.A., Navarro, A.B., Graipel, M.E., Peroni, N., 2019. Modeling the frugivory of a plant with inconstant productivity and solid interaction with relictual vertebrate biota. Ecol. Model. 408: 108728. https://doi.org/10.1016/j.ecolmodel.2019.108728

Bogoni, J.A., Tagliari, M.M., 2021. Potential distribution of piscivores across the Atlantic Forest: From bats and marsupials to large-bodied mammals under a trophic-guild viewpoint. Ecol. Inform. 64: 101357. https://doi.org/10.1016/j.ecoinf.2021.101357

Bogoni, J.A., Peres, C.A., Ferraz, K.M.P.M.B., 2021. Medium- to large-bodied mammal surveys across the Neotropics are heavily biased against the most faunally intact assemblages. Mammal Rev. 52(2): 221–235. https://doi.org/10.1111/mam.12274

Bivand, R.S., Pebesma, E., Gomez-Rubio, V., 2013. Applied Spatial Data Analysis with R. 2nd ed. Springer, NY, USA.

Bivand, R.S., Lewin-Koh, N., 2021. Maptools: Tools for Handling Spatial Objects. R package version 1.1-1. https://CRAN.R-project.org/package=maptools (accessed 2 Mar 2023).

Breiman, L., 2001. Statistical modeling: the two cultures. Stat. Sci. 16: 199–231.

Brown, J.H., 1995. Macroecology. 1st ed. University of Chicago Press, Chicago, IL, USA.

Carnaval, A.C., Moritz, C., 2008. Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. J. Biogeogr. 35: 1187–1201. https://doi.org/10.1111/j.1365-2699.2007.01870.x

Ceballos, G., Ehrlich, P.R., Raven, P.H., 2020. Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. Proc. Natl. Acad. Sci. USA 117: 13596–13602. https://doi.org/10.1111/j.1365-2699.2007.01870.x

Chaves, J., 2013. The problem of pattern and scale in ecology: what have we learned in 20 years? Ecol. Lett. 16(1): 4–16. https://doi.org/10.1111/ele.12048

Chen, G., Peterson, A.T., 2002. Prioritization of areas in China for the conservation of endangered birds using modeled geographical distributions. Bird Conserv. Int. 12: 197–209. https://doi.org/10.1017/S0959270902002125

Christin, S., Hervet, É., Lecomte, N., 2019. Applications for deep learning in ecology. Methods Ecol. Evol. 10(10): 1632–1644. https://doi.org/10.1111/2041-210X.13256

Conway, G.R., 1977. Mathematical models in applied ecology. Nature 269: 291–297.

Dobson, A.J., 1990. An Introduction to Generalized Linear Models. Chapman and Hall.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40: 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17: 43–57.

Elton, C.S., 1924. Periodic fluctuations in the numbers of animals: their causes and effects. Br. J. Exp. Biol. 2: 119–163. https://doi.org/10.1111/j.1472-4642.2010.00725.x

Escobar, H. 2017. In Brazil, researchers struggle to fend off deepening budget cuts. Scienceinsider. Available at: https://www.science.org/content/article/brazil-researchers-struggle-fend-deepening-budget-cuts

Faleiro, F.V., Machado, R.B., Loyola, R.D., 2013. Defining spatial conservation priorities in the face of land-use and climate change. Biol. Conserv. 158: 248–257. https://doi.org/10.1016/j.biocon.2012.09.020

Ficetola, G.F., Thuiller, W., Miaud, C., 2007. Prediction and validation of the potential global distribution of a problematic alien invasive species – the American bullfrog. Divers. Distrib. 13: 476–485. https://doi.org/10.1111/j.1472-4642.2007.00377.x

Franklin, J., 2009. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press.

García, A., 2006. Using ecological niche modeling to identify diversity hotspots for the herpetofauna of Pacific lowlands and adjacent interior valleys of Mexico. Biol. Conserv. 130: 25–46. https://doi.org/10.1016/j.biocon.2005.11.030

Giovanelli, J.G.R., Haddad, C.F.B., Alexandrino, J., 2008. Predicting the potential distribution of the alien invasive American bullfrog (*Lithobates catesbeianus*) in Brazil. Biol. Invasions 10: 585–590. https://doi.org/10.1007/s10530-007-9154-5

Haeckel, E., 1866. Generelle Morphologie der Organismen. G. Reimer, Berlin, Germany.

Hansen, M.C., Stehman, S.V., Potapov, P.V., 2010. Quantification of global gross forest cover loss. Proc. Natl. Acad. Sci. USA 107(19): 8650–8655. https://doi.org/10.1073/pnas.0912668107

Hassell, M.P., 2000. Host-parasitoid population dynamics. J. Anim. Ecol. 69: 543–566. https://doi.org/10.1046/j.1365-2656.2000.00445.x

Hugall, A., Moritz, C., Moussalli, A., Stanisic, J., 2002. Reconciling paleo distribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). Proc. Natl. Acad. Sci. USA 99: 6112–6117. https://doi.org/10.1073/pnas.092538699

Hughes, A.C., Orr, M.C., Ma, K., Cotello, M.J., Waller, J., Provoost, P., Yang, Q., Zhu, C., Qiao, H., 2021. Sampling biases shape our view of the natural world. Ecography 44(9): 1259–1269. https://doi.org/10.1111/ecog.05926

Hutchinson, G.E., 1957. Concluding remarks. Cold Spring Harb. Symp. Quant. Biol. 22: 415–427.

Jordan, M.I., Mitchell, T.M., 2015. Machine learning: trends, perspectives, and prospects. Science 349(6245): 255–260. https://www.science.org/doi/10.1126/science.aaa8415

Kampichler, C., Wieland, R., Calmé, S., Weissenberger, H., et al., 2010. Classification in conservation biology: A comparison of five machine-learning methods. Ecol. Inform. 5(6): 441–450. https://doi.org/10.1016/j.ecoinf.2010.06.003

Kar, T.K., Chakraborty, K., Pahari, U.K., 2010. A prey–predator model with alternative prey: Mathematical model and analysis. Can. Appl. Math. Q. 18(2): 137–168.

Lotka, A.J., 1925. Elements of Physical Biology. Williams and Wilkins Co., Baltimore, MD, USA.

McLaren, J.D., Buler, J.J., Schreckengost, T., Smolinsky, J.A., et al., 2018. Artificial light at night confounds broad-scale habitat use by migrating birds. Ecol. Lett. 21(3): 356–364. https://doi.org/10.1111/ele.12902

Magioli, M., Ribeiro, M.C., Ferraz, K.M.P.M.B., Rodrigues, M.G., 2015. Thresholds in the relationship between functional diversity and patch size for mammals in the Brazilian Atlantic Forest. Anim. Conserv. 18(6): 499–511. https://doi.org/10.1111/acv.12201

Magurran, A.E., 2004. Measuring Biological Diversity. Blackwell Publishing, Malden, MA, USA.

May, R.M., 1976. Thresholds and breakpoints in ecosystems with a multiplicity of stable states. Nature 269: 471–477. https://doi.org/10.1038/269471a0

Moura, M.R., Jetz, W., 2021. Shortfalls and opportunities in terrestrial vertebrate species discovery. Nat. Ecol. Evol. 5: 631–639. https://doi.org/10.1038/s41559-021-01411-5

Olden, J.D., Lawler, J.J., Poff, N.L., 2008. Machine learning methods without tears: a primer for ecologists. Q. Rev. Biol. 83: 171–193. https://www.journals.uchicago.edu/doi/10.1086/587826

Pebesma, E.J., Bivand, R.S., 2005. Classes and methods for spatial data in R. R News 5(2). https://cran.r-project.org/doc/Rnews/ (accessed 2 Mar 2023).

Pittman, S.J., Brown, K.A., 2011. Multi-scale approach for predicting fish species distributions across coral reef seascapes. PLoS One 6(5): e20583. https://doi.org/10.1371/journal.pone.0020583

R Core Team, 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing. http://www.R-project.org/ (accessed 2 Mar 2023).

Rammer, W., Seidl, R., 2019. Harnessing deep learning in ecology: An example predicting bark beetle outbreaks. Front. Plant Sci. 10: 1327. https://doi.org/10.3389/fpls.2019.01327

Sherratt, J.A., Eagan, B.T., Lewis, M.A., 1997. Oscillations and chaos behind predator-prey invasion: mathematical artifact or ecological reality? Philos. Trans. R. Soc. Lond. B Biol. Sci. 352: 21–38. https://doi.org/10.1098/rstb.1997.0003

Siqueira, M.F., Peterson, A.T., 2003. Consequences of global climate change for geographic distributions of Cerrado tree species. Biota Neotrop. 3(2): 1–14. https://doi.org/10.1590/S1676-06032003000200005

Smith, A.T., Xie, Y., Hoffmann, R.S., Lunde, D., MacKinnon, J., Wilson, D.E., Wozencraft, W.C., Gemma, F. (Illustrator), Wang, S. (Honorary Editor), 2013. Mammals of China. Princeton University Press. (Princeton Pocket Guides).

Stowell, D., Plumbley, M.D., 2014. Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. PeerJ 2: e488. https://doi.org/10.7717/peerj.488

Tagliari, M.M., Bogoni, J.A., Blanco, G.D., Cruz, A.P., Peroni, N., 2021. Disrupting a socio-ecological system: could traditional ecological knowledge be the key to preserving the Araucaria Forest in Brazil under climate change? Clim. Change 176: 1–20. https://doi.org/10.1007/s10584-022-03477-x

Volterra, V., 1926. Variazioni e fluttuazioni del numero d’individui in specie animali conviventi. Mem. Acad. Lincei Roma 2: 31–113.

Table 1. Journal, the impact factor (IF), year of publication (yr), authors, main country, number of citations, primary aim, machine-learning methods or framework, and main approach of the top-cited manuscripts (> 100 citations; N = 14) using machine-learning approach in ecology and conservation across the globe.

Journal	Impact factor	Year	Authors	Main country	N citation	Primary aim	Machine learning method(s) or framework	Main approach
Ecography	6.802	2006	Elith et al.	AUS	6282	To provide effective guidance on how best to use this information in the context of numerous approaches for modelling distributions	Maximum entropy models (MAXENT and MAXENT-T) and boosted regression trees (BRT, also called stochastic gradient boosting)	Landscape ecology & species distribution
Ecological Modelling	3.512	2009	Kocev et al.	SVN	139	To applied various machine learning methods to the problem of predicting the condition or quality of the remnant indigenous vegetation across an extensive area of south-eastern Australia	Regression trees, Multi-target regression trees, Ensembles, Bagging, Random forests	Landscape ecology & species distribution
Ecological Informatics	4.498	2010	Kampichler et al.	MEX	120	To compare five machine-learning based classification techniques (classification trees, random forests, artificial neural networks, support vector machines, and automatically induced rulebased fuzzy models) in a biological conservation context	Techniques of classification trees, random forests, artificial neural networks, support vector machines, and automatically generated fuzzy classifiers	Bird ecology & conservation
PLoS ONE	3.752	2011	Pittman & Brown	USA	181	to: (1) Determine whether the influence of environmental predictors on species' distribution was scale-dependent; (2) evaluate the utility of environmental data from a single remote sensing device combined with metrics for surface morphology to predict and map fish species distributions across a complex coral reef ecosystem; (3) determine which omponents of remotely sensed seafloor structure contribute most to the species distribution models; (4) identify threshold effects where changes in environmental variables abruptly influence species occurrence; and (5) evaluate the performance of two different machine-learning modelling algorithms for spatial predictions of marine fish distributions	Boosted regression trees (BRT) and maximum entropy modelling (MaxEnt)	Fish ecology & distribution

Remote Sensing of Environment	13.85	2012	Dronova et al.	USA	139	Classify Poyang Lakewetland PFTs using OBIA and to determine which image segmentation scales and machine-learning algorithms optimize class discrimination	Six machine-learning algorith representing machine-learning principles: a probabilistic Bayes method (NaïveBayesSimple in Weka notation), a logistic regression method (SimpleLogistic in Weka), an artificial neural network algorithm (MultilayerPerceptron (MLP)), a support vector machine tool with polynomial kernel and complexity parameter value of 10 (SMO inWeka), a K-Nearest Neighbors (IBk inWeka) and a tree-based classifier (RandomForest).	Landscape ecology & species distribution
Journal of Biogeography	4.81	2012	Werneck et al.	USA	174	To investigate the historical distribution of the Cerrado across Quaternary climatic fluctuations and to generate historical stability maps to test: (1) whether the ‘historical climate’ stability hypothesis explains squamate reptile richness in the Cerrado; and (2) the hypothesis of Pleistocene connections between savannas located north and south of Amazonia	Maximum-entropy	Climate change
Methods in Ecology and Evolution	8.335	2012	Barbet-Massin et al.	FRA	1401	To conduct a comprehensive comparative analysis based on simple simulated species distributions to propose guidelines on how, where and how many pseudo-absences should be generated to build reliable species distribution models.	Boosted regression trees (BRT) and random forest (RF)	Landscape ecology & species distribution
Integrative Zoology	2.083	2013	Li & Wang	CHN	120	To applying various algorithms for species distribution modelling	Multivariate adaptive regression splines, Mixture discriminant analysis, Artificial neural networks, Generalized boosting models, Classification and regression tree, Random forest, Hierarchical modelling, Genetic algorithm for rule set production, and Maximum entropy	Landscape ecology & species distribution

PeerJ	3.061	2014	Stowell & Plumbley	GBT	185	To introduce a technique for feature learning from large volumes of bird sound recordings, inspired by techniques that have proven useful in other domains	Spectral features and feature learning	Bird ecology & conservation
PLoS ONE	3.752	2016	de Souza et al.	CAN	152	To assess the accuracy of satellite-based AIS (Automatic Identification System) and VMS (Vessel Monitoring System) in correctly identifying individual fishing events or 'sets' by comparing against expert-labeled data	Hidden Markov Model (HMM) and Data Mining (DM)	Fish ecology & distribution
Energy and Buildings	7.201	2018	Touzani et al.	USA	212	To propose an energy consumption baseline modeling method based on a gradient boosting machine wasproposed.	Decision trees and Gradient boosting machine	Green energy & energy economy
Ecology Letters	11.274	2018	McLaren et al.	USA	102	To disentangle anthropogenic and landscape-related factors affecting stopover density, and thereby assess whether artificial light at night might be affecting selection of stopover habitat, we estimated responses in seasonal-mean reflectivity to geographic, land cover and anthropogenic predictors using additive regression models fit by gradient boosting, a machine-learning technique	Additive regression models fit by gradient boosting	Bird ecology & conservation

Proceedings of the National Academy of Sciences	12.799	2018	Norouzzadeh et al.	USA	493	Test how well deep learning can automate information extraction from camera-trap images	Deep learning architectures: AlexNet, NiN, VGG, GoogLeNet, ResNet	Wildlife ecology
Methods in Ecology and Evolution	8.335	2019	Christin et al.	CAN	187	To review existing implementations and show that deep learning has been used successfully to identify species, classify animal behaviour and estimate biodiversity in large datasets like camera-trap images, audio recordings and videos	Tensorflow, PyTorch, Keras, Microsoft Cognitive Toolkit (CNTK), Deeplearning4J, MATLAB + Deep Learning Toolbox, Apache MXNET, PlaidML	Wildlife ecology